ANTECEDENTS OF ESG-RELATED CORPORATE MISCONDUCT: THEORETICAL CONSIDERATIONS AND MACHINE LEARNING APPLICATIONS

Inauguraldissertation
zur
Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

2023

vorgelegt von
Lars Gemmer
aus
Duisburg

Referent: Prof. Dr. Marc Fischer

Korreferent: Prof. Dr. Werner Reinartz

Tag der Promotion: 29.08.2023

DEDICATION

To my father Dirk and my mother Barbara –

Your trust and support are the antecedents of this work.

CONTENTS

LIST OF FIGURES

List of Tables

Synopsis

Paper 1: When Will We Get in Trouble (Again)? Predicting the Occurrence of Severe Negative Publicity for Firms

PAPER 2: WHEN IS COMPETITION REALLY HEALTHY? ANALYZING THE IMPACT OF THE FIRM'S COMPETITIVE SITUATION ON THE DISCLOSURE OF UNETHICAL FIRM BEHAVIOR

PAPER 3: MACHINE LEARNING IN MARKETING – A REVIEW OF RECURRING PROBLEMS AND HOW TO SOLVE THEM

LIST OF ABBREVIATION

AIC	Akaike information criterion
ANN	Artificial neural network
AR	Abnormal stock returns
AUC	Area under the curve
Bal. accuracy	Balanced accuracy
BIC	Bayesian information criterion
DMEC	Direct minimum expected cost
Eret	Expected stock market returns
ESG	Environmental, social, and governance
FN	False negatives
FNR	False negative rate
FP	False positives
FPR	False positive rate
FTC	Federal Trade Commission
HNB	Hurdle negative binomial
LIME	Local interpretable model-agnostic explanations
m	Million
MC	Average market capitalization
ML	Machine learning
MSCI	MSCI world index
n.s.	Not significant
NAICS	North American industry classification system
PCC	Percentage correctly classified
PR	Precision recall
R&D	Research and development
ROC	Receiver operating characteristic
SD	Standard deviation
SE	Standard error
SG&A	Selling, general and administrative
SHAP	Shapley additive explanation
SIC	Standard industrial classification
SR	Stock market returns

TN	True negatives
TNIC	Text-based network industry classifications
TNR	True negative rate
TP	True positives
TPR	True positive rate
TR	Threshold
UPB	Unethical pro-organizational behavior
VIF	Variance inflation factor
XGB	XGBoost
ZTNB	Zero-truncated negative binomial

SYNOPSIS

1 Overview

The core objective of this cumulative dissertation is to generate new insights in the occurrence and prediction of unethical firm behavior disclosure. The first two papers investigate predictors and antecedents of (severe) unethical firm behavior disclosure. The third paper addresses frequently occurring methodological issues when applying machine learning approaches within marketing research. Hence, the three papers of this dissertation contribute to two recent topics within the field of marketing: First, marketing research has already focused intensively on the consequences of *corporate misconduct* and the accompanying media coverage (e.g., Stäbler and Fischer 2020). Meanwhile, the prediction [Paper 1] and the process of occurrence [Paper 2] of such threatening events have been examined only sporadically so far. Second, companies and researchers are increasingly implementing *machine learning* as a methodology to solve marketing-specific tasks (Ma and Sun 2020). In this context, the users of machine learning methods often face methodological challenges, for which this dissertation reviews possible solutions [Paper 3]. Table 1 presents an overview of the three articles as well as the involved authors and the publication-status information.

Table 1: Overview of Presented Dissertation Projects

Paper	Title	Author(s)	Status
1	When Will We Get in Trouble (Again)? Predicting the Occurrence of Severe Negative Publicity for Firms	Lars Gemmer, Samuel Stäbler, and Marc Fischer	Prepared to submit to: *Management Science*
2	When is Competition Really Healthy? Analyzing the Impact of the Firm's Competitive Situation on the Disclosure of Unethical Firm Behavior	Lars Gemmer, Alexander Edeling, and Marc Fischer	Prepared to submit to: *Journal of Marketing*
3	Machine Learning in Marketing – A Review of Recurring Problems and How to Solve Them	Lars Gemmer	Prepared to submit to: *International Journal of Research in Marketing*

Notes: Lars Gemmer made a major and substantial contribution to all three papers of this dissertation, including idea development, development of the empirical designs, data collection, data analyses, and writing up the manuscript.

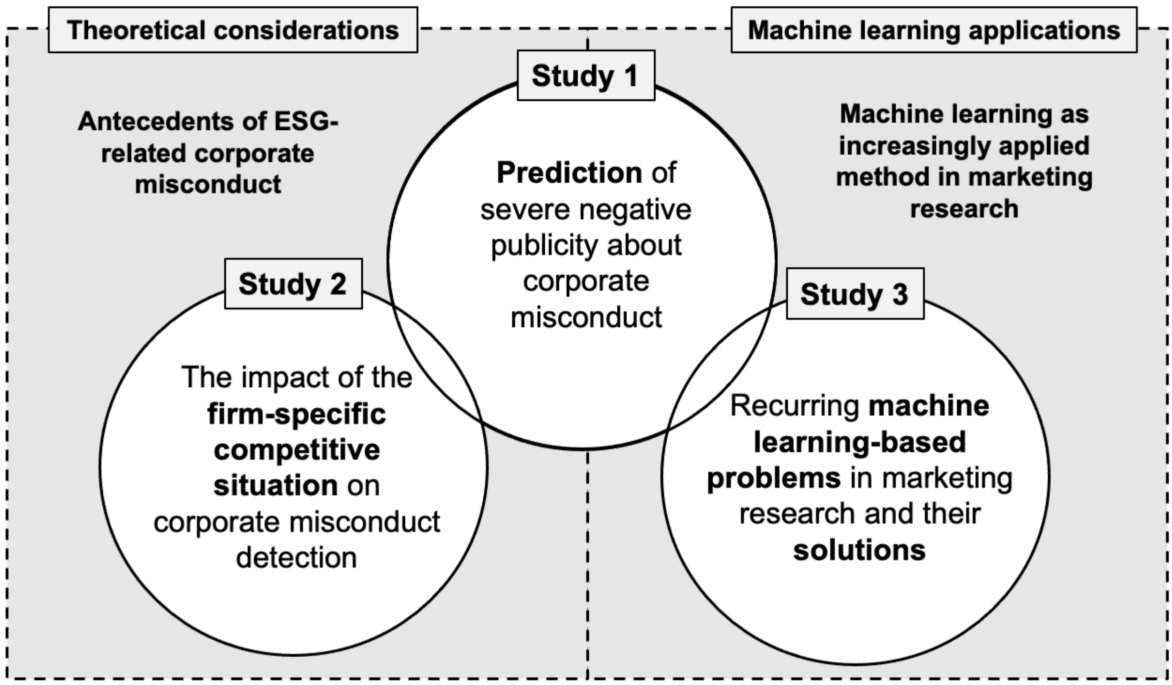
2 Introduction

Marketing research is characterized by two ongoing trends reflected in the three dissertation papers. First, marketing researchers increasingly address the subject of unethical misconduct committed by companies and resulting corporate crises (Dinner, Kushwaha, and Steenkamp 2019; Kang, Germann, and Grewal 2016). The importance of this topic has risen notably in recent years. The disclosure of corporate misconduct has the potential to trigger a serious corporate crisis with major negative consequences on key performance metrics of firms including sales figures (Van Heerde, Helsen, and Dekimpe 2007; Zhao, Zhao, and Helsen 2011) and firm value (Flammer 2013; Kang, Germann, and Grewal 2016).

The second trend in marketing research refers to the methodology of machine learning (Ma and Sun 2020; Ngai and Wu 2022). Machine learning is a process of mathematical algorithms learning patterns in data observations and then making predictions or classifications (Kirasich, Smith, and Sadler 2018). Machine learning-based methods have been established in the scientific context for decades (Rumelhart, Hinton, and Williams 1986; Quinlan 1986). The currently observed trend towards the increasing application of machine learning methods in research and in analytics departments of companies is driven by a better access to data, more computing power, and the continuous development of algorithms (Lantz 2019). An increased interest in machine learning applications is also observed in marketing science, though "the use of machine learning methods in marketing is still at an early stage" (Ma and Sun 2020, p. 482). As Ma and Sun (2020) explain, machine learning affects practical decision-making processes across the whole marketing mix.

The classification of the studies in Figure 1 demonstrates the interrelationships of the respective dissertation papers and illustrates how they connect to the two described thematic trends. Both, study 1 and study 2 of this dissertation cover the substantive topic of ESG-related unethical firm behavior and its disclosure. In study 1, machine learning algorithms are used to predict the future occurrence of severe threatening news coverage of corporate misconduct.

Figure 1: Classification of Presented Dissertation Papers and their Interdependencies



Study 2 identifies relationships between the specific competitive situation of a company within its industry and unethical firm behavior disclosure. Both studies contribute to the theory on unethical firm behavior and resulting corporate crises as they shed light on the so far poorly investigated antecedents of unethical firm behavior.

Machine learning plays a crucial role for the empirical analysis of the data in study 1. In this context, the paper exemplifies that marketing researchers face recurring challenges when applying machine learning (e.g., the prediction of rarely occurring events, the selection of the "right" performance metrics, or the incorporation of different misclassification costs). Paper 3 addresses these machine learning-based issues for marketing researchers and presents possible solutions by reviewing the computer science literature.

The first paper, titled "When Will We Get in Trouble (Again)? Predicting the Occurrence of Severe Negative Publicity for Firms" (co-authored by Lars Gemmer, Samuel Stäbler, and

Marc Fischer) introduces a cost-sensitive machine learning-based model that predicts when the next threatening information about a firm's unethical firm behavior is likely to be disclosed. Based on an analysis of 3,271 companies and their negative news coverage, this study shows how the forecast can be used in the short term to analyze whether severe negative publicity should be expected in the upcoming week or not. The model correctly identifies 8.6 out of 10 severe negative publicity events and achieves a balanced accuracy of about 85.5%. In addition, the study shows how this week-to-week forecast can be used by managers with longer lead times to address upcoming threats. The prediction model enables companies to identify upcoming threats that require appropriate attention.

The second paper, titled "When is Competition Really Healthy? Analyzing the Impact of the Firm's Competitive Situation on the Disclosure of Unethical Firm Behavior" is co-authored by Lars Gemmer, Alexander Edeling, and Marc Fischer. This study introduces new variables that measure the competitive situation of firms within one industry and identifies which competitive constellations intensify the pressure and force managers to act unethically. The findings from the analysis of more than 2,777 companies and 68,992 disclosed unethical firm behavior incidents contribute to the ongoing debate regarding the role of market share and market share-based competition for firms' behavior and financial performance. It shows that unethical firm behavior disclosure reduces the market share, and thus, weakens the competitive situation of firms. Hence, an extreme focus on market share maximization (by unethical means) is potentially counterproductive for succeeding in competition.

Paper 3, titled "Machine Learning in Marketing – A Review of Recurring Problems and How to Solve Them" (by Lars Gemmer), identifies recurring challenges that marketing researchers frequently face when applying machine learning approaches. Based on a review of representative marketing-related studies applying machine learning algorithms, this study overviews solutions from computer science literature for each of the identified issues. The results of this study offer practitioners as well as researchers fruitful ways to apply machine learning more efficiently.

The following section summarizes relevance, research aims, main results as well as implications of the three dissertation projects.

3 Summary of Dissertation Projects

3.1 Paper 1: When Will We Get in Trouble (Again)? Predicting the Occurrence of Severe Negative Publicity for Firms

Companies are increasingly confronted with negative publicity covered across a wide range of channels from low-reach local blogs to high-reach international newspapers. Such a negative publication may trigger a severe corporate crisis with negative consequences for the firm. Threatening news articles cover many different topics like malfunctioning products, environmental pollution, corruption, poor working conditions as well as issues of diversity and inclusion.

In line with this high practical relevance, the related area of corporate crisis research has grown to a major field over the last decades and studied the consequences of crisis events as well as important mediators and moderators of this process (Borah and Tellis 2016; Cleeren, Dekimpe, and Helsen 2008). For example, existing research focuses on the analysis of the consequences of firm crises on various outcome variables such as firm value (Flammer 2013; Kang, Germann, and Grewal 2016), sales (Van Heerde, Helsen, and Dekimpe 2007; Zhao, Zhao, and Helsen 2011), perceived quality (Gijsenberg, Van Heerde, and Verhoef 2015), and marketing effectiveness (Liu and Shankar 2015). In addition, numerous studies have identified moderators (Cleeren, Van Heerde, and Dekimpe 2013; Gao et al. 2015; Rubel, Naik, and Srinivasan 2011) and mediators (Berger, Sorenson, and Rasmussen 2010; Stäbler and Fischer 2020) of these relationships. Little knowledge, however, exists about the antecedences and the predictors of important crisis events.

This study fills this research gap by answering the research question of whether it is possible to predict the occurrence of severe negative publicity at a reasonable level of accuracy by using only available structured data. In addition, the study deals with the question how to control for the different sizes of the misclassification costs (i.e., costs due to falsely predicting a severe negative publicity and costs due to missing a severe negative publicity).

The introduced prediction model is calibrated on a unique dataset of 11 years of weekly data from 3,271 companies and their negative news coverage in more than 100,000 media resources and 23 languages. Even though the analyzed data set contains 14,440 severe negative publicity events, these are still very rare events at a frequency of less than 1%.

Results show that, within the setting of the study, XGBoost is the most promising machine learning approach for this extremely imbalanced prediction task as it outperforms all other tested methods. Independent of a specific classification threshold, the XGBoost algorithm performs 13.79 times better than a random classifier. Incorporating the different misclassification costs of potential users of the forecasting model by applying the cost optimal threshold for an exemplary cost ratio of 2:1, the XGBoost model identifies 8.6 out of 10 severe negative publicities achieving a balanced accuracy of 85.50%.

The study further reveals that the history of negative publicity (e.g., the time since the last negative publicity event occurred) is the most relevant information when predicting future event occurrence. However, marketing-related variables like the advertising spending or the customer relationship equity also matter for the prediction model.

The results have important implications for managers and for investors. The study shows that trading stocks according to the recommendations of the prediction model reduces the costs arising from unanticipated negative news coverage by 42.57%. Furthermore, the study reveals substantive insights into the drivers of severe negative publicity occurrence probability. Such findings on the variable effect size as well as the potential direction of the effect of the individual variables are of high interest to shareholders and managers of the companies concerned.

Next-week forecasts are hardly actionable for managers because they do not give managers sufficient time to act in response to an upcoming threat. Therefore, the study develops an early warning tool for managers that makes use of the time-series of the weekly forecasts to determine the duration in which the next severe negative publicity is likely to occur. In periods of increased risk, managers may reduce the probability of severe negative publicity. For example, corporate risk management may strive to prevent decision makers' unethical behavior through communication and monitoring activities.

3.2 *Paper 2: When is Competition Really Healthy? Analyzing the Impact of the Firm's Competitive Situation on the Disclosure of Unethical Firm Behavior*

Shortly after Volkswagen's market share became larger than Toyota's (Financial Review 2015), the Volkswagen emissions scandal became public (Siano et al. 2017) leading to the company's lowest market share since the financial crisis (Financial Times 2016). This example indicates the potential relationship between an intense market share-based competition orientation (i.e., race for a better position within the industry), disclosed unethical firm behavior (i.e., emissions scandal), and a resulting setback with regard to the competitive position (i.e., reduction of market share).

For many managers, one of the primary goals is to increase market share (Bhattacharya, Morgan, and Rego 2022; Farris et al. 2010). However, numerous studies question the pursuit of higher market share as a panacea. For instance, a higher market share may result in a decrease of customer satisfaction (Rego, Morgan, and Fornell 2013). Results from a meta-analysis by Edeling and Himme (2018, p. 4) question a strategy of a firm that "focuses too strongly on retaining and increasing market share as a business objective". Armstrong and Collopy (1996) suggest that firms should ignore their competitors and focus on profit maximization. This ongoing discussion about the role of market share for firms raises the question to what extent the effects of the Volkswagen example are generalizable.

This study fills this knowledge gap by answering the following research questions: (1) Are there competitive situations of firms within one industry that influence the pressure on decision-makers in firms and trigger unethical actions and their disclosure? (2) If yes, which specific situations of competitive performance exist and how do they drive unethical firm behavior disclosure? (3) Can striving for an improvement of the competitive situation actually turn into a disadvantage resulting in a worse competitive position?

In order to answer these research questions, the study introduces new market share-related variables that reflect the competitive situation of a firm within the industry and analyzes the unethical behavior disclosure of 2,777 international companies from 79 different industries in the time window from 2007 to 2017.

The results show that information regarding three different factors are relevant to explain the unethical behavior disclosure. These include the global ranking-related position of a company within an industry, the market-share based proximity of direct competitors, and dynamic changes of the global position and the proximities. In addition, the study finds that one additional unethical firm behavior disclosure per year significantly decreases the market share by .82% on average, and thus, leads to an actual setback in the battle for a better competitive position.

These new empirical generalizations imply, for example, that firms should include intra-industry competitive constellations when implementing internal monitoring measures that warn of corporate misconduct (disclosure). Investors, especially those that take ESG criteria into consideration (McGee 2022), may use this study's insights to optimize their investment decisions and to adapt their investment portfolio proactively by analyzing the competitive situations of focal firms.

Furthermore, managers' attempts to strive for a better competitive situation by unethical means are not necessarily crowned with success. It can even be counterproductive and counteract an improvement of the competitive position. Based on these findings, companies may rethink the incentives for their decision-makers and define other more sustainable goals such as increasing brand equity or improving customer relationships, instead of focusing on the potentially detrimental market share maximization.

3.3 Paper 3: Machine Learning in Marketing – A Review of Recurring Problems and How to Solve Them

The application opportunities of machine learning methodologies are diverse and the existing algorithms provide solutions across various research disciplines. For example, machine learning is used for the prediction of earthquakes (Asim et al. 2018; Mignan and Broccardo 2020), medical diagnosis (Kononenko 2001), crime prediction (Jordan and Mitchell 2015; Wang, Gerber, and Brown 2012) as well as bankruptcy prediction (Devi and Radhika 2018; Wang 2017).

Machine learning is also of increasing relevance for marketing researchers (e.g., Lemmens and Croux 2006; Lemmens and Gupta 2020). Several existing review studies underpin the essential role as well as the complexity of machine learning within the field of marketing (Ma and Sun 2020; Ngai and Wu 2022). Thereby, these reviews focus on the analysis of the research problem or the type of applied algorithms.

The challenges that marketing researchers face when applying machine learning methods are often comparable and of recurring nature. For example, highly skewed (or imbalanced) data that involve only a few observations within one of the classes are quite common (e.g., churn prediction and business failure prediction). For these recurring hurdles, there exists a multitude of potential solutions in computer science, which researchers and practitioners can only hardly overview.

The study aims to present potential solutions for recurring machine learning-based problems typically occurring in marketing tasks, such as imbalanced target variables, selection of variables, cost-sensitive learning, tuning of hyperparameters, interpretable machine learning as well as the selection of the performance metrics. Based on an initial identification of frequently occurring challenges of machine learning applications in marketing, this study overviews various approaches to solve these obstacles.

The study presents representative research dealing with marketing-related classification tasks and applying machine learning methods. Based on the general learning process, the recurring problems are grouped into three different categories: First, *data-driven challenges* originate from characteristics of the analyzed data and the variables. Second, *abstraction-based challenges* mainly refer to challenges within the model creation process or the training of the machine learning algorithms. Third, *generalization-based challenges* include challenges referring to actionable implications or performance evaluation.

The results reveal six major hurdles for marketing researchers applying machine learning for classifications. Within data-driven challenges researchers often need to handle highly imbalanced data sets as well as to identify the optimal set of variables to build the models. Abstraction-based challenges include cost-sensitive machine learning and the tuning of hyperparameters. Generalization-based challenges deal with the selection of the right metric to evaluate the classification performance as well as interpretable machine learning. For all of these six problems, the study synthesizes computer science literature that offers potential methodological solutions.

The insights provided by this review offer guidance for both, researchers and practitioners within the field of marketing and beyond. From a methodological point of view, the study overviews solutions for researchers to overcome recurring hurdles in machine learning. In this way, the review facilitates scientific working by providing concrete directions. Furthermore, the study combines machine learning-based problems from the marketing field with methodological solution approaches from computer science.

The findings enable companies to apply machine learning more efficiently. In this way, the analytical foundation of marketing mix decisions can be improved, which creates additional monetary value for companies (e.g., through cost-sensitive machine learning or the selection of the right performance metrics).

REFERENCES SYNOPSIS

Armstrong, J. Scott and Fred Collopy (1996), “Competitor Orientation: Effects of Objectives and Information on Managerial Decisions and Profitability,” *Journal of Marketing Research*, 33 (2), 188–99.

Asim Khawaja M., Adnan Idris, Talat Iqbal, and Francisco Martínez-Álvarez (2018), “Earthquake Prediction Model Using Support Vector Regressor and Hybrid Neural Networks,” *PLoS ONE,* 13 (7), https://doi.org/10.1371/journal.pone.0199004.

Berger, Jonah, Alen T. Sorenson, and Scott J. Rasmussen (2010), ”Positive Effects of Negative Publicity: When Negative Reviews Increase Sales,” *Marketing Science*, 29 (5), 815–27.

Bhattacharya, Abhi, Neil A. Morgan, and Lopo L. Rego (2022), “Examining Why and When Market Share Drives Profit,” *Journal of Marketing*, 86 (4), 73–94.

Borah, Abhishek and Gerard J. Tellis (2016), “Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands?,” *Journal of Marketing Research*, 53 (2), 143–60.

Cleeren, Kathleen, Marnik G. Dekimpe, and Kristiaan Helsen (2008), “Weathering Product-Harm Crises,” *Journal of the Academy of Marketing Science*, 36 (2), 262–70.

Cleeren, Kathleen, Harald J. van Heerde, and Marnik G. Dekimpe (2013), “Rising from the Ashes: How Brands and Categories Can Overcome Product-Harm Crises,” *Journal of Marketing*, 77 (2), 58–77.

Devi, S. Sarojini and Y. Radhika (2018), “A Survey on m´Machine Learning and Statistical Techniques in Bankruptcy Prediction,” *International Journal of Machine Learning and Computing*, 8 (2), 133–39.

Dinner, Isaac M., Tarun Kushwaha, and Jan-Benedict E. M. Steenkamp (2019), “Psychic Distance and Performance of MNCs During Marketing Crises,” *Journal of International Business Studies*, 50 (3), 339–364.

Edeling, Alexander and Alexander Himme (2018), “When Does Market Share Matter? New Empirical Generalizations from a Meta-Analysis of the Market Share–Performance Relationship,” *Journal of Marketing*, 82 (3), 1–24.

Farris, Paul W., Neil T. Bendle, Phillip E. Pfeifer, and David J. Reibstein (2010), *Marketing Metrics*, 2nd ed. Pearson Education.

Financial Review (2015), “VW Passes Toyota to Take Global Sales Lead in First Half of 2015” by Craig Trudell and Masatsugu Horie, (accessed June 15th, 2023), [available at https://www.afr.com/world/vw-passes-toyota-to-take-global-sales-lead-in-first-half-of-2015-20150728-gim6sd].

Financial Times (2016), “Volkswagen’s Market Share Falls After Scandal,” by Peter Campbell, London (accessed June 15th, 2023), [available at https://www.ft.com/content/35575f80-4a75-11e6-b387-64ab0a67014c].

Flammer, Caroline (2013), “Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors,” *Academy of Management Journal*, 56 (3), 758–81.

Gao, Haibing, Jinhong Xie, Qi Wang, and Kenneth C. Wilbur (2015), “Should Ad Spending Increase or Decrease Before a Recall Announcement? The Marketing–Finance Interface in Product-Harm Crisis Management,” *Journal of Marketing*, 79 (5), 80–99.

Gijsenberg, Maarten J., Harald J. van Heerde, and Peter C. Verhoef (2015), "Losses Loom Longer Than Gains: Modeling the Impact of Service Crises on Perceived Service Quality Over Time," *Journal of Marketing Research*, 52 (5), 642–56.

Jordan, Michael I. and Tom M. Mitchell (2015), "Machine Learning: Trends, Perspectives, and Prospects," *Science*, 349.6245, 255−60.

Kang, Charles, Frank Germann, and Rajdeep Grewal (2016), "Washing Away Your Sins? Corporate Social Responsibility, Corporate Social Irresponsibility, and Firm Performance," *Journal of Marketing*, 80 (2), 59–79.

Kirasich, Kaitlin, Trace Smith, and Bivin Sadler (2018), "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets," *SMU Data Science Review*, 1 (3), article 9.

Kononenko, Igor (2001), "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective," *Artificial Intelligence in Medicine*, 23 (1), 89–109.

Lantz, Brett (2019), *Machine Learning with R*, 3rd ed. Packt Publishing.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (2), 276–86.

Lemmens, Aurélie and Sunil Gupta (2020), "Managing Churn to Maximize Profits," *Marketing Science*, 2020, 39 (5), 956–73.

Liu, Yan and Venkatesh Shankar (2015), "The Dynamic Impact of Product-Harm Crises on Brand Preference and Advertising Effectiveness: An Empirical Analysis of the Automobile Industry," *Management Science*, 61 (10), 2514–35.

Ma, Liye and Baohong Sun (2020), "Machine Learning and AI in Marketing–Connecting Computing Power to Human Insights," *International Journal of Research in Marketing*, 37 (3), 481–504.

McGee, Suzanne (2022), "ESG Investors' Dilemma: Think Positive or Negative? It's a Question of Shunning All Companies in a Sector or Taking a More Nuanced Approach," *Wall Street Journal*, November 3, R. 5.

Mignan, Arnaud and Marco Broccardo (2020), "Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on their Limitations," *Seismological Research Letters*, 91 (4), 2330–42.

Ngai, Eric W. and Yuanyuan Wu (2022) "Machine Learning in Marketing: A Literature Review, Conceptual Framework, and Research Agenda," *Journal of Business Research*, 145, 35–48.

Quinlan, J. Ross (1986), "Induction of Decision Trees," *Machine learning*, 1, 81−106.

Rego, Lopo L., Neil A. Morgan, and Claes Fornell (2013), "Reexamining the Market Share–Customer Satisfaction Relationship," *Journal of Marketing*, 77 (5), 1–20.

Rubel, Oliver, Prasad A. Naik, and Shuba Srinivasan (2011), "Optimal Advertising when Envisioning a Product-Harm Crisis," *Marketing Science*, 30 (6), 1048–65.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986), "Learning Representations by Back-Propagating Errors," *Nature*, 323.6088, 533–36.

Siano, Alfonso, Agostino Vollero, Francesca Conte, and Sara Amabile (2017), "'More Than Words'": Expanding the Taxonomy of Greenwashing After the Volkswagen Scandal," *Journal of Business Research*, 71, 27–37.

Stäbler, Samuel and Marc Fischer (2020), “When Does Corporate Social Irresponsibility Become News? Evidence from More Than 1,000 Brand Transgressions Across Five Countries,” *Journal of Marketing*, 84 (3), 46–67.

Van Heerde, Harald J., Kristiaan Helsen, and Marnik G. Dekimpe (2007), “The Impact of a Product-Harm Crisis on Marketing Effectiveness,” *Marketing Science*, 26 (2), 230–45.

Wang, Nanxi (2017), “Bankruptcy Prediction Using Machine Learning,” *Journal of Mathematical Finance*, 7 (04), 908–18.

Wang, Xiaofeng, Matthew S. Gerber, and Donald E. Brown (2012), “Automatic Crime Prediction Using Events Extracted from Twitter Posts,” *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 231–38.

Zhao, Yi, Ying Zhao, and Kristiaan Helsen (2011), “Consumer Learning in a Turbulent Market Environment: Modeling Consumer Choice Dynamics After a Product-Harm Crisis,” *Journal of Marketing Research*, 48 (2), 255–67.

PAPER 1: WHEN WILL WE GET IN TROUBLE (AGAIN)? PREDICTING THE OCCURRENCE OF SEVERE NEGATIVE PUBLICITY FOR FIRMS

Authors: Lars Gemmer, Samuel Stäbler, and Marc Fischer

ABSTRACT

Companies are increasingly confronted with negative publicity related to, for example, malfunctioning products, pollution, corruption, diversity, inclusion, and other controversies. Indeed, the average yearly growth rate of negative news has reached 21.8%. Such negative press can trigger a corporate crisis with severe adverse consequences. Whereas most prior research has focused on the potential consequences of such trigger events, this study focuses on predicting these events on a weekly basis. The authors introduce a cost-sensitive machine learning-based prediction model calibrated on a unique data set of 11 years of weekly data from 3,271 companies and their negative news coverage. The model produces a time-varying trigger event forecast for each company. In the short term, this forecast can be used to analyze whether a trigger event should be expected in the upcoming week. Applying the cost-optimal classification threshold, the model identifies 8.6 of 10 trigger events achieving a balanced accuracy of approximately 85.5%. In addition, the study shows how managers can use this week-to-week forecast when they need longer lead time to address threats. The prediction model enables managers to identify trends that are indicative of upcoming threats and require appropriate attention.

Keywords: Cost-sensitive prediction, corporate crises, negative publicity, unethical firm behavior

1 Introduction

Companies are increasingly confronted with negative publicity generated from a wide range of channels, from low-reach local blogs to high-reach international newspapers. Such negative press can trigger a crisis with potentially severe financial and other adverse consequences for firms. Threatening news articles cover a variety of topics (e.g., malfunctioning products; environmental pollution; corruption; poor working conditions; issues of diversity, inclusion, and equity).

In line with this emerging trend, the related area of corporate crisis research has grown into a major field over the past few decades, investigating the consequences of crisis events as well as important mediators and moderators of this process (Borah and Tellis 2016; Cleeren, Van Heerde, and Dekimpe 2013). For example, existing research focuses on analyzing the consequences of firm crises on outcome variables such as firm value (Kang, Germann, and Grewal 2016), sales (Van Heerde, Helsen, and Dekimpe 2007; Zhao, Zhao, and Helsen 2011), perceived quality (Gijsenberg, Van Heerde, and Verhoef 2015), and marketing effectiveness (Liu and Shankar 2015). In addition, numerous studies have identified moderators (Cleeren, Van Heerde, and Dekimpe 2013; Gao et al. 2015; Rubel, Naik, and Srinivasan 2011) and mediators (Berger, Sorensen, and Rasmussen 2010; Stäbler and Fischer 2020) of these relationships.

Little knowledge, however, exists about the antecedents and the predictors of important crisis events. An exception is Rubel, Naik, and Srinivasan's (2011) study, which shows that firms potentially change their advertising expenditures strategically in anticipation of changes in economic performance. Another work is Campbell and Shang's (2022) recent accounting study, which explores employee comments in social media to evaluate the risk that a company will be prosecuted by U.S. regulatory bodies in the future due to misconduct. While prosecution misconduct cases in the United States are important events for firms, they represent a specific and rather small group of threatening negative events, and they typically do not receive media attention. In fact, negative publicity includes a much wider range of topics, including actions that do not break any laws and simply violate societal norms, as well as fake news stories. The scope of this work is all potentially relevant issues of corporate misbehavior, with or without cause, covered in thousands of media sources across the world. Specifically, the purpose of this study is to predict such negative news that may have severe consequences for the firm.

A corporate crisis is often associated with a date when information is released that triggers the evolution of a crisis, which we refer to as a "trigger event". A well-known example of such an event with significant negative consequences is the widely publicized complaint about poor customer service at United Airlines (e.g., Li, Juric, and Brodie 2017). A dissatisfied customer produced a music video describing United's mishandling of his guitar and the company's bad customer service, and he posted the video on social media. The video went viral, and as a consequence, the company's reputation and stock price were significantly damaged.

Oftentimes, a trigger event results in further news events presenting new pieces of information that are not necessarily related to the prior issue but have the capacity to significantly harm the firm. These news events could mark a new crisis or be part of an ongoing one. For our study, the conceptual separation of crises and thus the sequence of events is not of interest as long as the newly revealed information is of relevance (i.e., the negative news event involves a negative feedback effect from the stock market).

Such trigger events represent the focal event to be predicted in this study. Note that although we focus only on news that offers new information and does not repeat prior news, not every piece of negative publicity has financial consequences for the firm. In fact, our data show that only about one of three negative news items ultimately adversely affects the firm in form of a negative effect on stock return. As we show subsequently, however, the other two-thirds of harmless negative publicity still bear important predictive content.

We introduce a prediction model calibrated on a unique data set of 11 years of weekly data from 3,271 companies and their negative news coverage in more than 100,000 media resources and 23 languages. Although we identified 14,440 of such trigger events, we note that they are still very rare events within this large set of companies and long time period, occurring at a frequency of less than 1%. This situation creates special challenges for prediction comparable to the prediction of earthquakes and crimes.

Our prediction approach uses only *structured* information in a standardized format that is easily accessible before the trigger event. The accessibility of the data is important so that various stakeholders can make use of this prediction approach. The forecasting model involves a machine learning (ML) algorithm that addresses the challenges of highly imbalanced outcomes. Note that we do not introduce a new machine learning method; rather, we make use of four already existing and well-established methods, including logistic regression, to build our forecasting model.

The prediction involves two types of errors – missing a trigger event and falsely predicting one – that are likely to produce different perceived costs for the user of the forecast. Incorporating these varying misclassification costs into our prediction makes the forecasting approach cost-sensitive. Based on the ratio of the costs, we obtain a user-specific classification threshold that maximizes performance measured by the balanced accuracy. In our application, the prediction model correctly identifies 8.6 of 10 trigger events and achieves a balanced accuracy of 85.5%.

The main outcome of the proposed model produces a *probability* that a firm will face a trigger event in the upcoming week. We use this information in two ways. Stakeholders like investors, for example, are interested in being alerted of an event with negative stock value consequences at very short notice so they can flexibly adjust their investments. Firm managers, in contrast, need a longer lead time to prepare for and counter potential threats to the company.

To address their needs, we show how managers can use the short-term forecast of the ML-based model as an alert index that warns them in advance of a trigger event in the longer future. This prewarn system monitors the development of the predicted weekly trigger event probability over the last quarter and alerts management of a change in the long-term horizon until the next trigger event occurs at a given probability. A significant shortening of the horizon arises if the probability exceeds one or more thresholds that are specific to the industry, firm size, and region. With this long-term perspective, managers can transparently determine how likely a trigger event is to occur in a given future time period that requires appropriate attention and preemptive action.

The reminder of this paper is organized as follows. We first present the data, as the characteristics of these data involve specific challenges that drive the subsequent development of our model. Next, we conceptually develop the prediction model and describe the applied methodology. In the subsequent sections, we present the prediction results and explain the managerial value of our study. We conclude with a discussion.

2 Data

The prediction of trigger events on a weekly basis is a challenging process because these events occur extremely infrequently. The aim of this section is to introduce the data and their characteristics. The specific characteristics of the data are relevant for the subsequent theoretical derivation of individual predictor variables and the identification of an appropriate methodological approach (e.g., a powerful machine learning algorithm).

2.1 Definition of Trigger Events

We define the trigger event period as the calendar week in which the first report on the threatening news is published in a news outlet. A negative news event is a result of two conditions. First, a cause or reason for the negative publicity must exist, which can be based on evidence or rumor. Note that we also consider fake news. Stäbler and Fischer (2020) show that only about 43% of negative news reports are evidence based. Even if the event is of speculative nature or fake news, it still could represent threatening negative news for the company with potentially severe consequences. For example, the unfounded rumor that McDonalds processed worms in its products led to a local drop in sales of 30% (Tybout, Calder, and Sternthal 1981). Thus, these types of events also need to be considered in the prediction. Second, media and other public sources and stakeholders must become aware of an issue and evaluate it as news-worthy. Unknown and unselected issues generate no negative publicity and are not relevant for our prediction task.

2.2 Population of Negative News Events

RepRisk, a global intelligence firm, provided the data on negative news coverage of companies. Prior studies have used the RepRisk database, and it is well established in business research (e.g., Dinner, Kushwaha, and Steenkamp 2019). The company screens a broad range of over 100,000 media, stakeholders, and other third-party sources in 23 languages (reaching more than 95% of the world's gross domestic product based on the official language indicated per country) to identify negative news incidents (RepRisk 2023). The media screened by RepRisk include print and online media (including local, regional, national, and international), communications from nongovernmental organizations, government bodies, regulators, and think tanks as well as newsletters and social media including Twitter, blogs, and other online sources.

RepRisk identifies the very first news report on all potential issues that threaten corporate reputation. These news reports cover more than 70 issues and topics, which relate to not only unethical firm behavior but also product failures and controversial topics (e.g., child labor, fraud, poor employment conditions, waste issues, supply chain issues, water scarcity, pornography, privacy issues; for a full list of covered topics, see Appendix A).

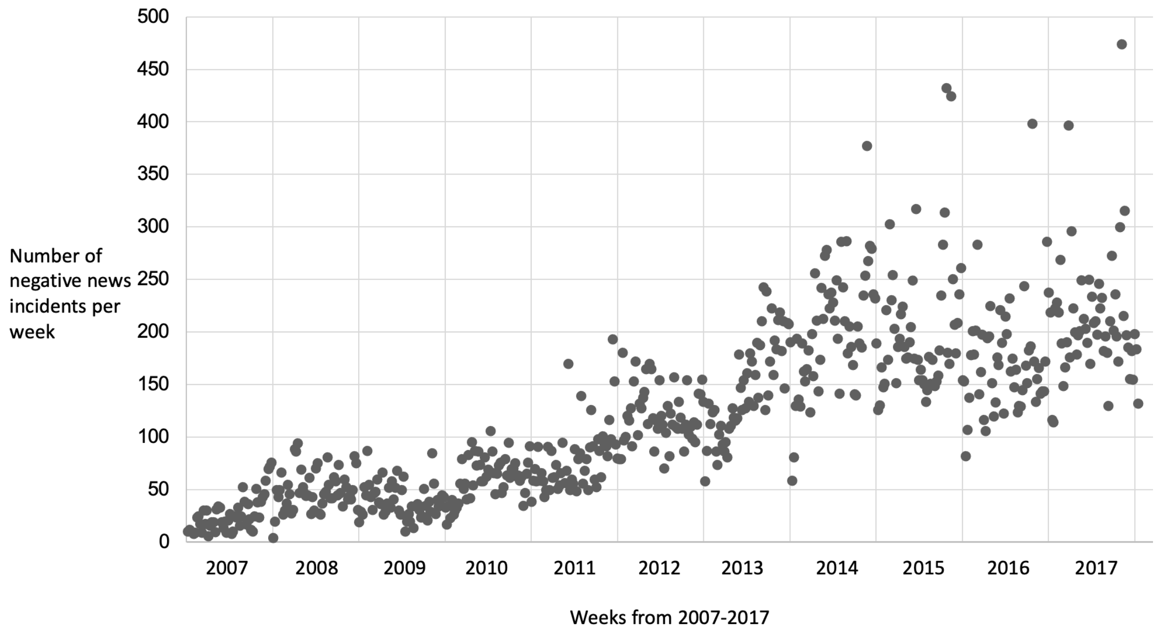
Importantly, RepRisk does not restrict itself to a pre-defined sample of firms. Specifically, it does not search for specific companies in the media but rather identifies news items that cover negative issues and topics and then includes all companies associated with these topics in its database. Consequently, the data set contains companies from all over the world, and any existing company has the potential to be included in the RepRisk data set as soon as unfavorable media coverage occurs about it. RepRisk's data include more than 205,000 companies from all industries and countries: 33% of the firms are from Asia, 26% from Europe, 24% from North America, and 17% from other areas, and 7% are publicly listed. The news events are accessible via public sources; thus, they are replicable and freely accessible, and companies can easily monitor them.

Figure 1 graphically demonstrates the practical relevance of negative publications about firms. The graph shows the sharp increase in negative publicity per week for 3,271 companies analyzed in this study in the time period 2007–2017 (574 weeks): only about 26 negative news events per week occurred on average in 2007, but the number increased to an average of 214 in 2017.

We aggregate daily news media events on a firm-week basis and create a dummy variable that measures whether a company is exposed to at least one negative news event in a firm-week from Monday to Sunday over the period 2007–2017. In 31.8% of the firm-weeks with negative news events, firms faced more than one negative news event across the world. We treat them as one event because they cannot be separated at the weekly level and this separation is not relevant to us; we focus on week-by-week predictions.[1]

[1] We also considered the daily level. However, this level is not meaningful for practical reasons. First, the daily level makes the highly unbalanced dataset even more extreme (by factor 4). Second, given our global scale, the day becomes a fuzzy concept for event prediction due to the time difference. A Monday in one time zone is already a Tuesday in another zone.

Figure 1: Number of Negative News Incidents During 2007–2017



2.3 Population of Firms

Our firm population consists of publicly listed firms included in Compustat (North America and Global) and for which we have advertising data from Kantar's database available. We need these firms for the purpose of separating out the threatening news that have significant negative consequences for the firm (see the subsequent section for details). Advertising spending is an important mandatory variable because it can have strong predictive power of news coverage of negative events (Stäbler and Fischer 2020). The intersection of these two databases still results in a large final sample of 3,271 companies. In total, the data set yields 1,588,306 firm-week observations.

Only about 50% of the companies (=1,595) faced one or more negative news events during the 11-year observation period. Including companies with no negative news report is especially valuable to avoid calibrating the prediction model only on the population of firms facing negative events.

2.4 Identification of Trigger Events

Not every negative news event turns into a trigger event with negative consequences for the firm involved. Negative consequences can be diverse, and managers' perceptions of them can subjectively vary. Although measuring all consequences is not possible, we were able to track investors' reactions. Thus, we only predict negative news events that have a negative financial impact for firms – that is, those events associated with negative abnormal stock market returns in the firm-week of occurrence (e.g., Gao et al. 2015).

We apply the market model to identify the trigger events (e.g., Dinner, Kushwaha, and Steenkamp 2019). The stock market return of company i in firm-week w (SR_{iw}) measures the relative change in the stock price from the previous week to the next week. The market model estimates the expected stock market return ($Eret_{iw}$) for a company in a given firm-week. We calculate the expected stock return as follows:

$$\mathrm{Eret}_{\mathrm{iw}} = \alpha_{\mathrm{iw}} + \beta_{\mathrm{iw}}\mathrm{MSCI}_{\mathrm{w}}\,. \tag{1}$$

We use the weekly stock market return of the previous 12 months before the negative news event occurred in form of rolling time window regression to estimate the week and company-specific α and β coefficients (Mazodier and Rezaee 2013). We use the weekly global benchmark stock market return from the MSCI World Index ($MSCI_w$) to predict the expected stock market return for a given week and company. The MSCI World Index represents about 85% of the market capitalization in 23 countries and is considered one of the world's most important stock indices. A global index is suitable as our sample is global, which is also the reason why we cannot apply the Fama-French factors that are only available for the U.S. stock market. We only consider negative news events a trigger if the following condition is fulfilled for the respective week: $SR_{iw} – Eret_{iw} < 0$.

It is not possible to determine the financial impact of every negative news incident because some stock market returns are missing or a share is not traded each period for various reasons (e.g., no public ownership, mergers and acquisitions, stock suspensions). To be conservative, we do not classify these firm-weeks as trigger events.

In total, we identify 14,440 firm-weeks with at least one trigger event out of 41,879 firm-weeks with negative news events, and we found 1,004 (31.2%) companies that faced at least one trigger event between 2007 and 2017. The average abnormal return for the identified trigger events equals −3.3% (median = −2.0%).

2.5 Evolution of Negative News and Trigger Events

Table 1 shows the distributions for firm-weeks with negative news and with trigger events over the 11 years covered in the data set. In line with the increase of news incidents across firms per week shown in Figure 1, we observe a strong increase in the number of firm-weeks with negative news and trigger events over time.

Table 1: Evolution of Firm-Weeks with Negative News Events and Trigger Events

Year	Number of firm-weeks with negative news event	Number of firm-weeks with trigger events	Growth rate of firm-weeks with negative news (in %)	Growth rate of firm-weeks with trigger events (in %)	Share of trigger events (in %)
2007	1,080	270	—	—	25.0
2008	2,010	660	86.1	144.4	32.8
2009	1,602	517	−20.3	−21.7	32.3
2010	2,327	788	45.3	52.4	33.9
2011	2,971	933	27.7	18.4	31.4
2012	4,102	1,349	38.1	44.6	32.9
2013	4,771	1,700	16.3	26.0	35.6
2014	6,056	2,171	26.9	27.7	35.9
2015	5,839	2,295	−3.6	5.7	39.3
2016	5,268	1,650	−9.8	−28.1	31.3
2017	5,853	2,107	11.1	27.7	36.0
Average	3,807.18	1,312.73	21.8	29.7	34.5

On average, the number of firm-weeks with negative news events per year increased by 21.8% and the number of firm-weeks with trigger events by 29.7% from one year to the next, which strongly underscores the rising importance of negative publicity for companies. The average share of trigger events among all negative news is at 34.5%.

Table 2: Industry-Specific Descriptives

Industry[a]	**Number of weeks with negative news event per company and year**	**Number of weeks with trigger event per company and year**	**Share of weeks with trigger event to weeks with negative news event (in %)**[b]	**Share of companies with negative news events (in %)**	**Average abnormal return for trigger event**	**Total sales per year and firm (in $1m)**	**Average monthly advertising budget (in US$000)**
Utilities	3.14	1.01	32.17	82.43	−.024	9,117.73	15.95
Transportation and Warehousing	2.42	1.05	43.61	80.00	−.021	31,072.60	2,050.36
Nonstore Retailers	2.13	.87	40.69	63.16	−.028	14,148.26	2,883.47
Air, Water, Railway Transportation	1.83	.57	31.06	75.33	−.031	8,026.92	264.88
Food and Textile Manufacturing	1.67	.58	34.79	79.59	−.030	6,241.55	2,328.06
Finance and Insurance	1.66	.65	39.33	58.66	−.030	5,658.92	656.57
Mining	1.62	.49	30.10	50.73	−.038	3,708.57	6.14
Wood and Chemical Manufacturing	1.48	.44	29.79	53.78	−.824	9,861.50	1,789.39
Steel and Machine Manufacturing	1.11	.39	34.75	49.49	−.039	6,925.58	938.31
Grocery Retailer	.98	.43	43.55	67.53	−.030	8,287.50	1,374.70
Wholesale Trade	.97	.34	34.56	56.76	−.108	19,549.43	500.40
Accommodation and Food Service	.93	.36	38.49	62.26	−.028	2,898.70	2,464.24
Construction	.71	.28	40.26	61.67	−.037	4,873.32	18.76
Administrative and Support Service	.70	.27	38.49	53.45	−.025	3,002.02	114.27
Information	.50	.17	34.48	33.53	−.028	3,781.93	1,614.85
Professional Services	.41	.09	22.54	42.52	−.038	2,596.00	162.38
Health Care	.38	.14	37.50	54.84	−.041	3,753.35	170.07
Educational Services	.21	.08	36.36	50.00	−.151	1,169.38	1,801.05
Arts and Entertainment	.19	.05	29.03	33.33	−.023	838.01	126.90
Real Estate	.15	.05	32.60	43.93	−.042	1,377.39	218.05
Other Services	.07	.02	25.00	20.00	−.037	1,137.41	460.51

Notes: [a] Industries are identified according to two-digit NAICS code classification.
[b] Share is calculated with the exact values and not with the rounded values presented in columns 2 and 3.

The number of firm-weeks with negative news and trigger events also strongly depends on the industry to which the individual company belongs. Table 2 shows the average yearly number of firm-weeks with negative news and trigger events per company for different industries in the observation period. For example, Information companies (e.g., BlackBerry, Yelp) faced on average .5 negative news events per year; of these incidents, .17 events were trigger events. Utility companies (e.g., American Electric Power, Black Hills) were exposed to the greatest risk of negative news coverage. Firms in this industry faced on average more than one trigger event each year.

The proportion of negative news items that represent trigger events also varies considerably between industries. For example, in the Professional Services industry (e.g., FalconStor, Nielsen), only 22.5% of firm-weeks with negative news events also experience trigger events. In contrast, the Transportation and Warehousing industry (e.g., American Airlines, FedEx) has a much larger proportion, at 43.6%. These large differences are an indication of the relevance of accounting for the industry code when determining the risk exposure.

3 Conceptual Development of the Prediction Model

Before we introduce our conceptual model for the prediction, we discuss insights from other business-related (e.g., churn prediction, bankruptcy prediction) and non-business-related (e.g., quantitative criminology, seismology) scientific disciplines that also must address the challenge of predicting random and rare events with negative impact on economy and society (Asim et al. 2018; Chen, Cho, and Jang 2015; Devi and Radhika 2018; Lemmens and Gupta 2020; Mignan and Broccardo 2020; Schmittlein, Morrison, and Colombo 1987). Our aim is to obtain information about variables used in these fields and to evaluate their potential for our prediction task.

Table 3 summarizes the literature streams from these fields and overviews important predictors used in these prediction tasks (column 2). Table 3 also shows which predictor sets we derive from the analyses of other disciplines for our prediction model (column 4) and introduces typical prediction performance results (column 3) that may serve as a point of reference to evaluate the performance of our prediction model.

3.1 Insights from Related Research Fields

3.1.1 The prediction of earthquakes and crimes. Earthquakes seem to occur randomly in time and space. However, what appears to be a random sequence of events actually follows a stochastic process that can be modeled and parametrized with information on the magnitude, location, frequency, and recency of historical earthquake activity (Mignan and Broccardo 2020). For our prediction model, we similarly keep in mind that the trigger events may also be closely interrelated in time and that previous events may provide valuable information for the future occurrence of events.

As noted previously, we do not predict all negative news events but rather only those with a negative financial impact. A similar procedure of filtering events is applied in earthquake prediction literature, in which researchers only focus on severe earthquakes that reach a magnitude with severe social-economic consequences (Asim et al. 2018; Harte 2016).

Like an earthquake, the occurrence of a crime (e.g., residential burglary) increases the likelihood of another crime nearby in space and time. Consequently, researchers forecast crime patterns by using historical information on the location, frequency, and recency of crimes (Mohler et al. 2011).

Recent model applications (e.g., Chen, Cho, and Jang 2015) have added social media data from Twitter and weather data as weather conditions influence aggressive behavior. We follow these studies and integrate geographical, online-based consumer data, and weather data to better predict the occurrence of trigger events.

Table 3: Related Literature and Sets of Predictors

Literature stream	Set of predictor variables	Exemplary reported prediction performances	Predictor sets used in this model
Prediction of earthquakes (e.g., Asim et al. 2018; DeVries et al. 2018; Freund et al. 2021; Harte 2016; Marsan and Lengliné 2008; Mignan and Broccardo 2020)	- Event history: recency and frequency of earthquakes - Geo-physical and chemical lead signals - Behavior of animals/financial markets	- ROC AUC: .85 - Accuracy: .82 - Bal. accuracy: .75	- News event history: Recency and frequency of negative news events and trigger events - Location (of headquarters)
Prediction of crimes (e.g., Chen, Cho, and Jang 2015; Johnson et al. 2007; Mohler et al. 2011)	- Event history: recency and frequency of crimes - Location characteristics - Weather data - Online data (e.g., Twitter)	- ROC AUC: .67	- News event history - Location (of headquarters) - Weather data - Online data (i.e., Google search volume)
Prediction tasks in marketing and other business disciplines - Business failure/ bankruptcy (e.g., Barth, Beaver, and Landsman 1998; Naumzik, Feuerriegel, and Weinmann 2022) - Customer churn (e.g., Donkers, Franses, and Verhoef 2003; Lemmens and Croux 2006; Lemmens and Gupta 2020; Schmittlein, Morrison, and Colombo 1987)	- Customer purchase history: recency and frequency - Geographic area - Consumer data (i.e., ratings) - Historic performance of stock price - Financial health of company/ estimated income of potential churner - Relationship with customer - Marketing activities (e.g., customer care calls for churn prediction)	- ROC AUC: .85	- News event history - Historic performance stock price - Location (of headquarters) - Online-based consumer data (i.e., Google search volume) - Financial health (e.g., leverage, profitability, market share) - Industry - Relationship equity - Marketing variables (e.g., advertising spending, SG&A expenses)

Notes: ROC = receiver operating characteristic; AUC = area under the curve; Bal. accuracy = balanced accuracy. These are common metrics used to evaluate the classification performance, where 1 indicates perfect prediction and 0 worst prediction. See Appendix B for more details and a formal definition of the metrics.

3.1.2 The prediction of risks in marketing and other business disciplines. Prediction of risky events are also well known in finance and accounting (e.g., credit default, bankruptcy) as well as marketing (e.g., customer churn, business failure). Studies in this literature stream identify various types of drivers and prediction models; for example, Mahajan, Srinivasan, and Wind (2002) identify the stock price development as a potential performance indicator that explains retailer bankruptcy, whereas Naumzik, Feuerriegel, and Weinmann (2022) show the importance of customer data (i.e., customer ratings) for predicting business failure.

In customer churn prediction models, considering the previous purchase behavior of customers (e.g., the time since the last purchase and the number of previous purchases in a certain time period; Schmittlein, Morrison, and Colombo 1987) is essential. Lemmens and Croux (2006) include marketing activities in the form of customer care calls in their churn prediction models. Comparable to a company's financial health and headquarters location, they also control for the income of a potential churner as well as the geographical area. Donkers, Franses, and Verhoef (2003) include the type of relationship between a potential churner and the company. Building on these insights, we identify important predictors such as the history of event occurrence, the company's financial health, previous stock market developments, marketing activities, and consumer data.

3.1.3 Summary. In prediction modeling for rare events, the most important criteria are the history (i.e., frequency and recency of events). In addition, seemingly unrelated variables that are correlated with the events should be identified, (e.g., weather as a factor for criminal behavior prediction and financial market developments for earthquake prediction). In our study, we also use seemingly unrelated variables such as previous weather conditions to predict the occurrence of trigger events.

Building on our analysis of other disciplines, we identify four sets of predictors that could be relevant for a trigger event forecast. We explain the process and rationale for selecting each predictor variable in the following section.

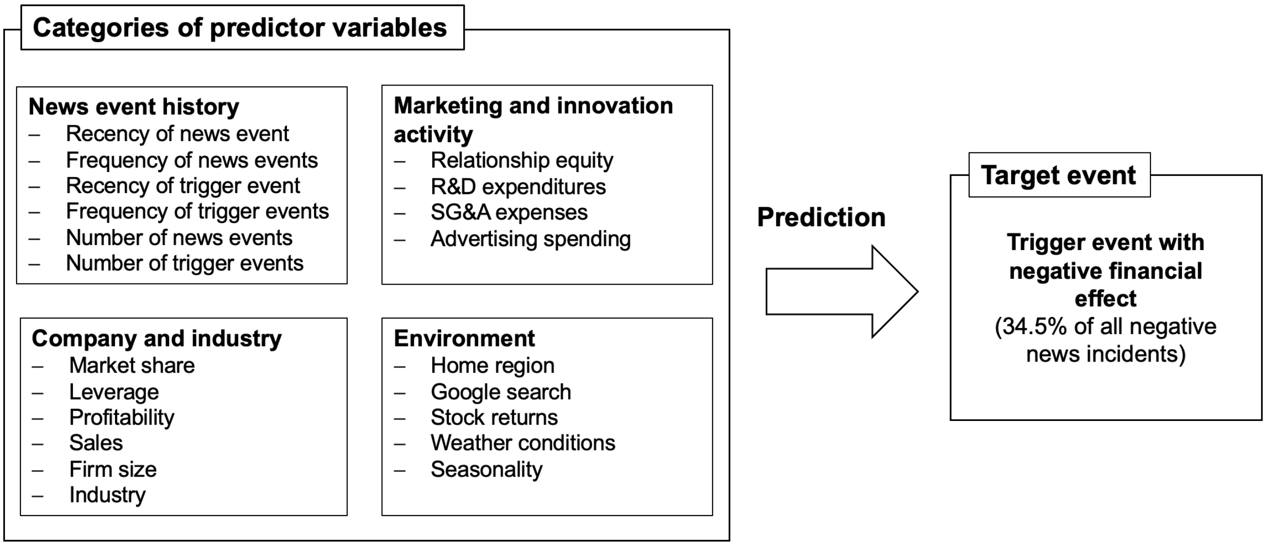
3.2 Selecting Variables for the Prediction of Trigger Events

Our prediction variables must meet one focal requirement to be considered in the model: the information should be publicly available (e.g., weather, Google search volume) or easily accessible (e.g., financial data). The data may be measured at different temporal aggregation levels (e.g., time-invariant, year, month, week). This mix is a strength as it exploits both the cross-sectional (e.g., firm characteristics) and the time variation (e.g., Google search). In addition, the data do not need to be complete and can have missing values over time; our prediction method can effectively handle these missing data. In line with our reasoning, we do not use qualitative, topic-based, and sentiment-based unstructured information because they are not easily accessible (e.g., Instagram posts), selective (e.g., product reviews), and require analysis techniques (e.g., NLP algorithms) to extract information that are prone to subjectivity.

Predictor variables may follow three main mechanisms to motivate the potential predictive capacity. First, the literature review shows that rare events often follow a stochastic process that can be described via variables of event frequency and recency. Second, predictors describe conditions that make unethical behavior more likely. For example, a company in a shaky financial situation may put pressure on its mangers to act unethically to reduce costs (Duanmu, Bu, and Pittman 2018). Third, variables explain why and how company issues become newsworthy such that media and other sources become more interested and focused on these companies. For example, Stäbler and Fischer (2020) show that the media are more likely to monitor and cover large firms, which increases the general probability that large firms' unethical behavior becomes public or that they are accused of such unproven behavior.

Figure 2 shows the four sets of predictors we use for the forecast: news event history, company and industry variables, marketing and innovation variables, and environmental variables. The following subsections elaborate on these predictors.

Figure 2: Conceptual Prediction Model



3.2.1 News event history. Forecasting models of seismology and criminology are based on the assumption that rare events can be predicted by prior event activity (e.g., Johnson et al. 2007; Mignan and Broccardo 2020; Mohler et al. 2011). Customer churn modeling (e.g., Schmittlein, Morrison, and Colombo 1987) follows similar principles. Therefore, the recency and the frequency of trigger events and negative news events in general are potentially powerful predictors. Although our model does not predict every negative news event, only those that represent a trigger event according to our definition, we still include the history of all negative news events of a company.

3.2.2 Company and industry variables. A firm's financial health is a predictor of risk such as bankruptcy (e.g., Barth, Beaver, and Landsman 1998). If managers envision such risks, they may increase pressure on employees to avoid them, which could lead to a higher propensity to engage in unethical and controversial behavior (e.g., reducing cost by increasing environmental pollution; Duanmu, Bu, and Pittman 2018; Dupire and Zali 2018). We thus measure the companies' financial health with firm profitability, leverage, and market share.

In addition, because firm size may influence the likelihood of trigger events, we also control for company size by including sales and the number of employees. For example, larger firms with many operations in different parts of the world increase the odds of employees engaging in unethical behavior simply because they have more employees. In addition, the media is more likely to monitor and cover large firms (Stäbler and Fischer 2020).

The industry may also differ in its attraction potential to media and the public. The public and the media observe industries with a lower reputation such as financial institutions or "dirty" industries (Dupire and Zali 2018, p. 608) to a higher degree, as they are more likely to be suspected of unethical behavior. Moreover, research has documented negative spillover effects from competitors within the same industry (Roehm and Tybout 2006), leading to higher news coverage. For these reasons, our model also considers the industry an important cross-sectional variable.

3.2.3 Firms' marketing and innovation activity. Higher SG&A expenses and investments in customer relationships, R&D, and advertising improve marketing performance, which builds brands and customer base and increases the salience and presence of companies and their brands (e.g., Morgan and Rego 2009; Rust et al. 2004). These activities, in turn, can increase media attention (e.g., Stäbler and Fischer 2020). These activities can also increase pressure for managers of successful companies to maintain the success (Mishina et al. 2010), which lowers the threshold for unethical firm behavior, resulting into negative publicity when it is revealed. In line with this argumentation, we include receivables investments as a proxy for relationship equity (Frennea, Han, and Mittal 2019), R&D investments, advertising investments, and SG&A expenses in the model.

3.2.4 Environmental variables. Environmental variables include the home region of headquarters, weather conditions, stock market returns, and Google search volume. On the one hand, headquarters are the heart of companies and have a major influence on company culture, value systems, and policies (Lee 2020). On the other hand, the headquarters are influenced by culture, value systems, and norms in the society of the geographic regions, which can influence managers' behavior. Hanlon and Slemrod (2009) show that when firms chose headquarters locations that are known for being tax havens for tax avoidance, they are more likely labeled as socially irresponsible. Following the argument of the important role of industry, this poor reputation could also lead to increased public awareness concerning the unethical behavior of these companies.

Similarly, research has shown that weather conditions have an impact on criminal behavior (e.g., Chen, Cho, and Jang 2015) in that it affects people's moods and human decision making in general; for example, air pollution deteriorates people's mood and intensifies their risk aversion (e.g., Lepori 2016). These findings suggest that weather may also have predictive quality with respect to the occurrence of threating corporate news.

The behavior of large social groups such as investors at the capital market and internet users who search for companies are also potentially relevant for the prediction. The capital market is highly efficient in processing information, regardless of source, provided it is relevant to the company's value. Online interest in brands and companies has been shown to impact media's attention to news about unethical firm behavior (Stäbler and Fischer 2020). Du and Kamakura (2012) point out the importance of trends in Google search volume in measuring shifts in consumer interest. Hence, we include online search volume as a relevant consumer index in our model. Finally, to control for potential seasonality within a calendar year, we enrich our model with monthly dummies.

3.3 Variable Operationalization and Descriptives

Table 4 provides information on variable measurement. We use historical data from a wide range of sources including RepRisk, Compustat, Kantar, Google, Yahoo Finance, and Visual Crossing[2]. The temporal aggregation level differs across variables (Table 4, column 5), which, as noted previously, our model can accommodate. Our model predicts the trigger event probability in the upcoming week. The column "Number of time lags used" in Table 4 illustrates how many time lags we build for each variable. For yearly data (e.g., Compustat data), we use one time lag (i.e., the previous year). For more disaggregated data, we consider more time lags. For example, we use the Google search volume of the last six months in the form of six time lags (1–6) and the stock returns of the previous four firm-weeks (1–4).

Table 4 also shows summary statistics. Because this information may be heavily inflated by outliers, we report the median and the 5% trimmed mean (mean after excluding the lowest 5% and highest 5% of the values) for each variable in addition to the mean and the standard deviation. The annual average relative R&D expenditures equals 4.4% and the companies spent US$1,051,980 per month on advertising. On average, 18,911 employees work in our sample of companies.

We include multiple variables to describe the weather at the company's physical headquarters location. For the sake of brevity, we explain these individual variables in Appendix C.

As mentioned previously, variables may have missing values. For example, not all companies are equally required to disclose balance sheet ratios, or weather data may not be collected in the same way in every location (e.g., entries on wind speeds at the company headquarters are missing in 4.0% of all cases). We describe how we handle missing values in the following sections.

[2] Visual Crossing is a leading provider of weather data to data scientists, business analysts, professionals, and academics. See https://www.visualcrossing.com for more information.

Table 4: Variable Operationalization and Descriptives

Variable	Operationalization	Source	Measurement unit	Temporal interval	Number of time lags used	Mean	Median	Standard deviation	5% trimmed mean
Target variable									
Trigger event	Dummy variable indicating whether the company faces a trigger event in a given firm-week	RepRisk	dummy	week		.009	0	.095	0
News event history variables									
Recency of news events	Firm-weeks since the last negative news event	RepRisk	count	week	1	184.051	137	166.256	174.424
Frequency of news events	Number of firm-weeks with negative news events within rolling time window of previous six months		count	week	1	.635	0	2.261	.220
Recency of trigger events	Firm-weeks since the last trigger event		count	week	1	224.836	201	170.038	219.097
Frequency of trigger events	Number of firm-weeks with trigger events within rolling time window of previous six months		count	week	1	.219	0	.940	.048
Number of news events	Number of firm-weeks with negative news events in previous calendar year		count	year	1	1.295	0	4.496	.474
Number of trigger events	Number of firm-weeks with trigger events in previous calendar year		count	year	1	.445	0	1.800	.116
Company and industry variables									
Market share	Sales/sales of industry[a]	Compustat	ratio	year	1	.103	.020	.190	.072
Leverage	(Long-term debt + debt in current liabilities)/total assets		ratio	year	1	.901	.177	37.462	.200
Profitability	EBIT/total assets		ratio	year	1	−.340	.057	17.043	.046
Sales	Total sales		count in $1m	year	1	6,612.802	871.192	21,655.202	3,100.254
Firm size	Number of employees		count in 000	year	1	18.911	3.300	49.666	10.385
Industry	Dummy variables indicating industry[b]		dummy						

Table 4: Variable Operationalization and Descriptives

Variable	Operationalization	Source	Measurement unit	Temporal interval	Number of time lags used	Mean	Median	Standard deviation	5% trimmed mean
Marketing and innovation variables (scaled by industry level):									
Relationship equity	Receivables/receivables of industry[a]	Compustat	ratio	year	1	.096	.014	.188	.064
R&D expenditures	R&D expenditures/R&D expenditures of industry[a]		ratio	year	1	.114	.016	.212	.080
SG&A expenses	(SG&A expenses/SG&A expenses of industry[a])/total assets in $1m		ratio	year	1	.002	$.019\times10^{-3}$	.096	$.058\times10^{-3}$
Advertising spending	• Monthly advertising spending	Kantar Media	count in 000$	month	1–4	3,137.916	66.800	11,208.856	1,111.653
	• Average monthly advertising spending per year		count in 000$	year	1	1,996.619	31.500	8,544.976	545.765
	• Variance of monthly advertising spending per year			year	1	$.832\times10^{+7}$	453.272	$.544\times10^{+8}$	$.690\times10^{+6}$
Environmental variables:									
Home region	Location of firm headquarters (10 regions)	Compustat	dummy						
Google search	Relative Google trends activity based on company name (benchmark name: Abercrombie & Fitch)	Google	continuous	month	1–6	33.828	0	218.018	5.589
Stock returns[c]	(Adjusted stock price$_w$ − Adjusted stock price$_{w-1}$)/Adjusted stock price$_{w-1}$	Yahoo finance	ratio	week	1–4	.063	0	12.176	.001
Weather conditions	12 variables including wind, heat, and temperature (see Appendix C for more details)	Visual Crossing	continuous	week	1–2				

Notes: The sample includes 3,271 firms and 1,588,306 firm-week observations. Within the observation period (January 2007–December 2017), 51.24% of the firms did not face a negative news event.
[a] Based on three-digit NAICS codes. [b] Based on two-digit NAICS codes.
[c] We use the daily stock market prices from Monday to Monday to match stock market reaction as closely as possible with our trigger event observation in week w.

4 Methodology

4.1 Trigger Forecast Model

We consider four established methods to predict the probability of trigger event occurrence in the upcoming week: a logit model, a random forest approach, an artificial neural network (ANN), and an XGBoost approach (XGB). We do not have an a priori preference for any method and use the logit model as a benchmark model. All approaches have particular strengths and weaknesses. To identify the approach that works best, we compare the prediction performances according to the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and the precision recall (PR) curve. Both metrics are independent of a specific classification threshold, as they evaluate the performance based on all potential classification threshold values and are well established across disciplines (e.g., Branco, Torgo, and Ribeiro 2016; Naumzik, Feuerriegel, and Weinmann 2022). Note that we report additional performance metrics (i.e., balanced accuracy and F_1-score) when we apply a specific threshold that classifies for each firm-week whether a trigger event is to be expected.

4.1.1 Benchmark model: logit model. Following prior literature, we use a logit model as our benchmark model (e.g., Lemmens and Croux 2006, p. 281). Logit models are by far the most frequently used method in binary econometric application settings because they are easy to implement and efficient to train, and researchers do not need to identify optimal parameter values in the training process of the algorithm.

The logit model can be written as follows:

$$\mathrm{Prob}_{\mathrm{i,wmy}}(\text{trigger event} \mid \mathbf{X}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \mathbf{FE}) = \frac{e^{f_{\mathrm{i,wmy}}(\mathbf{X})}}{1+e^{-f_{\mathrm{i,wmy}}(\mathbf{X})}}, \tag{2}$$

$$\text{with } f_{\mathrm{i,wmy}}(\mathbf{X}) = \alpha + \sum_{l=1}^{L_{\mathrm{week}}} \boldsymbol{\beta}'_{l}\mathbf{X}^{\mathrm{w}}_{\mathrm{i,my,w-l}} + \sum_{l=1}^{L_{\mathrm{month}}} \boldsymbol{\gamma}'_{l}\mathbf{X}^{\mathrm{m}}_{\mathrm{i,wy,m-l}} + \boldsymbol{\delta}'\mathbf{X}^{\mathrm{y}}_{\mathrm{i,wm,y-1}}$$

$$+\mathrm{FE}_{\mathrm{Month}} + \mathrm{FE}_{\mathrm{HR}} + \mathrm{FE}_{\mathrm{Ind}},$$

where $Prob_{i,wmy}$ denotes the probability of a trigger event to occur for firm i in week w, month m, and year y. **X** is a vector collecting predictors, α denotes the intercept, and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ are parameter vectors associated with predictors measured at the weekly, monthly, and yearly level, respectively. $L \in \mathrm{L}$ is an index for the time lag where the number of lags differs across variables and temporal aggregation level. FE denotes fixed effects related to month (Month), home region of the firm's headquarters (HR), and industry (Ind). We consider the same predictors for all four methods to create an equal comparison.

4.1.2 Additional machine learning methods. ML-based algorithms may outperform the logit model. Therefore, we additionally test the performance of a neural network, a random forest, and a XGB algorithm. These three approaches make it possible to flexibly analyze high-dimensional data with many different variables. We offer more detailed descriptions of these three methods in the Appendix D.

Random forest is a tree-based algorithm using bootstrap aggregation (Breiman 2001). The strength of the random forest comes from creating trees with different subsets of features. It is an ensemble-based method, in which the final prediction is made by a majority vote of trees. The feature selection for each tree is random. Compared with the logit model, it may require more computational power as well as resources as it builds numerous trees to combine their outputs. Thus, it often needs more time for training than fitting the logit model.

The *XGB algorithm* is also a tree-based approach that trains gradient-boosted decision trees (Chen and Guestrin 2016). In contrast to the random forest, the trees are not trained independently; rather, each tree incrementally incorporates and corrects the error produced by the previously trained tree. Both, random forest and gradient boosting are widely used and are among the methods with outstanding prediction performance (e.g., Chang, Chang, and Wu 2018, Olson et al. 2018). Depending on the application field, either approach could excel. For example, random forests tend to perform better in bioinformatics, which often deals with a great deal of noisy data (e.g., Strobl et al. 2008), whereas gradient boosting is better suited for unbalanced classification tasks (Chang, Chang, and Wu 2018). As our data are both, noisy and highly imbalanced, it is reasonable to test both approaches.

An *ANN* models the nonlinear relationship between a set of input signals and an output signal using a model derived from how a brain responds to stimuli (e.g., Kuhn and Johnson 2013, p. 141). As discussed previously, findings from earthquake research show that the application of neural networks can significantly contribute to the prediction of earthquakes (e.g., Mignan and Broccardo 2020).

4.1.3 Subsets to train models and test performance. We split the data set into three subsets – a training, a testing, and a holdout data set. We identify the optimal parameter settings of the algorithms by training the algorithms with the training data and maximizing the prediction performance of the testing data. The purpose of the holdout sample is to evaluate the final prediction performance of the four models.

We have a total firm population of 3,271 companies. Of these companies, we randomly select 495 companies (about 15%) whose data will serve only as the holdout validation. None of the observations of these holdout companies are included in the training or testing process of the algorithms. In addition, the holdout sample includes only data from 2016 and 2017. Next, we build the training data set, which contains all observations of the remaining companies in the time period 2007–2013. Finally, we create the testing data, which includes the observations of 2014 and 2015 of the same companies as the training data. Table 5 summarizes the most important descriptive statistics of the three subsets. It reports the number of companies and observations included in the subsets, the covered time window, the number of firm-weeks with trigger events, and the resulting incidence of the target variable.

Table 5: Descriptives of the Subsets for Training, Testing, and Validating

Type of subset	Number of companies	Number of observations	Time window	Number of firm-weeks with trigger events	Incidence of trigger event
Training	2,624	834,624	2007–2013	5,105	.61%
Testing	2,328[a]	239,898	2014–2015	3,611	1.51%
Holdout	495	50,752	2016–2017	738	1.45%
Total:	—	1,125,274	2007–2017	9,454	.84%

Notes: [a] The training data and the testing data include the same companies. The number of companies in the testing data is smaller because we observe company entries and exits over time.

By analyzing the prediction performance on the holdout observations, we evaluate the models by using information about companies as well as data from years that are unknown to the trained algorithm. This information provides both a company-based and a time-based holdout sample and thus a robust approach for evaluating the model performance. As a consequence, the model can be applied to new companies in future periods, although it is advisable to update and improve the algorithm constantly with new data as they arrive over time.

4.1.4 Data preparation. Each of the aforementioned algorithms needs specific data preparation because they process the data in different ways. For example, the XGB approach is capable of handling missing values by properties of the algorithm, whereas the other methods require the user to impute missing values (see Chen and Guestrin 2016). Specifically, the XGB does not ask for adding new information from the outside but exploits the patterns of missing values and learns how the occurrence of a missing value itself contributes to better prediction. The learned rules are easily transferred to the holdout sample prediction.

For the ANN, the random forest, and the logit model, we follow common practice and impute the missing values of each predictor by using the median values. The median imputation increases reliability because the mean of the variables in our study is affected by the existence of outliers (Brown and Kros 2003). We use the median values of the training data for the missing value imputation of the testing and holdout sample.

To counteract the unequal distribution of the trigger event variable, we balance the training data set for the logit model, the random forest, and the ANN so that the target variable is equally distributed. Prior research (e.g., Lemmens and Croux 2006) has shown that balancing the data set increases the prediction performance. For the XGB, no up- or down-sampling is required to balance the dataset. Instead, a hyperparameter is set to weight the classification error. We follow common practice and use the inverse of the class distribution to set this parameter, so that errors of predicting both positives and negatives are equally weighted (Chen and Guestrin 2016). As a result, the performance between the four prediction models is comparable and no model is disadvantaged.

4.1.5 Assessment criteria for the prediction and classification performance. In the first step, the four models produce a probability of an event occurrence. In the second step, this probability must be transformed into a 0 or 1 indicator of the event for each weekly observation by applying a classification threshold.

As overviewed in Table 6, binary classification tasks produce true positive predictions when a class (in our context, an event) is correctly predicted and true negative predictions when a class is correctly not predicted. Two errors may occur: a false positive result occurs when a class is falsely predicted and a false negative result occurs when a class is falsely not predicted.

Table 6: Outcomes for a Binary Classification Task

		Actual class	
		Trigger event	No trigger event
Predicted class	Trigger event	True positives	False positives
	No trigger event	False negatives	True negatives

We differentiate between general performance metrics for model prediction that are independent of a specific classification threshold and a classification performance that depends on a chosen threshold value. To identify the model with the best prediction performance in general, we evaluate the AUC value of both the ROC curve and the PR curve of the holdout sample (for a detailed explanation of these metrics and the underlying information, see Appendix B). The ROC curve summarizes the prediction performance in terms of true positive rate and true negative rate across each potential classification threshold value. Although the ROC AUC value is a popular and well-established measure to evaluate the performance of binary classifiers, the metric can be misleading in highly imbalanced classification scenarios like ours depending on the preferences of the user (Saito and Rehmsmeier 2015). In that scenario, PR curves add another important dimension of the classification performance because they also evaluate the fraction of true positives among all positive predictions (called precision). Thus, studies have also recommended using PR curves for comparing models based on imbalanced data (Saito and Rehmsmeier 2015). PR AUC values are usually much lower than ROC AUC values. Both metrics should be compared to their relevant random classifier benchmark, which is 50% for ROC AUC and the incidence of the event to be predicted for the PR AUC (1.45% in the holdout sample).

To investigate the classification performance for applications based on a pre-defined threshold, we use two established performance metrics, the balanced accuracy and F_1-score (e.g., Olson et al. 2017). Balanced accuracy is a normalized version of accuracy that accounts for class imbalance by calculating accuracy on a per-class basis and then averaging the per-class accuracies. Balanced accuracy as a function of threshold TR, with TR ∈ [0%, 100%], can be written as follows:

$$\text{Balanced accuracy } (TR) = \frac{\frac{\text{True positives } (TR)}{\text{True positives } (TR) + \text{False negatives } (TR)} + \frac{\text{True negatives } (TR)}{\text{True negatives } (TR) + \text{False positives } (TR)}}{2} \quad (3)$$

The equation for the F_1-score is as follows:

$$\text{F}_1 - \text{score } (TR) = \frac{2 \times \text{True positives } (TR)}{2 \times \text{True positives } (TR) + \text{False negatives } (TR) + \text{False positives } (TR)} \quad (4)$$

Both balanced accuracy and the F_1-score metrics consider the two resulting misclassifications (i.e., false positives and false negatives), which enables a weighting of the errors. We explain how we make our classification cost-sensitive in the following section.

4.2 Making the Classification Cost-Sensitive

Cost-sensitive learning is a type of machine learning that considers the misclassification costs caused by incorrect classifications. The goal is to minimize the overall costs resulting from the classification instead of simply achieving the highest possible classification accuracy. Most importantly, cost-sensitive learning treats the misclassifications (i.e., false positives and false negatives) differently because the costs they cause could vary in size (Elkan 2001).

The machine learning literature offers several approaches to make classifications cost-sensitive (e.g., Elkan 2001; Sheng and Ling 2006), and we chose to apply threshold adjusting (e.g., Sheng and Ling 2006). In the first step, threshold adjusting identifies the threshold in such a way that the costs arising from the resulting misclassifications are minimized using the observations of the training data set. This cost-optimal threshold from the training data set is then applied to holdout data.

Adjusting the threshold comes with an important benefit compared with other cost-sensitive learning methods: the algorithms' predicted probability remains unchanged regardless of the person using the prediction model; only the cost-optimal threshold differs between users in accordance to how they perceive and weight the misclassifications costs. This benefit enables a tailored adaptation to the risk preferences of every single user.

The classification performance as well as the size of the classification errors depend on the selected threshold value, which classifies each observation into one of two possible states. In marketing research, 50% is the common threshold for a binary classification task (e.g., Sismeiro and Bucklin 2004, p. 319); however, that threshold is not necessarily optimal. The two resulting classification errors are related in a trade-off relationship that is influenced by the choice of the threshold value. If a smaller threshold is applied for the classification, the proportion of false negatives (i.e., missed trigger events) will be lower and the proportion of false positives (i.e., falsely predicted trigger events) will be higher, and vice versa.

The question that arises next is how to set the threshold optimally for potential users. Because the trade-off relationship between these two prediction error costs is subjective and depends on the person's risk preferences, we do not impose a specific cost relation. For illustrative purposes, we weight one error higher (false negative) than the other error (false positive) following the principle of loss aversion as proposed by prospect theory (Kahneman and Tversky 1979), which indicates that decision-makers tend to place more weight on perceived losses than on perceived gains. A ratio of approximately 2:1 reflects the generalized value of the loss aversion parameter across studies (e.g., Kalyanaram and Winer 2022), and thus, we apply a perceived cost ratio of 2 (false negative) to 1 (false positive) for illustration.

For our classification task, we need to solve the following optimization problems. For balanced accuracy, we identify the optimal threshold, which maximizes the balanced accuracy of the training data by taking the cost differences into account. This is equivalent to minimizing the weighted sum of prediction error rates over all possible thresholds TR. Hence, using the information from Equation 3, we can formally express this as follows:

$$\min_{TR} \text{Classification Error Rate for Balanced Accuracy}$$
$$= \frac{\text{False positives } (TR)}{\text{False positives } (TR) + \text{True negatives } (TR)} + 2 \times \frac{\text{False negatives } (TR)}{\text{False negatives } (TR) + \text{True positives } (TR)}. \quad (5)$$

Further, we identify the cost-optimal threshold that maximizes the F_1-score of the training data according to Branco, Torgo, and Ribeiro (2016):

$$\min_{TR} \text{Classification Error Rate for } \mathrm{F}_1 - \text{score}$$
$$= \frac{1 + 2^2 \times \text{True positives } (TR)}{1 + 2^2 \times \text{True positives } (TR) + 2^2 \times \text{False negatives } (TR) + \text{False positives } (TR)}. \quad (6)$$

5 Empirical Results

5.1 Overall Prediction Performance

Table 7 reports the AUC of the ROC and PR curves of the four prediction approaches for the training set and for the holdout sample. Recall that these measures summarize the prediction performance across all possible thresholds. As shown, in terms of both ROC and PR AUC value, the XGB algorithm achieves the best prediction performance in the holdout sample, with a ROC AUC of 92.55% and a PR AUC of 20.00%.

Recall that the high ROC AUC value should be evaluated in light of the highly imbalanced target variable with the PR AUC (Saito and Rehmsmeier 2015). We report the PR AUC as well as the balanced accuracy and the F1-score in the following section. In addition, we report confusion matrices and additional metrics for all four model classifications in Appendix E.

Table 7: Overall Prediction Performance in Holdout Sample (495 firms, 2016–2017)

Approach	ROC AUC		PR AUC			
	Training	Holdout[a]	Training	Improvement multiplier[b]	Holdout[a]	Improvement multiplier[b]
XGB	**98.12%**	**92.55%**	20.84%	34.16	**20.00%**	**13.79**
Logit	93.45%	91.20%	n.a.[c]	-	17.76%	12.25
ANN	95.95%	89.40%	n.a.[c]	-	14.72%	10.15
Random forest	94.40%	91.12%	n.a.[c]	-	12.73%	8.78

Notes: Best values are indicated in boldface.
[a] Performance differences between models are significant (except between the logit and random forest models for ROC AUC) based on 1,000 bootstrap subsamples of the holdout sample ($p < .01$, two-sided t-test).
[b] The improvement multiplicator measures the improvement of the model compared with .61% (training data) and 1.45% (holdout data), respectively.
[c] Values are not meaningful because the training data set was rebalanced, which impacts the distribution of the target variable and the resulting PR AUC.

Especially for the PR AUC values, which offers more insights for the prediction of rare events, larger differences in performance between the models are evident. In contrast to the ROC AUC value, the PR AUC value must be compared and evaluated against a value resulting from the distribution of the target variable to be predicted. This value represents the classification performance of a random classifier. In the holdout sample, trigger events represent only 1.45% (see Table 5) of the observations, meaning that the XGB performs 13.79 times better than a random classifier.

Interestingly, the simple logit model also performs quite well. It outperforms the random forest and the ANN, but not the XGB. The PR AUC performance of the logit model is 12.25 times better than a random classifier, whereas the ANN performs only 10.15 times better and the random forest only 8.78 times better than a random classification.

5.2 Specific Prediction Performance

According to the ROC AUC and PR AUC, XGB is the strongest model to predict trigger events regardless of the selected threshold. However, choosing a specific meaningful threshold might change the picture. Table 8 presents the classification performance of the four models for the holdout sample applying the common threshold of 50% and ignoring different misclassification costs as well as the cost-optimal thresholds identified per Equations 5 and 6. The classification results of each model are consistent with the threshold-independent prediction performance. For both balanced accuracy and F_1-score, the XGB achieves the best results for a threshold of 50%. Applying the optimal thresholds to the holdout sample, the XGB still performs best with a balanced accuracy of 85.50% and an F_1-score of 31.35%.

The results further show how large the difference on the performance metrics can be when applying the cost-sensitive threshold compared with the traditional threshold of 50%. For the XGB, the most promising approach, the balanced accuracy increases by only 0.04%, but the F_1-score more than doubles from 15.58% to 31.35%. These results strongly support the recommendation that users adjust the threshold in accordance to their preferences. For both the intuitive threshold of 50% and the two cost-sensitive thresholds, the XGB algorithm consistently outperforms all other algorithms and achieves the highest balanced accuracy as well as the highest F_1-score.

Table 8: Specific Prediction Performance in Holdout Sample (495 firms, 2016–2017)

		XGB	ANN	Random forest	Logit
Exogenous threshold (50.00%)	Bal. accuracy	**85.46%**	84.01%	83.35%	84.61%
	F_1-score	**15.58%**	12.16%	9.85%	13.74%
	(Threshold)	(50.00%)	(50.00%)	(50.00%)	(50.00%)
Endogenous cost-optimal threshold	Bal. accuracy[a]	**85.50%**	82.00%	81.41%	83.67%
	(Threshold)	(43.68%)	(32.59%)	(32.58%)	(36.63%)
	F_1-score[b]	**31.35%**	26.65%	20.64%	29.54%
	(Threshold)	(74.93%)	(97.27%)	(99.53%)	(98.78%)

Notes: Best values are indicated in boldface.

[a] Threshold that maximizes cost-weighted balanced accuracy of training set, as explained in Equation 5.

[b] Threshold that maximizes cost-weighted F_1-score of training set, as explained in Equation 6.

6 Managerial Value of the Prediction Model

Our findings have important implications for managers and for investors. Considering the severe consequences that trigger events may have for firms, our all-encompassing approach to forecasting offers a valuable method of possibly attaining early warning before these events occur. We first quantify the value of our week-to-week trigger forecast financially for investors. We then identify the main drivers of the trigger event probability. Finally, we show how managers can use the predicted weekly trigger likelihood to track the weekly risk and anticipate a threatening event over a longer time horizon.

6.1 Investment Cost-Benefit Analysis

Trigger events seem to occur unpredictably and surprise stakeholders. For example, consider Volkswagen's dramatic overnight plunge in stock market value by 22% after its unethical misconduct (Volkswagen manipulated software, which was used to conceal the actual pollution its cars were emitting) became known to the public (Siano et al. 2017). The knowledge advantage generated by our forecasting model enables investors to anticipate the likely occurrence of such trigger events one week in advance, which has enormous potential to create financial value added for decision makers.

The week-to-week classification indicates whether a trigger event should be expected in the next week. If the predicted risk exceeds the individual cost-optimal threshold, investors can proactively adjust their actions (e.g., sell the stocks before the expected event to avoid losses and reinvest the money withdrawn in less riskier alternatives). We illustrate the financial added value for investors based on a simulation in which we compare the financial gains using our model versus not using it. Our simulation is based on 50,752 observations, 495 companies, and a total of 738 firm-weeks with trigger events from the holdout sample. In this simulation, the investor uses our predicted trigger probability and is prompted to sell a given stock when the probability exceeds the cost-weighted threshold of 43.68% and hold it otherwise. If the investor sells the shares, the money is directly reinvested into shares for which no trigger event is predicted. This investment strategy follows a straightforward, conservative rule. More profitable investment strategies (e.g., short-selling strategies) are likely to exist, and we leave these strategies to be explored in future research.

Using the average market capitalization of a company included in the MSCI index (US$35 billion) and the average abnormal stock market returns for weeks with trigger events in the holdout sample (−2.72%), investors would face total economic costs of US$700 billion because of trigger events (738 trigger events in the holdout sample) without using our prediction model (for an explanation of the detailed calculation, see Appendix F). These costs, however, can be reduced by 42.57% to US$402 billion when investors sell and purchase other stocks based on our forecast. Using our prediction model, the investors miss advance warning of only 102 of the 738 trigger events (false negatives). Note that the cost calculations already include the costs due to the second prediction error (false positives). In the simulation, a trigger event is falsely predicted in 7,597 cases (see Appendix F).

6.2 Substantive Insights into the Drivers of Trigger Event Probability

Although marketing research has dealt extensively with the consequences of negative publications and corporate crises, knowledge of the antecedents and information about how and when such corporate crises start is limited. We identify a large range of predictors. However, it is not yet clear how these variables contribute to the prediction of the likelihood to face a trigger event. The XGB algorithm is a black box method that lacks interpretation of the derived results. However, managers would like to understand how the individual predictors drive the trigger event probability so that they can implement appropriate actions.

We use a surrogate model (e.g., Burkart and Huber 2021) to gain insight into the effect size and direction of predictor variables. A surrogate model (e.g., a linear regression) is a simpler model to interpret than a black box model. Note that its results must not be interpreted causally. We use the predicted probability of the XGB algorithm for the training observations as the dependent variable. As independent variables, we include the 14 most important variables from the XGB model (see Appendix G for details on the selection of variables). We also include relevant interaction terms between the marketing-related variables and company variables to mimic the XGB algorithm more realistically and to gain additional substantive insights. We follow Stäbler and Fischer's (2020) approach to identify relevant interaction effects (see Appendix G for details).

Table 9 shows the results of this regression. We report the standardized coefficients to ensure comparability between the predictors. Even though this is not a causal model, it offers valuable insights into the relative contribution of variables to the prediction outcome. In general, we observe that variables from all four analyzed predictor categories (i.e., news event history, marketing, company, and environmental) matter for describing the trigger event probability.

Table 9: Surrogate Model

Dependent variable Predicted trigger probability (XGB)	Measure-ment unit	Standardized coefficients	[95% Confidence interval]	Variance inflation factor
News event history				
Recency of news event	week	**−.100**	[−.104, −.096]	3.833
Frequency of news events	week	**.352**	[.346, .357]	7.388
Recency of trigger event	week	**−.189**	[−.193, −.184]	3.796
Frequency of trigger events	week	**.141**	[.136, .146]	6.365
Marketing-related variables				
Advertising budget	month	**.162**	[.156, .169]	9.806
Relationship equity	year	.082	[.076, .088]	8.522
R&D expenditures	year	.033	[.028, .038]	5.935
SG&A expenses	year	.018	[.012, .023]	7.009
Other important variables				
Stock return	week	−.002	[−.005, $-.273\times10^{-03}$]	1.000
Sales	year	.077	[.073, .081]	4.450
Firm size	year	**.210**	[.206, .214]	4.529
Market share	year	.072	[.066, .078]	9.050
Leverage	year	$.116\times10^{-03}$	[−.002, .002]	1.012
Google search	month	.019	[.016, .021]	1.083
Relevant interactions between marketing and company-related variables[a]				
Advertising budget × Sales	month	**−.189**	[−.194, −.183]	6.666
Advertising budget × Firm size	month	−.009	[−.015, −.004]	6.563
Advertising budget × Market sh.	month	−.014	[−.017, −.010]	2.916
Advertising budget × Leverage	month	.062	[.057, .066]	4.652
Relationship equity × Sales	year	.042	[.037, .047]	4.903
Relationship equity × Firm size	year	−.054	[−.058, −.050]	3.908
Relationship equity × Market sh.	year	−.076	[−.083, −.070]	9.329
R&D expenditures × Sales	year	−.082	[−.086, −.078]	4.203
R&D expenditures × Market sh.	year	.009	[.004, .014]	6.425
SG&A expenses × Sales	year	.012	[.008, .016]	3.448
SG&A expenses × Firm size	year	−.009	[−.093, −.087]	2.051
SG&A expenses × Market share	year	−.028	[−.030, −.026]	1.384
SG&A expenses × Leverage	year	.008	[.002, .013]	7.101
Adjusted R-square / # observations		.709 / 255,636		

Notes: The seven largest effects are indicated in boldface.
[a] The selection of relevant interactions is explained in Appendix G.

The news event history predictors have by far the largest influence on the probability, similar to the fields of crime prediction (e.g., Mohler et al. 2011), earthquake prediction (e.g., Mignan and Broccardo 2020), and churn prediction (e.g., Schmittlein, Morrison, and Colombo 1987). Specifically, the information about the number of negative news events in the previous six months (frequency of news events), the time since the last trigger event (recency of trigger event), the number of trigger events in the previous six months (frequency of trigger events), and the time since the last negative news event (recency of news event) explain the trigger probability from the XGB prediction. Importantly, it is not only the past trigger events but the history of prior negative publicity that counts.

Among the other variables, the size of the firm and advertising budget stand out as most relevant predictors. This is not surprising but in line with Stäbler and Fischer's (2020) finding that media prefer to report unethical behavior of firms that are large, possess well-known brands, and show a higher advertising presence. All these factors drive attention to these firms, which is also reflected in the positive influence of the Google search index on the predicted probability.

The interaction terms offer additional substantive insights about the effect of marketing variables on the predicted trigger event probability. Most importantly, the influence of advertising is attenuated for larger firms with higher sales and market share.

6.3 "Early-Warning" System for Managers to Make Long-Term Forecasts

Next-week forecasts are hardly actionable for managers because they do not give them sufficient time to initiate and prepare appropriate actions in response to a trigger event. Consequently, managers are interested in long-term forecasts. Therefore, we suggest an early warning system for managers that makes a probabilistic forecast of the occurrence of a trigger event in the longer future. This warning system can yield substantive insights for managers as it helps them to initiate appropriate actions to diminish or avoid the negative consequences of a potential crisis in a timely manner.

6.3.1 The logic behind the system. The main purpose of the system is to provide managers with a continuously updated forecast of the time span within which the next trigger event is expected to occur with a given probability that we set at 80% for illustration. The system should be simple so that managers can understand and easily use it as well as leverage the outcomes of our ML-based prediction model. In fact, the weekly probability lends itself to being used as an alert index that is easy to follow and communicate within the organization. Monitoring this index helps management to better understand the conditions under which the company is exposed to higher negative publicity risk. The estimated trigger event probability combines the information content of the whole set of predictors in a single measure that we use as input in a hazard model to predict the time span until the next trigger event happens.

6.3.2 Alert index thresholds. By inspecting the evolution of trigger event probabilities, we learned that companies faced a higher frequency of trigger events if the index passed a certain threshold. This threshold, however, was not the same for each firm but varies across industries, firm sizes, and regions. To estimate firm-specific thresholds in these three categories based on the training and testing data, we obtained the average predicted trigger event probability that relates to a week in which an actual event occurred, as shown in Table 10. For each category, we then constructed an indicator variable that measures whether the firm's alert index has passed any of the three thresholds during the past 12 weeks. These indicators are the main variables in the hazard model and in the early warning system that uses the results of the hazard model for the probabilistic long-term forecast.

Model-free inspections of the data suggest that the identified thresholds are indeed informative. For example, the estimated probabilities of the fashion retailer Abercrombie & Fitch ranged from .34 to .71 between 2016 and 2017. However, weeks before mid-September 2016, when the firm was accused of firing and excluding from hiring Muslim women because of their religion, its estimated probabilities exceeded the firm size–, region-, and industry-specific thresholds. Crossing all three thresholds is associated with an 80% chance to face the next trigger event within the upcoming 3–4 months. Indeed, it took 5 months until the occurrence of the event. In contrast, the event management company Eventbrite, the German telecommunication provider Feenet AG, and the publishing company John Wiley & Sons never passed the thresholds of their industry, their region, or firm size in our observation period and were also never involved in a trigger event during that period. While the examples provide anecdotal evidence of the relevance of the three outlined threshold variables, our subsequent estimations derive precise estimates that can be used by managers to predict time periods in which future trigger events are likely to take place.

Table 10: Predicted Thresholds per Industry, Firm Size, and Region

Industry[a]	**Threshold (Prob. index in %)**	**Number of firm-weeks with trigger events**	**Firm size**[b]	**Threshold (Prob. index in %)**	**Number of firm-weeks with trigger events**
Department Stores and Warehouses	69.8	267	Super large	71.5	3,352
Wood and Chemical Manufacturing	69.0	987	Large	64.8	5,231
Utilities	68.6	568	Small-Medium	51.5	145
Finance and Insurance	68.2	1,599			
Mining	68.1	724			
Steel and Machine Manufacturing	67.8	2,074	**Region**		
Transportation and Warehousing	67.0	327	South Europe	71.8	585
Food and Textile Manufacturing	66.5	570	North Europe	70.6	1,559
Information	65.9	510	Asia	66.9	1,264
Wholesale Trade	65.1	190	North America	65.9	4,987
Grocery Retailer	61.6	205	Australia/ New Zeal.	65.8	133
Construction	60.8	148			
Administrative and Support Service	60.6	65	South America	60.3	101
Professional Services	57.7	92			
Accommodation and Food Service	57.7	102	Other	56.1	87
Other	51.4	125			

Notes: Threshold values are based on the average predicted event probability for a week with an actual trigger event within the respective category in the training and testing data sets. "Number of firm-week events" refers to the total number of trigger events within the respective training and testing sample.
[a] Industries are based on the two-digit NAICS code classification. However, if an industry includes less than 50 trigger events in the training and testing data, we include it in the industry "Other" to ensure that the created threshold values rely on enough observations. The industry classification in Table 10 thus differs from Table 2.
[b] For size classification, we distinguish between companies with up to 999 employees (small-medium), those with up to 99,999 employees (large), and those with 100,000 or more employees (super large). This classification into three size classes guarantees that each company size has at least 50 trigger events.

6.3.3 Hazard model specification. Based on the Bayesian information criterion, the Weibull distribution best describes our time spell data (see Appendix H). Hence, we use this distribution as underlying distribution for the hazard model. Following our prior logic, we calibrate the model only on the 495 firms in the holdout sample (in the period 2016–2017). Let t_{ik}

be the time spell between trigger events for event k of firm i. We estimate the following model on a weekly, rolling window basis:

$$f(t_{ikw}) = (\lambda_{ik} p)\,(\lambda_{ik} t)^{\,p-1} \exp(-(\lambda_{ik} t)^{p}, \quad \text{for } t_{ik}, \lambda_{ik}, p > 0 \qquad (7)$$

with $\lambda_{ik} = \exp[-(a + \boldsymbol{b}' \boldsymbol{Thresh}_{ikw}^{Prob} + c\, TriggerEventInThePast_{ikw})]$,

where $f(t_{ik})$ denotes the density of the time spell between trigger events, λ is the location parameter, p is the scale parameter that characterizes the moments of the distribution, and a, **b**, and c denote the parameter vectors to be estimated. $\boldsymbol{Thresh}_{ikw}^{Prob}$ is a vector that consists of three indicator variables, which take the value of 1 if the threshold was exceeded in the preceding 12 weeks and 0 otherwise. The underlying 'alert index' variable is the predicted probability $Prob_{ikw}$ using the focal XGB algorithm that changes from week to week. Hence, we estimate model 7 with a time-varying co-variate (Petersen 1986). Finally, we add the dummy variable $TriggerEventInThePast_{ikw}$ to our model, which indicates whether the focal firm i has been involved in a trigger event since 2007. The trigger event history is probably an important structural characteristic that separates firms from each other.

6.3.4 Results of the hazard model. Table 11 summarizes the estimation results. If a firm and its associated event probability exceeds the industry-specific (−1.095, $p < .01$), firm size–specific (−.949, $p < .01$), or region-specific (−.545, $p < .05$) threshold within the past 12 weeks, it has a significant negative effect on the survival duration. Firms that have been associated with trigger events in the past also have significantly shorter survival durations (−3.634, $p < .01$).

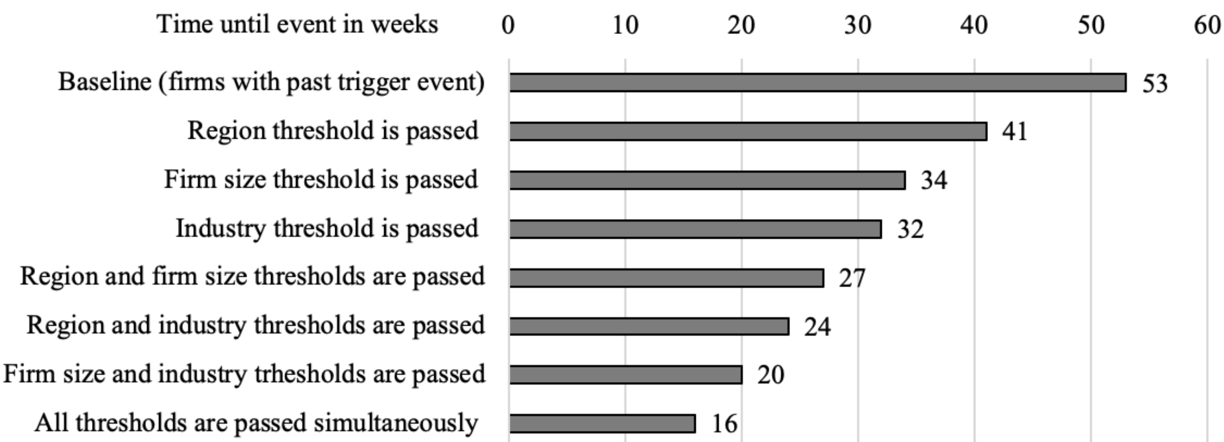
Table 11: Results of the Hazard Model

	Coefficient		S.E.
Constant (baseline)	8.379	***	.167
Trigger event in the past (Yes = 1, No = 0)	−3.634	***	.187
“Alert” index passed threshold of …			
Industry	−1.095	***	.211
Region	−.545	**	.249
Firm size	−.949	***	.235
1/Weibull scale parameter	1.21	***	.044
Log likelihood	−3,354		
Number of firms	495		
Number of observations	44,812		

Notes: To determine the duration until the next trigger event, we set a threshold (dummy variable) to 1 if the firm-relevant threshold was passed at least once in the proceeding 12 weeks. We use the first 12 weeks for initialization of the variables, and thus, exclude them from the investigation which reduces the holdout sample size to 44,812 observations.
** $p < .05$, *** $p < .01$ (two-sided t-test).

Figure 3 illustrates the application of the early warning system based on the estimates from the hazard model and for firms that experienced a past trigger event[3] in the period 2007–2015. In the base scenario, it takes more than a year until the next trigger event is expected to occur for these firms with a probability of 80%.

Figure 3: Estimated Duration until Next Trigger Event with a Given Probability of 80%



Notes: We simulated the effects by setting the respective threshold variable to 1 if passed, keeping the other variables at their empirical mean.

This time window reduces systematically when one, two, or all three firm-specific thresholds are passed. A trigger event is likely to occur after 7–9.5 months depending on which single

[3] The time spells for firms that have not experienced a trigger event, yet, are significantly longer. Most of these firms (91%) also do not face a trigger event in our holdout observation period (2016–2017).

threshold is passed. A combination of two passed thresholds reduces the time window further to 4.5–6 months. However, if all three thresholds are passed, there is an 80% chance that the next trigger event happens within the next 3–4 months.

7 Discussion

Negative media coverage of unethical corporate practices or product defects can trigger a serious corporate crisis with severe financial consequences. We introduce an ML-based prediction model that successfully predicts the probability of facing such a trigger event. The results stem from a rich and unique data set of 3,271 companies across the globe and various data sources.

The model performance we achieve with the extreme gradient boosting algorithm is impressive. Applying our cost-optimal classification threshold, the model correctly identifies 8.6 out of 10 trigger events per week with a balanced accuracy of about 85.5%. This performance compares very well with ML-based models in earthquake prediction that rarely achieve a balanced accuracy rate better than 75% (Mignan and Broccardo 2020).

7.1 Contributions to Theory and Practice

Our study offers practical, substantive, and methodological contributions that extend various literature streams. We add to the large field of research that deals with the social responsibility of firms and with corporate unethical or questionable behavior that often results into firm crises. While many important questions have been studied so far, they almost exclusively deal with the consequences of firm crisis events (e.g., Gao et al. 2015; Van Heerde, Helsen, and Dekimpe 2007). There is an obvious need to better understand the conditions from which these events arise and how to predict them. Our study offers an important contribution to satisfy this need. While we do not claim to have modeled causal drivers of the occurrence of trigger events, the prediction model is a relevant first step into this direction and offers high practical value. The implications from the large body of existing event studies on the effects of negative firm events related to ESG and product issues provide valuable insights for investors but they are not directly actionable. We offer an actionable model that investors can use to revise and fine-tune their investment strategies. The weekly model predictions help avoiding financial losses. But the model could also serve as a basis for developing active investment strategies.

In addition, we have demonstrated how the weekly prediction outcome can be integrated into a an early-warning system for management. Prior research has suggested various measures and practices that firms should follow to reduce the likelihood of firm scandals and crises (e.g., Rubel, Naik, and Srinivasan 2011). We cannot derive recommendations for specific firm actions based on our study. However, the proposed pre-warning system provides a valuable diagnostic tool for firms. It helps them to anticipate the future risks on a quantitative basis via the probabilistic long-term forecasts. In addition, management should learn from monitoring the alert index over time. The evolution of the predicted weekly trigger event probability helps managers to better understand the conditions under which the company is exposed to a higher negative publicity risk, for example, by actively searching for explanations for an observed increasing trend. The transparency created by our model offers a ground for connecting new and possibly overseen strands of interpretation. The alert index is also an easy instrument to communicate potential risks within the organization.

Methodologically, we add to the research on predicting rare events. Our findings on the importance of the event history as source for the predictions is very much in line with findings from prediction models across disciplines as different as earthquakes, customer churn, and crimes. We find that XGB is the most promising approach for our extremely imbalanced prediction task, as it outperforms all other methods. This insight confirms previous studies that also identify gradient boosting as the most powerful prediction approach in many application cases (e.g., Olson et al. 2018). However, we do not want to generalize this finding as other prediction tasks might be better addressed with other approaches such as neural networks that are the dominant ML approach used in earthquake prediction. We rather recommend to test a variety of alternative models including simpler models such as logistic regression, which turns out to be second-best in our application.

Substantively, we conclude that without any history, it is probably hard to come up with a good prediction – a fact, which is quite well known from other prediction tasks. Prediction models for burglary and robbery crimes, as an example, are not easily transferable to districts that are not known for a high crime intensity (e.g., Johnson et al. 2007). Our model shares this limitation and is probably best suited for a constantly rising number of firms that already faced a trigger event. Nevertheless, we identified several other important predictors that do not depend on the event history and improve the evaluation and prediction of potential risks.

Among those variables are variables that allude to the popularity of firms which correlates with their size, competitive strength, and marketing power and which is reflected in online searches for the firm. Marketing investments in brand (advertising budget) and customers (relationship equity) turn out to have a side effect as they increase the marketing power of the firm but also increase its attraction to those who look for signs of firm behavior that violates ethical principles and norms. This finding is in line with prior research on the mechanisms of the media business and that provides a theoretical explanation for drivers of media coverage of negative firm behavior (Stäbler and Fischer 2020). It also suggests that the disclosed trigger events do not necessarily must have an evidence base. Our study thus adds to the emerging stream of research that tries to understand how marketing and firm relevant information propagate within and across the various media channels (e.g., Shi, Liu, and Srinivasan 2022).

7.2 Limitations and Further Research Avenues

Limitations of this paper offer fruitful research opportunities. First, while our study uses a wide range of variables to precisely predict the occurrence of a trigger event, we do not provide causal evidence on the direction of their effect. Although the surrogate model provides initial indications of how the predictors drive the probability of the XGB, future research should also investigate how managers can influence potential drivers to diminish the likelihood of trigger events and accompanying corporate crises. These analyses do not have to be limited to structured data, as in our case; unstructured data in the form of text and image analyses could also be considered.

Second, our research indicates that users may determine an individual threshold that classifies the predicted risk index into zeros and ones. The question that then arises is how investors may decide on a threshold – which has not been investigated in the literature. Future research could explore the underlying mechanisms on how investors identify a threshold and quantify the perceived error costs.

REFERENCES PAPER 1

Asim, Khawaja M., Adnan Idris, Talat Iqbal T, and Francisco Martínez-Álvarez (2018), "Earthquake Prediction Model Using Support Vector Regressor and Hybrid Neural Networks," *PLoS ONE*, 13 (7), https://doi.org/10.1371/journal.pone.0199004.

Barth, Mary E., William H. Beaver, and Wayne R. Landsman (1998), "Relative Valuation Roles of Equity Book Value and Net Income as a Function of Financial Health," *Journal of Accounting and Economics*, 25 (1), 1–34.

Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science*, 29 (5), 815–27.

Borah, Abhishek and Gerard J. Tellis (2016), "Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands?" *Journal of Marketing Research*, 53 (2), 143–60.

Branco, Paula, Luís Torgo, and Rita P. Ribeiro (2016), "A Survey of Predictive Modeling on Imbalanced Domains," *ACM Computing Surveys*, 49 (2), https://doi.org/10.1145/2907070.

Breiman, Leo (2001), "Random Forests," *Machine Learning*, 45 (1), 5–32.

Brown, Marvin L. and John F. Kros (2003), "Data Mining and the Impact of Missing Data," *Industrial Management & Data Systems*, 103 (8), 611–21.

Burkart, Nadia and Marco F. Huber (2021), "A Survey on the Explainability of Supervised Machine Learning," *Journal of Artificial Intelligence Research*, 70, 245–317.

Campbell, Dennis W. and Ruidi Shang (2022), "Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews," *Management Science*, 68 (9), 7034–53.

Chang, Yung-Chia, Kuei-Hu Chang, and Guan-Jhih Wu (2018), "Application of eXtreme Gradient Boosting Trees in the Construction of Credit Risk Assessment Models for Financial Institutions," *Applied Soft Computing*, 73 (December), 914–20.

Chen, Tianqi and Carlos Guestrin (2016), "XGBoost: A Scalable Tree Boosting System," in *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 785–94.

Chen, Xinyu, Youngwoon Cho, and Suk Young Jang (2015), "Crime Prediction Using Twitter Sentiment and Weather," in *2015 Systems and Information Engineering Design Symposium*, 63–68.

Cleeren, Kathleen, Harald J. van Heerde, and Marnik G. Dekimpe (2013), "Rising from the Ashes: How Brands and Categories Can Overcome Product-harm Crises," *Journal of Marketing*, 77 (2), 58–77.

Devi, S. Sarojini and Y. Radhika (2018), "A Survey on Machine Learning and Statistical Techniques in Bankruptcy Prediction," *International Journal of Machine Learning and Computing*, 8 (2), 133–39.

DeVries, Phoebe M. R., Fernanda Viégas, Martin Wattenberg, and Brendan J. Meade (2018), "Deep Learning of Aftershock Patterns Following Large Earthquakes," *Nature*, 560 (7720), 632–34.

Dinner, Isaac M., Tarun Kushwaha, and Jan-Benedict E. M. Steenkamp (2019), "Psychic Distance and Performance of MNCs During Marketing Crises," *Journal of International Business Studies,* 50 (3), 339–64.

Donkers, Bas, Philip Hans Franses, and Peter C. Verhoef (2003), "Selective Sampling for Binary Choice Models," *Journal of Marketing Research*, 40 (4), 492–97.

Du, Rex Yuxing and Wagner A. Kamakura (2012), "Quantitative Trendspotting," *Journal of Marketing Research*, 49 (4), 514–36.

Duanmu, Jing-Lin, Maoliang Bu, and Russell Pittman (2018), "Does Market Competition Dampen Environmental Performance? Evidence from China," *Strategic Management Journal*, 39 (11), 3006–30.

Dupire, Marion and Bouchra M'Zali (2018), "CSR Strategies in Response to Competitive Pressures," *Journal of Business Ethics*, 148 (3), 603–23.

Elkan, Charles (2001), "The Foundations of Cost-Sensitive Learning," *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 17 (1), 973–78.

Frennea, Carly, Kyuhong Han, and Vikas Mittal (2019), "Value Appropriation and Firm Shareholder Value: Role of Advertising and Receivables Management," *Journal of Marketing Research*, 56 (2), 291–309.

Freund, Friedemann, Guy Ouillon, John Scoville, and Didier Sornette (2021), "Earthquake Precursors in the Light of Peroxy Defects Theory: Critical Review of Systematic Observations," *The European Physical Journal Special Topics*, 230 (1), 7–46.

Gao, Haibing, Jinhong Xie, Qi Wang, and Kenneth C. Wilbur (2015), "Should Ad Spending Increase or Decrease Before a Recall Announcement? The Marketing–Finance Interface in Product-Harm Crisis Management," *Journal of Marketing*, 79 (5), 80–99.

Gijsenberg, Maarten J., Harald J. van Heerde, and Peter C. Verhoef, (2015), "Losses Loom Longer Than Gains: Modeling the Impact of Service Crises on Perceived Service Quality Over Time," *Journal of Marketing Research*, 52 (5), 642–56.

Hanlon, Michelle and Joel Slemrod (2009), "What Does Tax Aggressiveness Signal? Evidence from Stock Price Reactions to News about Tax Shelter Involvement," *Journal of Public Economics*, 93 (1–2), 126–41.

Harte, David S. (2016), "Model Parameter Estimation Bias Induced by Earthquake Magnitude Cut-Off," *Geophysical Journal International*, 204 (2), 1266–87.

Johnson, Shane D., Wim Bernasco, Kate J. Bowers, Henk Elffers, Jerry Ratcliffe, Georg Rengert, et al. (2007), "Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization," *Journal of Quantitative Criminology*, 23 (3), 201–19.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–91.

Kalyanaram, Gurumurthy and Russell S. Winer (2022), "Behavioral Response to Price: Data-based Insights and Future Research for Retailing," *Journal of Retailing*, 98 (1), 46–70.

Kang, Charles, Frank Germann, and Rajdeep Grewal (2016), "Washing Away Your Sins? Corporate Social Responsibility, Corporate Social Irresponsibility, and Firm Performance," *Journal of Marketing*, 80 (2), 59–79.

Kuhn, Max and Kjell Johnson (2013), *Applied Predictive Modelling*, Vol. 26. Springer.

Lee, Dongyoung (2020), "Corporate Social Responsibility of US-Listed Firms Headquartered in Tax Havens," *Strategic Management Journal*, 41 (9), 1547–71.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (2), 276–86.

Lemmens, Aurélie and Sunil Gupta (2020), "Managing Churn to Maximize Profits," *Marketing Science*, 39 (5), 956–73.

Lepori, Gabriele M. (2016), "Air Pollution and Stock Returns: Evidence from a Natural Experiment," *Journal of Empirical Finance*, 35 (January), 25–42.

Li, Loic Pengtao, Biljana Juric, and Roderick J. Brodie (2017), "Dynamic Multi-Actor Engagement in Networks: the Case of United Breaks Guitars," *Journal of Service Theory and Practice*, 27 (4), 738–60.

Liu, Yan and Venkatesh Shankar (2015), "The Dynamic Impact of Product-Harm Crises on Brand Preference and Advertising Effectiveness: An Empirical Analysis of the Automobile Industry," *Management Science*, 61 (10), 2514–35.

Mahajan, Vijay, Raji Srinivasan, and Jerry Wind (2002), " The Dot.com Retail Failures of 2000: Were There Any Winners?" *Journal of the Academy of Marketing Science*, 30 (4), 474–86.

Marsan, David and Olivier Lengliné (2008), "Extending Earthquakes' Reach Through Cascading," *Science*, 319 (5866), 1076–79.

Mazodier, Marc and Amir Rezaee (2013), "Are Sponsorship Announcements Good News for the Shareholders? Evidence from International Stock Exchanges," *Journal of the Academy of Marketing Science*, 41 (5), 586–600.

Mignan, Arnaud and Marco Broccardo (2020), "Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on their Limitations," *Seismological Research Letters*, 91 (4), 2330–42.

Mishina, Yuri, Bernadine J. Dykes, Emily S. Block, and Timothy G. Pollock (2010), "Why "Good" Firms Do Bad Things: The Effects of High Aspirations, High Expectations, and Prominence on the Incidence of Corporate Illegality," *Academy of Management Journal*, 53 (4), 701–22.

Mohler, George O., Martin B. Short, P. Jeffrey Brantingham, Frederic P. Schoenberg, and George E. Tita (2011), "Self-Exciting Point Process Modeling of Crime," *Journal of the American Statistical Association*, 106 (493), 100–108.

Morgan, Neil A. and Lopo L. Rego (2009), "Brand Portfolio Strategy and Firm Performance," *Journal of Marketing*, 73 (1), 59–74.

Naumzik, Christof, Stefan Feuerriegel, and Markus Weinmann (2022), "I Will Survive: Predicting Business Failures from Customer Ratings," *Marketing Science*, 41 (1), 188–207.

Olson, Randal S., William La Cava, Patryk Orzechowski, Ryan J. Urbanowicz, and Jason H. Moore (2017), "PMLB: a Large Benchmark Suite for Machine Learning Evaluation and Comparison," *BioData Mining*, 10 (1), 1–13.

Olson, Randal S., William La Cava, Zairah Mustahsan, Akshay Varik, and Jason H. Moore (2018), "Data-Driven Advice for Applying Machine Learning to Bioinformatics Problems," in *Pacific Symposium on Biocomputing 2018*, Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Marylyn D. Ritchie, Tiffany A. Murray, and Teri E. Klein, eds. World Scientific, 192–203.

Petersen, Trond (1986), "Estimating Fully Parametric Hazard Rate Models with Time-Dependent Covariates," *Sociological Methods and Research*, 14, 219–46.

RepRisk (2023), "RepRisk Methodology Overview," (accessed June 15th, 2023) [available at https://www.reprisk.com/news-research/resources/methodology].

Roehm, Michelle L. and Alice M. Tybout (2006), "When Will a Brand Scandal Spill Over, and How Should Competitors Respond?" *Journal of Marketing Research*, 43 (3), 366–73.

Rubel, Olivier, Prasad A. Naik, and Shuba Srinivasan (2011), "Optimal Advertising When Envisioning a Product-Harm Crisis," *Marketing Science*, 30 (6), 1048–65.

Rust, Roland T., Tim Ambler, Gregory S. Carpenter, V. Kumar, and Rajendra K. Srivastava (2004), "Measuring Marketing Productivity: Current Knowledge and Future Directions," *Journal of Marketing*, 68 (4), 76–89.

Saito, Takaya and Marc Rehmsmeier (2015), "The Precision-Recall Plot is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS ONE*, 10 (3), https://doi.org/10.1371/journal.pone.0118432.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?," *Management Science*, 33 (1), 1–24.

Sheng, Victor S. and Charles X. Ling (2006), "Thresholding for Making Classifiers Cost-Sensitive," *American Association for Artificial Intelligence*, 6 (July), 476–81

Shi, Zijun, Xiao Liu, and Kannan Srinivasan (2022), "Hype News Diffusion and Risk of Misinformation: the Oz Effect in Health Care," *Journal of Marketing Research*, 59 (2), 327–352.

Siano, Alfonso, Agostino Vollero, Francesca Conte, and Sara Amabile (2017), ""More Than Words": Expanding the Taxonomy of Greenwashing After the Volkswagen Scandal," *Journal of Business Research*, 71 (February), 27–37.

Sismeiro, Catarina and Randolph E. Bucklin (2004), "Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach," *Journal of Marketing Research*, 41 (3), 306–23.

Stäbler, Samuel and Marc Fischer (2020), "When Does Corporate Social Irresponsibility Become News? Evidence from More Than 1,000 Brand Transgressions Across Five Countries," *Journal of Marketing*, 84 (3), 46–67.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis (2008), "Conditional Variable Importance for Random Forests," *BMC Bioinformatics*, 9 (1), 1–11.

Tybout, Alice M., Bobby J. Calder, and Brian Sternthal (1981), "Using Information Processing Theory to Design Marketing Strategies," *Journal of Marketing Research*, 18 (1), 73–79.

Van Heerde, Harald J., Kristiaan Helsen, and Marnik G. Dekimpe (2007), "The Impact of a Product-Harm Crisis on Marketing Effectiveness," *Marketing Science*, 26 (2), 230–45.

Zhao, Yi, Ying Zhao, and Kristiaan Helsen (2011), "Consumer Learning in a Turbulent Market Environment: Modeling Consumer Choice Dynamics After a Product-Harm Crisis," *Journal of Marketing Research*, 48 (2), 255–67.

APPENDIX PAPER 1

APPENDIX A: REPRISK'S RESEARCH SCOPE

The data provider RepRisk covers 28 issues (RepRisk 2023). These issues can be categorized as environment, social, governance, and cross-cutting issues as illustrated in Table A1. Every negative news incident covered by RepRisk is linked to at least one of these 28 issues. According to RepRisk the issues are "selected and defined in accordance with the key international standards related to ESG issues and business conduct, such as the World Bank Group Environmental, Health, and Safety Guidelines, the IFC Performance Standards, the Equator Principles, the OECD Guidelines for Multinational Enterprises, the ILO Conventions, and more" (RepRisk 2023).

Table A1: 28 Issues Covered by RepRisk

Environment	Social		Governance
Environmental Footprint	**Community Relations**	**Employee Relations**	**Corporate Governance**
Climate change, GHG emission, and global pollution	Human rights abuses and corporate complicity	Forced labor	Corruption, bribery, extortion, money laundering
Local pollution	Impacts on communities	Child labor	Executive compensation issues
Impacts on landscapes, ecosystems, and biodiversity	Local participation issues	Freedom of association and collective bargaining	Misleading communication
Overuse and wasting of resources	Social discrimination	Discrimination in employment	Fraud
Waste issues		Occupational health and safety issues	Tax evasion
Animal mistreatment		Poor employment conditions	Tax optimization
			Anti-competitive practices
Cross-cutting Issues			
Controversial products and services			
Products (health and environmental issues)			
Supply chain issues			
Violation of national legislation			
Violation of international standards			

In addition, RepRisk uses different topic tags, which complement the 28 issues. These topic tags are more specific than the broader issues and they refer to a concrete theme. Every topic tag can be linked to multiple issues. In contrast to the 28 issues, the topics tags are dynamic, which means that the list of topics expands over time in response to emerging trends. Table A2 overviews the topic tags.

Table A2: Topic Tags of RepRisk

Abusive/Illegal fishing	Agricultural commodity speculation	Alcohol	Animal transportation	Arctic drilling
Asbestos	Automatic and semi-automatic weapons	Cluster munitions	Coal-fired power plants	Conflict minerals
Coral reefs	Deep sea drilling	Depleted uranium munitions	Diamonds	Drones
Endangered species	Forest burning	Fracking	Gambling	Genetically modified organisms (GMOs)
Genocide/Ethnic cleansing	High conservation value forests	Human trafficking	Hydropower (dams)	Illegal logging
Indigenous people	Involuntary resettlements	Land grabbing	Land mines	Migrant labor
Monocultures	Mountaintop removal mining	Negligence	Nuclear power	Oil sands
Palm oil	Pornography	Predatory lending	Privacy violations	Protected areas
Rare earths	Sea-bed mining	Soy	Tobacco	Water scarcity

Notes: Illustrated topic tags refer to the analyzed time window between 2007-2017. As these topics are dynamic the list is constantly updated by RepRisk. In 2023, the list contained 74 topics (RepRisk 2023).

APPENDIX B: PERFORMANCE ASSESSMENT CRITERIA

The performance metrics used in our study are all based on the confusion matrix resulting from a binary classification. Table A3 represents this matrix (see Davis and Goadrich 2006 for a detailed explanation). Binary classifications produce true positive (TP) predictions when an event is correctly predicted and true negative (TN) predictions when an event is correctly not predicted. In addition, two different errors occur: a false positive (FP) result occurs when an event is falsely predicted and a false negative (FN) result occurs when an event is falsely not predicted.

Table A3: Outcomes for a Binary Classification Task

		Actual class	
		Trigger event	No trigger event
Predicted class	Trigger event	**True positives**	**False positives**
	No trigger event	**False negatives**	**True negatives**

The classification from Table A3 forms the basis for all performance metrics used in the study. Table A4 shows the metrics addressed in the study and their formula. Note that different formulas exist for some metrics (e.g., for the F_1-score).

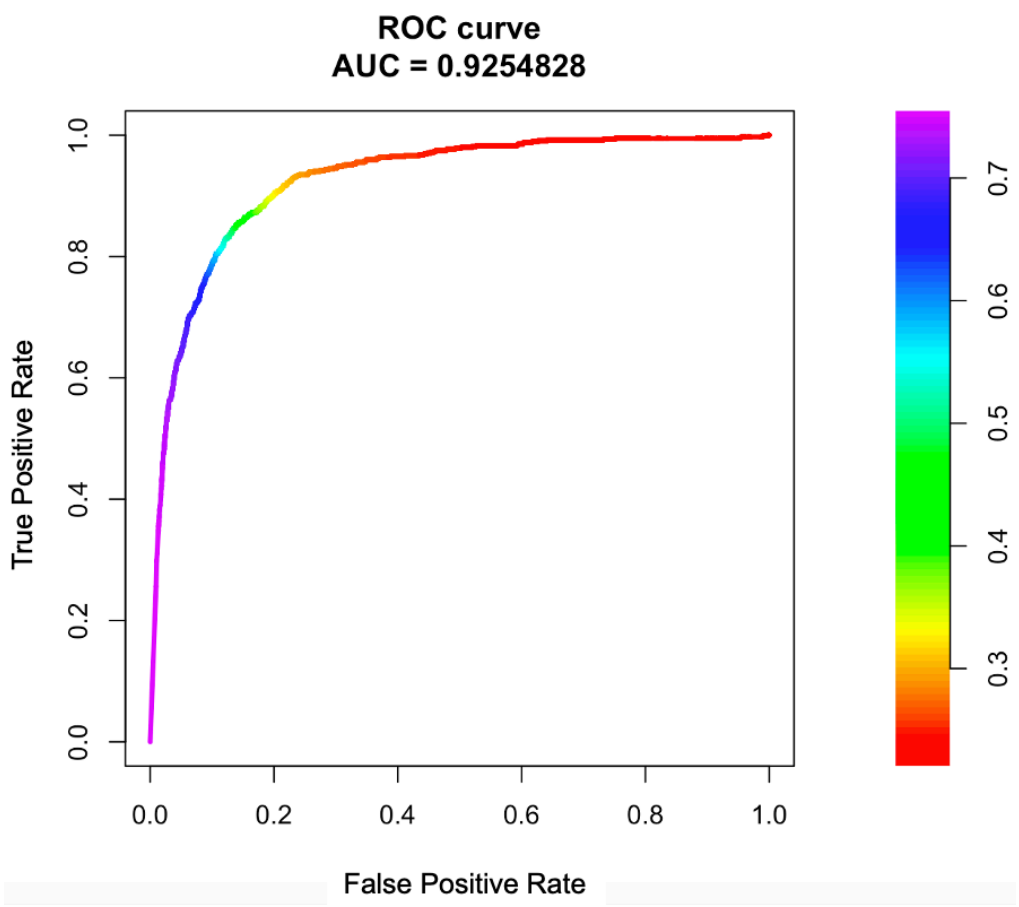
Table A4: Overview Performance Metrics

Accuracy	= (TP + TN) / (TP + TN + FP + FN)
Sensitivity/Recall (TPR)	= TP / (TP + FN)
Specificity (TNR)	= TN / (TN + FP)
Precision	= TP / (TP + FP)
R score	= sensitivity + specificity – 1
Balanced accuracy	= (sensitivity + specificity) / 2
F_1-score	= 2 X ((Precision x Recall) / (Precision + Recall))
False positive rate (FPR)	= FP / (FP + TN)
False negative rate (FNR)	= FN / (FN + TP)

Notes: True Positives = TP; False Positives = FP; True Negatives = TN; False Negatives = FN; TPR = True Positive Rate; TNR = True Negative Rate; FPR = False Positive Rate; FNR = False Negative Rate.

In addition, we analyze the area under the curve (AUC*)* of the receiver operating characteristic curve (ROC) curve and the precision recall (PR) curve. The ROC curve is an evaluation metric for binary classification problems (Lantz 2019, p. 312). It is a probability curve that plots the True Positive Rate (TPR) against the False Positive Rate (FPR) (1 – specificity) at various threshold values. Figure A1 shows the ROC curve of our XGB algorithm for the holdout sample. The different colors represent the different threshold values.

Figure A1: ROC Curve and AUC Value for the XGB Prediction

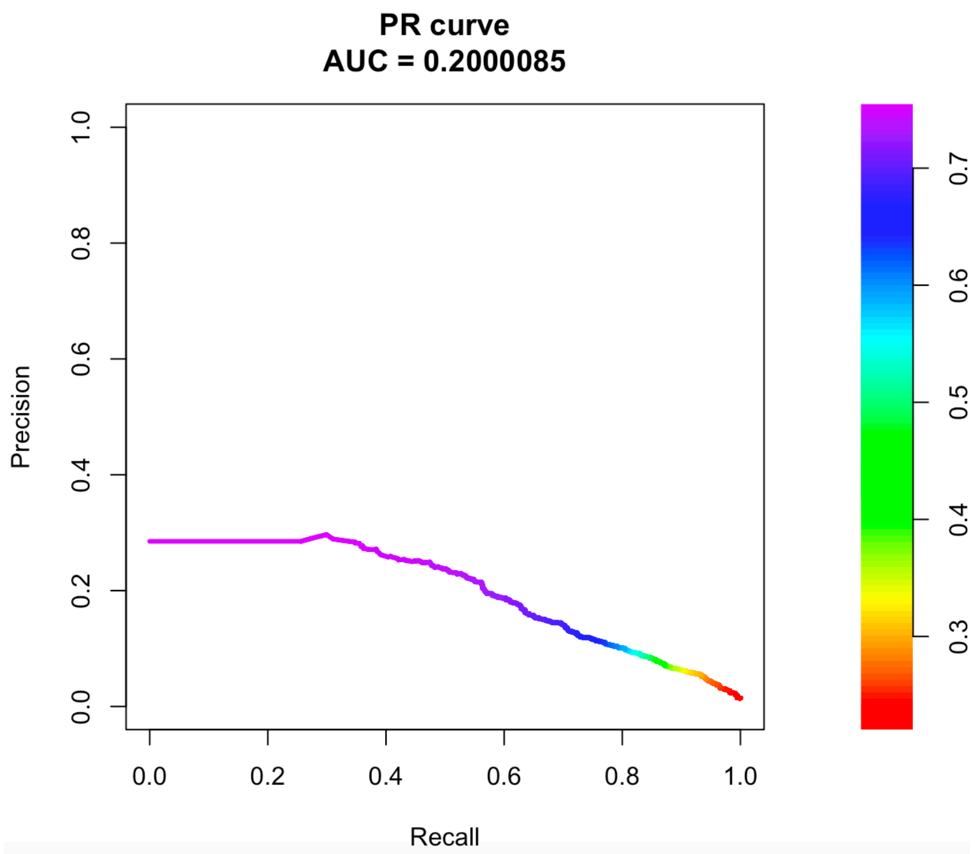


The ROC AUC measures the entire two-dimensional area underneath the ROC curve. Thus, it provides an aggregate measure of performance across all possible classification thresholds. The metric is considered as classification-threshold-invariant, because it measures the quality of the model's predictions irrespective of what classification threshold is chosen.

The maximum possible AUC value is 1 (100%) and the lowest value is 0 (0%). The higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. A model with 100% wrong predictions has an AUC of 0. A model whose predictions are 100% correct achieves an AUC of 1. An AUC value of .5 means the model has no class separation power at all (random classification). In case the AUC value of the model is lower than .5, the model performance is worse than randomly assigning the classes.

The PR curve and the corresponding AUC value are comparable to the ROC curve. However, this graph plots the Precision against the Recall like illustrated in Figure A2 (Boyd et al. 2012).

Figure A2: PR Curve and AUC Value for the XGB Prediction



This metric is specifically useful when data is heavily imbalanced (see Saito and Rehmsmeier 2015) since PR AUC focuses mainly on the positive class and cares less about the frequent negative class.

Unlike the ROC curve, the AUC value for a random classification is not automatically .5, but is based on the skewness of the binary target variable. For example, if the target variable contains 20% 1 and 80% 0, then the AUC corresponds to the PR curve .2 for a random classification (see for example Boyd et al. 2012). In this scenario, if an algorithm achieves a value higher .2, then it classifies better than a random classifier. Recall that in our study, the trigger events represent only 1.45% of all observations in the holdout sample. Thus, a random classifier would achieve a PR AUC value of .0145.

APPENDIX C: WEATHER DATA

Table A5: Weather Variables

Variable	Operationalization	Source	Unit measure	Temporal interval	Number of time lags used	Mean	Median	SD
Weather conditions	Weather conditions at headquarter location	Visual-crossing		weekly	1-2			
	- mean temperature		in celsius			14.92	15.69	9.18
	- maximum temperature		in celsius			23.41	24.80	9.02
	- minimum temperature		in celsius			7.48	8.00	10.23
	- snow depth		in centimeters			167.00	23.07	206.51
	- wind speed		in kilometers per hour			21.48	21.03	6.31
	- wind chill		in celsius			.67	2.53	6.89
	- wind gust		in kilometers per hour			54.92	53.60	17.78
	- visibility		in kilometers			14.34	14.47	5.30
	- relative humidity		ratio in %			67.42	68.72	12.34
	- precipitation		in millimeter			3.13	1.31	5.58
	- heat		in celsius			31.36	30.15	4.01
	- cloudiness		ratio in %			38.37	36.69	23.21

The weather data contains 12 different variables shown in Table A5. These variables describe the weather situation at the headquarter location of the company, which is stated in Compustat. In case there is no weather data available for the concrete location, we collect the weather data from the closest possible location to the headquarter available. More information is provided here: https://www.visualcrossing.com.

Appendix D: Description of Algorithms

Table A6: Applied Machine Learning Approaches

Approach	Model explanation	Exemplary sources
Artificial neural network[1]	- Models the relationship between a set of input signals and an output signal using a model derived from how a brain responds to stimuli from sensory inputs. - A typical artificial neuron with *n* inputs is represented by: $$y(x) = f\left(\sum_{i=1}^{n}(w_i x_i)\right).$$ - The *w* weights define how much each of the *n* inputs, (x), contributes to the sum of all input signals. The net total is used by the activation function *f(x)* and the resulting signal, *y(x)*, is the output axon. - We use the resilient backpropagation with weight backtracking as algorithm (Riedmiller 1994). - We use a sigmoid function as an activation function for the neural network: $$f(x) = \frac{1}{1+e^{-x}}$$ - Our artificial neural network includes two hidden layers. The first hidden layer consists of 12 hidden neurons and the second layer consists of 2 hidden neurons.	Lantz 2019; Kuhn and Johnson 2013; Riedmiller 1994
Random forest[2]	- Group of un-pruned classification made by randomly selecting samples and predictors of the training data. - "classifier consisting of a collection of tree-structured classifiers $\{h(\mathrm{x}, \Theta_k), k = 1, \ldots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x" (Breiman 2001, p. 6). - Ensemble-based method, in which final prediction is made by a majority vote of trees. - Most relevant tuning parameter is the number of randomly selected predictors to choose from at each split (in our case = 10). We train 500 trees in our random forest model. The minimum size of terminal nodes equals 5 and the maximum number of terminal nodes trees in the forest can have equals 10.	Breiman 2001; Lantz 2019

Table A6: Applied Machine Learning Approaches

Approach	Model explanation	Exemplary sources
XGBoost[3]	- Scalable tree boosting system that achieves best prediction performance for many classification problems: For example, XGBoost is a superior tool for the development of credit risk models (Chang, Chang, and Wu 2018). - In contrast to random forest, the trees are not trained independently but each tree incrementally incorporates and corrects the error produced by the previously trained tree. - Demands for tuning of several parameters, which may impact the performance. We apply a random search approach to identify the optimal combination of parameter values. Although grid search and manual search are the most widely used strategies for hyperparameter optimization, randomly chosen trials may be more efficient (Bergstra and Bengio 2012).	Bergstra and Bengio 2012; Chen and Guestrin 2016; Chang et al. 2018; Olson et al. 2018

Notes: [1] https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf leads to the R package used for the neural network.
[2] https://cran.r-project.org/web/packages/randomForest/randomForest.pdf leads to the R package used for the random forest.
[3] https://cran.r-project.org/web/packages/xgboost/xgboost.pdf leads to the R package used for the XGBoost.

Appendix E: Confusion Matrix for the Different Models

Table A7, Table A8, Table A9, and Table A10 present the confusion matrices for the four model classifications applying the optimal thresholds that maximize the weighted F_1-scores to the holdout sample. Table A11, Table A12, Table A13, and Table A14 overview the results applying the optimal thresholds that maximize the weighted balanced accuracy measure. Table A15 – Table A18 show the classification for a threshold value of 50%.

Table A7: Confusion Matrix for the XGB Classification (F_1-score Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	274	736	1,010
	No trigger event	464	49,278	49,742
	Sum	738	50,014	50,752

Notes: Classification threshold = 74.93%.

Table A8: Confusion Matrix for the ANN Classification (F_1-score Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	286	1,122	1,408
	No trigger event	452	48,892	49,344
	Sum	738	50,014	50,752

Notes: Classification threshold = 97.27%.

Table A9: Confusion Matrix for the Random Forest Classification (F_1-score Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	228	1,243	1,471
	No trigger event	510	48,771	49,281
	Sum	738	50,014	50,752

Notes: Classification threshold = 99.53%.

Table A10: Confusion Matrix for the Logit Classification (F_1-score Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	345	1,253	1,598
	No trigger event	393	48,761	49,154
	Sum	738	50,014	50,752

Notes: Classification threshold = 98.78%.

Table A11: Confusion Matrix for the XGB Classification (Bal. Accuracy Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	636	7,597	42,519
	No trigger event	102	42,417	8,233
	Sum	738	50,014	50,752

Notes: Classification threshold = 43.68%.

Table A12: Confusion Matrix for the ANN Classification (Bal. Accuracy Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	655	12,399	13,054
	No trigger event	83	37,615	37,698
	Sum	738	50,014	50,752

Notes: Classification threshold = 32.59%.

Table A13: Confusion Matrix for the Random Forest Classification (Bal. Accuracy Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	693	15,547	16,240
	No trigger event	45	34,467	34,512
	Sum	738	50,014	50,752

Notes: Classification threshold = 32.58%.

Table A14: Confusion Matrix for the Logit Classification (Bal. Accuracy Maximization)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	659	10,985	11,644
	No trigger event	79	39,029	39,108
	Sum	738	50,014	50,752

Notes: Classification threshold = 36.63%.

Table A15: Confusion Matrix for the XGB Classification (50%-Threshold)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	621	6,612	7,233
	No trigger event	117	43,402	43,519
	Sum	738	50,014	50,752

Notes: Classification threshold = 50.00%.

Table A16: Confusion Matrix for the ANN Classification (50%-Threshold)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	636	9,085	9,721
	No trigger event	102	40,929	41,031
	Sum	738	50,014	50,752

Notes: Classification threshold = 50.00%.

Table A17: Confusion Matrix for the Random Forest Classification (50%-Threshold)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	673	12,248	12,921
	No trigger event	65	37,766	37,831
	Sum	738	50,014	50,752

Notes: Classification threshold = 50.00%.

Table A18: Confusion Matrix for the Logit Classification (50%-Threshold)

		Actual trigger event		
		Trigger event	No trigger event	Sum
Predicted trigger event	Trigger event	625	7,738	8,363
	No trigger event	113	42,276	42,389
	Sum	738	50,014	50,752

Notes: Classification threshold = 50.00%.

In Table A19, we present the classification results using different performance metrics. All of these results base on the confusion matrices shown in Table A7 - Table A18.

Table A19: Different Performance Metrics for the Model Classifications

Thresholds that maximize the weighted F_1-score				
	XGB	**ANN**	**Random forest**	**Logit**
Accuracy	97.64%	96.90%	96.55%	96.76%
Sensitivity/Recall	37.13%	38.75%	30.89%	46.75%
Specificity	98.53%	97.76%	97.52%	97.50%
Precision	27.13%	20.31%	15.50%	21.59%
Balanced accuracy	67.83%	68.26%	64.21%	72.12%
F_1-score	31.35%	26.65%	20.64%	29.54%
Thresholds that maximize the weighted balanced accuracy				
	XGB	**ANN**	**Random forest**	**Logit**
Accuracy	84.83%	75.41%	69.28%	78.20%
Sensitivity/Recall	86.18%	88.75%	93.90%	89.30%
Specificity	84.81%	75.21%	68.92%	78.04%
Precision	7.73%	5.02%	4.27%	5.66%
Balanced accuracy	85.50%	82.00%	81.41%	83.67%
F_1-score	14.18%	9.50%	8.16%	10.64%
Traditional threshold of 50%				
	XGB	**ANN**	**Random forest**	**Logit**
Accuracy	86.74%	81.90%	75.74%	84.53%
Sensitivity/Recall	84.15%	86.18%	91.19%	84.69%
Specificity	86.78%	81.84%	75.51%	84.53%
Precision	8.59%	6.54%	5.21%	7.47%
Balanced accuracy	85.46%	84.01%	83.35%	84.61%
F_1-score	15.58%	12.16%	9.85%	13.74%

Appendix F: Cost-Benefit Analysis for Investors

Table A20 overviews the two scenarios and the calculations. In the baseline scenario (Scenario 1 in Table A20), we assume that our forecasting method is not applied. The shareholders are therefore not able to predict trigger events. On average, the abnormal stock returns (AR) in weeks with trigger events equals –2.72%. Stockholders would be exposed to this negative stock market development, as they would not predict the trigger events.

Table A20: Value Added for Investors: Simulation of Cumulated Financial Loss Across 495 Firms in Period 2016–2017

	Scenario 1: No risk forecast used	**Scenario 2: Risk forecast used**
Information needed		
Number of missed trigger events	738	102
Number of falsely predicted trigger events	0	7,597
Average abnormal stock return [AR] of missed trigger events	–2.72%	–3.08%
Average abnormal stock return of falsely predicted trigger events	-	0.29%
Average abnormal stock return of weeks for which no trigger event is predicted	-	0.18%
Average market capitalization [MC] of a company	US$35 billion	US$35 billion
Calculations		
Total costs of missing trigger events	**US$700 billion**	**US$110 billion**
Calculation of total costs of missing trigger events	738 trigger events x 2.72% [AR] x US$35 billion [MC]	102 trigger events x 3.08% [AR] x US$35 billion [MC]
Missed profits of falsely selling stocks	**US$0**	**US$771 billion**
Calculation of missed profits of falsely selling stocks		7,597 falsely predicted trigger events x 0.29% [AR] x US$35 billion [MC]
Additional profits of reinvestments	**US$0**	**US$479 billion**
Calculation of additional profits of reinvestments		7,597 x 0.18% [AR] x US$35 billion [MC]
Total economic costs	**US$700 billion**	**US$402 billion**
Calculation of total economic costs	US$700 billion + US$0 – US$0	US$110 billion + US$771 billion – US$479 billion
Total cost reduction in %	**42.57**	

Notes: All figures used for the calculations refer to the holdout sample observations.

Multiplying this negative abnormal return by the average market capitalization (MC) of the companies of about US$35 billion (Note: for illustrative purposes we calculate with average market capitalization of a company within the MSCI index) and number of trigger events results in total economic costs of US$700 billion.

In the second scenario, investors use the short-term prediction forecast. We exemplary apply the cost-optimal classification threshold of 43.68% from Table 7 using an error cost ratio of 2:1. At this threshold, only 102 of 738 trigger events are missed. These missed events lead to average negative abnormal returns of –3.08% which results in costs of US$110 billion due to false negative classifications. However, a second error is known to occur, in which a trigger event is incorrectly predicted although no event occurs (i.e., false positives). In this case, the investor may sell the shares although there is no actual negative impact due to trigger news. Thus, the investor loses the profits due to the mis-selling. In our scenario, this is the case for 7,597 observations. These cases show, on average, a positive abnormal return of .29% leading to missed profits of US$771 billion.

Since a rational investor would reinvest the money withdrawn, additional profits will be realized with reinvestments. In fact, our classification proposes alternative investment for which no trigger events are expected. The average abnormal returns for these alternative investments with no trigger event expected in the focal week equals .18%. These reinvestments lead to additional gains of US$479 billion.

To sum up, applying our short-term forecasting method leads to losses of US$110 billion and missed gains of US$771 billion because of the two different classification errors but also to new profits of US$479 billion due to reinvestments. Thus, the total economic costs can be decreased by 42.57% from US$700 billion without using the prediction tool to US$402 billion (110 + 771 – 479).

APPENDIX G: VARIABLE SELECTION FOR THE SURROGATE MODEL

Note that we do not impute missing values for the XGB approach because the algorithm can handle missing values. For the surrogate model (i.e., multiple regression), we use the predicted probability of the XGB from the training data set as dependent variable and the individual predictors as independent variables. However, since multiple regressions cannot handle missing values, we cannot include all predictor variables and need to make a selection. We only include the most important variables and the variables of special interest in the surrogate model. Otherwise, the number of missing values over the large number of predictors and time lags would be too high.

To help preselect the potentially most relevant variables, we use a feature of the XGB algorithm to evaluate variable importance (Chen and Guestrin 2016). This feature measures how useful and valuable each predictor variable is in constructing the decision trees within the model. The more often an attribute is used to arrive at important decisions, the higher is its relative importance. The importance is explicitly calculated for each attribute in the dataset, so that the attributes can be ranked and compared.

We consider the most important 25 pre-selected variables. Of these 25 variables, we exclude weather data because they contain a particularly large number of missing values and are also not actionable for managers.

Regardless of their importance for the XGB prediction, we include the four marketing variables, the four news events history variables (two recency and two frequency variables) as well as the google search variable due to our special interest in their direction of influence on the trigger event probability.

We only use the first time lag of each variable in case multiple time lags were used for the XGB. For example, we only used stock returns from the previous week in our multiple regression.

Further, we exclude profitability from the multiple regression because its VIF value is too high. In this way, we prevent multicollinearity between the independence variables from distorting the estimates. The VIF values of the main effect variables in our final surrogate model are not higher than 7.5. The final surrogate model includes 14 main effect variables (4 news event history variables, 4 marketing-related variables, and 6 other important variables).

In addition to these 14 variables, we consider two-way interactions between marketing-related variables and the company-related variables that are measured on a yearly basis. Table A21 summarizes our procedure and the results for identifying relevant interaction effects.

Table A21: Selection Process for Interaction Terms

Step 1: Specification test for each two-way interaction separately (F test: F-value)

Marketing-related variables:	**Company-related variables:**			
	Sales	**Firm size**	**Market share**	**Leverage**
Advertising budget	7,261.075	2,504.250	49.229	351.196
Relationship equity	24.638	724.600	1,519.897	n.s.
R&D expenditures	1,176.216	594.582	326.455	n.s.
SG&A expenses	317.199	4,176.707	1,144.623	10.603

Step 2: Collinearity check (√ if VIF < 10)

	Sales	**Firm size**	**Market share**	**Leverage**
Advertising budget	√	√	√	√
Relationship equity	√	√	√	-
R&D expenditures	√	√	√	-
SG&A expenses	√	√	√	√

Note: n.s. = not significant (p > .05).

In total, we have 16 interactions (4 x 4) to test for. To keep the model parsimonious and reduce the influence of collinearity that plagues interaction variables by construction, we follow a stepwise procedure. We first add each interaction effect separately to our base model and test for this model extension using the F test. We then check whether the remaining candidate interactions satisfy standard collinearity requirements (VIF < 10). This leaves us with 14 interactions. One out of the 14 interaction terms (R&D expenditures x Firm size) does not contribute to the model including all other interactions terms. Thus, we remove this interaction term, which leaves us with 13 interactions in our final surrogate model.

APPENDIX H: DISTRIBUTION SELECTION OF SURVIVAL MODEL

We tested underlying distributions of the survival function and come to the conclusion that the Weibull distribution outperforms the loglogistic, exponential, and lognormal distribution according to the information criterion AIC, the Bayes IC, and the Hannan Quinn.

Table A22: Determination of the Underlying Distribution of the Survival Model

Underlying distribution survival model	**Inf.Cr.AIC**	**Bayes IC**	**Hannan Quinn**
Weibull	**7108.6**	**7152.1**	**7122.3**
Loglogistic	7142.4	7186.0	7156.1
Exponential	7234.4	7269.3	7245.4
Lognormal	7142.4	7186.0	7156.1

REFERENCES APPENDIX PAPER 1

Davis, Jesse and Mark Goadrich (2006), "The Relationship Between Precision-Recall and ROC Curves," *Proceedings of the 23rd International Conference on Machine Learning*, 233–240.

Bergstra, James and Yoshua Bengio (2012), "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, 13 (2), 281–305.

Boyd, Kendrick, Victor Santos Costa, Jesse Davis, and C. David Page (2012), "Unachievable Region in Precision-Recall Space and its Effect on Empirical Evaluation," *arXiv Preprint arXiv:1206.4667*.

Breiman, Leo (2001), "Random Forests," *Machine Learning*, 45 (1), 5–32.

Chang, Yung-Chia, Kuei-Hu Chang, and Guan-Jhih Wu (2018), "Application of eXtreme Gradient Boosting Trees in the Construction of Credit Risk Assessment Models for Financial Institutions," *Applied Soft Computing*, 73, 914–20.

Chen, Tianqi and Carlos Guestrin (2016), "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, Association for Computing Machinery, 785–94.

Kuhn, Max and Kjell Johnson (2013), "Applied Predictive Modeling," Vol. 26. Springer.

Lantz, Brett (2019), *Machine Learning with R*, 3rd ed. Packt Publishing.

Olson, Randal S., Cava, William La, Mustahsan, Zairah, Varik, Akshay, and Jason H. Moore (2018), "Data-Driven Advice for Applying Machine Learning to Bioinformatics Problems," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*, 192–203.

RepRisk (2023), "RepRisk Methodology Overview," (accessed June 15h, 2023) [available at https://www.reprisk.com/news-research/resources/methodology].

Riedmiller Martin (1994), "Rprop - Description and Implementation Details - Technical Report," University of Karlsruhe.

Saito, Takaya and Marc Rehmsmeier (2015), "The Precision-Recall Plot Is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS ONE*, 10 (3), https://doi.org/10.1371/journal.pone.0118432.

Paper 2: When is Competition Really Healthy? Analyzing the Impact of the Firm's Competitive Situation on the Disclosure of Unethical Firm Behavior

Authors: Lars Gemmer, Alexander Edeling, and Marc Fischer

Abstract

There is an ongoing debate about the role of market share and market share-based competition for firms' behavior and financial performance. This study contributes to this discussion by analyzing the relationship between the specific competitive situation of a company and its unethical behavior disclosure. Analyzing a global sample of 2,777 companies and following an empirics-first approach, the authors introduce new variables that reflect the competitive situation of a firm and identify which competitive constellations increase competitive pressure and force companies to act unethically. Information regarding three different factors is relevant to explain unethical behavior disclosure. These include the exact ranking-related position of a company within an industry, the market-share-based proximity of direct competitors, and dynamic changes in these constellations. Furthermore, the authors show how the disclosure of unethical behavior, in turn, affects the future competitive position and thus reveal potential dynamics between the firm's competitive situation and unethical firm behavior disclosure. Although the initial goal of the ethical misconduct is likely to strengthen the competitive position, findings show that the market share decreases by .82% per each additionally disclosed incidence of unethical behavior. Managers should be aware of this relationship when developing competitive strategies and refrain from assuming that a market share orientation is infallible.

Keywords: ESG-related corporate misconduct, corporate social irresponsibility, market share-based competition, competitive pressure, empirics-first approach

1 Introduction

On 12th of March, 2015, Martin Winterkorn, the former CEO of Volkswagen, stated that "The Volkswagen Group is increasing the pace. In 2015, we intend to take the next step towards the top. In other words, we are now getting ready to overtake." (Volkswagen Group 2015). Indeed, a short time later, Volkswagen overtook Toyota and became the global market leader (Financial Review 2015). Only eight months later, the Volkswagen emissions scandal became public (Siano et al. 2017). As one consequence of the disclosed unethical firm behavior[4], the market share of Volkswagen in Europe fell to its lowest level since the financial crisis (Financial Times 2016). This highlights how the company suffered a consumer backlash after the emissions scandal. The example indicates the potential dynamic relationship between an intense market share-based competition (i.e., race for market leadership), disclosed unethical firm behavior (i.e., emissions scandal), and a resulting setback with regard to the competitive position (i.e., reduction of market share).

Volkswagen is not the only company focusing on market share (i.e., firm unit sales or revenues divided by market unit sales or revenues) as a key performance indicator. For many managers, one of the top goals is to increase market share (Bhattacharya, Morgan, and Rego 2022; Farris et al. 2010). Management decisions aimed at evaluating performance in relation to competitors can be referred to as competitor-oriented goals (Armstrong and Collopy 1996).

However, a growing number of studies question the pursuit of higher market share as a panacea. A higher market share may relate negatively to customer satisfaction due to the difficulty of pleasing larger and thus more heterogenous customer segments (Rego, Morgan, and Fornell 2013). Results from a meta-analysis by Edeling and Himme (2018, p. 4) question a competitive orientation of a firm that "focuses too strongly on retaining and increasing market

[4] Throughout this article, we consistently use the term (disclosed) unethical firm behavior to refer to "firm-induced incidents that appear to hurt the social good" (Kang, Germann, and Grewal 2016). Synonyms used in the literature are corporate social irresponsibility (e.g., Stäbler and Fischer 2020), corporate misconduct (e.g., Liu 2016), or ESG (environmental, social and governance)-related misconduct (e.g., Burke 2022). As we define later, we consider only unethical firm behavior that was disclosed. Undisclosed unethical firm behavior is not part of this study.

share as a business objective". They show that the positive effect of a high market share on performance is lowest in the US, an economy characterized by strong competitive pressure. Armstrong and Collopy (1996) even suggest that firms should ignore their competitors when setting objectives and, instead, focus directly on profit maximization. Countering these articles, a recent study by Bhattacharya, Morgan, and Rego (2022) does not support a negative mediating role of competitor orientation within the market-share profit relationship, which adds to the "competitor-orientation puzzle".

This fierce debate about the role of market share and market share-based competition orientation raises the question to what extent the effects of the Volkswagen example are generalizable. Are there competitive situations of firms within one industry that influence the pressure on decision-makers in firms and trigger unethical actions and their disclosure? If yes, which specific situations of competitive performance exist and how do they drive unethical firm behavior disclosure? Can striving for an improvement of the competitive situation turn into a disadvantage and even backfire leading to a weakening of the competitive position?

To answer these research questions, we introduce new competition-related variables that reflect the competitive situation of a firm and analyze the unethical behavior disclosure of 2,777 international companies from 79 different industries in a time window from 2007 to 2017. Our analyses consider all types of unethical firm behavior independent of its geographical origin and are not limited to large corporate crises (such as the mentioned VW emission scandal). Importantly, we follow an empirics-first approach. This approach describes research that "(1) is grounded in (originates from) a real-world marketing phenomenon, problem, or observation, (2) involves obtaining and analyzing data, and (3) produces valid marketing-relevant insights without necessarily developing or testing theory" (Golder et al. 2023, p. 319f.). Since there are no concrete existing insights about our new competition variables in the established literature and theories suggest mainly conflicting expectations regarding the effect direction of these new competition variables on disclosed unethical firm behavior, the application of an empirics-first approach is warranted (Golder et al. 2023, p. 330).

By developing two distinct models, we shed light on potential dynamics between the specific competitive situation and the disclosure of unethical firm behavior. First, we model the effect of the competitive situation of a firm within one industry on the likelihood of future unethical behavior disclosure by conducting a hurdle negative binomial (HNB) regression. HNB models assume that the data are a mixture of two separate data generation processes: one generates only zeros, and the other is a negative binomial data-generating process (count values) that truncates zeros (Gurmu and Trivedi 1996; Mullahy 1986). Both processes can be modeled independently. Secondly, we measure the impact of unethical firm behavior disclosure on future market share with a market share response model.

Compared to existing research, our study differs in important ways: First, we introduce numerous new market share-related variables that describe the firm-specific competitive situation and proxy the competitive pressure of companies and their managers. Specifically, we analyze the effect of a company's global position within an industry, the situation regarding the local (direct) competition, and the dynamics of these two factors. In this way, we gain important insights into the influence of a company's exact competitive situation within an industry on its unethical behavior.

Second, although research has already extensively focused on the consequences of unethical firm behavior disclosure, (e.g., Kang, Germann, and Grewal 2016; Kölbel, Busch, and Jancso 2017; Stäbler and Fischer 2020), the drivers of irresponsible firm behavior are still relatively unknown. To the best of our knowledge, no existing study analyzes the effect of the concrete competitive situation of a firm on unethical firm behavior as well as the potential negative feedback effects of unethical firm behavior disclosure on a firm's market share.

Our study provides important contributions to marketing research and practice: First, we contribute theoretically to the broad field of competition. We introduce several yet unstudied market share-related variables that describe the competitive situation of a company within one industry. In this way, we shed light on the sparse research on industry structure and business performance in marketing research (Uslay, Altintig, and Winsor 2010) and also on the current reinvigorated discussion about market share as a dominant corporate goal and its positive and negative consequences for future firm performance (Bhattacharya, Morgan, and Rego 2022; Edeling and Himme 2018; Rego, Morgan, and Fornell 2013).

Second, our study also contributes substantively to the extant research on unethical firm behavior. Although the topic has received growing attention from researchers in the last two decades (Borah and Tellis 2016; Cleeren, Dekimpe, and Helsen 2008), potential drivers of unethical firm behavior and its disclosure are still relatively unknown. Our study fills this research gap to a significant extent by identifying certain competitive constellations that increase the likelihood of unethical firm behavior disclosure.

Lastly, the findings contribute to managerial decision-making: Our study reveals that one additional unethical firm behavior disclosure leads to a relative decrease of the market share by .82%[5] on average, and thus, results in an actual setback in the battle for a better competitive position. Hence, we show that managers' attempt to strive for a better competitive situation by unethical means are not necessarily crowned with success. It can even be counterproductive and counteract an improvement of the competitive position. Based on these findings, companies should, for example, reevaluate the incentives for decision-makers and define other more sustainable goals such as increasing brand equity or improving customer relationships, instead of the potentially detrimental market share maximization, for which some managers use ethical misbehavior as a shortcut.

In addition, our findings imply that firms should include intra-industry competitive constellations when implementing internal monitoring measures that warn of corporate

[5] Note that it is important to distinguish between percentage changes and percentage point changes when talking about market shares. For example, an increase in market share from 5% to 6% implies a 20% increase in market share, but only a one percentage point increase.

misconduct. Investors, especially those that consider ESG criteria (so-called "ESG investors, see McGee 2022) may use the new insights from our study to optimize their investment decisions and to adapt their investment portfolio proactively by analyzing the competitive situations of firms they potentially invest in.

The remainder of this work is structured as follows: Firstly, we provide an overview of relevant existing studies to position our study and gain inspiration from the relevant fields of research. Secondly, we derive new variables that proxy the competitive pressure and develop a conceptual framework. Note that the empirics-first approach does not demand an overarching theoretical foundation or concrete a priori expectations about the effect directions presented. Third, we describe the data set and the methodology. Lastly, we summarize and discuss the empirical results, and provide potential theoretical explanations for our findings as well as practical implications.

2 Literature Review

Two research streams are relevant to this study. The first research stream attempts to explore the real-world consequences of unethical firm behavior and its disclosure. The second research stream zooms in on the impact of competitive pressure on specific types of unethical behavior, both at an individual (employee) and at an aggregate (firm) level. In the following, we will summarize these relevant research streams and present representative findings.

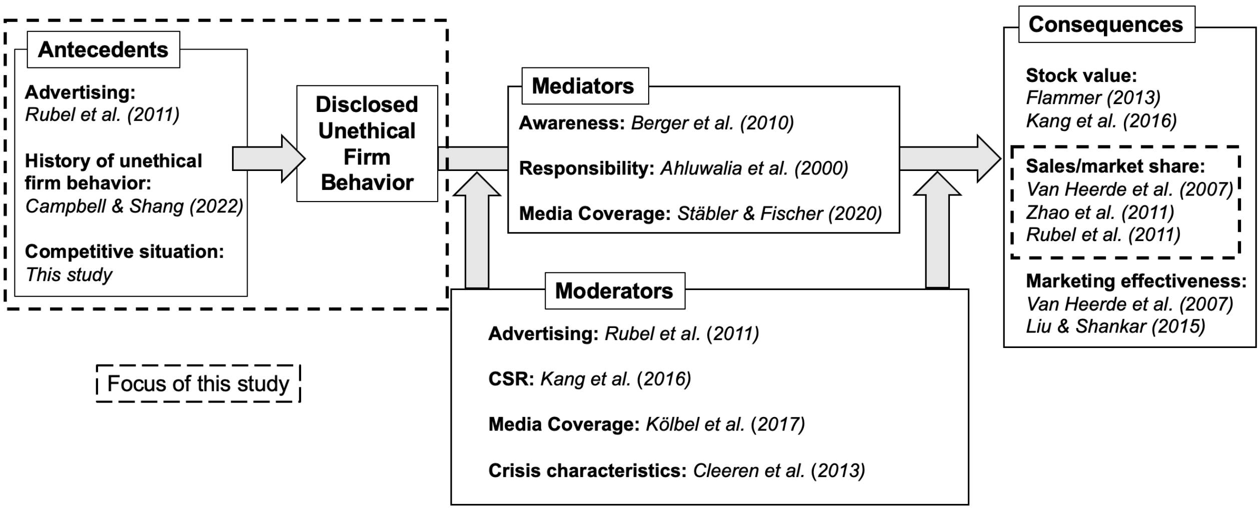
2.1 Literature on Unethical Firm Behavior (Disclosure)

In line with the high practical relevance of unethical firm behavior for companies, research about the topic has grown to a major field over the last decades. Researchers studied various real-world consequences of unethical firm behavior as well as important mediators and moderators of this process (e.g., Borah and Tellis 2016; Cleeren, Dekimpe, and Helsen 2008).

Figure 1 highlights representative studies and the identified relationships. As shown, existing research focuses mainly on the analysis of the consequences of unethical behavior disclosure on various outcome variables like the firm value (Flammer 2013; Kang, Germann, and

Grewal 2016), sales (Van Heerde, Helsen, and Dekimpe 2007; Zhao, Zhao, and Helsen 2011), or the marketing effectiveness (Liu and Shankar 2015).

Figure 1: Representative Studies on Unethical Firm Behavior (Disclosure)



In addition, numerous studies have identified moderators (Cleeren, Van Heerde, and Dekimpe 2013; Rubel, Naik, and Srinivasan 2011) and mediators (Berger, Sorenson, and Rasmussen 2010; Stäbler and Fischer 2020) of these relationships. There are also a few early attempts to investigate antecedents of unethical firm behavior. In their study, Rubel, Naik, and Srinivasan (2011) initially analyze the time window before the event occurs (i.e., the product harm crisis is published in the media). The authors look at pre-event advertising levels that might be strategically used by management envisioning a crisis to moderate the negative consequences.

In addition, a recent study by Campbell and Shang (2022) uses machine learning to predict firms' violations of rules and regulations issued by government institutions. However, the focus of that study is a prediction and not an explanation of unethical firm behavior. Also, there is a difference between unethical firm behavior and violations of regulations. Not every unethical action is also illegal.

2.2 Literature on Competitive Pressure and Unethical (Firm) Behavior

We categorize prior studies on competitive pressure and unethical behavior along two dimensions (see Table 1): First, prior studies examine either aggregate competitive pressure (i.e., competition intensity within the whole industry or the whole company) or individual competitive pressure (i.e., concrete entity-specific position or rank of a company within one industry or employee within one company).

Second, the existing studies take either an intra-firm perspective or an inter-firm perspective. The intra-firm perspective examines the unethical behavior of individual employees of a company. The inter-firm perspective, on the other hand, examines the competitive interactions of companies.

Table 1: Representative Studies on Competition and Unethical (Firm) Behavior

		Organizational perspective	
		Intra-firm	**Inter-firm**
Type of pressure	**Aggregate competitive pressure**	A Kulik, O'Fallon, and Salimath (2008) (T) Verbeke, Ouwerkerk, and Peelen (1996) (S)	B Bennett et al. (2013) (F) Branco and Villas-Boas (2015) (T) Shleifer (2004) (T)
	Individual competitive pressure	C Hegarty and Sims (1978) (E) Perry, Baranowski, and Parcel (1990) (F)	**D** **This study (F)**

Notes: (F) = Field study; (T) = Theoretical study; (E) = Experimental study; (S) = Survey study.

For example, the studies by Kulik, O'Fallon, and Salimath (2008) and Verbeke, Ouwerkerk, and Peelen (1996) take an intra-firm perspective and analyze competitive pressures on an aggregated level (Cell A). Kulik, O'Fallon, and Salimath (2008) develop a grassroots model to describe structural factors that may influence the emergence and spread of an employee's (un)ethical behavior in organizations by using the example of the company Enron. Verbeke, Ouwerkerk, and Peelen (1996) show how and when individual salespeople make unethical decisions based on the organization's environment and climate.

Bennett et al. (2013), Branco and Villas-Boas (2015), and Shleifer (2004) also examine the impact of aggregate competitive pressures, but these studies focus on the behavior between firms (Cell B). Bennett et al. (2013) show that competition can cause organizations to provide services that customers demand but violate government regulations, especially when price competition is restricted. Firms with more competitors pass customer vehicles at higher rates and are more likely to lose customers whom they fail, suggesting that competition intensifies pressure to provide illegal leniency. Branco and Villas-Boas (2015) theoretically show that the degree of competition is negatively related to the extent to which firms invest in behaving according to the rules of the marketplace. Using examples such as child labor and corruption, the study by Shleifer (2004) clarifies the potential negative (short-term) and positive (long-run) effects of competition in promoting unethical firm behavior (which the author terms "censured conduct").

Other studies explore the effect of individual competitive pressure and consider an intra-firm perspective (Cell C). Hegarty and Sims (1978), for example, look at ethical decision-making under different contingencies of reinforcement (e.g., knowledge of other performance or rank). A study by Perry, Baranowski, and Parcel (1990) distinguishes between type A and type B students. Type A students have been characterized by extremes of impatience, aggressiveness and hostility, competitive achievement striving, and time urgency that are evoked by a variety of environmental situations. The authors find that these students are more likely to cheat.

Interestingly, to the best of our knowledge, no study exists that links individual competitive pressures to an inter-firm perspective. We fill this research gap by introducing several variables that describe the company-specific competitive situation and the resulting competitive pressure of firms within an industry and by analyzing how these measures affect unethical firm behavior disclosure on a company level.

3 Competitive Pressure as Driver of Unethical Decisions

In the following section, we briefly present the categories of variables that are relevant to our analysis and show how we benefit from existing theories in deriving these variables. Note that after analyzing the data, we develop explanations for our empirical results.

Analyses of established theories reveal that three different factors – which are all related to a company's specific competitive situation – potentially influence competitive pressure: the global position of a firm within one industry, the direct local competition of a firm, and competitive dynamics.

3.1 The Comparative Position of Firms within One Industry

Theoretical evidence indicates that the position of a firm within one industry influences competitive pressure. For example, *tournament theory* is useful for describing behavior when reward structures are based on relative ranks rather than absolute levels of output (Connelly et al. 2014; Lazear and Rosen 1981). The theory is often used to describe inter-firm competition. Researchers are also incorporating uncooperative behavior by showing that contestants may benefit not only from increasing their productivity but also from reducing the productivity of their competitors.

Furthermore, based on *social comparison theory* (Festinger 1954), individuals and firms evaluate their performance compared to competitors (Armstrong and Collopy 1996; Greve 2008; Kilduff 2019). The drive to obtain or retain a higher relative position is characterized by comparative concerns, which result in competitive behavior (Garcia, Tor, and Schiff 2013, p. 635). Both, the insights from tournament theory and social comparison theory, indicate that the position of a firm within one industry plays an essential role for the competitive pressure and the resulting unethical behavior by decision-makers and companies.

Note that we also include information on whether a firm is a market leader and a member of the top three companies in an industry. Research has already demonstrated the crucial role of the market leader and the three largest companies within one industry (Edeling and Himme

2018; Uslay, Altintig, and Winsor 2010). For example, Uslay, Altintig, and Winsor (2010) show that there are often three generalist firms controlling the market.

3.2 Direct Competition

Connelly et al. (2014) conclude that *tournament theorists* incorporate aspects of actor heterogeneity (Nippa 2011) by forming sub-contests. Within these sub-contests, participants compete with a more homogeneous subgroup (Gomez-Mejia, Treviño, and Mixon Jr. 2009). Transferred to our research problem, this suggests that there may be subgroups of closer related firms within one industry. These subgroups may be specifically relevant for the competition orientation. We use the term direct local competition to describe the competitive situation within these subgroups.

Social comparison theory states social comparison can generally be viewed in two directions and is, therefore, divided into upward and downward social comparisons (e.g., Brown et al. 2007). In upward comparison, one’s performance is compared with that of the better performer, i.e., the firm with the next better position. This often results in negative feelings, which in turn can lead to competitive behavior (Garcia and Tor 2007; Garcia, Tor, and Gonzalez 2006; Garcia, Tor, and Schiff 2013; Tesser 1988). In contrast, downward comparison involves comparing with someone worse off, i.e., the company with the next worse position within one industry. This downward orientation can also lead to competitive behavior as the drive to keep a higher relative position may result in competitive behavior (Garcia, Tor, and Schiff 2013). We, therefore, consider the direct upward and downward competition of firms in our analyses.

3.3 Competitive Dynamics

Based on the *tournament theory* researchers examine how behavior dynamically changes in sequential tournaments (Rosen 1986). The fight for market share can be seen as a sequence of yearly tournaments suggesting to include information about competitive dynamics, i.e., the change of the competitive situation from one time period to another time period, in our study.

Prospect theory (Kahneman and Tversky 1979) suggests that performance improvements are associated with an increase in aspirational reference points. If the market share deteriorates and the firm even falls in position, this can lead to emotional and psychological incentives to regain market share to improve the position again (Ferrier et al. 2002). In addition, prospect theory suggests that decision-makers who have experienced loss are more risk-seeking (Fiegenbaum and Howard 1988; Mishina et al. 2010), which is expected to influence the amount of unethical behavior disclosure. For this reason, we consider also dynamic changes in the competitive situation (e.g., improvement or deterioration) as additional information.

4 Conceptual Framework

In the following section, we introduce the newly developed competition variables for the three previously identified categories. We describe our conceptual model as well as the dependent variables (i.e., the disclosed unethical firm behavior and the market share) that are relevant to answer our research questions. Finally, we explain important additional control variables that enrich our models.

4.1 Introduction of Competitive Pressure Variables

Guided by theoretical insights, we create several variables for each of the three identified proxies of competitive pressure. Table 2 groups the individual variables into one of the three categories.

Table 2: Measures of the Firm's Competitive Pressure

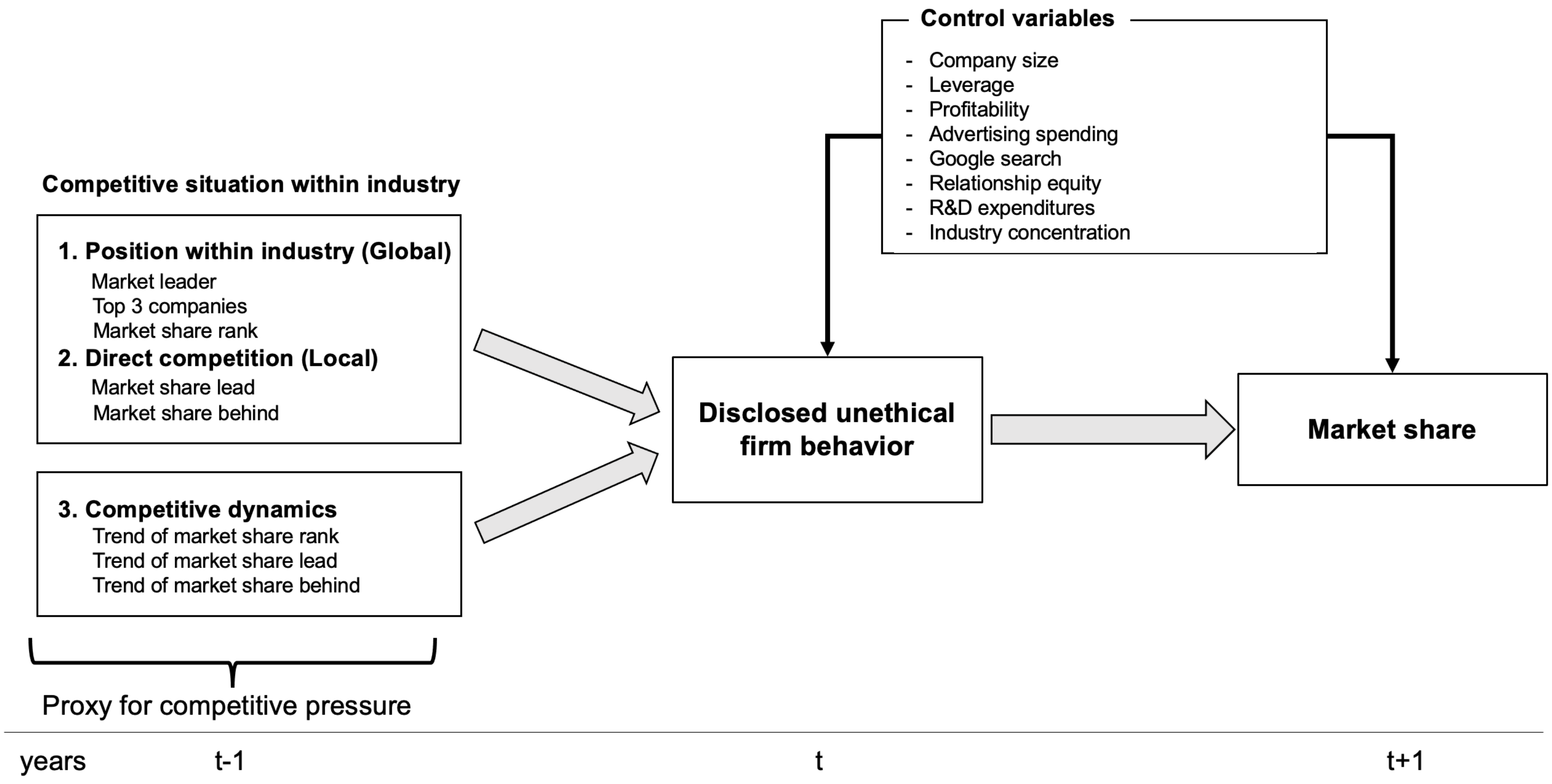
Position within industry (Global)	**Direct competition (Local)**	**Competitive dynamics**
• Company is the market leader • Company belongs to the three largest companies in the industry • Market share rank	• Difference in market shares between the company and the next smaller company ("market share lead") • Difference in market shares between the company and the next larger company ("market share behind")	• Trend of market share rank • Trend of market share lead • Trend of market share behind

The global position of a company within an industry can be characterized in terms of whether the company is a *market leader*, whether it belongs to the *three largest companies* in the industry, and which *market share rank* it occupies in the industry (see Table 4 for a detailed operationalization of the variables). The second category describes the direct competition between companies on a local level. This includes information on how close a company is to the *next larger* ("market share behind") and the *next smaller company* ("market share lead") in the industry with respect to its market share. The third category of variables covers dynamics. It includes the recent *trend of the market share rank* and the *trend of the market share differences* to the two closest competitors.

Figure 2 shows our conceptual framework including these competition variables. In the first step, there are specific competitive situations, in which companies are under rising pressure to increase their market share or to at least keep it constant. In these situations, companies try to improve or at least maintain their competitive position. Corporate misbehavior could be one considered strategy to achieve these goals in the product marketplace. These considerations could lead then to an actual increase in the disclosure of unethical misconduct.

In the second step, disclosed unethical firm behavior is expected to impact the future market share. Identifying this relationship empirically will show if unethical behavior disclosure will decrease the market share which is the opposite of what was initially intended in the first step. The initial goal of companies to compete and improve their situation through unethical behavior may ultimately fail and even, on the contrary, worsen the competitive situation.

Figure 2: Conceptual Framework



4.2 Disclosed Incidents of Unethical Firm Behavior

The Switzerland-based ESG data science firm RepRisk provides data on news coverage of companies regarding unethical firm behavior (RepRisk 2023). The database has been used by prior international strategy research (Dinner, Kushwaha, and Steenkamp 2019; Kölbel, Busch, and Jancso 2017) and is well established in practice as a leading ESG information source for various organizations from fields such as banking, hedge fund, and governments.

RepRisk screens a broad range of over 100,000 media, stakeholders, and other third-party sources in 23 languages (more than 95% of the world's GDP) to identify unethical firm behavior incidents. The media screened by RepRisk include print and online media (including local, regional, national, and international), communication from NGOs, government bodies, regulators, and think tanks as well as newsletters and social media including Twitter and blogs, and other online sources.

RepRisk identifies the very first news report on all potential issues that threaten corporate reputation. These news reports cover more than 70 different issues and topics which relate to unethical firm behavior. Examples cover child labor, fraud, poor employment conditions, waste issues, supply chain issues, water scarcity, and privacy issues. A full list of all topics included is given in Appendix A.

RepRisk does not restrict itself to a pre-defined sample of firms. Specifically, RepRisk does not search for specific companies in the media but identifies news that cover negative issues and topics, and based on this, includes all companies that are associated with these topics in its database. Consequently, the data set contains companies from all over the world, and any existing company (publicly listed or not) has the potential to be included in the RepRisk data set as soon as the company receives bad media coverage.

RepRisk covers more than 205,000 companies from all industries and countries. 33% of the firms are from Asia, 26% from Europe, 24% from North America, and 17% from other areas. 7% of the included companies are publicly listed.

As a measure of the actual unethical behavior of firms, the data provided by RepRisk is biased to some degree because of two main reasons. First, if the unethical firm behavior was never disclosed, it is not included in our analyses. Note that any research in the field on unethical firm behavior using observational data faces this issue. This bias could only be corrected with laboratory studies, in which, however, the complex realistic decisions of companies cannot be represented. Our study with observational data covering many companies from 79 industries ensures a high external validity. Second, disclosed incidents by RepRisk include accusations without evidence, rumor, or fake news. These incidents do not represent actual unethical firm behavior. However, data from previous research reveal a high correlation (= .920) between actual evidence-based unethical firm behavior and incidents without evidence (Stäbler and Fischer 2020), which strongly reduces the influence of this bias.

4.3 Controls

We include key financial measures such as leverage, profitability, and company size, but also marketing-specific factors such as the advertising budget, relationship equity, google search volume, and expenditures on research and development (R&D) as controls. The financial health of a firm is a predictor of risks such as credit default and bankruptcy (e.g., Barth et al. 1998; Lin, Ansell, and Andreeva 2012). If management envisions such risks, it increases the pressure to avoid them which might involve a higher propensity to engage in unethical and controversial behavior such as reducing costs by environmental pollution (Duanmu, Bu, and Pittman 2018; Dupire and Zali 2018). We measure the financial health of a company in terms of the firm's profitability and leverage.

In addition, we also control for the absolute size of companies in terms of revenues. Firm size may influence the likelihood of unethical decisions. Larger firms with many operations in different parts of the world increase the pool of employees engaged in unethical behavior just by the numbers. In addition, large firms are more likely to be monitored and covered by the media (Stäbler and Fischer 2020).

Investments in customer relationships, R&D, and advertising might improve marketing performance in terms of strong brands and a valuable customer base, which increases the salience and presence of companies and their brands (e.g., Rust et al. 2004). This in turn increases media attention and the likelihood that ethical misbehavior will be uncovered (e.g., Stäbler and Fischer 2020). At the same time, successful marketing performance may increase the pressure on managers to maintain the success (Mishina et al. 2010), which might influence the occurrence of unethical firm behavior and its disclosure. We include receivables investments as a proxy for relationship equity (Frennea, Han, and Mittal 2019), R&D investments as a proxy for innovation and product quality improvements, and advertising investments as a proxy for brand-building activities (Srinivasan, Lilien, and Sridhar 2011).

In addition, we control for the general industry concentration within an industry and past unethical behavior disclosure of companies. As our literature review has shown that less concentrated industries should be more competitive and that higher general competition could trigger unethical firm behavior (e.g., Bennett et al. 2013; Branco and Villas-Boas 2015; Shleifer 2004). Furthermore, a recent study by Campbell and Shang (2022) shows that the history of unethical behavior is a helpful predictor of future unethical behavior.

5 Data Description

5.1 Population of Firms

We create a unique panel data set including variables from the multiple data sources RepRisk, Compustat, Google Trends, and Kantar. Merging these data sets yields a rich sample of 20,063 yearly observations from 2,777 different companies from 61 countries active in 79 industries covering a time window from 2007 to 2017. Our sample of firms consists of all companies that are listed in Compustat (North America and Global) and are included the Kantar's Ad$pender data set at the same time. Consequently, all firms from our sample are listed on the stock market.

Importantly, only 61.36% of the companies (= 1,704) are included in the RepRisk data set, which indicates that the other share of companies (= 1,073) in our sample did not face disclosure of unethical firm behavior in the observed time period. Since RepRisk collects data issue-based and not company-based, it is reasonable to conclude that the companies which are not included in the RepRisk data set, did not face any unethical firm behavior incident in the observed time window. We keep these companies without a single incident in the data set so that the models developed are also transferable to new companies that have not been exposed to any unethical behavior before.

Google Trends data is available for all 2,777 companies in the analyzed time window.

5.2 Descriptives of Disclosed Incidents of Unethical Firm Behavior

We aggregate news reports about unethical firm behavior on a yearly basis and create a count variable that measures how often a company is exposed to negative news reports about unethical behavior in a year over the time period from 2007–2017.

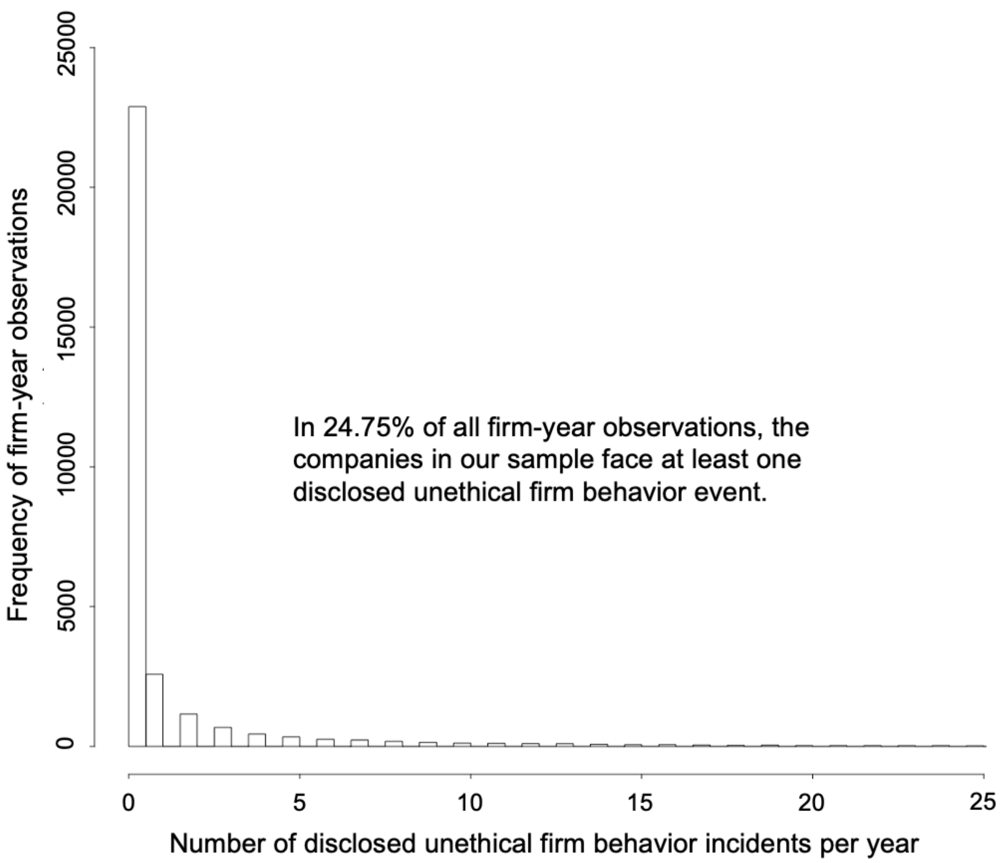
Table 3: Distribution of Unethical Firm Behavior Disclosure Over Time

Year	Number of disclosed uneth-ical firm behavior events	Growth rate of disclosed un-ethical firm behavior events
2007	1,363	-
2008	2,585	.897
2009	2,101	–.187
2010	3,226	.536
2011	4,190	.299
2012	6,284	.500
2013	7,723	.229
2014	10,622	.375
2015	10,626	.000
2016	9,158	–.138
2017	11,114	.214
Total:	68,992	7.154

Table 3 provides an overview of the number of disclosed unethical firm behavior incidents per year and the resulting growth rate. In total, our data set includes 68,992 disclosed unethical firm behavior incidents. The strong increase over the years (overall growth rate of 715% from the year 2007 to the year 2017) underpins the growing practical importance of the topic within the media landscape. On average, a company faces 2.5 unethical firm behavior incidents per year. The number of incidents per year varies between 0 and 385 per company. 75.25% of all firm-year observations contain no unethical firm behavior event so the count variable consists to a large extent of zeros.

Figure 3 shows the distribution of the count variable graphically. The variable exhibits overdispersion, with a variance much larger than the mean, which we account for with a negative binomial hurdle model (see Methodology section below).

Figure 3: Distribution of Disclosed Unethical Firm Behavior Events per Year



Notes: For illustrative purposes, the figure is limited to a maximum of 25 unethical firm behavior incidents. The actual maximum value equals 385 events per year. The distribution refers to the entire data set before missing values are deleted.

5.3 Competition within Homogeneous Groups of Firms

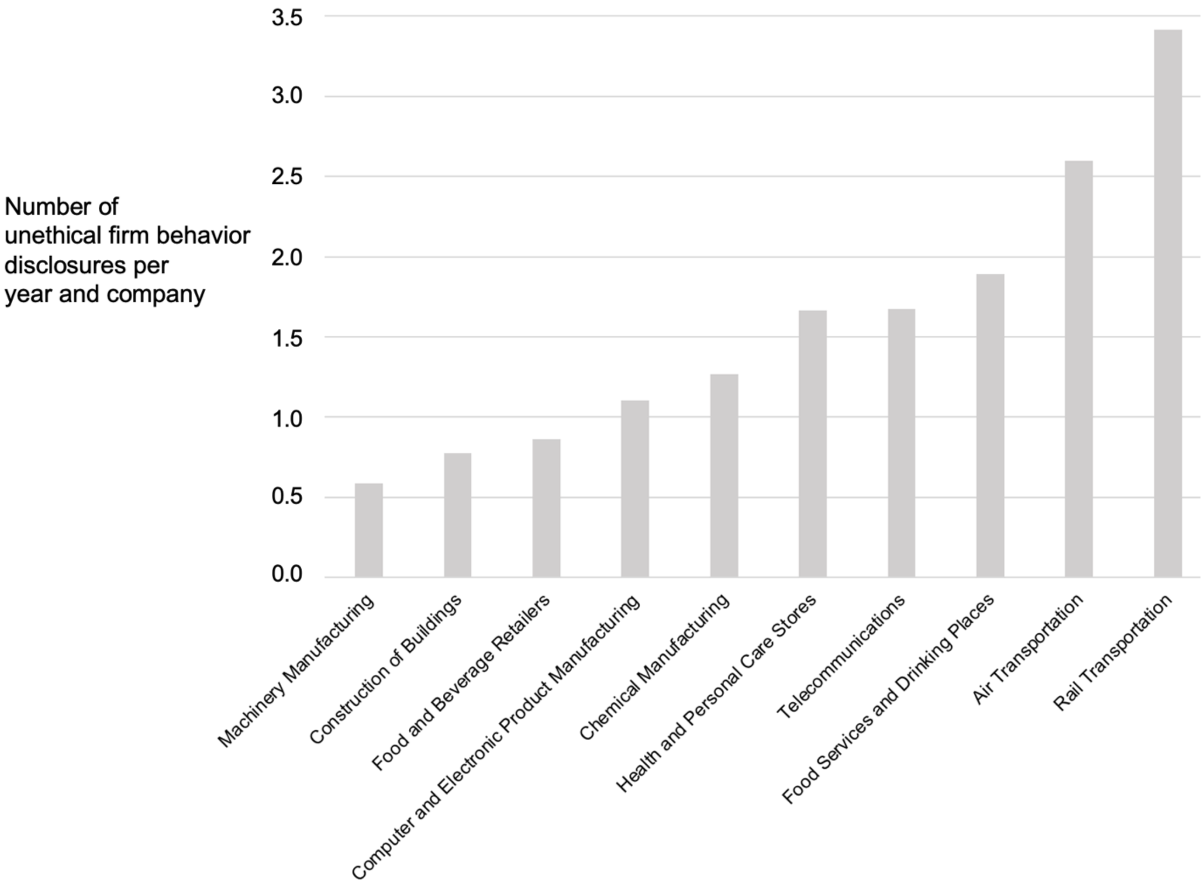
We use North American Industry Classification System codes to cluster the companies into subsectors (NAICS Association 2017). We are aware that across different business-related research fields there exist various approaches to divide companies into homogeneous clusters, e.g., Standard Industrial Classification (SIC), North American Industry Classification System (NAICS), or Text-based Network Industry Classifications (TNIC) (e.g., Bhojraj, Lee, and Oler 2003; Hoberg and Phillips 2010).

TNIC, however, is limited to a specific selection of companies ("database is based on all publicly traded firms (domestic firms traded on either NYSE, AMEX, or NASDAQ) for which we have COMPUSTAT and CRSP data.") (Hoberg and Phillips 2010, 2016). Recall that the setting of our study is global.

From a marketing research perspective, the division into industries based on SIC codes is widely established (e.g., Frennea, Han, and Mittal 2019; Steenkamp et al. 2011). However, NAICS aims to improve the SIC "by using a production-based framework throughout to eliminate definitional differences; identifying new industries and reorganizing industry groups to better reflect the dynamics of our economy; and allowing first-ever industry comparability across North America" (Saunders 1999, p. 37). Classifications based on NAICS are also increasingly used in marketing research (e.g., Uslay, Altintig, and Winsor 2010). Thus, for our study, it is reasonable to use 3-digit NAICS codes to classify the firms into homogenous industries.

All in all, our data set includes 79 different industries. On average, each of these industries in the data set contains 40.43 companies. Figure 4 shows how the average number of disclosed unethical firm behavior events for one company in one year differs between exemplary industries. We report the figures for all industries in Appendix B. From the selected industries, the industry "Rail Transportation" appears to have the greatest risk of unethical behavior disclosure. On average, a company belonging to this industry has 3.42 unethical incidents per year.

Figure 4: Number of Unethical Behavior Incidents per Company and Year for Exemplary Industries



Notes: Figure 4 shows exemplary industries. Figures for all included industries are shown in Appendix B. The number of unethical firm behavior events is relative to the number of companies belonging to the industries to allow comparability. The numbers refer to the entire data set before missing values are deleted.

Interestingly, the risk varies greatly between industries (we account for these differences by including industry-specific random effects in our model specifications). However, the graph does not reveal any information about how a company's particular situation within these industries drives the risk of unethical behavior disclosure.

5.4 Descriptives and Variable Operationalization

As displayed in Table 2, we develop eight market share-related variables from three variable categories (i.e., global position within the industry, direct local competition, and competitive dynamics) that describe the competitive situation of a firm. The denominator of our market share metric is the total annual sales figures (converted to US$ if otherwise stated in Compustat) of a NAICS 3-digit subsector within Compustat North America and Compustat Global. The market share of a company then represents its share of sales in these subsector total sales.

The average market share in our data set equals .91% with a minimum of 0% and a maximum value of 100%. We include a market leader dummy, a dummy that measures if a company belongs to the three largest companies within one industry in a given year, a rank variable (i.e., 1 = market leader, 2 = company with the second largest market share, and so forth) as well as two variables measuring the market share distance to the next larger (market share behind) and next smaller (market share lead) company.

Analogously to the market share calculation, the basis for calculating each of the competition variables is the total of all companies from Compustat North America and Compustat Global. However, the descriptive figures reported in Table 4 refer only to the companies included in our analyzed data set.

In 1.4% of our yearly observations, the company is a market leader and in 4.2% of the cases the company belongs to the three largest companies in the industry. The average rank of a company in our data set is 378 (Recall that this variable is calculated using every available company included in Compustat North America and Global).

We describe the detailed operationalization of the variables in Table 4. As the data is characterized by extreme values, we also report the median as well as the 5%-trimmed mean (mean of a variable after excluding the 5% smallest and 5% largest values) in addition to the mean, minimum, and maximum values.

Table 4: Variable Operationalization and Descriptives

Variable Category	Variable	Operationalization	Source	Mean	Median	SD	Min	Max	5% trimmed mean
Dependent variables	Disclosed unethical firm behavior	Number of disclosed incidents related to unethical firm behavior per year	RepRisk	2.478	0	11.848	0	385	.717
	Market share	Sales divided by total sales of industry	Compustat	$.910\text{x}10^{-02}$	$.110\text{x}10^{-02}$	.034	0	1.000	$.442\text{x}10^{-02}$
Competitive situation[1]									
Position within Industry (Global)	Market leader	Dummy indicating whether company is the market leader in industry	Compustat	.014	0	.116	0	1	.000
	Top 3 Companies	Dummy indicating whether company belongs to the group of the three biggest companies in industry	Compustat	.041	0	.198	0	1	.000
	Market share ranking	Variable indicating the rank of the company within one industry, 1 = market leader	Compustat	378.257	113	642.590	1	3,806	276.567
Direct competition (Local)	Market share lead	Absolute difference in market share of the company and the closest company with a lower market share	Compustat	$.214\text{x}10^{-2}$	$.880\text{x}10^{-05}$	.024	0	1.000	$.239\text{x}10^{-03}$
	Market share behind	Absolute difference in market share of the company and the closest company with a higher market share	Compustat	$.170\text{x}10^{-2}$	$.747\text{x}10^{-05}$	.013	0	.395	$.199\text{x}10^{-03}$
Competitive dynamics	Trend market share ranking	Absolute difference between the rank of the year t_{-1} and the year t_0	Compustat	–4.072	0	99.204	–2,505	1,809	–1.890
	Trend market share lead	Absolute difference between market share lead of the year t_{-1} and the year t_0	Compustat	$-.142\text{x}10^{-05}$	0	.005	–.238	.208	$-.219\text{x}10^{-05}$
	Trend market share behind	Absolute difference between market share behind of the year t_{-1} and the year t_0	Compustat	$-.119\text{x}10^{-03}$	0	.012	–.997	.233	$.125\text{x}10^{-07}$
Control variables	Relationship equity	Receivables/receivables of industry	Compustat	.098	.015	.192	0	1.000	.065
	Firm size	Total sales in $1m	Compustat	6,636.799	859.239	22,213.070	–5.074	470,171.000	3,026.281
	R&D expenditures	R&D expenditures/R&D expenditures of industry	Compustat	.056	$.320\text{x}10^{-04}$	.176	–6.455	1.000	.026
	Leverage	(Long-term debt + debt in current liabilities)/total assets	Compustat	1.193	.177	44.879	0	4,665.500	.199
	Profitability	EBIT/total assets	Compustat	–.352	.057	9.621	–591.000	1.632	.043
	Industry	Industry based on 3-digit NAICS classification	Compustat	79 different industries					
	Industry concentration	Herfindahl index (sum of all squared market shares within one industry)	Compustat	.037	.023	.044	.007	1.000	.031
	Advertising Spending	Yearly advertising budget in 000 US$	Kantar Media	1,173.909	.200	6,837.604	0	131,959.000	157.395
	Google search	Yearly google online search volume of a firm	Google trends	27.020	1.750	195.451	0	3424.241	4.620

Notes: The numbers refer to the data set without the observations lost due to time lags and missing values as these observations are relevant for our final models. As we consider outliers in our models, we report the minimum, maximum, and mean values in Table 4 although they are partly misleading and influenced by wrongly reported figures from data providers.
[1] Competition variables are calculated with all available companies from Compustat. This is the reason why, for example, the mean market share rank equals 378 although there are, on average, only 40 firms in each industry in the analyzed data.

Since the developed competition variables are all based on the market share of a company, there is a potential risk that (some of) the variables are highly correlated with each other. Table 5 addresses this concern and shows the correlations between the relevant competition variables and company size.

Table 5: Correlation Between Competition Variables and Company Size

	Company size	Market leader	Top 3 companies	Market share rank	Market share behind	Market share lead	Trend ranking	Trend market share behind	Trend market share lead
Company size	1								
Market leader	**.378**	1							
Top 3 companies	**.495**	**.570**	1						
Market share rank	**–.164**	**–.070**	**–.122**	1					
Market share behind	**.130**	–.013	**.365**	**–.064**	1				
Market share lead	**.231**	**.391**	**.340**	**–.051**	**.089**	1			
Trend ranking	.009	.004	.007	**.080**	.004	.003	1		
Trend market share behind	.000	**–.040**	.013	.001	**.449**	**–.022**	–.001	1	
Trend market share lead	–.004	**–.043**	–.005	–.001	–.009	**.108**	.000	**–.085**	1

Notes: Bold figures indicate a significant correlation at the .01 level.

Some competition variables show a significant correlation. However, the correlation only exceeds .5 in one single case: Naturally, the fact that a company is the market leader correlates with the top 3 companies variable. Although the correlations are quite low in general, we keep a potential correlation issue in mind and address it later in the results section of the study.

6 Methodology

6.1 Hurdle Negative Binomial Model

To analyze the relationship between a firm's competitive position and its unethical behavior disclosure, we develop two models: First, we examine the influence of different variables describing the competitive position of firms on their unethical behavior disclosure. Since the dependent variable is a count variable characterized by overdispersion and zero inflation, we adopt a hurdle negative binomial (HNB) model. HNB models provide a powerful way to model this type of situation (Gurmu and Trivedi 1996; Mullahy 1986). Note that zero-inflated negative binomial models (Greene 1994; Lambert 1992) may offer an alternative way to analyze the data. However, zero-inflated negative binomial models differentiate between two types of zeros, structural zeros (arising due to the specific structure of the data) and sampling zeros (in the counting process). In terms of our application case, this means that we analyze companies that generally do not face unethical firm behavior disclosure (structural zeros) and potentially unethical companies that, however, have years with 0 unethical behavior disclosure (sampling zeros). We assume that every firm has the potential to act unethically and to face disclosure meaning that there is only one source for zero unethical firm behavior disclosure (sampling zeros). Recall that RepRisk covers all companies worldwide and reports their unethical behavior incidents (Appendix A).

Formal test results that support overdispersion in our data are presented in Appendix C. We also compared the model performance of alternate models like a poisson and a negative binomial model with the performance of the HNB model (Appendix C). The HNB model performs best according to the AIC and BIC values.

HNB models assume that the data are a mixture of two separate data generation processes: one generates only zeros, and the other is a negative binomial data-generating process (count values) that truncates zeros (Gurmu and Trivedi 1996; Mullahy 1986). Both processes can be modeled independently: A binomial regression models the zeros and a zero-truncated negative binomial (ZTNB) regression models the count process:

$$\mathrm{P(Y = y)} = \begin{cases} Binomial(\pi) & for \;\; y = 0 \\ \mathrm{ZTNB}\ (\mu)\ with\ (1-\pi) & for \;\; y = 1, 2, \ldots \end{cases}, \tag{1}$$

Y represents the number of disclosed unethical firm behavior events, π is the probability of zero counts, $(1 - \pi)$ is the probability that the hurdle is crossed, and μ measures the number of unethical firm events when Y > 0.

In our context, a logit model identifies the years, in which companies generally do not face disclosure. The negative binomial regression then analyzes which factors determine the number of unethical firm behavior-related events if there is at least one disclosed misconduct. Note that even if a company faces many incidents over several years, there can be a year without any unethical firm behavior disclosure. We do not differentiate between firms without disclosure and firms with disclosure but we differentiate between years with zero incidents and with at least one incident.

The logit part of the HNB model can be written as:

$$P(\text{zero incidents} \mid G, \varphi, \delta, \nu, \gamma, , u) = \frac{e^{G_{it}}}{1+e^{-G_{it}}}, \tag{2}$$

$$G_{it} = \varphi_0 + \sum_{k=1}^{K} \varphi_k^X X_{it-1,k} + \sum_{l=1}^{L} \varphi_l^Y Y_{it-1,l}$$

$$+ \delta_{Year} + \nu_{Company} + \gamma_{industry} + u_{it},$$

where P measures the probability of having zero unethical firm behavior events. G represents the collection of all included independent variables, whereas X represents all focal competition variables and Y represents the remaining control variables. i is an index for the company and t is an index for the year. $k \in K$ and $l \in L$ are indices for the independent variables. u measures the error term. δ_{Year} measures the year-fixed effects. Equation 1 also includes two random error terms, $\nu_{Company}$ and $\gamma_{industry}$, which we assume to be normally distributed with zero mean and a variance to be estimated. By incorporating these error components, we account for unobserved effects that are specific to the company (ν) and the industry (γ). φ are the estimates of the logit model. We use a log link for the logit model.

For the negative binomial model, we specify the following model using the same notation and the same set of independent variables as in the logit part (2):

$$\text{Count}_{it} = \rho_0 + \sum_{k=1}^{K} \rho_k^X X_{it-1,k} + \sum_{l=1}^{L} \rho_l^Y Y_{it-1,l} \tag{3}$$

$$+ \delta_{Year} + \nu_{Company} + \gamma_{industry} + q_{it},$$

where ρ are the estimates of the negative binomial part and q measures the error term.

For the analysis, we delete all observations with at least one missing value in one of the variables considered. The final data set in the HNB model contains 20,063 observations from 2,777 companies belonging to 79 different industry sectors.

In line with the marketing literature, the independent variables are lagged by one year as already presented in the conceptual framework (see Germann, Ebbes, and Grewal 2015 for a detailed explanation of using time-lagged independent variables to reduce a potential endogeneity issue).

6.2 Market Share Response Model

Controlling for company size, leverage, R&D expenditures, advertising spending, profitability, industry concentration, relationship equity, google search, as well as year effects to account for unobserved heterogeneity, we develop the following market share response model:

$$\begin{aligned} \mathrm{MS_{it}} = {} & \beta_0 + \beta_1\,\mathrm{UNETHIC_{it-1}} + \beta_2\,\mathrm{SIZE_{it-1}} + \beta_2\,\mathrm{LEV_{it-1}} + \beta_3\,\mathrm{PROF_{it-1}} \\ & + \beta_4\,\mathrm{ADV_{it-1}} + \beta_5\,\mathrm{GS_{it-1}} + \beta_6\,\mathrm{RE_{it-1}} + \beta_6\,\mathrm{RD_{it-1}} + \beta_6\,\mathrm{ICON_{it-1}} \\ & + \delta_{\mathrm{Year}} + \nu_{\mathrm{Company}} + \gamma_{\mathrm{industry}} + \mathrm{z_{it}}, \end{aligned} \tag{4}$$

where

MS_{it}	=	Market share of firm i in year t,
$UNETHIC_{it}$	=	Number of unethical firm behavior events,
$SIZE_{it}$	=	Company size,
LEV_{it}	=	Leverage,
$PROF_{it}$	=	Profitability,
ADV_{it}	=	Advertising spending,
GS_{it}	=	Google search,
RE_{it}	=	Relationship equity,
RD_{it}	=	R&D expenditures,
$ICON_{it}$	=	Industry concentration,
β	=	unobserved parameter vectors,
z_{it}	=	stochastic error term.

Again, δ_{Year} measures the year-fixed effects. Equation 4 also includes two random error terms, ν_{Company} and $\gamma_{\mathrm{Industry}}$ to account for unobserved effects that are specific to the company (ν) and the industry (γ).

6.3 Addressing Potential Endogeneity

In our study, we do not claim to identify strict causal relationships between the specific competitive situation of a company within an industry and disclosed unethical firm behavior as well as the future market share. Recall that we use observational data that make it almost impossible to identify causal effects. However, we make a high effort to rule out alternate sources of unethical firm behavior disclosure. In the following section, we describe these efforts.

When analyzing observation data, there can be the problem that the focal variables (i.e., competition variables) correlate with the error term. This endogeneity issue would lead to biased estimates and the validity of the results would not be assured. Germann, Ebbes, and Grewal (2015) summarize various sources of endogeneity that are common and offer potential remedies. We subsequently describe these issues and explain how we address them in our study.

First, it is likely that disclosed unethical firm behavior is not only the result of the competitive situation of a company within one industry but also driven by other factors which are correlated with the competitive situation (e.g., financial pressure, firm size). We enrich our models with numerous control variables to tackle this source of endogeneity.

Second, unobservable time-invariant firm-specific effects such as specific characteristics of employees and executives could present another source of variable omission bias. To capture these and other unobserved time-invariant firm-specific effects, we exploit the advantages of panel data and estimate firm-specific random effects.

Third, there might be unobserved time-varying effects that impact both unethical firm behavior and the competitive situation of a company, e.g., higher efforts by media to identify unethical firm behavior. This should lead to serially correlated measures. By including the lagged dependent variable (and other control variables), we effectively control for such firm-specific time-varying unobserved influences.

Fourth, there may also be concerns that estimation results are affected by a simultaneity bias, for example, when a higher amount of anticipated unethical firm behavior disclosure in a given year influences the competitive situation of a company. To control for such simultaneity effects, the models include lagged values of the predictor variables. In addition, the competitive situation of a firm within one industry is a relative measure (see operationalization of competition variables). The focal variables are all based on the market share, which is a relative measure per definition. The final firm-specific situation as measured by our variables is the comparison between the situation of the focal firm and all other firms in that industry, and thus, again, a relative measure. This minimizes this potential simultaneity because the focal firm would need to anticipate the competitive situation and outcomes of each competitor to adjust its own situation. Note also that the number of sales (which is the main component of market share) is an intermediate performance variable that cannot be changed as directly as advertising expenditures or prices.

7 Empirical Results

7.1 Results of the Hurdle Negative Binomial Model

The results of the Hurdle negative binomial model are divided into two parts: The zero component contains logit coefficients for predicting non-zeros along with their standard errors and significance levels (left side of Table 6). The coefficients reflect how much the log odds of a non-zero change if covariates change by one unit. The results of the count component show how the independent variables affect the concrete number of disclosed unethical firm behavior incidents (right side of Table 6). The parameter values of the negative binomial regression indicate the effect of the independent variables on the log(number of disclosed unethical behavior incidents).

For both components of the HNB model, we report a main effects model. This model allows us to obtain information on the main effects of the focal variables. To obtain additional valuable information about potential moderators, we enrich the model with interactions. Specifically, we investigate whether the influence of the market share behind and the market share lead is moderated by the position of the company within an industry (i.e., ranking).

Table 6: Results of the HNB Model

	Zero inflation component **Dependent variable**[1]**:** At least one unethical behavior disclosure in the respective year				**Count component** **Dependent variable:** Number of unethical behavior disclosures in the respective year			
	Main effects		**Interaction effects**		**Main effects**		**Interaction effects**	
	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)	Estimate	(SE)
Focal Competition Variables:								
Global position:								
Market leader	–.335	(.410)	–.109	(.413)	.165*	(.090)	.242***	(.092)
Top 3 companies	.285	(.262)	.500*	(.265)	.111*	(.066)	.135**	(.066)
Market share ranking	–.002***	($.133 \times 10^{-03}$)	–.002***	($.132 \times 10^{-03}$)	–.001***	($.153 \times 10^{-03}$)	–.001***	($.151 \times 10^{-03}$)
Local direct competition:								
Market share lead	–7.031***	(2.002)	–15.540***	(3.500)	.729	(.873)	-3.002**	(1.355)
Market share behind	3.008	(2.707)	–23.670***	(6.299)	1.231**	(.610)	-3.609*	(2.030)
Competitive dynamics:								
Trend ranking	$.146 \times 10^{-03}$	($.362 \times 10^{-03}$)	$.141 \times 10^{-03}$	($.360 \times 10^{-03}$)	$.370 \times 10^{-04}$	($.257 \times 10^{-03}$)	$.353 \times 10^{-03}$	($.256 \times 10^{-03}$)
Trend market share lead	–1.002	(7.049)	–15.930**	(7.674)	–.509	(1.345)	–1.687	(1.391)
Trend market share behind	4.809**	(2.414)	4.343*	(2.484)	–.197	(.445)	-.211	(.445)
Interactions:								
Market share lead*Ranking			8.078***	(2.674)			2.801***	(.807)
Market share behind*Ranking			11.370***	(2.531)			2.373**	(.935)
Controls:								
Company size[2]	42.830***	(2.937)	39.300***	(2.937)	2.734***	(.381)	2.851***	(.380)
Leverage	–.007	(.015)	–.007	(.013)	–.069	(.145)	–.061	(.144)
Profitability	–.003	(.010)	–.003	(.010)	.432**	(.195)	.415**	(.194)
Advertising[2]	9.923***	(1.746)	9.539***	(1.727)	1.784***	(.352)	1.684***	(.351)
Google search[2]	1.224	(.760)	1.245*	(.757)	1.388***	(.428)	1.465***	(.425)
Relationship equity	1.023***	(.197)	.947***	(.197)	.671***	(.140)	.626***	(.140)
R&D	–.129	(.208)	–.111	(.206)	.168*	(.091)	.163*	(.091)
Industry concentration	–.604	(.385)	–.486	(.384)	–.414*	(.236)	–.309	(.236)
Previous unethical behavior	.521***	(.028)	.520***	(.028)	.009***	(.001)	.009***	(.001)
Yearly fixed effects	Yes		Yes		Yes		Yes	
Industry and company random intercepts	Yes		Yes		Yes		Yes	
Balanced Accuracy in %	83.314		83.319		–		–	
Pseudo R^2 (squared correlation)	–		–		.550		.563	
AIC value	37,766		37,721		37,766		37,721	

Notes: [1] We change the dependent variable so that it measures if there was at least one disclosure of unethical behavior to ensure simpler interpretability of estimates.
[2] We normalized the control variable between 0 (minimum value) and 1 (maximum value).
Number of observations = 20,063; number of industries = 79; number of years = 9; number of companies = 2,777.
*** p < .01 ** p < .05 * p < .1 (two-sided tests).

7.1.1 Zero component. Among the variables that describe the global position within the industry, only the ranking has a significant effect on the likelihood to have at least one unethical firm behavior-related incident (coefficient = –.002, $p < .01$). In other words, it is more likely that a company with a lower ranking – indicating a relatively larger company – faces at least one disclosure at all in a specific year.

Among the local competition variables, while we do not find a significant main effect for our market share behind variable, a significant effect emerges for the market share lead variable (coefficient = –7.031, $p < .01$). Firms that are leading their followers by only a small margin have a higher risk of facing unethical firm behavior disclosure.

Conversely, regarding our trend variables, the only significant variable in the HNB main effects model is the trend in market share behind (coefficient = 4.809, $p < .05$). Put differently, the trend in market share behind increases the probability of at least one disclosed event of unethical firm behavior in a given year. This means that if the distance to the next larger company increases – i.e., the focal company falls further behind – then the probability that there will be at least one disclosed incidence increases. Hence, losing ground leads to a higher tendency for unethical firm behavior disclosure than tightening the race.

Concerning control variables, larger companies and a higher advertising budget, higher relationship equity, and a higher number of previous unethical behavior events significantly increase the likelihood of at least one unethical incident in a year.

Interestingly, the interaction effects model reveals that the influence of direct competition is moderated by the ranking. The impact of a large distance to the next larger or smaller firm on the likelihood to have at least one unethical firm behavior incident is more pronounced if the ranking of the firm is high ($\text{coefficient}_{behind*rank} = 11.370$, $p < .01$; $\text{coefficient}_{lead*rank} = 8.078$, $p < .01$). This means in reverse that a tighter race at the top of an industry increases the likelihood of unethical firm behavior disclosure.

We already analyzed the correlation of the focal variables (see Table 5 again). To rule out the possibility that multicollinearity is a problem, we further analyze the resulting VIF values. The values of the focal variables in the main models are all below 2 and a multicollinearity problem is unlikely to exist (Appendix D).

7.1.2 Count component. As with the zero component, the main effects model in Table 6 shows the direct main effects of the focal variables on the number of disclosed annual unethical firm behavior incidents. All three global position-related variables have significant coefficients. A market leader has an expected log(number of disclosed unethical behavior incidents) of .165 higher than that of a non-market leader ($p < .10$). Thus, the fact that a company is a market leader increases the expected number of unethical firm behavior events by 17.94% (($e^{.165} - 1$) *100).

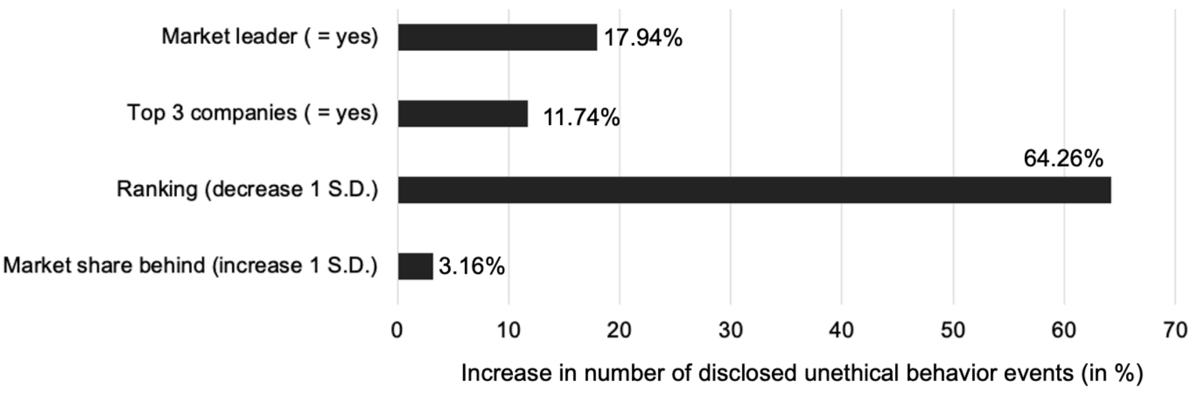
The two other variables describing the global position of a company are also significant. An increase in ranking by one unit – indicating a reduction of the market share size – reduces the number of unethical firm behavior disclosure by .10% (($e^{-.001} - 1$) *100, $p < .01$). This finding supports the insights about the effect of the ranking gained from the zero component of the HNB. In addition, if a company belongs to the group of three largest companies in an industry, the number of disclosed unethical incidents increases further by 11.74% (($e^{.111} - 1$) *100, $p < .10$). Note that this increase is the incremental effect of being a top three company. A market leader logically also belongs to the group of the three largest companies and is subject to both this effect and the incremental effect of market leadership.

Regarding direct competition variables, we find that in contrast to the zero component, the variable market share behind has a significant effect on the number of disclosed unethical incidents in the coming year. Interestingly, this effect is positive. A larger distance to the next larger competitor significantly increases the number of unethical behavior disclosure. One additional percentage of market share distance to the next larger company increases the number of incidents by 2.43% $(((e^{1.231} - 1) *100)/100, p < .05)$.

We find no significant impact of competitive dynamics on the number of unethical firm behavior disclosure. However, both included interaction effects are significant. Confirming the insights from the zero component of the HNB, the market share difference to both competitors, the next larger and the next smaller one, is significantly moderated by the market share ranking of a company ($coefficient_{behind*rank} = 2.373$, $p < .05$; $coefficient_{lead*rank} = 2.801$, $p < .01$). Thus, a tighter race at the top of an industry increases not only the likelihood that a firm faces disclosure at all but also the number of disclosed unethical firm behavior events per year.

Figure 5 illustrates the effect strength of the significant competition variables in the count component. For the two continuously measured variables ranking and market share, we show how the number of disclosed unethical behavior changes due to one standard deviation change of the competition variables.

Figure 5: Effect Strength of Significant Competition Variables (Count Component)



Consistent with the zero component model, we find that a larger firm size, a higher advertising budget, higher relationship equity, and more unethical firm behavior events in the past relate positively to unethical firm behavior. A higher industry concentration significantly decreases the number of unethical firm behavior events in the upcoming year. This effect shows that firms in more competitive industries (i.e., the market is not dominated by a few large companies) have on average more unethical firm behavior events confirming previous findings (e.g., Bennett et al. 2013).

7.2 Results of the Market Share Response Model

We find that unethical firm behavior disclosure reduces a company's market share (coefficient = –.0075, $p < .01$). As shown in Table 7, one additional unethical firm behavior incident reduces the market share in the next year by .0075.

This effect size seems to be relatively small. However, it is relevant to evaluate the effect size in relation to the average market share size of .91% (see Table 4). An additional unethical firm behavior disclosure, therefore, decreases this market share by .824% (.0075/.91)*100). On average, a company faces 2.478 disclosures per year resulting in a yearly market share reduction of 2.042% (.824%*2.478).

Table 7: Results of the Market Share Response Model

Dependent variable: Market share

	Estimate	**(SE)**
Focal Variable:		
Previous unethical firm behavior	$–.749x10^{–02}$***	($.759x10^{–03}$)
Controls:		
Company size	$.438x10^{–04}$***	($.752x10^{–06}$)
Leverage	$–.916x10^{–06}$	($.132x10^{–03}$)
Profitability	$.483x10^{–04}$	($.265x10^{–03}$)
Advertising spending	$.150x10^{–04}$***	($.274x10^{–05}$)
Google search	$.106x10^{–03}$	($.990x10^{–04}$)
Relationship equity	1.068***	(7.076)
R&D	.114**	(.054)
Industry concentration	.061	(.051)
Yearly fixed effects	Yes	
Industry and company random intercepts	Yes	
Pseudo R^2 (squared correlation)	.969	

Notes: Market share is scaled between 0 and 100.
Number of observations = 22,845.
*** $p < .01$ ** $p < .05$ * $p < .1$ (two-sided tests).

Recall that the number of unethical firm behavior incidents variable is characterized by an inflation of zeros, which reduces also the mean of the variable. Considering only years with at least one unethical firm behavior disclosure, the average number of incidents increases to

13.420. This would lead to a yearly market share reduction of 11.058% (.824%*13.420) on average. A company with an exemplary market share of 10% would suffer from a market share reduction that results in a market share of 8.894% (10% – (.11058*10%)) in the upcoming year.

A market share reduction weakens a firm's competitive situation, as the competitive variables are all based on market share. The initial goal of firms to become more successful in competing against other companies by unethical measures cannot be achieved as the market share becomes smaller (note that the market share is a relative metric that already controls for the performance of the firm's competitors in the same time period).

8 Discussion

8.1 Summary

Analyzing a rich and unique data set of 2,777 international companies and 68,992 disclosed unethical firm behavior-related incidences by applying an empirics-first approach, we show that certain competitive situations lead to an increase in disclosed unethical firm behavior. This, in turn, worsen the competitive situation for the company.

Our findings contribute theoretically to the literature on unethical firm behavior. Although the relationship between competitive pressure and unethical misconduct within a business context has already been shown (e.g., Bennett et al. 2013; Branco and Villas-Boas 2015; Shleifer 2004), there has been no theoretical derivation of the effect of the specific competitive situation of companies within one industry. In addition, marketing research is particularly sparse when it comes to understanding the general drivers of unethical firm behavior and its disclosure.

In our study, we derive three categories of factors – the global position of a firm within one industry, local competition, and competitive dynamics – that influence the pressure on decision-makers in companies and can thereby encourage unethical practices. All three factors describe the specific situation of a company within one industry. Like that, for the first time, we present the joint effect of several new competition variables on disclosed unethical firm behavior.

We also show how unethical behavior disclosure, in turn, affects the competitive position by reducing the market share by 2.04% per year on average and thus reveal potential dynamics between competitive pressure and disclosed unethical firm behavior. We see that the disclosure of unethical behavior sets back companies in the competitive arena. It inevitably follows that it is not useful to beat competitors by unfair means because it weakens the situation of one's own company relative to the competitors. By identifying these relationships, we contribute to the currently prevailing debate on the role of market share and market share-based competitive orientation.

8.2 Theoretical Explanations

Following the empirics-first approach, we aim to provide theoretical explanations for our findings. Importantly, we are not limited to one all-encompassing theory, but rather rely on "multiple angles" (Golder et al. 2023, p. 323).

8.2.1 The concept of unethical pro-organizational behavior. Several fundamental theories (e.g., tournament theory, competitor orientation theory, or social comparison theory) highlight the relevance of psychological mechanisms at an individual level to explain the decisions of managers. Conversely, our study explores unethical behavior disclosure at a company level and not at an individual level. Recall that we analyze unethical behavior disclosure from an inter-firm perspective, not at the individual employee level. However, key individuals, such as top managers, play a crucial role in making important (unethical) decisions in companies (e.g., Johnson, Sutton, and Theis 2020). Hence, it is necessary to shed light on the psychological mechanisms underlying their decision-making. The question arises if it is possible to explain the influence of company-related competitive pressure (i.e., market share maximization) with the psychological processes of individual decision-makers within that company. Can the competitive pressure the company as an organization is facing also affect the perceived pressure of the individual decision-making employees of that company so that they act unethically?

The concept of unethical pro-organizational behavior (UPB) provides an answer to this question. UPB is a well-established construct that describes the fact that managers and their employees act unethically on behalf of the organization (Mishra, Ghosh, and Sharma 2022; Umphress and Bingham 2011). UPB can be defined as “actions that are intended to promote the effective functioning of the organization or its members (e.g., leaders) and violate core societal values, mores, laws, or standards of proper conduct” (Mishra, Ghosh, and Sharma 2022, p. 622). A great body of scientific evidence confirms that employees conduct unethical behaviors *on behalf of* the organization (e.g., Ashforth and Anand 2003; Pinto, Leana, and Pil 2008).

Based on UPB, it is reasonable to look for theories explaining individual behavior and psychological processes (e.g., social comparison theory by Festinger 1954) to explain unethical misconduct caused by increased competitive pressure at a company level.

8.2.2 Explanations for the impact of the global competitive position. Table 8 summarizes the findings of the HNB model (main effects). The global position of a company within an industry plays an outstanding role in explaining the unethical behavior disclosure of companies. All three variables from this category have a significant influence on the occurrence of unethical firm behavior events. Recall that these effects are not due to company size effects, as we additionally control for company size in our models.

Managers conduct *cost-benefit analyses* regarding the decisions and weigh the perceived benefits of unethical actions against potential costs (Barsky 2008; Kilduff et al. 2016; Mishina et al. 2010; Schweitzer, Ordóñez, and Douma 2004). Considering the cost side, research about unethical misbehavior has shown that firms face the risk of negative consequences from being caught, such as a decline in marketing effectiveness or a decrease in the stock market value (e.g., Flammer 2013; Van Heerde, Helsen, and Dekimpe 2007). While unethical behavior disclosure can lead to these negative consequences, the likelihood of adopting unethical behavior grows when the value and attractiveness of the potential benefit rise (Barsky 2008; Kilduff et al. 2016; Schweitzer, Ordóñez, and Douma 2004). According to the cost-benefit analysis already successful firms have more to lose when acting unethically. For smaller firms at the bottom of the industry, the benefit of unethical behavior may be greater than the costs because they have less to lose.

Table 8: Summary of the Main Effects

	Zero-inflation component	**Count component**
Competition variable:		
Global position:		
Being the market leader	n.s.	More unethical events
Belonging to the Top 3 companies	n.s.	More unethical events
A better ranking	Higher chance of at least one unethical event	More unethical events
Local competition:		
Larger distance to the next smaller company (market share lead)	Lower chance of at least one unethical event	n.s.
Larger distance to the next larger company (market share behind)	n.s.	More unethical events
Competitive dynamics:		
Trend of ranking is increasing	n.s.	n.s.
Trend of distance to the next larger company is increasing	Higher chance of at least one unethical event	n.s.
Trend of distance to the next smaller company is increasing	n.s.	n.s.

Notes: n.s. = non-significant relationship.

However, a company with a larger market share is associated with stronger social comparison concerns and competitive behavior (e.g., Chen et al. 2012; Garcia and Tor 2007; Garcia, Tor, and Gonzalez 2006; Garcia, Tor, and Schiff 2013; Poortvliet et al. 2009; Vriend, Jordan, and Janssen 2016). Better positions are intensified by intrinsic (e.g., feelings of power and status) and extrinsic (e.g., economies of scale) values, which could trigger an increase in competitive orientation (Edelman and Larkin 2015; Vriend, Jordan, and Janssen 2016). In line with this argumentation, a *higher rank* may also decrease unethical firm behavior disclosure. Our empirics-first approach shows that the argument that smaller companies do not have anything to lose and are therefore exposed to an increased risk of unethical behavior disclosure seems to be incorrect.

Tournament theory explains the large prize differential at the highest level of a sequential tournament, such as for CEO pay in an internal promotion contest (Rosen 1986). This illustrates the special role of the winner in a competition. Transferring these insights to the results of our study, the *market leader* of an industry has significant strategic advantages over all of its competitors. For example, market leaders benefit from strategic advantages due to economies of scale, economies of scope, and learning curve effects (e.g., Edeling and Himme 2018; Jacobson and Aaker 1985; Ross 1986). All these advantages indicate that the relative competitive pressure on the market leader decreases and consequently, the unethical behavior disclosure should decrease. Our empirical findings show that this argumentation is wrong. Market leaders appear to be in a situation in which unfair means are still needed to survive in competition and defend their outstanding position. One potential explanation for the increase in disclosure risk may be that market share leaders, as holders of the position of power, should be highly motivated to maintain their outstanding position as they benefit from the highest intrinsic value of status and the highest extrinsic value such as leader compensation and reputational benefits for the firm (e.g., Vriend, Jordan, and Janssen 2016).

In addition, we confirm previous findings that the *three largest companies* within one industry play a special role (Uslay, Altintig, and Winsor 2010). Marketing literature suggests that most industries are characterized by the dominance of three large companies (coined the “Rule of Three” by Sheth and Sisodia 2002). As our empirical results suggest, it is fair to argue that these three large “generalists” compete particularly strongly for market leadership to benefit from its strategic advantages (e.g., Ross 1986). If a company belongs to the three largest companies, this leads to an increase in the number of disclosed unethical firm behavior events.

8.2.3 Explanations for the impact of local direct competition. We also find that direct competition is important for unethical corporate behavior disclosure (Table 8). A large body of research in social psychology shows that similarity between competitors can heighten social comparison pressures and raise the impact of the competition on their identities and thus enhance psychological involvement and objective threat (Chen, Su, and Tsai 2007; Garcia, Tor, and Schiff 2013; Kilduff 2019; Kilduff, Elfenbein, and Staw 2010; Tesser 1988). Hence, it can be inferred that leaders and the firms they run experience social comparison pressure vis-à-vis those who are similar regarding one particular performance dimension (e.g., market share). This may influence their competitive behavior and ultimately the likelihood to improve the current position by making use of unfair means. Thus, smaller differences in market shares between the focal company and the next smaller or the next larger company should increase unethical firm behavior.

We show indeed that if the *distance to the following company* is large, the probability that the company will behave unethically decreases. Thus, a large lead seems to reduce competitive pressure. With respect to the distance to the next larger company, however, the effects are surprising. We find that a *large distance to the next larger company* increases the number of disclosed unethical firm behavior events. It appears that the situation in which companies have to catch up a lot to come closer to the next larger competitor encourages unethical behavior. This may be explained by psychological mechanisms leading managers to evaluate the company's performance relative to that of competitors (Armstrong and Collopy 1996). As the relative performance equals the absolute performance of one's own company minus the absolute performance of the competitor (Harris and Bromiley 2007), the relative performance becomes worse if the distance to the next larger company increases. Harris and Bromiley (2007) show that the worse the relative performance, the higher the probability of accounting fraud, which is in line with our result that the probability of unethical misconduct is higher for more negative values of relative performance.

8.2.4 Explanations for the impact of competitive dynamics. Concerning the dynamics, we only find that an increasing *trend of distance to the next larger company* relates to a higher chance of at least one unethical behavior disclosure per year. If the market share deteriorates and the firm even falls in rank, this can lead to psychological motivation to maintain and increase the market share and rank (Ferrier et al. 2002). Decision-makers who have experienced loss are more risk-seeking (Fiegenbaum and Howard 1988; Mishina et al. 2010), which serves as one potential explanation for the higher chance of unethical behavior (disclosure).

8.3 Implications

8.3.1 Implications for management. Besides the theoretical contribution, the findings are of great practical value for managers. Previous studies show the strong negative impact of unethical corporate behavior on a wide range of metrics (e.g., Kang, Germann, and Grewal 2016; Kölbel, Busch, and Jancso 2017; Stäbler and Fischer 2020). These consequences underpin the practical relevance of unethical firm behavior and its disclosure for managers. The disclosure of unethical behavior has the potential to trigger an enormous corporate crisis (see the initial Volkswagen example). It should therefore be of utmost interest to a company's managers to prevent unethical behavior or at least mitigate the consequences.

Managers can use the insights from our study to improve internal corporate risk management: First, they may reduce the risk that unethical decisions are made in general. For example, firms may implement internal monitoring measures that warn of corporate misconduct based on our newly introduced competitive constellations.

Second, in periods of higher risk (e.g., when the company falls back compared to the larger competitor) they can plan (or initiate) appropriate actions in response to the (potential) unethical firm behavior disclosure to reduce negative consequences of unethical misconduct. For example, if managers are not able to find questionable activities early on, managers could consider the strategic launch of other neutral or positive brand news to diminish the occurrence of negative news in time periods of increased risk (Stäbler and Fischer 2020). Managers may also adjust their marketing plan and invest more heavily in advertising to restore trust, as advertising elasticities decrease in times of crisis (Landsman and Stremersch 2020).

In addition, corporate risk management may strive to prevent unethical behavior by decision-makers through communication strategies: It is important to convince managers that potential unethical behavior tends to set back the company in the competitive arena. Unethical decisions are counterproductive and increase the pressure on managers. We show instances when competitive pressure increases the risk of unethical behavior disclosure and that these situations create a negative feedback effect: Although the initial goal of the unethical behavior is actually to strengthen the competitive position, this position is weakened as the market share per additional unethical firm behavior event decreases on average. Managers should be aware of this paradoxical relationship when developing competitive strategies and should not assume that a market share orientation is infallible. Based on our findings, companies can, for example, rethink the incentives for decision-makers and define other more sustainable goals than short-term focus on competition which triggers unethical means.

8.3.2 Implications for other stakeholders. This study has value for shareholders and policymakers. Unethical firm behavior events are likely to hurt the stock market performance of companies (e.g., Flammer 2013). Investors may use the new insights to optimize their investment decisions and to adapt their investment portfolio proactively by analyzing the competitive situations of firms they potentially invest in. If they want to avoid potential share price losses due to unethical behavior, investors should, for example, avoid stocks of market leaders or top three companies and take the intensity of the industry battles (i.e., the proximity of the respective market shares) into account.

Besides this relevance for short-term stock trading, our study also generates added value for managers of long-term investment portfolios. The global financial market has experienced exponential growth in sustainable investing, an investment approach that considers environmental, social, and governance-related factors in portfolio selection and management (e.g., Avramov et al. 2022). Recall that ESG events are exactly the issues our models explain. So-called ESG investors are often very uncertain about the true ESG profile of a firm because of an absence of a reliable measure of the true ESG performance (Avramov et al. 2022, p. 643). Consequently, our work is also of high relevance for this stakeholder group as it offers one additional information criterion that can be used to make long-term investment decisions.

Political and other public institutions like federal regulatory agencies also have a major interest in understanding the emergence of unethical corporate behavior (see Campbell and Shang 2022). Using the findings of our study, they can better identify and monitor companies with a high risk of acting unethically.

Finally, we suggest new metrics (global competitive position, direct local competition, and competitive dynamics) that antitrust regulators (e.g., FTC) should consider when they determine the detrimental effects of "monopoly power" within a market.

8.4 Limitations and Further Research Avenues

The limitations of this paper offer fruitful avenues for future research. The first limitation goes along with the operationalization of unethical behavior in our study. As previously mentioned, only the unethical behavior of a company that has become public can be explained. Unethical behavior that has not been disclosed is not captured in the data set. However, this limitation concerns any research with observational data on unethical corporate behavior. Since the RepRisk data set includes any reporting of misconduct, no matter how small, and we additionally control for the company-specific random effects in our models (and thus also, for example, for the company-specific management ability to keep unethical behavior secret), disclosed unethical behavior is a suitable proxy. However, as the processes within the firm are inferred but not directly observed, future research could further investigate our suggested mechanism by collecting internal firm data or conducting controlled laboratory studies (e.g., confronting managers with different competitive situations and asking for their aptitude for unethical behavior).

Secondly, we focus on market share as the performance indicator managers aim to maximize. Market share gain is a key organizational objective and a measure of relative strength which managers often believe to be associated with better performance (Armstrong and Collopy 1996; Ferrier et al. 2002). However, performance is a multidimensional construct and future studies could analyze the effect of the competitive situation on unethical firm behavior based on other performance dimensions.

REFERENCES PAPER 2

Armstrong, J. Scott and Fred Collopy (1996), "Competitor Orientation: Effects of Objectives and Information on Managerial Decisions and Profitability," *Journal of Marketing Research*, 33 (2), 188–99.

Ashforth, Blake E. and Vikas Anand (2003), "The Normalization of Corruption in Organizations," *Research in Organizational Behavior*, 25, 1–52.

Avramov, Doron, Si Cheng, Abraham Lioui, and Andrea Tarelli (2022), "Sustainable Investing with ESG Rating Uncertainty," *Journal of Financial Economics*, 145 (2), 642–64.

Barsky, Adam (2008), "Understanding the Ethical Cost of Organizational Goal-Setting: A Review and Theory Development," *Journal of Business Ethics*, 81 (1), 63–81.

Barth, Mary E., Michael B. Clement, George Foster, and Ron Kasznik (1998), "Brand Values and Capital Market Valuation," *Review of Accounting Studies*, 3 (1), 41–68.

Bennett, Victor Manuel, Lamar Pierce, Jason A. Snyder, and Michael W. Toffel (2013), "Customer-Driven Misconduct: How Competition Corrupts Business Practices," *Management Science*, 59 (8), 1725–42.

Berger, Jonah, Alen T. Sorenson, and Scott J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science*, 29 (5), 815–27.

Bhattacharya, Abhi, Neil A. Morgan, and Lopo L. Rego (2022), "Examining Why and When Market Share Drives Profit," *Journal of Marketing*, 86 (4), 73–94.

Bhojraj, Sanjeev, Charles MC Lee, and Derek K. Oler (2003), "What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research," *Journal of Accounting Research* 41 (5): 745–74.

Borah, Abhishek and Gerard J. Tellis (2016), "Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands?," *Journal of Marketing Research*, 53 (2), 143–60.

Branco, Fernando and J. Miguel Villas-Boas (2015), "Competitive Vices," *Journal of Marketing Research*, 52 (6), 801–16.

Brown, Douglas J., Lance D. Ferris, Daniel Heller, and Lisa M. Keeping (2007), "Antecedents and Consequences of the Frequency of Upward and Downward Social Comparisons at Work," *Organizational Behavior and Human Decision Processes*, 102 (1), 59–75.

Burke, Jenna J. (2022), "Do Boards Take Environmental, Social, and Governance Issues Seriously? Evidence from Media Coverage and CEO Dismissals," *Journal of Business Ethics*, 176 (4), 647–71.

Campbell, Dennis W. and Ruidi Shang (2022), "Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews," *Management Science*, 68 (9), 7034–53.

Chen, Ming-Jer, Kuo-Hsien Su, and Wenpin Tsai (2007), "Competitive Tension: The Awareness-Motivation-Capability Perspective," *Academy of Management Journal*, 50 (1), 101–18.

Chen, Patricia, Christopher G. Myers, Shirli Kopelman, and Stephen M. Garcia (2012), "The Hierarchical Face: Higher Rankings Lead to Less Cooperative Looks," *Journal of Applied Psychology*, 97 (2), 479–86.

Cleeren, Kathleen, Marnik G. Dekimpe, and Kristiaan Helsen (2008), "Weathering Product-Harm Crises," *Journal of the Academy of Marketing Science*, 36 (2), 262–70.

Cleeren, Kathleen, Harald J. van Heerde, and Marnik G. Dekimpe (2013), "Rising from the Ashes: How Brands and Categories Can Overcome Product-Harm Crises," *Journal of Marketing*, 77 (2), 58–77.

Connelly, Brian L., Laszlo Tihanyi, T. Russell Crook, and K. Ashley Gangloff (2014), "Tournament Theory: Thirty Years of Contests and Competitions," *Journal of Management*, 40 (1), 16–47.

Dinner, Isaac M., Tarun Kushwaha, and Jan-Benedict E. M. Steenkamp (2019), "Psychic Distance and Performance of MNCs During Marketing Crises," *Journal of International Business Studies*, 50 (3), 339–64.

Duanmu, Jing-Lin, Maoliang Bu, and Russell Pittman (2018), "Does Market Competition Dampen Environmental Performance? Evidence from China," *Strategic Management Journal*, 39 (11), 3006–30.

Dupire, Marion and Bouchra M'Zali (2018), "CSR Strategies in Response to Competitive Pressures," *Journal of Business Ethics*, 148 (3), 603–23.

Edeling, Alexander and Alexander Himme (2018), "When Does Market Share Matter? New Empirical Generalizations from a Meta-Analysis of the Market Share–Performance Relationship," *Journal of Marketing*, 82 (3), 1–24.

Edelman, Benjamin and Ian Larkin (2015), "Social Comparisons and Deception Across Workplace Hierarchies: Field and Experimental Evidence," *Organization Science*, 26 (1), 78–98.

Farris, Paul W., Neil T. Bendle, Philip E. Pfeifer, and David Reibstein (2010), Marketing Metrics: The Definitive Guide to Measuring Marketing Performance, 2nd ed. Pearson Education.

Ferrier, Walter J., Cormac Mac Fhionnlaoich, Ken G. Smith, and Curtis M. Grimm (2002), "The Impact of Performance Distress on Aggressive Competitive Behavior: A Reconciliation of Conflicting Views," *Managerial and Decision Economics*, 23(4–5), 301–16.

Festinger, Leon (1954), "A Theory of Social Comparison Processes," *Human Relations*, 7 (2), 117–40.

Fiegenbaum, Avi and Thomas Howard (1988), "Attitudes Toward Risk and the Risk Return Paradox: Prospect Theory Explanations," *Academy of Management Journal*, 31 (2), 85–106.

Financial Review (2015), "VW Passes Toyota to Take Global Sales Lead in First Half of 2015" by Craig Trudell and Masatsugu Horie, (accessed June 15th, 2023), [available at https://www.afr.com/world/vw-passes-toyota-to-take-global-sales-lead-in-first-half-of-2015-20150728-gim6sd].

Financial Times (2016), "Volkswagen's Market Share Falls After Scandal," by Peter Campbell, London (accessed June 15th, 2023), [available at https://www.ft.com/content/35575f80-4a75-11e6-b387-64ab0a67014c].

Flammer, Caroline (2013), "Does Product Market Competition Foster Corporate Social Responsibility? Evidence from Trade Liberalization," *Strategic Management Journal*, 36 (10), 1469–85.

Frennea, Carly, Kyuhong Han, and Vikas Mittal (2019), "Value Appropriation and Firm Shareholder Value: Role of Advertising and Receivables Management," *Journal of Marketing Research*, 56 (2), 291–309.

Garcia, Stephen M. and Avishalom Tor (2007), "Rankings, Standards, and Competition: Task vs. Scale Comparisons," *Organizational Behavior and Human Decision Processes*, 102 (1), 95–108.

Garcia, Stephen M., Avishalom Tor, and Richard Gonzalez (2006), "Ranks and Rivals: A Theory of Competition," *Personality and Social Psychology Bulletin*, 32 (7), 970–82.

Garcia, Stephen M., Avishalom Tor, and Tyrone M. Schiff (2013), "The Psychology of Competition: A Social Comparison Perspective," *Perspectives on Psychological Science*, 8 (6), 634–50.

Germann, Frank, Peter Ebbes, and Rajdeep Grewal (2015), "The Chief Marketing Officer Matters!," *Journal of Marketing*, 79 (3), 1–22.

Golder, Peter N., Marnik G. Dekimpe, Jake T. An, Harald J. van Heerde, Darren S.U. Kim, and Joseph W. Alba (2023), "Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation," *Journal of Marketing*, 87 (3), 319–36.

Gomez-Mejia, Luis, Len J. Treviño, and Franklin G. Mixon Jr. (2009), "Winning the Tournament for Named Professorships in Management," *International Journal of Human Resource Management*, 20 (9), 1843–63.

Greene, William H. (1994), "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," *NYU Working Paper*, No. EC-94–10.

Greve, Henrich R. (2008), "A Behavioral Theory of Firm Growth: Sequential Attention to Size and Performance Goals," *Academy of Management Journal*, 51 (3), 476–94.

Gurmu, Shiferaw and Pravin K. Trivedi (1996), "Excess Zeros in Count Models for Recreational Trips," *Journal of Business & Economic Statistics*, 14 (4), 469–77.

Harris, Jared and Philip Bromiley (2007), "Incentives to Cheat: The Influence of Executive Compensation and Firm Performance on Financial Misrepresentation," *Organization Science*, 18 (3), 350–67.

Hegarty, W. Harvey and Henry P. Sims (1978), "Some Determinants of Unethical Decision Behavior: An Experiment," *Journal of Applied Psychology*, 63 (4), 451–57.

Hoberg, Gerard and Gordon Phillips (2010), "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis," *The Review of Financial Studies*, 23 (10), 3773–811.

Hoberg, Gerard and Gordon Phillips (2016), Text-Based Network Industries and Endogenous Product Differentiation, *Journal of Political Economy*, 124 (5), 1423–65.

Jacobson, Robert and David A. Aaker (1985), "Is Market Share All That It's Cracked Up to Be?," *Journal of Marketing*, 49 (4), 11–22.

Johnson, Joseph A., Steve G. Sutton, and Jochen C. Theis (2020), "Prioritizing Sustainability Issues: Insights from Corporate Managers About Key Decision-Makers, Reporting Models, and Stakeholder Communications," *Accounting and the Public Interest*, 20 (1), 28–60.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–92.

Kang, Charles, Frank Germann, and Rajdeep Grewal (2016), "Washing Away Your Sins? Corporate Social Responsibility, Corporate Social Irresponsibility, and Firm Performance," *Journal of Marketing*, 80 (2), 59–79.

Kilduff, Gavin J. (2019), "Interfirm Relational Rivalry: Implications for Competitive Strategy," *Academy of Management Review*, 44 (4), 775–99.

Kilduff, Gavin J., Hillary A. Elfenbein, and Barry M. Staw (2010), “The Psychology of Rivalry: A Relationally-Dependent Analysis of Competition,” *Academy of Management Journal*, 53 (5), 943–69.

Kilduff, Gavin J., Adam D. Galinsky, Edoardo Gallo, and James J. Reade (2016), “Whatever It Takes to Win: Rivalry Increases Unethical Behavior,” *Academy of Management Journal*, 59 (5), 1508–34.

Kölbel, Julian F., Timo Busch, and Leonhardt M. Jancso (2017), “How Media Coverage of Corporate Social Irresponsibility Increases Financial Risk,” *Strategic Management Journal*, 38 (11), 2266–84.

Kulik, Brian W., Michael J. O’Fallon and Manjula S. Salimath (2008), “Do Competitive Environments Lead to the Rise and Spread of Unethical Behavior? Parallels from Enron,” *Journal of Business Ethics*, 83 (4), 703–23.

Lambert, Diane (1992), “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing,” *Technometrics*, 34 (1), 1–14.

Landsman, Vardit and Stefan Stremersch (2020), “The Commercial Consequences of Collective Layoffs: Close the Plant, Lose the Brand?,” *Journal of Marketing*, 84 (3), 122–141.

Lazear, Edward P. and Sherwin Rosen (1981), “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89 (5), 841–65.

Lin, Simon M., Jake Ansell, and Galina Andreeva (2012), “Predicting Default of a Small Business Using Different Definitions of Financial Distress,” *Journal of the Operational Research Society*, 63 (4), 539–48.

Liu, Xiaoding (2016), “Corruption Culture and Corporate Misconduct,” *Journal of Financial Economics*, 122 (2), 307–27.

Liu, Yan and Venkatesh Shankar (2015), “The Dynamic Impact of Product-Harm Crises on Brand Preference and Advertising Effectiveness: An Empirical Analysis of the Automobile Industry,” *Management Science*, 61 (10), 2514–35.

McGee, Suzanne (2022), “ESG Investors' Dilemma: Think Positive or Negative? It's a Question of Shunning All Companies in a Sector or Taking a More Nuanced Approach,” *Wall Street Journal*, November 3, R. 5.

Mishina, Yuri, Bernadine J. Dykes, Emily S. Block, and Timothy G. Pollock (2010), “Why “Good” Firms Do Bad Things: The Effects of High Aspirations, High Expectations, and Prominence on the Incidence of Corporate Illegality,” *Academy of Management Journal*, 53 (4), 701–22.

Mishra, Madhurima, Koustab Ghosh, and Dheeraj Sharma (2022), “Unethical Pro-Organizational Behavior: A Systematic Review and Future Research Agenda,” *Journal of Business Ethics*, 179 (1), 63–87.

Mullahy, John (1986), “Specification and Testing of Some Modified Count Data Models,” *Journal of Econometrics*, 33 (3), 341–65.

NAICS Association (2017), “What is the Difference Between NAICS Codes and SIC Codes?,” (accessed June 15th, 2023), [available at https://www.naics.com/what-is-the-difference-between-naics-codes-and-sic-codes/].

Nippa, Michael (2011), “On the Need to Extend Tournament Theory Through Insights from Status Research,” in *Status in Management and Organizations*, Jone L. Pearce, ed. Cambridge University Press, 118–52.

Perry, Cheryl L., Tom Baranowski, and Guy S. Parcel (1990), "How Individuals, Environments, and Health Behavior Interact: Social Learning Theory," in *Health Behavior and Health Education: Theory, Research, and Practice*, Karen Glanz, F. M. Lewis, and Barbara K. Rimer, eds. Jossey-Bass/Wiley, 161–86.

Pinto, Jonathan, Carrie R. Leana, and Frits K. Pil (2008), "Corrupt Organizations or Organizations of Corrupt Individuals? Two Types of Organization-Level Corruption," *Academy of Management Review*, 33 (3), 685–709.

Poortvliet, P. Marijn, Onne Janssen, Nico W. Van Yperen, and Evert Van de Vliert (2009), "Low Ranks Make the Difference: How Achievement Goals and Ranking Information Affect Cooperation Intentions," *Journal of Experimental Social Psychology*, 45 (5), 1144–47.

RepRisk (2023), "RepRisk Methodology Overview," (accessed June 15th, 2023) [available at https://www.reprisk.com/news-research/resources/methodology].

Rego, Lopo L., Neil A. Morgan and Claes Fornell (2013), "Reexamining the Market Share–Customer Satisfaction Relationship," *Journal of Marketing*, 77 (5), 1–20.

Rosen, Sherwin (1986), "Prizes and Incentives in Elimination Tournaments," *American Economic Review*, 76, 701–15.

Ross, David R. (1986), "Learning to Dominate," *The Journal of Industrial Economics*, 34 (4), 337–53.

Rubel, Oliver, Prasad A. Naik, and Shuba Srinivasan (2011), "Optimal Advertising When Envisioning a Product-Harm Crisis," *Marketing Science*, 30 (6), 1048–65.

Rust, Roland T., Tim Ambler, Gregory S. Carpenter, V. Kumar, and Rajendra K. Srivastava (2004), "Measuring Marketing Productivity: Current Knowledge and Future Directions," *Journal of Marketing*, 68 (4), 76–89.

Saunders, Norman C. (1999), "The North American Industry Classification System: Change on the Horizon," *Occupational Outlook Quarterly*, 43 (3), 34–37.

Schweitzer, Maurice E., Lisa Ordóñez, and Bambi Douma (2004), "Goal Setting as a Motivator of Unethical Behavior," *Academy of Management Journal*, 47 (3), 422–32.

Sheth, Jagdish N. and Rajendra S. Sisodia (2002), *The Rule of Three: Surviving and Thriving in Competitive Markets*. The Free Press.

Shleifer, Andrei (2004), "Does Competition Destroy Ethical Behavior?," *The American Economic Review*, 94 (2), 414–18.

Siano, Alfonso, Agostino Vollero, Francesca Conte, and Sara Amabile (2017), "More Than Words": Expanding the Taxonomy of Greenwashing After the Volkswagen Scandal," *Journal of Business Research*, 71, 27–37.

Srinivasan, Raji, Gary L. Lilien, and Shrihari Sridhar (2011), "Should Firms Spend More on Research and Development and Advertising During Recessions?," *Journal of Marketing*, 75 (3), 49–65.

Stäbler, Samuel and Marc Fischer (2020), "When Does Corporate Social Irresponsibility Become News? Evidence from More Than 1,000 Brand Transgressions Across Five Countries," *Journal of Marketing*, 84 (3), 46–67.

Steenkamp, Jan-Benedict E. M., and Eric Fang (2011), "The Impact of Economic Contractions on the Effectiveness of R&D and Advertising: Evidence from U.S. Companies Spanning Three Decades," *Marketing Science*, 30 (4), 628–45.

Tesser, Abraham (1988), "Toward a Self-Evaluation Maintenance Framework of Social Behavior," in *Advances in Experimental Social Psychology*, Vol. 17, Leonard Berkowitz, ed. Academic Press, 181–227.

Umphress, Elizabeth E., and John B. Bingham (2011), "When Employees Do Bad Things for Good Reasons: Examining Unethical Pro-Organizational Behaviors," *Organization Science*, 22 (3), 621–40.

Uslay, Can, Z. Ayca Altintig, and Robert D. Winsor (2010), "An Empirical Examination of the "Rule of Three": Strategy Implications for Top Management, Marketers, and Investors," *Journal of Marketing*, 74 (3), 20–39.

Van Heerde, Harald, Kristiaan Helsen, and Marnik G. Dekimpe (2007), "The Impact of a Product-Harm Crisis on Marketing Effectiveness," *Marketing Science*, 26 (2), 230–45.

Verbeke, Willem, Cok Ouwerkerk, and Ed Peelen (1996), "Exploring the Contextual and Individual Factors on Ethical Decision Making of Salespeople," *Journal of Business Ethics*, 15 (11), 1175–87.

Volkswagen Group (2015), "Prof. Dr. Martin Winterkorn Speech at the Annual Media Conference and Investor Conference on March 12, 2015," (accessed April 26th, 2023), [available at https://www.volkswagenag.com/presence/investorrelation/publications/presentations/2015/03-march/2015-03-12_Teil%20I%20Jahrespressekonferenz%20&%20Investorenkonferenz%202015%20Rede%20Winterkorn%20ENGLISCH.pdf].

Vriend, Tim, Jennifer Jordan, and Onne Janssen (2016), "Reaching the Top and Avoiding the Bottom: How Ranking Motivates Unethical Intentions and Behavior," *Organizational Behavior and Human Decision Processes*, 137 (September), 142–55.

Zhao, Yi, Ying Zhao, and Kristiaan Helsen (2011), "Consumer Learning in a Turbulent Market Environment: Modeling Consumer Choice Dynamics after a Product-Harm Crisis," *Journal of Marketing Research*, 48 (2), 255–67.

APPENDIX PAPER 2

APPENDIX A: REPRISK'S RESEARCH SCOPE

The RepRisk data set has already been analyzed and described in detail in Paper 1 of this dissertation. To avoid repetition, see Appendix A of Paper 1 for a description of RepRisk's research scope.

APPENDIX B: NUMBER OF UNETHICAL BEHAVIOR INCIDENTS PER COMPANY AND YEAR

Table A1 reports the number of unethical behavior incidents per year and company for all industries. Note that Table A1 includes more than the 79 industries analyzed as we lose industries after deleting missing values.

Table A1: Number of Unethical Behavior Incidents per Company and Year

Industry	**Number of Unethical Behavior Incidents**
Rental and Leasing Services	.50
Merchant Wholesalers, Durable Goods	.55
Broadcasting (except Internet)	.56
Miscellaneous Manufacturing	.57
Professional, Scientific, and Technical Services	.57
Water Transportation	.58
Machinery Manufacturing	.59
Textile Product Mills	.59
Motor Vehicle and Parts Dealers	.65
Wood Product Manufacturing	.68
Web Search Portals, Libraries, Archives, and Other Information Services	.75
Accommodation	.76
Construction of Buildings	.78
Administrative and Support Services	.84
Food and Beverage Retailers	.86
Freight Transportation On The Great Lakes-st. (SIC)	.97
Insurance Carriers and Related Activities	.98
Heavy and Civil Engineering Construction	.99
Hospitals	1.04
Apparel Manufacturing	1.08
Computer and Electronic Product Manufacturing	1.10
Administrative and Support Services	1.10
Nonmetallic Mineral Product Manufacturing	1.12
Transit and Ground Passenger Transportation	1.16
Miscellaneous Store Retailers	1.22

Table A1: Number of Unethical Behavior Incidents per Company and Year

Industry	Number of Unethical Behavior Incidents
Deep Sea Domestic Transportation Of Freight (SIC)	1.23
Chemical Manufacturing	1.27
Merchant Wholesalers, Nondurable Goods	1.30
Plastics and Rubber Products Manufacturing	1.38
Paper Manufacturing	1.56
Health and Personal Care Stores	1.67
Telecommunications	1.67
Mining (except Oil and Gas)	1.72
Building Material and Garden Equipment and Supplies Dealers	1.88
Food Services and Drinking Places	1.89
Clothing and Clothing Accessories Stores	1.95
Electrical Equipment, Appliance, and Component Manufacturing	2.45
Food Manufacturing	2.51
Nonstore Retailers	2.52
Air Transportation	2.60
Beverage and Tobacco Product Manufacturing	2.71
Securities, Commodity Contracts, and Other Financial Investments and Related Activities	2.82
Nonclassifiable Establishments (SIC)	3.05
Couriers and Messengers	3.05
Oil and Gas Extraction	3.24
Rail Transportation	3.42
Credit Intermediation and Related Activities	4.09
Support Activities for Mining	4.12
Leather and Allied Product Manufacturing	4.19
Utilities	4.38
Primary Metal Manufacturing	5.61
Transportation Equipment Manufacturing	7.46
Wholesale Trade	8.11
Trucking And Courier Services, Except Air (SIC)	8.82
General Merchandise Stores	11.57
Pipeline Transportation	13.95
Petroleum and Coal Products Manufacturing	32.26

APPENDIX C: TESTS FOR ZERO INFLATION, OVERDISPERSION, AND ALTERNATE MODEL TYPES

In the first step, we check our model for overdispersion. Overdispersion occurs when the observed variance is higher than the variance of a theoretical model. For Poisson models, variance increases with the mean and, therefore, variance usually (roughly) equals the mean value. If the variance is much higher, the data are "overdispersed". If the dispersion ratio is close to one, a Poisson model fits well to the data. Dispersion ratios larger than one indicate overdispersion, thus, a negative binomial model fits better to the data. A p-value < .05 indicates overdispersion (Gelman and Hill 2007, p. 115).

We fit a Poisson model to check for overdispersion. Results are shown in Table A2. Overdispersion is detected for both cases, excluding and including zeros, and thus, a negative binomial model fits better to our data.

Table A2: Check for Overdispersion

	Including the zeros	**Excluding the zeros**
dispersion ratio	148.669	1.962
Pearson's Chi-Squared	2,978,732.732	10,686.139
p-value	< .001	< .001

In the second step, we test for zero inflation. Note that the existence of many zeros in our dependent variable is not necessarily proof of zero inflation (Warton 2005). If the number of observed zeros is larger than the number of predicted zeros, the model is underfitting zeros, which indicates a zero inflation in the data. In such cases, it is recommended to use a zero-inflated model. We detect a slight zero-inflation as the model includes 14,589 zeros and predicts 14,028 zeros showing that the data has more zeros than expected.

As we assume that there is only one potential source for zero observations, we fit a hurdle negative binomial model. As the model shows only a slight zero inflation, we also checked if the HNB model outperforms an ordinary poisson and a negative binomial. The hurdle negative model also produces a lower AIC (main model = 37,766) value, i.e., a better model fit, than other model types further supporting the hurdle negative binomial model (see Table A3 for AIC and BIC values). If the difference in AIC values of the two models is larger than 10, there is strong evidence to prefer the model with the smaller AIC value (Burnham and Anderson 2004).

Table A3: Comparison of Model Fits (Main Models)

Type of model	AIC value	BIC value
Hurdle negative binomial	37,766.0	38,216.7
Negative binomial	38,240.7	38,470.0
Poisson	43,679.1	43,900.5

APPENDIX D: VIF VALUES

Multicollinearity might pose a concern for the HNB. Variance inflation factor (VIF) values indicate how high the risk is that multicollinearity is an issue. Table A4 presents the VIF values of the HNB model, whereas Table A5 shows the VIF values of the market share response model. All values of the relevant competition are well below 10. A commonly used rule of thumb is that a VIF of 10 or more is evidence of severe multicollinearity (Cohen et al. 2003, p. 423; Kutner et al. 2004, p. 387). We note that in the count component model, previous unethical behavior and the yearly fixed effects show VIF values slightly above 10. As these are only control variables and both have a significant impact on unethical firm behavior, we keep them in our models. Note that a model without excluding one of the two variables also leads to worse model performance.

Table A4: VIF Values for the Main HNB Model

	VIF values	
	Zero component	**Count component**
Focal Competition Variables:		
Position within industry:		
Market leader	1.36	1.22
Top 3 companies	1.60	1.21
Ranking	1.00	1.07
Direct competition:		
Market share lead	1.31	1.41
Market share behind	1.27	2.23
Competitive dynamics:		
Trend ranking	1.00	1.02
Trend market share lead	1.05	1.19
Trend market share behind	1.03	1.83
Controls:		
Company size	1.40	1.89
Leverage	1.27	1.13
Profitability	1.27	1.22
Advertising	1.06	1.37
Google search	1.01	1.34
Relationship equity	1.13	1.09
R&D	1.10	1.07
Industry concentration	1.12	1.41
Previous unethical behavior	1.18	10.21

Table A5: VIF Values for the Market Share Response Model

	VIF values
Focal Variable:	
Previous unethical firm behavior	1.10
Controls:	
Company size	1.14
Leverage	1.07
Profitability	1.07
Advertising spending	1.09
Google search	1.02
Relationship equity	1.05
R&D	1.01
Industry concentration	1.09

REFERENCES APPENDIX PAPER 2

Burnham, Kenneth P. and David R. Anderson (2004), *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*, 2nd ed. Springer Science & Business Media.

Cohen, Jacob, Patricia Cohen, Stephen G. West, Leona S. Aiken (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. Erlbaum.

Gelman, Andrew and Jennifer Hill (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Kutner, Michael H., Christopher J. Nachtsheim, John Neter, William Li (2004), *Applied Linear Regression Models*, 4th ed. McGraw-Hill.

Warton, David I. (2005), "Many Zeros Does Not Mean Zero Inflation: Comparing the Goodness-of-Fit of Parametric Models to Multivariate Abundance Data," *Environmetrics: The Official Journal of the International Environmetrics Society*, 16 (3), 275–89.

Paper 3: Machine Learning in Marketing – A Review of Recurring Problems and How to Solve Them

Author: Lars Gemmer

Abstract

Machine learning is an increasingly applied methodology to make classifications in marketing research. A large number of studies underpin the relevance and complexity of machine learning-driven methods. The challenges that marketing researchers face when applying machine learning methods are often comparable. For example, highly skewed (or imbalanced) data that involve only a few observations within one of the classes are quite common (e.g., within churn prediction, prediction of unethical firm behavior, and business failure). A few existing marketing-related machine learning reviews pertain to the analysis of the research problem or the applied algorithms. This literature review adds to these studies by identifying several recurring machine learning-based challenges in the field of marketing such as imbalanced target variables, cost-sensitive learning, or the selection of performance metrics and offering potential solutions. Thereby, marketing-related research questions are combined with insights from computer science. As the presented challenges are independent of the actual research question, the insights from this literature review can be easily transferred to other research problems. Hence, this study does not only provide value for future research but also for the practical application of machine learning in companies.

Keywords: Machine learning in marketing, review, imbalanced target variable, feature selection, performance metrics, hyperparameter tuning, cost-sensitive learning, interpretable machine learning

1 Introduction

According to Mitchell (1997, p. 2), machine learning can be described as follows: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". In other words, machine learning is the process of mathematical algorithms learning patterns or trends in recorded data and then making predictions for future data observations (Kirasich, Smith, and Sadler 2018). Nowadays, machine learning is typically considered a subfield of artificial intelligence (Kumar et al. 2021).

Machine learning-based methods have been established in the scientific context for decades (Rumelhart, Hinton, and Williams 1986; Quinlan 1986). However, the currently observable trend towards the increasing application of machine learning methods in research and analytics departments of companies is mainly driven by easier access to large amounts of data, higher computing power of computers, and the continuous development of machine learning-based algorithms (Lantz 2019).

Thereby, the application possibilities of machine learning are not limited to specific thematic fields but can provide solutions across various research areas. Cross-disciplinary application examples for machine learning in science can be found in earthquake prediction (Asim et al. 2018; Mignan and Broccardo 2020), medical diagnosis (Kononenko 2001), crime prediction (Jordan and Mitchell 2015; Wang, Gerber, and Brown 2012), or bankruptcy prediction (Devi and Radhika 2018; Wang 2017). Regardless of the application field and the specific research question or business problem, machine learning aims to make predictions or classifications.

Increased relevance of machine learning applications is also observed in marketing science, though it has to be noted that in the field of marketing, machine learning is still at an early stage (Ma and Sun 2020). However, an increasing number of marketing researchers apply machine learning-based methods in their studies. Lemmens and Croux (2006) and Lemmens and Gupta (2020), for example, show how machine learning can be implemented to predict customer churn. Churn prediction is generally one of the most important marketing tasks for the application of machine learning approaches.

In line with this academic importance, machine learning is also taking an essential role in marketing-related decisions in companies (Ngai and Wu 2022). As Ma and Sun (2020) illustrate, machine learning affects practical decision-making processes across the whole marketing mix. As digitization has enabled companies access to large amounts of data that can be analyzed, they invest heavily in machine learning to enhance their marketing capabilities. Ngai and Wu (2022) elaborate on the prominent role of machine learning for businesses in their comprehensive literature review of machine learning applications in marketing. There are no signs that the methodology will lose any of its relevance within marketing in the future (Rust 2020).

Both, in marketing science and marketing practice, adopters of machine learning methods are frequently confronted with several recurring challenges. The above-introduced example of churn prediction highlights some of these challenges: The classification problem usually involves the prediction of rarely occurring events – the customer churns – which causes a high imbalance of the target variable (Lemmens and Croux 2006). Forecasting these very rare events often requires certain approaches to successfully enable a satisfying classification.

Moreover, existing studies dealing with churn prediction elaborate on two other recurring challenges: first, they highlight the role of interpretable machine learning (Lemmens and Croux 2006), and second, they introduce cost-sensitive machine learning to the classification task (Lemmens and Gupta 2020). For these hurdles, there exists a multitude of potential problem-solving approaches in computer science, which are often difficult for users to overview.

As mentioned, several recent studies already underpin the theoretical and practical relevance of the topic by comprehensively reviewing existing machine learning literature from the marketing field (Ma and Sun 2020; Ngai and Wu 2022). In contrast to these existing reviews on the general role of machine learning in marketing, this study aims to present potential approaches to solve recurring machine learning-based problems occurring in typical marketing tasks, such as imbalanced target variables, cost-sensitive learning, interpretable machine learning, as well as the selection of the performance metrics. Based on the initial identification of frequently occurring challenges of machine learning applications in marketing, this study overviews various approaches to solve these obstacles. In this way, the study guides both, researchers and users from the practice within the field of marketing and beyond.

The existing reviews on the machine learning literature highlight the complexity and multi-faceted nature of the topic. To reduce this complexity, the following work focuses on the challenges of classification tasks and proposes solutions. Challenges within regression problems are not covered. Classification tasks are very common in marketing (Lemmens and Croux 2006) and thus specifically relevant for marketing research and practice but also for other disciplines (e.g., Viaene and Dedene 2005). The solution approaches outlined in this review are mainly limited to the handling of structured data. However, it cannot be ruled out that they can also be applied to the analysis of unstructured data, such as text, images, audio, and video.

My study obtains important methodological and practical contributions. From a methodological viewpoint, the study overviews (methodological) solutions for researchers to overcome common challenges in machine learning. In this way, the review facilitates scientific working by providing methodological guidance for researchers. Additionally, the study contributes to marketing research, as it combines relevant machine learning-based problems from the marketing field with methodological solution approaches from computer science. Note that the terminology in marketing and computer science differs. Since my study links research from both fields, it is important to clarify the terminology. I explain relevant machine learning expressions in Appendix A to enable a better understanding of the subsequent study.

From a practical point of view, companies can benefit from the approaches presented in this study. My findings enable companies to apply machine learning more efficiently and effectively. In this way, the analytical foundation of marketing mix decisions can be improved, which creates additional monetary value for companies (e.g., through cost-sensitive machine learning or the selection of the right performance metrics).

The remainder of this paper is organized as follows. I develop the conceptual model based on a summary of important classification tasks in marketing and their accompanying challenges in section 2. Next, section 3 presents an overview of potential solutions for these challenges. Section 4 concludes by summarizing the main findings and limitations of this study and giving an outlook for future research.

2 Conceptual Framework

2.1 The Basic Learning Process

Regardless of whether the learner is a human or a machine, the basic learning process is similar. As shown in Figure 1, it can be divided into three main components (Figure retrieved from Lantz 2019, p. 10): The data input, the abstraction, and the generalization. I use this general learning process and its three components as a guiding framework to categorize the machine learning-based challenges and to develop a conceptual model.

The *data input* component represents the factual basis for future decisions (see Lantz 2019 for more details). It includes observations, memory, and recall. In this first step of the learning process, raw input data is the quintessential task for a learning algorithm. Before this point, the data has no meaning. The *abstraction* component involves the translation of the input data. During this step, a meaning is assigned to the data. When the (machine learning) model has been trained, the data has been transformed into an abstract form that summarizes the original information. Within the *generalization*, the abstracted data form a basis for action. The learning process is not complete until the learner can use its abstract knowledge for future action. The final step in the generalization process is to evaluate the model's performance. Note that this basic learning approach is a simplified description presented by Lantz (2019). There exist much more complex learning processes (e.g., Gentile, Groves, and Gentile 2014). However, these complex descriptions of learning can also be broken down into the three above-described main components.

Figure 1: The Basic Learning Process

Notes: Figure from Lantz (2019, p. 10).

2.2 Identification of Machine Learning-Based Challenges within the Field of Marketing

I use the three components of the learning process to classify the machine learning-based problems that frequently arise in the field of marketing research. Table 1 provides an overview of these marketing-related studies which deal with machine learning-based challenges. Based on the presented literature, I identify six different challenges (see also Figure 2). For each challenge, the studies are a representative selection to illustrate the complexity and relevance of the specific challenge for marketing research. In addition to the information about the authors, the journal, and the publication date, Table 1 summarizes the concrete marketing-related business problem the studies deal with as well as the applied solution approach. Note that there remain other less frequently occurring issues (e.g., the selection of the machine learning algorithm), but they are beyond the scope of this study and may be addressed in future research.

Table 1: Representative Marketing Studies Dealing with Recurring Challenges of Machine Learning

Type of challenge	Authors	Journal, publication year	Marketing-related business problem	
				Degree of imbalance
Data-based: Imbalanced target variable	Donkers, Franses, and Verhoef	Journal of Marketing Research, 2003	Churn prediction	2.5%
	Lemmens and Croux	Journal of Marketing Research, 2006	Churn prediction	1.8%
	Neslin et al.	Journal of Marketing Research, 2006	Churn prediction	1.8%
	Berger and Magliozzi	Journal of Direct Marketing, 1992	Direct marketing for household tools	.75%
				Feature selection approach
Data-based: Feature selection	Buckinx and Van den Poel	European Journal of Operational Research, 2005	Churn prediction	Conceptual derivation
	Idris, Rizwan, and Khan	Computers & Electrical Engineering, 2012	Churn prediction	Principle component analysis, Fisher's ratio, F-score, minimum redundancy, and maximum relevance
	Huang, Kechadi, and Buckley	Expert Systems with Applications, 2012	Churn prediction	Conceptual derivation
	Lalwani et al.	Computing, 2022	Churn prediction	Gravitational search algorithm

Table 1: Representative Marketing Studies Dealing with Recurring Challenges of Machine Learning

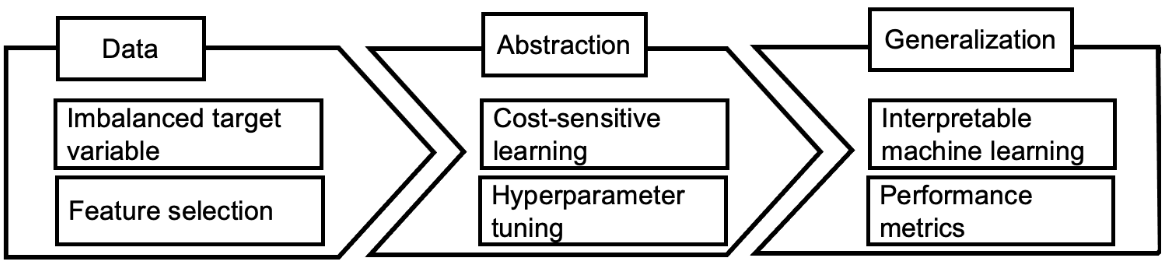
Type of challenge	Authors	Journal, publication year	Marketing-related business problem	
				Cost-sensitive learning approach
Abstraction-based: Cost-sensitive learning	Lemmens and Gupta	Marketing Science, 2020	Churn prediction	Profit-Based loss function
	Cui, Wong, and Wan	Journal of Management Information Systems, 2012	Direct marketing	Sampling
	Coussement	European Journal of Marketing, 2014	Churn prediction	Direct minimum expected cost (DMEC), metacost, thresholding, and weighting
	Bahnsen, Aouada, and Ottersten	Expert Systems with Applications, 2015	Churn prediction	Cost-sensitive machine learning algorithm
				Tuning method
Abstraction-based: Hyperparameter tuning	Coussement and Van den Poel	Expert Systems with Applications, 2008	Churn prediction	Grid search
	Kumar, Rao, and Soni	Marketing letters, 1995	Decisions by supermarket buyers whether to add a new product to their shelves or not	Trial and error
	Sarkar and De Bruyn	Journal of Interactive Marketing, 2021	Direct marketing	Bayesian optimization
	Sana et al.	PLoS ONE, 2022	Churn prediction	Grid search

Table 1: Representative Marketing Studies Dealing with Recurring Challenges of Machine Learning

Type of challenge	**Authors**	**Journal, publication year**	**Marketing-related business problem**	**Approach**
Generalization-based: Interpretable machine learning	Lemmens and Croux	Journal of Marketing Research, 2006	Predicting customer churn	Relative feature importance, Partial dependence plot
	Campbell and Shang	Management Science, 2022	Predicting corporate misconduct	Feature importance
	Coussement and Van den Poel	Expert Systems with Applications, 2008	Churn prediction	Feature importance
	Wang et al.	Information Systems Research, 2022	Marketing campaigns	e.g., SHAP
	Naumzik, Feuerriegel, Weinmann	Marketing Science, 2022	Predicting business failure	Feature importance
				Applied performance metrics
Generalization-based: Selection of performance metric	Campbell and Shang	Management Science, 2022	Predicting corporate misconduct	Pseudo-R^2, ROC AUC
	Lemmens and Croux	Journal of Marketing Research, 2006	Predicting customer churn	Top-Decile lift, gini coefficient, error rate
	Naumzik, Feuerriegel, Weinmann	Marketing Science, 2022	Predicting business failure	ROC AUC, balanced accuracy, F_1-score, specificity, sensitivity
	Lemmens and Gupta	Marketing Science, 2020	Churn prediction	Gini coefficient, top-decile lift
	Buckinx and Van den Poel	European Journal of Operational Research, 2005	Churn prediction	Percentage correctly classified (PCC), ROC AUC
	Idris, Rizwan, and Khan	Computers & Electrical Engineering, 2012	Churn prediction	Sensitivity, specificity, and ROC AUC

Figure 2 overviews the six challenges and shows to which components of the general learning process the challenges belong. The data input component includes the issue of an *imbalanced target variable* and the *variable selection* process. Both challenges are related to the data used to train the algorithms. The abstraction process deals with *cost-sensitive learning* and the *tuning of the algorithm-specific hyperparameters* (see Abstract 1 for definitions). The generalization component includes *interpretable machine learning* and the overview of different *metrics for the evaluation* of the model performance.

Figure 2: Challenges of Machine Learning Tasks within the General Learning Process



Note that the classification of these challenges into the learning process components is not always completely sharp. For example, an imbalanced target variable also affects the choice of a corresponding performance metric or the cost-sensitive learning process. Nevertheless, the origin of the problem lies in the structure of the data (e.g., the occurrence of a rare event) and therefore the classification of the problems in Figure 2 is still reasonable. However, it should be viewed with the appropriate amount of flexibility.

In the following section, I explain the six machine learning-based challenges by referring to the selected marketing literature. Then, I overview potential solution approaches for these challenges.

3 Review of Recurring Challenges

3.1 Data-Based Challenges

3.1.1 Imbalanced target variable. Many classifications task in marketing are characterized by a highly imbalanced target variable meaning that the class distribution (e.g., 0 and 1) is extremely unequal with one class occurring very rarely. One of the most frequently classification tasks in marketing is the prediction of customer churners. Customer churn describes the decision of a current customer to leave the current company (Lemmens and Croux 2006). A way to manage customer churn is to predict which customers are most likely to churn and then target incentives to those customers to make them stay (Neslin et al. 2006). This approach enables the firm to focus its efforts on customers who are likely to churn and it potentially saves money that would be wasted in providing incentives to customers who do not need them. However, the approach assumes that customer churn can be predicted with acceptable accuracy.

Most machine learning algorithms work best when the number of samples in each class is equal. The reason for the problem of an imbalanced target variable is that the minority class occurs very rarely in the training sample (Branco, Torgo, and Ribeiro 2016). The actual problem can be described by the following two assertions: First, the machine learning applicant assigns more importance to the predictive performance on the minority class (e.g., identifying the customer churns is more important than identifying the non-churners). Second, the classes that are more important for the applicant are poorly represented in the training set, which can cause non-optimal models (Branco, Torgo, and Ribeiro 2016). In other words, if the target variable is imbalanced then the machine learning models may achieve a very high classification accuracy just by predicting the majority class (e.g., no customer churn, no business failure), but they fail to predict the minority class (e.g., customer churn, business failure), which is the main goal of the model in the first place.

3.1.2 Feature selection. Users of machine learning methods face the challenge of identifying an optimal subset of possible features to be used in the models. Recall that in machine learning, a feature refers to a variable that describes some aspect of individual data objects and represents the input data to generate the prediction output (Dong and Liu 2018). Data in machine learning-based classifications are often multidimensional, which may be challenging for the analysis of the data (Cai et al. 2018; Sarkar and De Bruyn 2021).

Feature selection has been proven in both theory and practice effective in processing high-dimensional data and in enhancing learning efficiency (Blum and Langley 1997; Cai et al. 2018; Guyon and Elisseeff 2003). Feature selection describes the process of obtaining a subset from an original feature based on selection criteria to remove the irrelevant features. Feature selection techniques can pre-process learning algorithms, and good feature selection results can improve learning accuracy, reduce learning time, and simplify learning results (Cai et al. 2018; Kohavi and John 1997; Langley 1994; Zhao et al. 2010).

There are three main objectives of selecting a subset of features: First, feature selection can increase the prediction performance of the predictors. Second, a subset of features can provide faster and more cost-effective predictors. Lastly, feature selection can generate a better understanding of the underlying process that generated the data (Guyon and Elisseeff 2003).

Since machine learning users usually face the problem of deciding which features to include in their classification models, feature selection is also a recurring challenge in the field of marketing-related classifications (see Kohavi and John 1997, p. 275ff. for more details about the underlying problem of feature selection).

3.2 Abstraction-Based Challenges

3.2.1 Cost-sensitive learning. Cost-sensitive learning takes different costs, such as the different misclassification costs, resulting from a classification into consideration. It is one of the most active research areas in machine learning, and it plays an important role in real-world data mining applications (Domingos 1999; Ling and Sheng 2011), and thus, also for marketing tasks (e.g., Lemmens and Gupta 2020). The main aim of cost-sensitive learning is not to maximize the classification performance (e.g., the classification accuracy) but to minimize the resulting costs which are caused by the misclassification (Viaene and Dedene 2005). In that way, users are enabled to make cost-benefit-wise optimal decisions.

I explain the basic idea of cost-sensitive learning with the help of a churn classification. Table 2 overviews the misclassification costs for every potential outcome of the classification. In case of a correct classification (i.e., an actual churner is detected or a non-churner is classified as a non-churner), there are no misclassification costs. However, if an actual churner is classified as a non-churner, the company loses profits as it is not able to prevent the customer quits the business. In case a non-churner is classified as a churner, the company loses profits because it invests in unnecessary measures to prevent the churn (e.g., direct marketing activities). Both misclassification costs are of different sizes.

Table 2: Misclassification Costs for a Churn Classification

		Actual churn	
		Churn	**No churn**
Predicted churn	**Churn**	**No misclassification costs**	**Costs of falsely predicting a churn** (e.g., unnecessary actions to make the customer stay)
	No churn	**Costs of missing an actual churn** (e.g., losing profits because a customer churns)	**No misclassification costs**

Note that there is a strong relationship between imbalanced problems and cost-sensitive learning (Elkan 2001) as both result from the non-uniform preference biases of the user. However, a cost-sensitive problem may not be imbalanced if the more relevant cases (e.g., churners) are sufficiently represented in the data. This means that an imbalanced problem always involves unequal costs of the different misclassification errors, but the opposite is not always true (Branco, Torgo, and Ribeiro 2016). There is a variety of methodological ways to incorporate the different levels of misclassification costs to produce cost-optimal results (see Table 1).

3.2.2 Hyperparameter tuning. Most machine learning models have hyperparameters that require tuning (e.g., Snoek, Larochelle, and Adams 2012). Hyperparameters are those parameters within a model that are determined by the user to influence the learning process (see Appendix A). Hyperparameters have the potential to improve the learning of the model, and thus, they are determined before starting the learning process of the model (Kuhn and Johnson 2013). Examples of hyperparameters are the number of hidden units in a neural network, the batch size, the train-test split ratio, or the number of trees in a random forest (see Sarkar and De Bruyn 2021; Luo 2016, p. 2 for an overview of different hyperparameters for various machine learning approaches).

The process of selecting the best hyperparameter combination is called hyperparameter tuning (Goodfellow, Bengio, and Courville 2016). The challenge of identifying the optimal hyperparameter values is relevant for all machine learning-based applications and therefore also for the field of marketing research (e.g., Htet and Sein 2020; Sarkar and De Bruyn 2021). The parameter optimization procedure plays an important role for the predictive performance. For example, Coussement and Van den Poel (2008) show that the parameter-selection procedure matters for the identification of the machine learning algorithm with the highest prediction performance.

The number of hyperparameters as well as their range increase (e.g., Sarkar and De Bruyn 2021). As a consequence, the search space for the optimal hyperparameter combination becomes extremely large, which makes a manual search process challenging. In addition, the tuning process often requires knowledge of machine learning algorithms and the appropriate hyperparameter optimization solutions (Yang and Shami 2020). There is a wide range of methodological solutions available for identifying the best hyperparameter settings.

3.3 Generalization-Based Challenges

3.3.1 Interpretable machine learning. The core objective of machine learning is to predict as accurately as possible. So-called black-box (e.g., deep neural networks) algorithms are often used to maximize the prediction performance. A disadvantage of these black-box models is the fact that it is difficult or even impossible for the user to understand how the algorithm arrives at its final result (Burkart and Huber 2021; Marcinkevičs and Vogt 2023). However, to be able to make appropriate recommendations for action, it is often necessary to understand the relationships between the features and the target variable. Machine learning users usually need to trade-off between algorithms that are less complex and predict less well, but are easier to interpret (e.g., a logit model), and algorithms that predict very well but deliver less transparent results (e.g., random forest, deep neural networks) (e.g., Lundberg and Lee 2017; Rudin 2019).

In the context of interpretable machine learning, users are now seeking to understand how the algorithm arrived at its results. This is especially important for the field of marketing as this field constantly aims to produce findings that offer actionable value for managers (e.g., Burkart and Huber 2021, p. 250; Lemmens and Croux 2006; Naumzik, Feuerriegel, Weinmann 2022).

3.3.2 Performance metrics. Many different measures to evaluate the performance of a machine learning model exist (e.g., Branco, Torgo, and Ribeiro 2016). Certain measures reveal information about certain aspects of the machine learning model's performance. For rare events, as Morrison (1969) notes, the simple error rate or accuracy is often misleading. In addition, there are performance metrics that evaluate a binary classification independent of a predefined classification threshold (e.g., ROC AUC, PR AUC). These metrics measure the classification performance for all possible thresholds between 0 and 100 percent. Note that I define all performance metrics in detail in the subsequent solution section about the performance metrics. Other metrics, such as precision or recall, evaluate model performance with a focus on one of the existing classes of the target variable.

Applicants of machine learning algorithms face the challenge to identify the metric that fits best to their research problem and their individual preferences. Table 1 illustrates the complexity of the selection process by showing how numerous the different performance metrics in the marketing discipline are.

4 Review of Potential Solutions

4.1 Approaches for Imbalanced Target Variables

An imbalanced target variable often poses difficulties for machine learning applications. Many existing studies aim to tackle this problem. An all-encompassing literature review of every single approach is beyond the scope of this paper (see Bonas et al. 2021; Branco, Torgo, and Ribeiro 2016 for exemplary literature reviews). Hence, various methodological options provide an effective way of addressing an imbalanced target variable. Branco, Torgo, and Ribeiro (2016) identify four different types of solution strategies (Table 3).

Data pre-processing is one potential way of handling imbalanced target variables. Here, the distribution of the data is changed before it is analyzed by the machine learning algorithm, or different weights are assigned to the individual classes in accordance with user preferences (e.g., Zadrozny, Langford, and Abe 2003). Changing the distribution of the data can be done in two ways: Upsampling artificially increases the number of observations with the minority class (e.g., Bonas et al. 2021; Elkan 2001). Downsampling, on the other hand, reduces the number of observations of the majority class. Note that it is not automatically optimal to keep the number of observations of each class the same, but there is an optimal proportion of observations with the respective classes which corresponds to users' preferences (Lemmens and Croux 2006).

Table 3: Solution Strategies for Imbalanced Target Variables by Branco, Torgo, and Ribeiro (2016)

Data pre-processing	Special-purpose learning methods	Prediction post-processing	Hybrid methods
• Distribution change (e.g., down- or upsampling). • Weighting the data space (when cost-sensitive learning is applied).	• Solutions that modify existing algorithms to provide a better fit to the user preferences (e.g., assigning different penalties to false negatives and false positives). • Various solutions for different machine learning algorithms.	• Threshold Method • Cost-sensitive post-processing • Selection of performance metrics[a]	• Try to capitalize on some of the main advantages of the different approaches • Combine the use of pre-processing approaches with special-purpose learning algorithms.

Notes: Classification based on Branco, Torgo, and Ribeiro (2016). There is an overlap between solutions for imbalanced target variables and cost-sensitive learning approaches because both challenges are closely related.
[a] I added this point to the overview of Branco, Torgo, and Ribeiro (2016).

Special-purpose learning methods change the machine learning algorithms to learn from imbalanced target variables (Branco, Torgo, and Ribeiro 2016). For example, the magnitude of the penalties that the algorithm assigns to the misclassifications of the individual classes could be changed so that it corresponds to the preferences of the machine learning users. This strategy requires a high level of knowledge about the algorithms and also the preferences of the users. Furthermore, there is a high overlap with approaches for cost-sensitive learning.

The third solution strategy for imbalanced target variables, *prediction post-processing*, does not change the data or the algorithm but incorporates the users' preferences in the final predicted probabilities. For example, the chosen classification threshold (see Appendix A) can be changed in accordance with the preferences of the user (Sheng and Ling 2006). Varying the threshold value enables the user to decide on the trade-off between false positive and false negative classifications. In addition, the results of the "standard" evaluation metrics like the accuracy will not measure the performance of the model on these rare cases (Branco, Torgo, and Ribeiro 2016, see also section about performance metrics). Thus, it is essential to select the "right" performance metric if users face a classification task including an imbalanced target variable. I describe potential metrics for imbalanced data sets in the section about performance metrics.

Hybrid methods combine the advantages of the first three types of solutions for imbalanced target variables. For a more detailed overview of each solution strategy including its advantages and disadvantages as well as a review of relevant studies, see Branco, Torgo, and Ribeiro (2016).

4.2 Feature Selection Methods

In general, feature selection methods are divided into two overarching categories: *filter methods* and *wrapper methods* (Kohavi and John 1997; Michalak and Kwasnicka 2006). *Wrapper methods* evaluate feature sets based on the performance of a machine learning algorithm. *Filters,* on the other hand, consider only the characteristics of the features to identify the optimal subset of features (see Kohavi and John 1997, p. 280ff. for a review of different approaches).

The most general approach to narrow down the set of potentially relevant features for machine learning models is to look for *conceptual relationships* (e.g., Buckinx and Van den Poel 2005, Campbell and Shang 2022 for marketing-based examples). By examining the existing academic literature, users may identify features that are likely to affect the classification of the target variable. Another common method to limit the number of features is to analyze the *correlations* between the features (e.g., Michalak and Kwasnicka 2006).

Guyon and Elisseeff (2003, p. 1159) propose a heuristic checklist in which they summarize the steps that may be taken to solve a feature selection problem. This checklist includes ten questions and recommendations for action shown in Table 4 and offers a valuable point of orientation for users.

Table 4: Heuristic Checklist for Feature Selection by Guyon and Elisseeff (2003)

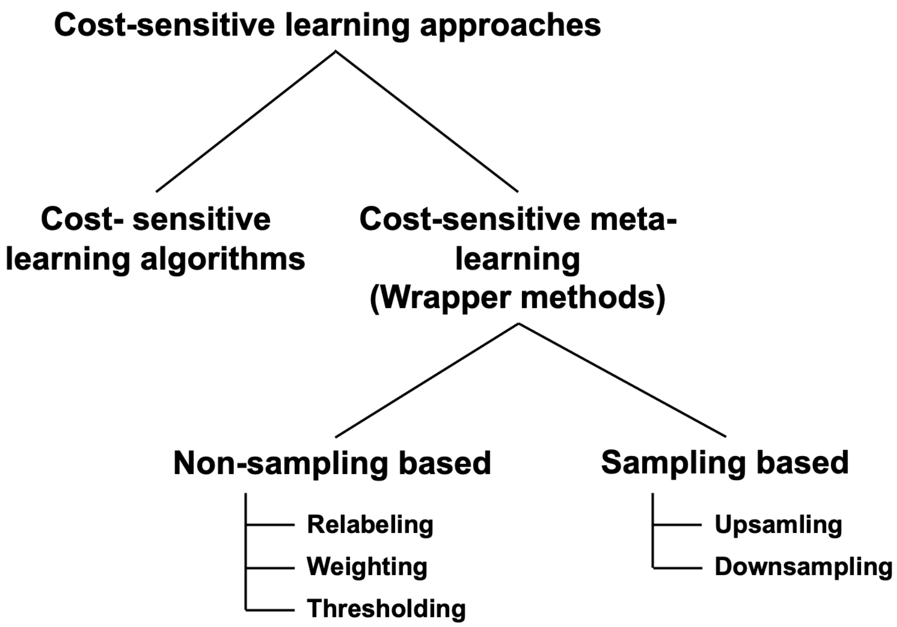
Question	Recommendation for feature selection
1. "Do you have domain knowledge?"	"If yes, construct a better set of "ad hoc" features."
2. "Are your features commensurate?"	"If no, consider normalizing them."
3. "Do you suspect interdependence of features?"	"If yes, expand your feature set by constructing conjunctive features or products of features, as much as your computer resources allow you […]."
4. "Do you need to prune the input variables (e.g., for cost, speed or data understanding reasons)?"	"If no, construct disjunctive features or weighted sums of features (e.g., by clustering or matrix factorization […])."
5. "Do you need to assess features individually (e.g., to understand their influence on the system or because their number is so large that you need to do a first filtering)?"	"If yes, use a variable ranking method […]; else, do it anyway to get baseline results."
6. "Do you need a predictor?"	"If no, stop."
7. "Do you suspect your data is "dirty" (has a few meaningless input patterns and/or noisy outputs or wrong class labels)?"	"If yes, detect the outlier examples using the top ranking variables obtained in step 5 as representation; check and/or discard them."
8. "Do you know what to try first?"	"If no, use a linear predictor. Use a forward selection method […] with the "probe" method as a stopping criterion […] or use the ℓ_0-norm embedded method […]. For comparison, following the ranking of step 5, construct a sequence of predictors of same nature using increasing subsets of features. Can you match or improve performance with a smaller subset? If yes, try a non-linear predictor with that subset."
9. "Do you have new ideas, time, computational resources, and enough examples?"	"If yes, compare several feature selection methods, including your new idea, correlation coefficients, backward selection and embedded methods […]. Use linear and non-linear predictors. Select the best approach with model selection […]."
10. "Do you want a stable solution (to improve performance and/or understanding)?"	"If yes, subsample your data and redo your analysis for several "bootstraps" […]."

Notes: Checklist from Guyon and Elisseeff (2003, p. 1159). See study from Guyon and Elisseeff (2003) for more details about individual methods mentioned in Table 4.

4.3 Making Classifications Cost-Sensitive

Various methods exist to make the classification cost-sensitive (Sheng and Ling 2006). Figure 3 classifies the different methods into *cost-sensitive learning algorithms* and *cost-sensitive meta-learning methods*. Cost-sensitive learning algorithms incorporate the different misclassification costs in the machine learning algorithm (Turney 1995, summarized in Sheng and Ling 2006). For example, Lemmens and Gupta (2020) implement a profit-based loss function in their algorithm to maximize the campaign profit for a churn prediction.

Figure 3: Types of Cost-Sensitive Learning Techniques Based on Sheng and Ling (2006)



Notes: Classification from Sheng and Ling (2006).

Cost-sensitive meta-learning algorithms are insensitive to the different costs. They incorporate the costs by pre-processing the data or by post-processing the predicted output (i.e., the predicted class probability). Within this category of methods, a further distinction can be made between *sampling-based* and *non-sampling-based* methods. *Sampling-based cost-sensitive learners* increase the proportion of one class in the training data set to improve prediction performance by duplicating one class (upsampling) or randomly deleting the other class (downsampling) (Elkan 2001; Viaene and Dedene 2005). Recall that this strategy is also applied in case of an imbalanced target variable. Sampling-based approaches have several disadvantages (Elkan 2001). For example, they distort the distribution of target classes, which may affect the performance of some classification algorithms. In addition, downsampling reduces the data available for training and upsampling may increases the learning time of the algorithms.

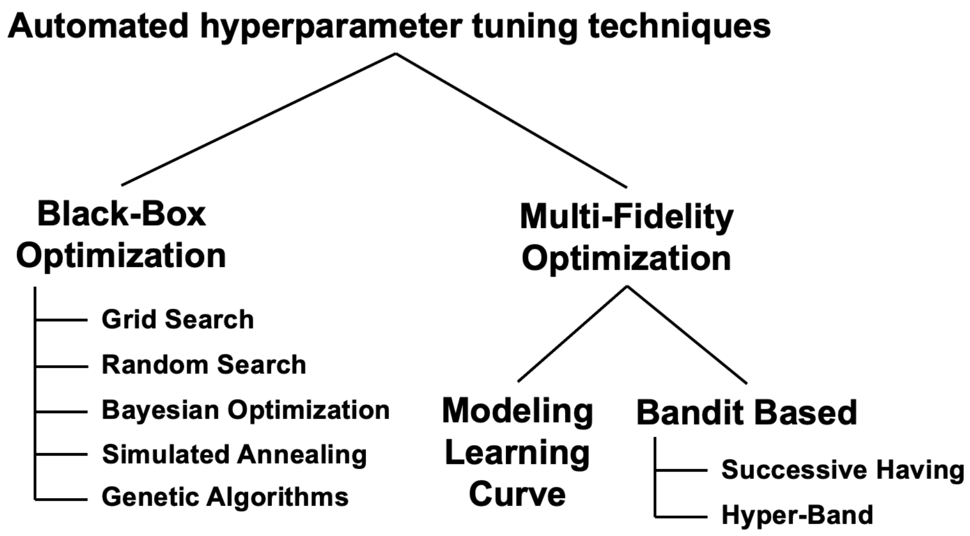
Within the type of *non-sampling-based* cost-sensitive learner, machine learning literature differentiates between *relabeling*, *threshold adjusting*, and *weighting*. *Relabeling* reassigns the class labels, e.g., churner or non-churner, of the individual observations by applying the *direct minimum expected cost criterion* that assigns an observation to the target class with the lowest misclassification costs (Viaene and Dedene 2005). As Coussement (2014) points out this can be done in the post-training phase (e.g., direct minimum expected cost classification, Duda, Hart, and Stork 2001) or in the pre-training phase (e.g., metacost algorithm by Domingos 1999). *Weighting* assigns different weights to the classes depending on misclassification costs with the aim that the algorithm favors the class with the higher weight (Ting 1998). *Threshold adjusting* searches for the cost-optimal threshold value on the training set. The threshold, which minimizes the misclassification costs of the training data, is then applied to the holdout sample (Sheng and Ling 2006).

4.4 Hyperparameter Optimization Techniques

Manual testing is a traditional way to tune hyperparameters (Bergstra and Bengio 2012; Yang and Shami 2020). As mentioned, the identification of the optimal hyperparameter combination is increasingly complex and manual tuning is very time-consuming. For this reason, the process is automated. The main aim of hyperparameter optimization is the automatization of the tuning process which enables machine learning users to solve prediction tasks effectively (Elshawi, Maher, and Sakr 2019; Yang and Shami 2020). Using an automatic selection method, the user of machine learning can skip the manual and iterative process of selecting a combination of hyperparameter values, which is labor intensive and requires a high skill set in machine learning (Luo 2016). Luo (2016) offers a valuable review of automatic selection methods for hyperparameter values.

In general, automated hyperparameter tuning techniques can be classified into *black-box optimization techniques* and *multi-fidelity optimization techniques* (Elshawi, Maher, and Sakr 2019). Figure 4 overviews the different sub-techniques in these two main categories.

Figure 4: Types of Automated Hyperparameter Tuning Techniques Based on Elshawi, Maher, and Sakr (2019)



Notes: Classification from Elshawi, Maher, and Sakr (2019).

A detailed explanation of each technique is beyond the scope of this study. A useful overview and explanations with reference to the relevant literature are provided in the study by Elshawi, Maher, and Sakr (2019) and Luo (2016). I focus on discussing the *grid search*, the *random search* method as well as *bayesian optimization* as three exemplary optimization approaches.

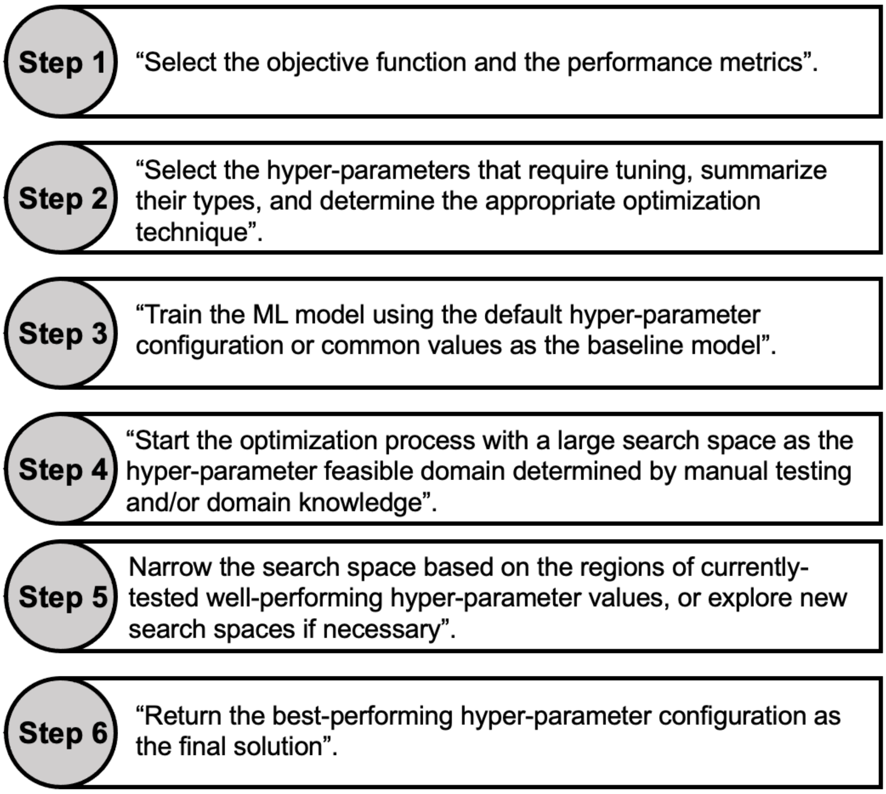
Grid search is a simple and one of the most widely used methods for hyperparameter optimization (Bergstra and Bengio 2012; Elshawi, Maher, and Sakr 2019). It involves the evaluation of all possible combinations of hyperparameters. Hence, this approach is computationally expensive as the number of potential combinations can become extremely high (e.g., Bergstra and Bengio 2012; Elshawi, Maher, and Sakr 2019).

Compared to grid search, *random search* is a more efficient approach to identify the right hyperparameter constellations (Bergstra and Bengio 2012). Random search samples hyperparameter combinations at random until a particular budget (e.g., number of trials, amount of computational time) is exhausted. Under this budget constraint, random search tends to find better solutions than grid search. The main reason why random search approaches are more efficient than grid search is that not all hyperparameters are equally important and grid search allocates many trials to dimensions that do not matter (see Bergstra and Bengio 2012 for details about hyperparameter importance).

Bayesian optimization is a sequential process that provides hyperparameters iteratively based on a loss function and previous hyperparameter performances which are then updated (see Shahriari et al. (2015) and Snoek, Larochelle, and Adams (2012) for a detailed explanation and a formal definition; Sarkar and De Bruyn 2021 apply the method in a marketing context). As Feurer, Springenberg, and Hutter (2015) state sequential model-based bayesian optimization is a successful hyperparameter optimization method in machine learning, which leads to better performances than grid and random search (e.g., Snoek, Larochelle, and Adams 2012).

Independent of the optimization technique, Yang and Shami (2020, p. 298) summarize the main steps of the hyperparameter optimization process as shown in Figure 5.

Figure 5: Hyperparameter Optimization Process Based on Yang and Shami (2020) and Luo (2016)



Notes: Process from Yang and Shami (2020, p. 298) and Luo (2016).
ML = Machine learning.

4.5 Interpretation of Machine Learning-Based Classifications

As shown in Table 5, two different types of interpretability can be distinguished (Burkart and Huber 2021; Lundberg, Erion, and Lee 2018). On the one hand, there is *global interpretability*, which tries to investigate the machine learning models as a whole, i.e., on a model level. This includes the *feature importance*, the *partial dependence plots*, and the *surrogate models*. *Local interpretability* aims to explain a single prediction of the model, i.e., on an observational level (Ribeiro, Singh, and Guestrin 2016b). On a local level, Shapley additive explanation (SHAP) and Local interpretable model-agnostic explanations (LIME) may be applied (Lundberg and Lee 2017; Ribeiro, Singh, and Guestrin 2016b).

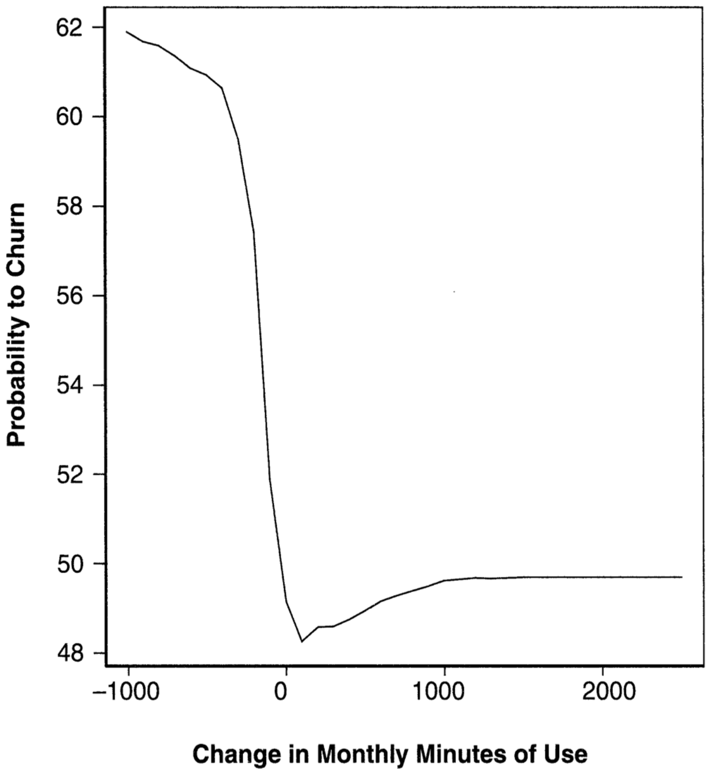
Table 5: Types of Interpretation within Machine Learning

Global interpretation	Local interpretation
• Feature importance	• Shapley additive explanation
• Partial dependence plot	• Local interpretable model-agnostic explanations
• Surrogate models	

4.5.1 Global interpretability. Breiman (2001) introduced the *feature importance* analysis in his paper on random forests. A feature is considered as important if it contributes to the model's performance. The application of feature importance analysis is already widely used in the marketing literature (Lemmens and Croux 2006; Naumzik, Feuerriegel, Weinmann 2022). However, the knowledge about the feature importance does not reveal anything about the direction of the effect of the individual features on the probabilities of the individual classes.

A *partial dependence plot* gives a graphical depiction of the marginal effect of a variable on the class probability (Friedman 2001). Partial dependence plots allow the analysis of non-linear relationships between the features and the target variable (Lemmens and Croux 2006). A partial dependence plot represents the impact of a predictor variable on the occurrence probability of one class (e.g., churn), conditional on all other predictors. Note that multicollinearity between features complicates interpretability and may bias results. Figure 6 provides an exemplary partial dependence plot from Lemmens and Croux (2006, p. 283). It shows how the churn probability changes when the consumption minutes of the customer increase or decrease.

Figure 6: Partial Dependence Plot from the Churn Prediction by Lemmens and Croux (2006)

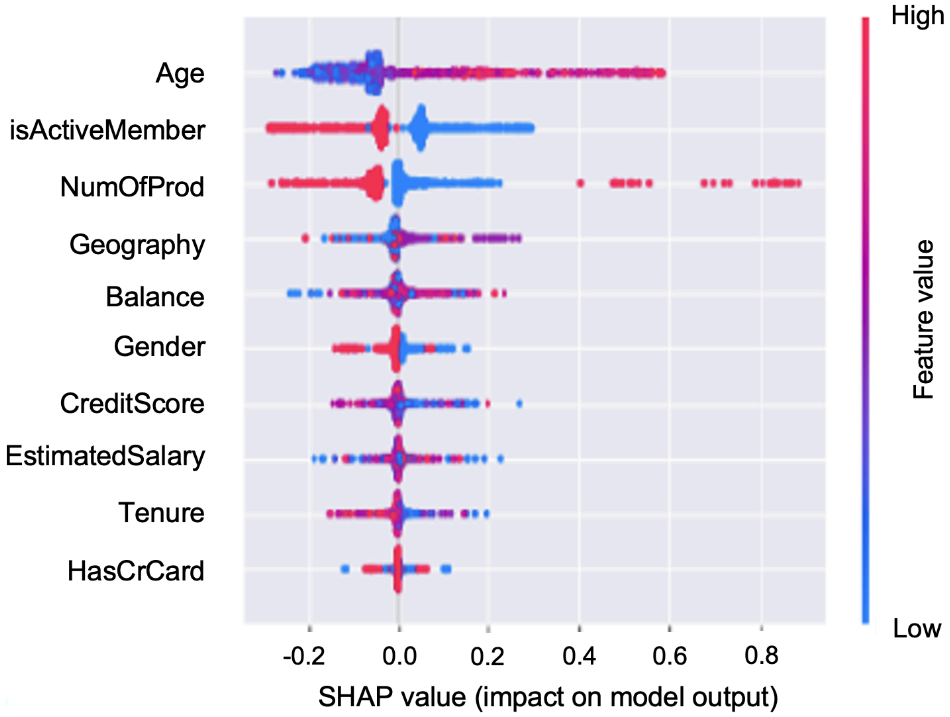


Notes: Figure from Lemmens and Croux (2006, p. 283).

An additional approach to produce insights about the effect strength and the effect direction of features is a *surrogate model* (Booker et al. 1999; Burkart and Huber 2021). A surrogate model is a model which is simpler to interpret (e.g., a linear regression) than a black-box model. The predicted class probability from the black-box algorithm represents the dependent variable of the surrogate model. The features used in the black-box machine learning algorithm are included as the independent variables. The beta coefficients of the surrogate model reveal insights about the effect size and the effect direction of the features on the predicted probability.

4.5.2 Local interpretability. Shapley additive explanation (SHAP) values "attribute to each feature the change in the expected model prediction when conditioning on that feature" (Lundberg and Lee 2017, p. 4772f.). SHAP values can replace the global feature importance in the form of *SHAP summary plots.* In contrast to the feature importance analysis on the global level, SHAP summary plots produce insights into the positive and negative relationships of the features with the target (Tékouabou et al. 2022). Figure 7 shows the exemplary SHAP summary plot from the churn prediction study by Tékouabou et al. (2022).

Figure 7: SHAP Summary Plot by Tékouabou et al. (2022)



Notes: Figure from Tékouabou et al. (2022, p. 12).

For example, Figure 7 reveals that an increase in age (red dots) is related to an increase in the churn probability. In addition, if the customer has a long-term relationship with the company (red dots for the "tenure" feature), this decreases the churn probability. Each colored dot of the individual features in Figure 7 represents one single observation of the training set.

In addition, *SHAP dependence plots* provide an alternative to partial dependence plots that better capture interaction effects (see Lundberg, Erion, and Lee 2018 for definitions, further explanations, and visualization examples).

Local interpretable model-agnostic explanations (LIME) aim "to identify an interpretable model over the interpretable representation that is locally faithful to the classifier" (Ribeiro, Singh, and Guestrin 2016a; Ribeiro, Singh, and Guestrin 2016b, p. 93). In their study, Ribeiro, Singh, and Guestrin (2016a) offer a formal definition of LIME and present graphical toy examples.

4.6 Presentation of Different Performance Metrics

Binary classification tasks (e.g., churn versus no churn, business failure versus no business failure) result in true positive predictions when a class, i.e., an event, is correctly predicted and true negative predictions when a class is correctly not predicted (see Table 6). In addition, two different errors may occur: a false positive result occurs when a class is falsely predicted and a false negative result occurs when a class is falsely not predicted (Branco, Torgo, and Ribeiro 2016; Cook and Ramadas 2020).

Table 6: Outcomes for a Churn Classification Task

		Actual class	
		Churn	No churn
Predicted class	Churn	True positives (TP)	False positives (FP)
	No churn	False negatives (FN)	True negatives (TN)

For the presentation of different performance metrics, I differentiate between metrics that are independent of a specific classification threshold and metrics that depend on a concrete threshold value. All of the presented performance metrics base on the confusion matrix in Table 6.

4.6.1 Threshold-dependent metrics. Table 7 (e.g., Branco, Torgo, and Ribeiro 2016) overviews the threshold-dependent performance metrics and their equations. Note that different equations exist for some metrics (e.g., for the F_1-score).

The *accuracy* is a commonly used performance criterion (Chawla et al. 2002; Mozer et al. 2000). It compares the a posteriori probability of the class occurrence with the true class occurrence. The resulting confusion matrix (Table 6) is used to calculate the accuracy of the models. A disadvantage of this standard measure is that it is not very robust concerning the chosen threshold value for the predicted probabilities (e.g., Baesens et al. 2002). In addition, the accuracy is poorly applicable to imbalanced target variables. For example, if a churn prediction contains only 1% churners, then an accuracy of 99% would be achieved by always classifying a non-churner. This is not in the interest of the machine learning applicants.

Table 7: Overview of Performance Metrics

Accuracy/ Percentage correctly classified	= (TP + TN) / (TP + TN + FP + FN)
Sensitivity/Recall (TPR)	= TP / (TP + FN)
Specificity (TNR)	= TN / (TN + FP)
Precision	= TP / (TP + FP)
R score	= sensitivity + specificity – 1
Balanced accuracy	= (sensitivity + specificity) / 2
F_1-score	= 2 x ((Precision x Recall) / (Precision + Recall))
False positive rate (FPR)	= FP / (FP + TN)
False negative rate (FNR)	= FN / (FN + TP)

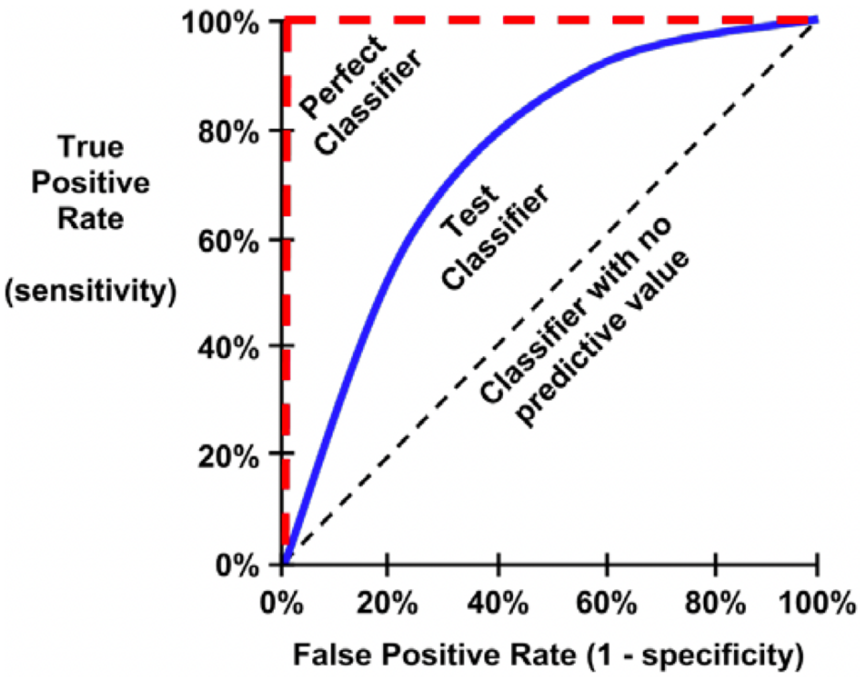
Notes: True Positives = TP; False Positives = FP; True Negatives = TN; False Negatives = FN; TPR = True Positive Rate; TNR = True Negative Rate; FPR = False Positive Rate; FNR = False Negative Rate.

Metrics that account for the specific distribution of the target variable are better suited to evaluate the performance on imbalanced data sets. For example, the *balanced accuracy* is useful for the analysis of imbalanced target variables (Bonas et al. 2021; Branco, Torgo, and Ribeiro 2016). Note that there exist approaches to incorporate the different misclassifications costs directly into the performance metrics. For example, the *F_β-score* enables a weighting of the different errors in accordance with the preferences of the user (Branco, Torgo, and Ribeiro 2016).

In addition to the metrics in Table 7, the *top decile* is one of the most managerially relevant success metrics (Sarkar and De Bruyn 2021). This evaluation measure focuses on the 10% events (e.g., churns) with the highest probability of occurrence (Lemmens and Croux 2006; Sarkar and De Bruyn 2021). Applied to the churn prediction example, the "proportion of real events in the top 10% most likely to churn is compared with the proportion of real events in the total dataset. This increase in density is called the top-decile lift" (Coussement and Van den Poel 2008, p. 317).

4.6.2 Threshold-independent metrics. Two relevant performance metrics evaluate the model performance independent of a chosen threshold. The *receiver operating characteristics* (ROC) curve summarizes the prediction performance in terms of true positive rate and false positive rate across each potential classification threshold value between 0% and 100% (Buckinx and Van den Poel 2005; Lantz 2019, p. 312). It is an evaluation metric for (mainly) binary classification problems.

Figure 8: ROC Curve from Lantz (2019)



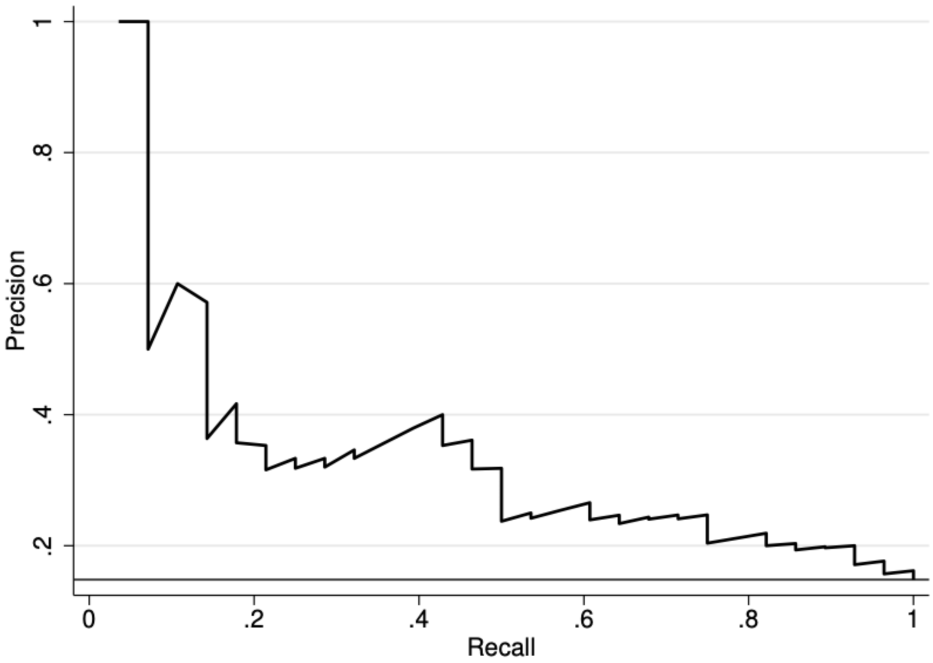
Notes: Figure from Lantz (2019, p. 312).

The area under the curve (AUC) of the ROC measures the entire two-dimensional area underneath the ROC curve. The metric is considered classification-threshold-invariant because it measures the quality of the model's predictions independent of what classification threshold is chosen. The maximum possible AUC value is 1 (red line in Figure 8) and the lowest value is 0. The higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. An AUC value of 0.5 indicates the model has no class separation power at all (black line in Figure 8). In case the AUC value of the model is lower than 0.5, the model performance is worse than randomly assigning the classes. The blue line in Figure 8 represents an exemplary classifier that performs better than a random classification but still results in false predictions as it does not achieve an AUC value of 1.

Although the ROC AUC value is a very popular and established measure to evaluate the performance of binary classifiers, the metric can be misleading in highly imbalanced classification scenarios (Saito and Rehmsmeier 2015). Precision-Recall (PR) curves provide a more accurate impression of the classification performance since they also evaluate the fraction of true positives among all positive predictions. Thus, the application of PR curves has been suggested for comparing models based on imbalanced data since PR AUC focuses mainly on the positive class and cares less about the frequently occurring negative class (Sofaer, Hoeting, and Jarnevich 2019).

The PR curve and the corresponding AUC value are comparable to the ROC curve. However, this graph plots the precision against the recall as illustrated in Figure 9 (Boyd et al. 2012; Figure 9 from Cook and Ramadas 2020, p. 145). Unlike the ROC curve, the AUC value for a random classification is not automatically 0.5 but is based on the skewness of the binary target variable. For example, if the target variable contains 14,81% of observations representing the minority class (value from Figure 9), then the AUC equals 0.1481 for a random classification (Boyd et al. 2012; Cook and Ramadas 2020). In this scenario, an algorithm achieving a value higher than 0.1481, classifies better than a random classifier. Figure 9 shows a PR curve with an AUC value of 30,89% (Cook and Ramadas 2020, p. 144), which outperforms a random classifier.

Figure 9: PR Curve from Cook and Ramadas (2020)



Notes: Figure from Cook and Ramadas (2020, p. 145).

5 Conclusion

5.1 Summary

In this study, I have identified six recurring machine learning-based challenges that researchers in the field of marketing frequently face. The challenges are rooted in the specific nature of the data or the selection of variables used for the classification, but also in the algorithms applied and their configuration as well as their performance evaluation and the derivation of recommendations for action.

Based on the general learning process, *imbalanced target variables* and *feature selection* are considered among the data-based challenges. Abstraction-based challenges include *hyperparameter tuning* and *cost-sensitive learning*. *Interpretable machine learning* and the *selection of performance metrics* belong to the generalization component of the learning process. After identifying the problems by analyzing representative marketing studies dealing with machine learning-based classification tasks, the study presents and structures potential solutions to each problem by reviewing literature from computer science.

5.2 Contributions

Linking the marketing discipline with computer science provides valuable contributions and guidance for machine learning users in academia and in practice. For example, the study provides methodological guidance for researchers and facilitates scientific work. Applicants of machine learning methods within marketing research and beyond certainly face several challenges identified in my study. While this study does not describe and explain in detail what each solution strategy looks like, it does structure and review different solution paths and point to relevant computer science literature that then provides further explanations. In this way, researchers are relieved of the time-consuming process of dealing with numerous ways of solving machine learning-based problems, as my study guides to possible solutions drawn from complex computer science literature.

Furthermore, for researchers, performance evaluation of their developed (machine learning-based) models is one of the main aspects of their studies. However, my review emphasizes the variety of existing performance metrics with different advantages and disadvantages. Thereby, an improvement of one metric may lead to a decline of the other metrics (see trade-off between false positive and false negative classifications in Table 6 and Table 7 again). Researchers need to take the relevance of the choice of the performance metric into account. Different performance metrics provide information on different elements of the classifications and should be applied in certain situations. In the marketing literature, for example, the Precision-Recall curve is rarely applied (see Table 1), although it is better suited for highly imbalanced data sets (Saito and Rehmsmeier 2015; Sofaer, Hoeting, and Jarnevich 2019). To obtain a holistic evaluation of the classification quality, researchers need to justify their selection of metrics or report multiple metrics.

From a practical point of view, companies can benefit from the approaches presented in this study. My findings enable companies to apply machine learning more efficiently and effectively. In this way, the analytical foundation of marketing mix decisions can be improved, which creates additional monetary value for companies. Problem areas like cost-sensitive learning or the selection of the right performance metric exemplify this contribution: To be able to make cost-optimal recommendations for action, companies must not maximize an arbitrarily selected performance metric, but must also take the costs in the machine learning process into account. In this way, actual cost savings may be generated which lead to higher profits. The overview of different methods that my study provides enables companies to quickly select the solutions that are most suitable for them.

5.3 Limitations and Future Research Avenues

My study has several limitations that provide opportunities for fruitful future research. First, the identified challenges are prominent because of their high frequency in marketing studies. However, machine learning is a broad methodological field that constantly develops. Future research could identify further machine learning-based problems and provide solutions for them. For example, a large number of different machine learning approaches exist with different strengths and weaknesses. Guiding machine learning users in their choice of the right approach would be of high value. In addition, this study focuses on classifications and the analysis of structured data. In other use cases (e.g., regressions or unsupervised machine learning), users face different challenges which need to be reviewed.

Second, this study provides an overview of potential solutions for selected problems. However, these individual problem areas are complex and the corresponding literature is extensive. Future research can focus on individual problem areas to be able to address and present the computer science literature in even more detail.

Regardless of the limitations of this review, the study also generates ideas for future research. For example, while it is already well established in the computer science discipline to control for the costs of misclassifications, the precise quantification of these costs is often complex and difficult (Elkan 2001). Moreover, costs of different misclassifications are subjectively weighted by users (see prospect theory by Kahneman and Tversky (1979) which points out that losses are weighted differently than gains). Future research could, therefore, address how to measure the perceived costs of misclassifications (e.g., with a conjoint analysis) to enable cost-optimal machine learning applications.

REFERENCES PAPER 3

Asim Khawaja M., Adnan Idris, Talat Iqbal, Francisco Martínez-Álvarez (2018), "Earthquake Prediction Model Using Support Vector Regressor and Hybrid Neural Networks," *PLoS ONE*, 13 (7), https://doi.org/10.1371/journal.pone.0199004.

Baesens, Bart, Stijn Viaene, Dirk van den Poel, Jan Vanthienen, and Guido Dedene (2002), "Bayesian Neural Network Learning for Repeat Purchase Modelling in Direct Marketing," *European Journal of Operational Research*, 138 (1), 191–211.

Bahnsen, Alejandro Correa, Djamila Aouada, and Björn Ottersten (2015), "Example-Dependent Cost-Sensitive Decision Trees," *Expert Systems with Applications*, 42 (19), 6609–19.

Berger, Paul and Thomas Magliozzi (1992), "The Effect of Sample Size and Proportion of Buyers in the Sample on the Performance of List Segmentation Equations Generated by Regression Analysis," *Academic Journal Journal of Direct Marketing*, 6 (1), 13–22.

Bergstra, James and Yoshua Bengio (2012) "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, 13 (2), 281–305.

Blum, Avrim L. and Pat Langley (1997), "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, 97, 245–71.

Bonas, Matthew, Son Nguyen, Alan Olinsky, John Quinn, and Phyllis Schumacher (2021), "A Method to Determine the Size of the Resampled Data in Imbalanced Classification," in *Contemporary Perspectives in Data Mining*, Vol. 4, Kenneth D. Lawrence and Ronald Klimberg, eds. Information Age Publishing Inc., 119–42.

Booker, Andrew J., John E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset (1999), "A Rigorous Framework for Optimization of Expensive Functions by Surrogates," *Structural Optimization*, 17(1), 1–13.

Boyd, Kendrick, Vítor Santos Costa, Jesse Davis, and C. David Page (2012), "Unachievable Region in Precision-Recall Space and Its Effect on Empirical Evaluation," in *Proceedings of the Twenty-Ninth International Conference on Machine Learning*, John Langford and Joelle Pineau, eds. Omnipress, 1619–26.

Branco, Paula, Luís Torgo, and Rita P. Ribeiro (2016), "A Survey of Predictive Modeling on Imbalanced Domains," *ACM Computing Surveys (CSUR)*, 49 (2), 1–50.

Breiman, Leo (2001), "Random Forests," *Machine Learning*, 45 (1), 5–32.

Buckinx, Wouter and Dirk Van den Poel (2005), "Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting," *European Journal of Operational Research*, 164 (1), 252–68.

Burkart, Nadia and Marco F. Huber (2021), "A Survey on the Explainability of Supervised Machine Learning," *Journal of Artificial Intelligence Research*, 70, 245–317.

Cai, Jie, Jiawei Luo, Shulin Wang, and Yang Sheng, (2018), "Feature Selection in Machine Learning: A New Perspective," *Neurocomputing*, 300, 70–79.

Campbell, Dennis W. and Ruidi Shang (2022), "Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews," *Management Science*, 68 (9), 7034–53.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer (2002), "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, 16, 321–35.

Cook, Jonathan and Vikram Ramadas (2020), "When to Consult Precision-Recall Curves," *The Stata Journal*, 20 (1), 131–48.

Coussement, Kristof (2014), "Improving Customer Retention Management Through Cost-Sensitive Learning," *European Journal of Marketing,* 48 (3/4), 477–95.

Coussement, Kristof and Dirk van den Poel (2008), "Churn Prediction in Subscription Services: An Application of Support Vector Machines While Comparing Two Parameter-Selection Techniques," *Expert Systems with Applications*, 34 (1), 313–27.

Cui, Geng, Man Leung Wong, and Xiang Wan (2012), "Cost-Sensitive Learning via Priority Sampling to Improve the Return on Marketing and CRM Investment," *Journal of Management Information Systems*, 29 (1), 341–74.

Devi, S. Sarojini and Y. Radhika (2018), "A Survey on Machine Learning and Statistical Techniques in Bankruptcy Prediction," *International Journal of Machine Learning and Computing*, 8 (2), 133–39.

Domingos, Pedro (1999), "Metacost: A General Method for Making Classifiers Cost-Sensitive," *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '99. ACM, 155–64.

Dong, Guozhu and Huan Liu (2018), *Feature Engineering for Machine Learning and Data Analytics*, CRC Press.

Donkers, Bas, Philip Hans Franses, and Peter C. Verhoef (2003), "Selective Sampling for Binary Choice Models," *Journal of Marketing Research*, 40 (4), 492–97.

Duda, Richard O., Peter E. Hart, and David G. Stork (2001), *Pattern Classification*, 2nd ed., John Wiley & Sons.

Elkan, Charles (2001), "The Foundations of Cost-Sensitive Learning," *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 17 (1), 973–78.

Elshawi, Radwa, Mohamed Maher, and Sherif Sakr (2019), "Automated Machine Learning: State-of-the-Art and Open Challenges," *arXiv preprint arXiv:1906.02287*, https://doi.org/10.48550/arXiv.1906.02287.

Feurer, Matthias, Jost Springenberg, and Frank Hutter (2015), "Initializing Bayesian Hyperparameter Optimization via Meta-Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, 29 (1), 1128–35.

Friedman, Jerome H. (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 29 (5), 1189–232.

Gentile, Douglas A., Christopher L. Groves, and J. Ronald Gentile (2014), "The General Learning Model: Unveiling the Teaching Potential of Video Games," in *Learning by Playing: Video Gaming in Education*, Fran C. Blumberg, ed. Univesity Press, 121–42.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016), *Deep Learning*, MIT Press.

Guyon, Isabelle and André Elisseeff (2003), "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, 3, 1157–82.

Htet, Khaing Suu and Myint Myint Sein (2020), "Effective Marketing Analysis on Gender and Age Classification with Hyperparameter Tuning," *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech)*, 247–48.

Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley (2012), "Customer Churn Prediction in Telecommunications," *Expert Systems with Applications*, 39 (1), 1414–25.

Idris, Adnan, Muhammad Rizwan, and Asifullah Khan (2012), "Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies," *Computers & Electrical Engineering*, 38 (6), 1808–19.

Jordan, Michael I. and Tom M. Mitchell (2015), "Machine Learning: Trends, Perspectives, and Prospects," *Science*, 349.6245, 255–60.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," Econometrica, 47 (2), 263–91.

Kirasich, Kaitlin, Trace Smith, and Bivin Sadler (2018), "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets," *SMU Data Science Review*, 1 (3), article 9.

Kohavi, Ron and George H. John (1997), "Wrappers for Feature Subset Selection," *Artificial Intelligence*, (97), 273–324.

Kononenko, Igor (2001), "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective," *Artificial Intelligence in Medicine*, 23 (1), 89–109.

Kuhn, Max and Kjell Johnson (2013), *Applied Predictive Modeling*, Vol. 26. Springer.

Kumar, Yogesh, Apeksha Koul, Pushpendra Singh Sisodia, Jana Shafi, Kavita Verma, Mehdi Gheisari, and Mohamad Bagher Davoodi (2021), "Heart Failure Detection Using Quantum-enhanced Machine Learning and Traditional Machine Learning Techniques for Internet of Artificially Intelligent Medical Things," *Wireless Communications and Mobile Computing*, 1–16.

Kumar, Akhil, Vithala R. Rao, and Harsh Soni (1995), "An Empirical Comparison of Neural Network and Logistic Regression Models," *Marketing Letters,* 6, 251–63.

Lalwani, Praveen, Manas Kumar Mishra, Jasroop Singh Chadha, and Pratyush Sethi (2022), "Customer Churn Prediction System: A Machine Learning Approach." *Computing*, 104, 1–24.

Langley, Pat (1994), "Selection of Relevant Features in Machine Learning," in *Proceedings of the AAAI Fall Symposium on Relevance*, Vol. 184, 245–71.

Lantz, Brett (2019), *Machine Learning with R*, 3rd ed. Packt Publishing.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (2), 276–86.

Lemmens, Aurélie and Sunil Gupta (2020), "Managing Churn to Maximize Profits," *Marketing Science*, 2020, 39 (5), 956–73.

Ling, Charles X. and Victor S. Sheng (2011), "Cost-Sensitive Learning and the Class Imbalance Problem," in *Encyclopedia of Machine Learning*, Claude Sammut and Geoffrey I. Webb, eds. Springer, 231–35.

Lundberg, Scott M. and Su-In Lee (2017), "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 30. Curran Associates, Inc., 4768–77.

Lundberg, Scott M., Gabriel G. Erion, and Su-In Lee (2018), "Consistent Individualized Feature Attribution for Tree Ensembles," *arXiv preprint arXiv:1802.03888*.

Luo, Gang (2016), "A Review of Automatic Selection Methods for Machine Learning Algorithms and Hyper-Parameter Values," *Network Modeling Analysis in Health Informatics and Bioinformatics*, 5, 1–16.

Ma, Liye and Baohong Sun (2020), "Machine Learning and AI in Marketing - Connecting Computing Power to Human Insights," *International Journal of Research in Marketing*, 37 (3), 481–504.

Marcinkevičs, Ričards and Julia E. Vogt (2023), "Interpretable and Explainable Machine Learning: A Methods-Centric Overview with Concrete Examples," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1493, 1–32.

Michalak, Krzysztof and Halina Kwasnicka (2006), "Correlation-Based Feature Selection Strategy in Neural Classification," *Sixth International Conference on Intelligent Systems Design and Applications*, 1, IEEE.

Mignan, Arnaud and Marco Broccardo (2020), "Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on Their Limitations," *Seismological Research Letters*, 91 (4), 2330–42.

Mitchell, Tom M. (1997), *Machine Learning*, McGraw Hill.

Morrison, Donald G. (1969), "On the Interpretation of Discriminant Analysis," *Journal of Marketing Research*, 6 (2), 156–63.

Mozer, Michael C., Richard Wolniewicz, David B. Grimes, Eric Johnson, and Howard Kaushansky (2000), "Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry," *IEEE Transactions on Neural Networks*, 11 (3), 690–96.

Naumzik, Christof, Stefan Feuerriegel, and Markus Weinmann (2022), "I Will Survive: Predicting Business Failures from Customer Ratings," *Marketing Science*, 41 (1), 188–207.

Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 43 (2), 204–11.

Ngai, Eric W. and Yuanyuan Wu (2022) "Machine Learning in Marketing: A Literature Review, Conceptual Framework, and Research Agenda," *Journal of Business Research*, 145, 35–48.

Quinlan, J. Ross (1986), "Induction of Decision Trees," *Machine Learning*, 1, 81–106.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016a), "Why Should I Trust You?: Explaining the Predictions of Any Classier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–44.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016b), "Model-Agnostic Interpretability of Machine Learning," *arXiv preprint arXiv:1606.05386*, 91–95, https://doi.org/10.48550/arXiv.1606.05386.

Rudin, Cynthia (2019), "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, 1 (5), 206–15.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986), "Learning Representations by Back-Propagating Errors," *Nature*, 323.6088, 533–36.

Rust, Roland T. (2020), "The Future of Marketing," *International Journal of Research in Marketing*, 37 (1), 15–26.

Saito, Takaya and Marc Rehmsmeier (2015), "The Precision-Recall Plot is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PloS ONE*, 10 (3), https://doi.org/10.1371/journal.pone.0118432.

Sana, Joydeb Kumar, Mohammad Zoynul Abedin, M. Sohel Rahman, and M. Saifur Rahman (2022), "A Novel Customer Churn Prediction Model for the Telecommunication Industry Using Data Transformation Methods and Feature Selection," *PloS ONE*, 17 (12), https://doi.org/10.1371/journal.pone.0278095.

Sarkar, Mainak and Arnaud De Bruyn (2021), "LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning," *Journal of Interactive Marketing*, 53 (1), 80–95.

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas (2015), "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE*, 104 (1), 148–75.

Sheng, Victor S. and Charles X. Ling (2006), "Thresholding for Making Classifiers Cost Sensitive," in *Proceedings of the 21st National Conference on Artificial Intelligence*, 476–81.

Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams (2012), "Practical Bayesian Optimization of Machine Learning Algorithms," in *Advances in Neural Information Processing Systems 25*, 2960–68.

Sofaer, Helen R., Jennifer A. Hoeting, and Catherine S. Jarnevich (2019), "The Area Under the Precision-Recall Curve as a Performance Metric for Rare Binary Events," *Methods in Ecology and Evolution*, 10 (4), 565–77.

Tékouabou, Stéphane C. K., Ștefan Cristian Gherghina, Hamza Toulni, Pedro Neves Mata, and José Moleiro Martins (2022), "Towards Explainable Machine Learning for Bank Churn Prediction Using Data Balancing and Ensemble-Based Methods," *Mathematics*, 10 (14), 2379.

Ting, Kai Ming (1998), "Inducing Cost-Sensitive Trees via Instance Weighting," in *Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery,* 23–26.

Turney, Peter D. (1995), "Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm," *Journal of Artificial Intelligence Research,* 2, 369–409.

Viaene, Stijn and Guido Dedene (2005), "Cost-Sensitive Learning and Decision Making Revisited," *European Journal of Operational Research*, 166 (1), 212–20.

Wang, Nanxi (2017), "Bankruptcy Prediction Using Machine Learning," *Journal of Mathematical Finance*, 7 (04), 908–18.

Wang, Tong, Cheng He, Fujie Jin, and Yu Jeffrey Hu (2022), "Evaluating the Effectiveness of Marketing Campaigns for Malls Using a Novel Interpretable Machine Learning Model," *Information Systems Research*, 33 (2), 659–77.

Wang, Xiaofeng, Matthew S. Gerber, and Donald E. Brown (2012), "Automatic Crime Prediction Using Events Extracted from Twitter Posts," *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 231–38.

Yang, Li and Abdallah Shami (2020), "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," *Neurocomputing,* 415, 295–316.

Zadrozny, Bianca, John Langford, and Naoki Abe (2003), "Cost-Sensitive Learning by Cost-Proportionate Example Weighting," in *Third IEEE International Conference on Data Mining*, 435–42.

Zhao, Zheng, Fred Morstatter, Shashvata Sharma, Salem Alelyani, Aneeth Anand, Huan Liu (2010), "Advancing Feature Selection Research," *ASU Feature Selection Repository*, 1–28.

APPENDIX PAPER 3

APPENDIX A: MACHINE LEARNING TERMINOLOGY

The terminology in marketing and computer science may differ. Since my study links research from both fields, it is necessary to clarify the machine learning terminology. Table A1 "translates" the expression commonly used in machine learning into the "language" spoken by marketing researchers, when necessary, and defines the most important machine learning expression to enable a better understanding of the study.

Table A1: Machine Learning Terminology

Expression (predominantly) used in computer science	**Expression (predominantly) used marketing research**	**Explanation**	**Exemplary source**
Target; label; output	Dependent variable	Output that is to be predicted or classified.	Lantz (2019)
Predictor; feature; input	Independent variable	Input data to generate the output (i.e., the target).	Dong and Liu (2018)
Training		In the training process, the model produces a mathematical representation of the relationship between the input and the output (in the case of supervised machine learning).	Lantz (2019)
Training data	Calibration data	Subset of data from which the machine learning algorithm learns patterns and relationships between the input and the output.	Lantz (2019); Sarkar and De Bruyn (2021)
Validation data[1]		Subset of data used to evaluate a model fit on the training dataset while tuning model hyperparameters.	Lantz (2019)
Testing data[1]	Holdout data	Subset of data that is completely new to the model and is used to evaluate the machine learning model's performance after it has been trained and validated.	Lantz (2019)
Baseline model		A simple model that acts as a reference for more sophisticated and complex machine learning models.	Sarkar and De Bruyn (2021)
Class		Values of discrete output variables (e.g., no churn vs. churn).	Lemmens and Croux (2006)
(Classification) Threshold		A cut-off value between 0% and 100% that transforms the predicted probability of each observation into one of the classes (e.g., no churn vs. churn).	Branco, Torgo, and Ribeiro (2016)

Notes: [1] The literature on machine learning often reverses the meaning of "validation" and "test" sets.
The listed expressions represent a selection based on their relevance for this study. In addition, many other expressions are typically used in machine learning. See also https://developers.google.com/machine-learning/glossary for a collection of additional definitions that are less relevant for this study.

Table A1: Terminology of Machine Learning

Expression (predominantly) used in computer science	Expression (predominantly) used marketing research	Explanation	Exemplary source
(Class-) Imbalanced data set		A data set with skewed class proportions.	Branco, Torgo, and Ribeiro (2016)
Downsampling; undersampling		Randomly removing observations from the majority class to make the data set less skewed.	Branco, Torgo, and Ribeiro (2016)
Feature importance	Variable importance	Measures how much each feature contributes to the model prediction. It determines the degree of usefulness of a feature for a model.	Lemmens and Croux 2006
Hyperparameter		Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that the algorithm learns.	Kuhn and Johnson (2013); Luo (2016); Sarkar and De Bruyn (2021)
Interpretability; explainability		Degree to which machine learning users understand how the algorithm makes predictions.	Burkart and Huber (2021)
Majority class		Classes that make up the larger proportion of the data set.	Branco, Torgo, and Ribeiro (2016)
Minority class		Classes that make up the smaller proportion of the data set.	Branco, Torgo, and Ribeiro (2016)
Overfitting		Model learns the details and noise in the training data to the extent that it negatively impacts the performance of the model on unknown data (i.e., holdout data). Model is not generalizable.	Goodfellow, Bengio, and Courville (2016); Kirasich, Smith, and Sadler (2018); Sarkar and De Bruyn (2021)
Oversampling; upsampling		Randomly increasing the number of observations from the minority class to make the data set less skewed.	Branco, Torgo, and Ribeiro (2016)

Notes: The listed expressions represent a selection based on their relevance for this study. In addition, many other expressions are typically used in machine learning. See also https://developers.google.com/machine-learning/glossary for a collection of additional definitions that are less relevant for this study.

Table A1: Terminology of Machine Learning

Expression (predominantly) used in computer science	Expression (predominantly) used marketing research	Expression within machine learning	Exemplary source
Parameter		Estimated from the data during training as the algorithm maps the relationship between the features and the labels.	Alibrahim and Ludwig (2021)
Performance metrics		Metrics that evaluate how well the model performs for the given data set.	Branco, Torgo, and Ribeiro (2016)
Supervised machine learning		Algorithm is trained with input data that has a particular output that needs to be predicted.	Lantz (2019)
Unsupervised machine learning (not covered in this study)		Algorithm clusters unlabeled observations without a predefined output.	Lantz (2019)

Notes: The listed expressions represent a selection based on their relevance for this study. In addition, many other expressions are typically used in machine learning. See also https://developers.google.com/machine-learning/glossary for a collection of additional definitions that are less relevant for this study.

REFERENCES APPENDIX PAPER 3

Alibrahim, Hussain and Simone A. Ludwig (2021), "Hyperparameter Optimization: Comparing Genetic Algorithm Against Grid Search and Bayesian Optimization," *2021 IEEE Congress on Evolutionary Computation.*

Branco, Paula, Luís Torgo, and Rita P. Ribeiro (2016), "A Survey of Predictive Modeling on Imbalanced Domains," *ACM Computing Surveys (CSUR)*, 49 (2), 1–50.

Burkart, Nadia and Marco F. Huber (2021), "A Survey on the Explainability of Supervised Machine Learning," *Journal of Artificial Intelligence Research*, 70, 245–317.

Dong, Guozhu and Huan Liu (2018), *Feature Engineering for Machine Learning and Data Analytics*, CRC Press.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016), *Deep Learning*, MIT Press.

Kirasich, Kaitlin, Trace Smith, and Bivin Sadler (2018), "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets," *SMU Data Science Review*, 1 (3), article 9.

Kuhn, Max and Kjell Johnson (2013), *Applied Predictive Modeling*, Vol. 26. Springer.

Lantz, Brett (2019), *Machine Learning with R*, 3rd ed. Packt Publishing.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (2), 276–86.

Luo, Gang (2016), "A Review of Automatic Selection Methods for Machine Learning Algorithms and Hyper-Parameter Values," *Network Modeling Analysis in Health Informatics and Bioinformatics*, 5, 1–16.

Sarkar, Mainak and Arnaud De Bruyn (2021), "LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning," *Journal of Interactive Marketing*, 53 (1), 80–95.

EIDESSTAATLICHE ERKLÄRUNG

nach § 8 Abs. 3 der Promotionsordnung vom 17.02.2015

"Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes unterstreichen) geholfen:

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Köln, den 20. Juni 2023 Unterschrift