Journal of Scientometric Research, 2024; 13(2):365-381.
https://www.jscires.org

Research Article

Exploring the Predictive Analytics Frontier in Business: A Bibliometric Journey

Joshy John[1], Jeena Joseph[2,*], Liance Mathew[3], Shamini James[4], Jobin Jose[5]

[1]Department of Business Administration, Marian College Kuttikkanam Autonomous, Idukki, Kerala, INDIA.
[2]Department of Computer Applications, Marian College Kuttikkanam Autonomous, Idukki, Kerala, INDIA.
[3]Department of Commerce, The Cochin College, Kochi, Kerala, INDIA.
[4]Department of Humanities, Amal Jyothi College of Engineering, Kottayam, Kerala, INDIA.
[5]Department of Library, Marian College Kuttikkanam Autonomous, Idukki, Kerala, INDIA

ABSTRACT

Predictive analytics has gained significant attention in business as organizations seek to leverage data-driven insights for informed decision-making. The study uses rigorous bibliometric analysis to map the landscape of predictive analytics in business, revealing its evolution, trends, and research tendencies. A sizeable selection of academic publications from the Scopus database is chosen to identify key themes, notable authors, well-known journals, and emerging research areas in predictive analytics. The study extracts and analyses bibliographic data, including publication trends, citation patterns, and co-authorship networks, using cutting-edge tools like CiteSpace, VOSviewer, and Biblioshiny. The findings of this study provide important information regarding how predictive analytics are now being employed in the business sector. The most significant research findings, innovative breakthroughs, and major research clusters are highlighted, exposing the leading research paths and areas of interest. Additionally, emerging trends, cross-disciplinary collaborations, and potential research needs are identified in the study. The intellectual landscape of predictive analytics in business is mapped in this study to offer a comprehensive picture of the corpus of knowledge. This information will help scholars, practitioners, and policymakers navigate and contribute to this fast-developing area. The insights gained from this bibliometric analysis can guide future research endeavors, inform strategic decision-making, and foster stakeholder collaboration in leveraging predictive analytics for business success.

Keywords: Predictive analytics, Business, Bibliometric Analysis, CiteSpace, VOSviewer, Biblioshiny.

Correspondence:
Jeena Joseph
Assistant Professor, Department of Computer Applications, Marian College Kuttikkanam Autonomous, Idukki, Kerala, INDIA.
Email: jeena.joseph@mariancollege.org
ORCID: 0000-0003-4070-5868

Received: 27-06-2023;
Revised: 21-12-2023;
Accepted: 05-07-2024.

INTRODUCTION

Predictive analytics in business refers to the practice of using data, statistical algorithms, and machine learning techniques to analyze past and present information in order to make predictions about future events, trends, and outcomes.[1] Organizations may use predictive analytics to get valuable insights, forecast consumer behavior, optimize operations, reduce risks, and make wise business decisions through leveraging historical and real-time data. In predictive analytics, relevant data is gathered from various sources, cleaned up, preprocessed, statistically analyzed, and then built prediction models.[2] These models are trained using historical data that includes relevant variables and their associated outcomes. The models may forecast future occurrences or outcomes based on newly acquired inputs once constructed and validated. Regression analysis, time series analysis, data mining, machine learning, and artificial intelligence are just a few of the techniques and algorithms used in predictive analytics.[2-6] These approaches allow companies to discover trends, detect associations, and extract hidden information that may be utilized for making precise forecasts.[3,5]

The use of predictive analytics in business is pervasive across numerous sectors and functional divisions. Predictive analytics, for instance, can be used in marketing to forecast customer lifetime value, identify target customers, and personalize advertising campaigns.[7] Credit scoring, fraud detection, and portfolio optimization are all possible uses for it in the financial sector. It can help with demand forecasting, inventory optimization, and logistics planning in supply chain management.[8] Predictive analytics can also be used in manufacturing, healthcare, human resources, and many other industries to improve operational effectiveness and strategic decision-making.[9,10]

Predictive analytics has many advantages for businesses. It enables businesses to make data-driven decisions, enhance customer

DOI: 10.5530/jscires.13.2.29

Publishing Partner : EManuscript Tech. [www.emanuscript.in]

happiness, better allocate resources, minimize expenses, and gain a competitive edge in the market.[11] By predicting future results, businesses may respond proactively to problems, take advantage of opportunities, and optimize their operations. It is crucial to remember that predictive analytics is not flawless. It greatly depends on the accuracy of the models and algorithms and the data's quality and relevancy. Furthermore, ethical issues like biases and data privacy must be considered when deploying predictive analytics systems.

To evaluate publication patterns, citation, and collaboration, bibliometrics is the quantitative analysis of scientific publications, such as articles, papers, or patents.[12-14] It could assess the level of publications, probe into the key research areas, and even forecast the course of future research.[15,16] It also makes inferences from abundant data possible by its capacity to manage massive databases.[17] This method of study examines many scholarly articles to discover how research on a specific topic has evolved.[18,19] It uses various quantitative techniques to uncover novel patterns and trends, including co-authorship analysis, keyword distribution, citation counts, and other available data.[20,21]

Science-related literature can be visualized and examined for trends and patterns using Citespace. It is primarily used in bibliometrics and scientometrics, which assess scientific publications' structure, characteristics, and citation patterns.[22] Bibliometric networks are constructed and exhibited using a tool called VOSviewer. The term stands for "Visualization of Similarities" viewer. VOSviewer enables academics to analyze enormous volumes of scientific literature by building network visualizations based on bibliographic data. Co-authorship networks, citation networks, keyword co-occurrence networks, and more may all be explored using it.[16,23-27] A user-friendly software program that offers many capabilities for the quantitative analysis of bibliometric data is the bibliometrix R-package. It has many algorithms explicitly designed for statistical analysis and producing scientific visualizations. Moreover, this package includes a user-friendly web interface tool called Biblioshiny, which was specifically created to help people who are not proficient in coding to conduct bibliometric analysis.[28,29]

The objectives of conducting a bibliometric analysis on Predictive Analytics in Business are:

Identify Key Trends and Themes: Analyze the literature to identify key trends and themes within the field of Predictive Analytics in Business. This can help researchers and practitioners understand the evolving landscape of predictive analytics and its applications in business.

Mapping Research Networks: Explore the network of researchers, institutions, and collaborations in the field. Comprehending the associations and collaborative efforts can offer valuable perspectives on the information-sharing process and the collaborative nature of Predictive Analytics research.

Assessing Research Productivity: Measure the productivity of researchers and institutions in terms of publications and citations. This can assist in highlighting significant efforts and identifying key figures in the subject.

Identifying Core Journals and Publications: Determine the most influential journals and publications in the area of Predictive Analytics in Business. Practitioners may use this information to remain up to speed on the most recent advancements and researchers can use it to help choose where to publish their work.

Tracking Citation Patterns: Analyze citation patterns to identify seminal works and highly cited papers.

Benchmarking and Comparative Analysis: Compare the research output and impact of different countries, institutions, or researchers. This can help identify leaders in the field and assess the global distribution of research contributions.

Informing Decision-Making: Provide a data-driven foundation for decision-making by academic institutions, funding agencies, and businesses interested in predictive analytics. This can guide resource allocation and strategic planning.

The bibliometric analysis on Predictive Analytics in Business addresses a notable gap in the existing literature. Although Predictive Analytics in business has attracted increased research attention, there is a recognized need for a comprehensive and current synthesis of scholarly contributions to this domain. This study employs advanced bibliometric tools such as CiteSpace, VOSviewer, and Bibloshiny, offering a unique and in-depth perspective on the landscape of predictive analytics research. To the best of available knowledge, there is a scarcity of reviews utilizing similar bibliometric methodologies within this specific intersection of predictive analytics and business. This analysis aims to fill this gap by providing insights into the evolving trends, influential works, and knowledge networks within Predictive Analytics in Business, thereby contributing to a more holistic understanding of the field.

The analysis highlights a robust growth in the application of predictive analytics across general business practices, there is a notable paucity of research focusing on its impact within niche markets and specialized industries. This gap suggests a significant opportunity for future research aimed at understanding the unique challenges and benefits of predictive analytics in sectors such as Small and Medium-sized Enterprises (SMEs), which have been largely overlooked. Additionally, while the research productivity assessment identifies key contributors and institutions driving the field forward, there is a conspicuous absence of interdisciplinary research collaborations that integrate business insights with technological advancements. This gap underscores the potential for groundbreaking discoveries through

cross-disciplinary research, bridging predictive analytics with fields such as behavioral science and ethics. The benchmarking and comparative analysis conducted reveal a geographical concentration of research efforts, with limited contributions from developing countries. This disparity highlights a critical need for fostering predictive analytics research in diverse economic contexts, ensuring that the benefits of these technologies are universally accessible and tailored to local business ecosystems. These identified gaps, firmly rooted in this bibliometric analysis and supported by recent scholarly contributions, underscore the rich avenues for future research in Predictive Analytics in Business.

MATERIALS AND METHODS

Scopus was chosen as the primary bibliographical data source for this study[30] because it covers a wider range of journals compared to other databases. The publications were retrieved on 20th June 2023 using the keywords "Predictive Analytics" and "business." The inclusion and exclusion criteria for this bibliometric analysis on Predictive Analytics in Business were meticulously defined to ensure the selection of studies that align with the specific focus and scholarly standards of the research. Included in the analysis are journal articles and conference papers published between 1989 and 2023, encompassing a diverse range of sources. Language restrictions were not imposed to capture a global perspective on the subject. The PRISMA approach for the selection of papers for bibliometric analysis is illustrated in Figure 1. It is a three-phase procedure beginning with the identification and extraction of data for analysis from the Scopus database. The key inclusion criterion emphasizes the direct relevance of studies to the application of Predictive Analytics within a business context. This encompasses research exploring predictive modelling, data analytics, machine learning, and similar techniques, specifically within the realm of business operations. The types of documents considered for inclusion are Articles and Conference papers. Conversely, exclusion criteria were established in the second phase to maintain methodological rigour. Non-peer-reviewed materials, such as books, book chapters, reviews, and editorials, were excluded, as were incomplete or inaccessible articles. Additionally, duplicates were eliminated to prevent redundancy in the dataset. By adhering to these criteria, the analysis aims to provide a comprehensive and rigorous examination of scholarly contributions to Predictive Analytics in Business. The findings were stored as "CSV" files and bibliometric analysis was performed on the data using CiteSpace version 6.2.R3 (Advanced), VOSviewer and Biblioshiny software. The main aspects of this investigation are summarized in Table 1.

RESULTS

Annual Scientific Production

Figure 2 illustrates the relationship between the number of published works and the corresponding years using the Biblioshiny visual representation. The graph reveals a steady increase in publication volume within the specific domain from 2006 to 2015, followed by a subsequent decline. However, from 2016 to 2020, there is a significant resurgence in the quantity of publications, which is then followed by a substantial decrease.

Most Significant Authors

The investigation into predictive analytics in business involved a group of 3656 authors who wrote articles on the topic. The significance of these authors was assessed based on the quantity of articles they had published. Loos P emerged as the most influential author, having published 9 articles. Krumeich J and Zhang Y followed closely with 8 articles each, while Cheng Y and Werth D had 7 articles each. Table 2 presents an overview of the publication counts for these noteworthy authors who consistently produced more than five articles over a specific time period. Through their extensive experience and expertise, they have established themselves as prominent figures in their respective fields, thus exerting significant influence. Figure 3 visually represents the productivity of these authors by showing the number of articles they have published over time.

Most relevant sources and affiliations

In our analysis, we investigated 744 different journal sources and collected 1210 publications. Among these, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) emerged as the

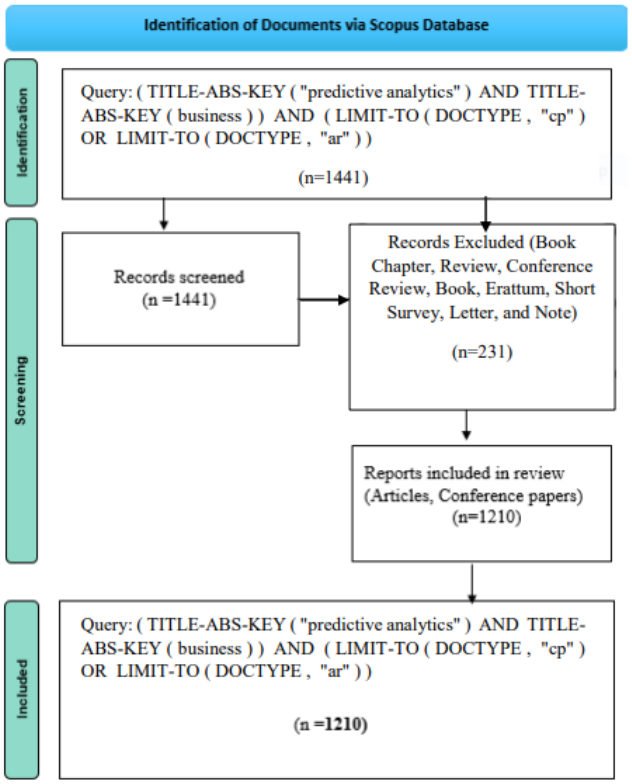

Figure 1: The PRISMA approach for the selection of papers.

most productive journal, publishing a maximum of 40 articles. The ACM International Conference Proceeding Series came in a close second with 37 publications. Figure 4 lists the top 10 journals that contributed the most research papers on predictive analytics in the business field.

Figure 5 illustrates the primary entities that have made significant contributions to research papers on predictive analytics in the business domain. Queensland University of Technology has published a maximum of 23 papers leading the list. Following closely are University of Minho with 20 publications and Hangzhou Dianzi University with 14 publications. Additionally, notable institutions such as Ghent University, Korea Institute of Science and Technology Information, University of Ljubljana, and University of Rome La Sapienza have also conducted substantial research in this field.

Three Field Plot of keyword, author and source

Figure 6 showcases an illustration that examines the relationship between keywords (on the left), authors (in the middle), and sources (on the right) in the realm of predictive analytics within business literature. The objective of the investigation was to identify commonly used keywords in the literature by various authors and published journals. The analysis of the primary keywords, authors, and sources unveiled several significant phrases, including "predictive analytics," "big data," "business analytics," "data science," and "deep learning." It was noted that many authors, such as Krumeich J, Loos P, and Werth D, utilized these keywords in their work and published in sources such as Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) and Lecture Notes in Business Information Processing.

Table 1: Essential aspects of the investigation.

Description	Results
Search Query	(TITLE-ABS-KEY ("predictive analytics") AND TITLE-ABS-KEY (business)) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar"))
Main Information About Data	
Timespan	1989:2023
Sources (Journals, Books, Etc.,)	744
Documents	1210
Annual Growth Rate %	11.38
Document Average Age	4.55
Average Citations Per Doc	10.97
References	36541
Document Contents	
Keywords Plus (Id)	6825
Author's Keywords (De)	3334
Authors	
Authors	3656
Authors Of Single-Authored Docs	139
Authors Collaboration	
Single-Authored Docs	145
Co-Authors Per Doc	3.39
International Co-Authorships %	18.35
Document Types	
Article	511
Conference Paper	699

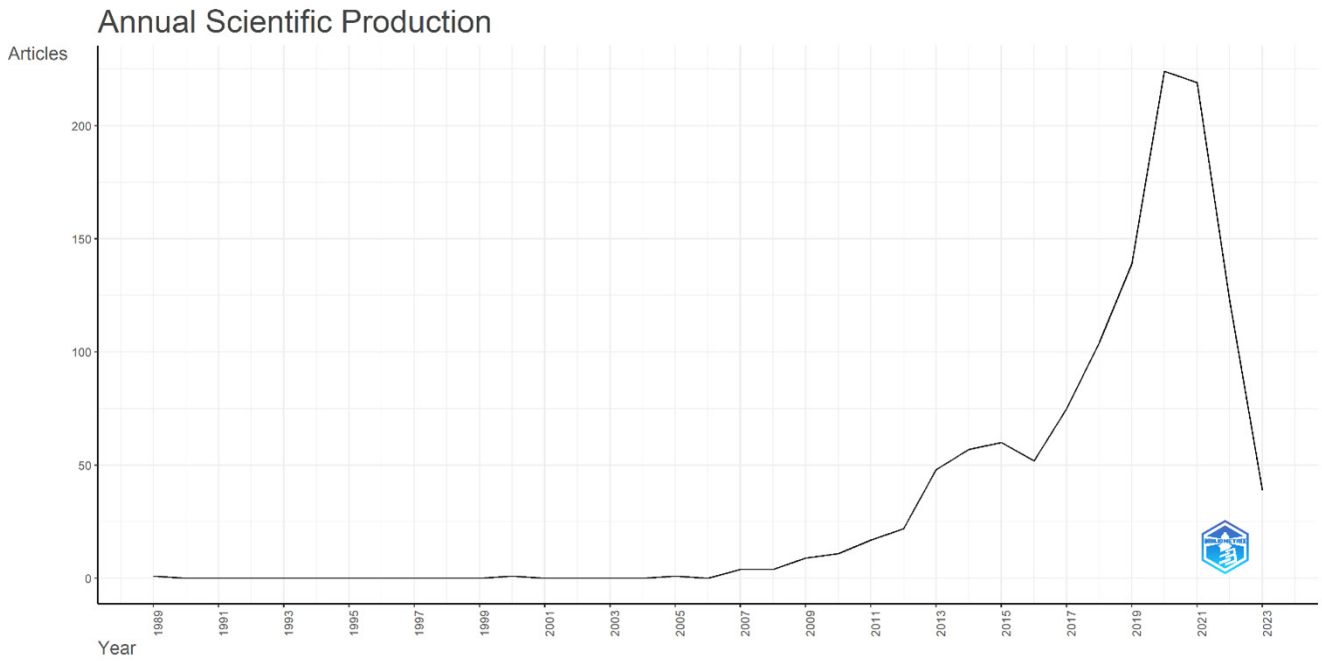

Figure 2: Annual scientific production from 1989 to 2023 visualized using Biblioshiny.

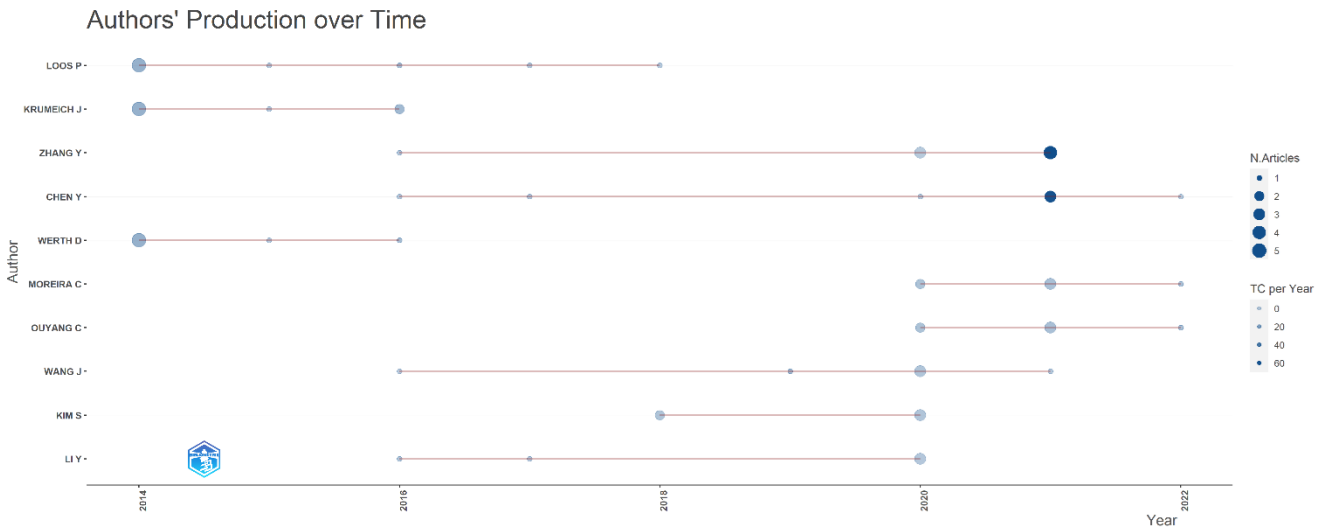

Figure 3: Authors' Production over Time.

Table 2: The authors having more than 5 articles.

Authors	Articles
Loos P	9
Krumeich J	8
Zhang Y	8
Chen Y	7
Werth D	7
Moreira C	6
Ouyang C	6
Wang J	6

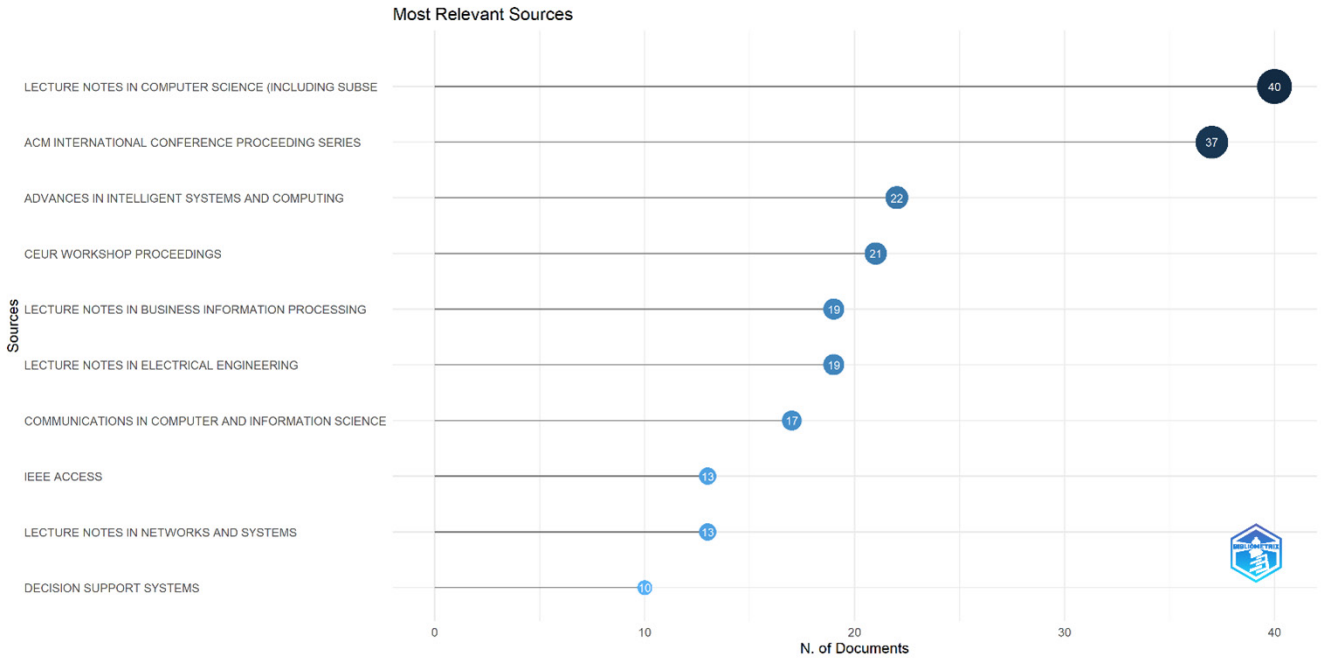

Figure 4: The top ten sources in terms of the number of publications.

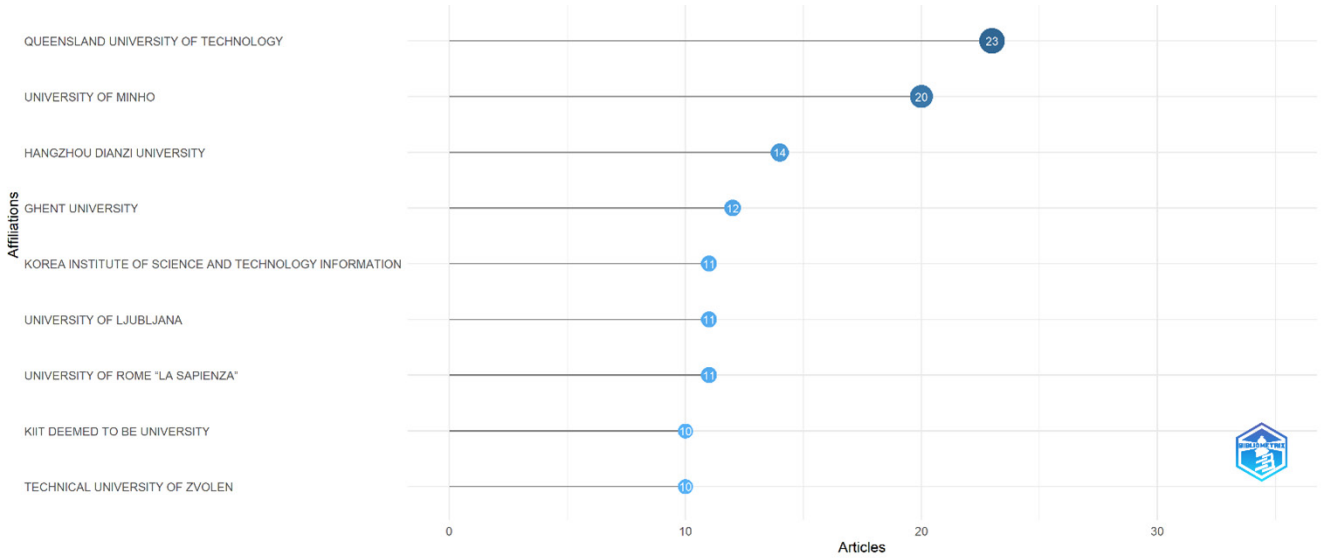

Figure 5: Most relevant affiliations in terms of the number of publications.

Most Frequent Words and Word Cloud of the keywords

Figure 7 presents a visual representation illustrating the common phrases frequently used along with their respective frequencies. The graph emphasizes the top ten keywords that are most commonly utilized, such as "predictive analytics (954)," "forecasting (347)," "decision making (185)," "data analytics (174)," "data mining (164)," "machine learning (162)," "big data (157)," "sales (138)," learning systems (132)," and "decision trees (104)."

Figure 8 and Figure 9 exhibit word clouds that illustrate the authors' keywords and keywords plus. These word clouds delve into commonly used phrases found in the articles being studied, with a focus on specific subject areas. A word cloud is a visualization method that condenses textual input into succinct terms and represents their significance by their size within the cloud. This approach establishes connections between the primary keywords employed by the authors, which are interspersed among the more prominent keywords like "predictive analytics," "machine learning," "forecasting," "decision making," "big data," "data mining," and related terms.

Thematic Map

We have identified several research areas that can significantly enhance the understanding of our findings. These areas can be

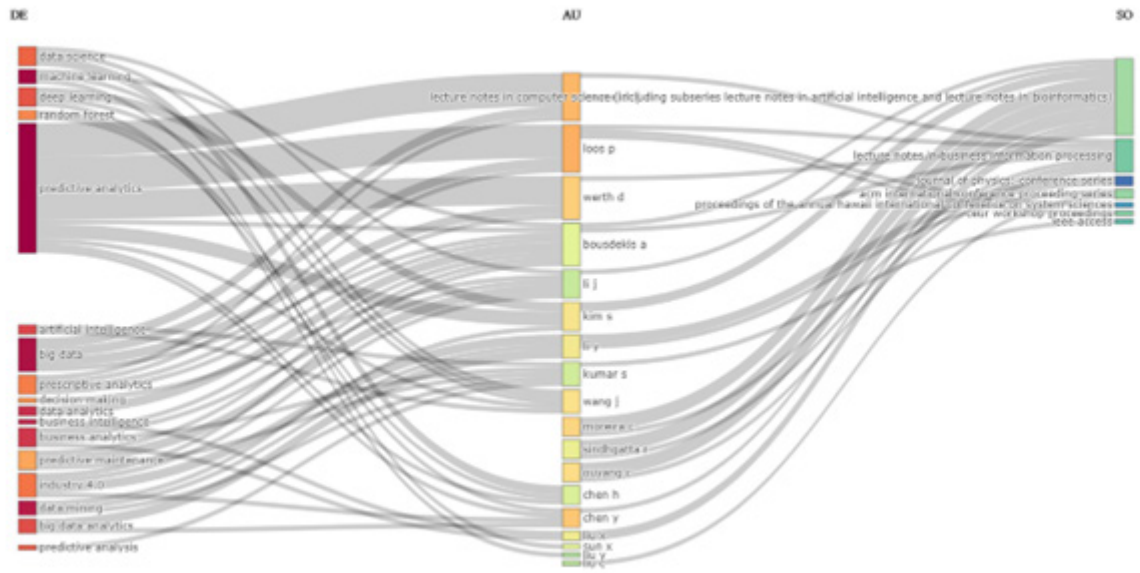

Figure 6: Three Field Plot (keyword (left), author (middle) and source (right) using Biblioshiny).

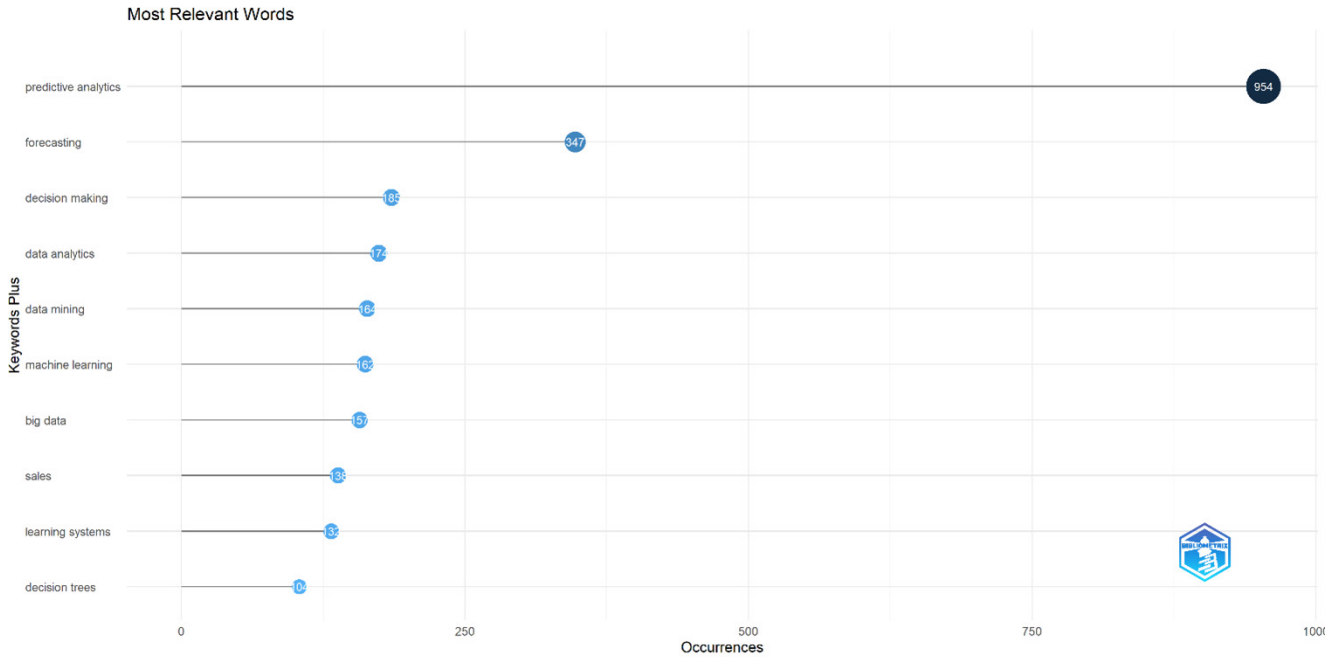

Figure 7: The most frequently used author keywords created using Biblioshiny

analyzed and grouped using a strategic diagram, shown in Figure 10. The diagram presents a thematic map that is based on two main aspects: density (vertical axis) and centrality (horizontal axis). Centrality indicates the importance of a specific research area, while density reflects the level of development it has achieved. The diagram is divided into four distinct sections, each representing different types of themes. The lower left section contains emerging or declining themes, which are new topics that may either gain importance or lose relevance in the research field. Some examples of these themes are "human," "humans," "prediction," "algorithms," and "article." Moving to the lower right section of the diagram, we find basic or transversal themes. These themes have been extensively researched, as indicated by their low density, but they also possess high centrality, indicating their fundamental nature and broad applicability. In the upper left part of the diagram, we see themes with high density but lower centrality. These themes are highly developed but somewhat disconnected from other areas of research. Finally, the upper right part of the diagram represents high density and centrality themes. These themes, referred to as motor themes, are

Figure 8: A Visualized Word-cloud of Authors Keywords Created using Biblioshiny.

Figure 9: A Visualized Word-cloud of Keywords Plus Created using Biblioshiny.

crucial and extensively developed, playing a central role in the research landscape. In this study, the motor themes identified are "predictive analytics," "decision making," "data mining," "big data," and "data analytics." The factors linked to each specific theme determine the size of its representation on the thematic map.

Co-Occurrence of Keywords

Keyword co-occurrence analysis has the potential to be an effective method for discovering significant patterns within textual data and obtaining a deeper understanding of the connections between various concepts and keywords. In this analysis, the nodes' sizes correspond to the frequency of the keywords, with larger nodes representing higher frequencies. The thickness and length of the edges indicate the proximity of interactions between two nodes. Furthermore, the colors of the keyword nodes denote the clusters to which they belong.[31] The VOSviewer software detects 9 separate groups among the 513 items shown in Figure 11. The first group comprises 92 items and includes terms such as "predictive analytics," "classification algorithm," and "machine learning." The second group, referred to as cluster 2, consists of 91 items, while cluster 3 contains 78 items. Both cluster 4 and cluster 5 consist of 63 items each. Cluster 6 contains 38 items, cluster 7 contains 31 items, cluster 8 contains 30 items, and cluster 9 contains 27 items.

Author Co-Authorship Analysis

VOSviewer generated a visual representation of author co-authorship analysis called a cluster map (Figure 12). In this analysis, 312 authors (who have published more than 2 papers) were included, and based on their co-authorship patterns, they were grouped into 10 distinct clusters. Authors who had a strong collaborative relationship were assigned the same color within the map. The author with the highest Total Link Strength (TLS), which measures the overall strength of connections, is Loos P with a TLS value of 23.

Country Co-Authorship Analysis

We utilized VOSviewer, a tool for visualizing networks, to create a visual representation of the collaborative relationships between countries in the field of predictive analytics within business publications. The distribution maps generated from this analysis offer valuable insights, aiding researchers in identifying potential partners for collaboration. Among the various clusters formed, the largest cluster consists of 59 countries grouped into nine clusters. Figure 13 provides a graphical depiction of the collaborative network among countries that have published

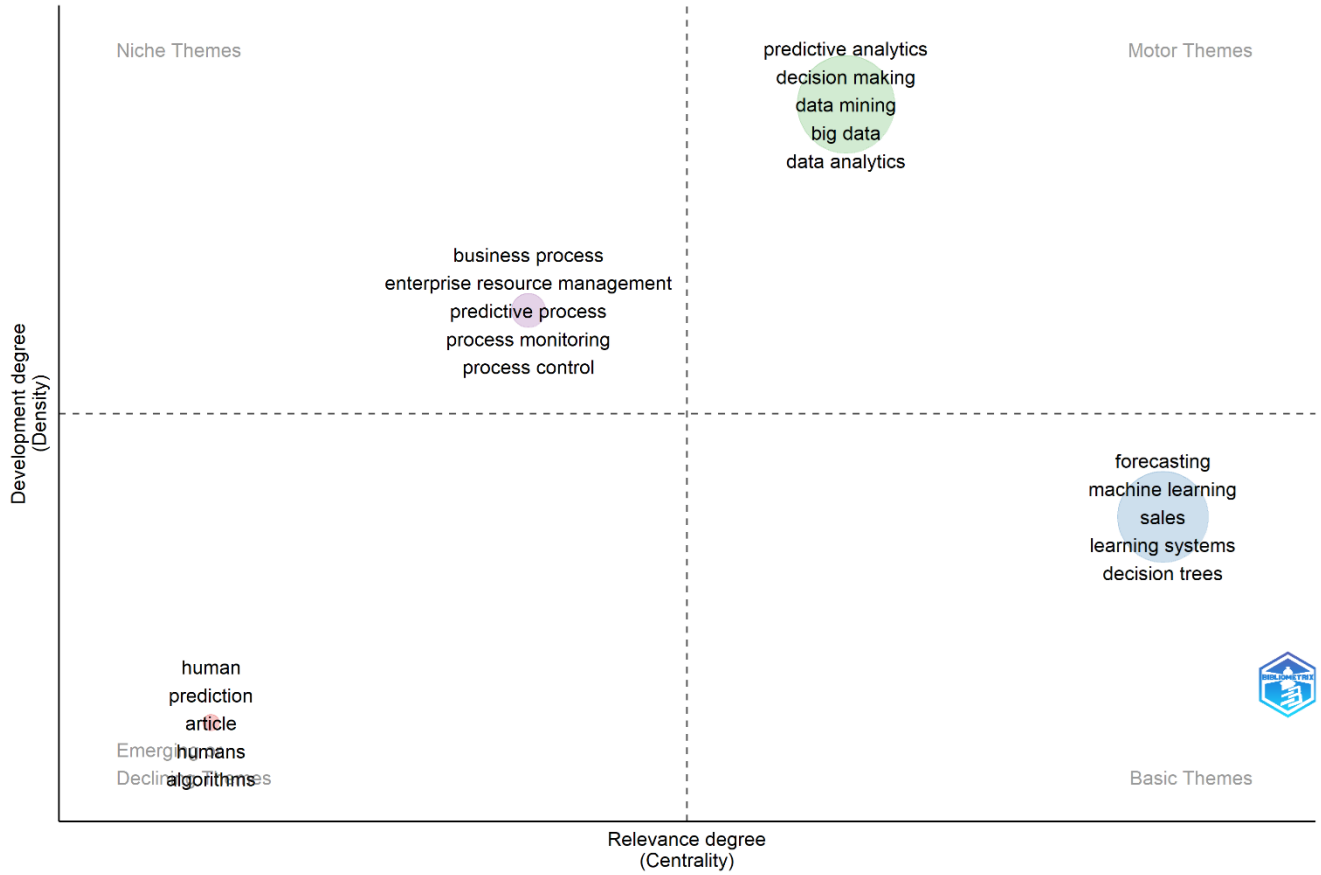

Figure 10: Thematic Map using Biblioshiny.

more than three documents. Clusters are formed based on the frequency of co-occurring terms associated with each country. The more frequently these terms co-occur, the stronger the association, and they are represented by specific colors assigned to each cluster. The size of the nodes in the visualization corresponds to the number of publications produced by each country, while the thickness of the lines connecting them signifies the degree of collaboration. For instance, the collaboration between the United States and the United Kingdom is represented by a thick line with a link strength of 12, indicating a strong partnership. In contrast, the link strength between the United States and India is 10. Countries sharing the same color are grouped together within a cluster. Notably, the United States exhibits the highest level of collaboration with other countries across the globe.

Country/Region distribution of published papers

Examining the publication venues of articles can uncover the academic connections between countries or regions, presenting fresh perspectives for assessing the influence of a particular country or region on scholarly advancements. This analysis also facilitates a broader comprehension of how research is distributed across various locations within the field of predictive analytics in business studies. To visually represent the significant contributors to research output from 2005 to 2023, Figure 13 illustrates the countries or regions involved. It is evident from the data depicted in Figure 14 and Table 3 that the United States holds a prominent position, leading in the number of published research papers. The country name appears notably larger than others in the figure, symbolizing its dominance. Following closely behind, India ranks second with a frequency of 206, while China closely trails with a frequency of 114. Furthermore, Table 3 offers insights into the ten most productive countries in predictive analytics in business research. The table reveals active research engagement in this field by both developed and developing countries, including the United States, the United Kingdom, India, China, and Germany.

In addition to how often it appears, centrality is another important factor that determines how well a node can establish connections with other nodes. Nodes that have high centrality, which falls within the range of 0 to 1, serve as essential bridges between different groups that display patterns of change. This measure of centrality also reveals the primary themes and popular subjects within a particular network. According to Table 3, the United States stands out with a centrality score of 0.44, indicating its crucial role in fostering connections and collaborations among researchers from different countries in the field of predictive analytics in business.

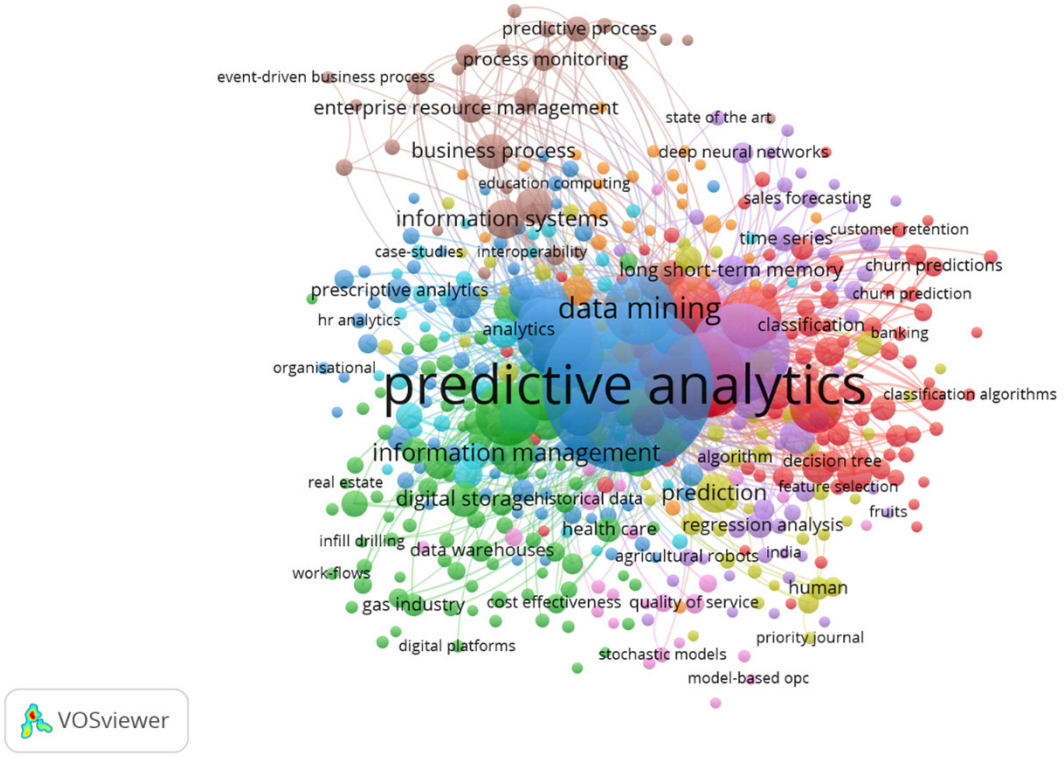

Figure 11: Co-occurrence network generated using VOSviewer.

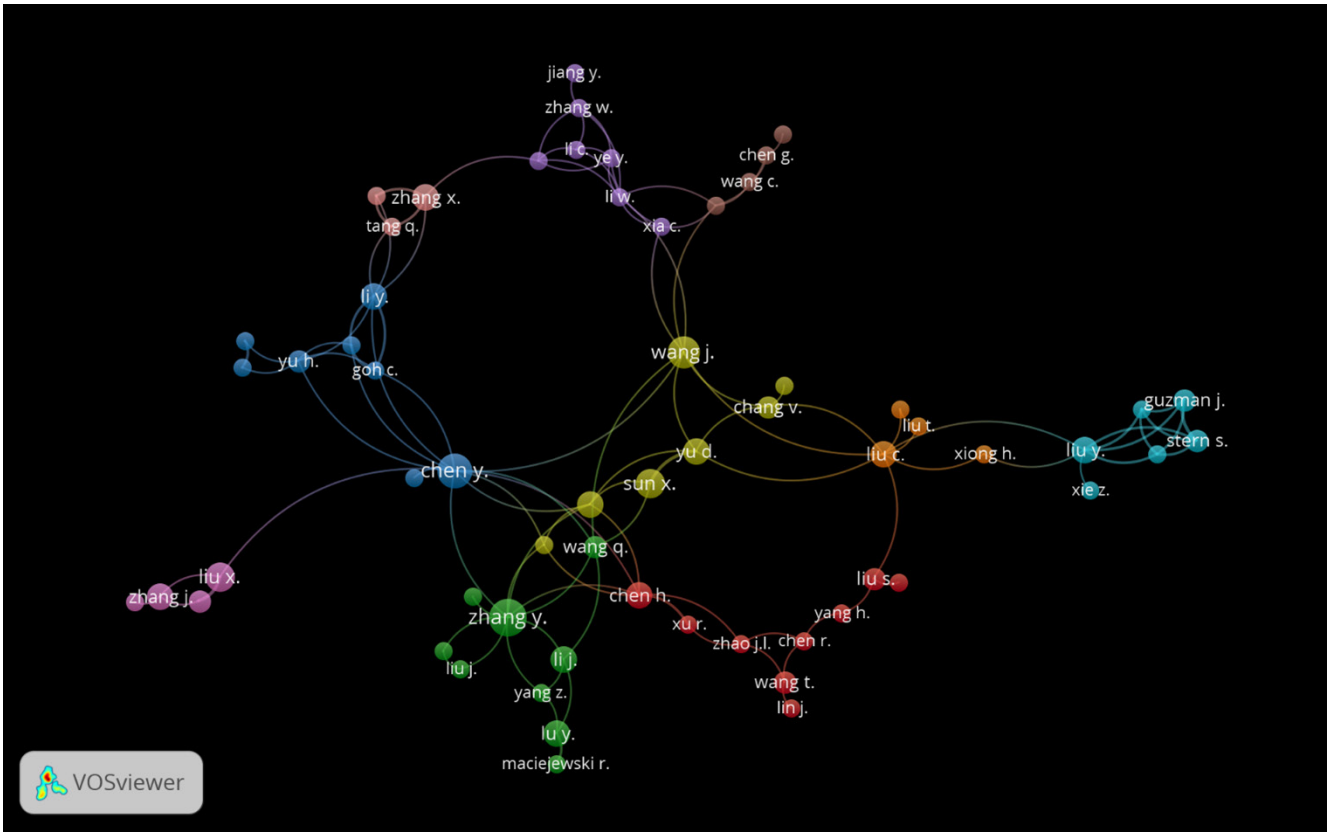

Figure 12: Author co-authorship analysis by VOSviewer.

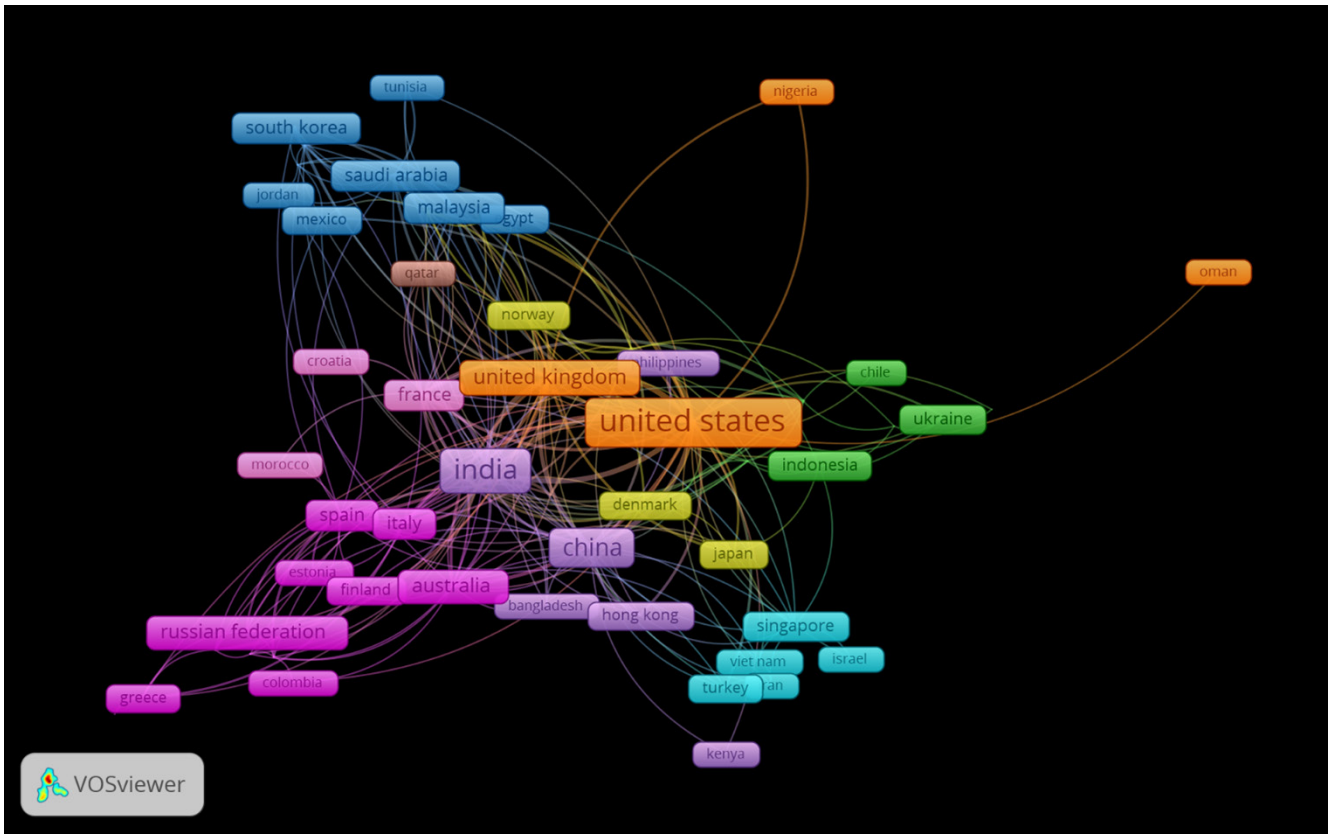

Figure 13: VOSviewer network visualization map of country co-authorship.

Figure 14: Map of countries'/regions' distribution of published papers from 2005 to 2023 created using CiteSpace.

Table 3: Top 10 Productive Regions/Countries.

Sl. No.	Frequency	Centrality	Country
	263	0.44	United States
	206	0.1	India
	114	0.15	China
	94	0.06	Germany
	71	0.15	United Kingdom
	44	0.1	Russian Federation
	41	0.07	Australia
	34	0.03	Italy
	32	0.04	Canada
	29	0.19	Spain

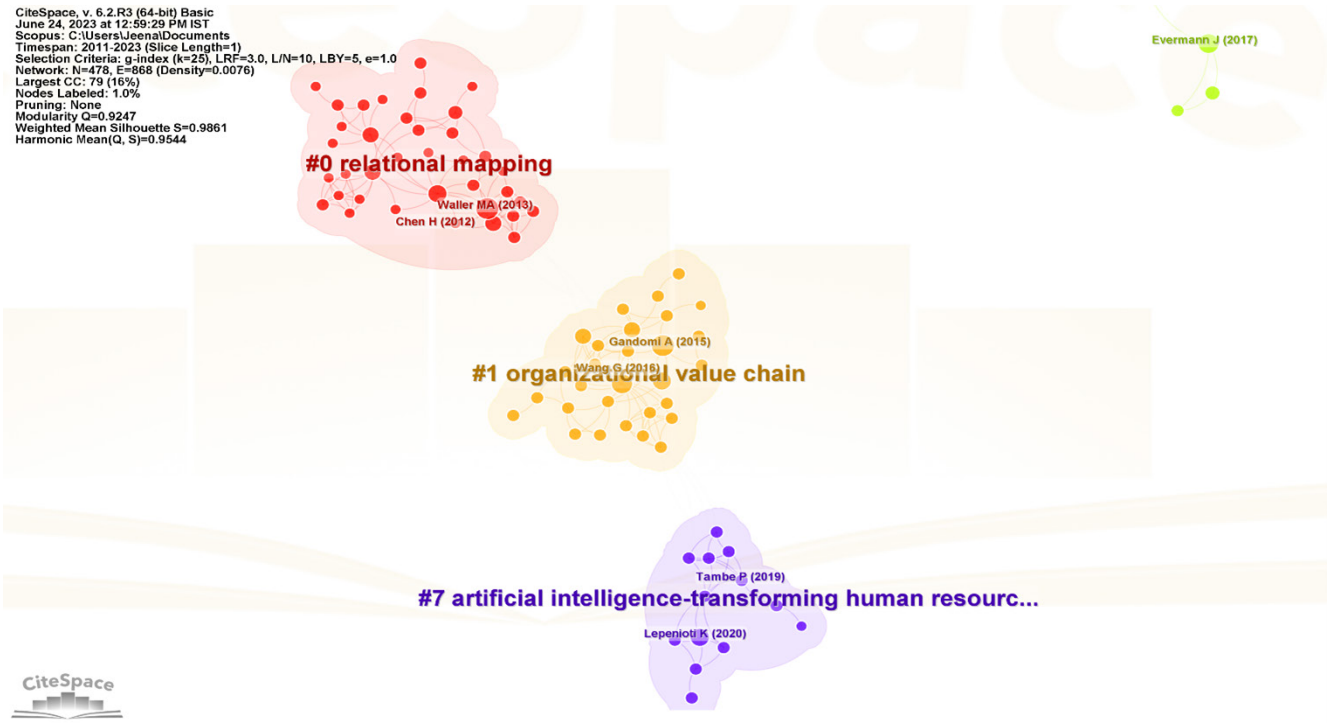

Figure 15: Visual representation of the co-cited reference from 2011 to 2023 created using CiteSpace.

Reference co-citation analysis

Reference co-citation refers to a situation where subsequent papers cite multiple papers simultaneously, creating a connection between the original papers. The strength of this connection depends on the number of authors who cite these papers. Figure 15 depicts the cited references, with lines connecting the nodes to represent co-citation relationships. The size of each node indicates the frequency of co-citation, with larger nodes representing more frequent co-citations. To gain a better understanding of the evolution of predictive analytics in business research over time, a chronological representation of the clusters has been generated. Figure 16 presents a timeline view that provides additional insights by highlighting the most frequently cited articles and the emergence of new issues within specific timeframes. Table 4 presents the authors of the most frequently referenced papers, comprising the top four publications. Out of the authors mentioned, Gandomi A, Waller MA, Wang G, and Evermann J exhibit the most frequent occurrences.

Journal co-citation analysis

Two journals are considered to have a co-citation when their titles are mentioned in subsequent works. This type of analysis, known as journal co-citation analysis, uncovers the connections and relationships between journals. In Figure 17, nodes represent the journals, and the links between the nodes indicate the degree of co-citation. The proximity of the nodes reflects the frequency

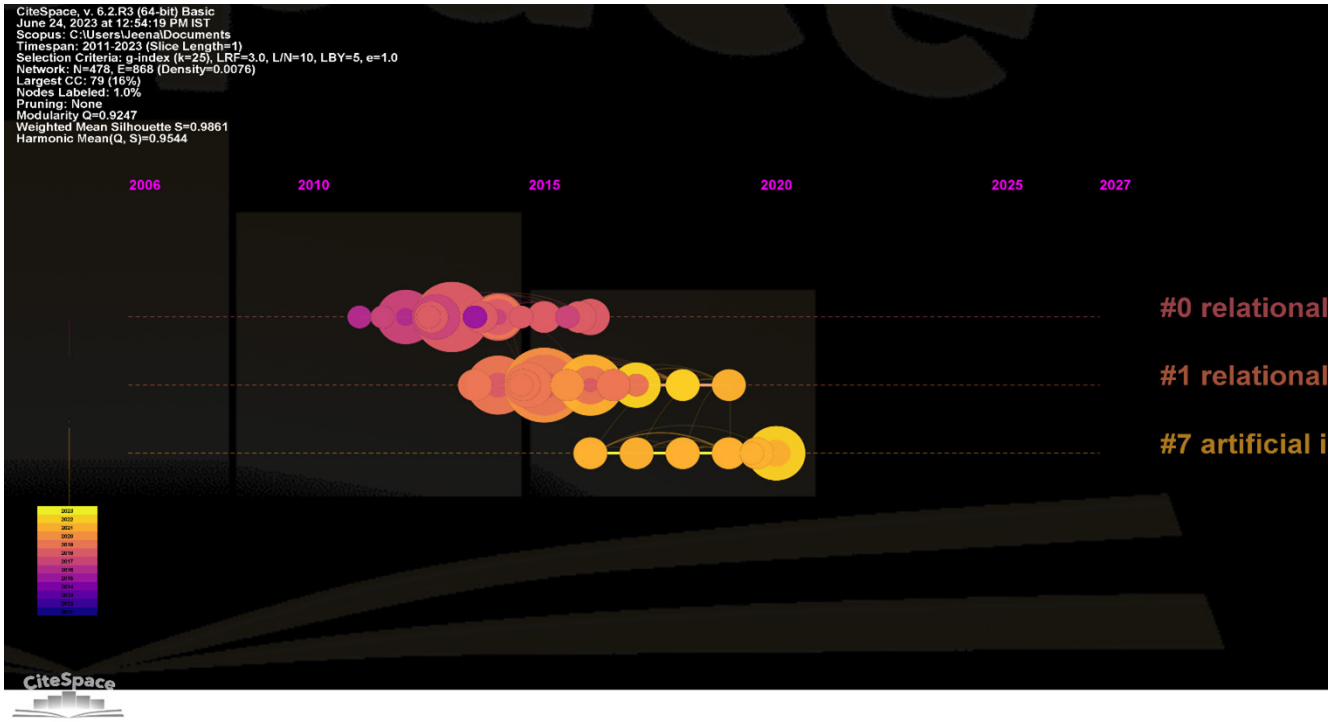

Figure 16: A co-cited reference timeline generated by CiteSpace.

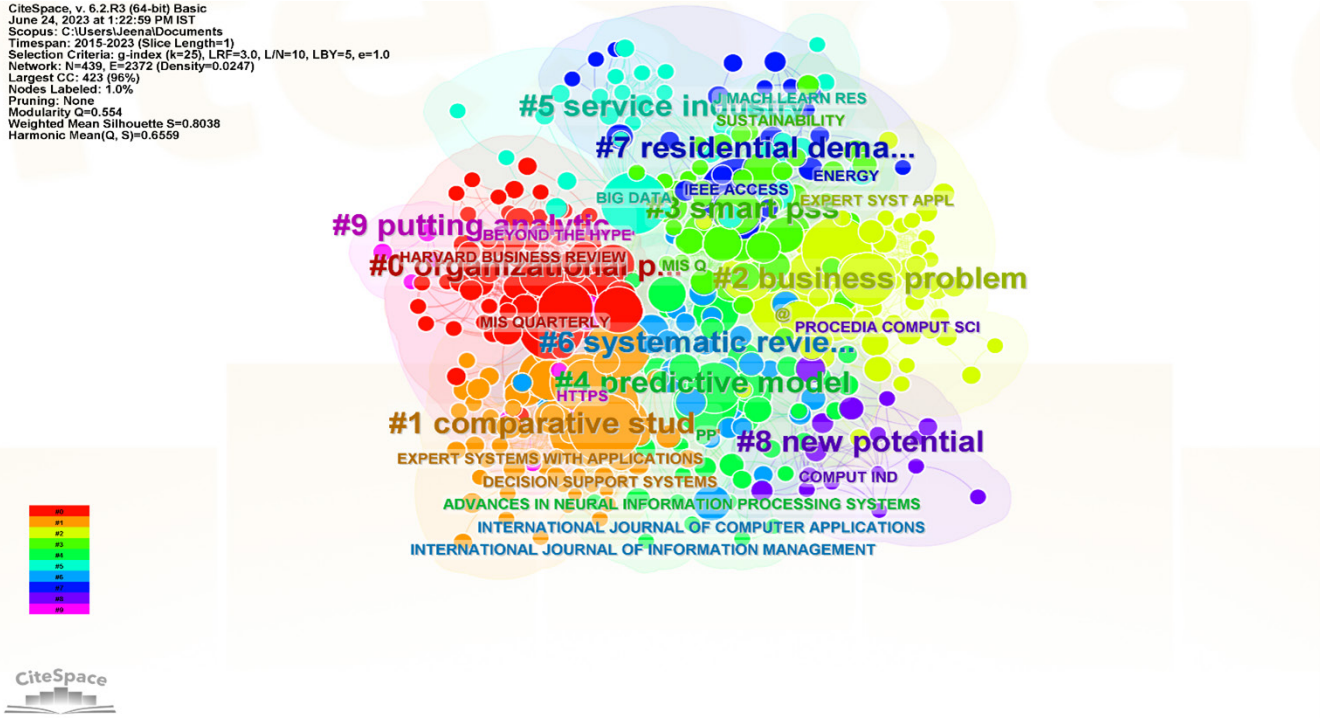

Figure 17: Map of journal co-citation analysis from 2015 to 2023 created using CiteSpace.

of co-citation, while the size of a node represents the number of citations received by a journal. Generally, as the size of a node increases, the co-citation frequency between the journals also increases. Figure 17 specifically highlights journals that have received a high number of citations.

Table 5 offers a comprehensive overview of the most frequently cited journals, providing insights into the importance and presence of core journals in predictive analytics within the field of business research and literature. The table displays the citation counts for these journals. Among them, "IEEE Access" stands out as the most cited journal with 81 citations, closely followed by "Expert Systems with Applications" with 74 citations and "Decision Support Systems" with 69 citations. This indicates that these journals serve as prominent platforms for cutting-edge research in predictive analytics within the business field.

Table 4: Top 4 Most Cited References (Frequency ≥ 8).

Frequency	Author	Year
11	Gandomi A	2015
10	Waller MA	2013
8	Wang G	2016
8	Evermann J	2017

Table 5: Top Journals with Higher Citation Frequency.

Frequency	Label	Year
81	IEEE Access	2019
74	Expert Systems with Applications	2015
69	Decision Support Systems	2015
64	Expert Systems with Applications	2017
60	Big Data	2015
54	Decision Support Systems	2016

Timeline view of author keywords

The timeline view of keywords illustrates the progression of hotspots in predictive analytics within the field of business research between 2020 and 2023. Figure 18 displays various keywords that are linked to specific subareas. This particular section presents the timeline view (refer to Figure 18) of these keywords, aiming to assist the reader in comprehending the evolving patterns of development. By adopting this perspective, all the terms are categorized into nine distinct groups or clusters.

DISCUSSION

A total of 1,210 articles were collected from 744 different sources spanning the period from 1989 to 2023. The domain being studied experienced a consistent growth in publication volume from 2006 to 2015. However, after 2015, there was a decline in the number of published works. This downward trend was reversed from 2016 to 2020, during which there was a significant increase in the quantity of publications. However, following this resurgence, there was a subsequent substantial decrease in the number of published works. Overall, the data suggests a fluctuating pattern in the publication volume within the specific domain, with periods of growth, decline, and resurgence.

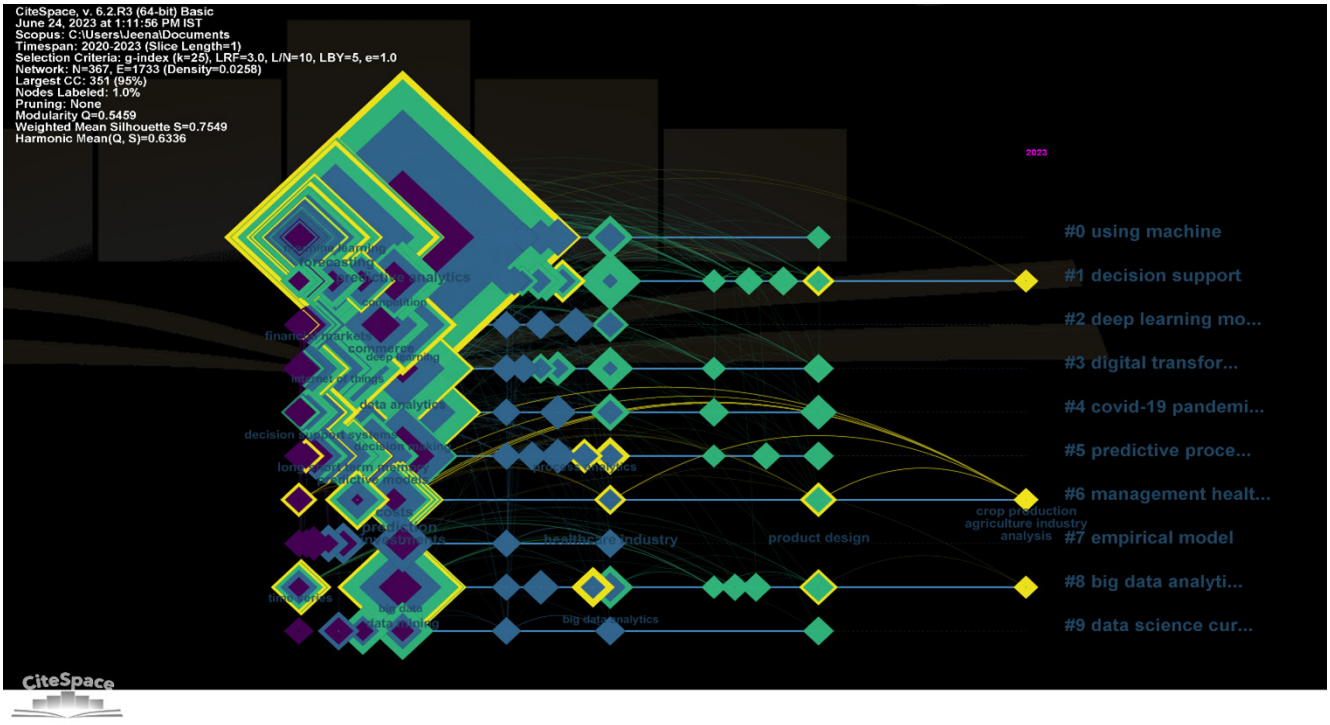

Figure 18: Timeline view of author keywords from 2020 to 2023 created using CiteSpace.

Technological advancements, changes in industry demands, or evolving methodologies influenced such patterns.

Based on their publication counts, consistent productivity, and established expertise, authors such as Loos P, Krumeich J, Zhang Y, Cheng Y, and Werth D have emerged as influential figures in the field of predictive analytics in business. The identification of influential authors provides insights into the individuals shaping the discourse on predictive analytics in business. These authors, with notable publication counts, contribute significantly to the field, showcasing sustained commitment and expertise. Their work not only reflects the depth of their contributions but also establishes them as thought leaders in the domain.

In the field of predictive analytics in business, the journal 'Lecture Notes in Computer Science' (including its subseries in Artificial Intelligence and Bioinformatics) is the most productive journal, publishing the highest number of articles, with a total of 40 publications. It is closely followed by the ACM International Conference Proceeding Series, which published 37 research papers. The field of predictive analytics in the business domain has seen significant contributions from various institutions. Among them, Queensland University of Technology has made the most significant impact, leading the list with a maximum of 23 published papers. The University of Minho closely follows with 20 publications, while Hangzhou Dianzi University has contributed 14 publications. Additionally, several other notable institutions, including Ghent University, Korea Institute of Science and Technology Information, University of Ljubljana, and University of Rome La Sapienza, have also conducted substantial research in the field of predictive analytics in the business domain. Overall, these findings suggest that these institutions have actively contributed to the advancement of predictive analytics research in the business domain, showcasing their expertise and dedication to this area of study. Exploring the most relevant sources and affiliations highlights the diversity of journals and institutions contributing to predictive analytics research in business.

The three-field plot linking keywords, authors, and sources provides a holistic view of the thematic landscape. Commonly used phrases like "predictive analytics," "big data," and "business analytics" emerge as central themes, reflecting prevailing concepts in the literature. This interconnected approach enhances our understanding of the relationships between different elements within the scholarly discourse. There is a substantial body of literature in the realm of predictive analytics within business that focuses on keywords such as "predictive analytics," "big data," "business analytics," "data science," and "deep learning." This literature includes works by authors like Krumeich J, Loos P, and Werth D, who have contributed significantly to this field. The sources where these works are published primarily include Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) and Lecture Notes in Business Information Processing.

The keyword co-occurrence analysis, with the visualization of clusters and the use of node size, edge thickness, edge length, and colour coding, provides insights into the relationships and emphasis on different topics within the analyzed textual data. It allows for a deeper understanding of the connections between various concepts and keywords, highlighting distinct research focus areas. The most commonly utilized phrase is "predictive analytics" with a frequency of 954, indicating its significant presence in the analyzed text. Following that, "forecasting" appears 347 times, and "decision making" appears 185 times. The strategic diagram helps categorize research areas based on their density and centrality, providing insights into the level of development and importance of different themes.

Co-citation analysis identifies shared citations between pairs or groups of journals rather than individual articles. Based on the co-citation analysis, "IEEE Access," "Expert Systems with Applications," and "Decision Support Systems" emerge as the top journals in the field of predictive analytics within the business research and literature, indicating their importance and presence as core journals in this area. Academics and researchers can regard these journals as essential for accessing recent developments and information on business predictive analytics.

Our investigation yielded some noteworthy results, one of which was that global collaboration is at its maximum level in the United States. This indicates that the US actively participates in international collaborative studies and significantly influences predictive analytics in business publications. In terms of research output, the United States retains a prominent position. It takes the lead, with India and China having significant contributions to predictive analytics for business studies. The analysis highlights the cross-national and cross-regional academic relationships and offers insightful information on the distribution and effect of this field of study. This global collaboration contributes to a diverse and comprehensive understanding of the subject.

The authors Gandomi A, Waller MA, Wang G, and Evermann J have received numerous citations in business research using predictive analytics. These works have attracted much interest and are regarded as influential within the scientific world. The number of publications cited after these authors' original work indicates how strong the relationship is.

The comprehensive discussion of the bibliometric analysis results illuminates the multifaceted aspects of predictive analytics in business research. The temporal trends, influential authors, collaborative networks, thematic content, and citation patterns collectively contribute to a holistic understanding of the current state and future directions of research in this dynamic and evolving field.

The addition of content analysis bridges the observed quantitative data with the thematic core of research contributions within predictive analytics in business. By examining the themes,

theoretical frameworks, and methodological approaches of the most influential papers, the evolution of key concepts and the intellectual progression of the field are revealed. This methodology not only enriches the understanding of trends observed in publication patterns, authorship, and journal contributions but also illuminates the qualitative dimensions of research advancements. Through the combination of quantitative analysis and qualitative examination, the discussion on the significance of findings is deepened, articulating how shifts in thematic focus, methodological innovations, and theoretical discussions reflect broader trends in the application of predictive analytics within business. This synthesis provides a more nuanced insight into the field's dynamics, highlighting the integration of bibliometric mapping with scholarly discourse, thereby addressing the initial critique of the study's depth.

Research Gaps and Implications of the Findings

The study has uncovered notable research gaps and practical implications. Firstly, a need exists for a deeper understanding of how organizations integrate predictive analytics into decision-making processes, encompassing both technical and organizational dimensions. Cross-industry comparative studies are limited, hindering insights into sector-specific challenges and opportunities. Ethical and legal considerations, such as privacy and bias in algorithmic decision-making, present unexplored areas. Long-term impact assessments of predictive analytics and the factors influencing user adoption and resistance within organizations are additional research gaps. On a practical level, customization of predictive analytics for industry-specific needs, investment in explainable AI to address model transparency concerns, and strategic talent acquisition at the intersection of data science and business emerge as key implications. Moreover, fostering collaboration between IT and business teams, ensuring continuous monitoring and updating of predictive models, and establishing robust frameworks for regulatory compliance and data governance are essential practical considerations. Synthesizing the interpretation of the results, it is evident that the identified sublines for future research offer promising avenues for further exploration. The practical applications of predictive analytics in decision-making processes, the integration of big data and machine learning techniques, and the evolving landscape of business analytics within different industry sectors present rich areas for in-depth investigations. Additionally, understanding the factors contributing to the observed cyclical patterns in scientific production opens avenues for future studies that delve into the external influences shaping the ebb and flow of research interests.

CONCLUSION

The study represents a significant contribution to the understanding of the evolving landscape of research in predictive analytics within the business domain. The comprehensive analysis of temporal trends, influential authors, significant sources, and prevalent themes, conducted through a meticulous bibliometric journey, provides valuable insights that extend beyond the quantitative metrics.

The observed cyclical nature of research interests, characterized by a steady increase in publications from 2006 to 2015, followed by a subsequent decline and resurgence, prompts reflection on the driving forces behind these patterns. This temporal analysis underscores the dynamism of the field and suggests that external factors such as technological advancements, shifts in business needs, or methodological innovations may play pivotal roles in shaping research trajectories. The identification of influential authors, including thought leaders such as Loos P, Krumeich J, and Zhang Y, highlights the intellectual pillars shaping the discourse. These authors, along with the most cited documents by Gandomi A, Waller MA, Wang G, and Evermann J, signify foundational contributions that have left a lasting impact on the predictive analytics community. A deeper dive into the content of these highly cited works could unveil specific themes, methodologies, and insights that have resonated strongly within the research community. The collaborative networks depicted through co-authorship and country co-authorship analyses emphasize the global nature of predictive analytics research, with the United States playing a prominent role in fostering collaborative efforts. This global perspective underscores the interdisciplinary and interconnected nature of research endeavors in predictive analytics.

In conclusion, this study not only serves as a foundational exploration but also sets the stage for future investigations that can delve deeper into specific themes, ultimately advancing our understanding of the dynamic and evolving frontier of predictive analytics in the business domain. The identified patterns and influential factors provide a robust framework for scholars, practitioners, and decision-makers to navigate and contribute to the ongoing discourse in this critical area of research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Zhao X. Research on market consumption prediction based on machine learning. Ahn Y, Wu F, editors. E3S Web Conf. 2020;214:02014.
2. Calixto N, Ferreira J. Salespeople Performance Evaluation with Predictive Analytics in B2B. Appl Sci. 2020;10(11):4036.
3. Yeboah-Ofori A, Islam S, Lee SW, Shamszaman ZU, Muhammad K, Altaf M, *et al.* Cyber Threat Predictive Analytics for Improving Cyber Supply Chain Security. IEEE Access. 2021;9:94318-37.
4. Lee CS, Cheang PYS, Moslehpour M. Predictive analytics in business analytics: decision tree. Adv Decis Sci. 2022;26(1):1-29.
5. Anagnostopoulos T, Kyriakopoulos GL, Ntanos S, Gkika E, Asonitou S. Intelligent Predictive Analytics for Sustainable Business Investment in Renewable Energy Sources. Sustainability. 2020;12(7):2817.
6. Zulaikha S, Mohamed H, Kurniawati M, Rusgianto S, Rusmita SA. Customer predictive analytics using artificial intelligence. Singap Econ Rev. 2020;1-12.
7. Ahsan M, Ahmad K, Ahamed J, Omar M, Ahmad KAB. PAPQ: Predictive analytics of product quality in industry 4.0. Sustain Oper Comput. 2023;4:53-61.

8. Stefanovic N. Proactive Supply Chain Performance Management with Predictive Analytics. Sci World J. 2014; 2014:1-17.
9. Chornous GO, Gura VL. Integration of Information Systems for Predictive Workforce Analytics: Models, Synergy, Security of Entrepreneurship. Eur J Sustain Dev. 2020;9(1):83.
10. Wang T. An Intelligent Passenger Flow Prediction Method for Pricing Strategy and Hotel Operations. Hassanien AEIB, editor. Complexity. 2021; 2021:1-11.
11. Sprongl P. Gaining competitive advantage through business analytics. Acta Univ Agric Silvic Mendel Brun. 2013;61(7):2779-85.
12. Chen C, Dubin R, Kim MC. Emerging trends and new developments in regenerative medicine: a scientometric update (2000–2014). Expert Opin Biol Ther. 2014;14(9):1295-317.
13. Mukherjee B. Analysis of Global Research Trends in Coronaviruses: A Bibliometric Investigation. J Scientometr Res. 2020;9(2):185-94.
14. Singh S, Solkhe A, Gautam P. What do we know about Employee Productivity?: Insights from Bibliometric Analysis. J Scientometr Res. 2022;11(2):183-98.
15. Sharma P, Rohatgi S, Jasuja D. Scientific Mapping of Gravity Model of International Trade Literature: A Bibliometric Analysis. J Scientometr Res. 2023;11(3):447-57.
16. Yu Y, Li Y, Zhang Z, Gu Z, Zhong H, Zha Q, *et al.* A bibliometric analysis using VOSviewer of publications on COVID-19. Ann Transl Med. 2020;8(13):816.
17. Donthu N, Kumar S, Pandey N, Gupta P. Forty years of the International Journal of Information Management: A bibliometric analysis. Int J Inf Manag. 2021;57:102307.
18. Godin B. On the origins of bibliometrics. Scientometrics. 2006;68(1):109-33.
19. Hood WW, Wilson CS. The Literature of Bibliometrics, Scientometrics, and Informetrics. 2001.
20. Ellegaard O, Wallin JA. The bibliometric analysis of scholarly production: How great is the impact? Scientometrics. 2015;105(3):1809-31.
21. Khakimova AKh, Zolotarev OV, Berberova MA. Coronavirus infection study: bibliometric analysis of publications on COVID-19 using PubMed and Dimensions databases. Sci Vis [Internet]. 2020;12(5). Available from: http://sv-journal.org/2020-5/10/
22. Chen C. Science Mapping: A Systematic Review of the Literature. J Data Inf Sci. 2017;2(2):1-40.
23. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr. 2017;11(4):959-75.
24. Le TV, Pham HH, Tran VB. A Bibliometric Analysis of Studies on 'Start-up Success' Covering the Period 1981-2019. J Scientometr Res. 2022;11(2):212-25.
25. McAllister JT, Lennertz L, Atencio Mojica Z. Mapping A Discipline: A Guide to Using VOSviewer for Bibliometric and Visual Analysis. Sci Technol Libr. 2022;41(3):319-48.
26. Utami N, Setiawan A, Hamidah I. A bibliometric analysis of augmented reality in higher education. 2023;18(3):1599-613.
27. Wan G, Ahmad YD, Nopasit C. A Bibliometric and Visual Analysis in the Field of Environment Social and Governance (ESG) Between 2004 and 2021. Int J Inf Sci Manag IJISM [Internet]. 2023;21(2). Available from: https://doi.org/10.22034/ijism.2023.1977765.0
28. Agbo FJ, Oyelere SS, Suhonen J, Tukiainen M. Scientific production and thematic breakthroughs in smart learning environments: a bibliometric analysis. Smart Learn Environ. 2021;8(1):1.
29. Prakash C, Dwivedi S. Trends in Political Campaigning Research: A Bibliometric Literature Analysis. J Scientometr Res. 2022;11(2):262-71.
30. Mishra M, Dash MK, Sudarsan D, Santos CAG, Mishra SK, Kar D, *et al.* Assessment of trend and current pattern of open educational resources: A bibliometric analysis. J Acad Librariansh. 2022;48(3):102520.
31. Bruni A, Serra FG, Gallo V, Deregibus A, Castroflorio T. The 50 most-cited articles on clear aligner treatment: A bibliometric and visualized analysis. Am J Orthod Dentofacial Orthop. 2021;159(4):e343-62.

Cite this article: John J, Joseph J, Mathew L, James S, Jose J. Exploring the Predictive Analytics Frontier in Business: A Bibliometric Journey. J Scientometric Res. 2024;13(2):365-81.