Bringing Big Data into Media: A Decision-Making Model for Targeting Digital News Content

Zhan Liu
Institute of Information Systems
HES-SO Valais-Wallis
University of Applied Sciences and Arts Western Switzerland
Sierre, Switzerland
zhan.liu@hevs.ch
ORCID: 0000-0003-3367-3204

Nicole Glassey Balet
Institute of Information Systems
HES-SO Valais-Wallis
University of Applied Sciences and Arts Western Switzerland
Sierre, Switzerland
nicole.glassey@hevs.ch
ORCID: 0000-0003-3268-8375

***Abstract*—The purpose of this study is to investigate how big data analytics technology affects decision-making in the media industry, with a focus on digital newspapers. To achieve this goal, we propose a decision-making model to identify the relationship between audiences and news topics using big data analysis and classification, intending to help news practitioners optimize their marketing strategies. To evaluate our model, we conducted a case study using a Swiss local newspaper. Preliminary results indicate that audience reading time and volume have been significantly increased after implementing the decision that based on the model's analysis. The study also provides some guidelines for editors and journalists to target their digital news content.**

***Keywords*—decision-making, big data analytics, digital newspaper industry, news content, user behavior**

I. Introduction

By definition, the news is information about recent events. For many years the news was considered a distinct commodity in which journalists and media organizations emphasized novelty, producing truth claims with an appropriate tone and an agreed set of values [1]. Since then, however, the news has grown increasingly complex, as the well-documented changes in news production show, including new journalism practices and opportunities for media users to produce and share content with a focus on digital news distribution (e.g., mobile media applications and social media platforms). The changes in presenting the content have entailed novel ways of consuming the news, and the ever-changing use of news has become a topic of interest for academicians, politicians, and the general public [2] [3].

One challenge for today's digital news providers is to create content that attracts younger audiences [4]. In some cases, designing news content tailored for specific readers, such as young readers, is the principal goal. However, such personalized news formats require big data analytics and predictive statistical techniques to support the decision-making process. Even before the news is presented, media companies need a comprehensive understanding of their readers' news consumption behaviors and particularly of how they consume digital news on smartphones and social media platforms. For this, most conventional media are ill-prepared. Either they cannot obtain the necessary data or because the data is so unstructured it has little value for them. Even when publishers have the appropriate data in hand, they are often unable to capture valuable insights, such as reader clicks, shares, and likes. As a result, they struggle with the rapid changes in the media market.

Compared to digital-born media, most traditional forms of media and readers lack interaction, or even readers themselves. This leads to a low data-driven environment in which many publishers have almost no understanding about their readers., nor to mention real-time feedbacks. The absence of real-time responses makes it difficult for traditional media news providers to adapt to the market because their response is slow. Moreover, unlike the new media, in which businesses interact with readers (i.e., inbound marketing), traditional media, both paper and digital, are forms of outbound marketing that rely on messages sent to consumers. Research has shown that inbound marketing outperforms outbound marketing because it lets consumers engage with the product [5].

Although readers are more likely to trust news from traditional media such as newspapers, TV, or radio than from social media like Facebook [6] [7] [8], they can also be easily tired of searching for information on traditional media. Some feel that there are too many articles about uninteresting topics. Entire sections of print newspapers are often skipped. This is an area in which new media companies such as Facebook excel—although their news content may not be seen as trustworthy. Readers like story recommendations based on their past reading preferences. Indeed, they may like media reporters to hear their voices in the hope that the news media will value their suggestions. Many also like being able to choose between free and paid content. For these reasons, a better understanding of reader preferences and behaviors us need to help publishers and journalists to come up with a targeted audience strategy.

A data-driven approach is considered as an effective way for media companies to better serve their readers. Big data strategies can include reader analytics, which leads to a deeper understanding of customer preferences. Specifically, information about reader behaviors from online news media can identify the relationships between news topics and the reader's profile. However, a lack of available analytical techniques and an understandable decision-making model prevent news media from turning data into useful information. In many cases, news providers still rely on instinct to predict what is good news content [9] [10].

In this study, we propose a decision-making model that brings big data into media marketing to get a better understanding of how the audience feels about the news content, as well as which user segments are worth communicating with and which news topics convert audiences. We use unexplored datasets within the media, including statistics and unstructured data, to identify the relationship between the audience and news topics using big data analytics algorithms-based text mining and classification. Moreover, we build a case study to evaluate our model's development. The case study is based on media datasets from a local Swiss newspaper company. Preliminary results indicate that decisions proposed by our model increase readership and news viewing times.

This paper is organized as follows. We begin by presenting the background and related works. In Section 3, we introduce the approach and the decision-making model. Section 4 explains the evaluation and results of our methods through a case study. Finally, in Section 5, we conclude with a summary of the current work and present suggestions for future research.

II. Related Work

Big data analysis of news media is of growing interest to researchers. Most studies have focused on text analysis with topic modelling. For example, Yang et al., [11] proposed topic modelling to extract the textual factors from news publishing platforms that can improve predicting news popularity. Lin and He [12] proposed a supervised sentiment-topic model, using a document-specific sentiment distribution to classify documents to obtain more coherent and informative topics. Krasnashchok and Jouili [13] applied domain-specific terms for news-centric content and presented a new weighting model for Latent Dirichlet Allocation (LDA). Their results indicate that involving more named entities in topic descriptors positively influences the overall quality of topics and hence improves their interpretability, specificity, and diversity. Nikolenko et al., [14] explained the approach of topic modelling techniques based on probabilistic latent semantic analysis to improve the news topics' quality. While valuable, these works have largely applied the news texts as analysis datasets but they did not consider the relationship between readers and the news topics.

A data-driven decision support system (DDS) is an information system that analyses large datasets and allows decision-makers to extract relevant information [15] [16]. Such systems are used in different areas to support business or organizational decision-making activities. The existing research on DDS has focused on predicting social media's popularity. For example, Petrovic et al., [17] predicted the number of retweets using both Tweet content (e.g., number of words or hashtags) and author social features (e.g., number of followers or user verification). Their system helps authors define the key factors affecting the number of retweets and thereby determine the writing style and skills in Twitter messages to maximize information dissemination. Similarly, Zohourian et al., [18] evaluated and predicted the popularity of Instagram images and videos. Their results identified those characteristics that a post needs to be popular. With a 90.77% prediction rate, they have been integrated into a decision-making system that guides authors on how best to publish posts.

The decision-making models have been widely used to investigate consumer behaviors in the last years. Simon [19] argues that decision-making is a cognitive process that can be separated into simple, sequential steps. He conceptualised the decision-making model in three stages of activities: intelligence activity, design activity, and choice activity. Based on this conception, Kollat et al., [20] extended the decision model into five components, including input, information processing, decision process, and variables influencing the decision process. Belch and Belch [21] generalized and discussed relevant internal psychological processes for each stage of model, they introduced that the five-stage decision-making model could be transformed into motivation, perception, attitude formation, integration, and learning. Hereafter, Kotler and Keller [22] went further and proposed "Five-stage of consumer buying decision-model". Their model involves five different stages - need recognition, information search, evaluation of alternatives, purchase and postpurchase behavior, that consumers move through buying and after the purchase a product or service. As the name suggests, this model allows to gain a better understanding about their customers and their behaviors. There is, however, little research on decision-making models in news media content, that involves big data analytics, editors and journalists can use. We intend to fill this gap.

III. Decision-making Model and Methods

In this section, we integrated big data analysis process into traditional decision-making model which was widely discussed in literature [22] in the context of news media. We also involved five stages that a media company might face when making decision on news offerings. As indicated in Figure 1, those five stages are business objectives, data collection, data analysis, decision making, and post-decision behavior.

A. Business objectives

The first stage of the model involves the business need and problem recognition. For news media companies, typical objectives could include: grow their customer base, improve engagement and time on site, target their products and contents to the right audience, therefore eventually increase their revenue. None of these can be achieved without knowing about their readers. In other words, learning about their customers, identify who their best customers are, determine what they like, what they dislike, what problems get their way are the key. All these can be managed through the second stage: data collection.

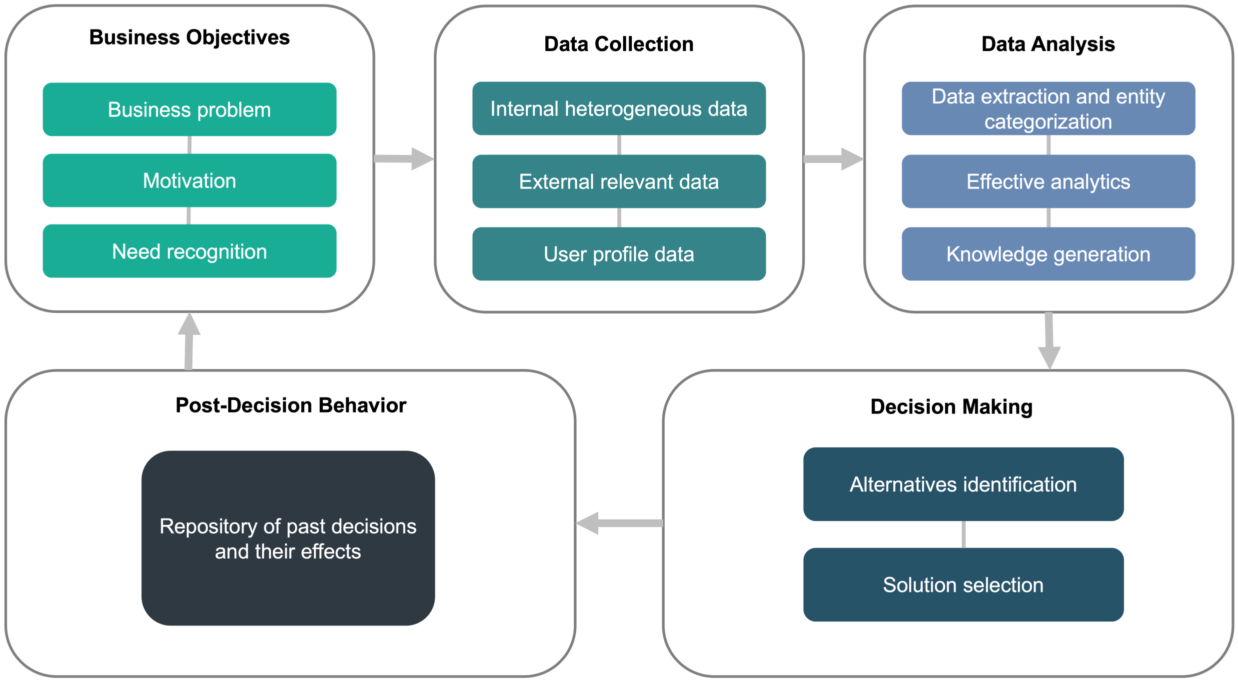

Fig. 1. Decision-making model of the big data analytics lifecycle

B. Data Collection

A data collection stage is dedicated to identifying and gathering the necessary datasets according to business objectives. Given the scope and nature of business problems, the required datasets should include both internal and external datasets. In the case of internal datasets, information about the news article content is collected, such as the publication's date and time, the number of page views and content categories, as well as information about the users, such as the reader's profile, reading interests, and reading times. In the case of external datasets, we collect a set of data from third-party news-related service providers, such as search engines and social media. The acquired data is then subjected to automated filtering to remove corrupt data or data considered to have no value to the analysis objectives.

C. Data Analysis

In the data analysis stage, we first implement data extraction and entity categorization algorithms to build the news clustering. The main tasks are to identify and annotate the entities as key elements from the text. Entity extraction based on semantic technologies disambiguate meaning and understands context, making unstructured data machine-readable and thus available for standard processing actions such as retrieving information and extracting facts. The objective is to understand the subject category of each news content (e.g., sports, economics, local news) in this stage. We then use statistical analyses to extract disparate data and transforming them into a format ready to implement a big data solution. Meanwhile, we also focus on an exploratory data analysis, in which the analysis is repeated until the appropriate pattern or correlation is revealed. The results obtained from data analysis are communicated graphically to journalists by using data visualization techniques and tools, and new knowledge and insights are generated that will be available for decision-making in the next stage.

D. Decision Making

In the decision-making stage, media editors evaluate and rank the solutions have been proposed from the data analysis stage. They choose the best solutions according to their business objectives and shape their marketing strategies. It should be also considered the implementation time of the chosen solution, which is an important factor in our model. A media decision support system with a user interface is developed to provide key performance indicators to enhance and optimize performance. Finally, these decisions, such as advertising focus, and customer relationship development, can be planned and executed to refine the business process.

E. Post-Decision Behavior

The decision-making process continues even after the actual decision is made. This stage allows marketers to understand whether they are making the right decisions. Post-decision behavior can be divided into two subgroups: post-decision satisfaction and post-decision effects. The importance of satisfaction is related to customer behavior, including number of reads and time spent reading. These behaviors help increase customer loyalty. In addition to satisfaction, important issues in post-decision effects include increased revenue for media companies. This is reflected in the user's purchase intention and becomes a registered user to read paid articles and exclusive reports. The results generated by some unsatisfied decisions can be used as an input to our model, which will be re-collected and analyzed by our system, and eventually replaced by improved decisions.

IV. Case study

The case example is a leading newspaper company (we call MediaCo) in a French-speaking region in Switzerland. Switzerland has a media landscape characterized by regional complexity. As in other countries, the scale of traditional journalism in Switzerland, which is concentrated in public broadcasting and conventional print media, is shrinking [23]. The evidence shows that, in Switzerland, the internet has become the most important source of media information, ahead of newspapers and television [24]. Consumers there particularly like free and timely news, so the dissemination of information is now often immediate, albeit sometimes superficial [25]. In recent years, almost all traditional Swiss media, including newspapers and TV, have added online portals and exclusive mobile applications, which help to maintain their consumer base in a fast-changing competitive market. However, newsmakers should be aware of emerging trends in the next round of disruption. First, the cognitive process of online news differs for readers with different backgrounds. Research shows that experienced web users benefit from online features only if the news content is difficult, whereas the presence of online features results in a drop in inexperienced users' knowledge [26]. Second, exposure to too much information can subject readers to information overload, which may, in turn, compromise their understanding of the content and result in a superficial understanding of it [27] [28] [29]. Therefore, a way to obtain useful information from the data, particularly of user interactions, is vital for publishers to develop effective news content strategies.

MediaCo's main product is a daily newspaper, published in French, and the company has been in business for more than a century. In 2022, MediaCo has over 600'000 monthly online active users. While it is still the dominant newspaper in that region, the paper faced the same challenges as many traditional media companies face: ad revenues migrated to other media. As a result, MediaCo embarked on a digital transformation journey to provide digital offerings – online and mobile versions. As part of its marketing campaign, the company sends a standard daily newsletter at 6 am to all subscribers, regardless of whether they have a paid subscription.

Since the beginning of 2018, the company collected anonymously its customers' online behavior data on various platforms including Google Analytics. However, the data are rather unstructured or siloed and are thus rarely used. MediaCo is investing in key digital capabilities in big data analytics to produce more personalized newspaper solutions.

We believe MediaCo fits well in our context because (1) data is available to collect and analyze and (2) the company is looking for data-driven solutions. It is thus willing to take actions based on data-driven results. Hence, we can evaluate our lifecycle decision-making model.

A. Dataset Description

The dataset mined in this study was obtained anonymously from Google Analytics. The data collection process contains three parts: first, we used Google Analytics API to get user profiles and information about their activity on media site. Second, we applied data cleaning techniques to remove irrelevant data (e.g., HTML tags, excessive blank space between text), convert data types (e.g., date formats), and handle missing values. Third, we developed a data converter with Python to transform the original JSON format to CSV and RDF. Our dataset covered audiences located in Switzerland over 7 days in January 2022 and contains 190,206 activities from 400 of the most active users on a local newspaper website. The dataset consists of profiles of users and their activity trajectories while browsing news on the website. The user's profile includes information on the device through which the user browses the website (e.g., mobile, desktop), the channel by which the user was acquired (e.g., direct, Google search engine, or social media), type of user (e.g., subscriber, free), and the user's location (e.g., city). For each user's daily activities, we identify the number of sessions (how many times the person has read the content), average session duration (how much time they spent on the site), bounce rate (how quickly they leave), and goal conversion rate (the percentage of the goals completed). Moreover, the user's interactions are also included in our dataset, which is grouped into four categories: "pageview", "goal completions", "e-commerce" and "event". This information allows us to see the detailed session activity. For instance, we can know when users visited the media website, which content they read, and how much time they spent reading it.

B. Data Analysis and Interpretation

Our interest in this case study is to optimize news content and media marketing strategies by analyzing different types of user preferences for news topics. Therefore, we focused on information that arises from "pageview" only. We labeled the topic of each news content using entity categorization method, which allows us to classify the different entities into the same category based on the hierarchy from Wikipedia ontology. We first applied DBpedia Spotlight [30] to complete the named entity extraction work. We then used linked data techniques to connect categories coming from annotated entities, transform them into trees of senses for each concept, and compare the trees to discover hierarchical relations between such concepts. This effort provided similar entities with the same upper-level context to establish an effective knowledge base for news category classifications. In addition, we consider the user's location and their subscription status as key drivers. Therefore, only four user categories are included: subscribers, free users, local users, and non-local users.

We first recoded the data in a format that can be used for further analysis. We dummied the user's location (e.g., whether the user is located in Canton of Valais) and the user's subscription status (e.g., whether the user has subscribed for paid services for the local media). We also dummied the daily data for the number of sessions that each user viewed for seven categories:

1) Local news (VS)
2) Swiss news (CH)
3) Sports news (SPORT)
4) Economic news (ECONOMY)
5) Worldwide news (WORLD)
6) Lifestyle news
7) Obituary

We had to exclude the last two categories for further analysis due to relatively low counts. Instead of daily data, we used weekly data for visiting topics because weekly data can better represent a typical user's behavior and diminishes the issue with missing data (e.g., one user did not visit the website every day).

Table I reports the correlation of variables used in this study. We found that the scores of the independent variables are relatively low and at an acceptable level, which indicates no severe multi-collinearity among the dependent variables.

Next, we conducted a series of multiple regressions and *t*-tests to examine the relationship between the readers' profiles – location and subscription status (independent variables) and their visiting topics (dependent variable). In the multiple regression model, we also include the devices and channels used (e.g., organic search, Facebook, etc.) as control variables.

The results in Table II show that local readers are more likely to read local news ($p<0.05$), sports news ($p<0.05$), and international news ($p<0.1$) than non-local readers. We also found that readers who had paid subscriptions are more interested in reading economic news ($p<0.05$) but less interested in Swiss news ($p<0.1$). None of the control variables are significantly related to the news preference, indicating that devices and channels used do not make a difference in news readership. Our *t*-tests confirm that the differences between the two (local and non-local users, paid vs free users) are significant.

C. Results-based Decision Making

As the results in Table II show, different types of users have different news-consuming preferences. Not surprisingly, local readers are more interested in local news, but there is no significant difference in national news viewership between them and non-local readers. Interestingly, paid subscribers are more interested in economic news than are non-paid subscribers. This should help MediaCo to better understand its readers and design appropriate marketing campaigns.

As previously noted, MediaCo sends a standard newsletter to all of its subscribers. They did not send a tailored newsletter because the company does not know how to analyze and use the data. Now, based on our analysis, we know that local users are more interested in local news and are more likely to open a newsletter if the headline story is local. On the other hand, a non-local user is less likely to find local news interesting, so the newsletter's headline story should avoid local news – a national news story might be more interesting to them. In either case, once hooked, other related news stories can be recommended.

D. Evaluation

MediaCo sent a tailored newsletter to different users. For local users, the headline is often a combination of local, sports, and world news. For non-local users, it is more likely to reflect national and economic-related news. For paid users, economic-related stories are frequently offered. For free users, headlines about national news stories are more often mentioned.

Since the implementation of the new marketing campaign in February 2022, we found that the weekly total page view volumes increased by 4%. More importantly, the average time per page view increased by 32%, from 37 seconds to 43 seconds.

Similarly, other results from big data can continuously support decision-making, which may provide additional value to the company (e.g., an increase in newsletter click rates and more time spent on digital content).

V. Conclusion and Future Work

In this paper, we presented a big data-assisted media news content analytics model, which combines big data and analytics into decision-making stages. The study extended the five-stage model of the consumer buying process [22] with the combination of the decision-making mechanism and big data analytics. In our case study, an analytical approach refers to collecting and evaluating the relationship between news content and audiences for strategic decision-making. This technique goes beyond a fundamental statistical analysis to establish the behavioral profiles of news consumers. Comparing to previous studies, our method integrates the data annotation techniques to analyze the relationship between news readers and the corresponding topics. It largely increases the accuracy of topic classification of news articles. Moreover, our proposed model enables coordinating decision-makers and the analytics group, to reveal needs and choices. It is worth noting that our model has a self-evaluation mechanism: each decision-making implementation is continuously monitored and evaluated to ensure the decision's effectiveness.

In the future, we intend to scale up the data collection to refine news topics and user segments to improve decision-making performance. Other factors such as user demographics (e.g., age and gender) and reading time could be integrated into the datasets and support decision-making. Furthermore, we plan to develop a decision-making system based on our model to collect real-time data automatically, analyze the data and produce easy-to-understand results for decision-makers to implement.

Acknowledgment

The work described in this paper is supported by the University of Applied Sciences and Arts Western Switzerland (HES-SO) under grant number 113014. We thank our media partner to the contribution of the case study.

TABLE I
CORRELATION MATRIX FOR KEY VARIABLES

	VS	CH	SPORT	ECONOMY	WORLD
VS	1				
CH	-0.0024	1			
SPORT	0.2738*	-0.0384	1		
ECONOMY	0.061	0.11312*	0.1102*	1	
WORLD	0.1411*	-0.2037*	0.1450*	0.2668*	1

* at 5% significance level

TABLE II
REGRESSION RESULTS

	VS	CH	SPORT	ECONOMY	WORLD
Independent variables					
LOCATION	1.588**	–	.906**	–	.192*
SUBSCRIPTION_STATUS	–	-1.348*	–	.104**	–
Control variables					
DEVICE, CHANNEL	–	–	–	–	–

* at 10% significance level, ** at 5% significance level

REFERENCES

[1] S. Edgerly and E. K. Vraga, "Deciding what's news: News-ness as an audience concept for the hybrid media environment," *Journalism & Mass Communication Quarterly*, vol. 97, no. 2, pp. 416–434, 2020.

[2] S. Bengtsson and S. Johansson, "A phenomenology of news: Understanding news in digital culture," *Journalism*, vol. 22, no. 11, pp. 2873–2889, 2021.

[3] I. Picone, C. Courtois, and S. Paulussen, "When news is everywhere: Understanding participation, cross-mediality and mobility in journalism from a radical user perspective," *Journalism practice*, vol. 9, no. 1, pp. 35–49, 2015.

[4] L. Galan, J. Osserman, T. Parker, and M. Taylor, "How young people consume news and the implications for mainstream media," 2019.

[5] L. Patrutiu-Baltes, "Inbound marketing-the most important digital marketing strategy," *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, vol. 9, no. 2, p. 61, 2016.

[6] A. Ardèvol-Abreu and H. Gil de Zúñiga, "Effects of editorial media bias perception and media trust on the use of traditional, citizen, and social media news," *Journalism & mass communication quarterly*, vol. 94, no. 3, pp. 703–724, 2017.

[7] Z. Liu, J. Shan, M. Delaloye, J.-G. Piguet, and N. Glassey Balet, "The role of public trust and media in managing the dissemination of covid-19-related news in switzerland," *Journalism and media*, vol. 1, no. 1, pp. 145–158, 2020.

[8] R. Fletcher and S. Park, "The impact of trust in the news media on online news consumption and participation," *Digital journalism*, vol. 5, no. 10, pp. 1281–1299, 2017.

[9] R. A. Latif, "A study on the decision making process in public and private broadcast newsrooms in malaysia," *Athens Institute for Education and Research. Retrieved from http://www. atiner. gr/papers/MED2014-0908. pdf*, 2014.

[10] D. O'neill and T. Harcup, "News values and selectivity," in *The handbook of journalism studies*. Routledge, 2009, pp. 181–194.

[11] Y. Yang, Y. Liu, X. Lu, J. Xu, and F. Wang, "A named entity topic model for news popularity prediction," *Knowledge-Based Systems*, vol. 208, p. 106430, 2020.

[12] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proceedings of the 18th ACM conference on Information and knowledge management*, 2009, pp. 375–384.

[13] K. Krasnashchok and S. Jouili, "Improving topic quality by promoting named entities in topic modeling," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 247–253.

[14] S. I. Nikolenko, S. Koltcov, and O. Koltsova, "Topic modelling for qualitative studies," *Journal of Information Science*, vol. 43, no. 1, pp. 88–102, 2017.

[15] D. J. Power, "Understanding data-driven decision support systems," *Information Systems Management*, vol. 25, no. 2, pp. 149–154, 2008.

[16] J. Lu, A. Liu, Y. Song, and G. Zhang, "Data-driven decision support under concept drift in streamed big data," *Complex & Intelligent Systems*, vol. 6, no. 1, pp. 157–163, 2020.

[17] S. Petrovic, M. Osborne, and V. Lavrenko, "Rt to win! predicting message propagation in twitter," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, 2011, pp. 586–589.

[18] A. Zohourian, H. Sajedi, and A. Yavary, "Popularity prediction of images and videos on instagram," in *2018 4th International Conference on Web Research (ICWR)*. IEEE, 2018, pp. 111–117.

[19] H. A. Simon, "Theories of decision-making in economics and behavioural science," in *Surveys of economic theory*. Springer, 1966, pp. 1–28.

[20] D. T. Kollat, J. F. Engel, and R. D. Blackwell, "Current problems in consumer behavior research," *Journal of Marketing Research*, vol. 7, no. 3, pp. 327–332, 1970.

[21] G. E. Belch and M. A. Belch, "Advertising and promotion: An integrated marketing communications perspective 6th," *New York: McGraw-Hil l*, 2004.

[22] P. Kotler and K. L. Keller, "Marketing management 12e," *New Jersey*, vol. 143, 2006.

[23] M. Reveilhac and D. Morselli, "Digital shift in swiss media consumption practices," *Swiss Journal of Sociology*, vol. 46, no. 3, pp. 535–558, 2020.

[24] M. Latzer. (2017) Strong digital well-being in switzerland. [Online]. Available: https://www.media.uzh.ch/en/Press-Releases/2017/Digital-Well-Being-.html

[25] S. Flaxman, S. Goel, and J. M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public opinion quarterly*, vol. 80, no. S1, pp. 298–320, 2016.

[26] M. Opgenhaffen and L. d'Haenens, "The impact of online news features on learning from news. a knowledge experiment," *International Journal of Internet Science*, vol. 6, no. 1, pp. 8–28, 2011.

[27] T. Jiang, Y. Hou, and Q. Wang, "Does micro-blogging make us "shallow"? sharing information online interferes with information comprehension," *Computers in Human Behavior*, vol. 59, pp. 210–214, 2016.

[28] C. York, "Overloaded by the news: Effects of news exposure and enjoyment on reporting information overload," *Communication Research Reports*, vol. 30, no. 4, pp. 282–292, 2013.

[29] J. B. Schmitt, C. A. Debbelt, and F. M. Schneider, "Too much information? predictors of information overload in the context of online news exposure," *Information, Communication & Society*, vol. 21, no. 8, pp. 1151–1167, 2018.

[30] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer, "Dbpedia spotlight: shedding light on the web of documents," in *Proceedings of the 7th international conference on semantic systems*, 2011, pp. 1–8.