ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/380237469

A Bibliometric Analysis of Customer Churn Prediction in the Telecommunications Industry Authors

Article *in* ET journal · April 2024

CITATIONS
0

READS
96

4 authors, including:

Minwir Al-Shammari
University of Bahrain
99 PUBLICATIONS **774** CITATIONS

SEE PROFILE

Habes Ali Altairey
University of Bahrain
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Najla Aljawder
University of Bahrain
2 PUBLICATIONS **11** CITATIONS

SEE PROFILE

All content following this page was uploaded by Najla Aljawder on 01 May 2024.

The user has requested enhancement of the downloaded file.

IJSRIS journal

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND INNOVATIVE STUDIES

ISSN: 2820-7157
www.ijsrisjournal.com

April 2024 **Volume 3** **Number 2** **Pages 01-07**

A Bibliometric Analysis of Customer Churn Prediction in the Telecommunications Industry

Minwir Al-Shammari[1], Habes Al-Tairey[2], Mohamed Manea[3], and Najla Aljawder[4]
[1]College of Business Administration, University of Bahrain, Kingdom of Bahrain, malshammari@uob.edu.bh
[2,3,4] College of Business Administration, University of Bahrain, Kingdom of Bahrain, 20001670@stu.uob.edu.bh; 19983941@stu.uob.edu.bh; 19912319@stu.uob.edu.bh

ABSTRACT: Analyzing and exploring vast amounts of scientific data is expected in bibliometric analysis. Statistical and graphical categorized tests are conducted to highlight the spatiotemporal elements of the paper's data and summarize it. Bibliometric analysis helps determine the most influential authors, organizations, and publications in a particular subject. It may also monitor how research trends change over time. This study involved the bibliometric evaluation of papers on customer churn prediction using machine learning (ML) in the Scopus databases. The study used the analytic tools VosViewer and R-Bibliometrics (an open-source R's package for bibliometric analysis powered by the R programming language). Researchers often use a combination of VOS viewer and R-Bibliometrics to harness the strengths of both tools. VOS viewer excels in generating intuitive network visualizations, aiding in identifying research communities and thematic clusters. R-Bibliometrics offers unparalleled flexibility and customization, enabling in-depth statistical analyses and reproducibility in research workflows. Integrating these tools provides a comprehensive approach, leveraging the user-friendly visualization capabilities of VOS viewer analytic features.

Keywords: Big Data, Customer Churn, Performance Analytics, Telecommunications Industry.

INTRODUCTION

Churn management involves identifying potential customers who may switch to a competitor's service provider [1]. Customer churn is a common issue faced by businesses across various industries such as banking [2-4], news [5], insurance [6], online mobile gaming [7], telecommunications [8-10], and online websites [11,12], and the issue has a significant impact on highly competitive businesses like the telecom sector [13,14]; hence the companies must address the issue of customer churn, which refers to the sudden shift of customers from one provider to another [15]. The Customer Relationship Management (CRM) system is utilized to analyze customer behavior to enhance profitability [16]; CRM and churn analysis are related fields, aiming to improve customer value and churn research identifying reasons for service discontinuation. [17,18].

Customer churn management involves forecasting the likelihood of churn and understanding customer preferences. Predictive capabilities are crucial for preserving recurring income, enhancing customer retention, and promoting business growth. Factors such as unavoidable and unavoidable reasons can influence customer decisions [19, 20]. In consumer-oriented marketing, a method that effectively guides consumers through the marketing funnel, encompassing stages from brand awareness to purchase decisions, is paramount for businesses seeking sustained success [21]. As clients traverse this multifaceted journey, it becomes evident that they must be ushered through each stage, with particular attention paid to cultivating customer retention—a pivotal factor in transforming one-time customers into devoted patrons and safeguarding against potential defection to competitors.

Various approaches have been posited to enhance revenue streams, including assessing client acquisition, upselling, and retention strategies regarding their return on investment [22]. Notably, findings from the study underscore client retention as the most financially lucrative strategy among these, affirming its central role in bolstering a company's bottom line. Consequently, the industry's strategic focus has been twofold, underscored by a concerted effort to attract new customers and safeguard and nurture existing relationships [23]. This dual approach acknowledges the symbiotic relationship between customer acquisition and retention in driving sustained business growth.

In navigating the dynamic landscape of customer churn, businesses proactively deploy strategies to identify and address client attrition at its nascent stages [24]. This forward-thinking approach enables companies to mitigate potential losses and underscores the significance of a proactive stance in preserving and fortifying their customer base. In the sector's contemporary landscape, a notable competitiveness surge has been observed, primarily attributed to the proliferation of operators and rapid technological advancements [25]. This heightened competition is further exacerbated by globalization and the widespread accessibility of communication technologies, compelling customers to contemplate the possibility of transferring services between different providers [26].

Reducing customer churn has become a critical objective for telecom companies in recent years, and predictive analytics has emerged as a powerful tool for achieving this goal. Bibliometric analyses help determine research objectives and gaps, monitor science and technology advancement, and recognize and honor outstanding scientific work [27]. This study aims to address these objectives:

- To explore the gaps in the research on customer churn prediction regarding the title, country, and type of machine learning (ML) models.
- To identify the significant journals/conference proceedings, researchers, affiliations, countries, articles, and terminology in ML in customer churn prediction research.
- To explore the relationships based on co-authorship, co-occurrence, citation, and bibliographic coupling in the literature on ML in customer churn prediction research.

1. METHODOLOGY

Bibliometrics can be used to identify scientific gaps in recent publications and cluster them to find gaps in publishing countries' maps. This strategy can help improve research by examining and discovering cutting-edge research subjects.

In the first stage, the study started with selecting the Scopus database. Scopus is the most extensive database available for academics, agencies, and commercial enterprises [27]. The study set a broad period extending over 20 years from 2001 to 2023 and ran multiple searches. The search was conducted in the search fields: "Titles, Abstract, and Author Keywords" using the following search string in the Scopus database:

Query Preview: (TITLE-ABS-KEY (("customer churn" OR "Customer attrition") AND ("Prediction" OR "Predict") AND ("Machine Learning" OR "Data Mining" OR "Artificial Intelligence") AND telecom*)).

Analyzing and exploring vast amounts of scientific data is expected in bibliometric analysis. Statistical and graphical categorized tests are conducted to highlight the spatiotemporal elements of the paper's data and summarize it [28]. Bibliometric analysis helps determine the most influential authors, organizations, and publications in a particular subject. It may also monitor how research trends change over time [28]. This study's technique involved bibliometric evaluation of the papers in the Scopus databases. The study used the analytic tools VosViewer and R-bibliometrics (an open-source R's package for bibliometric analysis powered by the R programming language). Researchers often use a combination of VOS viewer and R-bibliometrics for bibliometric analyses to harness the strengths of both tools. VOS viewer excels in generating intuitive network visualizations, aiding in identifying research communities and thematic clusters.

On the other hand, R-bibliometrics offers unparalleled flexibility and customization, enabling in-depth statistical analyses and ensuring reproducibility in research workflows. Integrating these tools provides a comprehensive approach, leveraging the user-friendly visualization capabilities of VOS viewer alongside the robust analytical features of R-bibliometrics [29]. Singh et al. [30] 's bibliometric tools, adopted in this paper with some modifications, are data statistics, keyword analysis, publication years, citation analysis, co-authorship analysis, co-occurrence analysis of keywords, and bibliographic coupling of documents.

2. ANALYSIS OF RESULTS

Figure 1 represents the primary information about the data obtained from the Scopus databases. It shows the publication timeline of the papers on the selected database from January 2001 to December 3, 2023. After 2015, the number of published documents exceeded ten, and the Scopus records reached 45 by that year.

Three hundred twenty papers were published in the last 21 years (from 2001 to 2023) using the data imported from the Scopus database. There are 597 keywords and 222 sources. A conceptual map was created to illustrate the relationships between the keywords in the database. Gaps in the literature can be identified by analyzing keywords to find underrepresented or unused keywords. This analysis can help us propose research questions to fill these gaps and add to our knowledge. For example, customer churn in telecom has yet to be studied using ensemble ML models, hybrid ML models, reinforcement learning, and blended ML models.

Figure 2 of the co-occurrence analysis shows a close relationship between "data mining," "decision trees," and "churn management." Also, "Churn prediction, SVM, logistic regression, and customer churn or churn prediction." In 2021, deep learning, xgboost, random forest, and ensemble learning were the most popular ML applications.

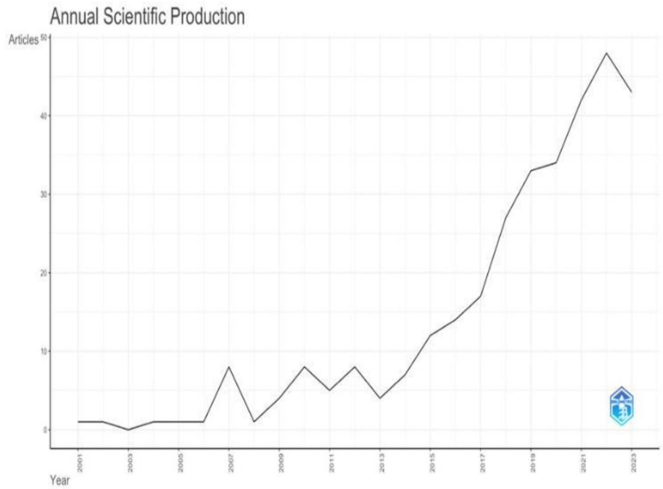

Figure 1: Number of yearly documents published

According to the co-occurrence analysis of keywords, the terms "sales," "forecasting," "data mining," "customer churn prediction," and "churn predictions" are the most frequently occurring. This list represents the most often used terms in the subject field's current research, which aspiring writers can consult.

Figure 3 visualizes countries' co-authorship. The countries are divided into four groups, each comprising closely connected countries to the study's topic: blue (Belgium, South Korea, Pakistan, and the United Kingdom), green (India, Saudi Arabia, United Arab Emirates, and Jordan), red (Turkey, United States, Nigeria, Malaysia), and yellow (China). Table 1 details the top ten countries' scientific production for Scopus.

Figure 3 represents the co-authorship of affiliations in Scopus. Its analysis highlights the Department of Information Technology at Manipal University Jaipur and the School of Computer Science and Informatics at University College Dublin. The results show significant research contributions from various global affiliations.

Figure 4 displays the citations for articles published between 2001 and 2023 (until December 3). The database recorded the highest number of research citations in 2004, but Scopus recorded more than 75 citations. In Figure 7, a network map shows the collaborations of leading academics in the field of ML-customer churn. The research studies were conducted in cooperation with these authors. Using classification techniques, the authors of the red and green clusters forecast client attrition.

The bibliometric data in Scopus have 222 citation sources. Figure 5 shows the density of published document sources. The database recorded the highest number of research citations in 2004, but Scopus recorded more than 75 citations. Academics may use these sources to acquire excellent research for future research. Figure 6 shows the density of the most frequently cited studies. 82 out of the 321 articles in Scopus have at least ten citations, according to the list of the top 10 mentioned documents. Verbeke [31] has the most citations in Scopus, with 284, followed by Hwang [32], with 276.

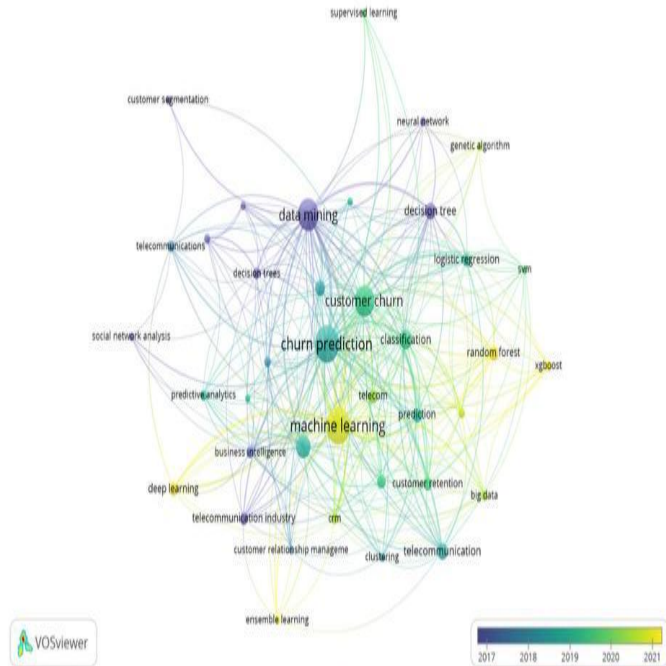

Figure 2: Visualisation of Co-occurrence of Keywords

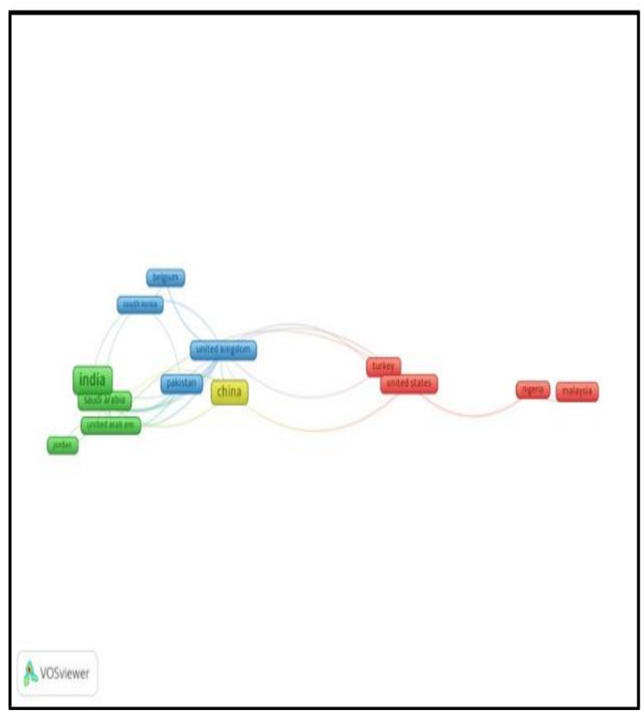

Figure 3: Co-authorship by Countries

Table 1: Top Ten Countries' Details for Scopus

Country-Scopus	TC	Average Article Citations
BELGIUM	347	115.70
INDIA	333	6.90
KOREA	301	100.30
CHINA	300	8.10
PAKISTAN	300	60.00
IRELAND	234	29.20
FRANCE	144	72.00
USA	144	24.00
SERBIA	54	54.00
GERMANY	49	24.50

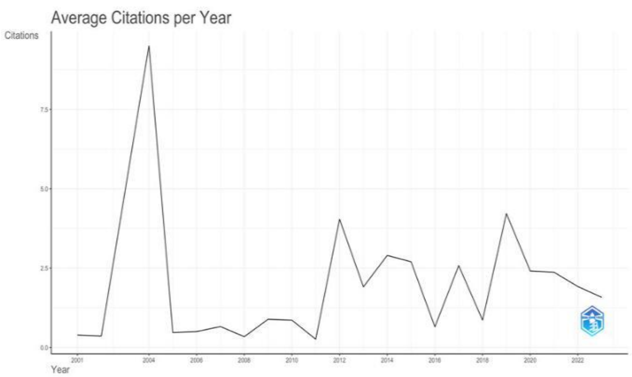

Figure 4: Yearly Citations of the Documents

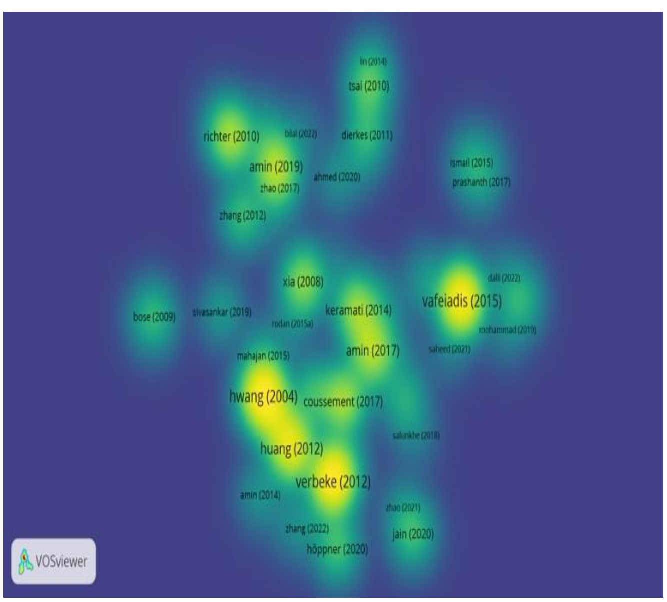

Figure 5: Density Visualization of the Source

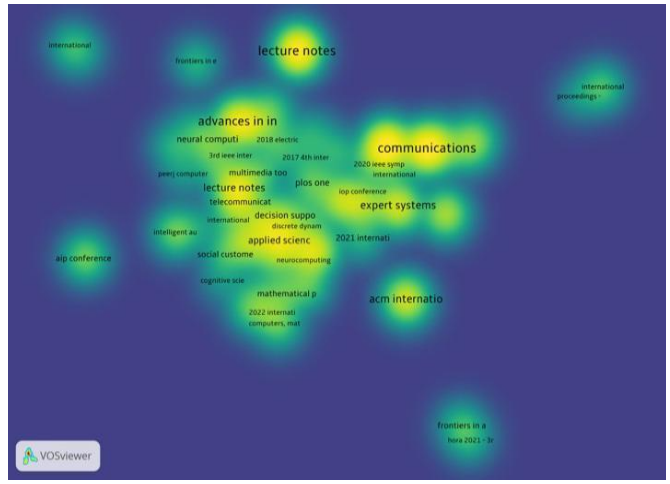

Figure 6: The Density of the Most Frequently Cited Studies

The dimension of the node represents the numb
er of times the keyword appears or the total number of publications where it seems. The lines show the relationships among the keywords, and the terms adjacent to one another are associated. The Scopus has six keywords: red, green, blue, yellow, light blue, and purple. The blue cluster applied several ML models such as "artificial neural network, "logistic regression," decagon tree," support vector machine," random forest," xgboost," and genetic algorithms."

In Figure 7, a network map shows the collaborations of leading academics in the field of ML-customer churn. The research studies were conducted in cooperation with these authors. Using classification techniques, the authors of the red and green clusters forecast client attrition. The dimension of the node represents the number of times the keyword appears or the total number of publications where it seems. The lines show the relationships among the keywords, and the terms adjacent to one another are associated. The Scopus has six keywords: red, green, blue, yellow, light blue, and purple. The blue cluster applied several ML models such as "artificial neural network, "logistic regression," decagon tree," support vector machine," random forest," xgboost," and genetic algorithms."

Figure 8 illustrates the papers' bibliographic connections to the Scopus database with at least one citation. Ten orange, red, green, blue, light blue, pink, brown, yellow, and purple clusters are used to group the documents in the Scopus database. The papers' bibliographic connection for the Scopus database with a minimum of one citation is shown in Figure 8. The link thickness represents the probability that pairs will co-occur, whereas the node size is associated with the paper's total link strength. The two nodes, or documents, are close to one another, indicating they have many citations in common [33].

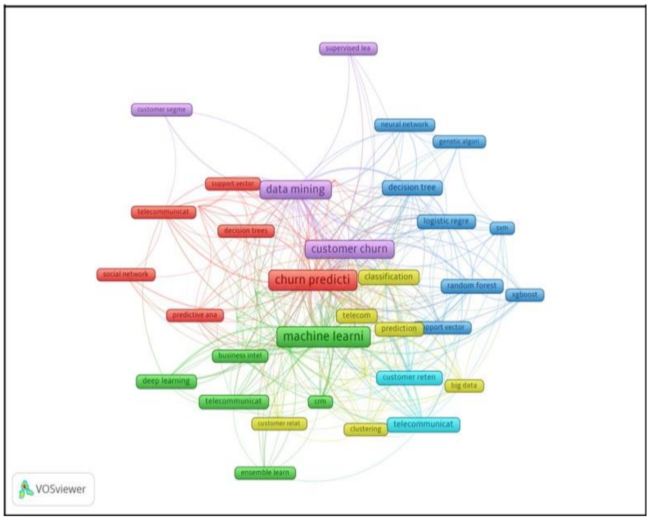

Figure 7: Keyword Network Map

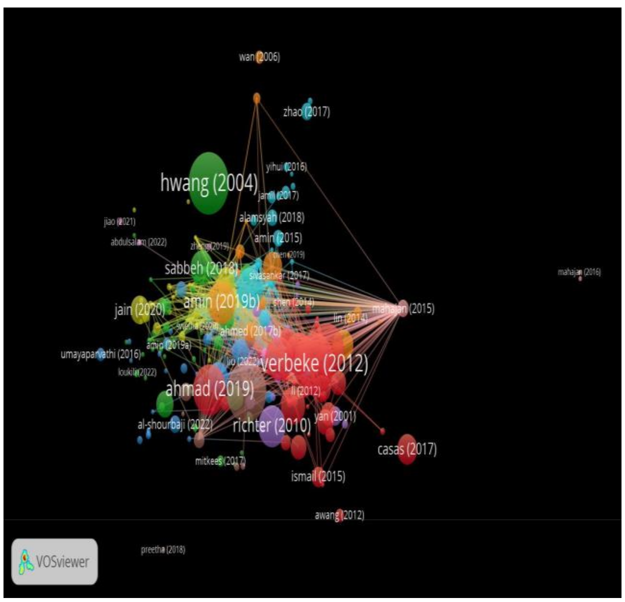

Figure 8: Bibliographic Coupling of Documents

The brown cluster in Figure 8 was by Vafeiadis et al. [34]. The paper presents a comparative analysis of the most widely used ML techniques to tackle the difficult task of predicting customer attrition in the telecom sector. The red cluster had a paper by Verbeke et al. [31] that conducts a novel approach to gauging the success of retention efforts by optimizing revenue from targeting high-churn clients. It also applied an experiment with benchmarking: a comparison of various classification methods utilizing statistical and profit-centric parameters on eleven telecom providers' data sets. Classification methods use statistical and profit-centric parameters on eleven telecom provider data sets.

The orange cluster contains the paper of Amin et al. [9]. The distance factor, which gauges the classifier's confidence for various data zones, is the basis for the paper's innovative approach to churn prediction. The study demonstrates that the approach obtains excellent accuracy for precise data. The green cluster is the paper by Hwang et al. [32]. This study suggests a novel method for estimating customer value using customer lifetime value (CLV), which considers a customer's likelihood of defecting, previous profit, and projected benefit. The article also presents a paradigm for evaluating and classifying clients based on their present, prospective, and loyalty value. It uses the framework and model of a wireless communication corporation as a case study.

CONCLUSION

Bibliographic analysis can help customer churn prediction researchers. The paper covers big data and ML customer churn predictions. The study's findings can help the researcher to create effective ML about customer retention plans and choices. Keyword analysis can reveal literature gaps by showing underrepresented or unused keywords. This analysis can help us propose research questions to fill gaps and expand knowledge. Ensemble, hybrid, reinforcement, and blended ML models have not been used to study telecom customer churn. Bibliometrics can cluster recent scientific gaps to find publishing country map gaps. This strategy can improve research by finding new research topics. The study found a lack of GCC customer churn studies. Future studies could investigate using other ML algorithms and techniques, such as hybrid, reinforcement, and blended ML models, to predict customer churn.

REFERENCES

[1] Hadden J, Tiwari A, Roy R, Ruta D (2007). **Computer-assisted customer churn management: state-of-the-art and future trends,** *Comput Operat Res*, Vol 34, No. 10, pp. 2902–2917.

[2] Kaya E, Dong X, Suhara Y, Balcisoy S, Bozkaya B (2018). **Behavioral attributes and financial churn prediction,** *EPJ Data Sci*, Vol. 7, No. 1, 41.

[3] Cil F, Cetinyokus T, Gokcen H (2018). **Knowledge discovery on investment fund transaction histories and socio-demographic characteristics for customer churn,** *Int J Intell Syst Appl Eng*, Vol. 6, No. 4, pp. 262–270.

[4] De Caigny A, Coussement K, De Bock KW, Lessmann S (2020). **Incorporating textual information in customer churn prediction models based on a convolutional neural network**, *Int J Forecast*, Vol. 36, No. 4, pp. 1563– 1578.

[5] Ballings, M, Van den Poel D (2012). **Customer event history for churn prediction: how long is long enough?** *Expert Syst Appl* Vol. 39, No. 18, pp. 13517–13522.

[6] Günther CC, Tvete IF, K, Sandnes GI, Borgan O (2014). **Modelling and predicting customer churn from an insurance company**, *Scand Actuar J*, Vol. 1, pp. 58–71.

[7] Perišić A, Jung DŠ, Pahor M (2022). **Churn in the mobile gaming field: establishing churn definitions and measuring classification similarities,** *Expert Syst Appl*, Vol. 191, 116277.

[8] Beeharry Y, Tsokizep Fokone R (2022). **Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry**, *Concur Comput: Pract Exp*, Vol. 34, No. 4, e6627.

[9] Amin A, Al-Obeidat F, Shah B, Adnan A, Loo J, Anwar S (2019). **Customer churn prediction in the telecommunication industry using data certainty**, *J Bus Res*, Vol. 94, pp. 290–301.

[10] Kandel IHA (2019). **A comparative study of tree-based models for churn prediction: a case study in the telecommunication sector**, *Master's Dissertation*, NOVA Information Management School, FL, USA.

[11] Coussement K, De Bock KW (2013). **Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning,** *J Bus Res*, Vol. 66, No. 9, pp. 1629–1636.

[12] Berger P, Kompan M (2019). **User modeling for churn prediction in E-commerce**, *IEEE Intell Syst*, Vol. 34, No. 2, pp. 44–52.

[13] S.K. Routray (2021), **Marketing strategy through machine learning techniques: A case study at telecom industry,** *International Journal of Innovation Engineering and Science Research*, Vol. 5, No. 3, pp. 21-30.

[14] Vrushabh Jinde and Savyanavar Amit (2020). **Customer churn prediction system using machine learning,** *International Journal of Advanced Science and Technology*, Vol. 29, No. 5, pp. 7957–7964.

[15] D. Bell, and C. Mgbemena (2018). **Data-driven agent-based exploration of customer behavior**, *Simulation*, Vol. 94, No. 3, pp. 195-212.

[16] Pustokhina, V. Irina, et al. (2021). **Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms**, *Information Processing & Management*, Vol. 58, No. 6, pp. 102706.

[17] Ascarza E, Neslin SA, Netzer O, Anderson Z, Fader PS, Gupta S, Schrift R (2018). **In pursuit of enhanced customer retention management: review, key issues, and future directions,** *Cust Needs Solut*, Vol. 5, No. 1, pp. 65–81.

[18] Krishna GJ, Ravi V (2016). **Evolutionary computing applied to customer relationship management: a survey,** *Eng Appl Artif Intell*, Vol. 56, pp. 30–59.

[19] Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K (2017). **Customer churn prediction in the telecommunication sector using a rough set approach,** *Neurocomputing*, Vol. 237, pp. 242– 254.

[20] Bhattacharyya J, Dash MK (2022). **What do we know about customer churn behavior in the telecommunication industry? A bibliometric analysis of research trends: 1985–2019**, *FIIB Bus Rev*, Vol. 11, No. 3, pp. 280–302.

[21] Lalwani, Praveen, et al. (2022). **Customer churn prediction system: a machine learning approach**, *Computing*, Vol. 104, No. 8, pp. 1-24.

[22] A.K. Ahmad, A. Jafar, K. Aljoumaa (2019). **Customer churn prediction in telecom using machine learning in big data platform**, *Journal of Big Data*, Vol. 6, No. 1, pp. 1-24.

[23] Wong K. (2010). **Fighting churn with rate plan right-sizing: a customer retention strategy for the wireless telecommunications industry**, *Serv Ind J*, Vol. 30, No. 13, pp. 2261–2271.

[24] Hosseini E, Rezvani MH (2021). **E-customer loyalty in gamified trusted store platforms: a case study analysis in Iran,** *Bull Electr Eng Inform* , Vol. 10, No. 5, pp. 2899–2909.

[25] J.Z. Zhang, and C.W. Chang (2021). **Consumer dynamics: Theories, methods, and emerging directions,** *Journal of the Academy of Marketing Science*, Vol. 49, pp. 166-196.

[26] Liu R, Ali S, Bilal SF, Sakhawat Z, Imran A, Almuhaimeed A, Sun G (2022). **An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms**, *ApplSci*, Vol. 12, No. 18, pp. 1-17.

[27] Pandey Y, Jha R (2022). **Customer churn analysis in telecom organization**, *J Posit School Psychol* Vol.6, No. 4, pp. 5475–5488.

[28] Ascarza, Eva, et al. (2018). **In pursuit of enhanced customer retention management: Review, key issues, and future direction**, *Customer Needs and Solutions*, Vol. 5, No. 17, pp. 65-81.

[29] Coussement, K.; Lessmann, S.; Verstraeten, G. (2017). **A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry**, *Decis. Support Syst.*, Vol. 95, pp. 27–36.

[30] Singh, C., Dash, M. K., Sahu, R., and Kumar, A. (2023). **Artificial intelligence in customer retention: a bibliometric analysis and future research framework,** *Kybernetes*, Vol. ahead-of-print, No. ahead-of-print.

[31] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., and Baesens, B. (2012). **New insights into churn prediction in the telecommunication sector: A profit-driven data mining approach**, *European Journal of Operational Research*, Vol. 218, No. 1, pp. 211–229.

[32] Hwang, H., Jung, T., and Suh, E. (2004). **An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry,** *Expert Systems with Applications*, Vol. 26, No. 2, pp. 181– 188.

[33] Marchiori, D., and Franco, M. (2020). **Knowledge transfer in the context of inter-organizational networks: Foundations and intellectual structures,** *Journal of Innovation and Knowledge*, Vol. 5, No. 2, pp. 130–139.

[34] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., and Chatzisavvas, K. Ch. (2015). **A comparison of machine learning techniques for customer churn prediction,** *Simulation Modelling Practice and Theory*, Vol. 55, pp. 1–9.

View publication stats