

The Last Decade of Retail Analytics: A Bibliometric Analysis

**Dr Richa Kalpesh Saxena¹, Vani Agarwal², Vansh Raja³, Veer Agarwal⁴,
Veer Sheth⁵, Yashas Dassani⁶**

¹Assistant Professor, Marketing & Operations, NMIMS, Anil Surendra Modi School of Commerce,
Mumbai Campus, India

^{2,3,4,5,6}BBA, NMIMS, Anil Surendra Modi School of Commerce, Mumbai Campus, India

ABSTRACT:

This bibliometric review maps the development and structure of research on retail analytics, with emphasis on big data, analytics and machine learning in retail and on their role in retail strategy and operational decision making. Publications were retrieved from the Scopus database on 17 September 2026 using the search string "Retail AND Analytics". After screening for time period (2016–2027), document type, language and source type, a final sample of 762 journal articles was analysed using Biblioshiny. The findings show that annual output grew from 19 articles in 2016 to 114 in 2025, a compound annual growth rate of about 22%, and a growth model suggests that the field has reached only about 10% of its estimated saturation level. The literature is dispersed across 435 sources, led by the Journal of Retailing and Consumer Services, with the USA and India as the largest contributing countries. The keyword and thematic analyses identify sales, machine learning, data mining, big data and predictive analytics as prominent themes and show a shift from data gathering towards artificial intelligence, data analytics and consumer behaviour. Research gaps are identified in data integration, model comparison, interpretability, the link between forecasts and decisions, new-product forecasting and multi-level forecasting. Future research directions include hybrid statistical–machine learning models, SKU-level and probabilistic forecasting, explainable AI, multi-channel data and closer integration of demand forecasting with inventory management. These insights are useful for researchers and retail practitioners seeking to align analytics with retail decision making.

KEYWORDS: Retail; Analytics

1. INTRODUCTION

Retail is an extremely competitive industry. The success of a retailer depends not only on the products it sells but also on its ability to understand what customers want and how much they are likely to buy. Estimating the demand for products is an important part of this process as it allows retailers to decide how much stock to maintain, when to replenish inventory and how to allocate resources. If the demand is underestimated then there may be stockouts and lost sales, whereas if it is overestimated then there may be excess inventory and unnecessary costs. These problems become even more complicated when customer preferences change, promotions influence purchasing behaviour and different products have different patterns of demand. As a result, retail demand forecasting has become an important area of

research as it helps retailers make better decisions using the information available to them (Fildes et al., 2022).

As the retail industry has developed, the amount of information available to retailers has increased significantly. This includes transactional data, promotional information, customer behaviour and information collected through digital platforms. For example, Ma, Fildes and Huang (2016) examine how promotional information within and across product categories can be used for forecasting retail sales. This is important because promotions do not only affect the demand for one product but may also influence the purchasing behaviour associated with other products. Similarly, Iftikhar and Khan (2020) examine the use of social media information alongside historical sales data for demand forecasting. This allows the understanding of demand to extend beyond what customers have purchased in the past and also considers what they are discussing and expressing interest in. As a result, retail analytics has developed into an area which combines different types of information with different analytical techniques to understand consumer behaviour and predict demand.

Despite the growing amount of research in this area, the literature covers several different approaches, data sources and applications. Some studies focus on traditional forecasting methods, while others examine big data analytics, machine learning, deep learning, promotional information and external digital signals. Furthermore, the studies may use different datasets, forecasting periods and evaluation measures, making it difficult to compare their findings directly. The literature also indicates that forecast accuracy is not the only factor which needs to be considered, as the usefulness of a forecast depends on whether it can support decisions relating to inventory, replenishment and other retail operations. As a result, it becomes important to understand how research in this field has developed over time, which themes have become prominent and how different areas of research are connected.

The purpose of this study is to examine the development and structure of research on big data, analytics and machine learning in retail through bibliometric analysis. More specifically, the study aims to examine the annual growth of publications, identify the most relevant publication sources, understand the geographical distribution of research and investigate the relationships between authors, countries and keywords. Additionally, the study examines the evolution of research themes in order to understand how the focus of the literature has changed over time. Bibliometric analysis is useful in this context because it allows a large collection of publications to be analysed quantitatively and helps identify patterns which may not be apparent when individual studies are considered separately (Aria and Cuccurullo, 2017).

The study therefore seeks to provide an overview of how research on retail analytics has developed and how its different themes are connected. By examining publication patterns alongside the existing literature on forecasting, big data and machine learning, it becomes possible to understand both the development of the research field and the challenges which continue to remain.

The bibliometric review is organized as the following. The next section presents the literature review, followed by the research methodology. The review analysis and future agendas are discussed.

2. LITERATURE REVIEW

At the initial stage, retail forecasting was a specialized problem. People attempted to make assumptions about sales in the next period based on previous demand. Today, this task has become broader and is associated with the large amounts of data. As explained by Fildes, Ma and Kolassa (2022), there are different applications of retail forecasting including strategic planning, overall forecasting, product level forecasting, and store level forecasting. In addition, they point out that there are aggregation problems,

which are discussed when it comes to the analysis of results obtained in the process of making predictions on the level of product, store, and an entire organization. In addition to it, there are two extra features discussed in their review, which are significant for the field. Demand can change due to promotion activity, and some products do not have enough historical data. These features are important for the field because the accuracy of predictions remains the core issue, while the real challenge lies in the existence of good quality demand data and its structuring. Papers in this area form a chain of research directions. Initially, traditional literature considered demand patterns, data structures, and explanatory variables. Later, big data and analytics provided additional. From then on, techniques drawn from machine learning and deep learning extend the toolkit, including predictive models based on multi-level hierarchical forecasting. In more recent research, predictions are more often tied to such aspects as inventory and replenishment, among others related to operations on a daily basis. Further developments will be discussed in the following sections.

There is a large amount of literature on the topic, but there are no links between them. Various researchers employ various techniques, data, and applications. Having reviewed the papers, some gaps can be found, which include:

Table 1: Research gaps emerging from the literature

Research gap	Evidence from the literature	Implication for future literature
Data integration gap	Sales deals, promo messages, customer details, and outside digital activity are usually looked at in different channels.	More focus is needed on linked forecast systems. These systems should pull in several demand signals that matter.
Model comparison gap	Work on machine and deep learning looks promising. But the older forecasting research is more convincing for several classic methods.	Make the comparison on the same data, with the same benchmarks and the same demand setup. Do not just assume that newer methods will always be better.
Interpretability gap	Some recent projects are starting to use SHAP and LIME. Still, a lot of advanced forecasting setups are hard to understand. For many of these tools, it is not clear why a prediction comes out the way it does.	Forecasting studies should look at how accurate their results are. They should also show whether managers can understand what the results mean. Accuracy alone is not enough for real decisions.
Forecast-to-decision gap	Forecast research and inventory plans often link together. Even so, a lot of forecasting work still checks results using prediction error alone.	Forecasting studies and stock plans are tied closely. But many teams still judge the work mainly by looking at prediction error.

Research gap	Evidence from the literature	Implication for future literature
New-product and changing-demand gap	New products often launch with not much history to draw on. Demand can also shift quickly, which makes planning harder.	Study ways to predict outcomes when we have little past data. Also look at cases where demand changes fast. Include situations with sudden disruptions.
Multi-level forecasting gap	Hierarchical forecasting helps when you need forecasts that match across time and across groups. Still, the research behind it is not settled yet.	Forecasts still need more work across each level. This includes SKU, category, and store. It also covers online and the total view.

Retail Analytics research has become more comprehensive. In the past, it was focused mainly on historical sales data. Nowadays, it relies more on data-driven methodologies and decision support systems. Previous research is important because it helps create the foundation through tests and comparisons. In addition, it includes issues related to combining data and researching demand factors. The later literature has introduced new variables to forecasting models. For instance, it uses big data, scheduling of promotions, social media cues, and customer data. The literature also analyzes digital channels in the context of demand analysis. Moreover, the application of new approaches to forecasting research has expanded the toolkit. Machine learning and deep learning techniques have been applied in the research of retail demand, especially in cases when there is a lot of data and multi-level forecasting is required. At the same time, it does not believe that increased complexity is equivalent to better results. There are a number of questions that have not yet been solved. For example, how to integrate various data flows, how to compare models, and interpret the results of their operation. What occurs when there is limited historical data for the product? What ensures consistency among forecasting? How does forecasting align with the decisions made on a daily basis at operations? Collectively, all these studies reveal a much more cohesive approach. Data, modeling, and Operations Research decision making must be viewed as a single integrated system rather than separate processes.

3. RESEARCH OBJECTIVE

The main purpose of this review is to conduct an analysis of development and expansion of the research area of retail analytics by using bibliometric techniques. This review will help to discover publication patterns, main sources and countries that contribute to the development of the area, networking among researchers in this area, and changing research themes. The purposes of this research are as follows:

- To study the evolution of research publications in the area of retail analytics over time and discover publication trends.
- To determine main journals, sources, and countries that develop literature in retail analytics and study relationships between authors, countries and keywords in this area.
- To analyze major themes and research areas in retail analytics and their evolution over time.
- To assess the existing body of literature and find under-researched issues in this literature.
- To offer new research directions based on the results of bibliometric analysis and literature assessment.

4. RESEARCH METHODOLOGY

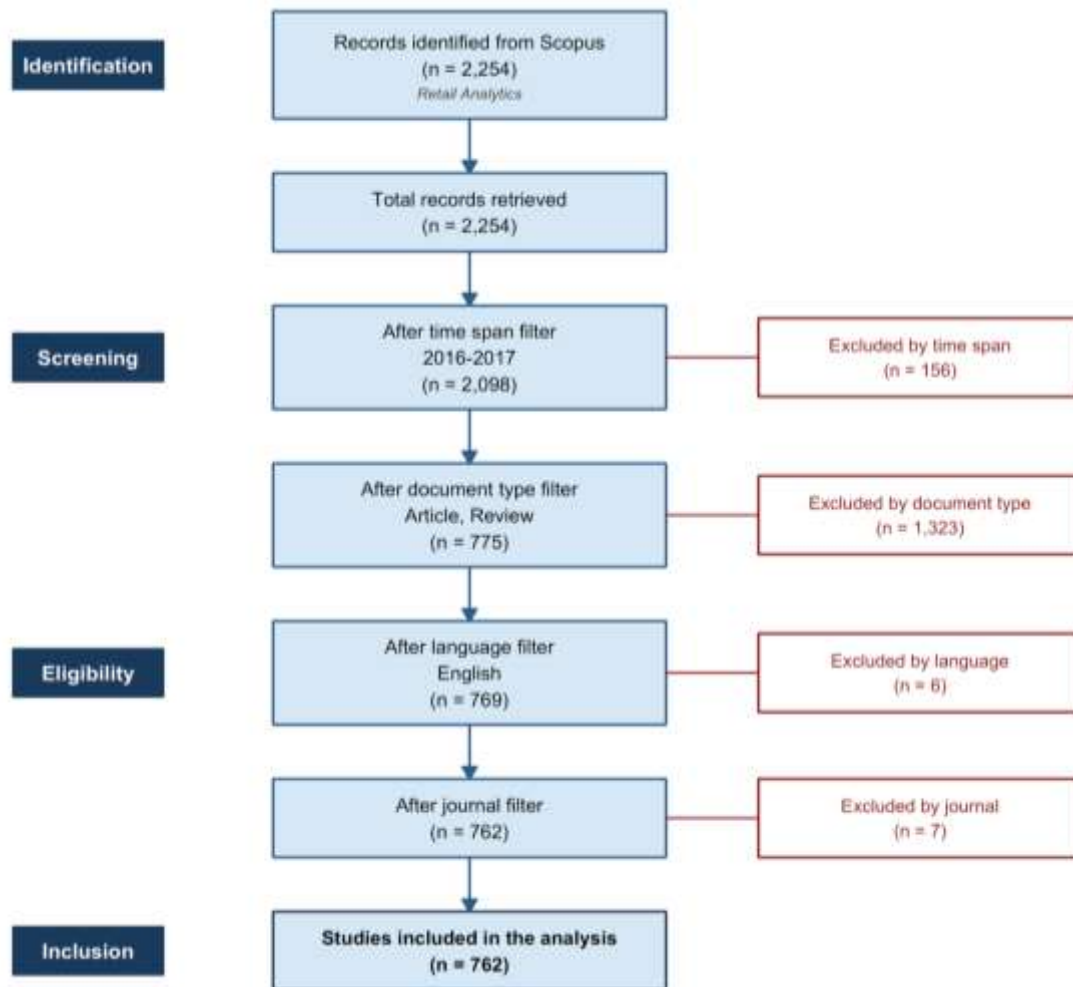
Bibliometric analysis is one of the best ways that can be used to explore trends and developments within any research field. The use of bibliometric analysis makes it possible to evaluate the existing research, identifying influential studies, emerging themes, and research gaps for future investigation (Paul & Criado, 2020). With the increased role of data analytics, artificial intelligence, and predictive analysis within the retail industry, the research on Retail Analytics has been on the rise in recent years. This study uses bibliometric analysis to identify the trends of research within the field of Retail Analytics in the last decade.

A systematic search of the literature was performed on 17 September 2026 using the Scopus database, which is one of the most reliable abstract and citation databases of peer-reviewed literature. In order to perform a systematic search, the Scopus database was chosen as the main source because it provides extensive coverage of scholarly literature in multiple disciplines and allows obtaining structured bibliographic information needed for bibliometric analysis (Pranckutė, 2021). For this purpose, the search string that was developed for the literature search on the topic of Retail Analytics was the following: “Retail AND Analytics”. The search resulted in a raw base sample of 2,254 publications.

It is noteworthy that publications obtained during the search were represented in a variety of subject areas including Business, Management, Sciences, Economics, amongst other categories. Given that Retail Analytics is a process of applying analytics in different functional areas and technologies, all obtained subject areas were kept rather than limiting the dataset to some specific discipline (Lorente-Martínez et al., 2022). The dataset was filtered to include publications from 2016 to 2027, retaining early access publications indexed for 2027. The initial dataset was reduced to 2,098 publications from 2,254. The next stage included filtering document type only to articles and reviews keeping consistent types of the academic papers in the bibliometric dataset. Thus, through the process of the filtering, 775 documents have remained in the scope of the analysis. In order to keep the consistency within the dataset, the language restriction has been applied reducing the dataset from 775 to 769 publications. A further criterion was then applied on the type of source, with this study concentrating solely on articles from journals as they fall within the scope of the academic journal literature covered in this study. After applying this final criterion, a total of 762 publications remained in the database and were used as the final sample in the bibliometric analysis.

This final dataset of 762 publications was analyzed using Biblioshiny, which is the graphical user interface of the Bibliometrix package, to identify development, structure, and research trends in Retail Analytics. The relevant publications have been downloaded in RIS (.ris) format and prepared for bibliometric analysis. Biblioshiny has been employed for the analysis of the bibliographic data and has produced major findings related to publication trends, source performance, author productivity, citations, keywords, and themes (Aria & Cuccurullo, 2017). The results thus obtained have been exported and arranged in the Microsoft Excel program, which has been used for calculations, tables, and figures. The entire process of publication search, screening, eligibility, and inclusion was clearly depicted using the PRISMA flow diagram as shown in figure 1 (Moher et al., 2010a).

Figure 1: PRISMA Flow Diagram adapted for bibliometric studies



5. DATA ANALYSIS AND FINDINGS

The analysis is based on 762 articles retrieved from the Scopus database on 17 September 2026 and processed using Biblioshiny. Since 2026 is an incomplete year, the annual trend and trend line are restricted to the complete years 2016 to 2025. The growth model constitutes the only exception, as Biblioshiny includes the 2026 count (187 articles to date).

Annual Growth of Publications

Annual scientific output between 2016 and 2025 is illustrated in Figure 2. The production began small, with a mere number of only 19 papers published in 2016 and a reduction down to 15 papers a year later. The subsequent growth avenue started from 39 papers produced in 2018 and 61 in 2020, and peaked at 78 articles in 2022. In 2023 some downturn was noted in performance, but outputs regained the upward trend in 2024 to reach 80 articles and finished even stronger at 114 papers issued in 2025. The linear line is a representer of the average increase of 9 papers a year. In relative terms, we can talk about the growth of annual production from a low number of only 19 in 2016 up to an impressive number of 114 papers in 2025. In terms of compound annual growth rates (CAGR), this is approximately a good 22% figure, more than double than the 10% benchmark commonly used to indicate a solid growth for a new emerging research area. However, it should be noted that these results are sensitive to both the starting and ending

years and a small number of 2016 is likely to distort the general picture of development of the newest scientific fields.

Figure 2: Annual scientific production, 2016–2025, with a linear trend line.

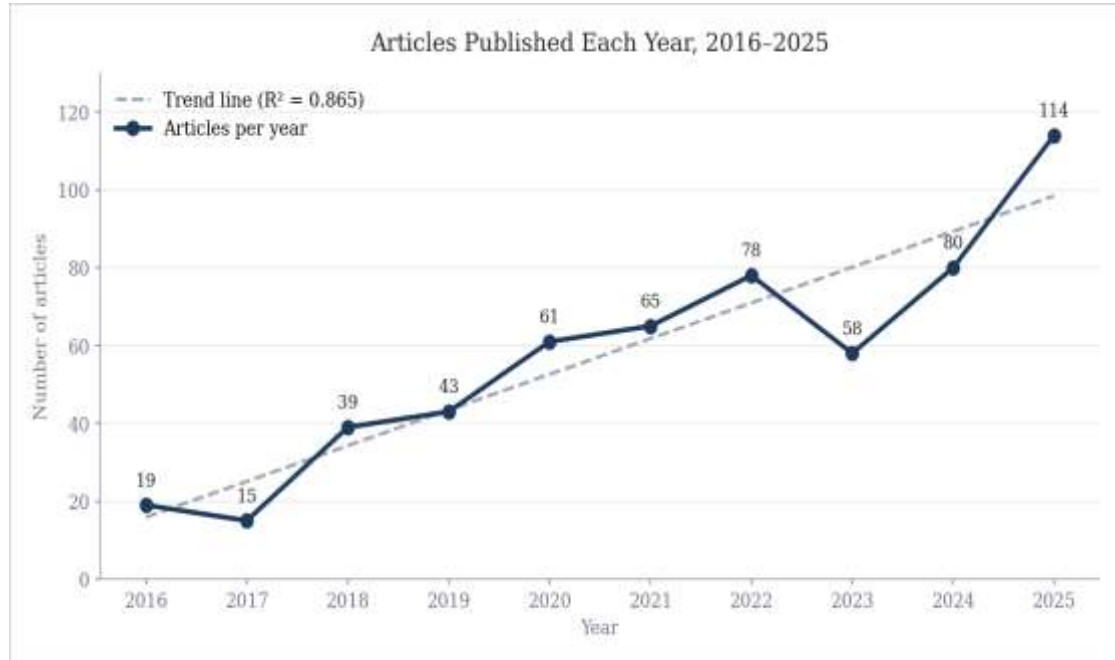


Table 2: Annual number of articles

2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
19	15	39	43	61	65	78	58	80	114

Interpretation of R^2 : R^2 (the coefficient of determination) ranges from 0 to 1 and indicates how closely a fitted line corresponds to the observed data. Values closer to 1 denote a stronger fit. For interpretation, this study treats values above 0.8 as good, values between 0.5 and 0.8 as moderate, and values below 0.5 as weak.

Is the R^2 value good? Yes. The linear trend in Figure 2 yields $R^2 = 0.865$, which indicates a good fit to the observed data. The fit is not perfect, mainly owing to the decline in 2023 and the sharp increase in 2025, which lies above the trend line. Nevertheless, the overall trajectory is clearly upward (Velásquez, 2025). In addition, the growth model provided by Biblioshiny was applied. This model assumes that a new research field grows slowly at first, accelerates rapidly, and then slows as it approaches saturation. Annual output therefore follows a bell-shaped curve (Figure 10, Section 6), whereas cumulative output follows an S-shaped curve. The model is estimated using data from 2016 to 2026.

Figure 3: Summary of the growth model: goodness of fit, saturation level, peak year and milestone years.



Table 3: Main results of the growth model

Item	Value	Interpretation
R^2	0.862	Good fit; the model corresponds closely to the observed data.
RMSE	17.09	On average, the model deviates from observed output by about 17 articles per year.
Articles in 2026 to date	187	Already exceeds the total for 2025 (114).
Total articles to date	759	Cumulative output from 2016 to 2026.
Maximum total (saturation)	7,590	Only about 10% of this level has been reached.
Peak year	2035	Approximately 458 articles in that year.
Fast-growth period	18.2 years	Time required to move from 10% to 90% of the total.

The growth model has an R^2 value of 0.862, confirming its efficiency. The RMSE is equal to 17.09 which can be interpreted as a low value considering that the research output was 114 articles in 2025 and it will increase to 187 articles in 2026, confirming that the model is adequate. The model predicts that the discipline has achieved only about 10% of its saturation level. Moreover, the model predicts that the figure will reach 50% in 2035, 90% in 2044, and finally 99% in 2054. Additionally, it is estimated that the research output will be about 353 articles in 2031 and 455 articles in 2036, thus making the cumulative figures equal to about 1,979 and 4,106 articles respectively.

However, these estimates should be treated carefully. The reason for this is the limited amount of available data for the model to rely on, specifically, only around eleven years and 2026 being incomplete. Therefore, the values for the years after 2026 are predictions and not well-grounded figures from the data. Nonetheless, the results suggest that the research is going through its growth period.

Most Relevant Sources

The sources are the journals and other venues that published the articles in this research and ranked according to the number of articles published, as this figure indicates where research is concentrated. In total, 435 sources published 762 articles and this reflects a scattered, multidisciplinary field, rather than the situation where research is concentrated in a few main journals (Roederkerk et al., 2022; Dekimpe, 2020). The results are summarized in Table 4.

Table 4: Ten most relevant sources

Rank	Source	Articles	Type
1	Journal of Retailing and Consumer Services	28	Retail journal
2	IEEE Access	14	General technology
3	Linguistic and Philosophical Investigations	12	Not related (see below)
4	Manufacturing and Service Operations Management	10	Operations research
5	Applied Marketing Analytics	9	Marketing / retail
6	International Journal of Retail and Distribution Management	9	Retail journal
7	International Journal of Production Economics	8	Operations research
8	Management Science	8	Operations research
9	Review of Contemporary Philosophy	8	Not related (see below)
10	Sustainability (Switzerland)	8	General / sustainability

There are three groups of sources. To begin with, the most notable sources are the retail and marketing journals, particularly the Journal of Retailing and Consumer Services, which has the highest number of published articles (28). The second group consists of operations research and management journals, which also rank very high. The last group includes various technology and sustainability journals, such as IEEE Access and Sustainability, which reflect a strong emphasis on the use of machine learning in the field. Nevertheless, the total number of articles published in these ten journals is only 114 or around 15% of the overall sample, confirming this dispersion trend.

It should be mentioned that such titles as Linguistic and Philosophical Investigations and Review of Contemporary Philosophy do not fall within the retail category. Their presence in the results may be

explained by the use of such generalized terms as "analytics" and "decision making," which can be relevant to philosophy papers as well. Hence, the records must be removed from the analysis afterwards.

Scientific Production by Country

The cumulative article count for the five major countries is depicted in Figure 4. The USA is in first rank while India is in the second position. India has shown a steep rise towards the end of the observation period. As the curves for the UK and Turkey have become very close to each other in the same diagram. Figure 5 illustrates the values for every country separately for the full years of 2016-2025. The R^2 value for each country, presented in each panel, indicates how uniform the increase was: numbers close to 1 show smooth growth while smaller ones show more turbulent motion.

Figure 4: Cumulative number of articles for the five leading countries, 2016–2026.

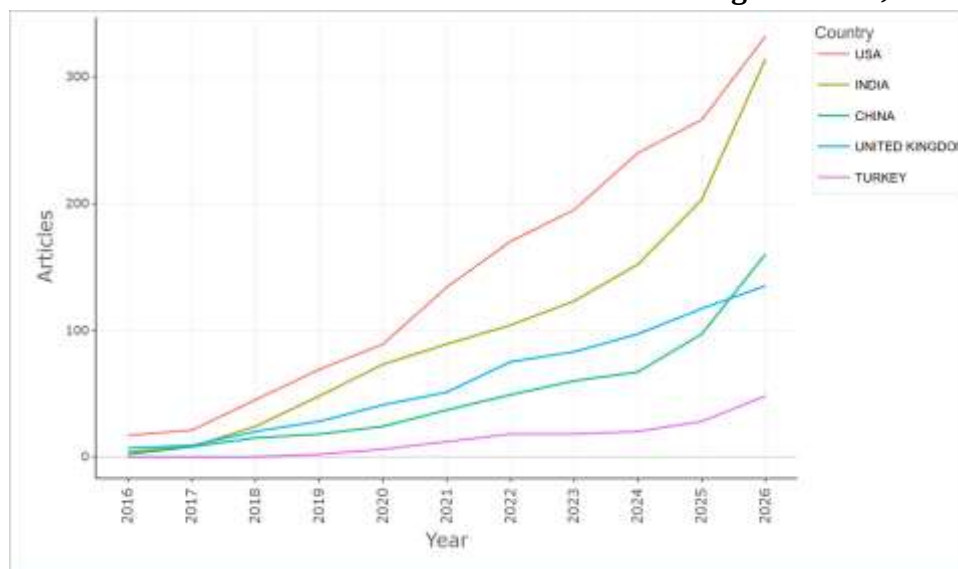
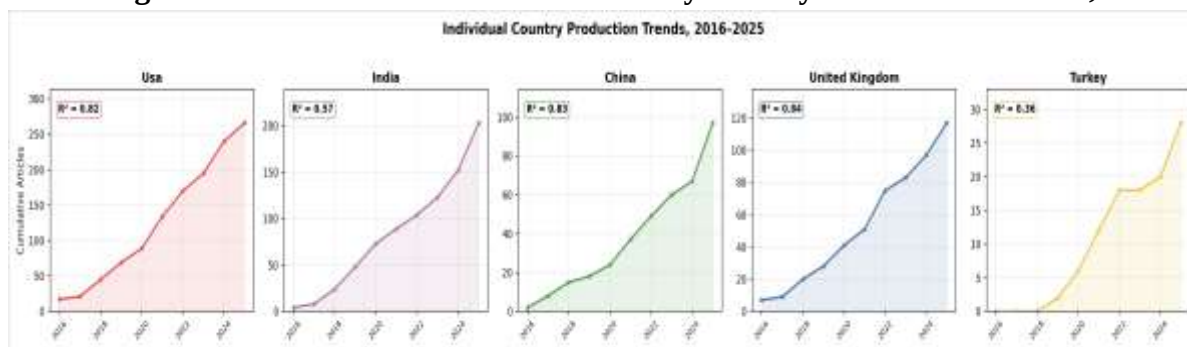


Figure 5: Cumulative number of articles by country on individual scales, 2016–2025.



- USA had the largest contribution, going from 17 outputs in 2016 to a total of 266 in 2025 ($R^2 = 0.82$, good).
- India started with very low numbers (4 articles in 2016) but grew rapidly after 2022, reaching 203 by 2025 and almost matching the USA. Its R^2 of 0.57 (moderate) shows that there were fluctuations in its growth year-by-year.
- China had low levels until 2020 and then rose sharply to 97 articles in 2025 ($R^2 = 0.83$, good).
- United Kingdom also shows steady growth, reaching about 117 articles ($R^2 = 0.84$, good) but without

the spike that happened in India and China.

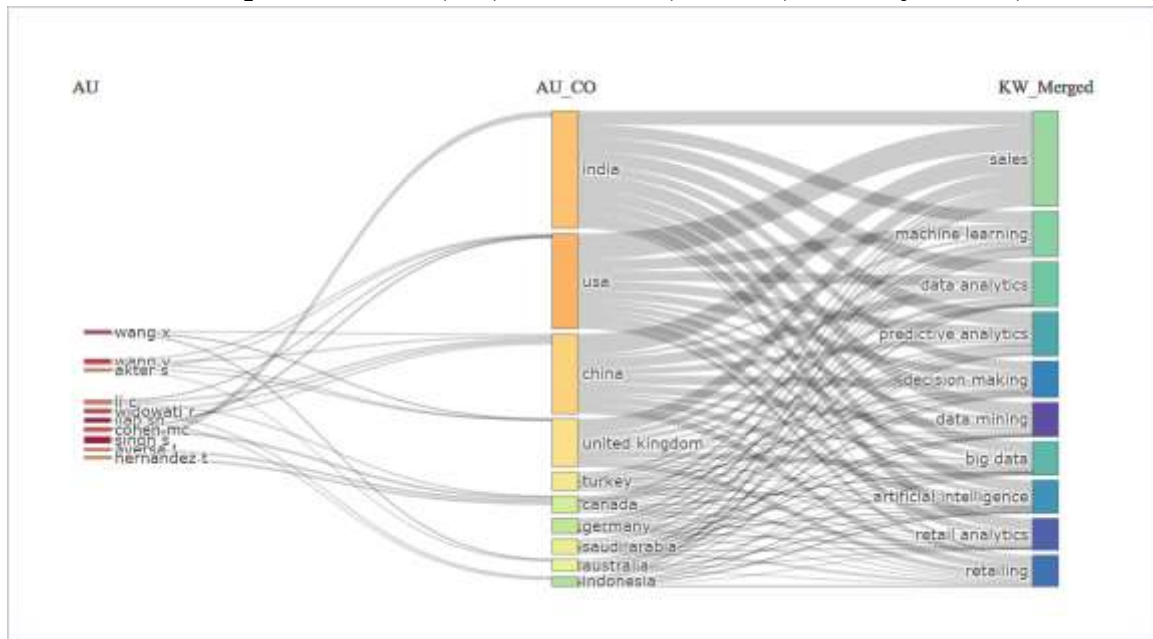
- Turkey is the smallest contributor with nearly no output before 2019, achieving a total of 28 papers in 2025. Its R^2 of 0.36 (weak) indicates that small level contributions are highly volatile, as small changes in input yield corresponding in output changes.

Overall, the USA and India are the leading contributors, whereas China, the UK and Turkey follow from different starting points and at different growth rates (Zhuang, 2021; Youssef et al., 2022).

Three Field Plot: Authors, Countries and Keywords

The three-field plot is a type of Sankey diagram which illustrates the connections between entries from three specific bibliographic fields. In this context, authors (AU) are located on the left, author countries (AU_CO) are in the centre and keywords (KW_Merged) are placed on the right. The size of the ribbons illustrates how frequently two entries are encountered together, so thicker ribbons suggest stronger connections. Since only the most common items are shown in the diagram, less frequently seen references may not be represented. However, the ribbons illustrate the concept of links between the research elements but do not show the cause and effect of the data.

Figure 6: Three-field plot: Authors (AU), Countries (AU_CO) and Keywords (KW_Merged).

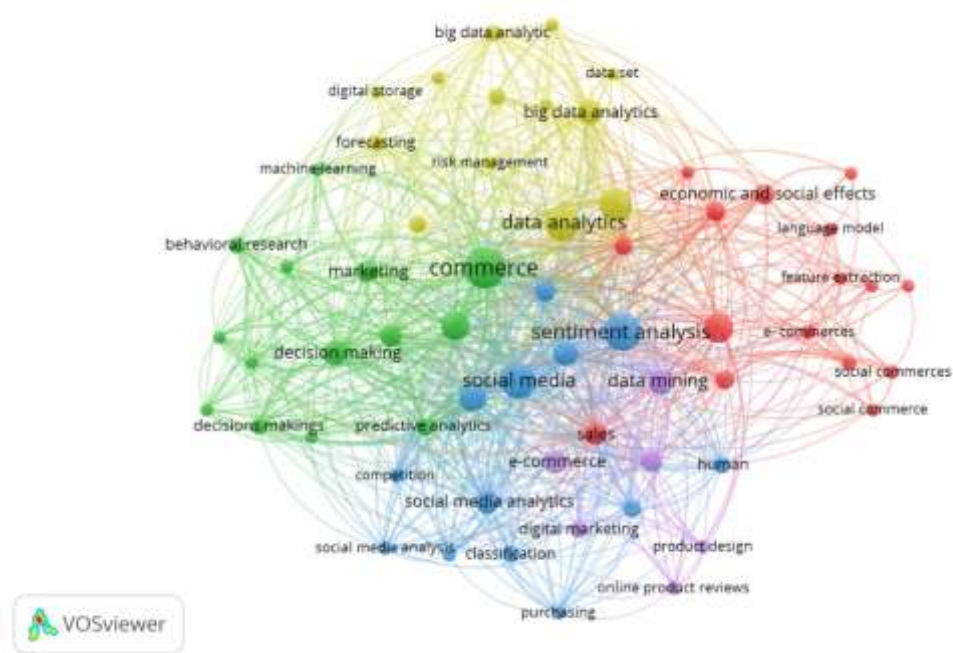


The observations from Figure 6 can be summarized as follows:

- **Countries.** The country of India has contributed the maximum in the second column among countries followed by USA, China and UK, whereas the countries like Turkey, Canada, Germany, Saudi Arabia, Australia and Indonesia make a considerably lower contribution. This finding is consistent with what is reported in Section 5.3.
- **Keywords.** Sales emerged as the highest keyword followed by machine learning, data analysis, predictive analysis, decision making, data mining, big data, artificial intelligence, retail analysis and retail.
- **Authors.** The author bars are smaller and of comparable length (e.g., Wang X, Wang Y, Akter S, Li C, Cohen MC and Singh S), indicating that there is no single author who is leading and thus the research output is shared among many.

The top countries are identified with almost all keywords, which means they explore similar subjects. On the contrary, only a few thin lines are drawn between authors and smaller countries, meaning that international cooperation is limited.

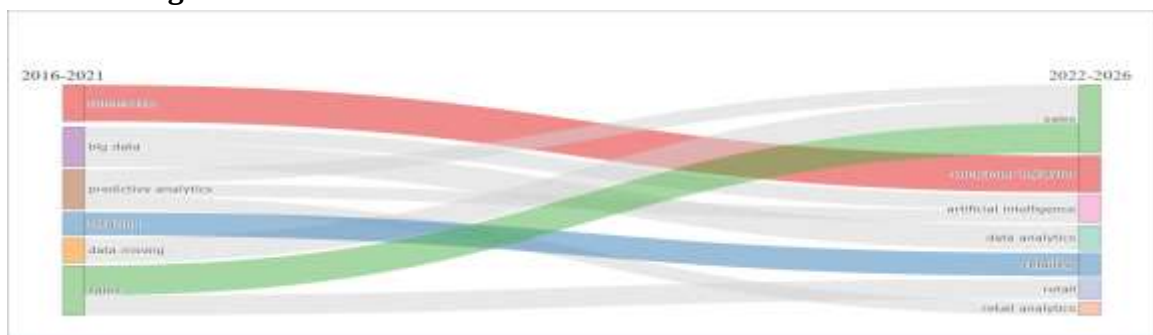
Figure 7: Keyword co-occurrence network (extracted from VOSviewer)



Thematic Evolution and Thematic Map

Figure 8 shows how research themes change across consecutive time slices. The analysis follows the network-based framework of Aria et al. (2025), which builds on Cobo et al. (2011) and is implemented in bibliometrix (Aria & Cuccurullo, 2017). First, themes are identified within each slice from keyword co-occurrence. Each publication is then given a graded membership in every theme, so that interdisciplinary papers can belong to more than one. Finally, themes in consecutive slices are linked by a lineage strength that reflects both the overlap and the structural importance of their shared terms, and only significant links are retained. Keywords were analysed over two slices, 2016–2021 and 2022–2026. The first slice yields six themes and the second seven, as labelled in Figure 8.

Figure 8: Thematic evolution between 2016–2021 and 2022–2026.



Node height reflects the fuzzy cardinality of each theme; band width reflects lineage strength. Node height indicates the weight of a theme, and band width indicates the lineage strength between two themes. Table 5 summarises the main evolutionary patterns.

Table 5: Evolutionary patterns in the thematic evolution analysis

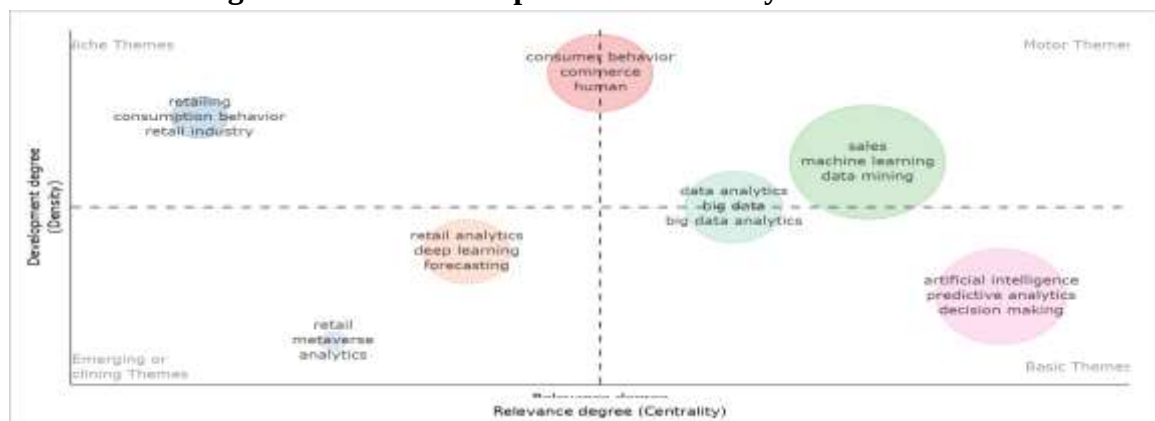
Pattern	Meaning	Observation in Figure 8
Continuation	One incoming and one outgoing edge	“Retailing” and “sales” persist; “commerce” continues as “consumer behavior”.
Splitting	Several outgoing edges	“Big data” and “predictive analytics” feed several later themes.
Merging	Several incoming edges	“Sales” and “artificial intelligence” draw on more than one earlier theme.
Emerging	No incoming edge	No clearly new theme; “retail analytics” has only a thin link.
Disappearing	No outgoing edge	None; “big data” and “data mining” are absorbed into newer themes.

In essence, the fundamental business ideas remain unchanged while the terms have been updated. The most evident word change is that "commerce" is replaced by "consumer behavior," showing a transition from general trading issues to customer-oriented study. The words "sales" and "retail" have kept their meaning in both periods; the themes of "big data" and "predictive analytics" have turned into "artificial intelligence," "data analytics," and "retail analytics." Therefore, the industry has gone from data gathering to using state-of-the-art technologies to study and forecast customer behavior (Tudor et al., 2025).

At the same time, two limitations should be mentioned here. Only two time periods have been analyzed so that the long-term development cannot be evaluated, and 2026 is not yet complete.

The thematic map (Figure 9) presents the main themes of the total studied period. Keywords are organized into themes and arranged in four quadrants in terms of relevance (significance for the area) and development. The size of the bubbles shows the frequency of words (Velásquez-Henao, 2025).

Figure 9: Thematic map of the retail analytics literature.



Motor themes (distinguished and well elaborated): The largest and dominant theme, consisting of "Sales, machine learning, data mining", has the shape of a big bubble. "Data analytics, big data, big data analytics" has almost the same size, thus positioned near the center, being the borderline between motor and basic themes. The third theme "consumer behavior, commerce, human" is in the top of the map despite its position being lower on the vertical axis.

Basic themes (important, but less developed): "Artificial intelligence, predictive analytics, decision making" comes along with the already proven usefulness in the field (Adulyasak et al., 2024).

Niche theme: "Retailing, consumption behavior, retail industry" is good at developing but has a distant location.

Emerging or declining: The themes "Retail analytics, deep learning, forecasting" and "retail, metaverse, analytics" show weaknesses on the map (Punia & Shankar, 2022).

Findings

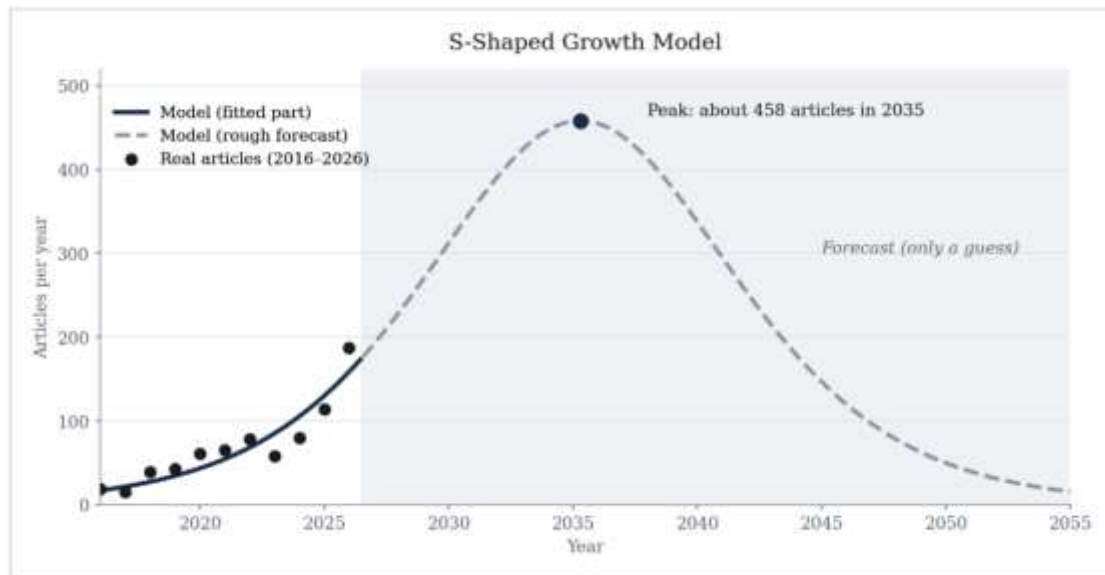
1. Growth in Publications: There has been a huge growth in the field of retail analytics. The number of publications has risen from 19 in 2016 to 114 in 2025.
2. Leading Research Sources: Journal of Retailing and Consumer Services became the top source of literature with 28 papers published. Still, the literature has been distributed over several journals as the subject matter of retail analytics is multidisciplinary in nature.
3. Country-wise Contribution: USA is the top country in terms of contribution of research papers in retail analytics. The next big contributors are India, China, and United Kingdom. The country which showed huge growth is India.
4. Author and Keyword Analysis: The distribution of research papers among authors shows that there is no particular author whose contribution in research is significantly higher than others. Keywords like sales, machine learning, data analysis, predictive analytics, decision making, and artificial intelligence were found.
5. Evolution of Research Themes: It has been observed that research has slowly moved from conventional topics like commerce and sales to consumer behaviour, data analytics, artificial intelligence, and predictive analytics.
6. Prominent Research Themes: Sales, machine learning, and data mining have been identified as significant research themes. Analytics and consumer behaviour are other significant themes, while artificial intelligence and predictive analytics continue to be significant themes for future research.
7. Future Research Themes: Research themes like retail analytics, deep learning, forecasting, and metaverse applications seem underdeveloped or emerging. However, one can conduct research on these themes, but caution must be taken in interpreting the results because the study analyses just two thematic periods, and 2026 is an incomplete period.

6. FUTURE RESEARCH AGENDA

The outcomes of the bibliometric study reveal that the study of big data, analytics and machine learning in retail has become a multidisciplinary field in the field of operations research, closely related to demand forecasting, retail management, supply chain management, AI and data-driven decision-making. The trends in publications reveal significant growth in the field from 19 articles in 2016 to 114 in 2025. The linear trend also shows a clear upward trend, and the growth model suggests that the research field will continue to experience rapid growth, with a peak expected around 2035. The projections beyond 2026,

however, should be viewed with some caution as they are based on models, not on actual publication activity.

Figure 10: Growth model: real articles (dots) and the model curve. The part after 2026 is only a rough forecast drawn from the model values.



Integration of Statistical and Machine Learning Approaches

The future studies include the combination of classical statistical forecasting techniques with machine learning and Deep learning techniques, which is one of the most promising fields of future research. The bibliometric analysis highlights key themes in the area such as machine learning, data analytics and predictive analytics, while the thematic map shows that predictive analytics and artificial intelligence are prominent but somewhat less-developed themes. This implies the shift towards more sophisticated analytical approaches over traditional forecasting is a continuing process. A natural progression of this work is to analyse hybrid forecasting models that combine statistical techniques such as ARIMA and exponential smoothing with machine learning approaches. Future studies should evaluate these approaches under comparable datasets, forecasting horizons and performance measures in order to determine when hybrid methods provide meaningful improvements over individual models.

SKU Level Forecasting and Heterogeneous Demand

Another important research direction concerns the development of more granular SKU-level forecasting approaches. Retail businesses often require forecasts for individual products and stores, yet demand can vary substantially between products because of differences in volatility, intermittency, seasonality, product life cycle and sales frequency. Liao et al. (2020) highlighted limitations in approaches that do not sufficiently incorporate individual product characteristics, while others demonstrated that different items can display substantially different demand patterns. Several questions still remain to be answered. A further area of investigation concerns the role of promotional information and hierarchical relationships between products and stores. Teixeira et al. (2024), for example, demonstrated the relevance of hierarchical sales forecasting when promotional information is incorporated into forecasting models. This

would be a fruitful area for further work, particularly in developing adaptive systems capable of selecting or adjusting forecasting models according to the characteristics of individual SKUs.

Forecast Accuracy and Operational Performance

Future research also needs to rethink the ways of assessing forecasting performance. The traditional accuracy measures are useful for comparing models, but are not sufficient to determine if a forecast is useful from an operational point of view. More broadly, research is also needed to determine if stockout reductions, excess inventory reductions, holding cost reductions, emergency restocking reductions and/or service level improvements are achieved as a result of the increase in forecast accuracy. The literature already shows that demand forecasting, in addition to statistical accuracy, has been related to inventory and supply chain performance. A further study could assess the long-term effects of forecasting models based on a combination of statistical measures (MAPE, RMSE), and operational measures (inventory cost, service level and replenishment performance).

Integration with Inventory Management and Stock Replenishment

Demand forecasting and supply chain management are part of the conceptual framework for inventory management and time-series research are important areas within the field identified by thematic analysis. It is a clear sign of the relevance of the linkage between forecasting and operational decision making. The next generation of research should, therefore, consider demand forecasting and its relationship to inventory management, inventory control and stock replenishment, instead of treating these as distinct processes. The usefulness of the demand forecast depends on the way in which the forecasted demand is used when making the following inventory decision, as these studies illustrate. We propose to investigate integrated optimisation models where a demand forecast directly contributes to the reorder points, safety-stock levels and order quantities.

Probabilistic Forecasting and Risk Aware Decision Making

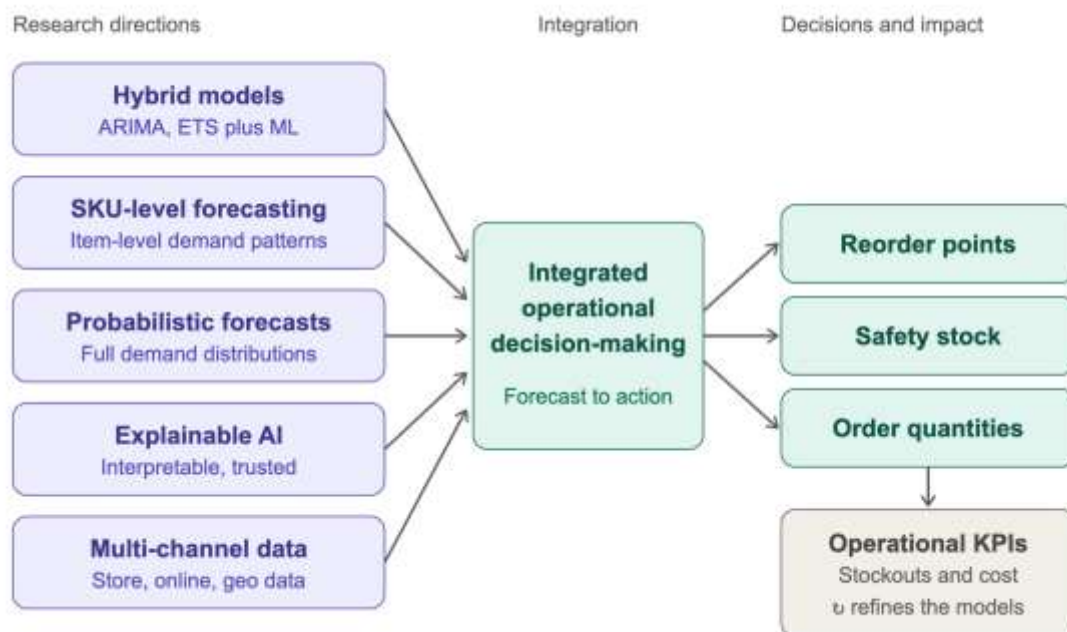
A further promising avenue for research is probabilistic forecasting. Conventional forecasting methods often provide a single expected value, whereas probabilistic approaches attempt to represent the uncertainty surrounding future demand. Wang et al. (2024), for example, proposed transformer-based probabilistic demand forecasting with adaptive online learning, highlighting the limitations of traditional methods in representing demand uncertainty. More research using controlled trials is needed to determine whether probabilistic forecasting produces better risk-adjusted outcomes than conventional point forecasting. Further work could also examine how forecast distributions can be incorporated into optimisation models for safety-stock determination, replenishment policies and other risk-sensitive inventory decisions.

Multi-Channel Retail and New Data Sources

The development of omni-channel retailing and digital commerce provides another important avenue for future research. The growing volume of literature has linked retail analytics to customer behaviour and made use of data analytics and predictive strategies, and research presented by Wu et al. (2024) and Hu et al. (2026), respectively, suggests that deep learning, automated machine learning and geospatial intelligence have the potential to be used for multi-channel retail forecasting. Such forecasting models for the future could be developed for combining the in-store and online sales data, along with customer behaviour, geographic information, promotion activity and other contextual information. More information on these contextual variables would help us to establish a greater degree of accuracy on this matter and expand the field of retail analytics beyond traditional sales information. Based on the bibliometric analysis, the authors can conclude that machine learning, data analytics and predictive

analytics/artificial intelligence are emerging themes, whereas sales, retailing and data mining are important established themes. Meanwhile, forecasting, deep learning and retail analytics seem to be additional areas that could be developed. These findings provide the following insights for future research: Hybrid statistical-machine learning models, SKU-level forecasting, Probabilistic and stochastic forecasting, Promotion-based forecasting, New-product forecasting, Intermittent demand and Multi-channel data analysis and explainable AI. These directions have the potential to ultimately contribute to closer links between demand forecasting and inventory management, supply-chain optimisation, inventory control and stock replenishment.

Figure 11: Proposed future research framework for integrated demand forecasting.



Accordingly, the central research question may need to move beyond “What will demand be?” towards “How can uncertain demand be predicted and subsequently used to support the best possible operational decision?” This represents a promising field for future research in operations research and retail management, particularly as predictive analytics and data-driven decision-making continue to gain importance.

7. CONCLUSION AND LIMITATIONS

After having collected and analysed the bibliographic data, it can be concluded that research on big data, analytics and machine learning in retail has developed into a multidisciplinary area which connects retail management, data analytics, machine learning and operational decision making. The analysis examines the development of publications, the distribution of research across different sources and countries, and the relationships between keywords and research themes. The results provide an understanding of how the field has developed over the selected period and how its focus has changed. However, it must be kept in mind that bibliometric analysis measures the patterns within published research and does not directly establish how effective these technologies are when applied in actual retail operations.

When these findings are considered alongside the literature review, it becomes clear that the development of retail analytics involves more than the increasing use of machine learning and artificial intelligence. Fildes, Ma and Kolassa (2022) explain that retail forecasting involves several challenges relating to demand drivers, aggregation and the evaluation of forecasting methods. Hofmann and Rutschmann (2018) also highlight the importance of treating big data analytics and forecasting as an integrated process. This is because the collection and processing of data can influence the usefulness of the final forecast. Similarly, Ma, Fildes and Huang (2016) demonstrate the importance of promotional information in retail sales forecasting, while Iftikhar and Khan (2020) examine how information from social media can be used alongside historical sales data. These studies show that the development of forecasting depends not only on the methods being used but also on the type and quality of information available.

However, there are still several challenges which need to be addressed. One of the main issues is the integration of different data sources, as sales information, promotional data and external digital signals may not always be consistent or equally useful. Additionally, advanced forecasting methods need to be compared with traditional methods under similar conditions before conclusions can be drawn about their relative performance. There is also a need to improve the interpretability of forecasting models so that retailers can understand the factors influencing their predictions. Finally, greater attention needs to be given to connecting forecast accuracy with actual operational outcomes such as inventory costs, stockouts and replenishment decisions. Future research should therefore focus on evaluating forecasting methods across different retail settings and understanding whether the improvements in prediction lead to meaningful improvements in decision making.

There are also limitations to the present study which must be considered. The analysis is based on publications retrieved from Scopus using the search string “Retail AND Analytics”, and therefore the results depend on the database and search strategy selected. Relevant publications which use different terminology may not have been included, while the broad search terms may also have retrieved some records which are not directly related to retail analytics. Additionally, the analysis uses bibliometric indicators and keyword relationships to examine the structure of the field, which means that it cannot directly establish the effectiveness of particular forecasting methods. The use of two periods in the thematic evolution analysis and the incomplete 2026 publication data also restrict the conclusions that can be drawn about long term changes. These limitations do not mean that the analysis is not useful, but they must be kept in mind when interpreting the results and identifying directions for future research.

Overall, the findings suggest that retail analytics is continuing to develop through the combination of big data, machine learning, predictive modelling and operational decision making. The increasing number of publications and the variety of research themes show that the field covers several interconnected problems rather than one single area of study. However, the availability of more data and advanced analytical methods does not necessarily guarantee better results. The quality of the data, the suitability of the forecasting method and the ability to use the results in practical situations remain important factors. Therefore, future research should not only focus on developing more advanced forecasting models but also on ensuring that these models can be compared fairly, interpreted correctly and connected with the decisions retailers need to make. Ultimately, the value of retail analytics lies not simply in predicting what may happen but in using those predictions to make better decisions.

REFERENCES

1. Adulyasak, Y., Cohen, M. C., Khern-Am-Nuai, W., & Krause, M. (2024). Retail Analytics in the New Normal: The Influence of Artificial Intelligence and the Covid-19 Pandemic. *IEEE Engineering Management Review*, 268–280.
2. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
3. Castillo-Vergara, M., Alvarez-Marin, A., & Placencio-Hidalgo, D. (2018). A bibliometric analysis of creativity in the field of business economics. *Journal of Business Research*, 85.
4. Chabowski, B. R., Samiee, S., & Hult, G. T. M. (2013). A bibliometric analysis of the global branding literature and a research agenda. *Journal of International Business Studies*, 44, 622–634.
5. Dekimpe, M. G. (2020). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*, 3–14.
6. Deng, L., Hu, X., & Xiao, Y. (2026). Blockchain-enabled strategic optimization: A decision analytics framework for multinational enterprises' e-commerce entry. *PLOS ONE*, 21(5).
7. Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*.
8. Hodzic, K., Hasic, H., & Cogo, E. (2024). Tactical sales forecasting using a very large set of macroeconomic indicators. *Expert Systems with Applications*, 13(1), 1–14.
9. Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: A conceptual analysis. *International Journal of Logistics Management*.
10. Hossain, M. A., Akter, S., Yanamandram, V., & Gunasekaran, A. (2022). Operationalizing artificial intelligence-enabled customer analytics capability in retailing. *Journal of Global Information Management*.
11. Iftikhar, R., & Khan, M. S. (2020). Social media big data analytics for demand forecasting: Development and case implementation of an innovative framework. *Journal of Global Information Management*.
12. Liao, S., & Yang, L. (2020). Mobile payment and online to offline retail business models. *Journal of Retailing and Consumer Services*, 57.
13. Lorente-Martínez, J., Navío-Marco, J., & Rodrigo-Moya, B. (2022). Are retailers leveraging in-store analytics? An exploratory study. *International Journal of Retail & Distribution Management*.
14. Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*.
15. Madanchian, M. (2024). The impact of artificial intelligence marketing on e-commerce sales. *Systems*.
16. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta analyses: The PRISMA statement. *International Journal of Surgery*, 8(5), 336–341.
17. Papanagnou, C. I., & Matthews-Amune, O. (2018). Coping with demand volatility in retail pharmacies with the aid of big data exploration. *Computers & Operations Research*.
18. Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know? *International Business Review*, 29(4).

19. Prancutè, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications*, 9(1).
20. Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*.
21. Punia, S., Singh, S. P., & Madaan, J. K. (2020a). A cross-temporal hierarchical framework and deep learning for supply chain forecasting. *Computers & Industrial Engineering*.
22. Punia, S., Singh, S. P., & Madaan, J. K. (2020b). From predictive to prescriptive analytics: A data-driven multi-item newsvendor model. *Decision Support Systems*.
23. Rooderkerk, R. P., DeHoratius, N., & Musalem, A. (2022). The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners. *Production and Operations Management*, 3727–3748.
24. Thomé, A. M. T., Scavarda, L. F., & Scavarda, A. J. (2016). Conducting systematic literature review in operations management. *Production Planning & Control*, 27(5), 408–420.
25. Tudor, C., Girlovan, A., & Botoroga, C. (2025). Where Business Meets Location Intelligence: A Bibliometric Analysis of Geomarketing Research in Retail. *ISPRS International Journal of Geo-Information*.
26. Velásquez, J. D. (2025). An analysis of trends, challenges, and opportunities in retail analytics. *International Journal of Market Research*, 394–422.
27. Velásquez-Henao, J. D. (2025). Emerging trends in Retail analytics: a bibliometric analysis of the last decade. *DYNA (Colombia)*, 16–29.
28. Youssef, M. A. E., Eid, R., & Agag, G. (2022). Cross-national differences in big data analytics adoption in the retail industry. *Journal of Retailing and Consumer Services*.
29. Zhuang, W. (2021). The Influence of Big Data Analytics on E-Commerce: Case Study of the U.S. And China. *Wireless Communications and Mobile Computing*.