**The Digital Transformation of the News Media Business –
Paid Content and Entrepreneurship in Digital Journalism**

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

2022

presented
by

M.Sc. Daniel O'Brien

from

Hamm (Westf.)

First reviewer: Dr. Christian-Mathias Wellbrock

Second reviewer: Prof. Dr. Reinhard Kunz

Date of oral defense: 17th of November 2022

Abstract

The digital transformation of the news business continues to agitate publishers. Concerned about declining sales in the print segment, legacy outlets, local news companies and freelance journalists alike search for ways to monetize digital journalism properly. At first glance, digital journalism and its monetisation as paid content seem a promising effort. The digitisation of the news business enabled distribution at a marginal cost of almost zero while giving journalists access to new research technologies and lowering the cost of entry for smaller companies.

However, while digital journalism enjoys broad popularity and use, online news are gaining few paying customers. Furthermore, online news compete within a larger digital media complex, comprising movies, games, and social media. After 25 years of experimentation, the digital future of journalism is still heavily debated in media management.

Concerning the reconstitution as a digital medium, this research examines conditions of success and obstacles for the digital news media business to be successful as a business venture. Therefore, the research question reads *What factors enable the viability and entrepreneurial success of the news media business in light of the consequences of digital transformation?* The overarching research question is considered from two angles: The first angle concerns the demand side by looking at the antecedents of the audience's willingness to pay for paid content. The second angle focuses on the supply side and therefore examines antecedents of success in the context of digital journalistic start-ups and founders.

In four studies, this thesis develops an analysis of the online news business with a local focus on the German news market. For this purpose, a variety of methods ranging from qualitative work and literature review to empirical research employing path analysis and predictive analytics are applied. Theoretically, digital transformation, free mentality and other peculiarities of information goods inform the frame of this work. Thus, this research aims at contributing to a financially sustainable news media business.

Preface and Acknowledgements

The research of this thesis has been developed during my three years in the department of media and technology management at the University of Cologne. I would like to take this opportunity for some acknowledgements. First, I am grateful for the financial support from the Volkswagen Foundation, which allowed me to devote myself to this research under the project title “Success Conditions and Factors for Digital Entrepreneurial Journalism” with grant number 94903. In addition, I would like to thank many individuals for their unwavering support. I would like to express my special thanks to my first supervisor, Dr. Christian-Mathias Wellbrock, who trusted me with the skills to take on this research project and repeatedly gave me the opportunity to develop myself scientifically. I am also very grateful for the guidance, advice and substantive rigour of Prof. Reinhard Kunz, who was always very keen to get the best out of this research. In addition, I would like to express my thanks to the staff and colleagues at the University of Cologne, especially Lea Püchel, with whom I was able to exchange ideas at every stage of the dissertation. A special thank also goes to Nicola Kleer, who helped me through part of this project. Furthermore, a general thanks also goes, of course, to the numerous students who have supported and challenged me through discussions in always exciting courses. Furthermore, I would also like to thank Prof. Dr. Christopher Buschow for the constructive exchange from which my work has benefited. In my private life, too, I must thank countless people who put up with my talk about the dissertation and stood by me with advice and support. Dr. Johannes Breuer, Dr. Linda Wulf, Fabian Wedekind and Nadine Kwelik: I cannot thank you enough. I would also like to thank my family, especially my mother, who against all odds made this step possible for me. But my special and heartfelt thanks goes to my husband, Glen O’Brien, without whose endless patience, discipline and loving compassion I would not have managed a month of this effort. Your love is the source of all the hard work and success reflected in this thesis.

Table of Contents

List of Figures

List of Tables

A. Introduction

„The world has changed less since the time of Jesus Christ than it has in the last thirty years.“

(Charles Péguy, quoted in: Hughes & Walsh, 1991)

“There’s a huge layer of the economy unseen in the official data and, for that matter, unaccounted for on the income statements and balance sheets of most companies. Free digital goods, the sharing economy, intangibles and changes in our relationships have already had big effects on our well-being. They also call for new organizational structures, new skills, new institutions, and perhaps even a reassessment of some of our values.”

(Brynjolfsson & McAfee, 2014, pp. 108–109)

1. Motivation and Purpose

Digitisation is omnipresent and continues to permeate everyday life. Broadly speaking, digitization has changed the way society operates. The continuous transformation of formerly analogue areas of life into the digital sphere has everlasting effects on every aspect of human interaction. Whether science or art, communication or production, private life or work – digitization has shaped society like hardly any other technology before (Berman, 2012; Loebbecke & Picot, 2015; Teece, 2018). While this proposition is now a commonplace of public discourse, its consequences are still emerging and being discussed – inconclusively in society and academia.

This applies to transformations of entire industries. In this respect, the term “digital transformation” describes the impact of digital technologies on the economy, companies and business models and related changes in production, organization and strategy (Hess et al., 2016). Specifically, Vial (2019) describes these changes as “a process where digital technologies create disruptions triggering strategic responses from organizations that seek to alter their value creation paths while managing the structural changes and organizational barriers that affect the positive and negative outcomes of this process” (p. 118).

Numerous companies and industries are busy coping with the opportunities and risks of digital transformation. While some are benefiting greatly, individual industries are still struggling to adapt their once analogue business model to the new, thoroughly digital world. This development is particularly striking in the media industry. Winner-takes-all dynamics, standardisation of demand and the proliferation of the commons make it difficult for numerous providers of information goods to monetise their products (Kutzner et al., 2018; Loebbecke & Picot, 2015).

One of the industries still struggling in this context turns out to be journalism. Journalism has been ascribed crucial functions for a working democracy: As informing, controlling and criticizing institution, the "fourth estate" enables the general public to make educated, informed decisions concerning public matters (McQuail, 1992; Wellbrock & Klein, 2014). However, this self-image of journalism as an enterprise with a definite public value is threatened by its business difficulties (Sraml Gonzalez & Gulbrandsen, 2022). The proverbial death of newspapers continues unabated and circulations are declining further (Chyi & Ng, 2020; Evens & Van Damme, 2016). Once considered a license to print money (Fletcher & Nielsen, 2020), the formerly highly lucrative news business is in deep crisis.

At the same time, digital news products seem to elicit only a comparatively low willingness to pay and are insufficient to cope with the losses of print media (Berger et al., 2015). Thus, in recent decades, while digitization has been radically transformative for the entire economy, it has been challenging for journalism in particular (Chyi & Ng, 2020; Goyanes et al., 2021; Himma-Kadakas & Kõuts, 2015).

On the one hand, information technologies enable unprecedented possibilities for the production and distribution of journalism online, while consumers have ubiquitous and cheap or free access to an unprecedented abundance of information goods (Kammer et al., 2015; Shapiro & Varian, 1999). Thus, the internet has globally become one of the most important news sources (Hölig & Hasebrink, 2020; Newman et al., 2018). Accordingly, in analogy to print journalism, digital journalism is often described as highly relevant for the functioning of democratic societies (Hensmans, 2021; Levy & Nielsen, 2010; Martens et al., 2018).

However, on the other hand, even though digital news circulations are growing at double-digit rates and about 69% of newspapers have adopted a paywall for their online offerings (Fletcher & Nielsen, 2020), traditional print media still account for a shrinking, but decisive majority of media and publishing revenues (Campbell et al., 2017; PwC, 2018). In 2021, 142 billion dollars (~80%) of the 178 billion dollars in global newspaper and magazine revenues was based on print (Statista, 2021). At the same time, revenues in the advertising segment are declining for print legacy outlets as well as online outlets and remain uncertain as a permanent revenue stream in the digital future (Chyi & Tenenboim, 2019; Olsen et al., 2020; Sridhar & Sriram, 2015). Prospectively, this trend will intensify, as advertising revenue is increasingly monopolized by tech corporations (Myllylahti, 2020).

Both, the relevance of journalism for a stable democracy and the insufficient success of paid content pose a striking dilemma for business and society. Therefore, while digital journalistic products are fashionable and highly in use (Salaverría, 2019), the aforementioned problems combined with low revenue in the online segment (Berger et al., 2015; Chyi & Ng, 2020) amount to severe financial difficulties, and the industry continues to struggle for a viable business model leading into a sustainable future (Chyi & Lee, 2013; Himma-Kadakas & Kõuts, 2015). The reluctance of the customer to pay for online news clashes directly with the need of journalistic enterprises to acquire new revenue streams, independent of advertising. Therefore, publishers ask why consumers' intention to pay is so low, especially compared to its print counterpart.

2. State of the Literature and Research Question

2.1. The Online News Media Business

Researchers from business administration and media management, journalism and communication studies alike are concerned with the monetisation crisis of news media (Guo & Volz, 2019; Olsen et al., 2021; Robinson et al., 2019). Still, after 25 years of research and experimentation, the concept of digital journalism itself remains remarkably vague (see study 1 and 3). Waisbord (2019) defines the term as "networked production, distribution, and consumption of news and information about public affairs" (p. 352). Some definitions focus on the technological quality to include "all digital forms of journalism, not only those related to interactive networks" (Salaverría, 2019, p. 3) and emphasize attributes of digital technologies (Robinson et al., 2019). Others highlight the journalistic properties of online news media, pointing to the ethos of journalistic quality, in line with its relevance for an informed public in a democracy (Zelizer, 2019). In the latter view, digital journalism is mainly provided by legacy media outlets, which apply reliable standards of quality journalism to the internet. Accordingly, they consider the focus on technological features as limiting and emphasize the contingency of the medium.

While the body of research on the digital news media business is growing steadily, researchers and practitioners alike tend to cling to the blueprint of traditional print journalism. Managerial and conceptual differences are sometimes neglected in this view. This is hardly surprising, considering the harsh reality that many formerly profitable analogue industries are facing (Hanelt et al., 2015).

Online news are not (yet) a mass paid-for product (Chyi & Tenenboim, 2019; Chyi & Yang, 2009; Kammer et al., 2015; see also: chapter B.). The paying intent and actual payments, even for high-quality offerings from legacy outlets, are far lower than for the traditional print newspaper (Berger et al., 2015). The majority refuses to pay for digital journalism in general

(Buschow & Wellbrock, 2019; Casero-Ripollés, 2012). Accordingly, subscriptions to digital news services still account for a fraction of total reader revenue for newspapers in the US (Chyi & Ng, 2020), and the expected disruption of journalism by digital means still has not materialised.

Notably, the empirical studies on payments for digital journalism exhibit a wide range of different shares of respondents that are generally willing to spend money on digital journalism: Himma-Kadakas and Kõuts (2015) report this share to be 7% for Estonia, while the estimates of Reichmann and Klimmt (2012) are situated at the other extreme of the spectrum with 47% of the respondents in Germany being willing to spend money on online news in general (Goyanes, 2015; Gundlach & Hofmann, 2017; Himma-Kadakas & Kõuts, 2015; PwC, 2019; Reichmann & Klimmt, 2012). Some reasons for these differing numbers might be sample selection as well as cross-national differences.

A large share of the audience acts particularly price-sensitive in case of digital journalism, while at the same time being presented with free alternatives online (Gundlach & Hofmann, 2017). Consequently, about half of the people state they would not pay because they prefer cost-free alternatives, while about a quarter indicates that online news are inherently undeserving of paying for (Newman et al., 2017).

Thus, some researchers emphasize the importance of traditional print news media and oppose the trend toward the common "Digital First" strategy (Chyi & Tenenboim, 2019; Pattabhiramaiah et al., 2018; Shafer, 2016). Beyond, some researchers argue, regarding technologies and business models, that journalism has rarely distinguished itself as innovative in a business sense (Usher, 2017).

Previous research on the factors of willingness to pay paints an unclear picture of the demands on digital journalism. Berger et al. (2015) demonstrate in a conjoint study that format is the second most relevant factor (after price) influencing the choice of a news product. They also point to a substantially higher willingness to pay for print products than for digital products (Berger et al., 2015). Beyond, research has addressed features of digital news products. Reichmann and Klimmt (2012) deem the mode of access (via smartphone, desktop, tablets) positively related to paying if it includes as many access modes as possible. One further study examines the preference of an offline access/archive access and states only low part-worth utility for this attribute in a conjoint study, meaning it contributes little to the overall consumer decision (Oechslein, 2014). Since research on this topic applies conjoint analysis, the format/medium in terms of accessibility seems to be a relevant factor. The *perceived quality* of digital journalism or the authoring journalist appears to have a significant positive impact on paying as well (Goyanes et al., 2021; Himma-Kadakas & Kõuts, 2015; Jere & Borain, 2018).

Two things stand out in the research. First, apart from the study of socio-demographic variables and specific product attributes, little research has been done on the psychological-motivational characteristics of the online news business. Second, concrete research on the success of digital journalism companies and start-ups is limited. This paper aims to contribute to both areas. Therefore, the overarching research question of this thesis is: *What factors enable the viability and entrepreneurial success of the news media business in light of the consequences of digital transformation?*

2.2. Digital Transformation of the News Media Business

As in other industries, the digital transformation of journalistic industry comprises chances and risks. Hanelt et al. (2021) assert four aspects of digital transformation as being paramount to its understanding: technology impact, compartmentalized adaptation, systemic shift and holistic co-evolution. Sraml Gonzalez & Gulbrandsen (2022) describe the generation of new solutions, organisational roles and users, new ways of interacting with users, and new opportunities for creativity and innovation as positive changes associated with the transformation of journalism. However, they also expose the negative effects on journalism, such as the devaluation of news content through its free provision online. Guo and Volz (2019) share this ambivalent view on online news, stating that the digital transformation leads to unprecedented changes in “news collection, production, reporting, and dissemination” (p. 5). However, they recognize the tendency of media convergence, where individual journalists must display a whole range of skills related to the digital transformation and cover different tasks and roles in their organisation, while employment in US newsrooms decreased 23% in the decade from 2008 to 2017. The plurality of tasks in the newsroom is brought about by the increased use of digital production and analytical methods, for instance, automated newsrooms and content production. In addition, they discuss the trend towards higher advertising spending, from which news companies do not benefit. Donders et al. (2018) point to digitalisation, internationalisation and pressure on the business model as crucial problems in the modern-day news industry. They also emphasise shrinking circulations, add-money and skipping of ads in the digital environment.

In the broader context of digital transformation, Kosterich (2021) describes the multi-skilled journalistic entrepreneur as an opportunity for the flourishing of journalism. Other research points to the possibility of local online news entrepreneurs bridging the supply gap in wake of dying local print news organizations. However, research also recognises the difficulty to find an appropriate business model in the digital transformation of news (Olsen et al., 2021). They emphasise that the diversification of the news industry into print and online has decreased total business revenue. While they complain that advertising money will not be enough to finance journalism, they recommend focusing on paid content.

The digital transformation is also preparing changes in the perception of news in terms of content. In the context of digital transformation, researchers focus on the problem of fake news (Achtenhagen et al., 2018; Guo & Volz, 2019). While the former business model, determined by gatekeepers in print journalism, limits the velocity and volume of fake news, the speed and fragmentation of online content providers and news distribution via the internet have become a problem for news quality (Martens et al., 2018). Therefore, a functioning online news media business has also a great responsibility from a journalistic quality point of view, in which high journalistic standards should help to reduce disinformation online.

2.3. Online News as Media Products with Information Goods Characteristics

Driven by the digital transformation of the news media business and the accompanying digitisation of content, the nature of the original news product is changing. Formerly a physical, haptic bundle of different print segments in form of a newspaper, the content of online news media appears as pure information online. Changes of the nature of the journalistic product should not be underestimated in the following work.

The seminal monograph “Information Rules” describes information as “(e)ssentially, anything that can be digitized—encoded as a stream of bits (…). For our purposes, baseball scores, books, databases, magazines, movies, music, stock quotes, and Web pages are all information goods fundamental observations in this regard.” (Shapiro & Varian, 1999, p. 3). Under digital journalism’s modus operandi, news are now to be understood as part of the broad tent of information goods. Information goods have the property of being expensive in initial production, but cheap – even increasingly so – in reproduction. Similarly, digital news media products have high fixed costs and almost no marginal costs. While already partly true in analogue times, digital technology renders the costs for the dissemination of information basically zero.

The ease of reproduction of information goods enables copying by non-authorized sources and free-riding of customers/users. Therefore, it is often hard for producers of information goods to cover production costs. Noteworthy, online news as information goods in many ways resemble public goods. Online news are difficult to make exclusive, while the digital attribute makes digital journalism appear as non-rival. As well This fact debilitates the ability to properly monetize digital journalism even further and is addressed in later parts of this thesis.

Finally, network effects, known from analogue business ventures, are accentuated for information goods on the internet. Large companies benefit from network effects that can accelerate concentration tendencies. Whether in the streaming sector, platforms like Facebook or operating systems: A natural tendency toward oligopolistic structures is emerging (Jones & Mendelson, 2011). Through these “winner-takes-all” dynamics, the tendency toward a few large providers serving the mass market and many small companies operating in the niche is strengthening. Supply and demand are both becoming more differentiated. For small entities and single individuals, entry into the market can look profitable due to small barriers from a technological perspective (Clemons & Lang, 2003).

2.4. Paid Content and Entrepreneurship

Concerning the business model of digitally transformed journalism, various approaches are emerging in research. Two recurrent themes were identified in the literature to address the problem: paid content (Chyi, 2012; Goyanes et al., 2021; Herbert & Thurman, 2007) and entrepreneurship (Buschow, 2018; Naldi & Picard, 2012).

Firstly, the approach via paid content prescribes converting non-paying users of free offers (decreasingly financed by advertising) to paying customers. Given the dwindling importance of advertising revenues, especially for small and new journalistic providers, Evens and Van Damme (Evens & Van Damme, 2016) describe the prominent role of paid content and multi-channel distribution of content. Several scholars emphasise the importance of paid-for content in response to the decline of advertising-based business models (Berger et al., 2015; Chyi & Lee, 2013; Himma-Kadakas & Kõuts, 2015). Business models derived from the industrial age, appropriate to physical goods, do not function properly in this new environment (Hanelt et al., 2015). The digital revolution has shattered old profitability regimes (Teece, 2018). They argue that reliance on advertising is not sustainable, therefore making users pay directly is the only practical way to make online news viable and independent (Herbert & Thurman, 2007).

However, although some countries have higher penetration rates of digital journalism, research is debating whether there is a fundamental problem standing in the way of the business success of online news. Dou (2004) demonstrates in an important paper the importance of the free mentality, that is, the attitude that certain (online) goods are available to everyone for free at all times. This attitude is discussed as a legacy of the former cross-financing of some information goods providers. Radio, TV and the internet, for example, mostly cross-financed their offerings through advertising. Correspondingly, Veermeer et al. (2020) point to the change in usage behaviour of news users online. Whereas previously a print news source was primary for most users, users are now able and willing to consume many different free news stories online in a variety of ways.

Consequently, one major problem of digital journalism appears to be the audience's free mentality concerning online news (see study 2). Thus, the audience behaves as payment-averse and thinks of generic online news as a common or public good, which everybody is entitled to consume (Goyanes et al., 2022; Sraml Gonzalez & Gulbrandsen, 2022). Further evidence for this "free-lunch" mentality (Lin et al., 2013) is given by a study, which demonstrates that the people not willing to pay are more convinced than those with stated willingness to pay that online news should stay free (Ye et al., 2004). Other scholars considered payments for Public Service Media (PSM) linked to free mentality, as PSM fees could appear as pre-emptive payments for news in general. However, cross-nationally, this effect was not observed (Fletcher & Nielsen, 2017).

To increase chances to monetise paid content in light of free mentality, knowledge of the customer via sophisticated analytics and machine learning algorithms in news production and distribution (Ahmed & Ahmed, 2021; Friedrichsen et al., 2017), marketing to specific groups and individuals (Koukova et al., 2008) and paywall management (Evens & Van Damme, 2016; Myllylahti, 2014; Rußell et al., 2020) are crucial ways to identify or convert potential customers.

Secondly, the approach via entrepreneurship (organizational/provider perspective) is based on scholarly research, which implies strong innovative potential for start-ups (Buschow, 2020; Khajeheian, 2017; Naldi & Picard, 2012). While digital market structures still feature size advantages and therefore tend to favour large companies on the distribution stage (Hindman, 2018), it is also true that decreasing fixed costs on the production stage have enabled smaller entities, entrepreneurs and start-ups to produce high-quality content (Kohn & Wewel, 2018). Still, research stresses the various organizational hurdles (Deuze & Witschge, 2018), the need for a network and collaboration (Mütterlein & Kunz, 2017), as well as various necessary abilities of a media entrepreneur (Achtenhagen et al., 2018; Gillmor, 2016; Phillips et al., 2009). Berman (2012) also points to the capabilities needed to modernize value propositions of companies in the process of digital transformation. All in all, research suggests the modern digital journalist founder to be a multi-talented and entrepreneurially orientated creature (Phillips et al., 2009).

While the research at hand partially aims at discerning factors of success, some critical remarks concerning the feasibility of such an endeavour are necessary (see also study 3). While research into corporate success factors is conducted extensively, scholars also repeatedly formulate fundamental criticism of such conceptualizations of success. Dömötör (2011) enumerates general, thematical and methodological criticism of success factor research. A general criticism relates to the intrinsic complexity and multi-causality of the subject in question, through which scholarly research has difficulty to establish causal relationships (March & Sutton, 1997). Paradoxically, Dömötör also problematizes the tendency that research related to business success is often integrated into an actual business. He explains that the measurable effects of the factors formulated in research would diminish over time, as companies would be homogenized by implementation in this respect, and thus no measurable effects would remain. Beyond, regarding the themes of success research, Dömötör criticises certain foci of success factor research (e.g. on marketing) and the mere diversity of independent variables in different studies, which make meaningful comparison difficult. Finally, methodological criticism relates to the application of online surveys and matters of data analysis.

In order to answer the research question, the aim of this thesis with regards to the digital news media business is therefore twofold: Firstly, to examine the factors which influence the willingness to pay for paid online content in the news media industry, and, secondly, to search for entrepreneurial antecedents of the success of start-up companies offering online news.

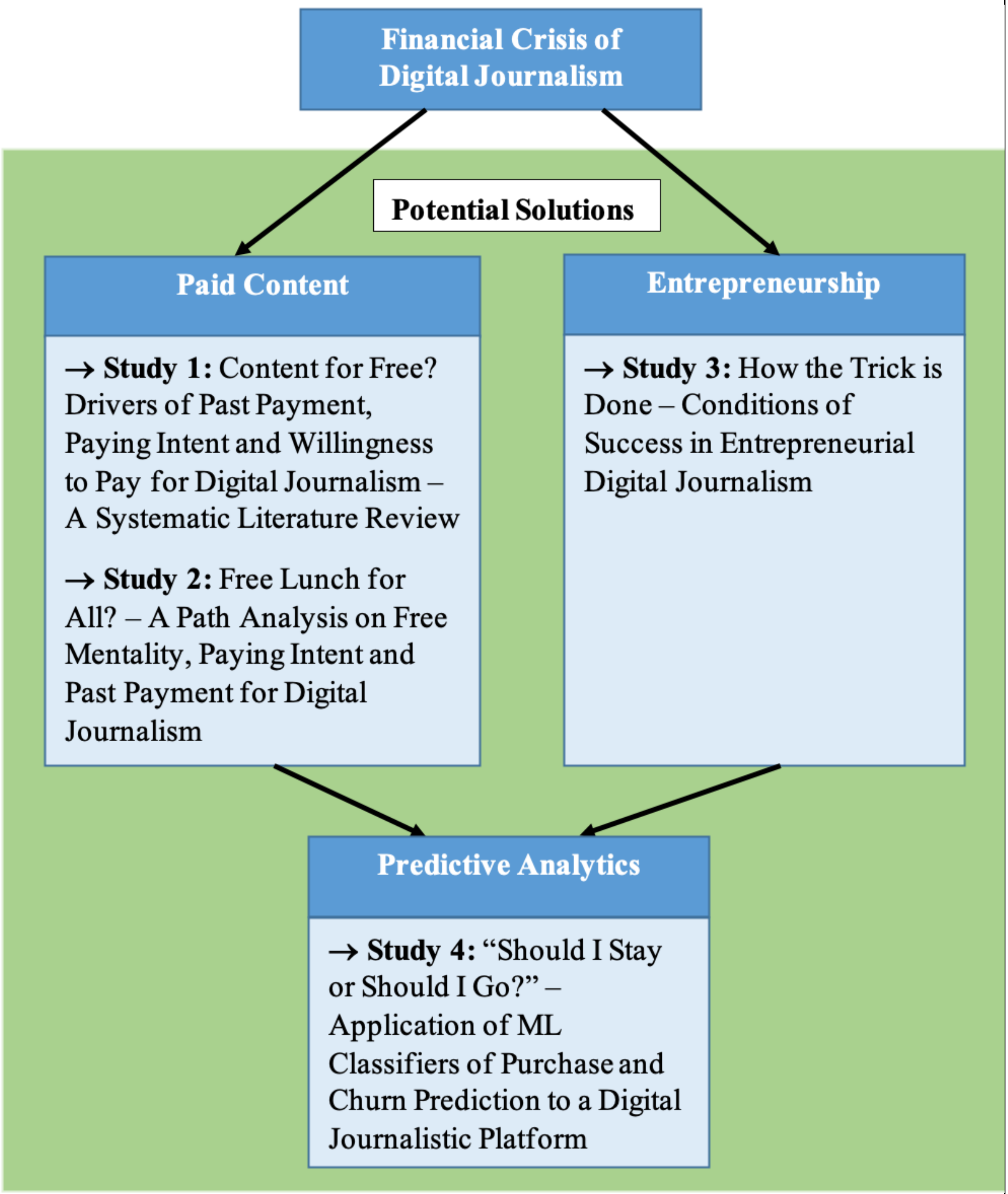
3. Structure of the Thesis

3.1. Overview

To holistically grasp the difficulties of the news media business and how it copes with the process of digital transformation, the work at hand relies on various approaches and topics relevant to digital journalism. As specified above, paid content as well as entrepreneurship are crucial topics in this regard. Accordingly, this dissertation addresses the “paid content” approach by (1) identifying the most important drivers of paid content using a systematic literature review and (2) examining the concept of free mentality, said to be a central inhibitor of paying intent, in more detail utilising survey data. Furthermore, this dissertation addresses the entrepreneurship approach by (3) empirically examining the key success conditions for digital journalism start-ups using a QCA (based on survey data). Finally, this research combines the two approaches/perspectives by (4) performing purchase and churn prediction using usage and payment data of a journalistic start-up via state-of-the-art machine learning algorithms. Figure A.1 depicts the design of the thesis at hand.

Various data sources are included, for example, cross-sectional survey data of online users, structured interviews with German start-ups, and user data for an empirical case study in a start-up setting. Methods applied in this research are, amongst others, Qualitative Comparative Analysis (QCA), path analysis, and predictive analytics via random forest classifiers and gradient boosting. The remainder of this chapter presents the research papers, which are part of the thesis. The papers are presented in the context of the respective research question, the applied method, as well as the publication state of the article.

Figure A.1: Research Scheme



3.2. Study 1: Content for Free? Drivers of Past Payment, Paying Intent and Willingness to Pay for Digital Journalism – A Systematic Literature Review

Published in: O'Brien, D., Wellbrock, C. M., & Kleer, N. (2020). Content for free? Drivers of past payment, paying intent and willingness to pay for digital journalism – A systematic literature review. *Digital Journalism*, *8*(5), 643–672. https://doi.org/10.1080/21670811.2020.1770112

History:

A different version was submitted to Association for Education in Journalism and Mass Communication (AEJMC, track "Graduate Student Interest Group") and presented at the respective conference in Toronto, August 2019. A first rudimentary version was published in German as a book chapter: Wellbrock, C.-M., & Buschow, C. (2020). Bestandsaufnahme: Stand der Forschung zur Zahlungsbereitschaft für digitalen Journalismus [A review: the state of research on willingness to pay for digital journalism]. In C.-M. Wellbrock & C. Buschow (Eds.), *Money for Nothing and Content for Free?: Paid Content, Plattformen und Zahlungsbereitschaft im digitalen Journalismus* (1 ed., pp. 23–38). Nomos. https://doi.org/10.5771/9783748907251-23

Authors:

Daniel O'Brien, Christian-Mathias Wellbrock, Nicola Kleer

This literature review served as a first conceptual anchor for the overarching research. The research focused on the factors that influence paying for digital journalism. Due to the exploratory nature of this paper, a systematic literature review based on the structured approach of Webster and Watson (2002) was applied. This article provides a review of 17 factors found in 37 articles that contribute to consumers' past payment (PP), paying intent (PI) and willingness to pay (WTP) for digital journalistic content.

It is important to note that researchers not only use different measurements but also different terms concerning this issue. On the one hand, some use the term PI to measure whether someone is inclined to buy digital journalistic products at all (in terms of a dichotomous "Yes"/"No" or a rating scale), while WTP is a quantitative measurement of how much (money) people are willing to pay (Chyi, 2005; Chyi & Yang, 2009). Goyanes (2014) confirms this view, as he defines WTP as the "maximum amount one is willing to pay for a product" (p. 746). Nonetheless, he uses the terms PI and WTP at times synonymously (Goyanes, 2015). Many other scholars adhere to the same inconsistencies in definition (Dou, 2004; Fletcher & Nielsen, 2017; Ye et al., 2004).

Despite the definitional vagueness in the literature, various antecedents of the audience's willingness to pay are already discernible. This research suggests that gender (being male), education, media use, news interest, format/medium (print or bundle), customization/personalization, (perceived) quality, specialization/niche (e.g. local) and income have a positive impact on at least one of the three dependent variables PP, PI, and WTP. In contrast, age, price and free mentality have a negative effect on at least one of them. While those factors are linked to either consumer, the product or its economics, little research examines the psychological needs and motives associated with demand for digital journalism. Nevertheless, free mentality as a consistent psychological attitude is discussed in this regard (Chyi & Lee, 2013; Dou, 2004; Lin et al., 2013), which is examined in detail empirically in study 2.

While confirming the low eagerness of the audience to pay for content online news media businesses, this work contributes also to the identification of broader trends in the literature. Firstly, the strong divergence in the methods and samples concerning measurements of PP, PI and WTP inform why numbers of the share of the audience willing to pay for content vary to such a degree. Thereby, this study brings structure in the highly dispersed literature. Secondly, regarding the overarching research question, a variety of factors were identified. However, price and free mentality appear as particularly strong determinants of paying for online news. This further informed the models examined in study 2.

Table A.1: Personal contribution Study 1
1. Intellectual input: ☐Less than 25% ☒25%-50% ☐51%-75% ☐76%-100% Comments: This research idea majorly stems from a project, at which both my co-authors and myself were part of. Additionally, it results from our joint observation that a concise literature review is missing. Thus, I conducted the literature review, in order to define the overall research problem.
2. Experimental set-up and results: ☐Less than 25% ☐25%-50% ☒51%-75% ☐76%-100% Comments: The idea to synthesize existing knowledge by literature review stemmed from my first co-author. The execution and the interpretation of the results of the literature review happened in collaboration with my second co-author.
3. Writing process: ☐Less than 25% ☐25%-50% ☒51%-75% ☐76%-100% Comments: Development of the paper was based on one version developed by me and presented at a conference. Once a first draft was established, all authors improved steadily on it.

3.3. *Study 2: Free Lunch for All? – A Path Analysis on Free Mentality, Paying Intent and Media Budget for Digital Journalism*

Published in: O’Brien, D. (2022). Free lunch for all? – A path analysis on free mentality, paying intent and media budget for digital journalism. *Journal of Media Economics*, *34*(1), 29–61. https://doi.org/10.1080/08997764.2022.2060241

History:
Different versions were submitted and accepted to the Journalism conference 2020 in Vienna (cancelled due to Corona), EMMA 2020 (cancelled due to Corona) and conference of International Communication Association (ICA, track: “Journalism studies”) 2021 (held online in June, 2021). The full paper in the present form was submitted to and won the “Nachwuchspreis Fachgruppe Medienökonomie, DGPuK”. A first rudimentary version was published in German as a book chapter: Wellbrock, C.-M., & Buschow, C. (2020). “Free or Nothing” – Gratis-Mentalität im Internet und Zahlungsbereitschaft für Digitaljournalismus ["Free or Nothing" - Free Mentality on the Internet and Willingness to Pay for Digital Journalism]. In C.-M. Wellbrock & C. Buschow (Eds.), *Money for Nothing and Content for Free?: Paid Content, Plattformen und Zahlungsbereitschaft im digitalen Journalismus* (1 ed., pp. 47–68). Nomos. https://doi.org/10.5771/9783748907251-47

Authors:
Daniel O’Brien
In light of declining print journalism circulation, job loss in journalism and increasing competition from tech giants, the digital news media business intends to monetize their content directly as paid content (Myllylahti, 2014; Newman et al., 2020). To compensate decreasing revenues, the willingness to pay for such content is a critical issue for the economic viability

of digital journalism. However, the reluctance of the customer to pay for online news clashes directly with the need of journalistic enterprises to acquire new revenue streams, independent of advertising. Therefore, publishers ask why consumers' willingness to pay is so low, especially compared to its print counterpart (Berger et al., 2015).

As derived from the literature review (study 1), free mentality is often cited as one explanation for the generalized low willingness to pay for online goods. In short, free mentality describes the aversion of the consumer to accept any price point other than zero (Niemand et al., 2019). The notion of a free mentality, or "free lunch" mentality, is discussed as a problem of various information goods, such as games (Hamari et al., 2017), online music (Lin et al., 2013), software (Niemand et al., 2015) and online clip art (Dou, 2004). Especially regarding the information good *news*, the emergence of the free mentality is a direct consequence, firstly, of the education of the previous cross-subsidised mass media, and, secondly, of the digital transformation of the industry, moving away from the haptic good.

However, the reasons for free mentality are not well understood and the term remains vague regarding its conceptual basis. Research tackles questions of convictions on the inherent value of information goods (Lin et al., 2013), the existence of free alternatives (Dou, 2004) or habits (Chyi, 2012; Fletcher & Nielsen, 2017; Goyanes et al., 2022), for example, advertising-based business models have heavily subsidized the recipient market in the past and mostly refrained from pricing the recipient side in the various sub-markets (radio, television, internet).

For this purpose, the data of a comprehensive user survey in Germany (n = 1,004) was analysed via path analysis. Results confirm low paying intent in the public and the role of free mentality as a relevant factor therein. The research further suggests that payments for public service broadcasters explain significant parts of this relationship. However, preference for advertising based finance models is not a significant reason for low paying intent.

Concerning the general research project, this study pinpoints a crucial problem of digital journalism; namely, free mentality, that is, the attitude of a significant part of the audience that news on the internet should be free. This tendency, which is particularly pronounced for general-interest news, implies that successful digital journalism must escape this generic-valuation trap through differentiation and offering specific added value to paid online news content. While the examination of free mentality for digital news media has already been hinted at by Dou (2004), it has not yet been empirically verified.

Table A.2: Personal contribution Study 2
1. Intellectual input: ☐Less than 25% ☐25%-50% ☐51%-75% ☒76%-100% Comments: The idea for this research project came from a study for the State Authority of in NRW, which I accompanied in its implementation. In this context, data was collected that I intended for the planned model of this study.
2. Experimental set-up and results: ☐Less than 25% ☐25%-50% ☐51%-75% ☒76%-100% Comments: The idea for this research project was inspired by the preceding literature review. In particular, the lack of focus on this obvious, yet overlooked, topic in the literature was striking. The data preparation and analysis was conducted completely on my own.
3. Writing process: ☐Less than 25% ☐25%-50% ☐51%-75% ☒76%-100% Comments: Outside of collegial discussions, the writing process and analysis in this research project was conducted solely by me.

3.4. *Study 3: How the Trick is Done – Conditions of Success in Entrepreneurial Digital Journalism*

Published in: O'Brien, D., & Wellbrock, C.-M. (2021). How the trick is done – Conditions of success in entrepreneurial digital journalism. *Digital Journalism* (Special Issue: The Business of Journalism, online first). https://doi.org/10.1080/21670811.2021.1987947

History:
Different versions were submitted and accepted to WMEMC 2020 (cancelled due to Corona), IMMAA 2020 (cancelled due to Corona), ECREA 2020 (Track "Media Industries and Cultural Production", cancelled due to Corona), and "DGPuK-Jahrestagung 2021" (held online in May, 2021).

Authors:
Daniel O'Brien, Christian-Mathias Wellbrock

One promising aspect of the ongoing digital transformation of journalism is the lowered barrier or entry for smaller companies and start-ups (Kosterich, 2021). This enabled the emergence of a multitude of smaller and independent providers of journalism, ranging from local news blogs to citizen journalists (Naldi & Picard, 2012). The digital transformation reinforces this shift towards a multitude of small providers and a few big players able to quench the public's thirst for news (Newman et al., 2021). In this sense, entrepreneurial journalism is often ascribed to the potential to partly close the supply gap in journalism, even though digital journalistic founders and start-ups are struggling frequently (Buschow, 2018; Wellbrock & Buschow, 2020).

This study aims to empirically investigate antecedents of success of start-up entrepreneurs in the digital news business. While there is a broad stream of literature on entrepreneurial success in general, empirical research says little about the reasons for digital journalistic entrepreneurs' success (or lack thereof).

Based on the literature, seven broader "super"-conditional prerequisites are considered: Experience, skills, personality, product, business model, company organization, and the broader environment. Many papers with regards to business success are either conducted with standard statistical analysis or in terms of various methods of qualitative research (Carsrud et al., 1989; Goedhuys & Sleuwaegen, 2010; Overall & Wise, 2016). In contrast, the data of this paper, collected through 49 interviews with digital journalistic entrepreneurs in Germany, will be analysed with Qualitative Comparative Analysis (QCA). QCA originated with Ragin (1987) and is a set-theoretical method to determine the necessity/sufficiency of preformulated conditions for a respective outcome (i.e. entrepreneurial (non-)success). The results indicate high importance of, firstly, the founder's experience, and, secondly, the functioning organization of the company, and, interestingly, thirdly, absence of support from the broader environment and/or government contributed to the success.

For practitioners in the news media industry, this research on success in start-ups provides important insights into necessary skills, motives, and resources. As our results demonstrate, particularly the role of comprehensive experience and a functional organization are relevant to success. To some extent, this is good news to practitioners: Experience is improvable. The same holds for a functional organization, which relies heavily on teachable administrative tasks or the organization of/networking with other people.

Overall, it is mainly business and management skills that lead to success. The moral ethos of many journalists, however, tends to assign lower priorities to the economic aspects of their business – a view that was expressed repeatedly by respondents. The present findings suggest that digital journalism founders should invest in (formal) training, competent mentors, and other support mechanisms as well as foster inclusive decision-making and strong professional networks outside the organization. Thereby, this study contributes to the under-researched antecedents of success for small and start-up companies.

Table A.3: Personal contribution Study 3
1. Intellectual input: ☐Less than 25%   ☐25%-50%   ☒51%-75%   ☐76%-100% Comments: This research idea majorly stems from a joint research project, of which both my co-author and myself were part of. It further results from my review of the literature and the identification of an empirical research gap.
2. Experimental set-up and results: ☐Less than 25%   ☐25%-50%   ☒51%-75%   ☐76%-100% Comments: The idea to use interviews and to apply this particular method did not stem from me. The idea for the empirical set-up and the method was co-developed with my co-author. The analysis and the interpretation of the results are more or less my contribution.
3. Writing process: ☐Less than 25%   ☐25%-50%   ☒51%-75%   ☐76%-100% Comments: Development of the paper was based on one version developed by me. Once a first draft was established, we jointly worked on it.

3.5. Study 4: "Should I Stay or Should I Go?" – Application of ML Classifiers of Purchase and Churn Prediction to a Digital Journalistic Platform

Submitted to: *Journal of Media Business Studies* (under Review since April 2022)

History:

An extended abstract of this paper was presented at EMMA 2022 in Munich and a version of the full paper was nominated for the Best Paper Award. A full paper is under review at the *Journal of Media Business Studies*.

Authors:

Daniel O'Brien, Reinhard Kunz

While many industries are fully embracing the digital transformation in all facets of production, distribution and management, journalism is sometimes a victim of its own "conservative" attitude towards innovation (Sylvie & Gade, 2009; Wilczek, 2019). Nevertheless, considering its financial problems, the news media business could benefit from embracing methods implied in the digital transformation of the industry. Researchers emphasized the importance of predictive analytics and data science applications in digital journalism (De Grove et al., 2020; Jiang et al., 2020; Neheli, 2018). In the face of journalism's monetisation crisis, a prudent and creative approach to new methods of analysis and customer management is called for. Thus, the final study investigates (1) whether and how digital journalistic companies and start-ups can benefit from the application of predictive ML methods and (2) which factors determine purchase and churn behaviour in such settings.

Machine learning-based customer decision prediction assumes that large data sets generated by different businesses are indicative of underlying patterns, which enable classification or prediction (Martínez et al., 2020). Ahmad et al. (2019) argue that steep competition in the digital telecommunication sector emphasizes the importance of the identification of loyal customers.

To this end, the present research applied random forest and gradient boosting models to a small dataset of a German platform for online news content. The platform enables freelance journalists to monetize their content by paying a fraction of their revenue to the platform provider, while the platform provides an integrated content management system. The platform operates in a thematic niche.

The good to excellent achieved prediction accuracy was comparable to the results in similarly small or even larger sample sizes. The most important features were concerned with recency, frequency, and time spent in terms of customer interaction. Keeping customers on the investigated platform for journalistic content and retaining them is therefore a major difficulty, as most customers only stay for a few actions and/or purchases. Possible solutions include free articles or free time on the platform to reduce uncertainty concerning the media service. Furthermore, news media companies should focus on user engagement via appropriate comment functions or networking via social media.

As the results of this study illustrate, even small entrepreneurial entities can benefit from applying algorithmic customer behaviour prediction. Those methods should be embraced to better sustain and manage digital transformation, even in small business contexts. As the media management literature asserts, the proliferation of such techniques to SMEs and start-ups lessens the competitive advantage of large companies and economies of scale (Buschow, 2020; Chan-Olmsted, 2019).

The difficult business situation of online news publishers will require a variety of solutions, of which the contemporary methods of analysing user data is one. While this study is one of few to address the application of predictive analytics in the news media business, potential solutions of this kind are increasingly feasible for large and small companies.

Table A.4: Personal contribution Study 4
1. Intellectual input: ☐Less than 25% ☐25%-50% ☒51%-75% ☐76%-100% Comments: This research idea majorly stems from thoughts gained during a research project at the University of Cologne, and in consultation with my co-author. It further results from my observation of a gap in the literature.
2. Experimental set-up and results: ☐Less than 25% ☐25%-50% ☐51%-75% ☒76%-100% Comments: The idea to use this particular data set did not stem from me, as the data emerged as part of a research project. The idea for the empirical set-up was co-developed with my co-author. The execution and the interpretation of the results are more or less my contribution.
3. Writing process: ☐Less than 25% ☐25%-50% ☒51%-75% ☐76%-100% Comments: A draft version of the paper was written by myself, however, it was also entirely changed and prepared for publication through both authors in the process.

4. Conclusion

4.1. Summary and Learnings

As the quotes at the beginning allude, digital transformation has changed our economic and cultural landscape fundamentally, rapidly, and with increasing velocity. In light of this new landscape, the research question of this thesis concerned the viability of the news media business under the impact of digital transformation. The present research tried to address the difficulties of entrepreneurial digital journalism in four studies through the lens of paid content (Berger et al., 2015; Herbert & Thurman, 2007; Rußell et al., 2020) and entrepreneurship (Buschow, 2020; Naldi & Picard, 2012). However, as the research considering digital transformation suggests, there is no one-size-fits-all answer. While it seems to be true that digital journalism services do not have the same broad-based demand, their popularity is undisputed (Hölig & Hasebrink, 2020).

To better understand what differentiates digital journalism from the old business model of print news, this research focuses on the specifics of the medium from a variety of perspectives. As it seems, the demand and valuation by the audience, and, consequently, the willingness to pay for digital counterparts to analogue news were significantly overestimated (Chyi, 2005; Chyi & Ng, 2020; Goyanes et al., 2022).

The systematic analysis of the literature shows that there are several factors that determine the audience's willingness to pay. However, the impact of these factors is often small and heavily debated in academia. It is also apparent that only a limited number of factors and theoretical frameworks have been studied in the literature. Attributes of the product and demographic factors were more in the foreground, followed by media attributes. However, they are less relevant when it comes to understanding what motivates people to pay for online news.

Therefore, one focus of this work has been to identify free mentality as the main barrier to paying for online news (Dou, 2004). The research at hand confirms the general role of free mentality for low online news spending. Beyond, while research up to this point mainly focused either on the general concept of free mentality in other business contexts or examined concepts peripherally related to free mentality (e.g. the zero-price effect), this thesis focuses firstly on the rationales behind free mentality and secondly proposes a broader model for the explanation of paying intent for online news. Concerning the first aspect, this research found that the ideal of the internet as a disseminator of ideas has a strong influence on free mentality. Additionally, within the context of the German market, this research argues that mandatory payments for PSM could have an increasing effect on free mentality. Noteworthy, the latter is highly dependent on the German media system. However, the idea to have a broader view of pre-emptive monetary and non-monetary costs, which a customer has paid before purchasing online news, deserves a closer examination in various economic and cultural environments.

By contrast, the traditional two-sided, advertising-based market model does not seem to play a significant role in explaining free mentality (Goyanes et al., 2022). Publishers subsequently try to adapt by adopting a multi-sided business model. It has become generally accepted that online news elicit only a low willingness to pay. As this study points out, it is crucial to focus on frequent users who are willing to pay and to target new user groups with differentiated products, which are not perceived as general information goods, therefore suffer less from the free mentality. Legacy outlets and start-up publishers will have to focus on different ways to add value to their goods, as mere information suffers rapid devaluation in the marketplace, and could be considered a public good. Differentiation via stardom or appeal of certain journalists, added value beyond information (such as clubs or preferable access to special treats), or the emphasis on the niche for smaller companies are possible strategies to tackle this issue (Cook & Sirkkunen, 2013).

While research has shown that free mentality is related to lower payments for online goods in general, also demonstrated for games (Hamari et al., 2020) and music (Lin et al., 2013), the situation seems to be even worse for online news. While games and music are highly demanded entertainment goods, news have always existed on the spectrum somewhere in between entertainment and information. While journalism is also a highly used product, it also suffers from rapid devaluation over time. The obvious fact of less time devaluation for music and movies compared to news allows for different business models. While the latter distinction makes it difficult to imagine business models similar to those of the entertainment industry (e.g. bundling), news is particularly controversial in terms of the monetary value that can be created. The old print model of journalism already considered the lower value of information compared to entertainment through cross-funding by bundling different segments (including puzzles, comics, ads, etc.) in one single newspaper.

Additionally, one relevant approach to successfully combating the supply gap in journalism focuses on the importance of new, smaller journalistic companies that can produce cheaper due to low entry barriers (Buschow, 2018; Naldi & Picard, 2012). Thus, they can deliver journalistic quality that is also in demand on the market, despite a smaller production value. For start-ups and freelance journalists, it becomes more attractive to serve a certain niche. Niche journalism, which in print used to be economically viable only for “large” niches, finds a possibility to further differentiate. The segmented taste of the internet is reflected in the equally segmented variety of media providers. This is a chance for entrepreneurship, which is why a part of this paper examined the conditions of success for such smaller entities. By applying QCA, a configurational method (Ragin, 1987), our research revealed that professional and personal experience associated with entrepreneurship, a functioning organization, and the absence of a supporting broader environment (by government and society) in conjunction could lead to success.

The factor experience appears to be of particular relevance for the economic survival of journalistic start-ups. This implies that biographical aspects such as having worked in the media industry, high educational degrees, managerial experience, having been abroad and having family members who are entrepreneurs tend to be helpful factors in becoming digital journalistic entrepreneurs. Beyond, organisational functions were revealed to be crucial. For practical purposes, this is good news. Voids in these areas can potentially be filled by seeking formal training and by consulting counsellors and mentors. Concerning the organizational capabilities, a constructive working environment that involves various members of the organization and stakeholders in decision-making processes and efforts in building a large professional network improve the likelihood of success.

Finally, a detailed look at a particular journalistic start-up operating in the described niche path with a focus on science journalism was intended to close the gap between paid content and entrepreneurship. The present work demonstrates a meaningful use-case of the use of predictive analytics in a start-up setting to improve customer management, marketing and product. Enabled by the digital transformation, such methods are ubiquitous (Ahmed & Ahmed, 2021). Still, even as digital journalism has difficulties monetising properly, it has been slow in adapting to such methods (Usher, 2017). As the results of this research illustrate, even small entrepreneurial entities can benefit from applying algorithmic customer behaviour prediction. Keeping customers on the platform and retaining them as paying customers is a significant difficulty, as most customers only stay for a few actions and purchases. However, by the means demonstrated in this work, such cases can be addressed promptly, for example, by providing several free articles or free time on the platform to reduce customer uncertainty.

4.2. Managerial Implications

There are opportunities and risks in the economic situation of digital journalism. As the present research demonstrates, various developments are taking place simultaneously. The winner-takes-all dynamics are amplified by the internet and much more pronounced than in analogue times. While it has always been the case that the economics of scope has helped give larger firms an advantage over smaller ones, the information goods quality of the news media business potentiates this effect (Varian, 2000). This trend will intensify over time. Already, news, like many other digital industries, fall into oligopolistic structures. “Love brands” and mass-market publishers, such as the New York Times, BILD and similar internationalised and/or brand-conscious companies, stand to benefit from this development, while the diverse environment of earlier times seems to be failing (de Oliveira et al., 2021).

This research reports typically only a small minority of users as potential customers. Only about 10% are found to be current payers for online news platforms, paying about €10 per month on average (comparable to subscribing to a streaming platform), which is in line with former research (Hölig & Hasebrink, 2020). Literature indicates that the bulk of the revenue of any online service is generated by a group of heavy users (Peskin & Glick, 2009). Hopes of achieving penetration rates comparable to print journalism seem utopian. However, considering the present results, a small, but digitally savvy, politically interested user group, which has a higher income and higher education, and consequently less free mentality, should be given special consideration in marketing and orientation. Publishers should think more carefully about their core audience, as wide held free mentality prohibits high paying intent in the broader audience.

Furthermore, a crucial general implication concerns the conceptualization of online news content as public good. The widespread free mentality and its anchoring in the ideal of free information, as well as potentially market-specific payments for PSM, points to the interpretation of generic online news as such a public good. Public goods are defined by being neither excludable nor rivalrous in their use, which is an increasingly debated attribute of various information goods (Shapiro & Varian, 1999). Since more parts of the modern economy, specifically media goods, are turned from physical goods to bits, public goods comprise an increasing share of the overall economy. Those, in turn, are widely distributed via the internet.

Therefore, one research stream states that the audience imagines sharing the cost of the production with a larger population, implying low to zero cost, and therefore possibly discounts the relatively high costs of production and distribution (van der Wurff, 2011).

While this proposition initially suggests negative prospects, it also seems to open possibilities on closer inspection. Publishers must circumvent the “generic news trap”. Specifically, targeted product differentiation and the implementation of additional added value beyond journalism are crucial for the successful branding of modern media companies. This is even more crucial because said companies are now often active in more than two markets and have diverse revenue streams. The differentiation of product range and business models contains strong opportunities for media companies that do not consider themselves mere news agencies. Success cases of international love brands like the New York Times (metered paywall) or focus news like the Wall Street Journal (hard paywall) show the importance of a brand that delivers more than mere news, but a specific focus, spin or lifestyle. On the other hand, past research (Boczkowski et al., 2011) demonstrates the gap between the value that journalists place on news in terms of information and the news that people actually seek out. Entertainment, sports, and consulting journalism are not necessarily subject to the same interpretation as a public good and consequently do not suffer the negative effects of free mentality to the same extent.

Beyond, technical, business, and content innovation can play a significant role, which emphasizes the role of experimentation, as some systems and models perform better than others, and interesting new possibilities arise. The research presented suggests that customers respond positively to digital journalistic content that is, for instance, easy payable via PayPal. The relevance of simple and secure payment systems for online services is emphasized in the literature (Dou, 2004). Online news providers should make it as painless as possible to buy their products. Besides, as the present research suggests, there are signs of higher paying intent for special interest publications and niche topics. Thriving in the “long tail” (Brynjolfsson et al., 2006) is a phenomenon that is especially interesting for small journalistic units, either start-ups or even individual journalists who can maintain their profit in a small niche. This is important, as a small niche can be quite profitable, and “living in the long tail” becomes a viable business model (Cook & Sirkkunen, 2013). While this research highlights the problems of monetising general interest news in the face of free mentality, there is a large but diverse market for niche topics and offerings that meet smaller demands with correspondingly low production value and effort.

For smaller companies, the last decades were inspirational, as production costs and entry barriers for smaller entities diminished over the years. This means that even small entities or single individuals can nowadays produce and distribute journalism of quality and relevance. While the economic situation for journalism, in general, is difficult, the economy of the long tail allows small companies and individuals to examine their niche and survive. For instance, a currently interesting model of monetizing news is the Substack platform. Here, users can sign up for individual subscriptions to journalists, so instead of paying an outlet, the consumer pays the content producer directly. Substack serves as a content management system and payment processor. Interestingly, this has brought financial success to some high-profile journalists.

Successful examples from different industries should be instructive: Media platforms like Spotify and Netflix have worked meticulously to get their audiences to pay for content. These (bundled) platforms and their recommendation systems provide strong value in terms of the convenience of their respective services. Still, there is only a small effort to bundle news products in a consumer-friendly way, in fear of cannibalization. Especially with regards to a powerful recommendation system, social media mostly provides this service of a bundled news aggregator for a large part of the audience. However, the incentives of a social media site are different from those of a comprehensive bundled news site that targets and possibly even counters readers' interests and attitudes.

Both legacy and more recent media outlets must focus on experimentation with subscription and paywall management. There is no "one size fits all" answer. The different models each benefit certain kinds of products and entities. Regardless of the business model, media companies should be encouraged to make paying or contributing as easy as possible, as complicated paying options deter consumers.

In summary, it is important that free mentality, as demonstrated here, is a problem of generic online news. Publishers need to think carefully about how to design a product that is not redundant in the immense amount of online information assets and that adds significant value to the user. Differentiation can be achieved in various ways, e.g. niche topics, prominent and individualistic journalistic voices, new technology, exclusive events and "clubs", or social causes. This is even more important as this contribution demonstrates that paying for PSM is part of the cause for free mentality: this means that media companies in countries with strong PSMs must differentiate themselves against those free products of often high production value.

However, as the last part of this work asserts, digital transformation brings new opportunities for analysing customer data. Given this flood of data, various possibilities arise, that is, firstly, to collect data in real-time, secondly, to analyse its consumers and thirdly to react to changes in demand and/or customer decisions of purchase or churn immediately. Concerning fake news and other distortions of reality, these methods represent sensible steps to counteract misinformation and improve the ability of democracies to decide what is right and what is wrong (Hensmans, 2021).

Nevertheless, while the barriers of entry for journalism have largely disappeared, there are still disadvantages in the market for journalistic entrepreneurs in the digital sphere, which make journalism in large suffer from the transformation of the news business. However, what the present research demonstrates is that many levers can help ensure the viability of the news media business.

4.3. Limitations & Future Research

Even the most comprehensive work on a topic can never answer all critical questions and cover all areas. Therefore, a few words about limitations are necessary. While this work yields crucial insights into the business of online news, it suffers from a lack of cultural diversity as it is focused solely on the German market. This certainly sets limitations on the possibility to draw generalisable conclusions for the industry in a cross-national and cross-cultural way and provides a promising avenue for future research.

In addition, data from cross-national panel studies (Newman et al., 2021) provide insight into the slowly but steadily increasing willingness to pay for parts of the population. This trend, although difficult to observe in compressed time, could turn out to be extraordinarily crucial with respect to the financial future of the online news media business. However, there are time limits to this study that do not allow such conclusions to be drawn.

Even though this study attempts to conceptualise the reasons for low willingness to pay, there are limitations here as well. The data in this study are observational and do not have an experimental or incentive-compatible design. Due to social desirability biases or current news about fake news, the actual willingness to pay in the broad population could be lower than asserted. Given market data hinting at low financial success of most digital journalistic products as well as distrust by the audience, the present approximation seems appropriate to describe the state of the digital news media business.

Another important aspect of the demand for online news media touches upon underlying psychological mechanisms, which might be different for certain products and presentation modes. As our work demonstrates, there are differences in attractiveness between formats and media. Thus, it might be important to conduct further research concerning the interactive and immersive attributes of recent news technologies, especially as there still seems to be a strong preference for print over digital news (Berger et al., 2015).

While the public good view of online news media is discussed in parts of the papers, future research should clarify what kind of goods specifically are perceived as public goods, for instance, how it relates to type of content or medium. In the same vain, a problem only indirectly addressed in this work spans the topic of paywall management. While this research focused on the general unwillingness to pay, future work could examine the relationship between free mentality and the implemented paywall type.

While free mentality is intensively examined in this work, research with respect to the effect of the existence of free alternatives to consume news in general is still sparse. Future research concerning free mentality should consider including free alternatives pull as part of the problem that online news media encounters in monetisation.

In addition, assessments of success factors and its antecedents depend strongly on cultural context and geography. Research on digital journalism and entrepreneurship in developing countries, for instance, highlights other aspects, such as the importance of innovation and uniqueness in the context of a particular region (Harlow, 2018; Salaverría, 2019). Furthermore, the potential of membership models is increasingly discussed in journalism practice and should receive academic research. In summary, it can be stated that a better understanding of underlying consumer needs and motives carries great potential for the development of better journalistic products.

In a similar line of thinking, the already discussed topic of learned habits through other technologies could be important (Goyanes et al., 2022; Niemand et al., 2019). Past media technologies, for instance, radio, television and, more recently, the internet, have strongly subsidized their business via advertising. While advertising revenue shrinks in the competition with tech giants, any additional direct pricing of the online news user would be perceived as particularly aversive. Therefore, habits and price perception formed by other media should be analysed in the context of the current monetization crisis of online news media businesses. In the future, scholars should emphasise this issue.

Furthermore, as some authors argue, the audience might experience not necessarily advertising, but instead data extraction by online content providers as a sufficiently high non-monetary payment to cover for respective services (Hüttel et al., 2018). Users of online news media might know the production costs of news but assumes to pay for with non-monetary means. Considering monetary and non-monetary costs linked to online news use for uses, future studies could take a closer look into data extraction as well as the budget for gadgets (smartphone, computer) and competing media (streaming services, print).

In the context of the application of sophisticated ML methods, online news media businesses and media management can be further developed. Content creation by AI, optimization, A/B-testing and customer segmentation are getting increasingly valuable.

Lastly, questions raised but not conclusively addressed in this work range from the role of community and audience engagement as well as aspects concerning cultural and societal differences between different markets and geographical areas. As Germany was the specific focus of this thesis, it is important to stress that the findings are not without caveats for different countries or the global industry.

5. References

Achtenhagen, L., Melesko, S., & Ots, M. (2018). Upholding the 4th estate—exploring the corporate governance of the media ownership form of business foundations. *International Journal on Media Management*, *20*(2), 129–150. https://doi.org/10.1080/14241277.2018.1482302

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1–24. https://doi.org/10.1186/s40537-019-0191-6

Ahmed, J., & Ahmed, M. (2021). Online news classification using machine learning techniques. *IIUM Engineering Journal*, *22*(2), 210–225. https://doi.org/10.31436/iiumej.v22i2.1662

Berger, B., Matt, C., Steininger, D. M., & Hess, T. (2015). It is not just about competition with "free": Differences between content formats in consumer preferences and willingness to pay. *Journal of Management Information Systems*, *32*(3), 105–128. https://doi.org/10.1080/07421222.2015.1095038

Berman, S. J. (2012). Digital transformation: Opportunities to create new business models. *Strategy & Leadership*, *40*(2), 16–24. https://doi.org/10.1108/10878571211209314

Boczkowski, P. J., Mitchelstein, E., & Walter, M. (2011). Convergence across divergence: Understanding the gap in the online news choices of journalists and consumers in Western Europe and Latin America. *Communication Research*, *38*(3), 376–396. https://doi.org/10.1177/0093650210384989

Brynjolfsson, E., Hu, Y. J., & Smith, M. D. (2006). From niches to riches: Anatomy of the long tail. *Sloan Management Review*, *47*(4), 67–71.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Buschow, C. (2018). *Die Neuordnung des Journalismus: Eine Studie zur Gründung neuer Medienorganisationen* [Re-organizing journalism: A study on the formation of new media organizations]. Springer VS. https://doi.org/10.1007/978-3-658-18872-6

Buschow, C. (2020). Why do digital native news media fail? An investigation of failure in the early start-up phase. *Media and Communication*, *8*(2), 51–61. https://doi.org/10.17645/mac.v8i2.2677

Buschow, C., & Wellbrock, C.-M. (2019). *Money for nothing and content for free? Willingness to pay for digital journalism.* Landesanstalt für Medien NRW. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-65275-7

Campbell, C., Henriksson, T., & Veseling, B. (2017). *World Press Trends*. World Association of Newspapers and News Publishers. http://www.wan-ifra.org/reports/2017/10/10/world-press-trends-2017

Carsrud, A. L., Olm, K. W., & Thomas, J. B. (1989). Predicting entrepreneurial success: Effects of multi-dimensional achievement motivation, levels of ownership, and cooperative relationships. *Entrepreneurship & Regional Development*, *1*(3), 237–244. https://doi.org/10.1080/08985628900000020

Casero-Ripollés, A. (2012). Beyond newspapers: News consumption among young people in the digital era. *Comunicar*, *20*(39), 151–158. https://doi.org/10.3916/c39-2012-03-05

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management*, *21*(3–4), 193–215. https://doi.org/10.1080/14241277.2019.1695619

Chyi, H. I. (2005). Willingness to pay for online news: An empirical study on the viability of the subscription model. *Journal of Media Economics*, *18*(2), 131–142. https://doi.org/10.1207/s15327736me1802_4

Chyi, H. I. (2012). Paying for what? How much? And why (not)? Predictors of paying intent for multiplatform newspapers. *International Journal on Media Management*, *14*(3), 227–250. https://doi.org/10.1080/14241277.2012.657284

Chyi, H. I., & Lee, A. M. (2013). Online news consumption: A structural model linking preference, use, and paying intent. *Digital Journalism*, *1*(2), 194–211. https://doi.org/10.1080/21670811.2012.753299

Chyi, H. I., & Ng, Y. M. M. (2020). Still unwilling to pay: An empirical analysis of 50 U.S. newspapers' digital subscription results. *Digital Journalism*, *8*(4), 526–547. https://doi.org/10.1080/21670811.2020.1732831

Chyi, H. I., & Tenenboim, O. (2019). Charging more and wondering why readership declined? A longitudinal study of U.S. newspapers' price hikes, 2008–2016. *Journalism Studies*, *20*(14), 2113–2129. https://doi.org/10.1080/1461670x.2019.1568903

Chyi, H. I., & Yang, M. J. (2009). Is online news an inferior good? Examining the economic nature of online news among users. *Journalism & Mass Communication Quarterly*, *86*(3), 594–612. https://doi.org/10.1177/107769900908600309

Clemons, E. K., & Lang, K. R. (2003). The decoupling of value creation from revenue: A strategic analysis of the markets for pure information goods. *Information Technology and Management*, *4*(2), 259–287. https://doi.org/10.1023/A:1022958530341

Cook, C., & Sirkkunen, E. (2013). What's in a niche? Exploring the business model of online journalism. *Journal of Media Business Studies*, *10*(4), 63–82. https://doi.org/10.1080/16522354.2013.11073576

De Grove, F., Boghe, K., & De Marez, L. (2020). (What) can journalism studies learn from supervised machine learning? *Journalism Studies*, *21*(7), 912–927. https://doi.org/10.1080/1461670X.2020.1743737 de Oliveira, S. S., Kovacs Grigg, M. H., & Doicaru, M. (2021). The influence of trustworthiness on brand love for news media brands: A cross-cultural examination of the Netherlands and Brazil. *Journal of Media Business Studies*. https://doi.org/10.1080/16522354.2021.1945877

Deuze, M., & Witschge, T. (2018). Beyond journalism: Theorizing the transformation of journalism. *Journalism*, *19*(2), 165–181. https://doi.org/10.1177/1464884916688550

Dömötör, R. (2011). Kritik an der Erfolgsfaktorenforschung [Criticism of success factor research]. In N. Franke, D. Harhoff, & J. Henkel (Eds.), *Erfolgsfaktoren der Innovativität von kleinen und mittleren Unternehmen* (pp. 55–63). Springer. https://doi.org/10.1007/978-3-8349-6552-3_4

Donders, K., Enli, G., Raats, T., & Syvertsen, T. (2018). Digitisation, internationalisation, and changing business models in local media markets: An analysis of commercial media's perceptions on challenges ahead. *Journal of Media Business Studies*, *15*(2), 89–107. https://doi.org/10.1080/16522354.2018.1470960

Dou, W. (2004). Will Internet users pay for online content? *Journal of Advertising Research*, *44*(4), 349–359. https://doi.org/10.1017/S0021849904040358

Evens, T., & Van Damme, K. (2016). Consumers' willingness to share personal data: Implications for newspapers' business models. *International Journal on Media Management*, *18*(1), 25–41. https://doi.org/10.1080/14241277.2016.1166429

Fletcher, R., & Nielsen, R. K. (2017). Paying for online news. *Digital Journalism*, *5*(9), 1173–1191. https://doi.org/10.1080/21670811.2016.1246373

Fletcher, R., & Nielsen, R. K. (2020). Are Netflix and Spotify subscribers more likely to pay for online news? Comparative analysis of data from six countries. *International Journal of Communication*, *14*, 3439–3457.

Friedrichsen, M., Kamalipour, Y. R., & Kamalipour, Y. (2017). *Digital transformation in journalism and news media*. Springer. https://doi.org/10.1007/978-3-319-27786-8

Gillmor, D. (2016). Towards a new model for journalism education. *Journalism Practice*, *10*(7), 815–819. https://doi.org/10.1080/17512786.2016.1164998

Goyanes, M. (2014). An empirical study of factors that influence the willingness to pay for online news. *Journalism Practice*, *8*(6), 742–757. https://doi.org/10.1080/17512786.2014.882056

Goedhuys, M., & Sleuwaegen, L. (2010). High-growth entrepreneurial firms in Africa: A quantile regression approach. *Small Business Economics*, *34*(1), 31–51. https://doi.org/10.1007/s11187-009-9193-7

Goyanes, M. (2015). The value of proximity: Examining the willingness to pay for online local news. *International Journal of Communication*, *9*, 1505–1522.

Goyanes, M., Artero, J. P., & Zapata, L. (2021). The effects of news authorship, exclusiveness and media type in readers' paying intent for online news: An experimental study. *Journalism*, *22*(7), 1720–1738. https://doi.org/10.1177/1464884918820741

Goyanes, M., Demeter, M., & de Grado, L. (2022). The culture of free: Construct explication and democratic ramifications for readers' willingness to pay for public affairs news. *Journalism*, *23*(1), 207–223. https://doi.org/10.1177/1464884920913436

Gundlach, H., & Hofmann, J. (2017). Preferences and willingness to pay for tablet news apps. *Journal of Media Business Studies*, *14*(4), 257–281. https://doi.org/10.1080/16522354.2017.1346948

Guo, L., & Volz, Y. (2019). (Re)defining journalistic expertise in the digital transformation: A content analysis of job announcements. *Journalism Practice*, *13*(10), 1294–1315. https://doi.org/10.1080/17512786.2019.1588147

Hamari, J., Hanner, N., & Koivisto, J. (2017). Service quality explains why people use freemium services but not if they go premium: An empirical study in free-to-play games. *International Journal of Information Management*, *37*(1, Part A), 1449–1459. https://doi.org/10.1016/j.ijinfomgt.2016.09.004

Hamari, J., Hanner, N., & Koivisto, J. (2020). "Why pay premium in freemium services?" A study on perceived value, continued use and purchase intentions in free-to-play games. *International Journal of Information Management*, *51*, 102040. https://doi.org/10.1016/j.ijinfomgt.2019.102040

Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, *58*(5), 1159–1197. https://doi.org/10.1111/joms.12639

Hanelt, A., Piccinini, E., Gregory, R. W., Hildebrandt, B., & Kolbe, L. M. (2015). Digital transformation of primarily physical industries-exploring the impact of digital trends on business models of automobile manufacturers. *Wirtschaftsinformatik*, *88*(3–5), 1313–1327.

Harlow, S. (2018). Quality, innovation, and financial sustainability: Central American entrepreneurial journalism through the lens of its audience. *Journalism Practice*, *12*(5), 543–564. https://doi.org/10.1080/17512786.2017.1330663

Hensmans, M. (2021). Exploring the dark and bright sides of Internet democracy: Ethos-reversing and ethos-renewing digital transformation. *Technological Forecasting and Social Change*, *168*, 120777. https://doi.org/10.1016/j.techfore.2021.120777

Herbert, J., & Thurman, N. (2007). Paid content strategies for news websites: An empirical study of British newspapers' online business models. *Journalism Practice*, *1*(2), 208–226. https://doi.org/10.1080/17512780701275523

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for formulating a digital transformation strategy. *MIS Quarterly Executive*, *15*(2), 123–139.

Himma-Kadakas, M., & Kõuts, R. (2015). Who is willing to pay for online journalistic content? *Media and Communication*, *3*(4), 106–115. https://doi.org/10.17645/mac.v3i4.345

Hindman, M. (2018). *The Internet trap: How the digital economy builds monopolies and undermines democracy*. Princeton University Press. https://doi.org/10.2307/j.ctv36zrf8

Hölig, S., & Hasebrink, U. (2020). *Reuters Institute digital news report 2020. Ergebnisse für Deutschland* [Results for Germany] (Arbeitspapiere des Hans-Bredow-Instituts, No. 50). Leibniz-Institut für Medienforschung | Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.71725

Hughes, R., & Walsh, M. (1991). *The shock of the new: Art and the century of change*. Thames and Hudson London.

Hüttel, B., Schumann, J., Mende, M., Scott, M., & Wagner, C. (2018). How consumers assess free e-services: The role of benefit-inflation and cost-deflation effects. *Journal of Service Research*, *21*(3), 267–283. https://doi.org/10.1177/1094670517746779

Jere, M., & Borain, R. (2018). Willing seller, unwilling buyer: Factors influencing intention to pay for online news in South Africa. In N. Malhotra & S. Burges (Eds.), *Proceedings of the 2018 Annual Conference of the Emerging Markets Conference Board* (pp. 97–106). University of the Witwatersrand Wits Business School.

Jiang, T., Guo, Q., Chen, S., & Yang, J. (2020). What prompts users to click on news headlines? Evidence from unobtrusive data analysis. *Aslib Journal of Information Management*, *72*(1), 49–66. https://doi.org/10.1108/AJIM-04-2019-0097

Jones, R., & Mendelson, H. (2011). Information goods vs. industrial goods: Cost structure and competition. *Management Science*, *57*(1), 164–176. https://doi.org/10.1287/mnsc.1100.1262

Kammer, A., Boeck, M., Hansen, J. V., & Hauschildt, L. J. H. (2015). The free-to-fee transition: audiences' attitudes toward paying for online news. *Journal of Media Business Studies*, *12*(2), 107–120. https://doi.org/10.1080/16522354.2015.1053345

Khajeheian, D. (2017). An introduction to entrepreneurship and innovation in media markets. *Global Media Journal–Canadian Edition*, *10*(1), 1–8.

Kohn, K., & Wewel, S. A. (2018). Skills, scope and success: An empirical look at the start-up process in creative industries in Germany. *Creativity & Innovation Management*, *27*(3), 295–318. https://doi.org/10.1111/caim.12279

Kosterich, A. (2021). Reengineering journalism: Product manager as news industry institutional entrepreneur. *Digital Journalism*. https://doi.org/10.1080/21670811.2021.1903959

Koukova, N. T., Kannan, P., & Ratchford, B. T. (2008). Product form bundling: Implications for marketing digital products. *Journal of Retailing*, *84*(2), 181–194. https://doi.org/10.1016/j.jretai.2008.04.001

Kutzner, K., Schoormann, T., & Knackstedt, R. (2018). Digital transformation in information systems research: A taxonomy-based approach to structure the field. *Proceedings of the European Conference on Information System (ECIS)*, *26*, 56.

Levy, D. A., & Nielsen, R. K. (2010). *The changing business of journalism and its implications for democracy*. Reuters Institute for the Study of Journalism.

Lin, T.-C., Hsu, J. S.-C., & Chen, H.-C. (2013). Customer willingness to pay for online music: The role of free mentality. *Journal of Electronic Commerce Research*, *14*(4), 315–333.

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*(3), 149–157. https://doi.org/10.1016/j.jsis.2015.08.002

March, J. G., & Sutton, R. I. (1997). Crossroads—Organizational performance as a dependent variable. *Organization Science*, *8*(6), 698–706. https://doi.org/10.1287/orsc.8.6.698

Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). *The digital transformation of news media and the rise of disinformation and fake news* (Digital Economy Working Paper 2018-02). Joint Research Centre. https://doi.org/10.2139/ssrn.3164170

Martínez, A., Schmuck, C., Pereverzyev, S. Jr, Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

McQuail, D. (1992). *Media performance: Mass communication and the public interest*. Sage.

Mütterlein, J., & Kunz, R. E. (2017). Innovate alone or with others? Influence of entrepreneurial orientation and alliance orientation on media business model innovation. *Journal of Media Business Studies*, *14*(3), 173–187. https://doi.org/10.1080/16522354.2018.1445162

Myllylahti, M. (2014). Newspaper Paywalls—the Hype and the Reality. *Digital Journalism*, *2*(2), 179–194. https://doi.org/10.1080/21670811.2013.813214

Myllylahti, M. (2020). Paying attention to attention: A conceptual framework for studying news reader revenue models related to platforms. *Digital Journalism*, *8*(5), 567–575. https://doi.org/10.1080/21670811.2019.1691926

Naldi, L., & Picard, R. G. (2012). "Let's start an online news site": Opportunities, resources, strategy, and formational myopia in startups. *Journal of Media Business Studies*, *9*(4), 69–97. https://doi.org/10.1080/16522354.2012.11073556

Neheli, N. B. (2018). News by numbers. *Digital Journalism*, *6*(8), 1041–1051. https://doi.org/10.1080/21670811.2018.1504626

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Reuters Institute digital news report 2017*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Reuters Institute digital news report 2018*. Reuters Insitute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf

Newman, N., Fletcher, R., Schulz, A., Andı, S., & Nielsen, R. K. (2020). *Reuters Institute digital news report 2020*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C., & Nielsen, R. K. (2021). *Reuters Institute digital news report 2021*. Reuters Institute for the Study of Journalism. https://ssrn.com/abstract=3873260

Niemand, T., Mai, R., & Kraus, S. (2019). The zero-price effect in freemium business models: The moderating effects of free mentality and price–quality inference. *Psychology & Marketing*, *36*(8), 773–790. https://doi.org/10.1002/mar.21211

Niemand, T., Tischer, S., Fritzsche, T., & Kraus, S. (2015, December 13–16). *The freemium effect: Why consumers perceive more value with free than with premium offers* [Paper presentation]. International Conference on Information Systems, Fort Worth, TX, USA.

Oechslein, O. (2014). On the search for new revenue models: An empirical investigation of personalized news aggregators. *BLED proceedings*, *2014*, 38–52. http://aisel.aisnet.org/bled2014/38

Olsen, R. K., Kalsnes, B., & Barland, J. (2021). Do small streams make a big river? Detailing the diversification of revenue streams in newspapers' transition to digital journalism businesses. *Digital Journalism*. https://doi.org/10.1080/21670811.2021.1973905

Olsen, R. K., Kammer, A., & Solvoll, M. K. (2020). Paywalls' impact on local news websites' traffic and their civic and business implications. *Journalism Studies*, *21*(2), 197–216. https://doi.org/10.1080/1461670X.2019.1633946

Overall, J., & Wise, S. (2016). The antecedents of entrepreneurial success: A mixed methods approach. *Journal of Enterprising Culture*, *24*(3), 209–241. https://doi.org/10.1142/s0218495816500096

Pattabhiramaiah, A., Sriram, S., & Manchanda, P. (2018). Paywalls: Monetizing online content. *Journal of Marketing*, *83*(2), 19–36. https://doi.org/10.1177/0022242918815163

Peskin, M., & Glick, M. (2009). *Paid content: Newspaper economic action plan*. American Press Institute. http://www.la-rem.com/wp-content/uploads/2013/11/paid-content.-US-Newspapers.pdf

Phillips, A., Singer, J. B., Vlad, T., & Becker, L. B. (2009). Implications of technological change for journalists' tasks and skills. *Journal of Media Business Studies*, *6*(1), 61–85. https://doi.org/10.1080/16522354.2009.11073479

PwC. (2018). *Global Entertainment & Media Outlook 2018–2022*. https://www.pwc.com/gx/en/entertainment-media/outlook/perspectives-from-the-global-entertainment-and-media-outlook-2018-2022.pdf

PwC. (2019). *Bevölkerungsbefragung zur Nutzung von Online-Medienangeboten* [Survey on the use of online media offerings]. https://www.pwc.de/de/technologie-medien-und-telekommunikation/pwc-befragung-nutzung-online-medienangebote.pdf

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.

Reichmann, A., & Klimmt, C. (2012). Die Refinanzierung von Online-Nachrichten: Eine Conjoint-Studie zur Zahlungsbereitschaft des Publikums [Title]. *Medienwirtschaft - Zeitschrift für Medienmanagement und Medienökonomie*, *9*(3), 11–21. https://doi.org/10.15358/1613-0669-2012-3-11

Robinson, S., Lewis, S. C., & Carlson, M. (2019). Locating the "digital" in digital journalism studies: Transformations in research. *Digital Journalism*, *7*(3), 368–377. https://doi.org/10.1080/21670811.2018.1557537

Rußell, R., Berger, B., Stich, L., Hess, T., & Spann, M. (2020). Monetizing online content: Digital paywall design and configuration. *Business & Information Systems Engineering*, *62*(3), 253–260. https://doi.org/10.1007/s12599-020-00632-5

Salaverría, R. (2019). Digital journalism: 25 years of research. Review article. *El Profesional de la Información*, *28*(1), Article e28010. https://doi.org/10.3145/epi.2019.ene.01

Shafer, J. (2016, October 17). What if the newspaper industry made a colossal mistake? *Politico Magazine*, *1*(1). https://www.politico.com/magazine/story/2016/10/newspapers-digital-first-214363

Shapiro, C., & Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business School Press.

Sraml Gonzalez, J., & Gulbrandsen, M. (2022). Innovation in established industries undergoing digital transformation: The role of collective identity and public values. *Innovation*, 24(1), 201–230. https://doi.org/10.1080/14479338.2021.1938578

Sridhar, S., & Sriram, S. (2015). Is online newspaper advertising cannibalizing print advertising? *Quantitative Marketing and Economics*, *13*(4), 283–318. https://doi.org/10.1007/s11129-015-9160-3

Statista. (2021). *Newspapers & Magazines: Worldwide*. https://www.statista.com/outlook/amo/media/newspapers-magazines/worldwide

Sylvie, G., & Gade, P. (2009). Changes in news work: Implications for newsroom managers. *Journal of Media Business Studies*, *6*(1), 113–148. https://doi.org/10.1080/16522354.2009.11073481

Teece, D. J. (2018). Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research Policy*, *47*(8), 1367–1387. https://doi.org/10.1016/j.respol.2017.01.015

Usher, N. (2017). Venture-backed news startups and the field of journalism. *Digital Journalism*, *5*(9), 1116–1133. https://doi.org/10.1080/21670811.2016.1272064

van der Wurff, R. (2011). Are news media substitutes? Gratifications, contents, and uses. *Journal of Media Economics*, *24*(3), 139–157. https://doi.org/10.1080/08997764.2011.601974

Varian, H. R. (2000). Buying, sharing and renting information goods. *The Journal of Industrial Economics*, *48*(4), 473–488. https://doi.org/10.1111/1467-6451.00133

Vermeer, S., Trilling, D., Kruikemeier, S., & de Vreese, C. (2020). Online news user journeys: The role of social media, news websites, and topics. *Digital Journalism*, *8*(9), 1114–1141. https://doi.org/10.1080/21670811.2020.1767509

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Waisbord, S. (2019). The 5Ws and 1H of digital journalism. *Digital Journalism*, *7*(3), 351–358. https://doi.org/10.1080/21670811.2018.1545592

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Wellbrock, C.-M., & Buschow, C. (2020). *Money for nothing and content for free? Paid Content, Plattformen und Zahlungsbereitschaft im digitalen Journalismus* [Money for nothing and content for free? Paid Content, Platforms and Willingness to Pay in Digital Journalism]. Nomos. https://doi.org/10.5771/9783748907251

Wellbrock, C.-M., & Klein, K. (2014). Journalistische Qualität – eine empirische Untersuchung des Konstrukts mithilfe der Concept Map Methode [Journalistic Quality - An Empirical Investigation of the Construct Using the Concept Map Method]. *Publizistik*, *59*(4), 387–410. https://doi.org/10.1007/s11616-014-0212-6

Wilczek, B. (2019). Complexity, uncertainty and change in news organizations. *International Journal on Media Management*, *21*(2), 88–129. https://doi.org/10.1080/14241277.2019.1590839

Ye, L. R., Zhang, Y. J., Nguyen, D.-D., & Chiu, J. (2004). Fee-based online services: Exploring consumers' willingness to pay. *Journal of International Information Management*, *13*(2), 133–141.

Zelizer, B. (2019). Why journalism is about more than digital technology. *Digital Journalism*, *7*(3), 343–350. https://doi.org/10.1080/21670811.2019.1571932

B. Content for Free? Drivers of Past Payment, Paying Intent and Willingness to Pay for Digital Journalism – A Systematic Literature Review

Published in: O'Brien, D., Wellbrock, C. M., & Kleer, N. (2020). Content for free? Drivers of past payment, paying intent and willingness to pay for digital journalism – A systematic literature review. *Digital Journalism*, *8*(5), 643–672. https://doi.org/10.1080/21670811.2020.1770112

C. Free Lunch for All? – A Path Analysis on Free Mentality, Paying Intent and Media Budget for Digital Journalism

Published in: O'Brien, D. (2022). Free lunch for all? – A path analysis on free mentality, paying intent and media budget for digital journalism. *Journal of Media Economics*, *34*(1), 29–61. https://doi.org/10.1080/08997764.2022.2060241

D. How the Trick is Done – Conditions of Success in Entrepreneurial Digital Journalism

Published in: O'Brien, D., & Wellbrock, C.-M. (2021). How the trick is done – Conditions of success in entrepreneurial digital journalism. *Digital Journalism* (Special Issue: The Business of Journalism, online first). https://doi.org/10.1080/21670811.2021.1987947

E. “Should I Stay or Should I Go?” – Application of ML Classifiers of Purchase and Churn Prediction to a Digital Journalistic Platform

Abstract

Considering the ubiquity of data, purchase and churn prediction have become increasingly crucial for data-driven decision-making. While many sectors have benefited from the abundance of data in the wake of digital transformation, the applications of predictive analytics remain limited in SMEs, and journalism. Thus, this study focuses on applying such techniques in the context of start-up journalism. For this purpose, article purchase and subscription data from a digital journalistic platform of over 2,700 individual customers were analysed in multiple models using the ensemble methods of random forest and gradient boosting, and logistic regression. Results indicate high accuracy scores and imply the applicability of such models in journalism, especially SMEs with limited databases. This paper demonstrates how to reduce classification thresholds in models to improve sensitivity.

1. Introduction

Automated driving, protein folding, social media, manufacturing, computer-aided design – machine learning (ML) has affected most areas of society profoundly (Berisha-Shaqiri, 2014; Ismagilova et al., 2019; Tayebi et al., 2019).

The described development holds especially true for the business world. In the economic context, digital transformation defines the transfer of formerly analogue areas of economic life into digital contexts (Berman, 2012; Vial, 2019). Entire business areas and sectors have undergone fundamental structural change, transforming the business models or strategic orientation of entire industries (Hess et al., 2016), from mining (Young & Rogers, 2019) to telecommunications and media (Ferrer-Conill & Tandoc, 2018). The all-encompassing and advancing digitalisation of planning, production and organisation of business has comprehensively determined the behaviour of companies and consumers in recent decades (Gobble, 2018; Reis et al., 2018; Ziyadin et al., 2019).

The flood of data accompanying digital transformation enables companies to understand customers better and predict behaviour with increasing accuracy. The term *predictive analytics* in particular covers various techniques for classifying and predicting future behaviour. Manufacturers (Hanelt et al., 2015), producers of consumer goods (Hamdani & Daim, 2020), actors in the global supply chain (Bowersox et al., 2005) and in the business-to-business (B2B) sector (Martínez et al., 2020) make use of such techniques. As Vial (2019) highlights, the accessibility of data, and with it the possibility of comprehensive analysis, are paramount to drive strategic responses and organisational alignment in the context of digital transformation. For instance, the ubiquity of metrics and business analytics, key performance indicators and algorithmic optimisation of production or product development play a crucial role for large and for small and medium-sized enterprises (SMEs).

Following this development, purchase and churn forecasting became ubiquitous in both practical applications and theoretical business studies (Bi et al., 2016). The importance of accurate analysis of existing customer data for strategic customer relationship management (CRM) and marketing is paramount (Martínez et al., 2020; Wixom et al., 2013). Research on retaining customers over acquiring new ones emphasises this, as it costs up to ten times less (Ahmad et al., 2019). Predictive analytics research is relatively advanced in established industries and sectors (Bradlow et al., 2017; Gunasekaran et al., 2017; Martínez et al., 2020), especially where there are significant amounts of data, such as in retail or telecommunications (Ahmad et al., 2019; Brânduşoiu et al., 2016; He et al., 2009). However, for SMEs and start-ups – often characterised by technical limitations and limited data – potentially helpful or even necessary techniques and applications remain under-researched (Howden, 2021; Kannan et al., 2021; Llave, 2017). Nevertheless, research shows that even with limited data, accurate predictions can be derived (Howden, 2021; Maroufkhani et al., 2020; Megahed et al., 2015). In real-world applications, practitioners must frequently rely on small data sets (Li & Liu, 2012). Nevertheless, predictive analytics have proven valuable across a range of enterprises from a corporate wellness programme provider and individual convenience stores (Kannan et al., 2021) to security systems in small businesses (Salih & Njenga, 2019).

The applicability of such techniques is promising for journalism. While digital transformation is opening new business areas and customer bases in many industries, the economic situation of journalism is often considered bleak (Berger et al., 2015; Chyi & Ng, 2020; Kammer et al., 2015; Lischka, 2019). Some authors warn not only of the cannibalisation of journalistic profits as the tech giants displace the classical advertising-based business model but also of a general devaluation of quality journalism in competition with the abundance of information goods (Brettel et al., 2015; Donders et al., 2018; Myllylahti, 2020; Sridhar &

Sriram, 2015). Therefore, while much of the world’s population consumes online news, only a fraction reports paying for such services (Chyi, 2005, 2012; Newman et al., 2018).

Essentially limited to distribution over the Internet, the digitalisation of journalism began in the early 1990s. Lischka (2019) describes the digital transformation of the news business along two lines: the digitization of journalism itself, and diversification into digital markets that are not news-related. However, all areas of contemporary journalism are now shaped by this development (Salaverría, 2019), from content creation, editing, production and dissemination to consumer analysis (Friedrichsen et al., 2017).

Concerns about the viability of journalism remain, including the consequences of the degeneration of the *fourth estate* (Anderson & Rainie, 2020; McQuail, 1992). These concerns are growing in the face of the spread of fake news and the increasing distrust of the press. (Park et al., 2020). Practitioners and scholars in the field are looking for meaningful solutions, ranging from paid content to contribution-based models and state-funded media (Rußell et al., 2020). Meanwhile, traditional journalism is torn between monetising itself and conveying the truth (Lischka, 2019).

However, where there are challenges there are opportunities. The digital world offers new ways for journalistic companies to interact with, retain and expand their customer bases. Analytics has enormous potential for the news business, although it is not yet fully understood. The monetisation crisis in digital journalism (Berger et al., 2015; Casero-Ripollés, 2012; Chyi & Lee, 2013) highlights the increasing importance of sophisticated approaches to analyse and exploit customer data. Advances in data science should help identify, understand and attract paying clientele. For example, it is possible to identify customers who are willing to pay and address cancellers in good time so that they can be retained as customers through individualised customer management. Regarding the management of existing or future customers and the optimisation of journalistic offerings, there are approaches for digital journalism to profit in the *brave new world* of ubiquitous information.

This study examines the extent to which predictive analytics is a reasonable means for SMEs to refine customer management. Beyond, we argue for the relevance of these techniques for the business of journalism, as it progresses into the digital age. Consequently, our research question is whether predictive analytics is a useful means in the context of SMEs with a particular focus on combating journalism's funding difficulties, whether purchase and churn can be reliably predicted, and which factors are most relevant for this prediction. We contribute to this debate in the literature by expanding journalism's views on the ever-growing field of predictive analytics and providing a meaningful use case. Prediction of purchase and subscription decisions based on anonymised customer data will play an increasingly important role in various industries. Given the funding crisis in print journalism on the one hand and the financial situation of digital journalism on the other, insights into customer buying or churn behaviour are crucial for both small and larger companies in the media sector and strengthen its broader role in society. Therefore, we pursue two goals with our work: first, to employ predictive analytics and extend its analytical power to the field of journalism, and second, to contribute to the analysis of relatively small data sets, especially in a start-up setting with a naturally limited database. This work's limitations involve problems typically associated with smaller data sets and unbalanced data, to which we propose concrete solutions. Thus, we demonstrate refined applications of classical prediction algorithms for small data sets.

This exploratory work is one of only a few to empirically examine how journalism can benefit from digital transformation by testing the application of predictive ML methods to real-world data. We use historical data from customers who pay for individual articles and customers who subscribe to predict which customers will pay for articles in the future and which customers will keep their subscriptions.

For this purpose, we analysed multiple data sets from a German journalism platform dealing with science, society and the environment. Data from over 2,700 individual customers were analysed in several models using both the ensemble method of random forest classifiers and gradient boosting, as well as logistic regression. The investigation of this digital journalistic platform promises insights into strategies for successfully monetising high-quality online journalism in a start-up environment.

2. Literature Review

2.1. Purchase and Churn Prediction

Customer purchase prediction and churn prediction, both essential instruments of CRM, have received intense scholarly attention (Fader & Hardie, 2009; Saarijärvi et al., 2013; Seddon et al., 2017). Intending to increase profitability and predictability, computer-assisted prediction of individual purchases, subscriptions and cancellation decisions have attracted some research and business scrutiny (Huang et al., 2015; Khodabandehlou & Rahman, 2017; J. Lee et al., 2021; Martínez et al., 2020). In retail, purchase predictions enable new ways to estimate the need for inventory or to increase production. In subscription-based businesses, the direct effect of churn is generally high as profits derive linearly from the number of subscribers and profit from network effects (Ahmad et al., 2019). Moreover, in the context of information goods, the marginal costs are often close to zero. Thus, while profit maximisation ordinarily stems from the output, in which marginal cost equals price, companies selling information goods must maximise their user base regardless. Therefore, identifying potential churners is crucial because customers who are potentially willing to cancel a service can be specifically addressed during the critical phase of their deliberations.

Customer decision prediction by ML is based on the assumption that large data sets generated by a business are indicative of underlying patterns to enable the classification and prediction of a discrete variable (Fader & Hardie, 2009; Martínez et al., 2020). Lasso regression, gradient boosting and random decision tree are prominent ML methods with a strong track record for binary classification problems (Ahmad et al., 2019; Martínez et al., 2020; Sabbeh, 2018). Lasso regression is a variation of standard regression techniques in which the inclusion of additional variables is discouraged to avoid overfitting. While frequently used as baseline models, those traditional regression techniques often perform worse than ensemble methods, which learn from any number of weak sub-models. Gradient boosting and random forests are ensemble methods in which the results of the prediction are averaged over multiple random splits of the data, thereby avoiding the overfitting of traditional decision tree models. For example, in predicting the financial performance of movies, Lash and Zhao (2016) demonstrated that ensemble methods combined with data about actors and genre vastly outperform regression models.

Martínez et al. (2020) stress that the cost of acquiring a new customer is up to one order of magnitude higher compared to keeping an existing one. Via lasso regression, gradient tree boosting and extreme learning machine models, the authors successfully classify customers in a non-contractual purchase context (B2B). Their predictions achieve accuracies of up to 88% for customers who purchase in the target period, and an area under the curve (AUC) of 0.949 via gradient tree boosting.

Likewise, concerning churn prediction, Ahmad et al. (2019) argue that fierce competition in the digital telecommunication sector emphasises the importance of identifying loyal customers. They used gradient boosting, random forest, and traditional decision tree models to predict churn with an AUC up to 93% for the gradient boosting model. In a similar study, Huang et al. (2015) achieved comparable results in the same industry with a random forest classification model.

Using data from a telecom provider, Sabbeh (2018) attested in her comparative analysis of different classifiers the best predictive power in churn rate to the random forest and adaptive boosting tree models. The latter is similar to gradient boosting but uses an exponential loss function and trains upon misclassified observations. Other approaches to identify customers’ willingness to buy or churn use clustering algorithms (Bi et al., 2016) or variations of logistic and linear regression (Ozan, 2018), which often serve as a baseline model.

2.2. *Predictions With Small Sample Sizes*

Big data covers many applications and gives the impression that it only makes sense with huge samples and in the domain of large, mature companies. Most of the literature regarding ML applications in classification problems is based on big data (for instance retail or telecommunications). This paper explicitly addresses a branch of ML implementations for small data sets, especially in the context of SMEs and start-ups, a context which to date has been explored only to a limited extent. In the literature, the repeated use of decision trees/random forests and gradient boosting methods in the context of small data sets stand out (Breiman, 2001; Han et al., 2021; Kannan et al., 2021; Olson et al., 2018).

Here, two research streams are of particular interest. Although few studies exist that apply predictive analytics to small data sets, or sparse data from small businesses, there are helpful case studies from other disciplines with limitations of accessible data or that are intrinsic to the logic of the discipline.

Kannan et al. (2021) addressed the question of how SMEs can benefit from ML applications, arguing that they give companies a competitive advantage and a better database for future decisions. In one case study concerning a corporate wellness programme provider for a bank, ML applications used decision trees and support vector machines to predict visceral fat levels (accuracy of about 0.85/0.89) for 1,079 participants in the study. This gave bank employees important information about their lifestyle and helped the employer plan for potential health hazards. Priyam et al. (2013) worked with different decision tree models to predict student performance in final exams, emphasising the appropriateness of decision tree models for small data sets.

Dittert et al. (2017) established a research framework for SMEs in which they emphasise the following steps: task definition; data collection and analysis; model selection; data formatting; evaluation of the results; and, depending on the acceptability of the results, either the iteration of the process starting with model selection or the report to decision-makers. Their case study used the sales data of a medium-sized petroleum retailer in Southern Germany to classify the sales potential (high/low) of retail sites (gas stations) via an artificial neural net. The best model reached an accuracy of 86.67%. While no sample size was given for the case study, they emphasise the small given database. Finally, W. J. Lee and Ong (2010) argued for using both random forests and gradient boosting in prediction tasks. Their study proposes a prediction model based on gradient boosting to determine the optimal recipe in a specific chip production procedure. They vary the sample size (100, 150, 300, 600, 2023) for three different production scenarios and obtain an accuracy of 64–84% even for the smallest sample sizes, with the highest accuracy of 99%.

Beyond the economic or manufacturing context, some chemical, medical and bioinformatics research is of interest in the context of prediction based on small sample sizes. In one study from the polymer industry, which examined 39 additive candidates to mitigate the degradation of polythene, random forest models achieved AUCs as high as 0.925 with the best feature setting (Liu et al., 2020). Especially in medicine, where only limited case numbers are often available for different classification tasks (for instance: patient sick or healthy, tumour benign or malignant), the techniques presented are important. A study by Han et al. (2021) used random forests in the context of small case numbers in vaccination trials concerning HIV infection risks with only 150 observations. Despite the small number of cases, they obtain an AUC of up to 0.831 through a combination of variable screening, class balancing, weighting and hyperparameter tuning. They emphasise the application of random forests, as this method is fast in model training and evaluation, robust to outliers, captures nonlinear relationships, can cope with class imbalances and has been shown to work well for small sample sizes. Olson et al. (2018) second this view and describe that "[t]he mechanism by which a random forest can generalise well on small data sets is straightforward: a random forest is an ensemble of lowbias, decorrelated trees. Randomisation combined with averaging reduces the ensemble’s variance, smoothing out the predictions from fully grown trees” (p. 3624).

Along with the realisation that more data and more complex models do not always produce better results, there is a growing trend in the literature to move away from these trends. Green and Amstrong (2015) argue in their research that scientists are rewarded for complex models that are not necessarily related to the topic being researched. Katsikopoulos et al. (2022) second this view, showing that simple heuristics and psychological models are superior to overcomplicated black-box models in many cases.

2.3. Predictive Analytics and Digital Transformation of Journalism

We now turn to the concrete application of the aforementioned techniques to journalism, where there is a research gap (Ferrer-Conill & Tandoc, 2018; Neheli, 2018). This is crucial given that journalism has social relevance but seems to be one of the biggest sufferers of digitalisation. According to Donders et al. (2018), digitisation, internationalisation, and pressure on business models are the main challenges of the media industry today. Consequently, digital journalism is increasingly understood as a significant target for business analytics and data science applications (De Grove et al., 2020; Jiang et al., 2020; Neheli, 2018), whether that be computer-assisted newsrooms (Ferrer-Conill & Tandoc, 2018; Tenenboim & Cohen, 2015; Zamith, 2018), AI-generated headlines (Gu et al., 2020), sentiment analysis of customer feedback data (Grljević & Bošnjak, 2018) or the application of algorithms to identify and deliver content that fits with the preference of the audience (Kanuri et al., 2014; Park et al., 2020). Digital technologies in journalism enable new revenue streams and new practices (Maijanen et al., 2019). In particular, Chan-Olmstedt (2019) highlights the challenge of finding a balance between the human and AI-based approaches to managing a newsroom. She identifies eight application areas relevant to the use of algorithmic methods in news media: audience content recommendations/discovery, audience engagement, augmented audience experience, message optimisation, content management, content creation, audience insights, and operational automation.

However, in a business context, implementation is often slow because of a disconnect between the analytics department and the rest of the organisation (Vidgen et al., 2017), and journalists themselves are slow to adopt such methods (Usher, 2017). Wilczek (2019) assumes that the avoidance of complexity and uncertainty is at the root of the inhibition of news organisations to carry out necessary transformations. As journalism was a profiting business with an established market position and audience for a long time, it is still slow to adapt to technological innovation (Sylvie & Gade, 2009). In this development, media companies must become more outwards oriented and further relations with audiences to cope with a new business environment. Many editors define engagement in terms of certain analytics, but they must also negotiate with the journalists what the results of those analytics mean (Ferrer-Conill & Tandoc, 2018). Some researchers argue against the dangers of a metricised newsroom (Zamith, 2018), while also underlining the importance of increasing predictive analytics and quantified newsrooms. Other scholars state that journalism can only gain economically and journalistically by increasing its use of software and IS (Bakke et al., 2020). In this view, journalistic entrepreneurs need to show maximum interest in the analytics of news, which has an impact on news content, journalistic values and editorial decision-making (Malcorps, 2019). Thus, algorithms and audiences act as a kind of additional gatekeeper. Diakopoulos (2016) emphasises that the news industry needs to improve its use of information technology to gain better insights into its customers and what they want and, thus, develop more effective business models. Based on 21 qualitative interviews with journalistic companies, Lamot and Paulussen (2020) argue for six relevant uses of advanced analytics in the newsroom: story placement, story packaging, story planning, story imitation, performance evaluation, and audience conception.

Speaking more generally about the use of predictive analytics in media, Kukkonen (2018) cautions that accurate knowledge and analysis of a customer’s past purchase history and click behaviour brings excellent benefits in evaluating further business opportunities, which holds especially true for strategic CRM. Besides any moral or technical aspects, others warn against the economically and substantively useless application of supervised ML methods. De Grove et al. (2020) outline the financial and time commitments needed to implement working predictive models, emphasising that they must be maintained and improved to benefit maximally from ML methods. The authors also claim that analysing smaller, more carefully constructed data sets is often more useful to journalism in identifying essential customer segments. Neheli (2018) cautions that analytics can lead to blind commodification of the audience although, if used carefully, they can consolidate customer relationships.

Overall, the literature dealing with predictive analytics seems to focus on large data sets in traditional industries (see Table E.1), while research emphasising the role of such technologies in journalism is essentially qualitative and focuses on the role of the editor. As journalists and the wider industry slowly begin to embrace digital monetisation and sophisticated CRM, we aim to address the specific problem of customer data analytics for smaller data sets.

Table E.1: Literature Review on Prediction Methods/Predictive Analytics

Reference	Sample			Method		Behavioural of Users		Prediction		
	Size	Product	Small data/ Start-up	Classi-fication	Multiple algo-rithms	Usage of pro-duct	Time spent on pro-duct	Pur-chase	Churn	Accuracy/ (AUC)
Ahmad et al., 2019	5,000,000 customer records	Telco	✗	✓	✓	✗	✗	✗	✓	93% (AUC)
Bi et al., 2015	1,800,000 subscribers	Telco	✗	✓	✓	✓	✓	✗	✓	95%
Dittert et al., 2017	70	Middle-sized oil-trader	✓	✓	✓	✓	✗	✓	✗	86%
Fernandes et al., 2015	39,000 news articles	Online news	✗	✓	✓	✗	✗	✗	✗	73% (AUC)
Han et al., 2021	150 patients	HIV infection risk	✓	✓	✗	✗	✗	✗	✗	83% (AUC)
Jääskeläinen et al., 2020	not made public in the paper	Online news	✗	✗	✗	✗	✗	✗	✗	75%
Jiang et al., 2020	98,016 clicks received by 7,120 users	Online news	✗	✗	✗	✗	✗	✗	✗	N.A.
Kannan et al., 2021	1,079 clients (bank workers)	Corporate wellness program provider	✓	✓	✓	✗	✗	✗	✗	85%/89%
Lash & Zhao, 2016	2,506 movies	Movies	✗	✓	✓	✗	✗	✓	✗	83%
W. J. Lee & Ong, 2010	100, 150, 300, 600, 2023	Chip manufacturing	✓	✓	✗	✗	✗	✗	✗	Up to 99%
Liu et al., 2020	39	Polyethylene stabilizer	✓	✓	✗	✗	✗	✗	✗	92% (AUC)
Sabbeh, 2018	3,333 customers	Telco	✗	✓	✓	✓	✓	✗	✓	96%
This study	~69,000 logins, ~13,000 transactions, ~4,300 customers	Online news	✓	✓	✓	✓	✓	✓	✓	96%/85%

3. Method

The platform studied in this paper acts as a content management and publishing platform for individual freelance journalists. The decision to work with this specific platform was based first, on the specific design of the platform’s business model and second, on its focus on high-quality, independent science and society journalism, meaning the communication and classification of scientific findings to broader society.

The company enables freelance journalists to monetise their content by paying only a fraction of their profit to the platform operator. The platform offers an integrated content management system. The participating journalists decide on the price of their articles or their subscription. The customer can decide between purchasing individual articles and subscribing to all articles of specific journalists. Furthermore, a bundled option is available, where users pay a monthly subscription fee for access to all articles. Bundling that is combining different goods into a mixed bundle is often described as a viable and profitable strategy for information goods in general (Bakos & Brynjolfsson, 1999; Wu et al., 2008) and online news in particular (Bakos & Brynjolfsson, 2000; Kammer et al., 2015; Koukova et al., 2008). Platforms and their specific economics are of increasing interest. Facebook, Airbnb, Uber, etc., dominate the new market by exploiting network effects and first-mover advantages. However, smaller entities such as the company studied here can also benefit significantly from a platform model. We define a platform as an intermediary for two or more parties to exchange goods and services or, following Rysman (2009): “1) two sets of agents interact through an intermediary or platform, and 2) the decisions of each set of agents affects the outcomes of the other set of agents, typically through an externality” (p. 125).

The data sets provided by the examined platform comprised five large data dumps, which we used to construct all features in the final data sets for both learning tasks. The dumps comprised one extensive login data set containing individual visits with over 2.8 million entries. Additionally, three distinct data sets are concerned with payment data, with the first one listing individual purchases of articles (*n* = 4,274), the second one subscriptions to individual authors (*n* = 1,016), and the third one subscriptions to the platform (i.e. the bundled version, *n* = 821). Finally, we included user data containing 4,296 registered users from 23 countries. The study period ranges from December 2016 to the end of March 2021. In line with the company’s business model, the data regarding individual purchases and subscriptions were dealt with separately.

The seminal works of Martínez et al. (2020) and Sabbeh (2018) helped significantly in the design and execution of our analysis. Additionally, we followed the mentioned framework of Dittert et al. (2017), comprising task definition, data collection and analysis, model selection, data formatting, evaluation of the results, and, depending on the results, the iteration of the process or report. In the data engineering process, we focused on non-personal data to work on fully anonymised data sets. Following the literature (Müller et al., 2016; Rußell et al., 2020), those data sets comprised mostly anonymous login and payment data, which were used to extract and engineer the relevant data set for the final models. The user ID was the one-to-one identifier for users across the different data dumps. Therefore, in the first step of data cleaning, we removed all login data that was not associated with a user ID, leaving us with 69,308 entries in the login data set. While the total amount of data is far from the scale usually associated with big data, we are in the broader range on the border between small and midsized data (Axelrod, 2016). Compared to several studies mentioned in the literature review, the data is not particularly small.

The data sets for the individual purchases and subscriptions to authors were cleaned from missing values. Additionally, data points later than 31 December 2020 were excluded for the training of the prediction model as they fall in the target period of the prediction. Therefore, those data entries were only used to determine whether a user purchased or churned up to Q1 2021.

The analysis was conducted with Python 3.8, explicitly with the scikit-learn library (Pedregosa, 2011). As the initial data sources were comprehensive SQL dumps provided by the examined company, we first extracted individual tables with the features of the customers, the purchases of single articles and payments for subscriptions. From those basic tables, all additional features were extracted or engineered. Categorical data was either label encoded (e.g. country or operating system) or encoded binary as 1 and 0 (e.g. voluntary contribution). The target variable *purchase* was encoded as 1 if a user made a purchase in the target period and 0 otherwise. The target variable *churn* was encoded as 1 if a subscription was cancelled and 0 otherwise (for descriptive statistics see Appendix E.1 and Appendix E.2, for correlations Appendix E.3 and Appendix E.4).

We split the data sets into a training and a test data set, on which we then performed and tested the predictions against reality. Train/test splits usually overweight the training class to provide enough variation in data for the model to generalise sufficiently (Martínez et al., 2020). Typical splits range between 60/40 to 90/10 for the training/test set (Sabbeh, 2018). We determined the train/test split by repeated application of the model with different splits from 60/40 to 90/10. Following the results, we used a 75/25 split for both models. The splits were stratified concerning the outcome variable.

Following the insights of the literature review, which demonstrated gradient boosting and random forest to be highly robust ensemble methods and frequently used regarding small data, both were applied to the finalised data sets for the purchase and churn prediction tasks. Additionally, we used logistic lasso regression as a baseline to demonstrate performance in the absence of advanced ML methods.

In prediction tasks with imbalanced data sets, where one of the possible outcomes happens far less frequently than the other, performance can be enhanced (in the case of the purchase prediction data, only ~21% purchased in Q1 2021). To increase the sensitivity for the minority class, the literature recommends resampling methods for unbalanced data sets (Huang et al., 2015; Rivera & Xanthopoulos, 2016; Sabbeh, 2018). Therefore, we oversampled the minority class in the training phase.

On the finalised data sets, a lasso regression, a random forest classifier and a gradient boosting classifier were applied to predict, first, purchase decision, and second, churn decision for Q1 2021. The categorical features of the model were label-encoded. The metric features were normalised.

The hyperparameters were derived via random grid search. They concerned the number of estimators, maximum number of features, maximum depth, minimum sample split and minimum sample leaf size for both algorithms. We calculated accuracy scores as the relative number of correctly identified cases to evaluate the performance of the model. Moreover, we constructed confusion matrices and visualised the results with a receiver operating characteristic (ROC) curve.

For the purchase prediction, the final data set contained 2,070 individual customers who purchased articles on the platform anytime between December 2016 and December 2020 to predict their purchase in Q1 2021. Table E.2 shows the data derived from the data set or by feature engineering, which are used for the purchase prediction.

Table E.2: Variables for the prediction of individual purchases

Variable name	Description	Data type
sum_price	Sum of price of all purchases per customer in Cent	numerical
num_purchases	Number of all purchases per customer	numerical
active_months	Number of months since a user is active	numerical
num_logins	Number of logins to the platform	numerical
actions_per_visit	Average number of actions per visit	numerical
time_per_visit	Average time spent on the platform per visit	numerical
config_os	Operating System of the customer	categorical
country	Country of customer	categorical
subscribed_channel	1 if user has at least one active author subscription, else 0	categorical (dichotomous)
subscribed_flat	1 if user has subscribed to the flat for the platform (bundle), else 0	categorical (dichotomous)
purchases_per_month	Number of purchases per month active on the platform	numerical
purchased	Binary target variable (1 if customer purchased ≥ 1.1.2021, else 0)	categorical (dichotomous)

Furthermore, the customer data concerning 703 subscriptions to journalists was analysed regarding customer churn (see Table E.3).

Table E.3: Variables for the prediction of subscription churn

Variable name	Description	Data type
price_sub	Price of each individual subscription in Cent	numerical
voluntary_sub	1, if customer included a voluntary payment	categorical (dichotomous)
channel_id_sub	ID of the journalist's channel	numerical
country_sub	Country of customer	categorical
sum_price_sub	Sum of all payments of a customer who pays for a specific subscription	numerical
num_payments_sub	Number of all payments per customer	numerical
active_since_sub	Number of days since a subscription is active	numerical
num_logins_sub	Number of logins to the platform	numerical
actions_per_visit_sub	Average number of actions per visit	numerical
time_per_visit	Average time spent on the platform per visit	numerical
config_os	Operating System of the customer	categorical
churn_sub	Binary target variable (1 if customer churned ≥ 1.1.2021, else 0)	categorical (dichotomous)

4. Results

For both data sets, we created an initial cursory baseline model with the lasso regression. In the second step, we added a random tree and gradient boosting model (1st model: purchase prediction, 2nd model: churn prediction). The first results of our analysis show relatively high values regarding the accuracy scores for the ensemble models (between 0.84 and 0.96). Table E.4 breaks down the confusion matrix for the test data sets in both purchase and churn prediction for random tree and gradient boosting (Martínez et al., 2020). The diagonal from the upper left to the lower right contains the correctly classified cases (both positive or negative), while the diagonal from the lower left to the upper right indicates falsely classified positives or negatives. Table E.5 includes the feature importance of the individually created model variables.

Table E.4: Confusion matrices for purchase prediction and churn prediction

	Purchase Model				**Churn Model**			
	Random forest		**Gradient boost**		**Random forest**		**Gradient boost**	
Accuracy		0.944		0.963		0.841		0.847
Precision		0.882		0.990		0.818		0.852
Recall		0.858		0.841		0.771		0.743
Classification	***Reality***				***Reality***			
Prediction	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Positive	97	13	95	1	54	12	52	9
Negative	16	392	18	404	16	94	18	97

Table E.5: Feature importance of the individual variables for overall prediction (in %)

Purchase			Churn		
	Random forest	**Gradient boost**		**Random forest**	**Gradient boost**
sum_price	4.73	0.39	**sum_price_sub**	15.35	15.37
num_purchases	2.68	0.16	**num_payments_sub**	16.32	24.10
num_logins	5.20	0.76	**num_logins_sub**	8.89	5.84
actions_per_visit	4.49	1.16	**actions_per_visit_sub**	8.18	3.33
time_per_visit	4.57	1.23	**time_per_visit**	9.28	5.96
config_os	1.66	0.22	**config_os**	9.32	10.05
country	0.51	0.12	**country_sub**	1.88	1.56
active_months	56.19	95.48			
subscribed_channel	4.39	0.05			
subscribed_flat	0.15	0.06			
purchases_per_month	15.43	0.48			
			price_sub	4.55	2.64
			voluntary_sub	1.10	0.02
			channel_id_sub	4.38	3.85
			active_days_sub	20.74	27.29

For a clearer picture of the effectiveness of each prediction, Figure E.1 and Figure E.2 show the ROC curves of the purchase and the churn prediction respectively. The ROC curve depicts the quality of the binary classification compared to chance, helping to identify optimal models with certain constraints (Sinha & May, 2004).

Figure E.1: Area under curve (AUC) purchase prediction

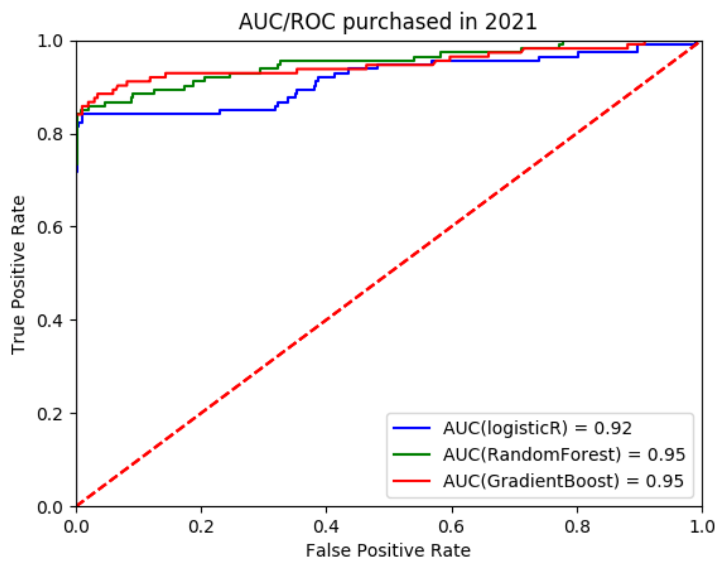
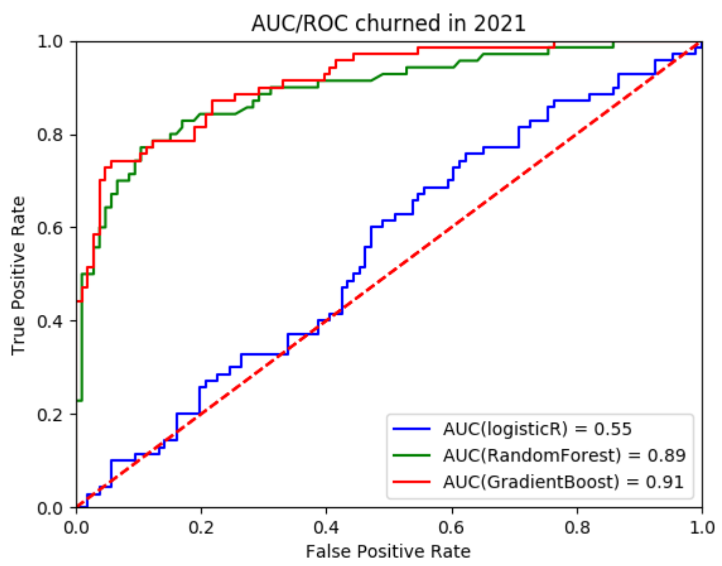


Figure E.2: Area under curve (AUC) churn prediction



5. Discussion

A comparison between the two ensemble models and the logistic lasso regression shows a substantial difference between the purchase and the churn model. While the purchase ensemble models perform only marginally better than the logistic regression, the logistic regression performs rather poorly compared to the random tree and gradient boosting algorithms in the case of the churn model. The achieved prediction accuracy was comparable to the results in similarly small or even larger sample sizes. Regarding the feature importance, for the purchase model, the most important feature by far is recency (described by the variable ‘active_months’), which suffices for a relatively good performance of the logistic regression. For the churn model, a handful of features contribute to the accuracy of the classifier (depending on the chosen ensemble method). In the latter case, the logistic regression performs significantly worse for the AUC/ROC curves and the remaining performance features.

Examination of the accuracy of the individual methods demonstrates that ML prediction has great potential for evaluating customer data for the digital journalism industry in a targeted and economical way. Given digital journalism’s constant struggle to achieve or maintain demand and good monetisation, predictive analytics are important tools to identify customers who are willing to buy or at risk of churn. While these methods have existed for some time, it is only the relatively recent and comprehensive digital transformation of the journalistic business that makes it possible to analyse this data economically; these techniques can ensure that individual users are managed optimally and profitably in real-time or can be engaged through CRM (e.g. through advertising and promotion).

However, closer examination of the confusion matrix shows a significant caveat in the quality of the purchase and churn prediction, which generally holds for the accurate evaluation of prediction tasks. The accuracy of both models is high and good-to-excellent compared with studies identified in the literature review, where accuracy often ranged lower. Nevertheless, identifying the actual minority cases (i.e. customers willing to purchase/willing to churn) is weaker. That is, while the precision and recall of both models are still high (between 0.75 and 0.99), the identification of the crucial customers is weaker than the overall model. This is often true in predictive ML models (Hanczar & Nadif, 2019; Juba & Le, 2019; Winkler et al., 2019). As a result, in the weakest model, ~25% of the crucial customers, who would be churners, would not be recognised by the model. The recall of the purchase prediction is therefore lower. Still, the results compare well against the state of the literature, not only in small data sets but even for large applications, as accuracy and recall are comparable.

Importantly, precision and recall are in a trade-off relationship, where one of the two values can be optimised at the expense of the other. Since identifying paying customers and churners is crucial in our classification task, we have tried such an optimisation by reducing the classification threshold (Buckland & Gey, 1994). In cases where the reliable identification of the target class (recall) is more important than the correct classification of the non-positive cases (precision), this trade-off may well be acceptable (a typical case is the detection of malignant tumours). The costs of misclassification as a potential buyer or churner are relatively low in our context. For instance, since it is a matter of addressing such cases via correct marketing or individualised messages, the damage caused by such false positives is manageable. Therefore, in our case, it is potentially desirable to increase the recall at the expense of precision. This procedure gave us a recall of 0.9 for the random forest and 0.78 for gradient boosting (see Table E.6). The recall values of the random forest are higher than the gradient boosting because the classifications of the latter are much more decisive (most cases are assigned to the target class with a value close to 0 or 1), which means that the shift in the threshold causes only minor changes. Therefore, the random forest method seems to be more adjustable in our situation.

Table E.6: Churn Model optimised for recall with random forest (threshold: ≤ 0.35) & gradient boosting (threshold: ≤ 0.01)

	Random forest		**Gradient boost**	
Accuracy	0.772		0.818	
Precision	0.656		0.764	
Recall	0.900		0.786	
Classification	***Reality***			
Prediction	Positive	Negative	Positive	Negative
Positive	63	33	55	17
Negative	7	73	15	89

Another interesting point concerns the features deemed to be necessary by the model. The purchase model is mainly informed by how long a customer is active on the platform. This recency effect is common in both non-contractual information goods and contractual settings with short notice periods. While the active months are the most important feature for churn prediction, several other variables have a non-negligible effect and account for much of the explanatory performance of the model. Behavioural data (duration and number of logins, number of actions) and the operating system used have an essential role in this prediction task. These variables, which can be interpreted as a proxy for the way people interact with the platform, make real-time analyses of customer data promising and economically feasible for journalistic platforms.

As the results of our research illustrate, even small entrepreneurial entities can benefit from applying algorithmic customer behaviour prediction. The results are stable – especially for churn prediction, despite the limited number of users and expandable database – and promising concerning the application to additional data sets.

The present study, which focuses on a small media business with an equally small database, was able to show that qualitative predictions can be based on limited data. The achieved prediction accuracy was comparable to the results in similarly small or even larger sample sizes. Factors that characterised the recent activities and the interaction of the users with the platform were of particular importance.

5.1. Implications

This research has both practical and theoretical implications for the field of media management, particularly customer relationship management of media companies. As the results of our research illustrate, even small entrepreneurial entities can benefit from applying algorithmic customer behaviour prediction. In practice, first, the recency effect seems to be decisive. Therefore, keeping customers on the platform and retaining them as paying customers is a significant difficulty, as most customers only stay for a few actions and purchases (J. B. Kim, 2013). Several free articles or more free time on the platform might be possible attractors. Thus, customer uncertainty concerning the product can be reduced (Dawson et al., 2010). In addition to this screening opportunity, the company could signal its added value by highlighting prizes won (Wellbrock & Wolfram, 2021).

Second, the platform’s advantages should be emphasised more. For example, the possibility of receiving a bundled product from several individual journalists and the convenience of the unified CMS are attractors that should be marketed (Bakos & Brynjolfsson, 2000; Kammer et al., 2015). Vukanovic (2009) emphasises the strategic advantage of bundling as a means to soften competition and take advantage of economies of scale. Beyond, Putzke et al. (2010) assert with view on online news that perceived ease to use is a strong driver of paying intent. The examined platform should therefore stress the advantages of its platform bundle. Literature also emphasises the strength of niche journalism, if the company’s focus is to achieve high competence and proficiency in one area, as is the case with science journalism (Cook & Sirkkunen, 2015). Gundlach and Hofmann (2021) distinguish different kinds of users of journalistic content and identify about one-third of the audience as “intellectual-journalistic consumers" with an orientation towards cognitive information and higher education. This is the target group for a journalistic platform like the one examined here. If the content of such an offering is focused, has sufficient depth, and includes information of high quality and scarcity, small but strongly motivated audiences can be highly profitable.

Third, the intensity of interaction with the website is a relevant variable, so the platform should focus on more robust user engagement. This could be done through appropriate comment functions or networking via social media (D. H. Kim & Desai, 2021). While some authors caution against the attention trap on social media (Myllylahti, 2018), others hint at the critical branding and marketing power of social media, especially for non-legacy digital natives and start-ups (Kosterich & Weber, 2019). Furthermore, research shows that news is more frequently shared via social media (Khuntia et al., 2016). At the same time, research on the German news market underlines the importance of direct media search and Google as access points (Myllylahti, 2018). Thus, brand recognition and proper marketing are of direct importance to a journalistic brand's success (Klaß, 2021). As journalistic brands and star appeal of individual journalists becomes increasingly important, online news offerings, like the one examined here, have to orient themselves towards traditional marketing to attract customers, as many readers do seek out news for the particular brand or journalist associated. This development is underlined by the recent success of individual journalists on platforms such as Substack.

The theoretical implications relate to expanding the application of predictive analytics to better sustain and manage digital transformation, even in small business contexts. As the media management literature asserts, the proliferation of such techniques to SMEs and start-ups lessens the competitive advantage of large companies and economies of scale (Buschow, 2020; Chan-Olmsted, 2019). Thus, mass personalization and niche products are now viable business options for these companies, which enables them to compete with scaled businesses.

Furthermore, we demonstrated technical aspects of the implementation of such techniques. As such, we have shown how applying balancing related to the underrepresented class and anonymised feature engineering can make potent predictions even with small data sets. Specifically, correct hyper-parametrisation and adjustment of the thresholds for classification together have non-negligible effects on correct identification in cases where recall is potentially more important than precision.

5.2. Limitations and Avenues for Research

Although our results have good predictive power to approach potential customers and churners, there are opportunities for future research. First, it is notable that the usual factors, such as frequency and number/scope of previous purchases, remain essential for prediction. However, different variables related to the content or even seasonality might add significantly to the model. Second, the amount of data for a journalistic start-up is by its nature limited. While the performance of the prediction models is satisfactory, more detailed data sets would achieve greater accuracy and nuanced predictions. Third, in terms of strategic planning for media companies, future research could apply multi-class prediction about highly valued content and topics.

6. References

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1–24. https://doi.org/10.1186/s40537-019-0191-6

Anderson, J., & Rainie, L. (2020, February 21). Many tech experts say digital disruption will hurt democracy. *Pew Research Center – Internet & Technology*. https://www.pewresearch.org/Internet/2020/02/21/many-tech-experts-say-digital-disruption-will-hurt-democracy/

Axelrod, C. (2016). Analyzing big, midsize, and small data for application security. *International Journal of Design & Nature and Ecodynamics*, *11*(3), 268–274. https://doi.org/10.2495/DNE-V11-N3-268-274

Bakke, N. A., Barland, J., & Fetveit, A. (2020). Business model transformation of legacy newspapers in the age of the digital attention economy. In I. J. Bastesen, B. K. Lange, H. E. Næss, & A. N. Thon (Eds.), *Ledelse av mennesker i det nye arbeidslivet* (pp. 19–52). Cappelen Damm Akademisk.

Bakos, Y., & Brynjolfsson, E. (1999). Bundling information goods: Pricing, profits, and efficiency. *Management Science*, *45*(12), 1613–1630. https://doi.org/10.1287/mnsc.45.12.1613

Bakos, Y., & Brynjolfsson, E. (2000). Bundling and competition on the internet. *Marketing science*, *19*(1), 63–82. https://doi.org/10.1287/mksc.19.1.63.15182

Berger, B., Matt, C., Steininger, D. M., & Hess, T. (2015). It is not just about competition with “free”: Differences between content formats in consumer preferences and willingness to pay. *Journal of Management Information Systems*, *32*(3), 105–128. https://doi.org/10.1080/07421222.2015.1095038

Berisha-Shaqiri, A. (2014). Impact of information technology and Internet in businesses. *Academic Journal of Business, Administration, Law and Social Sciences*, *1*(1), 73–79.

Berman, S. J. (2012). Digital transformation: Opportunities to create new business models. *Strategy & Leadership*, *40*(2), 16–24. https://doi.org/10.1108/10878571211209314

Bi, W., Cai, M., Liu, M., & Li, G. (2016). A big data clustering algorithm for mitigating the risk of customer churn. *IEEE Transactions on Industrial Informatics*, *12*(3), 1270–1281. https://doi.org/10.1109/TII.2016.2547584

Bowersox, D. J., Closs, D. J., & Drayer, R. W. (2005). The digital transformation: Technology and beyond. *Supply Chain Management Review*, *9*(1), 22–29.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, *93*(1), 79–95. https://doi.org/10.1016/j.jretai.2016.12.004

Brânduşoiu, I., Toderean, G., & Beleiu, H. (2016, June 9–10). *Methods for churn prediction in the pre-paid mobile telecommunications industry* [Paper presentation]. International conference on communications (COMM), Bucharest, Romania.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brettel, M., Reich, J.-C., Gavilanes, J. M., & Flatten, T. C. (2015). What drives advertising success on Facebook? An advertising-effectiveness model: Measuring the effects on sales of "Likes" and other social-network stimuli. *Journal of Advertising Research*, *55*(2), 162–175. https://doi.org/10.2501/JAR-55-2-162-175

Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, *45*(1), 12–19. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L

Buschow, C. (2020). Why do digital native news media fail? An investigation of failure in the early start-up phase. *Media and Communication*, *8*(2), 51–61. https://doi.org/10.17645/mac.v8i2.2677

Casero-Ripollés, A. (2012). Beyond newspapers: News consumption among young people in the digital era. *Comunicar*, *20*(39), 151–158. https://doi.org/10.3916/c39-2012-03-05

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management*, *21*(3–4), 193–215. https://doi.org/10.1080/14241277.2019.1695619

Chyi, H. I. (2005). Willingness to pay for online news: An empirical study on the viability of the subscription model. *Journal of Media Economics*, *18*(2), 131–142. https://doi.org/10.1207/s15327736me1802_4

Chyi, H. I. (2012). Paying for what? How much? And why (not)? Predictors of paying intent for multiplatform newspapers. *International Journal on Media Management*, *14*(3), 227–250. https://doi.org/10.1080/14241277.2012.657284

Chyi, H. I., & Lee, A. M. (2013). Online news consumption: A structural model linking preference, use, and paying intent. *Digital Journalism*, *1*(2), 194–211. https://doi.org/10.1080/21670811.2012.753299

Chyi, H. I., & Ng, Y. M. M. (2020). Still unwilling to pay: An empirical analysis of 50 U.S. newspapers' digital subscription results. *Digital Journalism*, *8*(4), 526–547. https://doi.org/10.1080/21670811.2020.1732831

Cook, C., & Sirkkunen, E. (2013). What's in a niche? Exploring the business model of online journalism. *Journal of Media Business Studies*, *10*(4), 63–82. https://doi.org/10.1080/16522354.2013.11073576

Dawson, G. S., Watson, R. T., & Boudreau, M.-C. (2010). Information asymmetry in information systems consulting: Toward a theory of relationship constraints. *Journal of Management Information Systems*, *27*(3), 143–178. https://doi.org/10.2753/MIS0742-1222270306

De Grove, F., Boghe, K., & De Marez, L. (2020). (What) can journalism studies learn from supervised machine learning? *Journalism Studies*, *21*(7), 912–927. https://doi.org/10.1080/1461670X.2020.1743737

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62. https://doi.org/10.1145/2844110

Dittert, M., Härting, R.-C., Reichstein, C., & Bayer, C. (2017). A data analytics framework for business in small and medium-sized organizations. *International conference on intelligent decision technologies*, 73(1), 169–181. https://doi.org/10.1007/978-3-319-59424-8_16

Donders, K., Enli, G., Raats, T., & Syvertsen, T. (2018). Digitisation, internationalisation, and changing business models in local media markets: An analysis of commercial media’s perceptions on challenges ahead. *Journal of Media Business Studies*, *15*(2), 89–107. https://doi.org/10.1080/16522354.2018.1470960

Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of interactive marketing*, *23*(1), 61–69. https://doi.org/10.1016/j.intmar.2008.11.003

Fernandes, K., Vinagre, P., & Cortez, P. (2015). A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News. In *Progress in Artificial Intelligence* (pp. 535-545). Springer.

Ferrer-Conill, R., & Tandoc, E. C. (2018). The audience-oriented editor. *Digital Journalism*, *6*(4), 436–453. https://doi.org/10.1080/21670811.2018.1440972

Friedrichsen, M., Kamalipour, Y. R., & Kamalipour, Y. (2017). *Digital transformation in journalism and news media*. Springer.

Gobble, M. M. (2018). Digital strategy and digital transformation. *Research-Technology Management*, *61*(5), 66–71. https://doi.org/10.1080/08956308.2018.1495969

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, *68*(8), 1678–1685. https://doi.org/10.1016/j.jbusres.2015.03.026

Grljević, O., & Bošnjak, Z. (2018). Sentiment analysis of customer data. *Strategic Management*, *23*(3), 38–49. https://doi.org/10.5937/StraMan1803038G

Gu, X., Mao, Y., Han, J., Liu, J., Wu, Y., Yu, C., Finnie, D., Yu, H., Zhai, J., & Zukoski, N. (2020). Generating representative headlines for news stories. In Y. Huang, I. King,T.-Y. Liu, & M. van Steen (Eds.), *Proceedings of The Web Conference 2020* (pp. 1773–1784). Association for Computing Machinery. https://doi.org/10.1145/3366423.3380247

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, *70*, 308–317. https://doi.org/10.1016/j.jbusres.2016.08.004

Gundlach, H., & Hofmann, U. (2021). Information search, behavioural economics, and relevance decisions in the online media industry: How strongly do the algorithms of intermediaries influence the relevance evaluation of information? *Journal of Media Business Studies*, *18*(3), 179–198. https://doi.org/10.1080/16522354.2020.1854602

Hamdani, Y., & Daim, T. (2020). Organizational transformation: Consumer goods. In T. Daim (Ed.), *Digital transformation: Evaluating emerging technologies* (pp. 523–575). World Scientific Publishing.

Han, S., Williamson, B. D., & Fong, Y. (2021). Improving random forest predictions in small datasets from two-phase sampling designs. *BMC Medical Informatics and Decision Making*, *21*(1), Article 322. https://doi.org/10.1186/s12911-021-01688-3

Hanczar, B., & Nadif, M. (2019). Controlling and visualizing the precision-recall tradeoff for external performance indices. In M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, & G. Ifrim (Eds.), *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2018* (pp. 687–702). Springer. https://doi.org/10.1007/978-3-030-10925-7_42

Hanelt, A., Piccinini, E., Gregory, R. W., Hildebrandt, B., & Kolbe, L. M. (2015). Digital transformation of primarily physical industries-exploring the impact of digital trends on business models of automobile manufacturers. *Wirtschaftsinformatik Proceedings*, *88*(3–5), 1313–1327.

He, Y., He, Z., & Zhang, D. (2009). A study on prediction of customer churn in fixed communication network based on data mining. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, *1*, 92–94. https://doi.org/10.1109/FSKD.2009.767

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for formulating a digital transformation strategy. *MIS Quarterly Executive*, *15*(2). https://doi.org/10.7892/BORIS.105447

Howden, W. J. (2021). *Business intelligence and data analytics in small business: Exploring the drivers of adoption* [Unpublished doctoral dissertation]. University of Maryland University College.

Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., & Zeng, J. (2015). Telco churn prediction with big data. In T. Sellis, S. B. Davidson, & Z. Ives (Eds.), *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (pp. 607–618). Association for Computing Machinery. https://doi.org/10.1145/2723372.2742794

Ismagilova, E., Hughes, L., Dwivedi, Y. K., & Raman, K. R. (2019). Smart cities: Advances in research—An information systems perspective. *International Journal of Information Management*, *47*, 88–100. https://doi.org/10.1016/j.ijinfomgt.2019.01.004

Jääskeläinen, A., Taimela, E., & Heiskanen, T. (2020). *Predicting the success of news: Using an ML-based language model in predicting the performance of news articles before publishing*. Proceedings of the 23rd International Conference on Academic Mindtrek, Tampere, Finland. https://doi.org/10.1145/3377290.3377299

Jiang, T., Guo, Q., Chen, S., & Yang, J. (2020). What prompts users to click on news headlines? Evidence from unobtrusive data analysis. *Aslib Journal of Information Management*, *72*(1), 49–66. https://doi.org/10.1108/AJIM-04-2019-0097

Juba, B., & Le, H. S. (2019). Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 4039–4048. https://doi.org/10.1609/aaai.v33i01.33014039

Kammer, A., Boeck, M., Hansen, J. V., & Hauschildt, L. J. H. (2015). The free-to-fee transition: Audiences’ attitudes toward paying for online news. *Journal of Media Business Studies*, *12*(2), 107–120. https://doi.org/10.1080/16522354.2015.1053345

Kannan, R., Rosdi, I. S., Ramakrishnan, K., Rasid, H. R. A., Rafy, M. H. I. M., Yusuf, S., & Salamun, S. N. A. M. (2021). Leveraging business data analytics and machine learning techniques for competitive advantage: Case study evidence from small businesses. *International Journal of Management, Finance and Accounting*, *2*(1), 73–87. https://doi.org/10.33093/ijomfa.2021.2.1.3

Kanuri, V. K., Thorson, E., & Mantrala, M. K. (2014). Using reader preferences to optimize news content: A method and a case study. *International Journal on Media Management*, *16*(2), 55–75. https://doi.org/10.1080/14241277.2014.943898

Katsikopoulos, K. V., Şimşek, Ö., Buckmann, M., & Gigerenzer, G. (2022). Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*, *38*(2), 613–619. https://doi.org/10.1016/j.ijforecast.2020.12.006

Khodabandehlou, S., & Rahman, M. Z. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, *19*(1/2), 65–93. https://doi.org/10.1108/JSIT-10-2016-0061

Khuntia, J., Sun, H., & Yim, D. (2016). Sharing news through social networks. *International Journal on Media Management*, *18*(1), 59–74. https://doi.org/10.1080/14241277.2016.1185429

Kim, D. H., & Desai, M. (2021). Are social media worth it for news media?: Explaining news engagement on Tumblr and digital traffic of news websites. *International Journal on Media Management*, *23*(1–2), 2–28. https://doi.org/10.1080/14241277.2021.1958820

Kim, J. B. (2013). Understanding the repurchase abandonment process of consumers in the online environment: Integrating dissatisfaction and distrust. *Journal of Internet commerce*, *12*(2), 154–194. https://doi.org/10.1080/15332861.2013.817863

Klaß, N. (2021). Technology builds the journalist star? An experiment on the effects of human star power and content exclusiveness on reading, sharing and paying intentions for online news. *Journal of Media Business Studies*. https://doi.org/10.1080/16522354.2021.1989972

Kosterich, A., & Weber, M. S. (2018). Starting up the news: The impact of venture capital on the digital news media ecosystem. *International Journal on Media Management*, *20*(4), 239–262. https://doi.org/10.1080/14241277.2018.1563547

Koukova, N. T., Kannan, P., & Ratchford, B. T. (2008). Product form bundling: Implications for marketing digital products. *Journal of Retailing*, *84*(2), 181–194. https://doi.org/10.1016/j.jretai.2008.04.001

Kukkonen, E. (2018). Organizing a framework for customer value management in online media relationships. *Marketing Management Journal*, *28*(1), 60–79.

Lamot, K., & Paulussen, S. (2020). Six uses of analytics: Digital editors' perceptions of audience analytics in the newsroom. *Journalism Practice*, *14*(3), 358–373. https://doi.org/10.1080/17512786.2019.1617043

Lash, M. T., & Zhao, K. (2016). Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, *33*(3), 874–903. https://doi.org/10.1080/07421222.2016.1243969

Lee, J., Jung, O., Lee, Y., Kim, O., & Park, C. (2021). A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(5), 1472–1491. https://doi.org/10.3390/jtaer16050083

Lee, W.-J., & Ong, S.-C. (2010). Learning from small data sets to improve assembly semiconductor manufacturing processes. *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, *2*, 50–54. https://doi.org/10.1109/ICCAE.2010.5451376

Li, D., & Liu, C. (2012). Extending attribute information for small data set classification. *IEEE Transactions on Knowledge and Data Engineering*, *24*(3), 452–464. https://doi.org/10.1109/TKDE.2010.254

Lischka, J. A. (2019). Strategic renewal during technology change: Tracking the digital journey of legacy news companies. *Journal of Media Business Studies*, *16*(3), 182–201. https://doi.org/10.1080/16522354.2019.1635349

Liu, A. L., Venkatesh, R., McBride, M., Reichmanis, E., Meredith, J. C., & Grover, M. A. (2020). Small data machine learning: Classification and prediction of poly (ethylene terephthalate) stabilizers using molecular descriptors. *ACS Applied Polymer Materials*, *2*(12), 5592–5601. https://doi.org/10.1021/acsapm.0c00921

Llave, M. R. (2017). Business intelligence and analytics in small and medium-sized enterprises: A systematic literature review. *Procedia Computer Science*, *121*, 194–205. https://doi.org/10.1016/j.ijinfomgt.2020.102190

Maijanen, P., Rimscha, B. von, & Głowacki, M. (2019). Beyond the surface of media disruption: digital technology boosting new business logics, professional practices and entrepreneurial identities. *Jorunal of Media Business Studies*, *16*(3), 163–165. https://doi.org/10.1080/16522354.2019.1700095

Malcorps, S. (2019). News website personalisation: The co-creation of content, audiences and services by online journalists and marketers. *Journal of Media Business Studies*, *16*(3), 230–247. https://doi.org/10.1080/16522354.2019.1689766

Maroufkhani, P., Tseng, M.-L., Iranmanesh, M., Ismail, W. K. W., & Khalid, H. (2020). Big data analytics adoption: Determinants and performances among small to medium-sized enterprises. *International Journal of Information Management*, *11*(4), 483–513. https://doi.org/10.1016/j.ijinfomgt.2020.102190

Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

McQuail, D. (1992). *Media performance: Mass communication and the public interest*. Sage.

Megahed, A., Ren, G., & Firth, M. (2015). Modeling business insights into predictive analytics for the outcome of IT service contracts. In P. P. Maglio, I. Paik, & W. Chou (Eds.), *IEEE International Conference on Services Computing* (pp. 515–521). IEEE. https://doi.org/10.1109/SCC.2015.76

Müller, O., Junglas, I., Brocke, J. vom, & Debortoli, S. (2016). Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *European Journal of Information Systems*, *25*(4), 289–302. https://doi.org/10.1057/ejis.2016.2

Myllylahti, M. (2018). An attention economy trap? An empirical investigation into four news companies’ Facebook traffic and social media revenue. *Journal of Media Business Studies*, *15*(4), 237–253. https://doi.org/10.1080/16522354.2018.1527521

Myllylahti, M. (2020). Paying attention to attention: A conceptual framework for studying news reader revenue models related to platforms. *Digital Journalism*, *8*(5), 567–575. https://doi.org/10.1080/21670811.2019.1691926

Neheli, N. B. (2018). News by numbers. *Digital Journalism*, *6*(8), 1041–1051. https://doi.org/10.1080/21670811.2018.1504626

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Reuters Institute digital news report 2018*. Reuters Insitute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf

Olson, M., Wyner, A. J., & Berk, R. (2018). Modern neural networks generalize on small data sets. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)* (pp. 3623–3632). Neural Information Processing Systems Foundation.

Ozan, Ş. (2018). A case study on customer segmentation by using machine learning methods. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1–6). IEEE. https://doi.org/10.1109/IDAP.2018.8620892

Pedregosa, E. A. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Priyam, A., Abhijeeta, G., Rathee, A., & Srivastava, S. (2013). Comparative analysis of decision tree classification algorithms. *International Journal of Current Engineering and Technology*, *3*(2), 334–337.

Park, S., Fisher, C., Flew, T., & Dulleck, U. (2020). Global mistrust in news: The impact of social media on trust. *International Journal on Media Management*, *22*(2), 83–96. https://doi.org/10.1080/14241277.2020.1799794

Putzke, J., Schoder, D., & Fischbach, K. (2010). Adoption of mass-customized newspapers: An augmented technology acceptance perspective. *Journal of Media Economics*, *23*(3), 143–164. https://doi.org/10.1080/08997764.2010.502514

Reis, J., Amorim, M., Melão, N., & Matos, P. (2018). Digital transformation: A literature review and guidelines for future research. In À. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Trends and Advances in Information Systems and Technologies: WorldCIST'18 2018* (pp. 411–421). Springer. https://doi.org/10.1007/978-3-319-77703-0_41

Rivera, W. A., & Xanthopoulos, P. (2016). A priori synthetic over-sampling methods for increasing classification sensitivity in imbalanced data sets. *Expert Systems with Applications*, *66*, 124–135. https://doi.org/10.1016/j.eswa.2016.09.010

Rußell, R., Berger, B., Stich, L., Hess, T., & Spann, M. (2020). Monetizing online content: Digital paywall design and configuration. *Business & Information Systems Engineering*, *62*(3), 253–260. https://doi.org/10.1007/s12599-020-00632-5

Rysman, M. (2009). The economics of two-sided markets. *Journal of Economic Perspectives*, *23*(3), 125–143. https://doi.org/10.1257/jep.23.3.125

Saarijärvi, H., Karjaluoto, H., & Kuusela, H. (2013). Customer relationship management: The evolving role of customer data. *Marketing Intelligence and Planning*, *31*(6), 584–600. https://doi.org/10.1108/MIP-05-2012-0055

Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications*, *9*(2), 273–281. https://doi.org/10.14569/IJACSA.2018.090238

Salaverría, R. (2019). Digital journalism: 25 years of research. Review article. *El Profesional de la Información*, *28*(1), Article e28010. https://doi.org/10.3145/epi.2019.ene.01

Salih, S., & Njenga, K. (2019). Big data for small businesses: Abstracting security and decision-making tools. *Proceedings of 4th International Conference on the Internet, Cyber Security and Information Systems*, *12*, 310–319. https://doi.org/10.29007/6tpw

Seddon, P. B., Constantinidis, D., Tamm, T., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, *27*(3), 237–269. https://doi.org/10.1111/isj.12101

Sinha, A. P., & May, J. H. (2004). Evaluating and tuning predictive data mining models using receiver operating characteristic curves. *Journal of Management Information Systems*, *21*(3), 249–280. https://doi.org/10.1080/07421222.2004.11045815

Sridhar, S., & Sriram, S. (2015). Is online newspaper advertising cannibalizing print advertising? *Quantitative Marketing and Economics*, *13*(4), 283–318. https://doi.org/10.1007/s11129-015-9160-3

Sylvie, G., & Gade, P. (2009). Changes in news work: Implications for newsroom managers. *Journal of Media Business Studies*, *6*(1), 113–148. https://doi.org/10.1080/16522354.2009.11073481

Tayebi, S., Manesh, S., Khalili, M., & Sadi-Nezhad, S. (2019). The role of information systems in communication through social media. *International Journal of Data and Network Science*, *3*(3), 245–268. https://doi.org/10.5267/j.ijdns.2019.2.002

Tenenboim, O., & Cohen, A. A. (2015). What prompts users to click and comment: A longitudinal study of online news. *Journalism*, *16*(2), 198–217. https://doi.org/10.1177/1464884913513996

Usher, N. (2017). Venture-backed news startups and the field of journalism. *Digital Journalism*, *5*(9), 1116–1133. https://doi.org/10.1080/21670811.2016.1272064

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639. https://doi.org/10.1016/j.ejor.2017.02.023

Vukanovic, Z. (2009). Global paradigm shift: Strategic management of new and digital media in new and digital economics. *International Journal on Media Management*, *11*(2), 81–90. https://doi.org/10.1080/14241270902844249

Wellbrock, C.-M., & Wolfram, M. (2021). Effects of journalism awards as quality signals on demand. *Journalism*, *22*(10), 2531–2547. https://doi.org/10.1177/1464884919876223

Wilczek, B. (2019). Complexity, uncertainty and change in news organizations. *International Journal on Media Management*, *21*(2), 88–129. https://doi.org/10.1080/14241277.2019.1590839

Winkler, J. P., Grönberg, J., & Vogelsang, A. (2019). Optimizing for recall in automatic requirements classification: An empirical study. In *2019 IEEE 27th International Requirements Engineering Conference (RE)* (pp. 40–50). IEEE. https://doi.org/10.1109/RE.2019.00016https://doi.org/0.14279/depositonce-8721

Wixom, B. H., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive*, *12*(2), Article 6.

Wu, S.-y., Hitt, L. M., Chen, P.-y., & Anandalingam, G. (2008). Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach. *Management Science*, *54*(3), 608–622. https://doi.org/10.1287/mnsc.1070.0812

Young, A., & Rogers, P. (2019). A review of digital transformation in mining. *Mining, Metallurgy & Exploration*, *36*(4), 683–699. https://doi.org/10.1007/s42461-019-00103-w

Zamith, R. (2018). Quantified audiences in news production. *Digital Journalism*, *6*(4), 418–435. https://doi.org/10.1080/21670811.2018.1444999

Ziyadin, S., Suieubayeva, S., & Utegenova, A. (2019). Digital transformation in business. In S. Ashmarina, M. Vochozka, & V. Mantulenko (Eds.), *Digital Age: Chances, Challenges and Future. ISCDTE 2019* (pp. 408–415). Springer. https://doi.org/10.1007/978-3-030-27015-5_49

Appendix E.

Appendix E.1 Descriptive Statistics for Model 1 (Purchase Prediction)

	sum_ price	num_ pur- chases	active_ months	num_ logins	actions_ per_visit	time_ per_visit	config_os	country	sub- scribed_ channel	sub- scribed_ flat	purchases _per _month	pur- chased
count	2070	2070	2070	2070	2070	2070	2070	2070	2070	2070	2070	2070
mean	581.6995	0.894203	14.15459	31.46957	4.734755	598.7323	4.831401	9.580193	0.171014	0.021739	0.134425	0.217874
std	2463.446	2.42229	10.93322	157.809	3.552512	547.3083	2.926534	4.804249	0.376613	0.145866	0.263984	0.412901
min	0	0	0	1	0.5	0	0	1	0	0	0	0
25%	0	0	6	1	2.487012	230.0958	3	9	0	0	0	0
50%	99	1	12	4	4	462.8125	5	9	0	0	0.055556	0
75%	349	1	19	17	6	783.2286	8	9	0	0	0.125	0
max	55000	87	90	4626	52	5115	8	37	1	1	4	1

Appendix E.2 Descriptive Statistics for Model 2 (Churn Prediction)

	price	volun- tary	Channel _id	country	sum_ price	num_ pay- ments	active_ since	num_ logins	actions_ per_visit	time_per _visit	config_os	churn
count	703	703	703	703	703	703	703	703	703	703	703	703
mean	409.9644	0.064011	11.45235	2.881935	18444.49	42.92745	653.5092	98.89616	4.039042	475.0412	3.082504	0.398293
std	308.4966	0.244947	13.4178	0.571399	120036.8	126.4467	331.4413	420.9544	3.582917	408.7347	2.337364	0.489895
min	50	0	1	0	249	1	13	1	0.647059	0	0	0
25%	349	0	4	3	2793	8	373	4	2.071574	217.2132	1	0
50%	399	0	4	3	7200	19	670	12	3	394.7647	3	0
75%	399	0	15	3	12967.5	32	944	45	4.875	622.3667	6	1
max	6000	1	66	7	3111822	1560	1146	4626	52	5115	6	1

Appendix E.3 Correlation of Variables of Model 1 (Purchase Prediction)

	sum_ price	**num_ purchases**	**active_ months**	**num_ logins**	**actions_ per_visit**	**time_ per_visit**	**config_os**	**country**	**subscribed _channel**	**subscribed _flat**	**Purchase s_per_ month**
sum_price	1	0.193153	0.00476	0.04819	0.011591	0.00366	0.043604	−0.02035	−0.04248	0.109206	0.114143
num_ purchases	0.193153	1	0.192554	0.388671	−0.0046	0.015269	−0.00818	−0.01611	0.015074	0.15288	0.287723
active_ months	0.00476	0.192554	1	0.236938	−0.19742	−0.12968	−0.00059	0.007217	0.364969	0.265803	−0.35982
num_logins	0.04819	0.388671	0.236938	1	−0.103	−0.05967	−0.00744	−0.01378	0.086949	0.118609	0.012471
actions_per _visit	0.011591	−0.0046	−0.19742	−0.103	1	0.561895	0.095766	−0.09178	−0.15566	−0.04287	0.126594
time_per_ visit	0.00366	0.015269	−0.12968	−0.05967	0.561895	1	0.083993	−0.07199	−0.13411	−0.03153	0.078072
config_os	0.043604	−0.00818	−0.00059	−0.00744	0.095766	0.083993	1	0.080458	0.002931	−0.00613	0.013147
country	−0.02035	−0.01611	0.007217	−0.01378	−0.09178	−0.07199	0.080458	1	0.147084	−0.0249	−0.01231
subscribed_ channel	−0.04248	0.015074	0.364969	0.086949	−0.15566	−0.13411	0.002931	0.147084	1	0.020274	−0.16483
subscribed_ flat	0.109206	0.15288	0.265803	0.118609	−0.04287	−0.03153	−0.00613	−0.0249	0.020274	1	−0.02528
purchases_ per_month	0.114143	0.287723	−0.35982	0.012471	0.126594	0.078072	0.013147	−0.01231	−0.16483	−0.02528	1

Appendix E.4 Correlation of Variables of Model 2 (Churn Prediction)

	price	**voluntary**	**channel _id**	**country**	**sum_ price**	**num_ payments**	**active_ since**	**num_ logins**	**actions_ per_visit**	**time_per _visit**	**config_os**
price	1	0.062296	−0.06088	0.024664	0.054224	−0.07067	0.064911	−0.05898	0.008595	−0.01264	0.067921
voluntary	0.062296	1	0.192718	0.023541	0.01672	0.119224	0.006055	0.040667	0.14277	0.028518	−0.08139
channel_id	−0.06088	0.192718	1	0.039305	−0.01668	0.004711	−0.32606	0.063639	0.126588	0.02539	0.036371
country	0.024664	0.023541	0.039305	1	0.021362	0.043729	0.081545	0.037686	−0.00755	−0.01505	−0.03216
sum_price	0.054224	0.01672	−0.01668	0.021362	1	0.330168	0.128761	0.078333	−0.0077	0.034583	−0.01277
num_ payments	−0.07067	0.119224	0.004711	0.043729	0.330168	1	0.279387	0.467067	−0.05194	−0.02722	−0.05537
active_since	0.064911	0.006055	−0.32606	0.081545	0.128761	0.279387	1	0.07284	−0.19408	−0.14217	−0.0691
num_logins	−0.05898	0.040667	0.063639	0.037686	0.078333	0.467067	0.07284	1	−0.07339	−0.03026	−0.03338
actions_per_ visit	0.008595	0.14277	0.126588	−0.00755	−0.0077	−0.05194	−0.19408	−0.07339	1	0.631366	0.131309
time_per_ visit	−0.01264	0.028518	0.02539	−0.01505	0.034583	−0.02722	−0.14217	−0.03026	0.631366	1	0.06237
config_os	0.067921	−0.08139	0.036371	−0.03216	−0.01277	−0.05537	−0.0691	−0.03338	0.131309	0.06237	1

Eidesstattliche 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, 22.07.2022
Ort, Datum

Unterschrift

Curriculum Vitae
Daniel O'Brien (born Kunkel)

Address	Mühlenbach 28, 50676 Cologne
Mobile	+49162 2914708
Mail	daniel.obrien@rub.de
Date of Birth	15.04.1991 in Hamm (Westf.)
Marital status	married

Employment

01/2022-current	Research Assistant, 'Schmalenbach-Institut für Wirtschaftswissenschaften', University of Applied Sciences Cologne
12/2018-11/2021	Research Assistant, 'Lehrstuhl für Medien- und Technologiemanagement', University of Cologne
09/2018-02/2019	Editorial staff, 'casting-network'
06/2015-08/2018	Teaching Assistant, 'Professur für Angewandte Medienwissenschaften', international film school cologne ifs

Education

12/2018-current	University of Cologne, 'Business Administration' (PhD)
2015-2018	international film school cologne ifs, 'Directing' (B.A.)
2012-2014	University of Bochum, 'Cognitive Science' (M.Sc.)
2010-2014	Scholarship 'Evangelischen Studienwerks Villigst e.V.'
2009-2012	University of Münster, 'Politics & Philosophy' (B.A.)
2001-2009	'Märkisches Gymnasium Hamm' (Abitur)

Publications

Peer-Reviewed Journal Article

O'Brien, D. (2022). Free lunch for all? – A path analysis on free mentality, paying intent and media budget for digital journalism. *Journal of Media Economics*, 34(1), 29–61. https://doi.org/10.1080/08997764.2022.2060241

O'Brien, D., & Wellbrock, C.-M. (2021). How the trick is done – Conditions of success in entrepreneurial digital journalism. *Digital Journalism* (Special Issue: The Business of Journalism, online first). https://doi.org/10.1080/21670811.2021.1987947

O'Brien, D., Wellbrock, C. M., & Kleer, N. (2020). Content for free? Drivers of past payment, paying intent and willingness to pay for digital journalism – A systematic literature review. *Digital Journalism*, 8(5), 643–672. https://doi.org/10.1080/21670811.2020.1770112

In Review

"Should I Stay or Should I Go? Application of ML Classifiers of Purchase and Churn Prediction to a Digital Journalistic Platform" (Under review in: *Journal of Media Business Studies*).

Awards

Young Talent Award, Media Economics Section, DGPuk 2022 for: O'Brien, D. (2022). Free lunch for all? – A path analysis on free mentality, paying intent and media budget for digital journalism.

Nomination for the Best Paper Award at EMMA Conference 2022 in Munich for: O'Brien, D., Kunz, R. (2022). Should I Stay or Should I Go? Application of ML Classifiers of Purchase and Churn Prediction to a Digital Journalistic Platform.

Book Chapters

Daniel O'Brien, Wellbrock, C.-M. & C. Buschow (2020): Bestandsaufnahme: Stand der Forschung zur Zahlungsbereitschaft für digitalen Journalismus, in: Wellbrock, C.-M. & C. Buschow (Hrsg.): Money for Nothing and Content for Free? Paid Content, Plattformen und Zahlungsbereitschaft im digitalen Journalismus, Nomos.

Daniel O'Brien, Wellbrock, C.-M. & C. Buschow (2020): „Free or Nothing" – Gratis-Mentalität im Internet und Zahlungsbereitschaft für Digitaljournalismus, in: Wellbrock, C.-M. & C. Buschow (Hrsg.): Money for Nothing and Content for Free? Paid Content, Plattformen und Zahlungsbereitschaft im digitalen Journalismus, Nomos.

Conferences

AEJMC, WMEMC, ECREA, ICA, EMMA, DGPuK Conference on Media Economics

Transfer Publications

Daniel O'Brien (2019): Quo vadis, Paid Content?, in: Timo Busch (Verl.): Das Meedia Jahrbuch 2019.

Reports

- Money for Nothing and Content for Free? – Willingness to Pay for Digital Journalistic Content (with C.-M. Wellbrock and C. Buschow, Landesanstalt für Medien NRW)
- Evaluation of the Funding Program of the Film and Media Foundation NRW (with G. Freyermuth, Film and Media Foundation NRW)
- XR (Cross / Extended Reality) in Germany 2022: Development of industry and network structures of XR companies in Germany (forthcoming)

Teaching

University of Cologne
(B.Sc. Business Administration, B.A. Media Studies, M.Sc. Media and Technology Management, M.A. Media Studies):

-Platforms, Information Goods, and Infrastructure
-Information Technology, Business and Society
-Digital Journalism - Consumer Behavior and Entrepreneurship (with C.-M. Wellbrock)
-Entrepreneurship in Digital Journalism (with C.-M. Wellbrock);
- Media Companies and Technologies: Introduction to Management Issues (with C.-M. Wellbrock)
-Digital Journalism – Preferences and Pricing (with C.-M. Wellbrock).

German Sports University Cologne (M.A. Communications):

- Selected Fields of Media Economics (mit C.-M. Wellbrock).

Hamburg Media School (M.B.A., Digital & Media Management):

- Quantitative Methods.

Summer Schools and Workshops

05.2021-09.2021	'Probability - The Science of Uncertainty and Data' (MITx)
24.03.-27.05.2021	'Introduction to Computational Thinking and Data Science' (MITx)
27.01.-01.04.2021	'Introduction to Computer Science and Programming Using Python' (MITx)
28.09. -01.10.2020	VHB-ProDok PhD Course, 'Data Science as a Research Method' (Online)
03.-07.08.2020	9th GESIS Summer, 'Mathematical Tools for Social Scientists'
20.-24.07.2020	1st Python Porto Summer School (Online), Católica Porto Business School

Programs

Python, R, SPSS, SPSS AMOS, SmartPLS, SQL, MS Office, Adobe Suite, DaVinci

Languages

English (fluently), French (basics knowledge), Spanish (basic knowledge), Latin (Latinum)

Memberships

DGPuK, VHB

Summary of Thesis

The digital transformation of the news business continues to agitate publishers. Concerned about declining sales in the print segment, legacy outlets, local news companies and freelance journalists alike search for ways to monetize digital journalism properly. At first glance, digital journalism and its monetisation as paid content seem a promising effort. The digitisation of the news business enabled distribution at a marginal cost of almost zero while giving journalists access to new research technologies and lowering the cost of entry for smaller companies.

However, while digital journalism enjoys broad popularity and use, online news are gaining few paying customers. Furthermore, online news compete within a larger digital media complex, comprising movies, games, and social media. After 25 years of experimentation, the digital future of journalism is still heavily debated in media management.

Concerning the reconstitution as a digital medium, this research examines conditions of success and obstacles for the digital news media business to be successful as a business venture. Therefore, the research question reads *What factors enable the viability and entrepreneurial success of the news media business in light of the consequences of digital transformation?* The overarching research question is considered from two angles: The first angle concerns the demand side by looking at the antecedents of the audience's willingness to pay for paid content. The second angle focuses on the supply side and therefore examines antecedents of success in the context of digital journalistic start-ups and founders.

In four studies, this thesis develops an analysis of the online news business with a local focus on the German news market. For this purpose, a variety of methods ranging from qualitative work and literature review to empirical research employing path analysis and predictive analytics are applied. Theoretically, digital transformation, free mentality and other

peculiarities of information goods inform the frame of this work. Thus, this research aims at contributing to a financially sustainable news media business.