Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 217–225

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

Towards a News Recommendation System to increase Reader Engagement through Newsletter Content Personalization

Elizabeth Fernandes[a,*], Sérgio Moro[b], Paulo Cortez[c]

[a,b]Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisboa, Portugal
[c]LASI/ALGORITMI, University of Minho, 4804-533 Guimarães, Portugal

Abstract

In the big data era, recommendation systems (RS) play a pivotal role to overcome information overload. In the digital landscape publishers need to optimize their editorial strategies to increase reader engagement and digital revenue. Newsletters emerged as an important conversion channel to engage readers as they provide a personalized experience by building habits. However, the lack of human resources and the need for more content assertiveness per reader lead publishers to search for an advanced analytics solution. We address this problem by proposing a research agenda on news recommendation algorithms inspired in the *table d'hôte approach* and the concept of 'personalized diversity'. Thus, the reader receives a personalized newsletter where he can discover informative and surprising content. The goal is to offer a self-contained package that retains readers, increases loyalty and consequently, the propensity to subscribe. A live controlled experiment with readers from the Portuguese newspaper *Público* was performed and a new approach is proposed. We study the effects of content recommendations on the behavior of newsletter subscribers. Findings reveal that serendipitous content tends to increase reader engagement. Finally, we propose a *table d'hôte approach* and new challenges are identified for future research.

Peer-review under responsibility of the scientific committee of the CENTERIS / ProjMAN / HCist 2023

Keywords: Data science, Digital journalism, News recommendation, Newsletters, Personalization

* Corresponding author. Tel.: +351 936 548 01.
E-mail address: elisabeth.ferna@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the CENTERIS / ProjMAN / HCist 2023
10.1016/j.procs.2024.06.165

1. Introduction

Nowadays, recommendation systems (RS) play a pivotal role to overcome information overload and fight news fatigue. Furthermore, newsletters (NLs) provide to the reader a personalized experience as it can control how and when to read the content. Moreover, NLs promote readers habit [1] that induces a loyalty increase [2]. Despite the NL's predictions about their demise, the medium continues to grow [3]. Moreover, NL represent an important channel to acquire subscribers [4]. Some successful examples have been published in the industry literature, at The Globe they proved that a reader is 10 times more likely to become online subscriber after subscribing the editorial NL [3]. Hence, 70% of the publishers say that they will put more resources into email NLs [5].

Researchers argue that news fatigue is becoming more recognized as a serious concern. As mentioned in the annual Reuters Institute Digital News Report most readers "*remain engaged and use the news regularly*" but there are readers "*that also increasingly choose to ration or limit their exposure to it- or at least to certain types of news*" [6]. As news fatigue is slowing down subscriptions sales, how to engage readers amid the news fatigue? News personalization can help people managing information overload [7] while publisher draw audiences to the website and keep them engaged [8]. An example of a successful personalization feature is myFT at the Financial Times. This feature allows the user to personalize their news services and has proven results on engagement KPI's [4].

To build an automated NL, authors were inspired by the *table d'hôte* approach presented by [9] coupled with the concept of '*personalized diversity*' presented by [10]. News personalization makes individuals' news diet unique [8], thus the goal it is to make news consumption more diverse and unique. The goal is to satisfy the news-diet needs [9] considering a particular channel. The reader can open the NL at any time after receiving the content, which induces a need of content with diverse lifetimes. Thus, a recommendation set can cover different needs of a reader by combining important stories to be informed, and unexpected surprises. Furthermore, as algorithms emerge, the notion of serendipity gains scholarly attention [11]. Serendipity refers to incidental news consumption [11], in RS refers to suggestions that are attractive and unexpected [9]. In this paper, we aim to offer a self-contained package that offers value and retain readers by developing an automated NL for readers. In this exploratory research, we aim to address the following research questions (RQs): How to cover reader needs with a NL? Can a NL impact positively the engagement KPI's? Thus, we did two live controlled experiments with real readers in which we measured the impact on reader's behavior and engagement. Then a more complex approach is proposed to be tested in the future.

2. Recommendation systems and NLs

Recommendation systems (RS) are widely used in retail and e-commerce to meet users' and business' needs [7]. RS were built based on the following items: users, items, preference [12]. Furthermore, other industries have proven the positive impact of RSs such as healthcare, transportation, agriculture, culture, or media [7]. A wide range of applications can be found across de literature [13], [14]. In the media context, RS usually focus on recommending stories in the front-page [15] or across the recirculation elements of the website [16]. Furthermore, to build an effective and efficient RS, it is more challenging for the news domain than other domains [16], [17].

Seeking to define the RS for this study, Table 1summarizes research on RS in news, published after 2020. Furthermore, a literature review presented at [18] details the top ten most cited articles in RS until the end of 2020. Moreover, [7] provides the landscape of RS research and identifies directions by presenting types of RS, challenges, limitations, and business adoptions. We further note that only one article [9] at Table 1 is focused on NLs in media, as the overall algorithms developed are usually adapted to NLs. Hence, more research has been made in the e-commerce context to improve NLs performance [19], [20].

According to [4], there are two kinds of NLs created by publishers: those that drive content back to a news website and those promoting content specifically created just for the NL. In the present approach, the goal is to drive the reader to the website by providing a bundle of content. A set of recommendations is a bundle with items interacting each other [20], a high-quality bundle boosts the user-experience and increases reading volume [19].

Table 1. Literature review on RS in news

Author	Title	Methods and Findings
[7]	Recommendation systems: Algorithms, challenges, metrics, and business opportunities	The work provides the landscape of RS research and identifies directions by presenting types of RS, challenges, limitations, and business adoptions.
[21]	News Recommendation based on Collaborative semantic Topic Models and Recommendation Adjustment	This research proposes a Collaborative Semantic Topic Model and an ensemble model to predict reader preferences. The algorithm improved the recommendation quality of articles.
[17]	A Review on Reinforcement Learning based News Recommendation Systems and its challenges	The article reviews the different Reinforcement algorithms to develop the news RS and also mentioned the challenges by those algorithms.
[22]	Research on News Recommendation Algorithm Based on User Interest and Timeliness Modeling	This work presents a model based on user interest and timeliness. Authors show that this model is superior to the traditional news RS in terms of accuracy and recall rate, also presents higher recommendation performance.
[23]	DebiasRec: Bias-aware User Modeling and Click Prediction for Personalized News Recommendation	The research presents a bias-aware personalized RS named DebiasRec to solve the problem of CTR variation influenced by position or size in the online news platform. The method includes a bias representation module, a bias-aware user modeling module, and a bias-aware click prediction module. The authors reduced the bias effect on user interest inference and model training.
[10]	Appreciating News Algorithms: Examining Audiences' Perceptions to Different News Selection Mechanisms	The study gives insights into audience's perception to news RS and news selection mechanisms. People prefer content-based similarity over collaborative similarity and content-based diversity. Concerns about information overload and missing viewpoints New concept 'personalized diversity' introduced to slowly guide audiences into a more diverse news diet
[9]	Toward the Next Generation of News Recommender Systems	*Table d'hôte* approach: set of articles that together can fulfill reader needs for information and joy.

Across the literature the most common news selection mechanisms are content-based similarity, collaborative similarity, content-based diversity [10], association rule-based, utility-based, knowledge-based [24], demographic-based, and hybrid-based [7]. RS are studied in [10] from their reader preference point of view. According to [10], people prefer content-based similarity over collaborative similarity and content-based diversity.

RS present some challenges that need to be detailed and considered when we aim to present a RS for NLs. RS can become a "warped mirror" as result of two types of bottlenecks: technical and moral [24]. In a technical perspective, the problem of data sparsity and the cold starts have been addresses by researchers by presenting sparse algorithms [7], [21], [25], [26], utility-based news RS [16], and cold start explicit and implicit solutions [27]–[29]. Regarding the moral perspective, some problems are detailed in the literature, the information imbalance or diversity issues, habituation effect [7], and information manipulation [24]. To avoid lack of diversity, surprisal and personalization measures have been proposed [30]. Recently the concept of 'personalized diversity' was introduced by [10] to slowly guide audiences into a more diverse news diet by using personalization features.

Email NLs offer to the publishers the chance to maintain a strong direct relationship with readers [31]. They provide user personality, provide context and value, content focused, consistency by using automation, despite to be personal it is also an important editorial marketing channel. Thus, *Público* aims to start automating NLs to create personalized experiences and activate habits in their digital subscribers. The articles of the recommendation package must be dissimilar as well as, complement each other for a certain purpose [9]. News articles must be recommended on what make a user read in a diverse way, rather than recommend considering clicks [16] and likes the content has [10].

Furthermore, the assumption that a click behaviour can indicate user interest can be erroneous as a click could be consequence of other factors, such as the bias of news presentation [23]. Moreover, the *table d'hôte* approach presented by [9] proposes a combination of articles that the reader needs to read (surveillant content) and articles that are not related to the reader desire of information but bring him joy and surprise (serendipitous content). Similarly, to how a chef put together a set of items in a balanced and enjoyable dining experience, the order of the content is also an important factor to give the user the best experience [9].

In the surveillance perspective, the popularity of news is often measured by considering the number of interactions in the web social network [32], [33]. However, other algorithms can be implemented to predict the reader needs [9]. In the serendipity perspective, content brings surprise, diversion or relaxation can be selected by a wide range of criterions [34] or automations [15], [35].

To increase the propensity to click in the NL, we aim to provide a selection of articles that are useful to the reader by providing a saving-time solution. The reader receives a personalized NL where he can discover surprising and informative content. The criteria selection to automatically recommend a story could be influenced or not by past reading history, personalized or non-personalized [36]. However, in the case of low engaged users, i.e., in a cold-start situation low historical browsing data are available [29]. Another important challenge in RS for NLs is related to how quickly content can become irrelevant due to new developments [34]. Thus, it is important to consider the lifecycle of a news story to develop a RS in a channel that can be accessed some minutes or hours later. The lifetime of a story is the interval between the moment when the story is published on the website and the instant when the lifecycle is over, as presented at Fig. 1**Erro! A origem da referência não foi encontrada.**. Furthermore, the lifetime-impact is the number of views that the story gets in its entire lifetime [36].

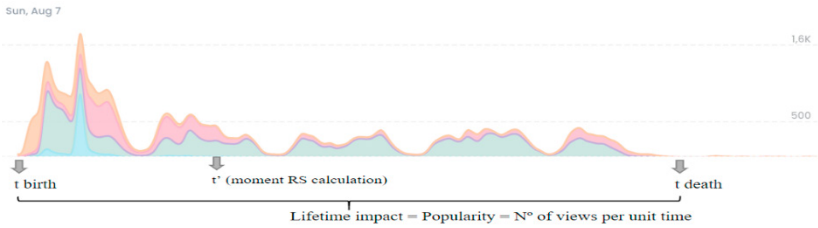

Fig. 1. Popularity lifecycle of a news story. The popularity is measured as the number of views per unit time, i.e., is the total area under the popularity curve (source: Compass analytics www.marfeel.com/audience)

3. Methodology

According to [1] churn is more important than acquisition. A small improvement on that will have a huge impact on company results [1]. Thus, to provide a data-driven solution to *Público*, this section presents an ongoing experiment on readers that present high propensity to churn. Our focus is to improve reader engagement from the NL channel to guarantee assertiveness and reader engagement increase on the website. Despite NLs only represent 2% of monthly users of the website, readers from NLs stay longer on the website (more than 6.5 minutes by visit), they see more than 4 pages by session, and visit the website approximately 4 times per month.

While the overall average of the website users see 2.4 pages by session, stays 3 minutes per visit, and do less than 2 website visits per month. Yet, research shows that micro-segments are efficient acquisition channels [1].

3.1. Research motivation, continuous experimentation, and improvement

The present experience started after running a subscriptions churn model at *Público*. Findings revealed that, when a *Público* subscriber presented less than 42% of active days (AD), for 60 days, the propensity to churn strongly increases. Thus, the team decided to send an automatic weekly NL to those subscribers. The first experiment, consisted in deliver a bundle of articles based on a simple news selection mechanism (SNSM), i.e., select N articles from the most read by all subscribers, that have high lifecycle (evergreen content published in the last 3 days) that those subscribers didn't visit yet. Those decisions were based into premises: one related to the engagement definition: reader *engagement* is a multidimensional phenomenon [37] related to the level of attention and involvement (emotional, cognitive and behavioral) with media [38] Thus, content that involved many subscribers will have high propensity to be of other subscribers interest. Furthermore, content that catch the attention of the subscriber, i.e., long articles mostly evergreen have preference to be recommended. Thus, as the goal was to provide serendipitous content, unexpected and diverse [39], the content was selected from the sub brands Fugas, íPsilon, P3, Culto and Azul. Respectively, soft news journalism related to travel, culture, wellbeing, environment, and opinion, that represents 35% to 40% of total (an example can be found in the link https://news.publico.pt/vl/-99-3174eb6ca75-0c63324342fdca1985c3c6c74e6gae0ep3we). To guarantee a good user experience, and the right content balance, on desktop and mobile, all the content is presented in the same format and size to avoid news bias regarding boxes size [23].

Results of 16 weekly NLs sent to an average of 9,905 low engaged subscribers, across the first semester of 2022, showed an average of 45% open rate (OR) and 6% of click through rate (CTR). A better performance than the average values registered at *Público* NLs and better than the benchmark values reported by [40] for the media industry (22% OR and 5% CTR). Furthermore, 55% of the subscribers that clicked on the NL increased AD, 18% kept the same AD, while 27% decreased AD in the next 60 days. Thus, subscribers were induced to read more and visit the website. Moreover, from those subscribers that had a subscription end in the next 60 days after open at least one NL, 70% renewed while those that did not open the NLs 66% renewed. Findings revealed that act on those low engaged subscribers with a content email induces an engagement recover.

Furthermore, to improve CTR and consequent website visits, a second experiment was performed for continuous improvement. The marketing team decided to ask a provider for a ML model to compete with our SNSM approach. The experiment consisted of applying a A/B test in four NLs sent in a weekly basis, between 27th July and 17th August 2022. The reader's sample was a group of register readers that made only one website visit in the last 30 days, low engaged registered readers. The provider ML model considers all the content published by *Público* to build the bundle of articles. Results revealed no statistically significant difference between both CTR. However, the SNSM achieves more 1p.p. in the CTR on average. Moreover, after two days of the NL delivery, the CTR of the NL with the SNSM approach still increasing, while CTR of the NL with ML decreases. This indicates that the content recommended by SNSM is more engaging across more days. Thus, the lifetime of the content delivered is an important factor as a NL is not a real-time channel that could be accessed some days after send. Those results drove the authors to propose a RS inspired on the *table d'hôte* RS [9].

The methodology followed was inspired on the Cross-Industry Standard Process model for Data Mining (CRISP-DM) methodology (Moro et al., 2011). CRISP-DM consists of a cycle that comprises six stages named: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Data analysis and modelling was performed through *Colaboratory* a product from Google Research that is a hosted Jupiter notebook service with access to Google hardware [41].

3.2. Data collection and data preparation

The data used in this research was collected from *Público* database. When a logged user reads an article, it is tracked on the website, and then is recorded as a hit that represents a row in the data warehouse (e.g. BigQuery by Google). Furthermore, a hit contains information like *date, time, user id, user environment variables, article url,* and *events* [42] (such as, social share, click in recirculation boxes, and etc). However, clickstream data collection contains unrelated information and some noisy data. Thus, data processing and cleaning is a pivotal step before data modelling. The list of attributes collected for each article is presented at Table 2. Three main tables of the database were combined: table of the articles and their main characteristics, table of website hits, and table of tags. By article we also added the new variable *homepage post* that indicates if journalists decided to post the article in the homepage, it is an indicator of journalistic relevance of the content.

3.3. Data modelling and future work

Creating a *table d'hôte* RS is a multi-objective process [9]. The bundle results of two main elements combination: *Surveillance* and *Serendipity*. As presented at Fig. 2, the first three articles are information that reader needs, and the last three are stories that offer a surprise to the reader. Calculations are detailed as follows. Firstly, to define the three articles that each reader needs, we calculate the expected value of views for new articles. From the articles expected to have high number of views, we excluded those that the reader already read. To predict the views by article, we apply daily the XGBoost machine learning algorithm to model the number of views considering the features highlighted at Table 2. XGBoost was applied as it has achieved superior results in several ML challenges [43]–[45]. Then, for articles published in the last N hours the number of views is predicted i.e., the lifetime impact expected. Some experiments could be performed to better define the better N, also that depends on the hour of the day that the NL is send. Secondly, to define the three surprising articles by user we use the SNSM approach. Then, by reader, we concatenate the respective articles (see Fig. 2).

Table 2. List of attributes

Feature group	Feature	Description
Article characteristics	Title characters	Number of characters in the title
	Tags	Number of tags
	Date day	Day of the week published
	Hour	Hour published
	Antiquity	Nº of article days
	Autor ID	Autor ID
	Premium	Article Premium (y/n)
	News type ID	Article type defined by the journalist from the list: *Noticia*, Listicle, Reportagem, Perguntaserespostas, Imagem_semana, Foto_legenda, Shopping, Depoimento, Opiniao, Entrevista, Cronica, Analise, Listicle _Number, Perfil, Cronica_de_jogo, NL, Prepublicacao, Critica, Comentario, Editorial, Investigacao, Ensaio, Ficha
	Section	Culure, World, Economics, Blue, Society, Polithics, Science, Local, Sports, Technology, Opinion, NLs, Iniciativas Público, Público na Escola, Terroir, Fugas P3, Culto
Traffic performance metrics	Users	Number of Users
	Views	Number of Views
	Logged Users	Number of logged users, users that are registered into the website
Article content relevance	Homepage post	Article posted at the Homepage (y/n)
	Views of the main tag	Average pageviews of the main tag

As presented at Table 3, two models were built. However, because of technological challenges, at the moment that we wrote this article, it was not possible to have ready the live controlled experiment with real readers. Future work will be to test this approach and measure the impact of the NL on reader's reading behavior.

Table 3. XGBoost model evaluation

	Mean absolute error	Mean squared error	Mean squared log error	Median absolute error	R2 score
Model 1	4,868	210,534,291	2.77	1,911	0.485
Model 2	**4,306**	**133,022,393**	**2.58**	**1,584**	**0.648**

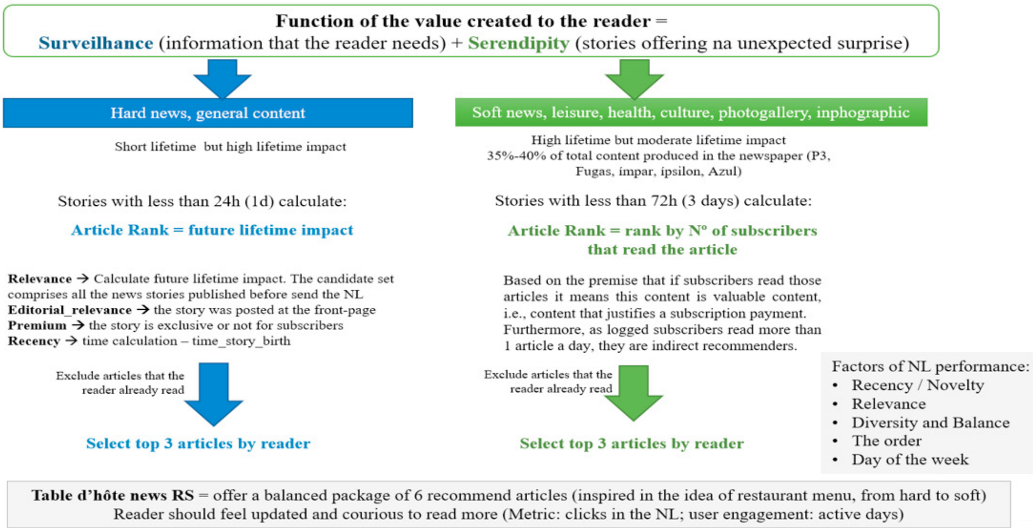

Fig. 2. Multi-objective process to build a news recommendation list by reader

4. Discussion and future research opportunities

The engagement NL, with the SNSM approach, has been sent automatically on a weekly basis to low engaged subscribers (around 12 thousand) since January 2022. On average, 256 users click at least once in the NL and 2,825 more pageviews are generated. This increases the AD on low loyal subscribers, increases engagement, as well as the perceived value of the subscription, without spending human resources. However, after, more than forty weeks, there was a slightly decrease in performance to 43.4% OR and 5.5% CTR that induces a need to improve the algorithm. The second experiment already demonstrated that long lifecycle articles with serendipitous content has higher CTR. As the reader can open the NL some hours and days after the sending date, content with short lifecycle could decrease reader NL loyalty. This finding is consistent with the work of [39]. Serendipitous content is relevant while novel content could be irrelevant [39]. As argued by [9], the NL should be a combination of two strategies to provide the news-diet that user needs with diversity and balance. Thus, to answer the two RQ raised, it was proven that a NL can impact positively reader engagement, but more must be done to answer the RQ: How to cover reader needs with a NL?

5. Conclusion and future research agenda

In this paper, we present two experiments with NLs content automation and propose a new approach to increase reader engagement to have a save-time solution for the newsroom. The main goal it is to use the *table d'hôte* and *personalized diversity* concepts adapted to *Público* context (that allow us to answer RQ1). Thus, by combining, surveillant content, that are the news that must be read, and serendipitous content, that bring them surprising news, we propose an approach that will increase reader engagement as the NL that only contains serendipitous content already shown an increase of AD. Future work should improve the assertiveness of the content selected by analyzing which are the serendipitous news that are more joyful, a text mining approach with tags and keywords study can provide new directions. Furthermore, to improve the bundle generation, the algorithm to predict article views could be improved, and other features could be added to increase CTR and OR. Also, data from the real-time software could be added. Moreover, we should consider an analysis of the best day and hour of the week to send the NL. Moreover, in the experiments the NLs contained six articles, but other bundle sizing can be tested considering desktop and mobile browser. Regarding NL subject, as it was always the same, a text mining approach could help to improve the subject and consequently increase OR.

Acknowledgements

This article was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2020 and UIDP/04466/2020], and [Project UIDB/00319/2020].

References

[1] E. Suárez, "How to build a successful subscription news business: lessons from Britain and Spain," 2020.
[2] D. Mcmullan, "How Newsletters Are Redefining Media Subscriptions," 2018.
[3] N. Newman, R. Fletcher, A. Kalogeropoulos, and R. K. Nielsen, "Reuters Institute Digital News Report 2019," 2019.
[4] D. Mcmullan, "How Newsletters Are Redefining Media Subscriptions INMA," Jun. 2018. Accessed: Dec. 26, 2022. [Online]. Available: https://www.inma.org/report/how-newsletters-are-redefining-media-subscriptions
[5] N. Newman, "Digital News Report: Journalism, Media, and Technology Trends and Predictions 2022," *Reuters Institute*, p. 46, 2022.
[6] N. Newman, R. Fletcher, A. Schulz, S. Andi, C. Robertson, and R. Nielsen, "Digital News Report 2021 10th Edition," 2022. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021
[7] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Applied Sciences (Switzerland)*, vol. 10, no. 21, pp. 1–20, 2020, doi: 10.3390/app10217748.
[8] A. Lafrance, "The Power of Personalization," *Nieman Reports*, 2017. https://niemanreports.org/articles/the-power-of-personalization/ (accessed Aug. 11, 2022).
[9] H. Abdollahpouri, E. C. Malthouse, J. A. Konstan, B. Mobasher, and J. Gilbert, "Toward the Next Generation of News Recommender Systems," *The Web Conference 2021 - Companion of the World Wide Web Conference, WWW 2021*, pp. 402–406, 2021, doi: 10.1145/3442442.3452327.
[10] G. Joris, F. De Grove, K. Van Damme, and L. De Marez, "Appreciating News Algorithms: Examining Audiences' Perceptions to Different News Selection Mechanisms," *Digital Journalism*, vol. 9, no. 5, pp. 589–618, 2021, doi: 10.1080/21670811.2021.1912626.
[11] K. van Damme, M. Martens, S. van Leuven, A. Vanden, and L. Marez, "Mapping the Mobile DNA of News. Understanding Incidental and Serendipitous Mobile News Consumption," *Digital Journalism* , vol. 8, no. 1, p. 4968, 2020.
[12] N. Omar, Y. M. K. Omar, and F. A. Maghraby, "Machine Learning Model for Personalizing Online Arabic Journalism," *Mach Learn*, vol. 11, no. 4, 2020, doi: 10.14569/IJACSA.2020.0110484.
[13] M. Nasir, C. I. Ezeife, and A. Gidado, "Improving e-commerce product recommendation using semantic context and sequential historical purchases," *Soc Netw Anal Min*, vol. 11, no. 1, Dec. 2021, doi: 10.1007/s13278-021-00784-6.
[14] D. R. Liu, K. Y. Chen, Y. C. Chou, and J. H. Lee, "Online recommendations based on dynamic adjustment of recommendation lists," *Knowl Based Syst*, vol. 161, pp. 375–389, 2018, doi: 10.1016/j.knosys.2018.07.038.
[15] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Optimizing the recency-relevancy trade-off in online news recommendations," *26th International World Wide Web Conference, WWW 2017*, no. i, pp. 837–846, 2017, doi: 10.1145/3038912.3052656.
[16] M. Zihayat *et al.*, "A utility-based news recommendation system," *Decis Support Syst*, vol. 117, no. December 2018, pp. 14–27, 2019, doi: 10.1016/j.dss.2018.12.001.
[17] S. Bangari, S. Nayak, L. Patel, and K. T. Rashmi, "A Review on Reinforcement Learning based News Recommendation Systems and its challenges," *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, pp. 260–265, 2021, doi: 10.1109/ICAIS50930.2021.9395812.
[18] E. Fernandes, S. Moro, and P. Cortez, "Data Science, Machine Learning and Big Data in Digital Journalism: A survey of state-of-

the-art, challenges and opportunities," *Expert Syst Appl*, vol. 221, 2023.

[19] J. Bai *et al.*, "Personalized Bundle List Recommendation," in *The World Wide Web Conference*, 2019, pp. 60–71.

[20] Zhu, P. Harrington, J. Li, and L. Tang, "Bundle Recommendation in eCommerce," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 657–666.

[21] Y.-S. Liao, J.-Y. Lu, and D.-R. Liu, "NEWS RECOMMENDATION BASED ON COLLABORATIVE SEMANTIC TOPIC MODELS AND RECOMMENDATION ADJUSTMENT," 2019.

[22] Z. Qin and M. Zhang, "Research on News Recommendation Algorithm Based on User Interest and Timeliness Modeling," *ACM International Conference Proceeding Series*, vol. PartF16898, 2021, doi: 10.1145/3448734.3450933.

[23] J. Yi, F. Wu, C. Wu, Q. Li, G. Sun, and X. Xie, *DebiasRec: Bias-aware User Modeling and Click Prediction for Personalized News Recommendation*, vol. 1, no. 1. Association for Computing Machinery, 2021. [Online]. Available: http://arxiv.org/abs/2104.07360

[24] S. Feng, J. Meng, and J. Zhang, "News recommendation systems in the Era of information overload," *Journal of Web Engineering*, vol. 20, no. 2, pp. 459–470, 2021, doi: 10.13052/jwe1540-9589.20210.

[25] D. Sun, T. Zhou, J. G. Liu, R. R. Liu, C. X. Jia, and B. H. Wang, "Information filtering based on transferring similarity," *Information filtering based on transferring similarity. Physical Review*, vol. 80, no. 1, 2009.

[26] Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," *Daniel* , vol. 22, no. 1, pp. 116–142, 2004.

[27] J. Gope and S. K. Jain, "A survey on solving cold start problem in recommender systems," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017, pp. 133–138.

[28] R. F. Rahmat, T. Z. Lini, S. Purnawati, R. Mulki, D. Sukarja, and Lubis, "Sparse matrix implementation on personal news recommendation for anonymous user," in *4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, 2020, pp. 140–145.

[29] J. Misztal-Radecka, B. Indurkhya, and A. Smywiński-Pohl, "Meta-User2Vec model for addressing the user and item cold-start problem in recommender systems," *User Model User-adapt Interact*, vol. 31, no. 2, pp. 261–286, 2021.

[30] Zhou, Z. Kuscsik, J. G. Liu, M. Medo, J. R. Wakeling, and Y. C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences* , vol. 107, pp. 4511–4515, 2010.

[31] A. Jack, "Editorial Email newsletters: The Medium Is Not the Only Message," 2016. [Online]. Available: https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civil wars_12December2010.pdf%0Ahttps://think-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625

[32] K. Fernandes, P. Vinagre, and P. Cortez, "A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News," in *Portuguese Conference on Artificial Intelligence*, 2015, pp. 535–546.

[33] R. Obiedat, "Predicting the popularity of online news using classification methods with feature filtering techniques," *J Theor Appl Inf Technol*, vol. 98, no. 8, pp. 1163–1172, 2020.

[34] J. Mizgajski and M. Morzy, "Affective recommender systems in online news industry : how emotions influence reading choices," *User Model User-adapt Interact*, vol. 29, no. 2, pp. 345–379, 2019, doi: 10.1007/s11257-018-9213-x.

[35] F. Garcin, C. Dimitrakakis, and B. Faltings, "Personalized News Recommendation with Context Trees," in *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 105–112.

[36] A. Chakraborty, S. Ghosh, N. Ganguly, and K. P. Gummadi, "Optimizing the recency - relevance - diversity trade - offs in non - personalized news recommendations," *Information Retrieval Journal*, vol. 22, no. 5, pp. 447–475, 2019, doi: 10.1007/s10791-019-09351-2.

[37] S. Steensen, R. Ferrer-Conill, and C. Peters, "(Against a) Theory of Audience Engagement with News," *Journal Stud*, vol. 20, pp. 1–19, 2020, doi: 10.1080/1461670X.2020.1788414.

[38] T. B. Ksiazek, L. Peer, and K. Lessard, "User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments," *New Media Soc*, vol. 18, no. 3, pp. 502–520, 2016, doi: 10.1177/1461444814545073.

[39] D. Kotkov, S. Wang, and J. Veijalainen, "A survey of serendipity in recommender systems," *Knowl Based Syst*, vol. 111, pp. 180–192, Nov. 2016, doi: 10.1016/j.knosys.2016.08.014.

[40] Mailchimp, "2022 Email Marketing Statistics & Benchmarks - Mailchimp," *https://mailchimp.com/pt-br/resources/email-marketing-benchmarks/*, 2022.

[41] M. J. Nelson and A. K. Hoover, "Notes on Using Google Colaboratory in AI Education," *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pp. 533–534, 2020, doi: 10.1145/3341525.3393997.

[42] Google, "Event Measurement," 2022. https://developers.google.com/analytics/devguides/collection/analyticsjs/events (accessed Jan. 04, 2021).

[43] J. Brownlee, *XGBoost With Python: Gradient Boosted Trees With XGBoost and scikit-learn*. Machine Learning Mastery, 2016.

[44] R. Wieland, T. Lakes, and C. Nendel, "Using SHAP to interpret XGBoost predictions of grassland degradation in Xilingol, China," *Geosci Model Dev*, vol. 14, no. 3, pp. 1493–1510, 2021.

[45] T. Chen and C. Guestin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd 611 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, A. Press, Ed., San Francisco, California, USA, 2016, pp. 785–794.