A COMPREHENSIVE BIBLIOMETRIC ANALYSIS OF CHURN PREDICTION RESEARCH: AN ESSAY ON TRENDS, KEY CONTRIBUTORS AND GLOBAL PARTICIPATION IN THE FIELD.

Zakariya, M. S. Mohammed[1], Faroug A. Abdalla[2], Mohyaldein Salih[1], Ashraf F. A. Mahmoud[2], Ali Satty[1*]
[1] Department of Mathematics, College of Science, Northern Border University, Arar, Saudi Arabia
[2] Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia
*Correspondence:(+966)561012510; alisatty1981@gmail.com

ABSTRACT: *Churn prediction is a common challenge across companies in many sectors. It has received great attention from researchers across various sectors due to its role in helping companies retain their customers. Bibliometric analysis identifies primary studies and researchers focusing on topics pertinent to churn prediction within the academic landscape. The main objective of this paper is to analyze the scientific landscape of churn prediction research using bibliometric analysis. The literature review of machine learning and statistical-based churn prediction methods is presented. Data were collected from the Web of Science database, comprising 517 articles on churn prediction published between 1999 and 2022. The extracted research was authored by 1,259 researchers across 66 countries and disseminated via 378 distinct sources. A total of 8,426 citations, resulting in an average of 16.3 citations per article were amassed. The analysis was conducted using the R bibliometrics library. Various bibliometric indicators were used, encompassing trend analyses, assessment of most cited publications, authorship analysis, co-authorship network examinations, evaluations of most prolific countries and sources along with h-index analyses, identification of most pertinent affiliations, and frequency of keywords usage. The results provide robust guidance for researchers and organizations to understand the historical trajectory and growth of churn prediction as a discipline, current research focal points, and areas warranting attention in future research.*

Keywords: Churn prediction; Prediction models; Customer churn; Customer retention; Bibliometric analysis

1. . INTRODUCTION

Customer churn presents a significant challenge for companies across various sectors, such as telecommunication industries where clients frequently switch between service providers [1]. It is defined as a customer's decision to discontinue their association with a service provider in favour of a new one [2]. As such, generating an effective and accurate churn prediction is a critical issue. According to Wang et al. [3], churn prediction refers to a methodology that identifies consumer behaviour and characteristics that indicate the timing of customers departing. The significance of churn prediction has led researchers to produce relevant scientific studies on the subject. Thus, a crucial task in advancing knowledge on churn prediction is the synthesis of previous research findings, which can be accomplished using bibliometric analysis methodology.

According to De Bellis [4] and Donthu et al. [5], bibliometric analysis is a computer-aided rigorous statistical method to explore and analyze large scientific publications. It can identify the research findings or authors along with their interconnections, about an entire collection of publications concerning a specific subject or discipline. Explicitly, it investigates the connection between various aspects including, but not limited to, disciplines, fields of knowledge, areas of expertise, and authors. Comprehensive details on bibliometric analysis can be acquired from references [6, 7, 8]. As discussed in Small [9], this analysis employs quantitative measures to describe, assess, and monitor published research, ascertaining its cognitive structure and development. As such, the analysis recognizes central research fields, providing researchers with a robust basis for situating significant current research contributions and discerning new directions for future research [10].

Bibliometric analysis has been utilized to analyze scientific productions in a variety of fields, such as machine learning [11], business [12], mathematics [13], and big data [14]. Despite the increase in research publications on churn prediction, there has been a distinct lack of comprehensive reviews focusing on the analysis of the scientific production within this field. To address this gap and introduce insights into how the research has approached churn prediction, this article employs a bibliometric analysis. The main objective of this article is to analyze the scientific landscape of churn prediction research using bibliometric analysis. Specifically, this article is carried out to enable researchers to efficiently comprehend the historical developments as well as future research avenues in the field of churn prediction. The importance of this study lies in its ability to provide researchers with valuable direction for future research, and for companies as well, by emphasizing the most eminent research contributions to the customer retention process. This study has only restricted attention to consider the Web of Science database due to its quality in publishing scientific research [15, 16].

The remainder of this article is arranged as follows. Section 2 introduces the methodology used in this research, encompassing data sources and search strategy as well as the data analysis technique. Section 3 presents the obtained results from the extracted data. Detailed discussions of the results are given in Section 4. Finally, Section 5 offers conclusive insights for future research.

2. LITERATURE REVIEW

2.1 Machine learning churn prediction-based methods

The literature reviews several methodologies for dealing with the churn prediction field. Machine learning methodologies have been widely used for customer churn prediction. These methods have shown to be successful in detecting customers who pose a risk. In what follows, we review various machine learning-related churn prediction methodologies.

Ning Lu [17] proposed dividing consumers into two clusters according to the weights supplied by the boosting algorithm before applying boosting methods to improve a churn prediction model. Benlan He [18] proposed a methodology for churn prediction using a support vector machine (SVM) model and random sampling to correct imbalances in customer datasets and enhance the performance of the SVM model. A few ensemble methods based on rotation-based techniques were proposed by Koen W. De Bock [19]. These techniques were offered as a means of balancing interpretability and performance in customer churn prediction models. To create precise and intelligible classification rule-based customer churn prediction models, Wouter Verbeke [20] recommended applying the Ant-Miner+ and ALBA algorithms on a publicly accessible customer churn dataset. Lee et al. [21] created a clear and precise predictive model. They employed a technique known as partial least squares, which works well with datasets in which the variables have a strong correlation with one another. The objective was to develop a model that could forecast whether or not a customer would discontinue doing business with a company. Koen [22] developed a novel type of predictive method for customer churn prediction. It is based on generalized additive models and is known as GAMensPlus. This method assesses understanding and reconciling performance for the churn prediction modelling problem. To improve the accuracy of their churn prediction model, Ning et al. [23] investigated the use of boosting in the prediction of customer churn. To sum up, they experimented with several methods to predict customer churn and, based on the findings, recommended boosting as a means of refining their churn prediction model. Pendharkar [24] developed two distinct approaches to predict customer churn by using neural networks and genetic algorithms to identify which customers are most likely to discontinue their relationship with a provider. Thus, it was a method of predicting customer churn by merging those two approaches. Predicting whether a consumer would discontinue doing business with a company is the goal of the approaches. On a big data platform, customer churn was examined by Huang et al. [25]. They focused on how big data may significantly enhance the process of predicting customer churn. Their study aimed to demonstrate how big data, which consists of a large volume of rapidly arriving data, improves churn prediction by providing an abundance of available information. Vafeiadis et al.'s study [26] examined various machine-learning techniques for predicting customer churn. They contrasted several classification models, including neural networks and decision trees. They examined boosted versions of the models as well as simple versions, combining or stacking the models in an attempt to increase accuracy. Wagh et al. [27] suggested a new methodology for predicting customer churn through the use of various machine learning algorithms. They used popular algorithms like random forests, K-nearest neighbours, and decision trees to process customer turnover data. The objective was to determine the most effective method for identifying the customers who would ultimately migrate to a different service provider. Samira and Zivari [28] investigated several machine-learning techniques for predicting customer churn. In essence, they were attempting to determine which customers could terminate their subscriptions. The framework they developed consisted of six steps. The main idea was to figure out a way to identify those customers who were most likely to cancel so the business could try to protect them from going out of business. A methodology was suggested by Abdelrahim and Assef [29] to assist telecom businesses in predicting customer churn. This methodology makes predictions using machine learning on big data. Four machine learning algorithms have been tested. A churn prediction model was presented by Irfan et al. [30] that identifies churn customers and offers the factors that contribute to customer churn in the telecom industry. This has been achieved using classification and clustering algorithms. Mohammed [31] determined the factors that affect customer churn, created a useful model for predicting customer churn, and offered the finest analysis of data visualization outcomes. Several machine learning methods were selected, such as Random Forest and Artificial Neural Network.

2.2 Statistical-based churn prediction methods

Statistical-based methods are commonly employed in churn prediction. In the following, we present the most commonly used statistical methodologies for churn prediction domain.

Among the methodologies commonly employed for churn prediction, logistic regression stands out as the predominant choice. Its widespread adoption is attributed to its simplicity, interpretability, resilience, and popularity among marketers and customer churn modellers [32, 33, 34]. Additionally, in scenarios where churn occurrence is infrequent, logistic regression maintains its dominance [35].

Pareto/Negative Binomial Distribution (Pareto/NBD) and beta-geometric/NBD (BG/NBD) models were also utilized for churn prediction. The appeal of Pareto/NBD and BG/NBD lies in their capacity to (1) leverage past transactional behaviour to construct the model, (2) predict an individual's future purchase levels, and (3) estimate the likelihood that a specific customer remains active after a certain duration. These characteristics have positioned these two models as the most renowned and recommended stochastic approaches for identifying customer churn and projecting future sales in non-contractual environments [36, 37, 38].

Survival analysis and hazard function are utilized to explore customers who exhibit a high likelihood of churning and the timing of their potential churn. Survival analysis assesses the probability of a customer's survival following an observation period,

while the hazard function is employed to anticipate whether a customer will churn within a specific timeframe [39, 40, 41, 42, 43]

3. METHODOLOGY

3.1 Data extraction and search strategy

Data on scientific publications for churn prediction were extracted from the Web of Science database ("Clarivate Analytics"). The keyword used for the search is “churn prediction”. A comprehensive search within the Web of Science database revealed 517 publications on churn prediction, spanning from 1999 to 2022. Solely English-language articles were taken into account. A total of 517 scientific publications (articles, proceeding papers, early access, correction and meeting abstracts) published from 2001 to 2022 were returned. These publications were authored by 1,259 researchers across 66 countries and disseminated via 378 distinct sources, such as journals and books. In addition, these publications amassed a total of 8,426 citations, resulting in an average of 16.3 citations per article. The analysis of the retrieved documents identified five primary document types. Among these, articles were the most prevalent with 255 publications, accounting for 49.3% of the overall production. Subsequently, proceedings papers constituted 48.2%, represented by 249 documents. Other document categories present in the search results comprised reviews, early access papers, meeting abstracts, and corrections, collectively amounting to 13 (2.5%).

3.2 Data analysis

As mentioned earlier, bibliometric analysis can help us earn insight into published research in many fields and can therefore be used for the same purpose in the case of churn prediction. In this article, it is utilized to explore and analyze a vast array of scientific publications on churn prediction using the bibliometric R-package proposed by Aria et al. [45] who stated that this package offers an assortment of tools designed for quantitative research. Various bibliometric indicators were used, encompassing trend analyses, assessment of most cited publications, authorship analysis, co-authorship network examinations, evaluations of most prolific countries and sources along with h-index analyses, identification of most pertinent affiliations, and frequency of keywords usage.

4. RESULTS

4.1 Trends analysis

As illustrated in Figure 1, there has been a growing interest in the field of churn prediction research since 2014, consistently resulting in over 25 publications annually. Notably, the pinnacle was reached in 2022 with a total of 77 publications, surpassing the figures of 53 in 2021, 52 in 2017, and 48 in 2018

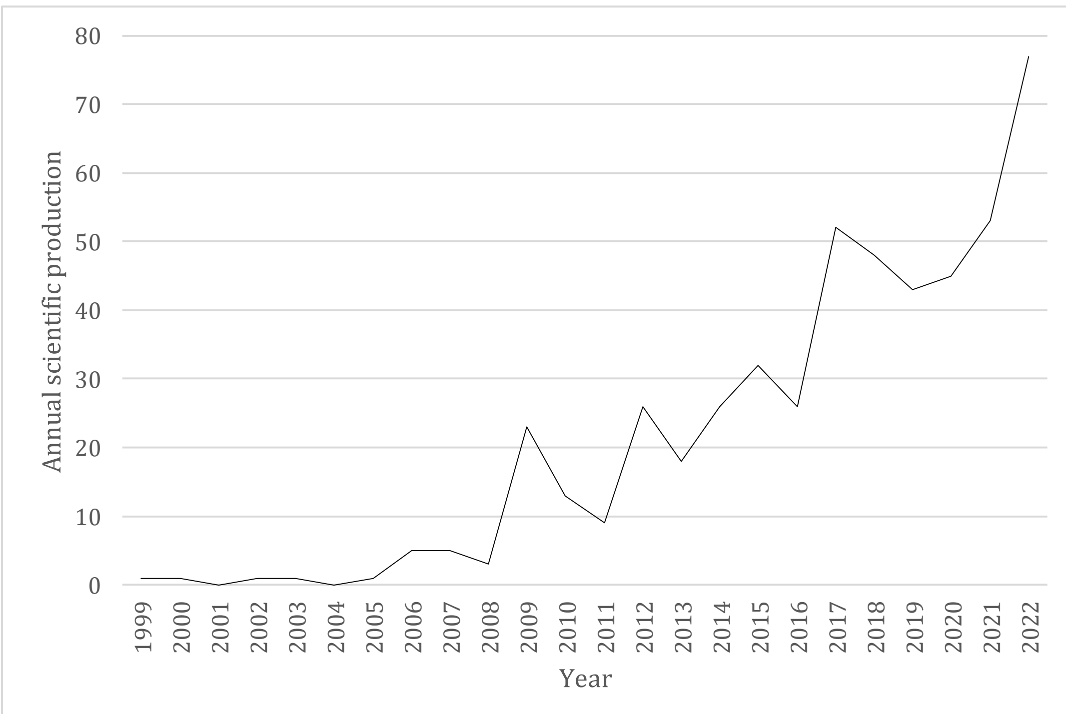

Figure 1: Trend in publications on churn prediction.

4.2 Most cited and influential publications

As depicted in Table 1, the top 10 most frequently referenced papers in the field of churn prediction are displayed. The total number of citations (Global citations) for these papers ranges between 149 and 272. The most cited paper, amassing 272 citations, was "Handling class imbalance in customer churn prediction" authored by Burez, J and Van den Poel, D.

TABLE 1: Top cited articles.

Paper title	Authors	DOI	Total citation	Year
Handling class imbalance in customer churn prediction	Burez, J and Van den Poel, D	10.1016/j.eswa.2008.05.027	272	2009
Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques	Coussement, K and Van den Poel, D	10.1016/j.eswa.2006.09.038	242	2008
Turning telecommunications call details to churn prediction: a data mining approach	Wei, CP and Chiu, IT	https://doi.org/10.1016/S0957-4174(02)00030-1	240	2002
New insights into churn prediction in the telecommunication sector: A profit-driven data mining approach	Verbeke, W; Dejaeger, K; Martens, D; Hur, J and Baesens, B	10.1016/j.ejor.2011.09.031	221	2012
Customer churn prediction using improved balanced random forests	Xie, YY; Li, X; Ngai, EWT and Ying, WY	10.1016/j.eswa.2008.06.121	204	2009
A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees	De Caigny, A; Coussement, K and De Bock, KW	10.1016/j.ejor.2018.02.009	202	2018
A novel evolutionary data mining algorithm with applications to churn prediction	Au, WH; Chan, KCC and Yao, X	10.1109/TEVC.2003.819264	192	2003
A comparison of machine learning techniques for customer churn prediction	Vafeiadis, T; Diamantaras, KI; Sarigiannidis, G and Chatzisavvas, KC	10.1016/j.simpat.2015.03.003	179	2015
Building comprehensible customer churn prediction models with advanced rule induction techniques	Verbeke, W; Martens, D; Mues, C and Baesens, B	10.1016/j.eswa.2010.08.023	169	2011
Customer churn prediction in telecommunications	Huang, BQ; Kechadi, MT and Buckley, B	10.1016/j.eswa.2011.08.024	149	2012

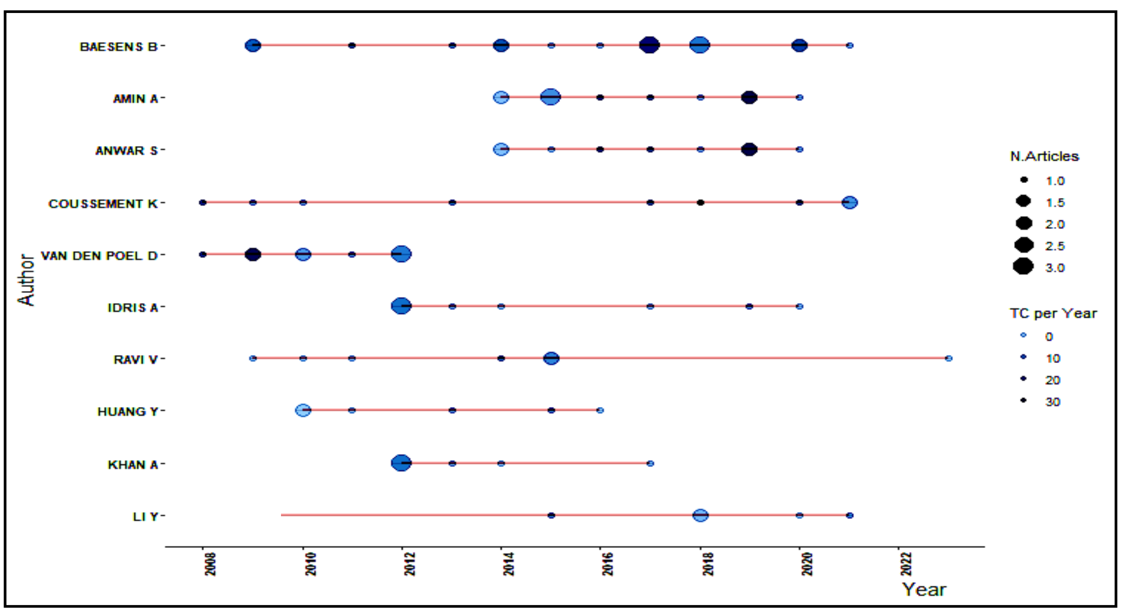

Figure 2: Author's production over time.

4.3 Authorship analysis

Table 2 illustrates the most distinguished and impactful authors in the field of churn prediction. The leading and most productive authors in the field of churn prediction were Baesens Bart with 17 publications, followed by Amin Adnan with 11 publications as well as Coussement Kristof, Van Den Poel and Anwar Sajid, each contributing 9 publications, ranking third. In addition, Figure 2 presents the author's production over time.

TABLE 2: Most influential authors.

Authors	Number of articles	H-index	G-index	Total citations	Year of start
Baesens Bart	17	12	17	771	2009
AMIN Adnan	11	9	11	444	2014
Anwar Sajid	9	8	9	405	2014
Coussement Kristof	9	8	9	834	2008
VAN DEN POEL D	9	7	9	813	2008
Idris Adnan	8	7	8	186	2012
Ravi Vadlamani	7	4	7	170	2009
Verbeke Wouter	6	6	6	608	2011
Khan Asifullah	6	5	6	153	2012
Oskarsdottir Maria	6	4	6	93	2017
LI Yong	6	3	6	92	2009
Qi Huang Jia	6	3	6	45	2006
Huang Yiqing	6	3	6	119	2010

TABLE 3: Co-authorship analysis.

From	To	Frequency
BELGIUM	UNITED KINGDOM	22
CHINA	USA	13
CHINA	UNITED KINGDOM	6
PAKISTAN	SAUDI ARABIA	6
PAKISTAN	UNITED KINGDOM	6
PAKISTAN	U ARAB EMIRATES	5
BELGIUM	FRANCE	4
PAKISTAN	KOREA	4
PAKISTAN	USA	4
CHINA	SAUDI ARABIA	3
UNITED KINGDOM	SAUDI ARABIA	3
UNITED KINGDOM	U ARAB EMIRATES	3
USA	CANADA	3
USA	TURKEY	3

4.4 Co-authorship network analysis

According to the information gathered from the Web of Science, a total of 1,556 authors contributed to the referenced documents. Table 3 illustrates a co-authorship analysis that included only authors who have published a minimum of three publications

4.5 Most Productive Countries

As per Table 4, among the 66 countries contributing to research publications in churn prediction, China has taken the lead with 91 papers (18%), followed by India with 82 publications (16%) and Belgium with 32 publications (7%). The countries of the corresponding authors are depicted in Figure 3

.

TABLE 4: Top most productive and cited countries.

Country	Number of articles	Percentage	Total citations	Average article citations	Single country publications (SCP)	Multiple country publications (MCP)	MCP Ratio
CHINA	91	0.18	1514	16.6	76	15	0.16
INDIA	82	0.16	658	8	78	4	0.05
BELGIUM	32	0.07	1859	58.1	12	20	0.63
PAKISTAN	32	0.07	781	24.4	16	16	0.50
TURKEY	24	0.05	155	6.5	19	5	0.21
USA	23	0.05	353	15.3	19	4	0.17
KOREA	20	0.04	234	11.7	14	6	0.30
IRAN	19	0.04	192	10.1	18	1	0.05
UNITED KINGDOM	14	0.03	206	14.7	13	1	0.07
FRANCE	12	0.02	531	44.2	6	6	0.50

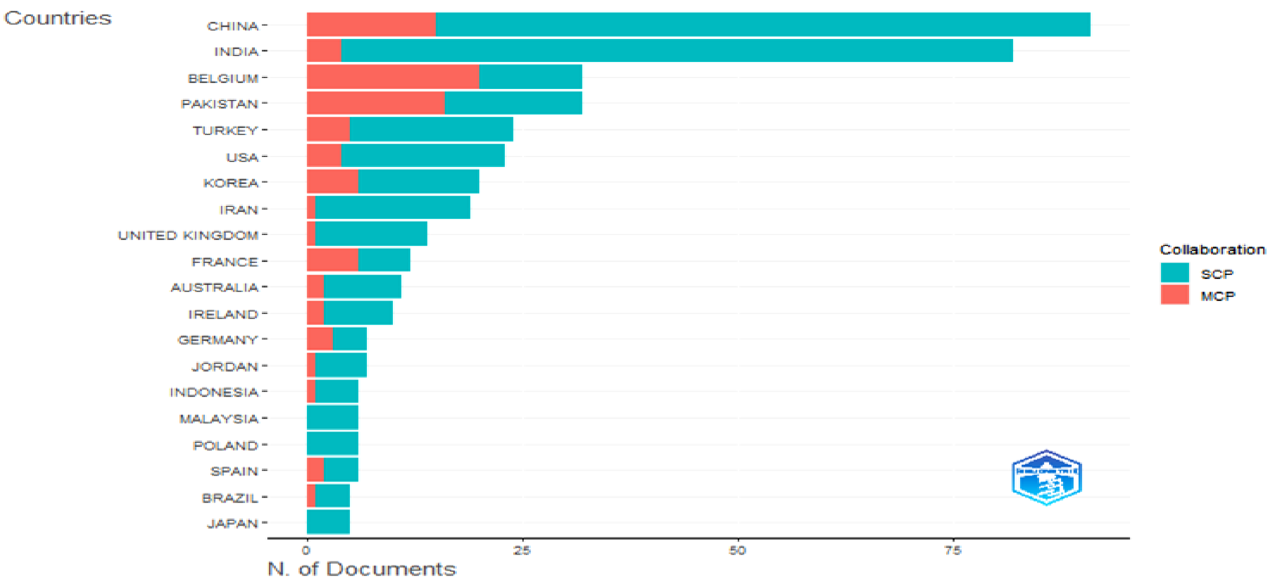

Figure 3: Corresponding author's countries

4.6 Sources and h-index analysis

As displayed in Table 5, in total, articles on churn prediction were published in 378 distinct journals. The leading journal was "Expert Systems with Applications" (h-index = 25) with 32 publications, followed by "The International Journal of Advanced Computer Science and Applications" (h-index = 4) consisting of 9 publications, and "The European Journal of Operational Research" (h-index = 8) having 8 publications.

TABLE 5: Top most relevant sources on churn prediction

Sources	Number of articles	H-index	Total citations	Year of start
EXPERT SYSTEMS WITH APPLICATIONS	32	25	2438	2002
INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS	9	4	40	2011
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH	8	8	693	2009
APPLIED SOFT COMPUTING	7	6	354	2014
NEURAL COMPUTING & APPLICATIONS	7	5	75	2018
IEEE ACCESS	6	4	243	2016
CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS	4	4	84	2017
DECISION SUPPORT SYSTEMS	4	4	193	2012
KNOWLEDGE-BASED SYSTEMS	4	4	103	2012
IEEE TRANSACTIONS ON GAMES	4	3	31	2019
JOURNAL OF BUSINESS RESEARCH	4	3	189	2013
APPLIED SCIENCES-BASEL	4	2	19	2021

4.7 Most relevant affiliations

Table 6 shows the affiliations associated with churn prediction publications between the years 1999 and 2022. Katholieke University Leuven emerged as the leader in this field, with 49 publications. The University of Southampton followed in second place, contributing 24 publications, and both Institute of Management Science and University College Dublin were ranked third, each having 18 publications

.

TABLE 6: Most relevant affiliations

Affiliation	Country	Number of publications	Percentage
KATHOLIEKE UNIVERSITY LEUVEN	Belgium	49	9.5
UNIVERSITY OF SOUTHAMPTON	United Kingdom	24	4.6
INSTITUTE OF MANAGEMENT SCIENCE	Pakistan	18	3.4
UNIVERSITY COLLEGE DUBLIN	Ireland	18	3.4
BEIJING UNIVERSITY OF POSTS AND TELECOMMUN	China	17	3.2
GHENT UNIVERSITY	Belgium	16	3.0
NATIONAL INSTITUTE TECHNOLOGY	India	14	2.7
TSINGHUA UNIVERSITY	China	14	2.7
UNIVERSITY OF TEHRAN	Iran	14	2.7
SICHUAN UNIVERSITY	China	13	2.5

4.8 Most frequent keywords

The word cloud, as shown in Figure 4, demonstrates the prevalence of the author's keywords. The word "models" was the most prevalent word with 50 frequencies, succeeded by "classification" (47) and "retention" (46). On the other hand, Table 7 introduces the first (Q1), median and third (Q3) quartiles per year, showing the trend topics concerning the churn prediction field. According to this table, the emerging topics were "model" and "algorithm" with the third quartile year being 2022 followed by "customer churn" and "retention" with the third quartile year being 2021

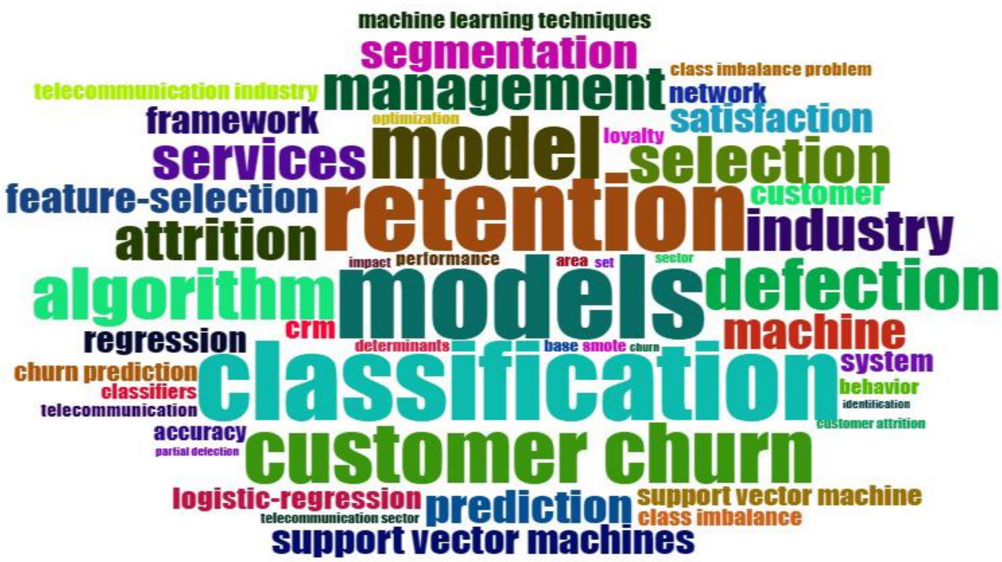

Figure 4: Most frequent keywords

TABLE 7: Trend topics

Item	Frequency	Q1	Median	Q3
Models	50	2014	2016	2020
Classification	47	2014	2018	2020
Retention	46	2012	2020	2021
Model	37	2018	2021	2022
Customer churn	36	2016	2018	2021
Defection	31	2012	2014	2019
Algorithm	31	2018	2019	2022
Selection	28	2012	2014	2019
Management	26	2012	2017	2020
Services	25	2012	2015	2017

5. DISCUSSION

This study conducted a bibliometric analysis of publications about churn prediction within the Web of Science database, spanning from 1999 to 2022. In order to obtain a comprehensive understanding of the research distribution within the field of churn prediction, the evaluation identifies trends in churn prediction research, most frequently cited research, most active authors, leading countries in productivity, most relevant affiliations, and most frequent words related to churn prediction.

Upon Examining trend analysis concerning churn prediction, it became evident that there has been a marked increase in publication volume on this subject matter in 2022, indicative of the heightened attention from the scientific community towards this field. Moreover, retrieved documents reveal an appreciable escalation in annual scientific production on churn prediction commencing from the year 2014. As such, it can be anticipated that this expansion of literature will persist into future years.

Monitoring the extent of journals' coverage or the frequency of citations for journal articles is a common method for evaluating publication quality. The most highly cited journal article was entitled "Handling Class Imbalance in Customer Churn Prediction", which ranked first among the top 10 most referenced journal articles in the field of churn prediction. The findings also indicated that Basens Bart was the author with the most publications and highest citation rate in the churn prediction field.

The co-authorship network analysis results introduced that Belgium to the United Kingdom had the largest number of co-authors in the field of churn prediction. This result demonstrates that these countries have prominent existence and superiority among authors concerning churn prediction. Furthermore, the results explained that Belgium held the prime position in terms of citation on churn prediction, with France following closely behind as a secondary country [46].

Upon examining the results of the most productive countries, it was observed that China emerged as the leading and most cited country within the field of churn prediction. These findings are consistent with reported results in (“SJR”, “National Science Board”).

An alternative method for evaluating the quality of publication citations in journals involves examining total citation counts. In this context, the results revealed that "Expert Systems with Applications" stands as a significant contributor to Web of Science indexed publications, holding the record for the highest quantity of articles published in churn prediction.

Moreover, the results exhibited that "models" represented the most frequent word, followed by "classification" as a close second. The emerging topics were led by "model" and "algorithms". These results indicate the increase in scientific community interest in modelling and quantifying the likelihood of churn as well as in utilizing computer-based methods and algorithms in predicting churn [49].

Generally, the results of this study explained the impact of research on churn prediction, encompassing contributions from over 517 distinct authors spanning diverse geographical locations. It also presented more comprehensive views on the churn prediction research. Furthermore, the study acts as an extension to previous studies while introducing an alternative perspective. Such a study presents a summarized view for future research to evaluate the field of churn prediction more thoroughly, by offering a wider understanding of the bibliometric analysis of research trends associated with churn prediction.

Conclusion

In this paper, the application of bibliometric analysis in churn prediction has been thoroughly examined. The main objective was to analyze the scientific landscape of churn prediction research using bibliometric analysis. The intention was to critically evaluate previous research on churn prediction in order to establish a solid foundation for further exploration as well as to introduce insights into how the research has approached churn prediction. A comprehensive search of electronic databases in Web of Science was conducted, considering studies published from 1999 to 2022. The findings of this study, in general, yielded valuable insights for both researchers and companies by emphasizing the most significant research contributions within the churn prediction field. Based on these findings, researchers and organizations can employ the documented research and cited authors as reference points for understanding the historical trajectory and growth of churn prediction as a discipline, current research focal points, and areas warranting attention in future research.

While our research offers numerous advantages, certain limitations necessitate further research in the field of churn prediction. This study only utilizes information indexed within the Web of Science. Consequently, churn prediction literature found in other databases is not encompassed herein. Future research should consider examining databases such as Scopus (https://www.scopus.com) and Google Scholar (https://scholar.google.com) alongside Web of Science. Additionally, as the analysis has confined its scope to English-language publications, there is an inherent bias present. It would be beneficial for future studies to consider the inclusion of additional languages for a more comprehensive analysis.

REFERENCES

1. Lalwani P, Mishra MK, Chadha JS, Sethi P. Customer churn prediction system: a machine learning approach. *Computing*. 2022 Feb 1:1-24.https://doi.org/10.1007/s00607-021-00908-y
2. Kumar V, Leszkiewicz A, Herbst A. Are you back for good or still shopping around? Investigating customers' repeat churn behaviour. *Journal of Marketing Research*. 2018 Apr;55(2):208-25.https://doi.org/10.1509/jmr.16.0623
3. Wang C, Han D, Fan W, Liu Q. Customer churn prediction with feature embedded convolutional neural network: An empirical study in the internet funds industry. *International journal of computational intelligence and applications*. 2019 Mar 15;18(01):1950003.https://doi.org/10.1142/S1469026819500032

4. De Bellis N. Bibliometrics and citation analysis: from the science citation index to cybermetrics. *scarecrow press*; 2009 Mar 9.

5. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*. 2021 Sep 1;133:285-96.https://doi.org/10.1016/j.jbusres.2021.04.070

6. Zhang J, Yu Q, Zheng F, Long C, Lu Z, Duan Z. Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the association for information science and technology*. 2016 Apr;67(4):967-72. https://doi.org/10.1002/asi.23437

7. Zupic I, Čater T. Bibliometric methods in management and organization. *Organizational research methods*. 2015 Jul;18(3):429-72. https://doi.org/10.1177/1094428114562629

8. Akintunde TY, Musa TH, Musa HH, Musa IH, Chen S, Ibrahim E, Tassang AE, Helmy MS. Bibliometric analysis of global scientific literature on effects of COVID-19 pandemic on mental health. *Asian journal of psychiatry*. 2021 Sep 1;63:102753. https://doi.org/10.1016/j.ajp.2021.102753

9. Small H. Visualizing science by citation mapping. *Journal of the American society for Information Science*. 1999;50(9):799-813. https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

10. Ferreira FA. Mapping the field of arts-based management: Bibliographic coupling and co-citation analyses. Journal of Business Research. 2018 Apr 1;85:348-57. https://doi.org/10.1016/j.jbusres.2017.03.026

11. Ahmed S, Alshater MM, El Ammari A, Hammami H. Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*. 2022 Oct 1;61:101646. https://doi.org/10.1016/j.ribaf.2022.101646

12. Donthu N, Kumar S, Pattnaik D. Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of business research*. 2020 Mar 1;109:1-4. https://doi.org/10.1016/j.jbusres.2019.10.039

13. Tsilika K. Exploring the Contributions to Mathematical Economics: A Bibliometric Analysis Using Bibliometrix and VOSviewer. *Mathematics*. 2023 Nov 20;11(22):4703. https://doi.org/10.3390/math11224703

14. Liao H, Tang M, Luo L, Li C, Chiclana F, Zeng XJ. A bibliometric analysis and visualization of medical big data research. Sustainability. 2018 Jan 11;10(1):166. https://doi.org/10.3390/su10010166

15. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*. 2021 Jun;126:5113-42. https://doi.org/10.1007/s11192-021-03948-5

16. Birkle C, Pendlebury DA, Schnell J, Adams J. Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*. 2020 Feb 1;1(1):363-76. https://doi.org/10.1162/qss_a_00018

17. Lu N, Lin H, Lu J, Zhang G. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*. 2012 Oct 12;10(2):1659-65. 10.1109/TII.2012.2224355

18. He B, Shi Y, Wan Q, Zhao X. Prediction of customer attrition of commercial banks based on SVM model. *Procedia computer science*. 2014 Jan 1;31:423-30. https://doi.org/10.1016/j.procs.2014.05.286

19. De Bock KW, Van den Poel D. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*. 2011 Sep 15;38(10):12293-301. https://doi.org/10.1016/j.eswa.2011.04.007

20. Verbeke W, Martens D, Mues C, Baesens B. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert systems with applications*. 2011 Mar 1;38(3):2354-64. https://doi.org/10.1016/j.eswa.2010.08.023

21. Lee H, Lee Y, Cho H, Im K, Kim YS. Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decision Support Systems*. 2011 Dec 1;52(1):207-16. https://doi.org/10.1016/j.dss.2011.07.005

22. De Bock KW, Van den Poel D. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications*. 2012 Jun 15;39(8):6816-26. https://doi.org/10.1016/j.eswa.2012.01.014

23. Lu N, Lin H, Lu J, Zhang G. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*. 2012 Oct 12;10(2):1659-65. 10.1109/TII.2012.2224355

24. Pendharkar PC. Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications*. 2009 Apr 1;36(3):6714-20. https://doi.org/10.1016/j.eswa.2008.08.050

25. Huang Y, Zhu F, Yuan M, Deng K, Li Y, Ni B, Dai W, Yang Q, Zeng J. Telco churn prediction with big data. InProceedings of the 2015 *ACM SIGMOD international conference on management of data* 2015 May 27 (pp. 607-618). https://doi.org/10.1145/2723372.2742794

26. Vafeiadis T, Diamantaras KI, Sarigiannidis G, Chatzisavvas KC. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*. 2015 Jun 1;55:1-9. https://doi.org/10.1016/j.simpat.2015.03.003

27. Wagh SK, Andhale AA, Wagh KS, Pansare JR, Ambadekar SP, Gawande SH. Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*. 2024 Mar 1;14:100342. https://doi.org/10.1016/j.rico.2023.100342

28. Khodabandehlou S, Zivari Rahman M. Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*. 2017 Mar 13;19(1/2):65-93.https://doi.org/10.1108/JSIT-10-2016-0061

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

30. Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW. A churn prediction model using random forest: analysis of

machine learning techniques for churn prediction and factor identification in telecom sector. IEEE access. 2019 May 6;7:60134-49. 10.1109/ACCESS.2019.2914999

31. Mohammad NI, Ismail SA, Kama MN, Yusop OM, Azmi A. Customer churn prediction in telecommunication industry using machine learning classifiers. *InProceedings of the 3rd international conference on vision, image and signal processing* 2019 Aug 26 (pp. 1-7).https://doi.org/10.1145/3387168.3387219
32. Buckinx W, Van den Poel D. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European journal of operational research*. 2005 Jul 1;164(1):252-68. https://doi.org/10.1016/j.ejor.2003.12.010
33. Naz NA, Shoaib U, Sarfraz MS. A review on customer churn prediction data mining modeling techniques. *Indian Journal of Science and Technology*. 2018 Jul;11(27):1-27.10.17485/ijst/2018/v11i27/121478, July 2018
34. Lalwani P, Mishra MK, Chadha JS, Sethi P. Customer churn prediction system: a machine learning approach. *Computing*. 2022 Feb 1:1-24. https://doi.org/10.1007/s00607-021-00908-y
35. Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of marketing research*. 2006 May;43(2):204-11.https://doi.org/10.1509/jmkr.43.2.204
36. Wübben M, Wangenheim FV. Instant customer base analysis: Managerial heuristics often “get it right”. *Journal of Marketing*. 2008 May;72(3):82-93.https://doi.org/10.1509/jmkg.72.3.82
37. Jerath K, Fader PS, Hardie BG. New perspectives on customer “death” using a generalization of the Pareto/NBD model. *Marketing Science*. 2011 Sep;30(5):866-80.https://doi.org/10.1287/mksc.1110.0654
38. Castéran H, Meyer-Waarden L, Reinartz W. Modeling customer lifetime value, retention, and churn. *InHandbook of market research* 2021 Dec 3 (pp. 1001-1033). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57413-4_21
39. Mohammed Z, Kotze D. Survival data mining in the telecommunications industries: usefulness and complications. *WIT Transactions on Information and Communication Technologies*. 2005 May 4;35. 10.2495/DATA050501
40. Mohammed Z, Maritz S, Kotze D. Estimation of the customer mean survival time in subscription-based businesses. Data mining, text mining and their business applications, *Wessex Institute of Technology Transaction on Information and Communication Technologies*. 2007;38:285-92.
41. Mohammed Z, Maritz JS, Kotze D. Customer survival time in subscription-based businesses (case of Internet service providers). Data mining, text mining and their business applications, *Wessex Institute of Technology Transaction on Information and Communication Technologies*. 2007;38:303-10.
42. Viljanen M, Airola A, Pahikkala T, Heikkonen J. Modelling user retention in mobile games. *In2016 IEEE Conference on Computational Intelligence and Games (CIG)* 2016 Sep 20 (pp. 1-8). IEEE. https://doi.org/10.1109/CIG.2016.7860393
43. Routh P, Roy A, Meyer J. Estimating customer churn under competing risks. *Journal of the Operational Research Society*. 2021 May 4;72(5):1138-55. https://doi.org/10.1080/01605682.2020.1776166
44. Clarivate Analytics. Web of Science master journal list, https://mjl.clarivate.com/home (2023, accessed 2 December 2023).
45. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informatics*. 2017 Nov 1;11(4):959-75. https://doi.org/10.1016/j.joi.2017.08.007
46. Kwiek M. What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. Studies in Higher Education. 2021 Dec 2;46(12):2629-49. https://doi.org/10.1080/03075079.2020.1749254
47. SJR - International Science Ranking. www.scimagojr.com. Retrieved 2023-12-30.
48. National Science Board, National Science Foundation. 2021. Publications Output: U.S. and International Comparisons. Science and Engineering Indicators 2022. NSB-2021-4. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20214/ .
49. Geiler L, Affeldt S, Nadif M. A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*. 2022 Sep;14(3):217-42. https://doi.org/10.1007/s41060-022-00312-5