International Journal of Innovative Research and Scientific Studies, 8(4) 2025, pages: 1189-1201

IJIRSS International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

A Bibliometric literature review on telecommunication customer churn

Lucy Yvonne Mathebula[1*], Modisane Seitshiro[2,3], Martin Chanza[1,3], Anupkumar Bongale[4]

[1]*Department of Business Statistics and Operations Research, North-West University, Mahikeng, South Africa.*
[2]*Centre for Business Mathematics and Informatics, North-West University, Potchefstroom, 2531, South Africa.*
[3]*National Institute for Theoretical and Computational Sciences (NITheCS), Potchefstroom, 2531, South Africa.*
[4]*Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Symbiosis International Deemed University, Pune 412115, Maharashtra, India.*

Corresponding author: Lucy Yvonne Mathebula (*Email: lucyy.mathebula@gmail.com*)

Abstract

This article aims to identify the most impactful research on customer churn in the telecommunications industry and assess the intellectual development, the characteristics of authors, and research trends in this industry. A bibliometric analysis was performed on 282 publications from 1999 to 2024, sourced from the Web of Science (WoS) database. Applied bibliometric techniques include co-word analysis, co-citations, bibliographic coupling, and co-authorship networks. The bibliometric analysis was conducted using the Biblioshiny package in R. The results indicated that the terms "customer churn," "churn management," and "churn prediction" were commonly used by the majority of churn researchers. The University of Southampton was identified as the affiliation with the highest number of published articles over time. China, India, and the USA emerged as the leading countries in terms of corresponding authors and the volume of published articles. Churn studies' most frequently used models include logistic regression, random forests, and various machine-learning techniques.

Keywords: Bibliometric analysis, Customer churn, Logistic regression, Telecommunication.

DOI: 10.53894/ijirss.v8i4.8045
Funding: This study received no specific financial support.
History: Received: 18 April 2025 / **Revised:** 23 May 2025 / **Accepted:** 26 May 2025 / **Published:** 24 June 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Predicting customer churn is vital for the telecommunications industry, as it can significantly impact market share [1]. Good prediction models will enable telecommunication companies to retain their good customers and attract new ones from other service providers. Churn in the telecommunication industry is defined as when a customer cancels their telecommunication service or their account is closed due to defaulting on their payments [2]. In banking, churn is defined as when a customer does not perform any transactions within a period of observed payments [2]. Churn, also known as customer turnover, has become a significant problem in the telecommunications industry. Customers will leave their existing provider and switch to another one that offers the same services, thus affecting the company's growth and revenue [1]. A higher rate of return is achievable through the analysis of customer churn to reinforce constant growth and business advancement [3]. The cause of network switches by customers is mostly due to the lack of network connectivity, which tends to be poor sometimes. Also, the package costs contribute to customer churning [4]. Literature suggests that it is cheaper to prevent existing customers from switching providers than to attract new customers [5]. Therefore, prediction models have been employed to identify customers who are likely to churn [1].

Datasets commonly used for predicting churn in the telecommunications sector generally come from telecom service providers. Variables that are considered include call usage details, specifically for voice, SMS, and data, roaming frequency of usage, frequency of SMS, frequency of voice, customer personal information such as but not limited to income and credit, customer interaction variables such as the number of contacts with the support center, subscription-related variables, e.g., type of tariff plan, type of services, subscription length, charge amount, customer demographics e.g., gender, age group, location, marketing-related attributes, and financial information attributes and churn status [6-10]. In addition, these variables can provide information about the time and date of customer churn for telecommunication service providers.

Dimension reduction was performed using principal component analysis (PCA) to predict churn for a telecom prepaid dataset [11]. The algorithms proposed by the researcher are neural networks (NN), support vector machines (SVM), and Bayesian networks (BN). According to the results produced by Brânduşoiu, et al. [11] all algorithms performed well in predicting customer churn, with an accuracy level that is around 99%.

A telecommunication dataset was also used by Ahmad, et al. [7] for predicting churn; tree-based machine learning algorithms, namely XGBoost, GSM, random forest (RF), and decision trees (DT), were compared. The XGBoost model outperformed all other models based on the area under the curve (AUC) Keramati, et al. [12] compared Artificial Neural Network (ANN), k-nearest neighbor (KNN), Decision Tree (DT), and Support Vector Machine (SVM), and the results revealed that ANN was the model that outperformed the other models based on recall and precision for churn prediction Fujo, et al. [1] compared XGBoost, logistic regression (LR), Naïve Bayes (NB), and KNN to the proposed Deep-BP-ANN model. The results revealed that the Deep-BP-ANN model outperforms the other churn predictive models. The LR, NB, SVM, DT, RF, XGBoost Classifier, CatBoost Classifier, AdaBoost Classifier, and Extra Trees Classifier machine learning models were compared by Lalwani et al. [13] to determine the best model to predict churn.

Hashmi, et al. [13] reviewed 61 articles from 2002 to June 2023 in telecommunication by year of publication, data mining technique, journal, and dataset type. The results reveal that the wireless dataset type and the decision tree technique are dominant. In terms of the year of publication, 2012 had the most research articles, and Expert Systems with Applications is the dominant journal in the review period.

A systematic review was conducted by Eria and Marikannan [14] from 2014 to 2017, in telecommunications, using existing data mining models, data preparation methods, and churn prediction challenges, their results revealed that the feature selection method is the dominant data preparation technique, and the dominant data mining techniques are SVM and NN. Furthermore, large datasets and class imbalances are the primary churn prediction challenges. However, simplicity and variable transformations suggested by Seitshiro and Govender [15] may be implemented.

Eria and Marikannan [14] reviewed techniques, datasets and performance measures used for churn prediction in telecom from the year 2005 to 2020. The results reveal that the data sets used were taken from telecommunication companies of different countries. The maximum size was 600,000 and has 722 features, the most common of which are related to recharges, demography, duration, service, and customer care. Furthermore, logistic regression and decision tree appeared to be the most used techniques, with accuracy being the primary performance measure. The authors noted that the confusion matrix, precision, recall and f-score must be considered when measuring the performance.

A bibliometric analysis was conducted by Bhattacharyya and Dash [16] from 1985 to 2019, and research on churn behavior in telecom has been extensively studied. The authors' results revealed that literature on churn has been cited 6,544 times. The United States of America (USA) was the most cited country, followed by the United Kingdom. The journal Expert Systems with Applications contributed the most, with University College Dublin being the most contributing affiliation, and Computer Science being the most contributing research field. Furthermore, based on keyword term occurrence map frequency, machine learning, data mining, classification, and variance-based statistical methods are the prevalent research topics. According to keyword co-occurrence analysis, churn prediction and modeling are the dominant topics.

Ribeiro, et al. [17] conducted a bibliometric review of churn from the year 1995 to 2020. Their results show that Expert Systems with Applications is the dominant publishing journal. China is the most productive country, followed by the USA, and Katholieke Universiteit Leuven was the most contributing organization. The dominant keywords are "churn prediction" and "satisfaction," respectively.

Peer-reviewed articles have been published that focused on building customer churn prediction models in the telecommunications industry [8]. These authors have proposed customer churn prediction models based on techniques such as decision trees, random forests, neural networks, and logistic regression, just to name a few. Despite an increase in published articles on telecommunications customer churn, a summarized review of the scientific landscape focuses on customer churn prediction in all industries. To the best of our knowledge, few studies have focused on analyzing the development of scientific production on telecommunication customer churn. To address this gap, this article uses a bibliometric analysis to assess the extent of use and gain an enhanced understanding.

This study seeks to identify, screen, summarize, and analyze existing studies on telecommunication customer churn to reveal patterns in this field. The research questions of this study are as follows:

i. What are the search trends associated with telecommunications customer churn?
ii. What is the intellectual structure of the field?
iii. What is the specific authorship structure of the field?
iv. What is the social structure of the field?

The bibliometric analysis in this study contributes to the development of the literature on customer churn in telecommunications. Firstly, it reveals the most trending topics and frequently used words through co-word analyses. Secondly, it identifies very active authors in this field and provides guidance for future research. Lastly, it reveals the growth of collaborations among authors and countries in the field.

The paper is organized as follows: Section 2 discusses the literature review on customer churn prediction in telecommunications, Section 3 details the models used in churn prediction, Section 4 discusses the research methodology, Section 5 presents the results observed from the study, and Section 6 concludes.

2. Related Churn Prediction Techniques

In this section, the commonly used data mining methods are discussed.

2.1. Decision Tree

The DT represents a tree structure and is widely used in data mining methods for developing prediction algorithms. This method classifies an event into branch-like outcomes that represent an inverted tree [18]. The outcome is derived from the input variable that splits the data into two homogeneous classes at each stage [19]. To assess the quality of this split, a Gini impurity measure is calculated as follows:

$$\boldsymbol{Gini(t)} = \mathbf{1} - \sum_{\mathbf{1=0}}^{c-\mathbf{1}}[\boldsymbol{r(i|n)^2}] \qquad (1)$$

where i is the class in which the fraction of records $r(i|n)$ belongs to at a given node n and the number of classes is c. Support Vector Machine

Support vector machine methodology is a two-classification model that is used in classification problems, regression, and outlier detection [19, 20]. It uses principles borrowed from statistical learning theory, where the best function that minimizes an observed risk, such as an estimated function that is different from the current function, is chosen from a set of given functions.

In a binary classification instance, the decision function maximizes the initial m-dimensional space of the problem input space $\boldsymbol{x}$, is mapped into the n-dimensional $(n \geq m)$ feature space $\boldsymbol{z}$, in which the classes become linearly separable through a kernel function [21] see line labeled 1 on Figure 1 to Figure 3. The upper and lower boundaries are $\Phi w^{\intercal}(x) + b = 1$ and $\Phi w^{\intercal}(x) + b = -1$. Thus, the decision function $D(x)$ is determined as $D(x) = \Phi w^{\intercal}(x) + b$ where $w \in \mathbb{R}^n$ is an $m-$ dimensional vector, the bias is represented by $b \in \mathbb{R}$ is a scalar and for , $i = 1, \dots, n$.

The hyperplane forms a separating hyperplane that separates $\boldsymbol{x}_i$, $i = 1, \dots, n$, is

$D(x) = \Phi w^{\intercal}(x) + b$ for $-1 < D(x) < 1$. The distance between the training instances nearest to the $D(x)$ and $D(x) = -1$ and the hyperplanes and the separating hyperplanes are called the margin. If the hyperplanes $D(x) = 1$ and $D(x) = -1$ have at least one training instance, the maximum margin for the hyperplane $D(x) = 0$ is achieved for $-1 < D(x) < 1$. Thus, the generalization region for the decision function is $\{x| -1 \leq D(x) \leq 1\}$. Thus, the hyperplane with the maximum margin is called the optimal separating hyperplane (Figure 1)

The SVM formula for classification problems is:

$$\sum_{i=1}^{n} \propto_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \propto_i \propto_j y_i y_j \Phi(x_i)^T \Phi(x_j) \text{ subject to } \sum_{i=1}^{n} \propto_i y_i = 0, \quad \propto_i \geq 0,\ i = 1,2, \dots, n$$

where $\alpha = (\propto_1, \propto_2, \quad, \propto_n)^{\intercal}$and $\propto_i$ are the non-negative Lagrange multipliers.

Logistic Regression

According to Nie, et al. [2] and Mestiri [19] logistic regression model is a commonly used statistical modelling technique that is used for classifying binary events. The outcome of an event is written as follows:

$$\pi(X) = \frac{exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}{1 + exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)} = \frac{exp(X\beta)}{1 + exp(X\beta)}$$

where π is the probability of the event happening, which ranges between 0 and 1, β_0 is the intercept or slope, $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients, *n* is the number of observations and $X_1, X_2, \dots, X_n$ are the independent or predictor variables.

2.2. Random Forest

Random forest (RF) is a type of decision tree whereby multiple decision trees are combined to make a prediction [15]. Its output category is determined by the mode of the output category of a single decision tree and the method is shown as follows:

i. A new training sample is created whereby samples are extracted from the original training samples using the bootstrap technique, thereby constructing a classification decision tree.
ii. Suppose that the total number of sample features is a, b features $v < a$ are randomly selected at the splitting nodes of each DT, and the information contained in each feature is calculated. Then, the ideal feature is selected among the features for splitting,

iii. The DT must fully develop before it can undergo the pruning process,
iv. The resulting DTs are then combined to form an RF, and the predicted results are output by the random forest. Also, the importance of every variable will be given [21].

2.3. Deep Neural Network

Deep neural network (DNN) is a machine learning technique that is comprised of a network layer of neurons that resembles the human brain [22]. These neurons are known as nodes, which are divided into an input layer, a hidden layer, and an output layer. The output in the output layer of the DNN is calculated as:

$$y(t) = \sum_{k=1}^{L} f\big(w_k + x_k(t)\big) + \in (t)$$

where w_k denotes the weights of a layer, $x_k(k = 1,2, \dots, L)$ is the total number of sequences of events and f is the activation function [19].

3. Research Design

3.1. Research Methodology

The Web of Science (WOS) electronic bibliographic database was used to collect publications on churn in telecommunications. This database contains data that includes article titles, article types, authors, author institutional affiliations, keywords, abstracts, number of citations, journal names, publisher names and addresses, years of publication, volumes, issue numbers, and lists of cited references. Table 1 shows the sample criterion that was followed to collect the publications. The first row of the table fully explains the rationale and objective of our study; the design of the study follows next, as it specifies the type of analysis applied. The study considered all the years available, and the retrieval date was 3 May 2024. A total of 2702 English records were selected, where all numbers of citations per publication were considered. The main keywords used to restrict the search are explained fully in the search strategy. Many bibliometric studies focus on churn but neglect to focus on customer churn in telecommunications.

Table 1.
Sample criterion.

Criterion	Description
Objective and rationale	The paper analyzes the prediction of churn in the telecommunications sector. The paper aims to investigate methods or techniques used in the telecom sector to predict churn.
Study design	The study applies a bibliometric analysis to summarize the existing evidence from the literature based on a rigorous process.
Eligibility criteria	We considered the eligibility criteria of articles by searching the Web of Science electronic database and applying the codes defined by the authors; we mapped and clustered bibliometric data.
Publication time frame	All years
Language	English
Search strategy	We selected the following codes to be searched in the source database: TS (("customer turnover" OR "customer attrition" OR "customer rotation" OR "customer defection" OR "customer switching" OR "churn*" AND ("telecom*) OR "telecommunication*")) Keywords added are: churn, churn analysis, churn management, churn prediction, customer churn, customer churn prediction, customer churn prediction CCP, and telecom churn prediction.
Sample	Results found 2,702 publications where the number of citations per item was considered.

3.2. Article Selection

The search flowchart used to select published articles is presented in Figure 1. It describes the stages of identifying, screening, and excluding articles from the analysis. All articles were read and analyzed in detail. As a result, 2410 articles were excluded, mainly because they were outside the scope of the study or addressed research topics not related to customer churn in telecommunications. Consequently, 292 articles were considered for this study.

Figure 2 presents the information summary of the selected 292 articles from the R Biblioshiny package. R is an open-source programming language. These articles have an average citation of 21.3 per document and were written by 830 authors, of whom 18 authors single-authored the articles.

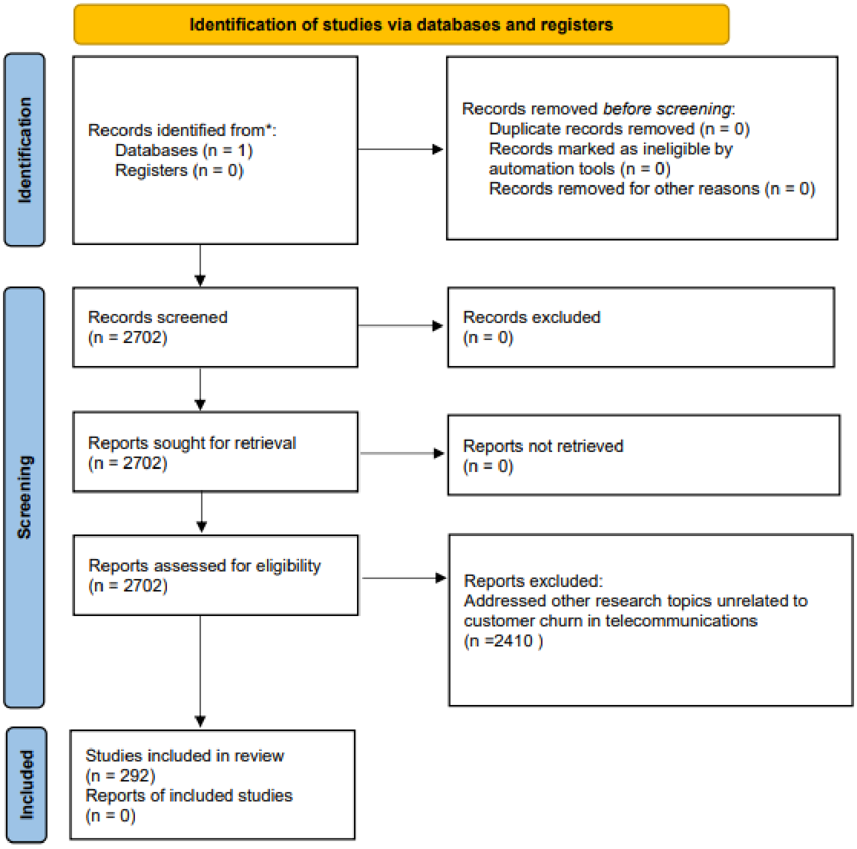

Figure 1.
Search flowchart used for article selection.

Figure 2.
Main information summary.

4. Results and Discussion

4.1. Descriptive Analysis

Only research articles are considered for this study. Figure 3 illustrates the publishing trend in churn prediction between 1999 and 2024 for 292 articles. The number of published articles prior to 2022 was quite low. The year 2022 saw the highest number of papers published at 40, but the number of published articles dropped to almost the same level as in 2020.

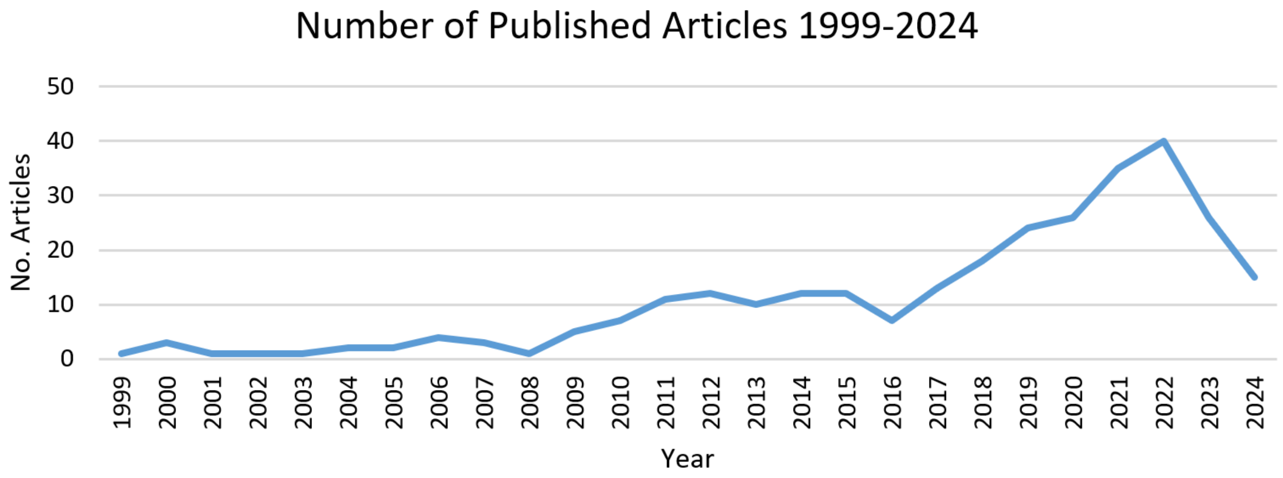

Figure 3.
Number of articles published over the years.

Figure 4 presents the average number of citations over the years. 2002 has the highest average number of cited articles, followed by 2004.

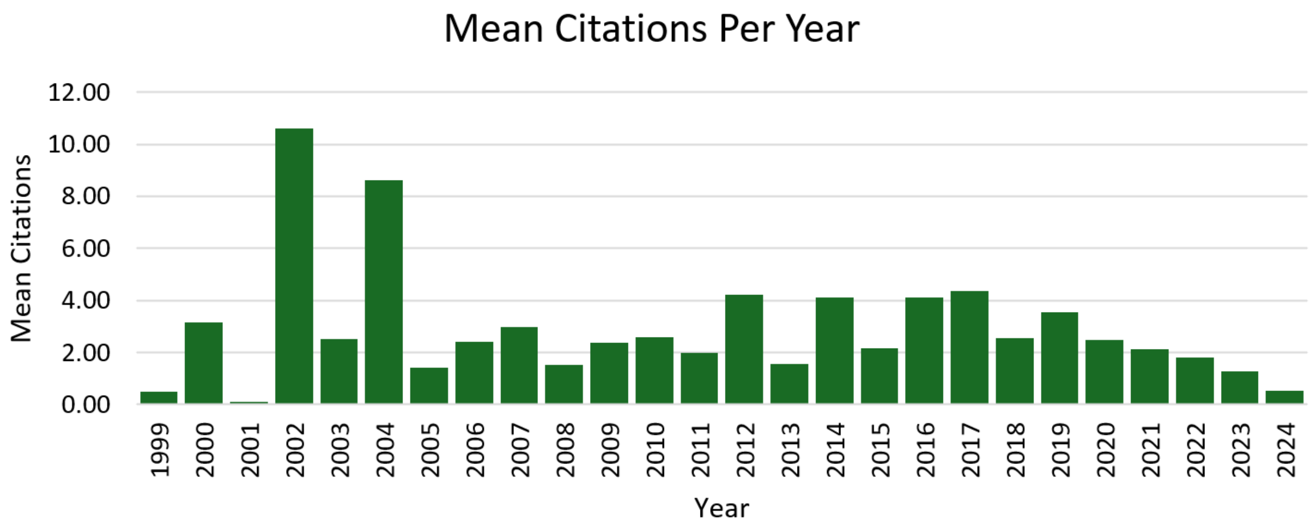

Figure 4.
The mean number of citations over the years.

Figure 4 presents the average number of citations over the years. The year 2002 has the highest average number of articles cited, followed by 2004.

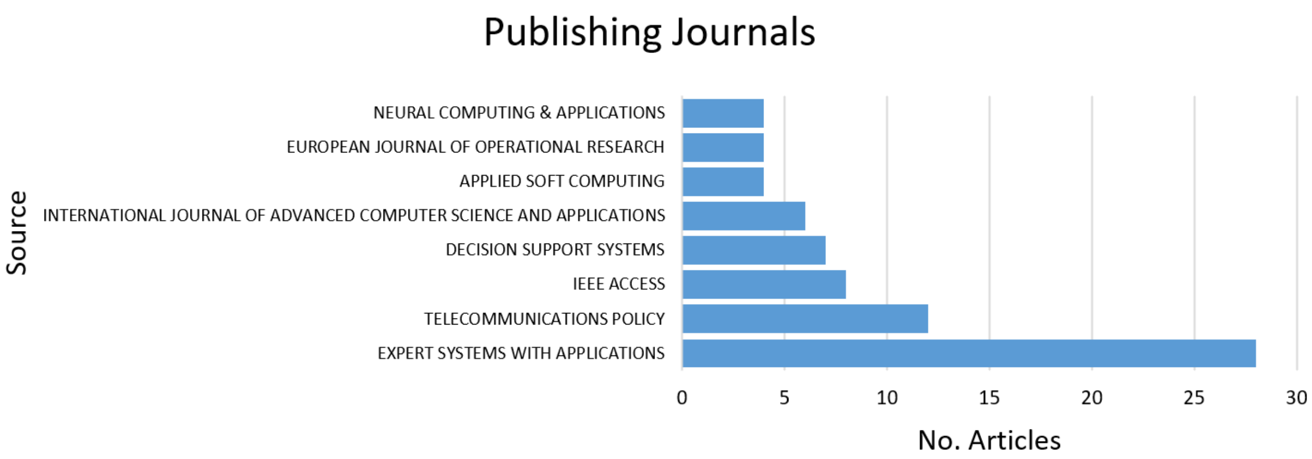

Figure 5.
Number of published articles per journal.

Journals that publish at least four articles are presented in Figure 5. The journal "Expert Systems with Applications" has the highest number of published articles, with 28 articles.

Figure 6 presents journals with at least 100 citations. The journal of "Expert Systems with Applications" has the highest number of published articles, with 933 cited articles.

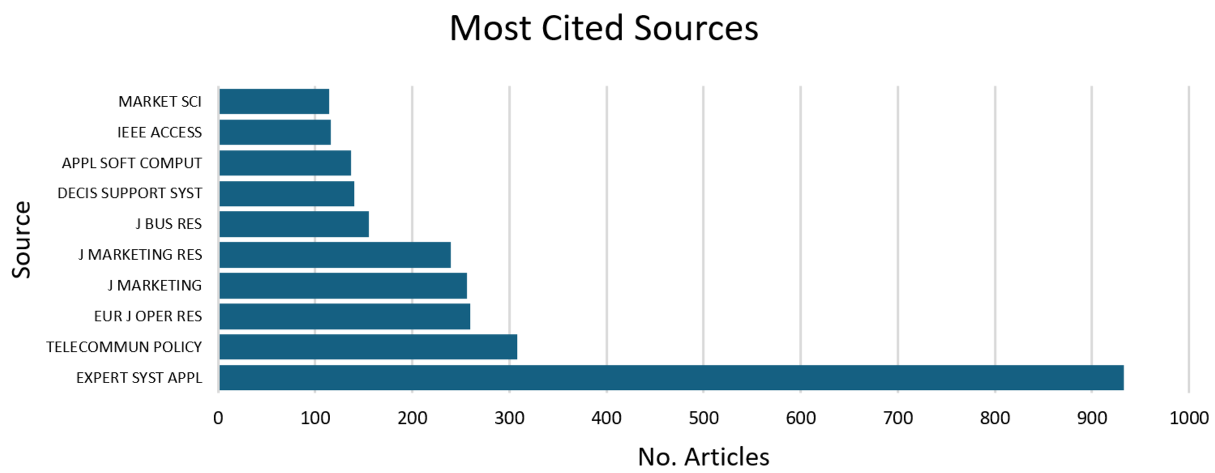

Figure 6.
Number of most cited journals.

Table 2 reveals journals with at least a local impact h-index. "Expert Systems with Applications" has the highest impact of 18, with 1352 published articles. The same journal has the highest cumulative occurrence over the years, see Figure 6 and Figure 7.

Table 2.
Source's local impact and total citations.

Sources	H-Index	TC
Expert Systems with Applications	18	1352
Telecommunications Policy	9	616
Decision Support Systems	6	256
IEEE Access	6	325
European Journal of Operational Research	4	298
Neural Computing & Applications	4	76

Figure 8 reveals the top 10 authors with at least four articles. The author "Baesens B" has the highest number, with 11 published articles.

The affiliation with the most published articles over time is "University of Southampton" with 12 articles, as illustrated in Figure 9.

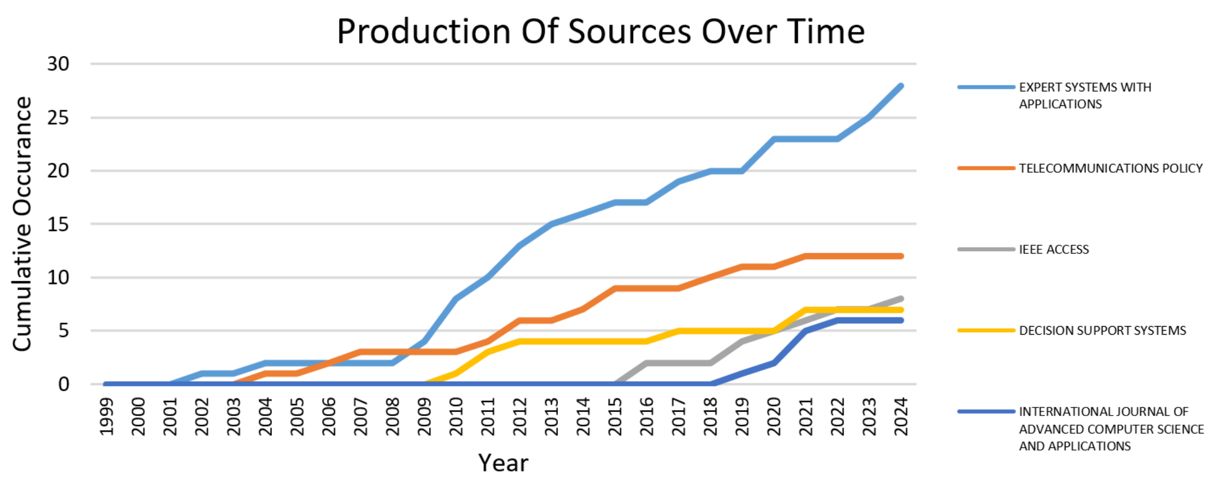

Figure 7.
Sources' production overtime.

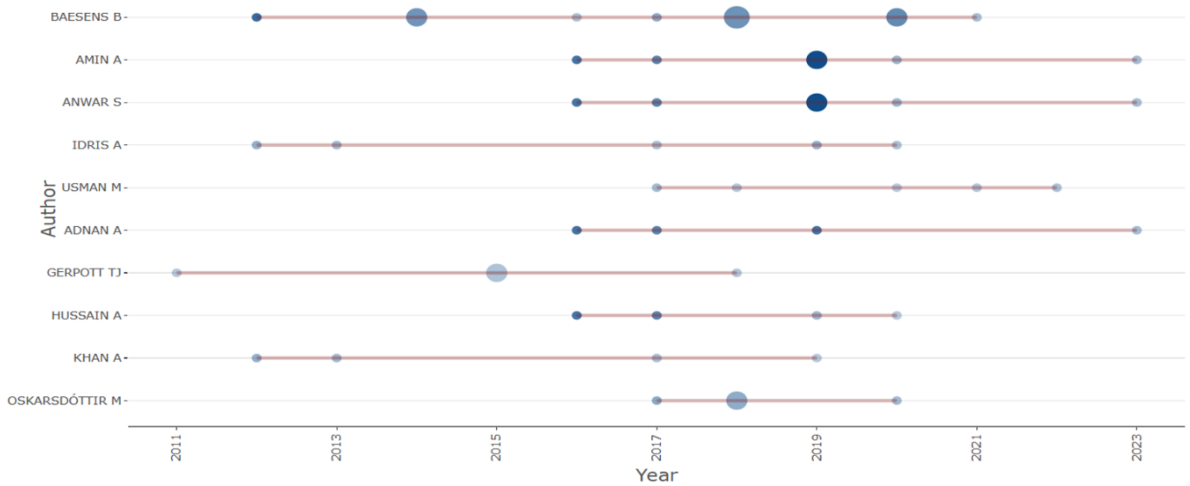

Figure 8.
Authors' production over time.

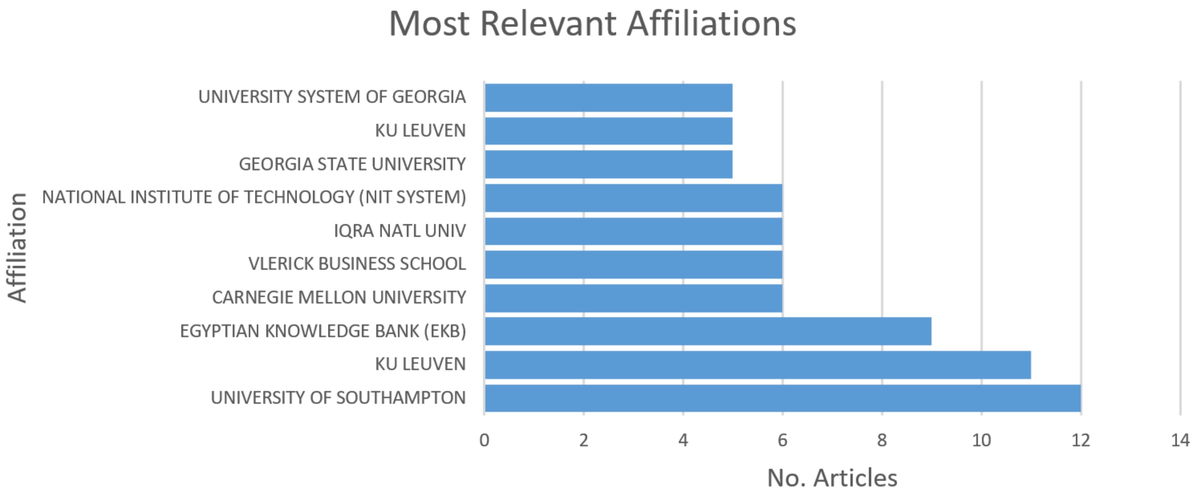

Figure 9.
Affiliations' production over time.

Figure 10 shows the Countries' scientific production, which is the number of published articles in countries around the world. The results show that China has the highest number of published articles, at 98, followed by India with 66 articles. Regarding citations, China comes out on top with 923 citations, as seen in Figure 11.

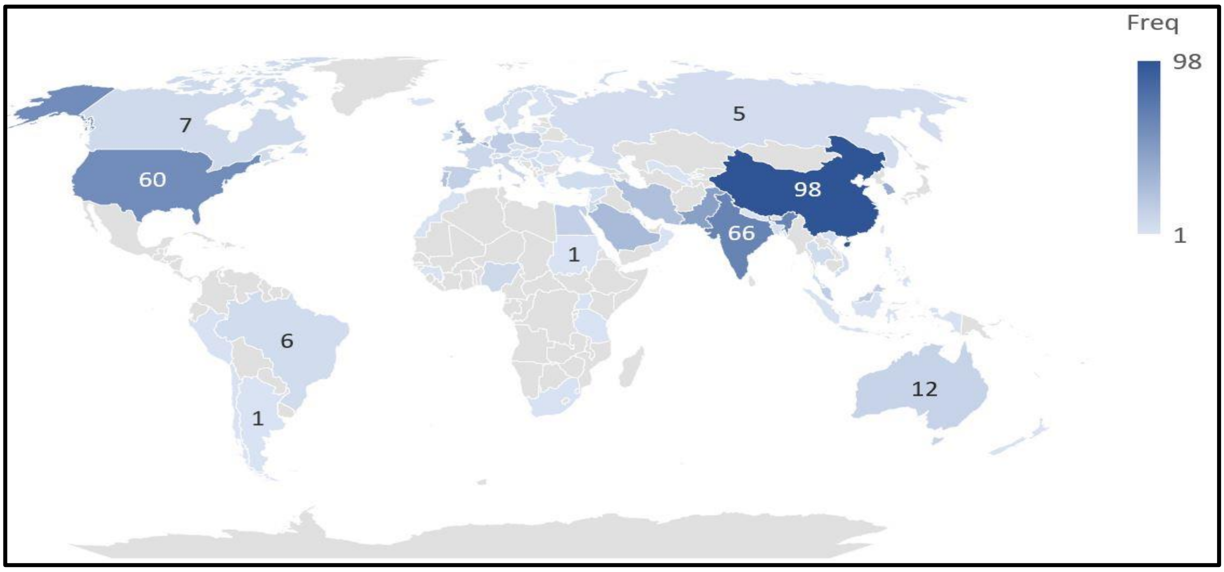

Figure 10.
Countries' scientific production.

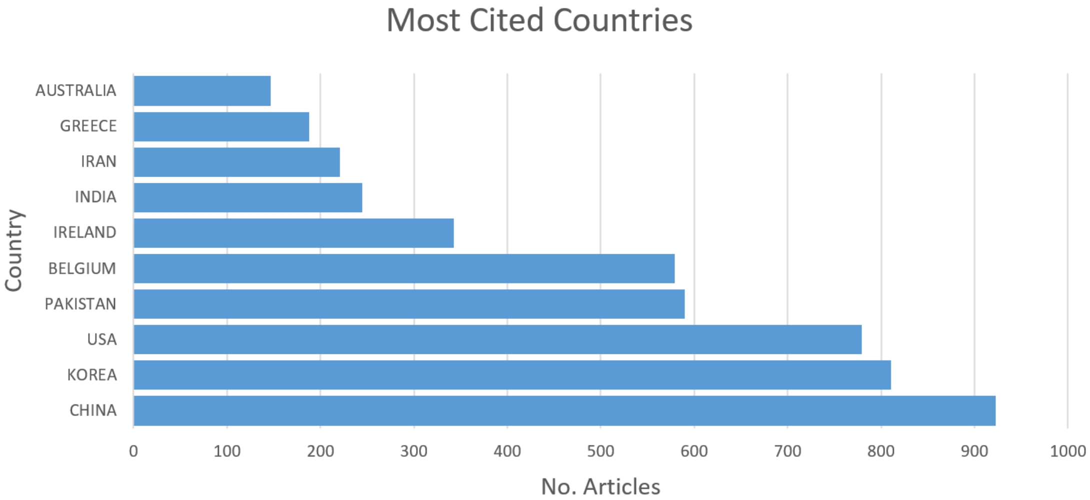

Figure 11.
The highest countries' citations.

Table 3 presents the top 10 most-cited documents. The author with the highest number of citations is Wei CP, who has 244 cited articles. Hung SY authored the reference with the highest number of citations, with 67 citations, see Table 4.

Table 3.
Most globally cited documents.

Paper	DOI	TC	TC per Year
Wei and Chiu [23]	10.1016/S0957-4174(02)00030-1	244	10,61
Verbeke, et al. [10]	10.1016/j.ejor.2011.09.031	231	17,77
Mestiri [19]	10.1109/72.846740	194	7,76
Hashmi, et al. [13]	10.1016/S0957-4174(03)00133-7	192	9,14
Verbeke, et al. [10]	10.1016/j.simpat.2015.03.003	188	18,80
Ki, et al. [22]	10.1016/j.telpol.2004.05.013	170	8,10
Hashmi, et al. [13]	10.1016/j.eswa.2011.08.024	154	11,85
Amin, et al. [9]	10.1016/j.telpol.2006.09.006	154	8,11
Amin, et al. [9]	10.1109/ACCESS.2016.2619719	145	16,11
Keramati, et al. [12]	10.1109/MWC.2014.6845044	118	10,73

Table 4.
Most local cited references.

Cited References	TC
Hashmi, et al. [13] EXPERT SYST APPL, V31, P515, DOI 10.1016/J.ESWA.2005.09.080	67
Wei and Chiu [23] EXPERT SYST APPL, V23, P103, DOI 10.1016/S0957-4174(02)00030-1	59
Verbeke, et al. [10] EUR J OPER RES, V218, P211, DOI 10.1016/J.EJOR.2011.09.031	54
Hashmi, et al. [13] EXPERT SYST APPL, V39, P1414, DOI 10.1016/J.ESWA.2011.08.024	53
Amin, et al. [9] NEUROCOMPUTING, V237, P242, DOI 10.1016/J.NEUCOM.2016.12.009	46
Brânduşoiu, et al. [11] EXPERT SYST APPL, V36, P4626, DOI 10.1016/J.ESWA.2008.05.027	42
Keramati, et al. [12] TELECOMMUN POLICY, V28, P751, DOI 10.1016/J.TELPOL.2004.05.013	42
Verbeke, et al. [10] SIMUL MODEL PRACT TH, V55, P1, DOI 10.1016/J.SIMPAT.2015.03.003	42
Ahmad, et al. [7] TELECOMMUN POLICY, V30, P552, DOI 10.1016/J.TELPOL.2006.09.006	41
Amin, et al. [9] J BUS RES, V94, P290, DOI 10.1016/J.JBUSRES.2018.03.003	41
De Caigny, et al. [4] EUR J OPER RES, V269, P760, DOI 10.1016/J.EJOR.2018.02.009	41

4.2. Network Analysis

This sub-section of the study used the R software to construct and visualize a bibliometric network. analysis.

4.2.1. Keyword Co-Occurrence

The keyword co-occurrence assists in determining the most frequently used terms in various papers in this study. Furthermore, it benefits researchers by helping them understand topics of great interest to authors. According to Figure 12, the top 5 words with the most cumulative occurrences are model, retention, loyalty, satisfaction, and prediction in terms of their order frequency. The topics authors found interesting in 2022 and 2023 are logistic regression, random forest, telecommunication, analytics, and machine learning techniques, see Figure 13**.**

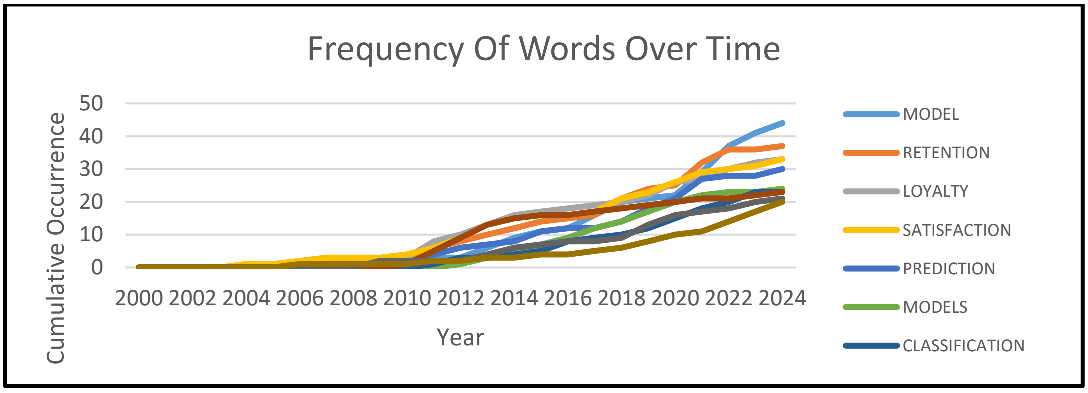

Figure 12.
Top 10 keywords cumulative occurrence.

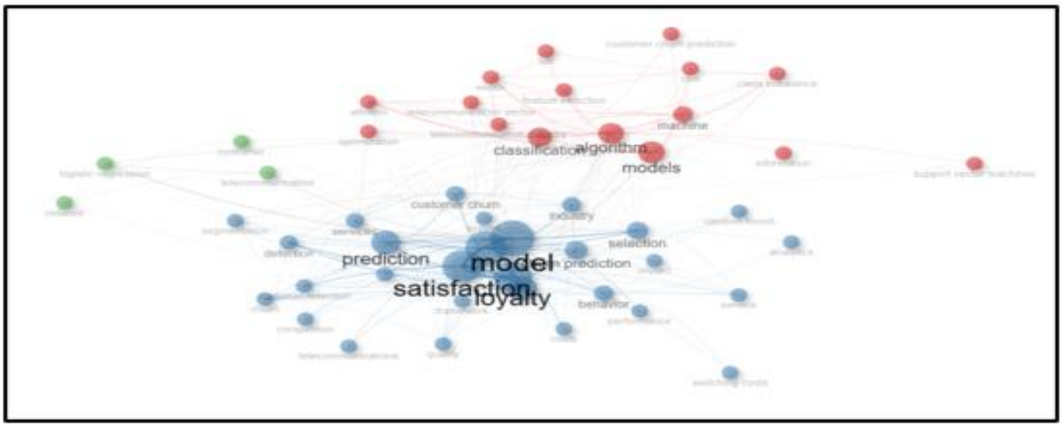

Figure 13.
Keyword co-occurrence network.

The map that represents several keywords connected through lines is illustrated in Figure 13. These lines indicate that these keywords appeared together in multiple papers in the dataset. The most frequently used are "model," "prediction," "loyalty," and "satisfaction," and these terms are an indication that more research has been conducted in these areas.
The most trending topic in 2023 was logistic regression and random forest, see Figure 14.

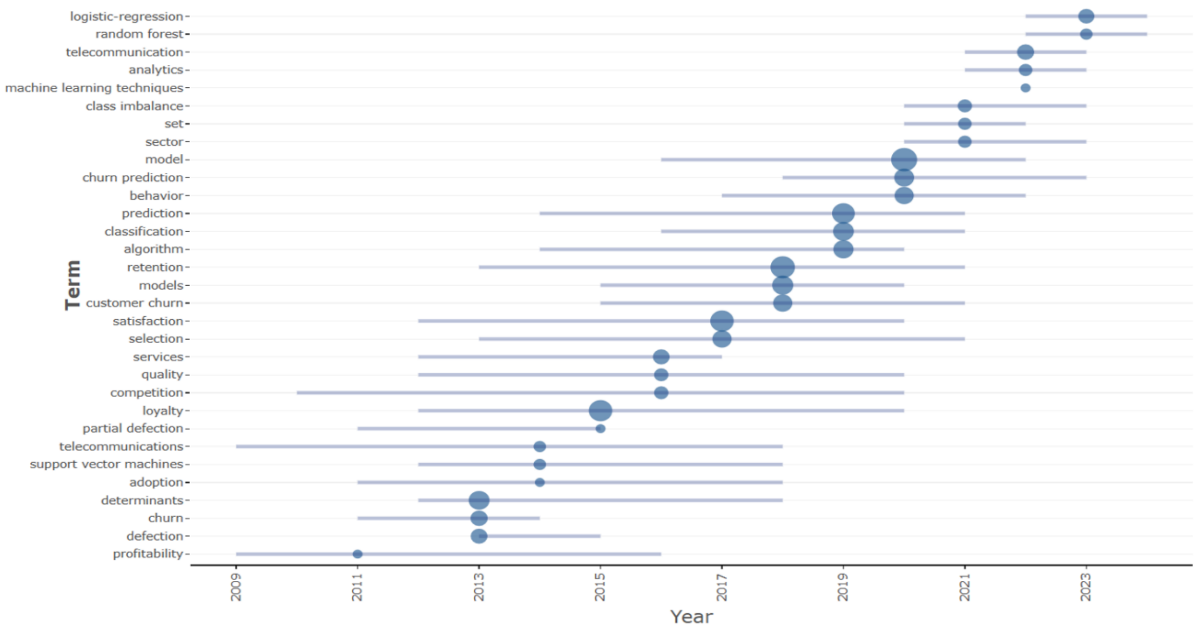

Figure 13.
Trending topics over time.

4.2.2. Co-Citation Network

Figure 15 and Figure 16 reveal the three clusters that stand out from the rest. The sphere size in the network is relative to the citation frequency; authors with similar co-citations appear in clusters. The closer the authors' names are to the central position, the more closely related the authors are.

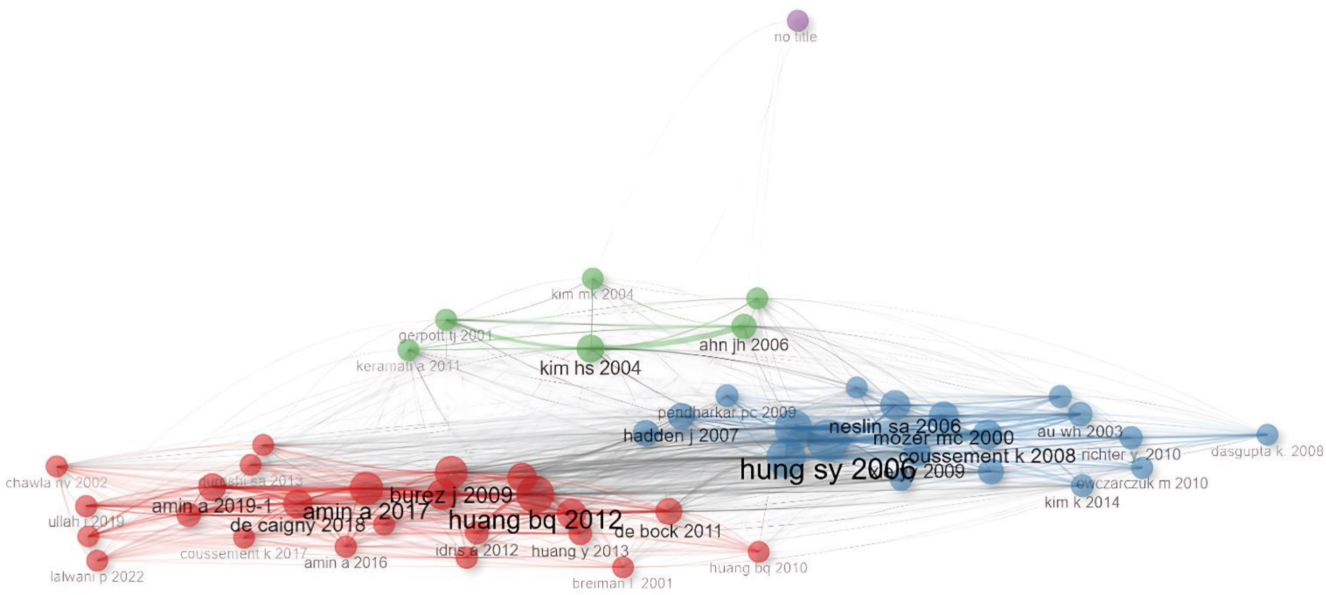

Figure 14.
Citation analysis network.

Figure 15 and Figure 16 reveal the three clusters that stand out from the rest. The sphere size in the network is relative to the citation frequency; authors with similar co-citations appear in clusters. The closer the authors' names are to the central position, the more closely related the authors are, while those that are not close appear outside near the border.

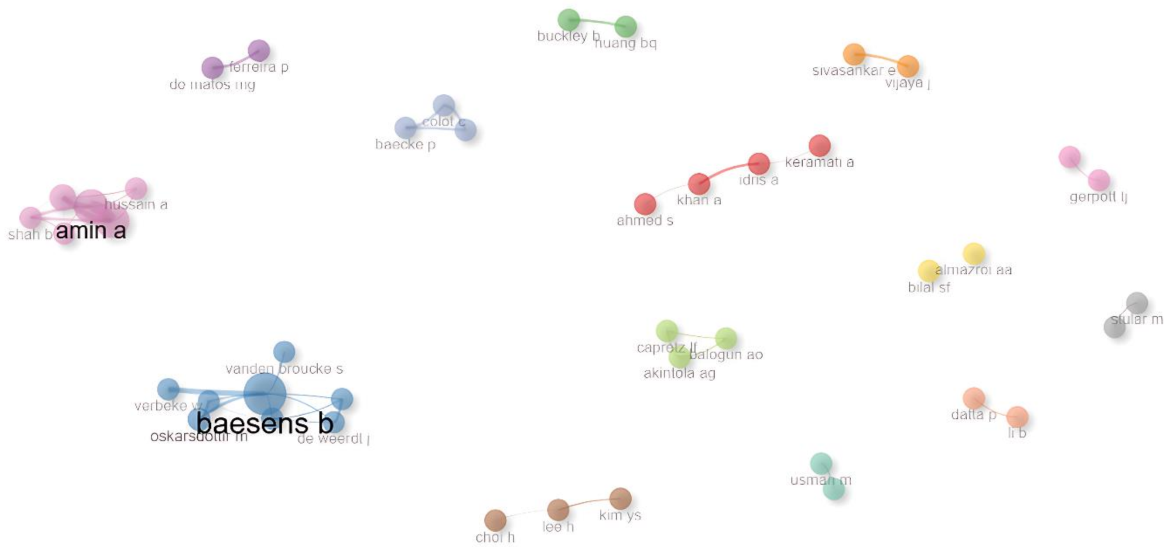

Figure 15.
Collaboration analysis network.

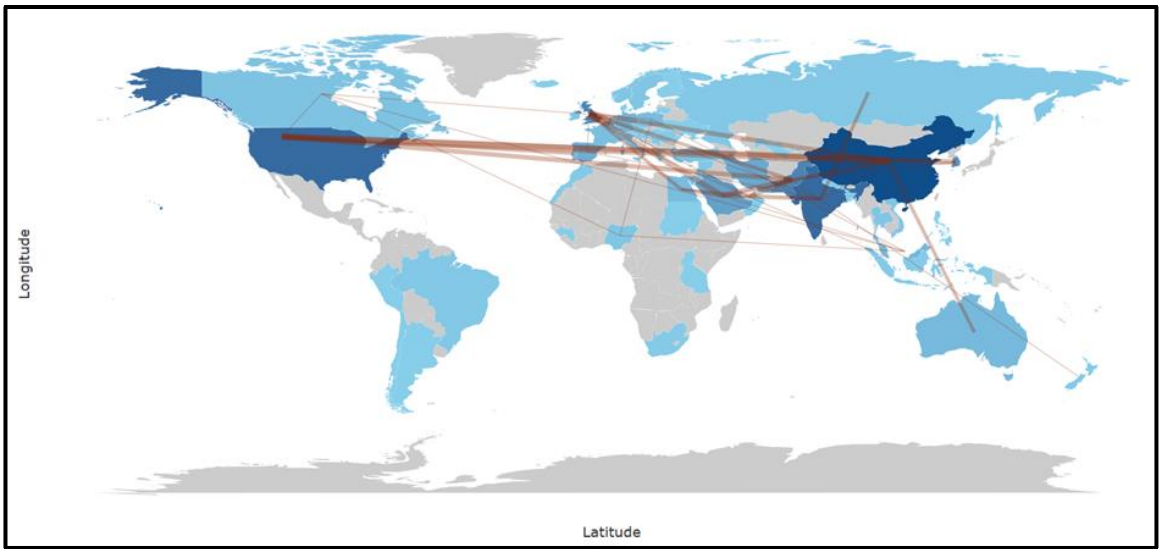

Figure 17.
Countries' collaboration analysis network.

Figure 17 presents the frequency of collaboration maps between countries. The map reveals that countries that stand out are Belgium and the United Kingdom, which are the top collaborators, followed by China and the USA.

5. Discussion

5.1. Descriptive Analysis

Recall that the "Expert Systems with Applications" journal is dominant in the number of published articles in terms of customer churn analytics in telecommunications. The results agree with the article published by Ribeiro et al. [18] and Hashmi et al. [14]. The "European Journal of Operational Research" has the second-highest number of published articles. This journal comes second, and the results agree with Ribeiro et al. [18]. However, according to results revealed by Hashmi et al. [14], the "Telecommunication Policy" journal was the second journal that dominated the number of publications in the field. Our study shows that most cited local articles come from the "Expert Systems with Applications" journal. However, the "European Journal of Operational Research" seems to be the dominant journal with the most locally cited articles on the study done in 2022 by Ribeiro et al. [18]. The dominant contributing organization in this study is the "University of Southampton". However, Katholieke Univ Leuven was the dominant contributing organization in 2022. China appears to be the dominant country in terms of article publication, as revealed in both this study and the study conducted by Ribeiro et al. [18].

5.2. Network Analysis

Regarding the most frequent keywords used, "logistic regression" and "random forest" are the dominant keywords used in the study. However, according to Ribeiro et al. [18], Churn prediction and data mining are frequently used words. Keywords common to both studies include logistic regression, telecommunication, and customer churn. In terms of the most cumulative occurrence of words, model, retention, loyalty, satisfaction, and prediction are the most frequent. Words appearing in both studies include churn prediction. Logistic regression is the most common model in this study and in studies conducted in previous years. Co-citation analysis results revealed that authors with the greatest closeness and betweenness centrality are Verbeke et al. [10] are the authors with the highest betweenness and closeness from previous years' studies.

6. Conclusions

The purpose of the review study was to identify, screen, summarize, and analyze existing studies on telecommunication customer churn to reveal patterns in this field. To the best of our knowledge, few bibliometric analysis studies have focused on examining the development of scientific production on telecommunication customer churn. This study adds to the existing knowledge in this field. A co-word analysis was performed on the first research question proposed for this paper, which was about the search trends associated with telecommunications customer churn. Logistic regression was found to be the most trending topic over time. This term relates to other concepts such as "random forest," "analytics," "telecommunication," and "machine learning techniques." However, the most frequently used term is "model," which relates to concepts such as "retention," "loyalty," "satisfaction," and "prediction."

Regarding the specific authorship structure of the field, which is the third proposed question of the article, the authors, affiliations, and countries were analyzed. The most relevant author is Baesens B., based on the number of articles published over time and the impact factor. The University of Southampton has been identified as the affiliation with the highest number of published articles over time. The top three countries with the most corresponding authors and the number of published articles are China, India, and the USA. Although China is revealed as the most cited country, Korea follows second, and the USA is third.

Regarding the intellectual development and knowledge structure of the field, which is the second research question proposed for this study, co-citation analysis and bibliographic coupling techniques were proposed. Co-citation analysis results revealed that Verbeke, et al. [10] BQ are the top five authors with the greatest closeness and betweenness centrality. The article that is most cited globally and locally is by Wei and Chiu [23].

The results reveal that Lalwani, et al. [24] are the top five authors with the greatest closeness and betweenness centrality. Furthermore, the collaboration between Belgium and the United Kingdom showed the highest frequency, followed by the collaboration between China and the USA. Following the answers to the research questions of the study, it becomes clear that the trending topics in telecommunication churn are not enough to offer a broader perspective in this field. More topics that cover "customer churn," "churn management," and "churn prediction" are some of the topics authors should consider. There seems to be less representation of African countries, their authors, and affiliated bodies in this field. Authorship and collaboration need to be extended to less-represented countries. The WoS database was the only source of articles for this study, so it is recommended that other databases be considered in future studies. Additionally, bibliometric analysis is the type of analysis considered for this study; other types of analysis should be considered to get more information about literature reviews in the field.

References

[1] S. W. Fujo, S. Subramanian, and M. A. Khder, "Customer churn prediction in telecommunication industry using deep learning," *Information Sciences Letters,* vol. 11, no. 1, p. 24, 2022.

[2] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Systems with Applications,* vol. 38, no. 12, pp. 15273-15285, 2011. doi.org/https://doi.org/10.1016/j.eswa.2011.04.041

[3] M. Chaudhary, L. Gaur, N. Jhanjhi, M. Masud, and S. Aljahdali, "Envisaging employee churn using MCDM and machine learning," *Intelligent Automation & Soft Computing,* vol. 33, no. 2, 2022.

[4] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European journal of operational research,* vol. 269, no. 2, pp. 760-772, 2018. doi.org/https://doi.org/10.1016/j.ejor.2017.12.007

[5] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Transactions on Industrial Informatics,* vol. 10, no. 2, pp. 1659-1665, 2012.

[6] K. Coussement, S. Lessmann, and G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," *Decision Support Systems,* vol. 95, pp. 27-36, 2017. doi.org/https://doi.org/10.1016/j.dss.2016.12.001

[7] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *Journal of Big Data,* vol. 6, no. 1, pp. 1-24, 2019. doi.org/https://doi.org/10.1186/s40537-019-0178-6

[8] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access,* vol. 7, pp. 60134-60149, 2019.

[9] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research,* vol. 94, pp. 290-301, 2019. doi.org/https://doi.org/10.1016/j.jbusres.2018.01.059

[10] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction," *Applied Soft Computing,* vol. 14, pp. 431-446, 2014.

[11] I. Brânduşoiu, G. Toderean, and H. Beleiu, "Methods for churn prediction in the pre-paid mobile telecommunications industry," in *2016 International Conference on Communications (COMM)*, 2016: IEEE, pp. 97-100.

[12] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Applied Soft Computing,* vol. 24, pp. 994-1012, 2014. doi.org/https://doi.org/10.1016/j.asoc.2014.07.018

[13] N. Hashmi, N. A. Butt, and M. Iqbal, "Customer churn prediction in telecommunication a decade review and classification," *International Journal of Computer Science Issues (IJCSI),* vol. 10, no. 5, p. 271, 2013.

[14] K. Eria and B. P. Marikannan, "Systematic review of customer churn prediction in the telecom sector," *Journal of Applied Technology and Innovation,* vol. 2, no. 1, 2018.

[15] M. B. Seitshiro and S. Govender, "Credit risk prediction with and without weights of evidence using quantitative learning models," *Cogent Economics & Finance,* vol. 12, no. 1, p. 2338971, 2024. doi.org/https://doi.org/10.1080/23322039.2024.2338971

[16] J. Bhattacharyya and M. K. Dash, "What do we know about customer churn behaviour in the telecommunication industry? A bibliometric analysis of research trends, 1985–2019," *FIIB Business Review,* vol. 11, no. 3, pp. 280-302, 2022.

[17] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. Rodrigues, "Churn in services–A bibliometric review," *Cuadernos de Gestion,* vol. 22, no. 2, pp. 97-121, 2022.

[18] Y.-Y. Song and L. Ying, "Decision tree methods: Applications for classification and prediction," *Shanghai Archives of Psychiatry,* vol. 27, no. 2, p. 130, 2015.

[19] S. Mestiri, "Financial applications of machine learning using R software," 2024.

[20] H. Liu, S. Huang, P. Wang, and Z. Li, "A review of data mining methods in financial markets," *Data Science in Finance and Economics,* vol. 1, no. 4, pp. 362-392, 2021.

[21] Y. Chen and R. Zhang, "Default prediction of automobile credit based on support vector machine," *Journal of Information Processing Systems,* vol. 17, no. 1, pp. 75-88, 2021. doi.org/https://doi.org/10.3745/JIPS.04.0221

[22] M. Ki, J. Shang, M. Ki, and J. Shang, "Prediction of minimum wages for countries with random forests and neural networks," *Data Sci. Financ. Econ,* vol. 4, pp. 309-332, 2024.

[23] C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," *Expert Systems with Applications,* vol. 23, no. 2, pp. 103-112, 2002.

[24] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing,* pp. 1-24, 2022.