



Machine Learning Based Customer Churn Prediction: A Bibliometric Analysis of Techniques, Key Methodologies, and sectoral Trends (2011-2025)

التنبؤ بفقدان العملاء القائم على التعلم الآلي: تحليل بيبليومتري للتقنيات والمنهجيات الرئيسية والاتجاهات القطاعية (2025-2011)

^{1*} Nour El Houda YAHIAOUI, email: nour_el_houda.yahiaoui@univ-sba.dz

¹ Senior Lecturer, Financial and Accounting Studies Laboratory, Djillali Liabes University of Sidi Bel Abbes (Algeria)

Received: 18/02/2026

Accepted: 30/05/2026

Published: 02/06/2026

Abstract

In a competitive business landscape, predictive models leveraging machine learning (ML) are indispensable for customer churn prediction. In order to represent the field's growth and future frontiers, the study analyzes 260 Scopus-indexed publications between 2011 and 2025 using RStudio and Biblioshiny. The data analysis reveals a dramatic shift away from the pre-2010 trends in comparative analyses and the basic methods of data mining into the dominant themes. Major focus is on Ensemble Classifiers such as XGBoost, Deep Learning architectures such as LSTM, and Big Data scalability models for challenging industries such as telecommunications, banking and e-commerce. The key journals for the major studies of high impact, IEEE Access, occupy significant positions; authors such as Ravi, V. and Vu, V.H. contributed to emergent clusters in support of predictive modeling. The regional patterns show India's productivity and influence, followed closely by China and impactful research from South Korea. This paper demonstrates the critical need for more XAI and qualitative analysis, including capturing ethical consequences, business actionability, and behavioral details.

JEL Classification Codes : M31 ; C45 ; C53 ; Y10

Keywords

Customer Churn Prediction;
Machine Learning (ML);
Bibliometric Analysis;
Deep Learning;
Explainable AI (XAI).

الكلمات المفتاحية

التنبؤ بفقدان العملاء؛ التعلم الآلي؛ التحليل الببليومتري؛ التعلم العميق؛ الذكاء الاصطناعي التفسيري.

في مشهد الأعمال التنافسي، يتم الاعتماد على النماذج التنبؤية التي تستفيد من التعلم الآلي للتنبؤ بخسارة العملاء. تهدف هذه الدراسة إلى تمثيل نمو هذا المجال وحدوده المستقبلية، تقوم بتحليل بيبليومتري لـ 260 منشورًا مفهرسًا ضمن Scopus بين عامي 2011 و 2025 باستخدام RStudio و Biblioshiny تكشف تحليل البيانات عن تحول كبير بعيدًا عن اتجاهات ما قبل 2010 في التحليلات المقارنة والأساليب الأساسية لاستخراج البيانات إلى الموضوعات السائدة. ينصب التركيز الرئيسي على مصنفات المجموعة مثل XGBoost، وبنيات التعلم العميق مثل LSTM، ونماذج قابلية التوسع للبيانات الضخمة للصناعات مثل الاتصالات والخدمات المصرفية والتجارة الإلكترونية. المجالات الرئيسية للدراسات ذات التأثير الكبير، IEEE Access، تحتل مواقع هامة؛ ساهم مؤلفون مثل Ravi و Vu في المجموعات الناشئة لدعم النمذجة التنبؤية. وتظهر الأنماط الإقليمية إنتاجية الهند ونفوذها، تليها الصين مباشرة والأبحاث المؤثرة من كوريا الجنوبية.

تصنيف JEL : Y10 ؛ C53 ؛ C45 ؛ M31

* Corresponding author: Nour El Houda YAHIAOUI, email: yahiaoui.houda@gmail.com

I.INTRODUCTION:

In today's competitive and data driven economy, the use of data science and machine learning (ML) is creating a major strategic change in corporate management (Ahmad et al., 2019; Lalwani et al., 2022). These disciplines are important to a successful decision-making process, in providing the most efficient approach to strategic planning and daily practice. ML and customer analytics are correlated with one another, in that vast customer datasets are collected, analyzed, and interpreted to inform strategic decision making and generate maximum long-run customer value (Chen & Yang, 2025). ML employs a mixture of statistical analysis to uncover hidden patterns, and artificial intelligence is used in descriptive, predictive, and predictive modelling to predict behaviors and to improve the work of business processes (Ullah et al., 2022; Vafeiadis et al., 2015).

These methods are used extensively to predict customer churn and thus to enhance retention and profitability (Vafeiadis et al., 2015). Prior to 2015, preliminary work focused on KDD, and on relatively simple classifications, such as "Support Vector Machines (SVMs)" and "Decision Trees" and primarily on comparative performance studies (Farquad et al., 2014; Sabbeh, 2018). The mid- and late 2010s, 2015-2020, witnessed a shift to "Ensemble Methods" (e.g. "Random Forest") that would enable the greatest accuracy and industrial scalability in Big Data environments (Ahmad et al., 2019; Ullah et al., 2019). Several recent trends (2021–2025) are for now exhibiting a rapid adoption of deep learning (DL) architectures such as "Recurrent Neural Networks (RNNs)" and "Long Short-Term Memory (LSTMs)" to analyze sequential transaction and usage data for better prediction (Sarkar & De Bruyn, 2021; Vo et al., 2021).

While it is useful to consider the breadth of literature on ML and churn analytics, not much remains of a complete, updated bibliometric review of the global development of the field. More specifically, the existing literature is unable to quantitatively correlate geographical distribution with significant methodological trends and structural anomalies that form as a consequence of the move into complex, imprecise models.

This paper addresses this lacuna by performing a bibliometric analysis of 260 Scopus-indexed publications from 2011 to 2025. Leveraging scientific mapping methodologies with Bibliometrix R and Biblioshiny, this work intends to deliver a comprehensive, holistic understanding of the field's transformation, thereby pinpointing core knowledge gaps, critical inflection points, and limitations for prospective research. The study is governed by the following research questions:

Main research question (RQ): What is the trajectory of scholarly output and thematic focus within the domain of ML-driven consumer churn prediction over the study period?

- **RQ1:** Which principal and emergent research themes have shaped the long-term focus of consumer churn prediction using ML over the study period?
- **RQ2:** Based on quantitative metrics of impact and citation frequency, which academic publications hold the highest relevance in the field of ML based churn prediction?
- **RQ3:** Who are the key researchers driving productivity and impact in the intersection of ML based customer churn prediction, and how are their contributions methodologically characterized?

- **RQ4:** What is the geographical distribution of ML based churn research globally, and what comparative trends are evident when analyzing output across countries and continental regions?
- **RQ5:** What core methodological concepts and future directions are identifiable from an in-depth content analysis of the world's most frequently cited articles on ML based churn prediction?
- **RQ6:** How do different industrial verticals utilize and adapt ML techniques for churn forecasting, and what sector-specific modeling challenges emerge?

The rest sections of this article are arranged as follows: The methodological approach is presented in Section 2. Section 3 is devoted to reporting and visualizing the core results. Section 4 analyzes the evolution and discussion, including implications and limitations. The final conclusions and future inquiry directions are aggregated in Section 5.

II. METHOD AND PROCEDURES:

The methodology employed a five-stage procedure for scientific mapping: study design, data collection, data filtering and preparation, data analysis and visualization, and interpretation, specified by established bibliometric guidelines.

1. Data Collection and Scope

The scientific database selected for this bibliometric review was Scopus, which provides the most comprehensive coverage for scientific mapping, including nearly 97% of publications found in Web of Science (WoS) (Burnham, 2006). The search was conducted using a comprehensive query across the TITLE, ABSTRACT, and AUTHOR KEYWORDS fields to maximize both recall and precision.

The core search strategy combined three components: The Problem terms related to customer attrition (e.g., "customer churn," "client attrition"). The Methods terms related to machine learning (e.g., "machine learning," "deep learning," "xgboost"). The Exclusion terms to filter out non-consumer research (e.g., "employee churn"). This initial query yielded 748 records (1999–2025). The data collection date was October 24, 2025.

2. Data Filtering and Preparation

The initial 748 records were filtered based on: Document Type: Limited exclusively to Articles (254) and Review Articles (6). Language: Restricted to English documents.

Table (01): Defining the Scope: Eligibility and Exclusion Parameters

Topic	TITLE-ABS-KEY(("customer churn" OR "client attrition" OR "customer defection" OR "churn prediction") AND ("machine learning" OR "deep learning" "AI" OR "data mining" OR "xgboost" OR "random forest" OR "SVM" OR "neural network*" OR "LSTM") AND NOT ("employee churn" OR "staff turnover" OR "personnel turnover" OR "employee attrition")) AND (LIMIT-TO (LANGUAGE , "English"))).
Date	October 24, 2025

Language	Only documents in English.
Categories	All disciplinary fields are included, regardless of research specialization.
Source	Scopus Database.
Initial Yield	748 records (full period 1999–October, 2025, various document types).
Type of Documents	Articles, Review Articles.
Final Yield	260 records (full period 2011– October, 2025, Articles and review articles).

Source: Author's elaboration

This filtering brought up 260 relevant documents from the 2011 to 2025 research study period. The units of data are in both RIS and CSV formats. The “RStudio software package” was used for data processing, cleaning, and analysis (R version 4.3.0, <https://www.R-project.org>, accessed October 24, 2025), the “Bibliometrix R package” (version 4.5.1, <http://www.bibliometrix.org>, accessed October 24, 2025), “the Biblioshiny application” (version 5.0), and “VOSviewer” version 1.6.20 of October 31, 2023. Its preparation ensured data integrity by combining terminology such as Deep Learning and DNN and minimizing incomplete or duplicate records.

3. Data Analysis, Metrics, and Visualization

The analysis focused on extracting significant structural and conceptual insights from the 260 documents. Multiple metrics and statistical laws were applied h-index, Total Citations (TC), Lotka's Law (for author productivity), Country/Continent Productivity, Centrality. Conceptual analysis was achieved by combining Author Keywords and terms extracted from the Title and Abstract to maximize the representation of underlying concepts, thus identifying connections between ML algorithms, industrial sectors (RQ6), and emerging themes (RQ5).

4. Interpretation and Conclusion

The quantitative bibliometric results were interpreted based on established bibliometric techniques. At this stage the results were analyzed qualitatively and focused on the structural and conceptual connections identified. Final phase presents conclusions, implications and limitations of findings and avenues for future studies.

III.METHOD AND PROCEDURES:

Table 2 provides main descriptive information derived from the 260 articles and review articles, demonstrating a significant upward trend in scientific output.

Table (02): Data descriptive information

Description	Results
DAT MAIN INFORMATION	
Period	2011:2025
Sources	179
Documents	260
Annual Growth Rate %	25.29

Document Average Age	2.64
Average citations per doc	20.6
DOCUMENT CONTENTS	
Keywords Plus (ID)	951
Author's Keywords (DE)	663
AUTHORS	
Authors	1013
Authors of single-authored docs	0
AUTHORS COLLABORATION	
Single-authored docs	0
Co-Authors per Doc	7.33
International co-authorships %	26.92
DOCUMENT TYPES	
article	254
review	6

Source: Adapted summary table from Biblioshiny.

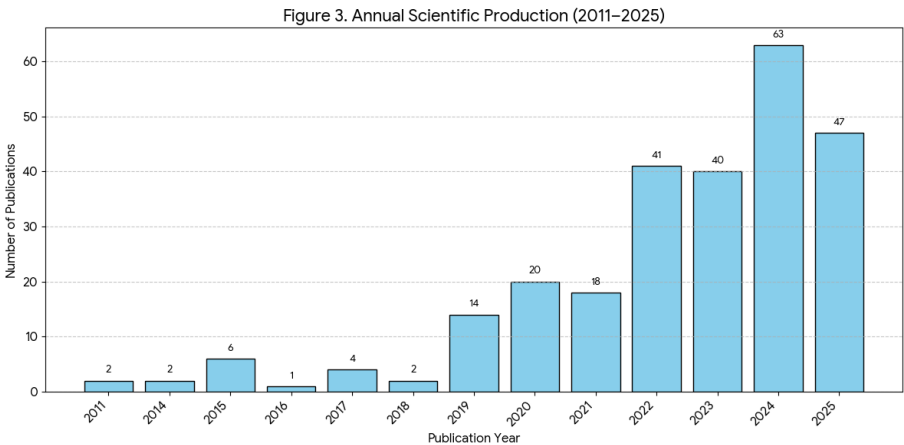
1. Main Information and Descriptive Statistical Overview

The analysis covers 260 documents published across a time span of 2011 to 2025, sourced mainly from 254 articles (97.7%) and 6 review papers (2.3%). This document type distribution reveals a heavy predominance of empirical and technical work, suggesting that while the field is rapidly advancing with new models and applications, there is also a necessary, albeit smaller, effort dedicated to synthesizing the state-of-the-art. The annual growth rate of 25.29% is high and validates the rapid acceleration observed in the field, particularly from 2019 onwards illustrated in figure 1, positioning the topic as one of the most rapidly expanding areas in applied data science over the period 2019-2025.

Furthermore, ML based customer churn prediction body is remarkably recent, with a low document average age of 2.64 years. This confirms the contemporary nature of the literature and indicates the continuous development of the ML based customer churn prediction research. Regarding impact, each manuscript obtained, on average, 20.60 citations. This high citation rate reflects the practical and strategic importance of effective ML based customer churn prediction models in the modern business environment.

In terms of authorship, in total, 1,013 researchers have contributed to the publications related to ML in customer churn, demonstrating wide participation. “The collaboration index (CI)” represents the average number of or co-authored or authors per jointly authored articles (Elango & Rajendran, 2012) reached a value of 7.33. Within this average, approximately 26.92% of the documents feature international co-authorship. Notably, the dataset contains zero single-authored documents, emphasizing the entirely collaborative nature of modern churn prediction research. Finally, the analysis discovered 951 “Keywords Plus (ID)” and 663 “Author Keywords (DE)”, providing a precise perspective on the topics addressed in publications on ML based customer churn prediction.

Figure (01): ML based customer churn prediction annual production

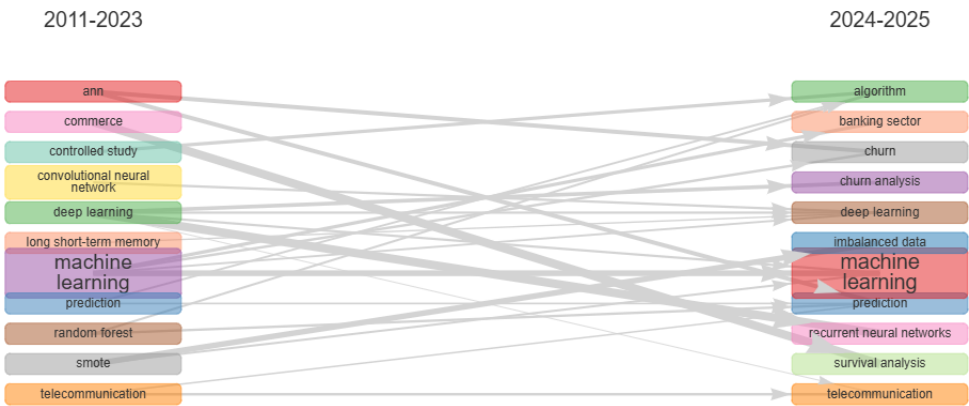


Source: Author's elaboration

2. ML based Customer Churn Prediction Evolutionary Analysis (RQ1)

Figure 2 shows the flow of the research themes from a longer period (2011-2023) to the most recent period (2024-2025). This perspective provides an opportunity to observe how the foundational themes evolved, blended and/or remained pertinent, which demonstrates the progression of research focus. While the concepts of machine learning, prediction, and deep learning have retained their opulence, the analysis primarily reveals how thematic flow was modified and applied to specific application areas over time. Telecommunication, on the other hand, remains a main industry in the field. The Rise of Deep Learning Techniques; the terms referring to such sophisticated neural designs as “convolutional neural network” (CNN) and “long short-term memory” (LSTM) have largely flown into the more general realm of “deep learning” as well as the new, specific aspect of “recurrent neural networks” in 2024–2025 (Zaghloul et al., 2025).

Figure (02): Research themes evolution from 2011-2025

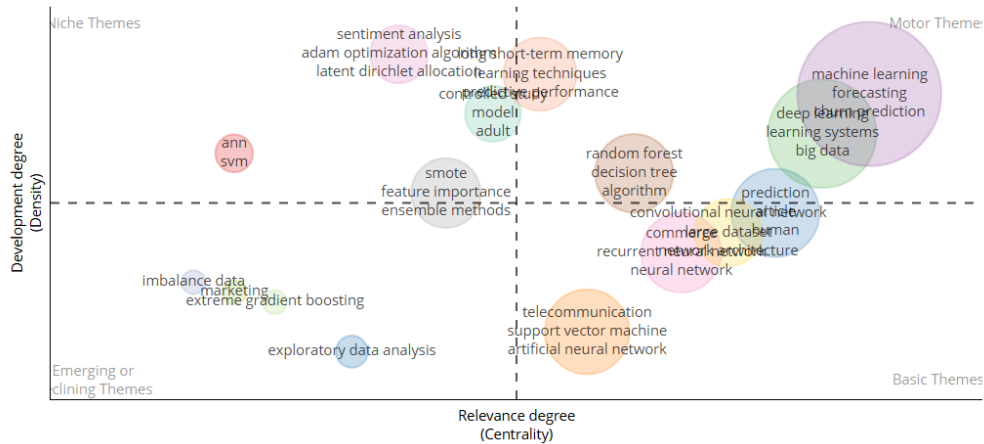


Source: Generated by the author utilizing R-bibliometrix and Biblioshiny.

The broad term "commerce" from the earlier period appears to have diversified, leading to the highly specific application themes of "banking sector" and "churn" or "churn analysis" in the newer

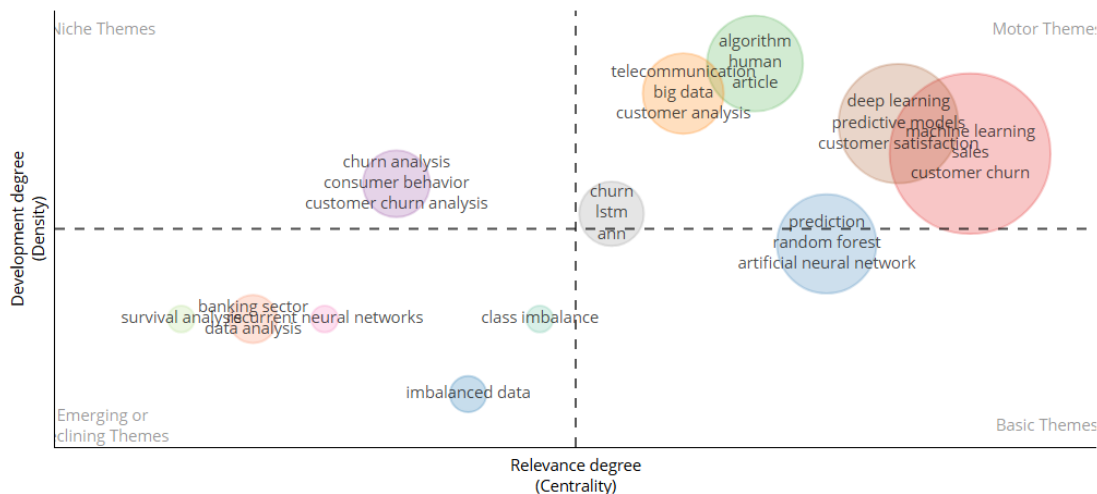
period. This indicates a refinement of research focus towards specific industry problems. For example, recent work is specifically focused on "Ensemble-based customer churn prediction in banking" (Zaghloul et al., 2025). "Random forest," a classical ML algorithm, remains highly relevant, with a clear flow into the newer period.

Figure (03): Thematic evolution map period 2011-2023



Source: Generated by the author using R-bibliometrix and Biblioshiny.

Figure (04): Thematic evolution map period 2024-2025



Source: Generated by the author using R-bibliometrix and Biblioshiny.

Figure 3 and 4 illustrate the thematic maps position themes based on their relevance (Centrality) and development degree (Density). During the period 2011–2023, the field was characterized by a balance of foundational methods and emerging applications. Motor themes were broad and highly interconnected, centered on "machine learning," "forecasting," "deep learning," and "big data". Basic themes included mature, foundational tools like "telecommunication" and "support vector machine". Critically, methodological challenges such as "imbalance data" and specific algorithms like "gradient boosting" were still underdeveloped, residing in the Emerging/Declining quadrant. Niche themes: "Artificial Neural Network" and "sentiment

analysis" are highly developed internally but not strongly connected to the core themes. This suggests focused, specialist communities studying these specific topics.

In the 2024–2025, the landscape became more concentrated and focused on practical outcomes. The motor themes tightened their focus around application-driven ML: "Machine Learning," "Deep Learning," "Customer Churn," and "Predictive Models" (Luong et al., 2025). This marks a definitive shift from generic ML toward predictive ML applied specifically to customer behavior. The specific act of "churn analysis" moved into the niche quadrant, indicating it became a highly specialized sub-field focused on actionable results, including interpretable counterfactual ensembles (Tonati et al., 2025). New frontiers, such as the "banking sector" and specific methods like "survival analysis," emerged (Chen & Yang, 2025), while classic algorithms like "random forest" and "artificial neural network" remain fundamental, transversal concepts.

2. The local Impact of Sources: Identifying the Most Relevant Journals (RQ2)

To evaluate the significance of the publication venues within ML based customer churn prediction, we analyzed the total number of publications (NP) and various impact indices ("h-index", "g-index", "m-index", and "Total Citations (TC)") in Table 3. The field's output is primarily driven by specialized computer science and open-access multidisciplinary journals.

IEEE Access leads the list with 9 publications (NP=9) and the highest influence, evidenced by 477 Total Citations (TC) and the highest h-index (h=6) among the top sources. This highlights its role as a major platform for applied computational research. Closely following are "the International Journal of Advanced Computer Science and Applications" (NP=7) and Applied Sciences (Switzerland) (NP=6), confirming that high-productivity journals favor the application and engineering aspects of ML-based churn prediction

Table (03): Top ten most relevant sources

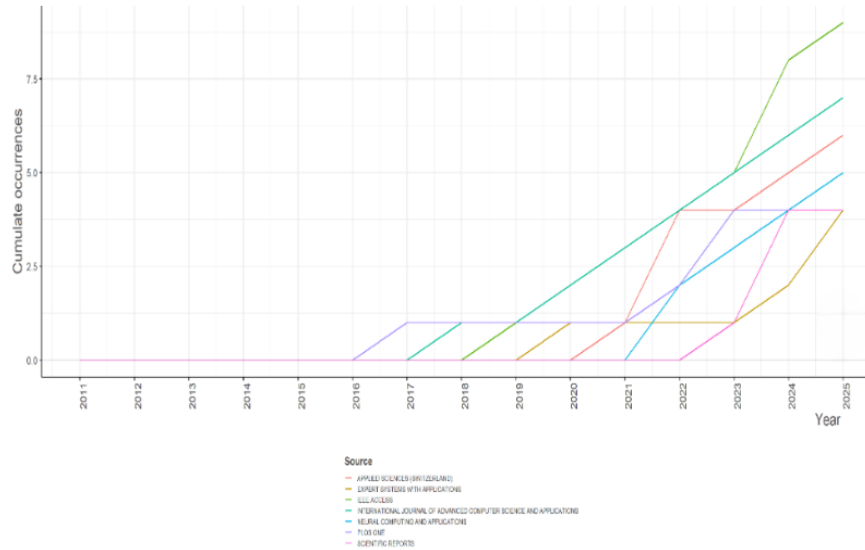
Source	h_index	g_index	m_index	TC	NP	PY_start
"IEEE Access"	6	9	0,857	477	9	2019
"International Journal of Advanced Computer Science and Applications"	3	7	0,375	162	7	2018
"Applied Sciences (Switzerland)"	4	6	0,8	153	6	2021
"Neural Computing and Applications"	4	5	1	105	5	2022
"Plos One"	4	4	0,444	117	4	2017
"Scientific Reports"	4	4	1,333	73	4	2023
"Expert Systems with Applications"	3	4	0,5	114	4	2020
"Journal of Forecasting"	3	3	1	25	3	2023
"Mathematics"	3	3	0,75	72	3	2022
"Applied Soft Computing"	2	2	0,167	187	2	2014

Source: Generated by the author using R-bibliometrix and Biblioshiny.

Based on Total Citations (TC), IEEE Access leads with 477 citations. The m-index, highlights the accelerating rate of highly cited work from younger, specialized sources. Scientific Reports achieved the highest m-index (1.333) despite starting in 2023, while Neural Computing and Applications also recorded a high value (m-index=1). This suggests a clear shift toward more

focused, high-impact venues for state-of-the-art ML churn papers. This high dynamism is supported by the chronological distribution (Figure 5), showing many top 10 journals starting their involvement in the field around 2019–2021, confirming that top journals are driving the field's rapid, exponential growth.

Figure (05): Evolution of source output from 2011-2025



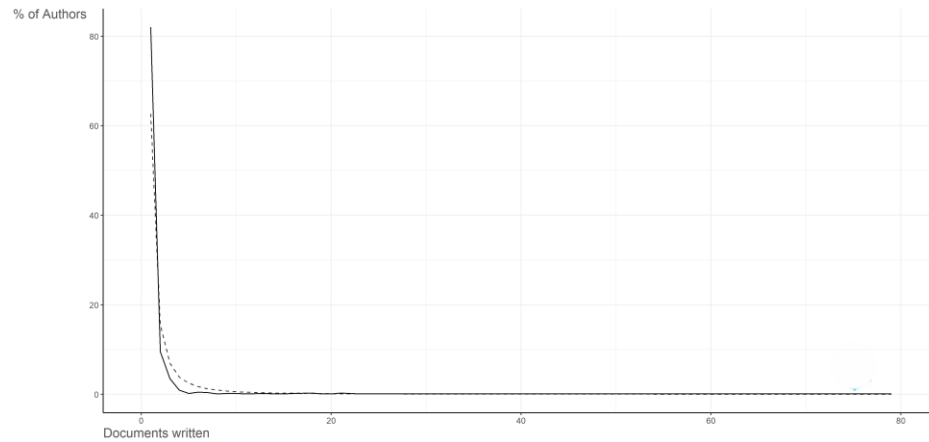
Source: Generated by the author using R-bibliometrix and Biblioshiny.

3. Research Influence Analysis: Productivity and Impact Metrics (RQ3)

A. Researcher Productivity Using Lotka's Law

To investigate the structural characteristics of the research community (RQ3), the distribution of author productivity was examined using Lotka's Law (Aria & Cuccurullo, 2017; Lotka, 1926). This law establishes that the rate at which authors contribute a specific quantity of publications is in a reciprocal relationship with a power function of that same publication count. The analysis of the 1,013 authors in the 260 documents confirms a strong conformance to this law (Figure 6), typical for rapidly evolving research fields. The data reveals a strong pattern of sporadic, single contributions: approximately 88.6% of all authors have contributed only a single article. Conversely, a small dedicated core is responsible for higher productivity, with 7.5% authoring two articles, and just 0.6% publishing four articles. This symbiosis suggests that, while many men are entering the domain, most work relies on several individual contributions with only a small core running the maintenance of publication.

Figure (06): Researcher's Productivity



Source: Author's elaboration using R-bibliometrix and Biblioshiny.

B. Researcher Impact: Productivity and Influence (RQ3)

Table 4 summarizes the top researchers, with a comprehensive list of measures including Number of Publications (NP), researcher impact (h-index), and overall influence (TCN, Total Citations Number in the data).

Table (04): Key authors' impact

Authors	Institution	Country	NP	h_index	TCN
Vu, V.H.	"Van Hieu, Vietnamese Academy of Science and Technology Institute of Information Technology, Hanoi"	Viet Nam	4	3	23
Ravi, V.	"Vadlamani, Institute for Development and Research in Banking Technology India, Hyderabad"	India	3	3	181
Usman-Hamza, F.E.	"Fatima Enehezei, Department of Computer Science, University of Ilorin, Ilorin"	Nigeria	3	3	41
Balogun, A.O.	"Abdullateef Oluwagbemiga, Department of Computer and Information Science, Universiti Teknologi PETRONAS, Seri Iskandar"	Malaysia	3	3	41
Capretz, L.F.	"Luiz Fernando, Department of Electrical & Computer Engineering, Western University, London"	Canada	3	3	41
Mojeed, H.A.	"Hammed Adeleye, Department of Computer Science, University of Ilorin, Ilorin, Nigeria, Department of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdansk"	Poland	3	3	41
Salihu, S.A.	"Shakirat Aderonke, Department of Computer Science, University of Ilorin, Ilorin"	Nigeria	3	3	41

Akintola, A.G.	“Abimbola Ganiyat, Department of Computer Science, University of Ilorin, Ilorin”	Nigeria	3	3	41
Loukili, M.	“Manal, School of Applied Sciences, Université Sidi Mohamed Ben Abdellah, Fez”	Morocco	3	3	27

Source: Author’s elaboration using R-bibliometrix and Biblioshiny.

Vu, V.H. (“Vietnamese Academy of Science and Technology Institute of Information Technology, Hanoi”) is the leader in terms productivity with 4 published articles (Ngo & Vu, 2024; Nguyen et al., 2025; Trinh et al., 2025; Vu, 2024a, 2024b) and an h-index of 3. With a TCN of 23, their research, while voluminous, has not achieved the highest citation impact among this group. Ravi, V. (from “the Institute for Development and Research in Banking Technology”, India) exhibits the highest Total Citations (TCN) of 181 from only 3 published articles (Farquad et al., 2014; Kate et al., 2023; Vasu & Ravi, 2011).

Figure (07): Co-authorship network



Source: Author's elaboration using VOSviewer

Figure 7. illustrates collaboration structure in ML based customer churn prediction visualized through the co-authorship network. The analysis of this network reveals a clear pattern of regionalized, institutional clusters rather than extensive global collaboration.

The largest cluster is the group of authors affiliated with the University of Ilorin, Nigeria such as “Usman-Hamza, F.E.” and “Balogun, A.O.” and “Mojeed, H.A.” “Salihu, S.A.” and “Akintola, A.G.” A second smaller cluster reveals strong relationships between Loukili, M. (Morocco) and the other authors, suggesting a separate, independent research area. In addition, authors like Ravi, V.

(India) and Zhang, X. (China) are highly cited, isolated nodes, or have only one peripheral co-author.

This separation suggests that while their work is highly influential (high TCN), they are important contributors in an independent or single-lab context of this data. The overall structure suggests that most productivity in this field arises from close-knit, institutional teams, with limited evidence of large-scale international or inter-institutional projects that span across the main collaborative groups.

4. Geographic Trends Analysis by Country and Continent (RQ4)

This section dedicated to mapping the geographic concentration of scholarly output, thereby illustrating the regions where studies on ML based customer churn prediction are most actively generated.

Table (05): Leading countries in publication and influence

Country	Total Citations	Number of Publications	Country	Number of Publications	Total Citations
India	1157	61	India	61	1157
South Korea	576	11	China	21	244
United States	487	18	Malaysia	19	277
Pakistan	408	9	United States	18	487
Saudi Arabia	326	16	Saudi Arabia	16	326
China	244	21	Iran	14	184
Germany	230	7	Egypt	12	218
Egypt	218	12	South Korea	11	576
Hong kong	211	5	Viet Nam	11	185
Iran	184	14	Indonesia	11	122

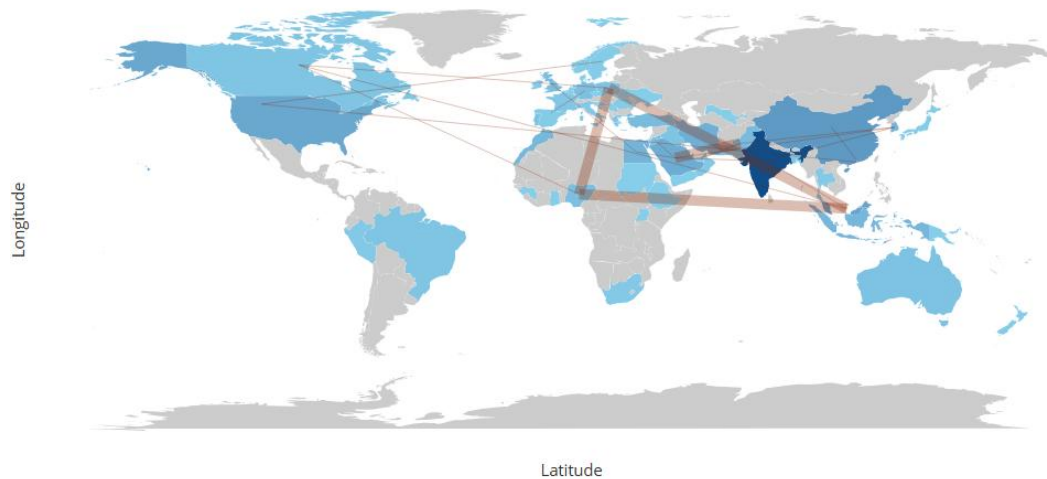
Source : Author's elaboration.

Table 5 provides an overview of the countries contributing most significantly to the ML-based customer churn prediction domain, based on both “Number of Publications” (productivity) and “Total Citations” (influence). The data clearly show India as the leader of the world with 61 publications and a total Citation count of 1157 (Arockia Panimalar et al., 2025; Bhuria et al., 2025; Subramanian et al., 2025). China is second in achieving the number of productive articles (Wang et al., 2025; Wei, 2025; R. Y. Wu et al., 2025; Yang et al., 2025). Malaysia is second with 19 articles (Gaidhani et al., 2025; Hammood et al., 2025; Ong et al., 2024) and the United States (18 articles). But, the descending order of influence (Total Citations) makes a big shift; South Korea gets the second highest citation score out of the five best contributors, with 576 citations from 11 publications and boasts the highest per-article citation ratio among the best contributors (Abdullayev et al., 2023; Bhuria et al., 2025; Fatima et al., 2024; Joy et al., 2024).

The United States is also fourth in productivity but has the third most influence with 487 citations (Chen & Yang, 2025; He & Ding, 2024; Imani et al., 2025; Rana et al., 2025). In contrast, Malaysia and Saudi Arabia are highly productive but have little influence in the global economy, suggesting that their research output is relatively large, but less frequently described by the international community than those of powerful countries such as South Korea and the United States. This analysis shows a geographically distinct focus on churn prediction that contributes to

significant outcomes not only in established research areas but also among rapidly growing cities in Asia.

Figure (08): Geographic Mapping of Research: Distribution of Publications Across countries and continent



Source: Generated by the author utilizing Biblioshiny.

The co-authorship network in Figure 8 and country productivity in Table 6 offer complementary views on the global research structure. Based on this analysis, research in ML based customer churn prediction is found to be global but highly fragmented, characterized by modest international collaboration. This is the case for the densely populated cluster of researchers from one institution: University of Ilorin, Nigeria, thus demonstrating that this research model is not restricted to a type of strong inter-institutional collaboration. The overwhelming leading Indian country in publications (61) and citations (1157), but, suggests that authors in India often work as nodes or in small, isolated teams. Also, although South Korea is the most frequently quoted country by number in terms of citations per article, it does not appear to be at the center of the large collaborative clusters and the idea is reinforced that it will make a significant contribution if it is driven by highly influencers in research rather than broad inter-country collaborations.

Internationally, research in this field is also uneven: 61 countries are estimated to have published studies, with 134 countries still waiting to contribute, suggesting a need for more knowledge sharing. The highest percentage of countries participating was Europe (18 countries) likely due to its good education and research infrastructures, followed by Asia (10 countries) due to the highly productive Indian, China, and South Korea. This network clearly displays a dense and linked cluster composed predominantly of students from only one institution – University of Ilorin, Nigeria, thus suggesting a research model heavily based on strong internal and institutional collaboration. Despite the India-dominated place in both publications (61) and citations (1157), the co-authorship network suggests that authors are often individual nodes or working in small, isolated teams. In addition, while South Korea has the highest number of citations per article, it lacks prominence in the major collaboration clusters, suggesting that the power of South Korea's research is driven by very influential internal research and not large-scale cross-border projects.

The development of research in this field also seems to have a disaggregated effect: as many as 61 countries have published research and 134 countries have yet to contribute; these shows that greater knowledge dissemination is warranted. Europe was the most active (18 countries), with Asia following (10 countries) led by the high output of India, China, and South Korea. Other regions exhibit moderate to low participation. While ML-based churn prediction is a globally relevant topic, the research effort is largely conducted within national or institutional silos. This suggests localized competition and parallel development of solutions, underscoring the urgency to enrich geographic diversity and boost international knowledge exchange.

5. The Most Cited Papers in ML Based Customer Churn Prediction (RQ5)

Table 6 demonstrates the top 10 most cited documents on ML based customer churn prediction identifies several research trends and core themes that define the field.

Table (06): Top-ranked cited documents

Rank	Most cited documents	Author	Total Citations	TC per Year	Normal ized TC
1	"A comparison of machine learning techniques for customer churn prediction"	Vafeiadis, T. et al. (2015)	357	32,45	4,22
2	"Customer churn prediction in telecom using machine learning in big data platform"	Ahmad, A. et al. (2019)	343	49,00	6,35
3	"A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques..."	Ullah, I. et al. (2019)	263	37,57	4,87
4	"Customer churn prediction system: a machine learning approach"	Lalwani, P. et al. (2022)	214	53,50	8,69
5	"Analyzing and predicting user participations in online health communities"	Wang, X. et al. (2017)	150	16,67	1,79
6	"Churn prediction using comprehensible support vector machine: An analytical CRM application"	Farquad, M. et al. (2014)	139	11,58	1,96
7	"Machine-learning techniques for customer retention: A comparative study"	Sabbeh. (2018)	127	15,88	1,97
8	"Incorporating textual information in customer churn prediction models based on a convolutional neural network"	De Caigny. et al. (2020)	123	20,50	4,42
9	"Leveraging unstructured call log data for customer churn prediction"	Vo, N. et al. (2021)	93	18,60	2,85
10	"Dynamic behavior-based churn prediction in mobile telecom"	Alboukaey, N. et al. (2020)	91	15,17	3,27

Source: Author's elaboration using Biblioshiny

The majority of the influential articles fall into three areas of concentration:

Comparative Model Performance and Selection where articles like the most cited, "A comparison of machine learning techniques for customer churn prediction" by Vafeiadis et al. (2015) with 357 citations, establish the baseline performance metrics and compare traditional algorithms (e.g., Logistic Regression, SVM, Random Forest) for churn, making it a pivotal reference for subsequent studies.

Industry-Specific Application in Telecommunications is a dominant sub-field. The second most cited article, Ahmad et al.'s (2019) study on "Customer churn prediction in telecom using machine learning in big data platform" (343 citations), is a key reference, rigorously demonstrating the application of ML on vast telecom datasets. Similarly, Ullah et al. (2019) focuses on a Random Forest model mainly for the Telecom Sector (263 citations), underscoring the field's maturity in this industry.

Advanced Data Sources and Modeling containing more recent, high-impact papers focus on leveraging non-traditional data and complex neural network models. de Caigny et al. (2020) made an important contribution by revealing the efficacy of using textual information (e.g., customer service notes) via Convolutional Neural Networks (123 citations). Vo et al. (2021) further push this boundary by Leveraging unstructured call log data (93 citations), demonstrating the shift towards holistic customer behavior analysis.

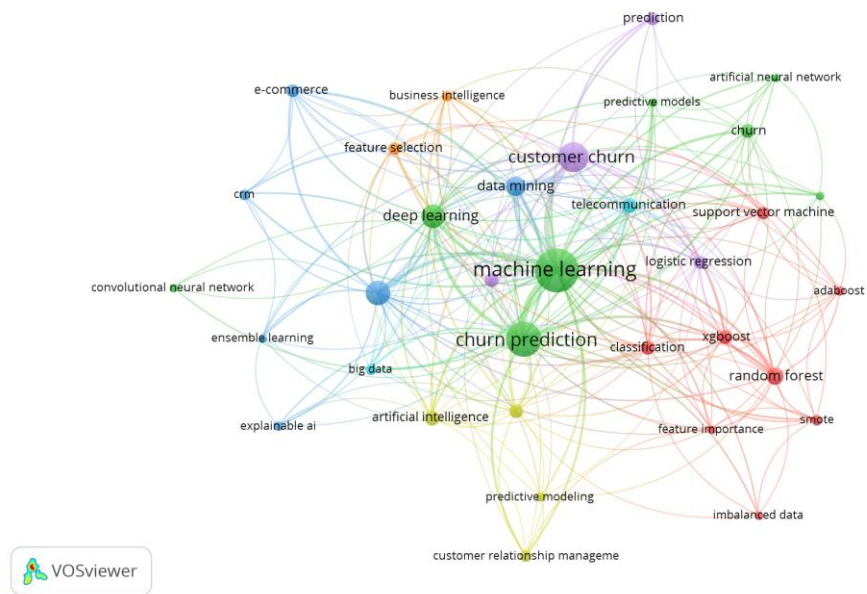
The primary focus of the field is the comparison of algorithms and telecom application, followed by Vafeiadis et al. (2015) and Ahmad et al. (2019). Lalwani et al. (2022) with 214 references proposes a complete machine learning approach that blends the previous approaches and shows that predictive modeling is mixed. Wang et al. (2017), which has 150 citations, is a valuable methodological contribution which extends prediction in advanced analytics beyond traditional industries and into online health communities and demonstrates the cross-disciplinary nature of costumer churn prediction.

In summary, the discipline of predictive customer churn modeling via Machine Learning represents a core focus of modern business intelligence. The most frequently cited articles serve as a definitive map, illustrating the key thematic developments—from early model selection to modern integration of complex data—essential for enhancing retention strategies. These accumulated findings offer authoritative guidance for both researchers and industry experts engaging with this vibrant field.

6. Exploring ML Based Churn Prediction: Sectors and Contexts (RQ6)

Identification of thematic clusters relevant to the research variable was performed using the established bibliometric approach of author keyword co-occurrence analysis. The criteria applied to the analysis (simulated via VOSviewer) were based on author keywords, counting occurrences ≥ 5 , of the 663 keywords, 34 meet the threshold. The result would be visualized in a network map Figure 9, and the data is presented in tabular form in Table 7, distinguishing between seven clusters related to ML-based customer churn prediction.

Figure (09): ML based costumer churn related keywords



Source: Generated by the author using [VOSviewer](#)

Table (07): Research Streams in ML-based Customer Churn Prediction literature

Cluster	Keywords (Core)	Sector and Related Context
1	Adaboost, Random Forest, XGBoost, Classification, SMOTE, Imbalanced Data, Feature Importance, SVM.	Algorithmic Refinement: Ensemble Techniques and Data Pre-processing
2	Deep Learning, CNN, ANN, Churn, Churn Prediction, Machine Learning, Telecommunications, Deep Learning, Predictive Models	Deep Learning & Application: Advanced Architectures in Telecom
3	CRM, Customer Churn Prediction, Data Mining, E-commerce, Ensemble Learning, Explainable AI (XAI)	Strategic Implementation: CRM, Prediction, and Model Interpretation
4	AI, Customer Relationship Management (CRM), Predictive Analytics, Predictive Modeling	High-Level Strategy: AI Integration and Executive Alignment
5	Customer Churn, Customer Retention, Logistic Regression, Prediction	Foundational Core: Basic Models and Business Objectives
6	Big Data, Telecommunication	Data Infrastructure: Scale and Primary Industry Domain
7	Business Intelligence (BI), Feature Selection	BI & Data Preparation: Integrating Analytics with Business Systems

Source: Author’s elaboration based on VOSviewer co-occurrence analysis

Cluster 1: Algorithmic Refinement and Data Pre-processing

highlighted by keywords such as XGBoost, Random Forest, and Adaboost. This reflects its focus on the algorithmic refinement to maximize predictive power. The inclusion of imbalanced data and “SMOTE” (“Synthetic Minority Over-Sampling Technique”) signifies a research avenue focused on solving the fundamental challenge of rare events in churn datasets (Ullah et al., 2019; Wang et al., 2017). The aim is to develop robust classification models that maintain high accuracy even when positive churn cases are scarce.

Cluster 2: Deep Learning and Primary Application

Focuses on concepts such as Deep Learning, “Convolutional Neural Network (CNN)”, and “Artificial Neural Network (ANN)” (Lalwani et al., 2022; Farquad et al., 2014). This outcome underscores the necessity of employing sophisticated model architectures to successfully interpret the intricate patterns inherent in customer data. The strong link between Machine Learning and the Telecommunications sector confirms the ongoing drive to apply cutting-edge methods to the field's historically largest application domain (Sabbeh, 2018). This cluster is crucial for creating high-performance predictive systems that advance beyond classic models to analyze complex features.

Cluster 3: Strategic Implementation and Explainability

Characterized by concepts of “CRM (Customer Relationship Management)”, Ensemble Learning, and Explainable AI (XAI) (Vo et al., 2021; Wu et al., 2021). This reflects a strategic research stream that connects algorithmic development with business utility. This, and similar results, were found by combining forecasting for customer churn and data mining with XAI, suggesting that we need to not only anticipate customer exit but also to understand the drivers of prediction (Sudharsan & Ganesh, 2022). This is a necessary feature in order to translate model output into actionable marketing and retention strategies for E-commerce and other digital platforms.

Cluster 4: High-Level Strategy and AI Integration

Focuses on high-level concepts such as “AI”, “Predictive Modeling”, and “Customer Relationship Management (CRM)” (Sarkar & De Bruyn, 2021; Zhang et al., 2023). This is the most strategic cluster, emphasizing the integration of predictive capabilities into overall business operations. This stream ensures that the technical outputs of predictive analytics are aligned with executive goals, fostering effective Predictive Analytics that inform long-term business strategy and optimize the entire Customer Relationship Management pipeline (Sarkar & De Bruyn, 2021).

Cluster 5: Foundational Core and Business Objective

Cluster 5 is distinguished by keywords such as “Customer Churn”, “Prediction”, “Customer Retention”, and Logistic Regression (Sunarya et al., 2024). This cluster represents the simplest and most direct line of inquiry in the field. It focuses on the fundamental customer retention and its most basic predictive tool (Logistic Regression), reflecting continuous research into straightforward, easily interpretable models. It highlights the importance of simple, accessible models for immediate decision-making and rapid deployment.

Cluster 6: Data Infrastructure and Primary Industry

Cluster 6 focuses exclusively on the environmental factors: “Big Data” and “Telecommunication” (Hammood et al., 2025; Krishna et al., 2024). This cluster highlights the scale and volume challenges that drive the need for sophisticated modeling. The strong link between

Telecommunication and Big Data confirms that this industry continues to define the requirements for handling massive customer interaction records, thereby establishing the necessary data infrastructure for the entire field of study (Sagming et al., 2024).

Cluster 7: BI Integration and Data Preparation

Cluster 7 is closely linked to “Business Intelligence” and “Feature Selection” (Wu et al., 2023). This highlights the crucial intersection between technical model input and organizational data systems. Feature Selection is vital for optimizing model efficiency and avoiding noise, while Business Intelligence is the organizational framework used to consume the final predictive output. This cluster confirms the importance of rigorous data preparation practices within an operational BI framework to ensure predictive success (Sagming et al., 2024)

IV.RESULTS DISCUSSION:

This bibliometric study provides a comprehensive key research patterns in Machine Learning (ML)-driven customer attrition prediction spanning the period 2011–2025. We have addressed our research inquiries by evaluating the fundamental themes that guided the field's progression, identifying the most impactful publications and authors, and discerning structural trends based on highly cited articles. The resultant data confirms a pronounced and continuous evolution in the core topics of ML-based customer churn prediction. The literature illustrates a thematic migration from early, foundational Data Mining approaches and comparative studies (prior to 2020) toward the dominant use and integration of advanced Ensemble Classifiers (e.g., “XGBoost”) and contemporary “Deep Learning architectures” (e.g., “LSTM”, “Bi-LSTM”), a shift mirroring the increasing competitive intensity within sectors such as e-commerce, and banking.

This growth is driven by influential journals such as “IEEE Access”. Key contributions have been consolidated by highly productive authors like “Vu, V.H.” and high-impact authors like “Ravi, V.” (TCN=181). Geographically, the research reveals that India is the undisputed global leader in both publication volume and citation count, though the co-authorship network shows the overall research structure is fragmented, conducted mainly within institutional silos. This highlights that while the domain is globally relevant, the leading activity is heavily concentrated, with countries like South Korea demonstrating high-quality, impactful work despite lower volume.

Despite the clear focus on methodological advances and enhancing predictive performance (which is confirmed by the themes in the most cited articles) a critical analysis reveals a potential overreliance on purely quantitative, "black-box" predictive models. In addition to the fact that Deep Learning and ensembles have received a strong emphasis, model interpretation (Explainable AI or XAI) may not be important due to the increased emphasis on accuracy. A critical discussion of the impact of these layered models on business decision making and ethics remains important (Rana et al., 2025; Zaghloul et al., 2025).

This gap hinders the ability of practitioners to turn predictive data into concrete marketing and retention strategy with confidence and transparency. A future research agenda is needed to fill this gap. Future research should seek to include XAI processes that help decision making, and a focus on transparency and trust in the most critical business applications (He & Ding, 2024). Applied research should also concentrate on cross-vertical studies (i.e., banking and e-commerce churn

drivers) in order to improve the generalizability of findings because lack of such comparative work limits the current model domain. Finally, there is a need for inclusion and research in regions that are not represented adequately (“Africa,” “South America”) that could capture regional customer behavior and give the field a global relevance and application (Imani et al., 2025).

The findings confirm that ML offers a strong capability for improving customer retention, but its deployment is highly uneven among sectors and firms, especially in areas struggling with technological access. A critical scholarly deficiency, however, lies in the failure to investigate the interaction between these sophisticated ML models and established theories from economics and human behavior, limiting the theoretical coherence of the field.

Considering the ethical and research design perspective, the exclusive focus on the Scopus index means that potentially pertinent articles from other platforms or localized databases may have been omitted, thereby imposing a constraint on the analysis's breadth. While comprehensive, the study's scope is confined to Scopus-indexed articles, potentially omitting cutting-edge research and gray literature such as conference papers or technical reports, which often reflect emerging advancements in ML-based customer churn prediction. Moreover, the field's prevailing emphasis on complex quantitative techniques (e.g., “Deep Learning” and “Ensemble Methods”) leads to a deficit in the qualitative comprehension of the psychological factors driving user departure. Incorporating qualitative studies and undertaking cross-cultural or cross-regulatory comparisons is essential to fully gauge the business consequences of these prediction systems.

V.CONCLUSION:

This bibliometric synthesis extracted defining patterns in Machine Learning-driven customer churn prediction research. Our analysis highlights the decisive role of influential publications (e.g., “*IEEE Access*”) and leading researchers (such as “Ravi, V.”) in both consolidating and advancing these technical domains. The study also maps the global distribution of research, clearly identifying centers of expertise and development, primarily concentrated in Asian nations. Through careful consideration of the more frequently quoted articles we identify crucial in-coming directions including the urgent need for Explainable AI (XAI) and ethical data governance. The review also clarifies the practical applications and scope across businesses. Finally, the findings show a serious shift in research towards advanced and powerful deep learning and ensemble techniques. The key, common challenge in the next wave of research is to turn high predictive accuracy into transparent, ethically based, contextually informed business intelligence for customer relationship management.

VI. REFERENCES:

1. Abdullayev, I. S., Prodanova, N. A., Ahmed, M. A., Laxmi Lydia, E., Shrestha, B., Joshi, G. P., & Cho, W. (2023). Leveraging metaheuristics with artificial intelligence for customer churn prediction in telecom industries. *Electronic Research Archive*, 31(8), 4443–4458. <https://doi.org/10.3934/ERA.2023227>
2. Ahmad, A., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0191-6>
3. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/J.JOI.2017.08.007>
4. Arockia Panimalar, S., Krishnakumar, A., & Senthil Kumar, S. (2025). Intensified Customer Churn Prediction: Connectivity with Weighted Multi-Layer Perceptron and Enhanced Multipath Back Propagation. *Expert Systems with Applications*, 265. <https://doi.org/10.1016/j.eswa.2024.125993>
5. Bhuria, R., Gupta, S., Kaur, U., Bharany, S., Rehman, A. U., Hussien, S., Tejani, G. G., & Jangir, P. K. (2025). Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data. *Discover Sustainability*, 6(1). <https://doi.org/10.1007/s43621-025-00807-8>
6. Burnham, J. F. (2006). Scopus database: a review. *Biomedical Digital Libraries*, 3(1), 1. <https://doi.org/10.1186/1742-5581-3-1>
7. Chen, P., & Yang, X. (2025). Premature casual carpooling in Texas: analyzing customer churn in the Metropia experiment with survival analysis and machine learning. *Case Studies on Transport Policy*, 22. <https://doi.org/10.1016/j.cstp.2025.101587>
8. Citation: Elango, B., & Rajendran, P. (2012). Authorship trends and collaboration pattern in the marine sciences literature : a scientometric study. *International Journal of Information Dissemination and Technology*, 2(3), 166–169. <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html>
9. Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*, 19, 31–40. <https://doi.org/10.1016/j.asoc.2014.01.031>
10. Fatima, G., Khan, S., Aadil, F., Kim, D., Atteia, G. E., & Alabdulhafith, M. I. (2024). An autonomous mixed data oversampling method for AIOT-based churn recognition and personalized recommendations using behavioral segmentation. *PeerJ Computer Science*, 10, 1–32. <https://doi.org/10.7717/peerj-cs.1756>
11. Gaidhani, Y. A., Ramesh, J. V. N., Singh, S., Dagar, R., Subha Mastan Rao, T. S. M., Godla, S. R., & Baker El-Ebiary, Y. A. (2025). AI-Driven Predictive Analytics for CRM to Enhance Retention Personalization and Decision-Making. *International Journal of Advanced Computer Science and Applications*, 16(4), 552–563. <https://doi.org/10.14569/IJACSA.2025.0160456>
12. Hammood, W. A., Hammood, O. A., Mohamad Asmara, S., Lah, M. S. C., & Rasyidah.

- (2025). Customer Behavior Analysis for Forecasting Customer Attrition: An Artificial Intelligence Approach. *International Journal on Advanced Science, Engineering and Information Technology*, 15(3), 680–685. <https://doi.org/10.18517/ijaseit.15.3.12816>
13. He, C., & Ding, C. H. Q. (2024). A novel classification algorithm for customer churn prediction based on hybrid Ensemble-Fusion model. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-71168-x>
 14. Imani, M., Joudaki, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning. *Machine Learning and Knowledge Extraction*, 7(3). <https://doi.org/10.3390/make7030105>
 15. Joy, U. G., Hoque, K. E., Nazim Uddin, M., Chowdhury, L., & Park, S. (2024). A Big Data-Driven Hybrid Model for Enhancing Streaming Service Customer Retention Through Churn Prediction Integrated With Explainable AI. *IEEE Access*, 12, 69130–69150. <https://doi.org/10.1109/ACCESS.2024.3401247>
 16. Kate, P., Ravi, V., & Gangwar, A. K. (2023). FinGAN: Chaotic generative adversarial network for analytical customer relationship management in banking and insurance. *Neural Computing and Applications*, 35(8), 6015–6028. <https://doi.org/10.1007/s00521-022-07968-x>
 17. Krishna, R., Jayanthi, D., Shylu Sam, D. S., Kavitha, K., Maurya, N. K., & Benil, T. (2024). Application of machine learning techniques for churn prediction in the telecom business. *Results in Engineering*, 24. <https://doi.org/10.1016/j.rineng.2024.103165>
 18. Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
 19. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317–323. <http://www.jstor.org/stable/24529203>
 20. Luong, T. T., Luong, V. G., Tran, A. H. T., & Nguyen, T. M. (2025). Application of Machine Learning Techniques for Customer Churn Prediction in the Banking Sector. *Interdisciplinary Journal of Information, Knowledge, and Management*, 20. <https://doi.org/10.28945/5469>
 21. Ngo, V. B., & Vu, V. H. (2024). Multi-Level Machine Learning Model to Improve the Effectiveness of Predicting Customers Churn Banks. *Cybernetics and Information Technologies*, 24(3), 3–20. <https://doi.org/10.2478/cait-2024-0022>
 22. Nguyen, T. Van, Ngo, V. B., & Vu, V. H. (2025). Enhancing Telecom Churn Prediction Using an Advanced Stacking Model. In *Lecture Notes in Networks and Systems: Vol. 1205 LNNS*. https://doi.org/10.1007/978-3-031-80943-9_11
 23. Ong, J. X., Tong, G. K., Kok-Chin, K. C., & Haw, S. C. (2024). Enhancing Customer Churn Prediction With Resampling: A Comparative Study. *TEM Journal*, 13(3), 1927–1936. <https://doi.org/10.18421/TEM133-20>
 24. Rana, M. S., Chouksey, A., Hossain, S., Sumsuzoha, M., Bhowmik, P. K., Hossain, M.,

- Rabby, M. F., Gurung, N., & Zeeshan, M. A. F. (2025). AI-Driven Predictive Modeling for Banking Customer Churn: Insights for the US Financial Sector. *Journal of Ecohumanism*, 4(1), 3478–3497. <https://doi.org/10.62754/joe.v4i1.6188>
25. Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications*, 9(2), 273–281. <https://doi.org/10.14569/IJACSA.2018.090238>
 26. Sagming, M. N., Heymann, R., & Visaya, M. V. V. (2024). Using topological data analysis and machine learning to predict customer churn. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-01020-6>
 27. Sarkar, M., & De Bruyn, A. (2021). LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning. *Journal of Interactive Marketing*, 53, 80–95. <https://doi.org/10.1016/j.intmar.2020.07.002>
 28. Subramanian, D., Ajitha, & Maidin, S. S. (2025). Unveiling Hybrid Model with Naive Bayes, Deep Learning, Logistic Regression for Predicting Customer Churn and Boost Retention. *Journal of Applied Data Sciences*, 6(2), 1379–1391. <https://doi.org/10.47738/jads.v6i2.675>
 29. Sudharsan, R., & Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connection Science*, 34(1), 1855–1876. <https://doi.org/10.1080/09540091.2022.2083584>
 30. Sunarya, P. A., Rahardja, U., Chen, S., Li, Y., & Hardini, M. (2024). Deciphering Digital Social Dynamics: A Comparative Study of Logistic Regression and Random Forest in Predicting E-Commerce Customer Behavior. *Journal of Applied Data Sciences*, 5(1), 100–113. <https://doi.org/10.47738/jads.v5i1.155>
 31. Tonati, S., Di Vece, M., Giannotti, F., & Pellungrini, R. (2025). Counterfactual ensembles for interpretable churn prediction: from real-world to privacy-preserving synthetic data. *Machine Learning*, 114(10). <https://doi.org/10.1007/s10994-025-06880-4>
 32. Trinh, T., Vu, V. H., & Nguyen, T. Van. (2025). A multi-task test case optimization framework with integrated explainable AI for customer churn prediction. *Journal of Supercomputing*, 81(14). <https://doi.org/10.1007/s11227-025-07816-4>
 33. Ullah, I., Raza, B., Malik, A. K. K., Imran, M. U. M., Islam, S. ul, & Kim, S. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, 7, 60134–60149. <https://doi.org/10.1109/ACCESS.2019.2914999>
 34. Ullah, I., Rios, A., Gala, V., & McKeever, S. (2022). Explaining deep learning models for tabular data using layer-wise relevance propagation. *Applied Sciences (Switzerland)*, 12(1). <https://doi.org/10.3390/app12010136>
 35. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9. <https://doi.org/10.1016/j.simpat.2015.03.003>
 36. Vasu, M., & Ravi, V. (2011). A hybrid under-sampling approach for mining unbalanced datasets: Applications to banking and insurance. *International Journal of Data Mining*,

- Modelling and Management*, 3(1), 75–105. <https://doi.org/10.1504/IJDMMM.2011.038812>
37. Vo, N. N. Y., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, 212. <https://doi.org/10.1016/j.knosys.2020.106586>
38. Vu, V. H. (2024a). An Efficient Customer Churn Prediction Technique Using Combined Machine Learning in Commercial Banks. *Operations Research Forum*, 5(3). <https://doi.org/10.1007/s43069-024-00345-5>
39. Vu, V. H. (2024b). Predict customer churn using combination deep learning networks model. *Neural Computing and Applications*, 36(9), 4867–4883. <https://doi.org/10.1007/s00521-023-09327-w>
40. Wang, X., Xie, L., Wang, H., Xing, X., Wan, W., Wu, Z., Ma, X., & Li, Q. (2025). Deciphering Explicit and Implicit Features for Reliable, Interpretable, and Actionable User Churn Prediction in Online Video Games. *IEEE Transactions on Visualization and Computer Graphics*, 31(9), 5990–6007. <https://doi.org/10.1109/TVCG.2024.3487974>
41. Wei, Q. (2025). Personalized and Contextualized Data Analysis for E-Commerce Customer Retention Improvement With Bi-LSTM Churn Prediction. *IEEE Transactions on Consumer Electronics*, 71(2), 4406–4414. <https://doi.org/10.1109/TCE.2024.3376672>
42. Wu, R. Y., Hu, Y., & Chou, E. (2025). Predicting the churn patterns of monetizers and non-monetizers: exploring the influence of behavioral variability in churn prediction. *Internet Research*. <https://doi.org/10.1108/INTR-05-2024-0747>
43. Wu, S., Yau, W. C., Ong, T. S., & Chong, S. C. (2021). Integrated Churn Prediction and Customer Segmentation Framework for Telco Business. *IEEE Access*, 9, 62118–62136. <https://doi.org/10.1109/ACCESS.2021.3073776>
44. Yang, C., Xia, G., Zheng, L., Zhang, X., & Yu, C. (2025). Customer Churn Prediction Based on Coordinate Attention Mechanism with CNN-BiLSTM. *Electronics (Switzerland)*, 14(10). <https://doi.org/10.3390/electronics14101916>
45. Zaghloul, M., Barakat, S. I., & Rezk, A. (2025). Enhancing customer retention in Online Retail through churn prediction: A hybrid RFM, K-means, and deep neural network approach. *Expert Systems with Applications*, 290. <https://doi.org/10.1016/j.eswa.2025.128465>
46. Zhang, X., Guo, F., Chen, T., Pan, L., Beliakov, G. V., & Wu, J. (2023). A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(4), 2188–2216. <https://doi.org/10.3390/jtaer18040110>