



The application of machine learning for demand prediction under macroeconomic volatility: a systematic literature review

Manuel Muth¹ · Michael Lingenfelder¹ · Gerd Nufer²

Received: 20 November 2023 / Accepted: 23 May 2024
© The Author(s) 2024

Abstract

In a contemporary context characterised by shifts in macroeconomic conditions and global uncertainty, predicting the future behaviour of demanders is critical for management science disciplines such as marketing. Despite the recognised potential of Machine Learning, there is a lack of reviews of the literature on the application of Machine Learning in predicting demanders' behaviour in a volatile environment. To fill this gap, the following systematic literature review provides an interdisciplinary overview of the research question: “How can Machine Learning be effectively applied to predict demand patterns under macroeconomic volatility?” Following a rigorous review protocol, a literature sample of studies (n=64) is identified and analysed based on a hybrid methodological approach. The findings of this systematic literature review yield novel insights into the conceptual structure of the field, recent publication trends, geographic centres of scientific activity, as well as leading sources. The research also discusses whether and in which ways Machine Learning can be used for demand prediction under dynamic market conditions. The review outlines various implementation strategies, such as the integration of forward-looking data with economic indicators, demand modelling using the Coefficient of Variation, or the application of combined algorithms and specific Artificial Neural Networks for accurate demand predictions.

Keywords Machine learning · Macroeconomic volatility · Demand forecasting · Marketing predictions · Systematic literature review

JEL Classification C53 · E32 · C45 · M31

✉ Manuel Muth
muthman@students.uni-marburg.de

¹ School of Business and Economics, Philipps-Universität Marburg, Universitätsstr. 24, 35037 Marburg, Germany

² ESB Business School, Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany

1 Introduction

In light of geopolitical instabilities in Eastern Europe and the Middle East, changes in inflation and interest rates, and disruptions in global supply chains (Dörner et al. 2023; European Central Bank 2023), various business functions face a complex task in accurately predicting the behaviour of demanders. The economic circumstances influencing managerial operations have hence undergone profound changes and many of the existing prediction approaches rely on substantially different circumstances than those currently prevailing (Durst et al. 2022; Durugbo and Al-Balushi 2022). As a result, “contemporary organizations face environments with unprecedented levels of volatility, uncertainty, complexity, and ambiguity” (Troise et al. 2022), presenting new challenges for predicting the future behaviour of demanders.

In scientific literature, several authors emphasise the general potential of Machine Learning (ML) for predictive analytical tasks in important business areas like marketing (Huang and Rust 2021; Ma and Sun 2020; Verma et al. 2021). However, there is an existing knowledge gap in terms of predicting future demand with ML in volatile environments (Ghoddusi et al. 2019), leaving a significant need for further research. The lack of comprehensive literature reviews on this particular question is substantiated by preliminary searches in the Web of Science-database across 12,000 available journals and screening the 205 query results that include the title words “systematic literature review” and “machine learning”, along with additional investigations in further electronic resources. While some existing literature reviews address the overarching intersection of ML and applied management disciplines, there is a tendency to focus more on general questions such as primary application categories or global trend developments (see further Mustak et al. 2021; Verma et al. 2021; Vlačić et al. 2021). Although there are already review papers that address the application of ML and Artificial Intelligence, each adopting different specific perspectives (e.g., Kaushal et al. 2023; Keding 2021), none of the available reviews aligns with the thematic focus of this systematic literature review.

This background provides the underlying foundation for defining the research question of this systematic literature review. The general scientific problem is structured according to the “CIMO” logic (Context, Intervention, Mechanism, Outcome), formulated by Denyer et al. (2008) and recently addressed by Kuckertz and Block (2021). As a result, the research question of this paper is defined as: “How can ML (I) be effectively applied (M) to predict demand patterns (O) under macroeconomic volatility (C)?”. The aim is to collect and synthesise state-of-the-art academic knowledge surrounding this question while adopting an application perspective from marketing science.

In line with the recommendation of Linnenluecke et al. (2019), the aim of this work is “to consider interdisciplinary contributions” to ensure an integrative overview covering the relevant sources of information. To foster such an interdisciplinary synopsis, methodological concepts from computer science are adopted with a practical business perspective of marketing, while maintaining an inclusive approach towards other management disciplines such as economics. The research itself is confined to a primarily analytical viewpoint with an application-oriented focus. In

strict accordance to the research question, the aim of this review paper is to address the task of demand prediction, and therefore no focus is placed on explanatory or descriptive tasks (see further Thommen et al. 2017). Consequently, intentionally outside the scope of this research are a precise modelling of individual decision-making mechanisms, an explanation of consumer behaviour or an understanding of the underlying motivations and preferences of demanders. Moreover, this review paper follows the general understanding that ML applications are inherently observation-driven (see further Ghoddusi et al. 2019; Xie 2020). Therefore, the demand predictions discussed here are based on an existing dataset and, consequently, not guided by general economic theories or the functional relationships of market participants that are not reflected in the underlying data.

The structure of this review paper is built on the methodology by Xiao and Watson (2019) for systematic literature reviews. The core stages of their methodology provide general guidance for the structuring of this review paper and consist of planning the review, conducting the review, and reporting on the review. Based on this, the review paper is structured as follows: After the introduction (1), a brief theoretical and terminological foundation is provided (2). Subsequently, the research methodology of the literature review is presented (3), illustrating the planning of the review. The following section (4) then deals with the implementation of the review and involves an analysis of the literature sample. In this way, the steps of searching and selecting literature, as well as quality evaluation and data extraction, are demonstrated. Afterwards, an analysis of the article characteristics (4.1), an analysis of the abstracts (4.2), and an analysis of the full texts (4.3) are provided for the final literature sample—followed by the general limitations (4.4) of the study. The final section (5) provides a conclusive report on the key components and offers a comprehensive conclusion.

2 Theoretical background

2.1 Machine learning

Lanquillon (2019) distinguishes ML by enabling the generation of a model from a dataset by a learning procedure instead of by an explicit programming instruction. A more general definition is provided by Mitchell (1997), who describes ML as being able “to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” Over the last decades, the understanding of ML as a sub-field of Artificial Intelligence (AI) has become generally accepted (Ma and Sun 2020; Shaikh et al. 2022; Verma et al. 2021). From an econometrics perspective, Ghoddusi et al. (2019) observe that economists traditionally tend to focus on theoretically guided modelling involving statistical analysis of individual exploratory variables, whereas the primary focus of ML is on generating a prediction using available data inputs. In view of this, ML is characterised as observation-driven modelling, which is generally different from the way parametric models are generated, as parametric models involve a marginalisation over combinations of parameters and thus implicitly rely on underlying presumptions (Xie 2020). In contrast, observation-driven

“ML models do not make any pre-specified assumptions about the functional form of the equation, the interaction between variables, and the statistical distribution of parameters” (Ghoddusi et al. 2019). In fact, ML algorithms generally adopt all data input for training and are thus able to fit the model across very different data structures without requiring specific a priori considerations (Xie 2020).

ML can be categorised according to its associated learning styles. Supervised, unsupervised, and reinforcement ML are recognised as general core categories (Ernst et al. 2020; Murphy 2012). In addition, further learning styles can be found in the research literature (see further Ma and Sun 2020; Zhang 2020). In particular, supervised ML is used for a variety of practical prediction tasks in areas such as marketing (Ma and Sun 2020), for which reason it is described in more detail. A major goal of supervised learning is to draw conclusions about future or yet unknown developments from existing information. For this purpose, one or more target variables $y = (y_1, \dots, y_n)$ are associated with a certain number of potential input variables $x = (X_1, \dots, X_n)$. Subsequently, a model is fitted in such a way that the values of the input variables correspond to the values of the target variable with a minimum error: $f: x \rightarrow y$. This provides the opportunity to introduce new input variables $\tilde{x} = (\tilde{X}_1, \dots, \tilde{X}_n)$ into the model and to derive a prediction about unknown target variables $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_n)$ for arbitrary values $f(\tilde{X}_i) = y_i$. This can be applied to make a prediction about the classification of a group—for example, who will be a potential buyer—which is defined as a discrete prediction problem. In addition, a prediction for metric values is also possible—for example, about the expected turnover—which is known as a metric prediction problem (Goodfellow et al. 2016; Murphy 2012).

2.2 Predictive modelling

Albrecht et al. (2021) underline the recent development of ML in the specific context of prediction, noting that while in the past more traditional statistical or further empirical methods were employed, “more recently, ML as a subset of AI has been added to the domains contributing effectively to business prediction problems.” Prediction is one of the main application problems of ML. However, there are others, such as the processing of images for object classification or natural language processing for speech generation (Black et al. 2022)—including Generative Pre-trained Transformers (GPT) such as ChatGPT (Esmaeilzadeh 2023), which are beyond the scope of this review paper. The term “prediction” is frequently used in business and marketing management in regard to projections about unknown future states based on past and ongoing data (Kozak et al. 2021; Seyedan and Mafakheri 2020). By its very nature, this term therefore implies a future-oriented perspective on the outcome. In order to distinguish the role of prediction from the closely related term “forecast”, Kmiecik and Zangana (2022) offer a differentiation. They specify that “forecasting is a type of prediction, and it bases the future outcomes on temporal recorded data [...] In essence, every forecast is a type of future prediction; however, not all future predictions are forecasts, as forecasts focus on not only a future occurrence but also the time of the occurrence.” Thus, when comparing prediction and forecasting, the authors emphasise for the latter the necessity of temporally recorded

data, also known as a time series, which is a sequence of chronologically arranged values (X_1, \dots, X_n), measured to a certain time (t_1, \dots, t_n).

Taking a more detailed view of the matter, various types of approaches can be considered, such as quantitative predictions which allow for a measurement of the predictive error, and which are the main focus of this review. In addition, qualitative predictions can be applied, for instance, based on the intuition of experts (Kmieciak and Zangana 2022). Furthermore, there are different types in terms of the number of input variables, also called predictors, whereby either one (univariate) or multiple (multivariate) predictors can be included in a model (Homburg 2020). In this respect, Seyedan and Mafakheri (2020) emphasise the relevance of data inputs that can provide additional explanatory potential for predicting future events. They suggest: “Incorporating existing driving factors outside the historical data, such as economic instability [...] could help adjust the predictions with respect to unseen future scenarios of demand.” This suggests that a demand behaviour analysis requires consideration of the impact exerted by the prevailing circumstances. For example, demand behaviour can be influenced by company-related micro-factors, such as promotional activities. In addition, external macro-factors, such as market conditions, shape the environment in which demanders operate (Arunraj and Ahrens 2015; Efat et al. 2022).

2.3 Demand under volatility

The term “volatility” is employed in particular in econometric analyses, as well as in the financial sector, and is commonly expressed as the standard deviation (σ). This metric is the square root of the variance, defined by the following formula for a given population dataset of x_1, x_2, \dots, x_n : $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$ (N =total number of observations; μ =mean; x_i = i th observation) (Mondello 2022). In order to determine standard deviation on a comparable basis for volatility measurement, it is typically benchmarked against a reference value, with considerations for additional components such as trend (Cariolle and Goujon 2013; Loayza et al. 2007; Raju and Acharya 2020). However, it should be emphasised that volatility can also be unpredictable, especially when causal relationships or circumstances were previously unknown or not evident (Angus et al. 2023). Demand is one of the areas characterised by volatility, as Abolghasemi et al. (2020) note: “the demand for a particular product or service is typically associated with different uncertainties that can make them volatile and challenging to predict”, making it a critical aspect of marketing predictions. Potential strategies that might be considered to cope with volatility in demand include increasing inventory levels or capacity to cover the fluctuation. However, this has direct business implications in terms of inventory management, the costs associated with overstocking and impacts on capital commitment and liquidity (Lin et al. 2022). As a result, these strategies can involve significant additional effort along a marketing supply chain (Kmieciak and Zangana 2022). On this point, Kmieciak and Zangana (2022) indicate that “demand fluctuations could imply supply management problems and create a tendency to keep excessive stocks as a buffer to production systems. Using flexible and precise forecasting procedures

gives possibilities to gain good results even in capricious markets.” This results in the need to deepen the understanding of how to anticipate volatile demand patterns and incorporate them to a certain extent into predictive methods for guiding informed decision-making (Lin et al. 2022).

In this context, a more detailed observation of the macroeconomic circumstances can provide insights, especially if they exert a significant impact on demand (Hasheminejad et al. 2022). This requires looking beyond the perspective of a single company as an individual economic unit and incorporating the general macroeconomic environment that encompasses the overall situation of the real and monetary economies (Conrad 2020). From a real-economy perspective, there is a systematic shift in the way consumers allocate their budget depending on economic developments (Kamakura and Du 2011). Kamakura and Du (2011) explain that during economic downturns, “consumers generally reduce their consumption budget [for non-essential commodities] either because their income is lower, or because they become more risk averse, allocating more of their income towards savings, which forces them to satisfy essential needs first”. Related to this, customers’ responses to price changes can be affected by economic growth rates (Gordon et al. 2013). For example, in many product categories, price sensitivity increases when the economy weakens (countercyclical behaviour). According to Gordon et al. (2013), this seems “consistent with the intuition that consumers become more price sensitive during weaker economic periods.” However, this is not necessarily always applicable. In regard to substitute goods, for instance, demand may even increase when a macroeconomic downturn occurs (procyclical behaviour). Hence, volatile macroeconomic circumstances create a dynamic environment for demanders’ behavioural responses.

2.4 Macroeconomic environment

From an academic viewpoint, Loayza et al. (2007) see volatility induced by the macroeconomic environment as being related to factors such as external shocks, economic policies or microeconomic and institutional distortions. In regard to the current situation, Taskan (2022) notes: “In recent years, organisations have faced large and unexpected events, such as financial crises, the COVID-19 pandemic, climate change and war, with a large impact on the world at several economic and societal levels, and the acronym VUCA [Volatility, Uncertainty, Complexity und Ambiguity] has been frequently used by scholars and practitioners to try to understand such environmental dynamics.” The recent example of COVID-19, with its major implications, is also taken up by Tudor (2022) and classified as a “black swan” event, referring to it as an external shock to the general environment with several multi-layered consequences. The associated effects on the behaviour of demanders are reflected, for example, in the increased confidence intervals of demand patterns during this period (Ahmed et al. 2022; Ma and Fildes 2020).

Another very recent and to some extent related example of the shifting current macroeconomic landscape is the emergence of high inflation rates, intertwined with changes in monetary policy, shaping the behaviour of demanders. The term

“inflation” here is conventionally understood as a sustained increase in the general price level (Conrad 2020). For example, annual inflation in the Eurozone was at a remarkable level of 8.4% in 2022 and is projected by the European Central Bank (2023) to remain above its 2% target until 2025. The reason for this high inflation is multi-causal, and the determining factors include price increases in the energy and food sectors. These are in turn caused by circumstances such as the war in Ukraine—most recently, also the conflicts in the Middle East—but also by ongoing disruptions in supply chains as a consequence of the COVID-19 pandemic (Ministry for Economic Affairs and Climate 2023). For consumers, the occurrence of high inflation usually implies that a given nominal amount of money can be used to buy fewer services and goods, which—according to general economic understanding—can lead to a loss of purchasing power following higher prices and the shifting of demand for goods and services (Sievering 2021). In addition, the current price may send ambiguous signals about future price developments in an inflationary environment, which can lead to volatile or misleading demand decisions. A further complicating factor is that the associated price increases usually do not occur at the same time but in a delayed sequence (Conrad 2020). Hence, consumer demand typically reflects not only the price of an individual item but also its price relative to the average of the category and, more broadly, to the average of its category relative to other categories (Danaher and Brodie 2000). However, not only do the actual prevailing conditions matter but also the expectations of households about the future development of the monetary situation, as well as their own perceived uncertainty. As shown in the research by Duca et al. (2010), consumers are more likely to make a larger purchase when a larger change in inflation is expected. As a conclusion of this discussion, it appears almost imperative to consider the macroeconomic situation and its effects in greater detail when dealing with demand predictions.

3 Research methodology of the literature review

As the academic contribution of a literature review is essentially determined by the compelling quality of its methodological design, Linnenluecke et al. (2019) posit that the employment of “rigorous methods and the clarity of reporting, as well as on the application of scientific strategies” are crucial components of a systematic literature review. Similarly, Walter (2021) indicates that only “a well-structured, transparent, and replicable methodology results in a reliable basis of knowledge” (see further Fink 2014; Fisch and Block 2018; Petticrew and Roberts 2006). For this reason, this review paper contains a description of its exact methodology, which encompasses an overview of the structural approach and a systematic explanation of the individual steps. Therefore, this systematic literature review differs in its approach to analytical literature evaluation from trial-and-error methodologies (Kaushal et al. 2023; Tranfield et al. 2003).

Adhering to the framework provided by Xiao and Watson (2019) for a systematic literature review, this methodology generally contains three main phases, which are defined as: planning the review, conducting the review, and reporting the

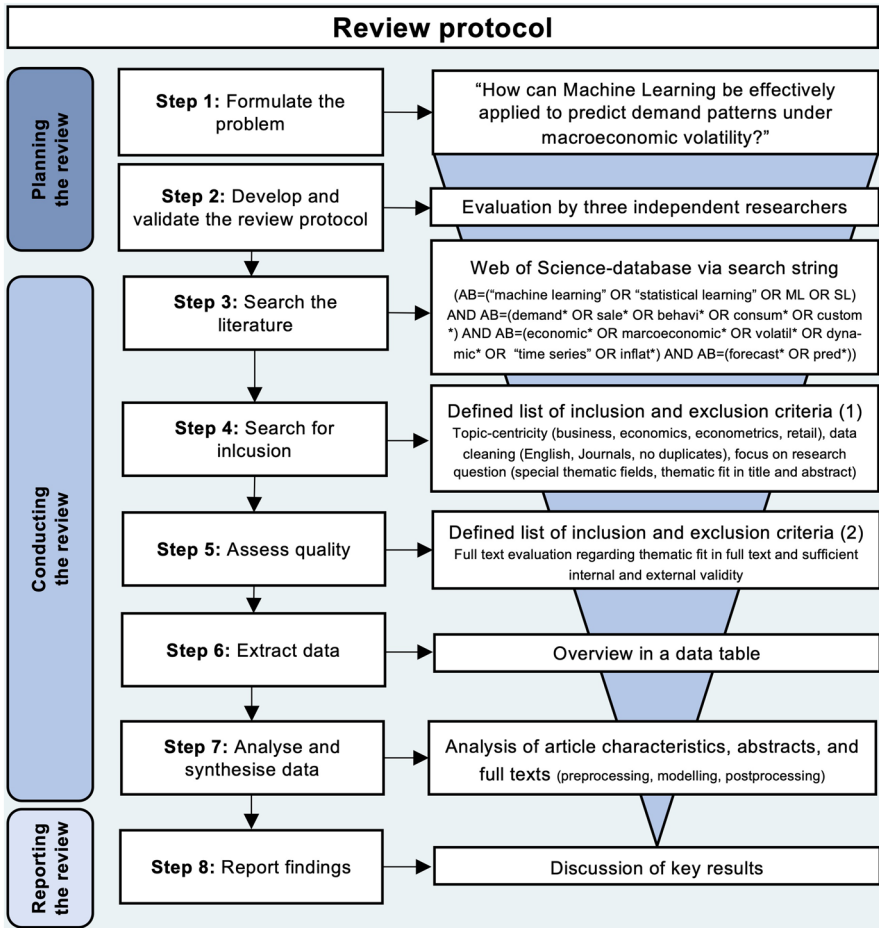


Fig. 1 Review protocol (adapted from Xiao and Watson 2019)

review (see Fig. 1). These stages show a high degree of similarity with those of Tranfield et al. (2003), which are widely used in the current management literature (Durugbo and Al-Balushi 2022; El Shoubaki et al. 2021; Göcke et al. 2021). Nonetheless, Xiao and Watson's (2019) suggestion is followed in this review paper, as it incorporates more recent scientific findings and provides a comprehensive notation and detailed description of each subcategory. The first main phase of planning the review includes two sub-steps. The first sub-step involves formulating the overarching research problem of "How can ML be effectively applied to predict demand patterns under macroeconomic volatility?" which, as previously explained, aligns with the structure of the "CIMO" logic. The second sub-step requires the development and validation of a review protocol, which represents a predefined plan detailing the research methodology. As depicted in Fig. 1, the review protocol serves to ensure the reliability of the entire study by enabling others to replicate and verify the results

(Kitchenham and Charters 2007; Xiao and Watson 2019). The research protocol is validated by three independent researchers to collectively ensure adherence to rigorous quality criteria.

Following the planning of the review, the second main phase, the conduct of the review, is initiated, which includes the process of searching pertinent literature. Electronic databases are utilised for this purpose, as they are usually used as primary sources for contemporary literature reviews (Booth et al. 2016; Xiao and Watson 2019). One of the central rules of ML is also applicable to literature reviews: the value of the results is dependent on the scope and, in particular, the quality of the data source (Lanquillon 2019). Therefore, the selection of an appropriate database in this study is founded on the empirical findings derived from the database analysis of Gusenbauer and Haddaway (2019). Utilising a list of 27 criteria, the authors assess the suitability of various databases for literature reviews. They conclude that Web of Science belongs to the academic search systems that are “well-suited to evidence synthesis in the form of systematic reviews in that they met all necessary performance requirements” (Gusenbauer and Haddaway 2019). Moreover, Web of Science facilitates multidisciplinary queries, a crucial feature given the interdisciplinary nature of the current research question, and a search in over 12,000 high-impact journals (Mustak et al. 2021).

In addition to selecting the search medium, also the definition of an appropriate search string necessitates a rigorous scientific approach. The search string is formulated in alignment with the research question, offering explicit guidelines concerning the intended articles (Xiao and Watson 2019). Thus, the keywords within this string are chosen to ensure that the results contain a high number of relevant records (high sensitivity) while minimising the presence of irrelevant records (high specificity) (Gusenbauer and Haddaway 2019). Therefore, the research question is divided into distinct keywords. To avoid excluding important results, pertinent synonyms, related terms, and abbreviations (e.g., Machine Learning = ML) are incorporated alongside the original keywords. Additionally, so-called truncations (e.g., *) are employed to accommodate various forms of the words (e.g., volatil* = volatility, volatile) (Booth et al. 2016; Denyer and Tranfield 2009; Kitchenham and Charters 2007; Rowley and Slack 2004; Xiao and Watson 2019). Ultimately, the keywords are logically connected using Boolean operators (e.g., AND, OR, NOT) (Gusenbauer and Haddaway 2019). For the review, this process yields the following search term, which is derived from a comparison with methodologies from other literature reviews (e.g., Mustak et al. 2021) and an iterative refinement process:

AB = (“machine learning” OR “statistical learning” OR ML OR SL).

AND

AB = (demand* OR sale* OR behavi* OR consum* OR custom*).

AND

AB = (economic* OR macroeconomic* OR volatil* OR dynamic* OR “time series” OR inflat*).

AND

AB = (forecast* OR pred*).

Utilising the aforementioned search string to query the Web of Science-database results in a total of 2877 results within the screening for literature phase. However, this does not yet constitute the final literature sample. The retrieved results necessitate a cleaning process to eliminate duplicates and false positives, specifically those results that appear relevant based on the search query but do not actually fulfil the criteria (Linnenluecke et al. 2019; Xiao and Watson 2019). This is undertaken within the screening for inclusion sub-step. When assessing individual studies, inclusion and exclusion criteria needed to be defined to minimise individual leeway, which may inadvertently introduce substantial bias due to the arbitrary exclusion of studies. Consequently, these criteria serve as objective guidelines for the screening process. In Table 1, a set is presented of criteria to ensure transparency and justification for the selection of the final literature sample (structure based on Keding 2021; Schmid and Dowling 2022; Xiao and Watson 2019). The initial inclusion criterion mandates that papers address the subjects of business, economics, econometrics, and retail, as these academic fields specifically represent the intended scholarly perspective of this study. Next, only journal articles in English are included, and duplicates are systematically excluded from the literature sample. Additionally, only papers from journals are included in the literature sample to ensure comparability and maintain the robustness of the data through the typically double-blind peer-review process of journals with extensive revision cycles. At this point, it is worth mentioning that an intentional decision was made to reject some widespread other exclusion criteria. For instance, no criterion was established concerning the quality of the publishing journal in order to encompass the full spectrum of scientific activity within the sample. This is in line with the opinion of Denyer and Tranfield (2009) who state, “Reviewers are best advised to guard against using proxies for research quality such as the quality rating of journals as a basis for exclusion.” Additionally, also the number of citations is not employed as a criterion to select the papers, as a significant correlation ($r=0.223$, $p=0.003$) exists within the literature sample between the elapsed time since the publication date and the citation count. Consequently, relying on citation count would tend to favour older publications, thereby undermining the clear purpose of a state-of-the-art review.

Subsequently, a selection of papers is conducted based on the relevance of their thematic fields in relation to the research question. Included are contributions with a primary Web of Science-category in Computer Science, Business, Economics, Engineering, Management, Mathematics, Information Science and Library Science, Agronomy, Operations Research and Management Science, Automation and Control Systems, and Multidisciplinary Sciences.¹ A stepwise approach is implemented to screen all identified studies. This process begins with an examination of the title,

¹ This results in the exclusion of the following precise categories: “Agricultural Economics & Policy”, “Agriculture”, “Behavioral Sciences”, “Biology”, “Biotechnology & Applied Microbiology”, “Chemistry”, “Construction & Building Technology”, “Education & Educational Research”, “Energy & Fuels”, “Environmental Sciences”, “Food Science & Technology”, “Geosciences”, “Green & Sustainable Science & Technology”, “Health Care Sciences & Services”, “Mathematical & Computational Biology”, “Mechanics”, “Medical Informatics”, “Medicine”, “Philosophy”, “Physics”, “Primary Health Care”, “Psychology”, “Robotics”, “Thermodynamics”, “Transportation”, and “Urban Studies”.

Table 1 Selection of the literature sample

Inclusion and exclusion criteria	Search string formulation	AB=(“machine learning” OR “statistical learning” OR ML OR SL) AND AB=(demand* OR sale* OR behavi* OR consum* OR custom*) AND AB=(economic* OR marcoeconomic* OR volatili* OR dynamic* OR “time series” OR inflat*) AND AB=(forecast* OR pred*)	n = 2877
	Database selection	Web of science	
Step 4: screen for inclusion (i)	Topic refinement	Inclusion of papers with general topics business, economics, econometrics or retail	(– 2706) n = 171
Step 4: screen for inclusion (ii)	Data cleaning	(a) Inclusion of papers written in English (b) Inclusion of papers in Journals (c) Exclusion of duplicates	(– 3) n = 168 (– 3) n = 165 (– 2) n = 163 (– 40) n = 123
Step 4: screen for inclusion (ii)	Research question narrowing	(a) Inclusion of thematic fields generally related to the research question, based on primary Web of Science category: “Computer Science”, “Business”, “Economics”, “Engineering”, “Management”, “Mathematics”, “Information Science and Library Science”, “Agronomy”, “Operations Research and Management Science”, “Automation and Control Systems”, and “Multidisciplinary Sciences” (b) Exclusions of titles with a lack of thematic fit to the research questions (c) Exclusions of abstracts with a lack of thematic fit to the research questions	(– 28) n = 95 (– 24) n = 71 (– 7) n = 64
Step 5: assess quality	Full-text assessment	Exclusions of full texts with a lack of thematic fit to the research questions or insufficient internal and external validity	

proceeds with the assessment of the abstract, and ultimately culminates in the comprehensive appraisal of the paper's entirety, as recommended by Xiao and Watson (2019). Finally, the full texts are assessed for thematic fit, sufficient internal validity—the absence of significant bias in the methodological approach—and external validity—the study's content is, at least in parts, suitable for deriving generally applicable recommendations (drawing upon Petticrew and Roberts 2006; Rousseau 2006; Schmid and Dowling 2022; Xiao and Watson 2019). This results in a final sample size of 64 studies.

4 Analysis of the literature sample

Based on the literature sample derived from the filtering process above, the most relevant information from each of the remaining studies ($n=64$) is extracted (Xiao and Watson 2019). To accomplish this, the papers from the Web of Science-database are organised in a table describing details such as the title, authors, publication year, author keywords, source or research area. Following this extraction step, the sample is subjected to further analysis in order to examine and synthesise the contents of the papers.

The findings are presented in three segments: general article characteristics, abstract, and full text. The first segment is focused on the publications per year, the geographical distribution, as well as the publishing journals in relation to the keywords and references of the articles. The second segment contains an examination of the abstracts and identifies generalisable patterns across them, taking into account the considerable word count across the included studies ($n=64$). The third segment involves consideration of the text part of the paper itself, with the provision of an overview of the literature sample related to the research question and a presentation of the findings in chronological order across the categories of preprocessing, modeling, and postprocessing. Finally, in the last segment of this chapter, limitations are presented for the analysis of the literate sample.

To conduct this analysis, a hybrid methodological approach is employed, which incorporates both quantitative evaluations—primarily based on numerical figures for a structural summary—and qualitative content analysis that reflects the essential aspects drawn from the literature sample for a differentiated insight. The quantitative analyses are primarily based on quantitative evaluations involving calculated values or diagrams, facilitating the identification of overarching trends to deliver a comprehensive overview (Trapp 2012). To visually structure the extensive information found in the papers, the R programming language is utilised within the RStudio programming environment (Sauer 2019). Data analysis and visualisation of the results are performed largely by using the R programming package “Bibliometrix” (see further Aria and Cuccurullo 2023, 2017; Kaushal et al. 2023). Furthermore, the analysis also includes qualitative content analysis, wherein the key messages of the individual studies in the literature sample are presented, critically discussed, and related to one another (Trapp 2012). The motivation for this procedure is, as Trapp (2012) points out, that relying solely on this quantitative approach offers only limited depth and lacks differentiated insight into the individual contributions. Furthermore,

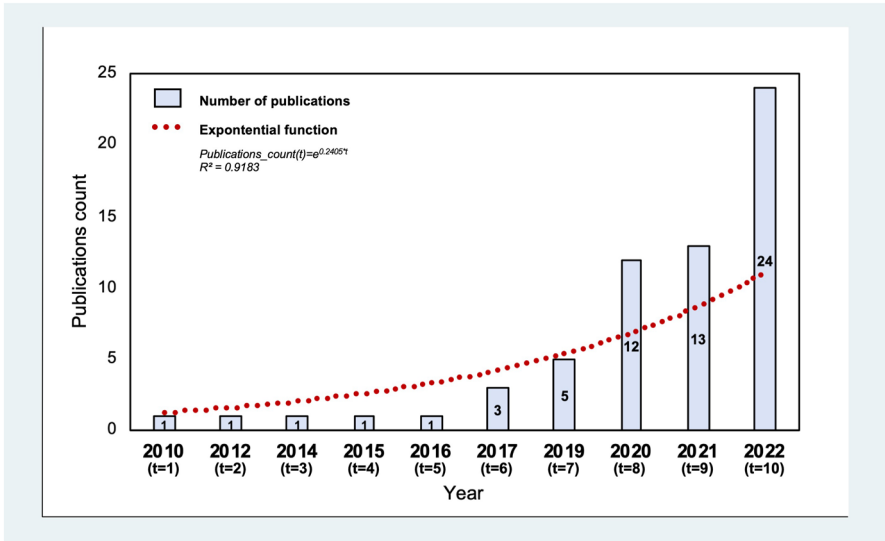


Fig. 2 Number of publications per year with publication trend (for years with recorded publications). Note: For figure 2, years with no publication (Publications_count (t)=0) are excluded when calculating the trend lines, as the logarithm of 0 is not defined, which is necessary for fitting an exponential trend line (Publications_count (t) = a·e^{bt}). All trendlines are set to intersect the y-axis at 1. The year 2023 is omitted, as for a correct interpretation only complete years are included in this calculation. The exponential function may not be extendable into the future, since the assumption of exponential growth might not hold in the event of a trend reversal in subsequent years. Notation: t = number of years with Publications_count(t) > 0 since the base year, Publications_count(t) = number of publications in t

Niedzela (2022) explicitly emphasises that multiple perspectives—such as quantitative analysis and qualitative synthesis—contribute to scientific versatility and thus can reinforce the progress of the research field.

4.1 Analysis of article characteristics

Initially, the number of papers in the literature sample (n=64) is assessed in relation to the year of publication and visualised in a bar chart (see Fig. 2). Within this visualisation, a substantial increase since 2020 can be observed. To systematically describe the detected trend, multiple functional relationships are examined. As a result, a modelled exponential function (R²_{exponential} = 0.9183) exhibits a higher goodness of fit than a linear function (R²_{linear} = 0.7113). Based on this finding, it can be inferred that the number of publications in years with recorded publications, currently tends to follow an exponential rather than a linear trend since 2010 (Publications_count(t) = e^{0.2405·t}). One possible explanation for this increase could be that it indicates a dynamic growth of interest in research on the interdisciplinary topic discussed in this literature review.

A further analysis is conducted to understand the geographical distribution of the papers. The country’s scientific production is measured in terms of the appearance

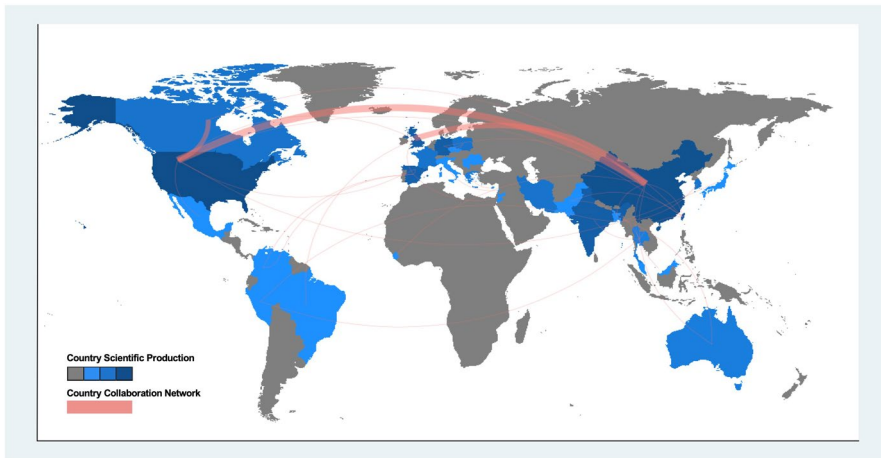


Fig. 3 Country scientific production and collaboration network (geographical distribution). Note: “Country Scientific Production” shows the amount of author appearances according to country affiliation (the intensity of blue indicates the strength of production). “Country Collaboration Networks by Region” represent co-authorships identified through a country collaboration matrix (visualised via red links between countries visualised with the “Map Parameter” set to “Min edges=2”) (Aria and Cuccurullo 2023; Aria and Cuccurullo 2017)

of authors by country affiliation. The highest numbers are recorded from the People’s Republic of China ($n=52$), the United States of America (U.S.) ($n=37$), Germany ($n=24$) and India ($n=20$). In terms of continents, most scientific production comes from Asia ($n=100$), followed by Europe ($n=86$) and North America ($n=46$). An indication of the number of publications across all countries is illustrated by the colouring of the world map in Fig. 3. Furthermore, Donthu et al. (2021) underline that “it is [...] important to understand how scholars interact amongst themselves” and consequently to reveal the intellectual collaboration of researchers in different countries. Therefore, the strength of collaboration across countries is also indicated on the map, visualised by red connections. According to this, collaboration for publications in the literature sample appears to be strongest between China and the U.S., followed by China and the United Kingdom (UK), and the U.S. and Canada.

As an additional analysis, the journals (left) from which the publications originate most frequently, the keywords (middle) defined by the authors, and the most cited references (right) within the papers are related (see Fig. 4). In the figure, a sankey diagram is intended to give insights into the interplay of these three dimensions with the aim of providing guidance in identifying relevant core literature and leading authors in the interdisciplinary research field examined in this paper (Trapp 2012). The connection between the elements represents the co-occurrence between them and the thickness specifies the number of times they co-occur (Donthu et al. 2021). Based on the evaluation illustrated in the figure, “Expert Systems with Applications” can be identified as the leading journal with the highest total number of mentions ($n=145$) among the reference lists of all studies in the sample. It is primarily connected with the most relevant keyword, ML, and also makes the most

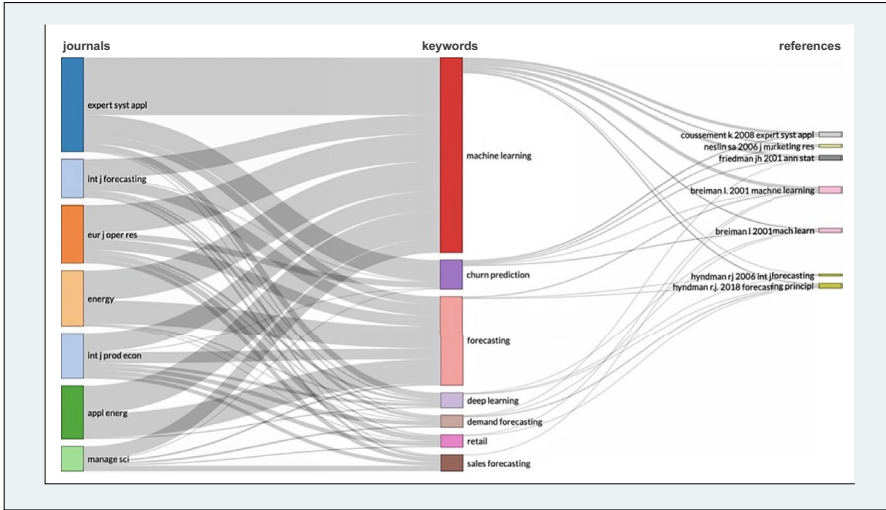


Fig. 4 Sankey diagram of journals—keywords—references. Note: Figure 4 visualises a “Three-Fields Plot”, with “Left Field”=Journals (Sources), “Middle Field”=Keywords (Authors’ Keyword) and “Right Field”=References (“Cited References in the papers”), using the parameter setting “Number of Items = 7” (Aria and Cuccurullo 2023; Aria and Cuccurullo 2017). The sankey diagram is limited to the most common journals, keywords, and references, and can thus oversimplify the complex interconnections between them. Only the first-named author of the reference is mentioned on the right-hand side to facilitate a concise visualisation

significant contribution to the churn prediction and the deep learning keywords. In contrast, “Energy” (n=53) and “Applied Energy” (n=49) are particularly high-contributing journals for the keyword forecasting. Furthermore, on the right-hand side of Fig. 4, it can be seen that the article by Breiman (2001) is mentioned most frequently in connection with the keyword ML and is cited a total of 16 times as reference. In addition, the contributions of Friedman (2001) (n = 7), Coussemont and Poel (2008) (n = 6) and Neslin et al. (2006) (n = 5) are cited with high frequency. For the keyword forecasting, Hyndman and Athanasopoulos (2018) (n = 6) and Hyndman and Koehler (2006) (n = 5) emerge as very important. Finally, it is evident from the sankey diagram that the leading source for the keyword demand forecasting is the “International Journal of Production Economics” (n = 53), with a pronounced emphasis on technical perspectives, while for the keyword sales forecasting, the primary source is the journal “Management Science” (n = 48), with a stronger business-oriented focus. Consequently, further attention is needed to monitor how certain terminologies for similar topics are more prevalent in specific academic (sub) disciplines than in others, particularly in regard to interdisciplinary studies.

4.2 Analysis of abstracts

In this section, an examination of abstracts of all publications in the literature sample (n = 64) is provided for the purpose of synthesising core information and

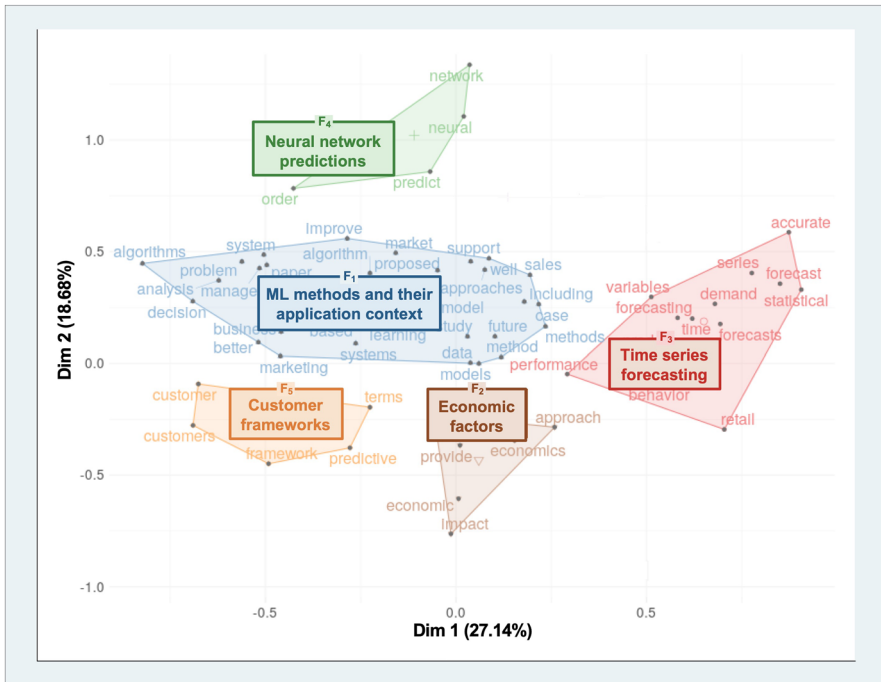


Fig. 5 Conceptual structure map. Note: Figure 5 visualises a “Word Map” via the “Conceptual Structure” command based on a word occurrence matrix and with the parameter setting “Method=Correspondence Analysis” and “Field=Abstracts”. In the K-Means Clustering, all clusters with fewer than 3 words were removed for clarity purposes (Aria and Cuccurullo 2023; Aria and Cuccurullo 2017)

discovering structural relationships among them (Trapp 2012). For this, a conceptual structure map (see Fig. 5) is generated to identify the specific knowledge profile in the abstracts (Aria and Cuccurullo 2017). At the outset, relevant terms are filtered from all abstracts, removing irrelevant terms such as determiners or conjunctions from the total set of words and applying Porter’s (1980) stemming algorithm to reduce all words to their root form. Then, a Correspondence Analysis is applied as a data reduction technique to map the thematic spaces in two dimensions, followed by K-Means Clustering to identify clusters of studies expressing common concepts (Aria and Cuccurullo 2017). Within the abstracts, five focal points (F_{1-5}) can be identified, which are assigned a superordinate description term. The largest one (blue) can be summarised under the category (F_1) ML methods and their application context, which is, in addition to methodological terminologies (e.g., algorithms, models), related to the focused area. This seems appropriate, considering that over 350 algorithm or model usages are registered in the studies of the literature sample. It is also apparent in the conceptual map that another focal point (brown) exists that deals with the (F_2) Economic factors. This can be attributed to the fact that a number of studies involve consideration of economic impact, market circumstances, sentiments, or macro-environment

in their datasets or in their predictive modelling approaches. Two further focal points covered in the abstracts are of a more technical nature. They involve (F₃) Time series forecasting (red) and the usage of (F₄) Neural network predictions (green). While 27 studies focus either entirely or partially on time series data for their empirical modelling, Artificial Neural Networks are employed 89 times for this purpose—sometimes in various ways within a single study—which underlines the relevance of this method. The remaining focal point relates to (F₅) Customer frameworks (orange), as customer-related activities and processes are a consistent research motive for ML applications in this specific research context.

4.3 Analysis of full texts

The analysis of the abstracts is followed by a review of the full texts, where the aim is to provide a topic-centred overview and recapitulate the key textual perspectives. The approach adheres to some of the instructions provided by Linnenluecke et al. (2019), who recommends researchers going “through ways that prior publications have contributed to developing [...] understanding of themes, concepts or phenomena of interest.” This process is an attempt to find possible approaches on how to managing the ML application process, adapting it to macroeconomic volatile conditions and the interdisciplinary requirements for predicting demand behaviour. To ensure concise reporting of the findings, they are provided in accordance with the common chronological order along the ML workflows, using the following three overarching categories: preprocessing, modelling, and postprocessing. This includes underlying questions such as which state-of-the-art model algorithms are most suitable in this context and which (macroeconomic) predictors are appropriate for accurate demand predictions. Beyond summarising the major findings of the papers within those subcategories, the review also involved an attempt to identify divergent views and inconsistent directions of research approaches (Trapp 2012).

4.3.1 Preprocessing

The following section contains an analysis of the preprocessing strategies that aim to enhance ML results in predicting demand in the marketing context within dynamic macroeconomic settings. Preprocessing is predominantly focused on collecting raw data as well as the subsequent handling procedure of this data to construct an optimal feature set for the predictive model (Kharfan et al. 2021; Punia and Shankar 2022; Wang 2022). Authors describe this stage as both a challenge and a critical determinant for the final quality of the model, with a primary focus on selecting an appropriate set of input variables (Raizada and Saini 2021). This section includes a comparison of the performances of univariate models that rely on a single input variable with complex models that factor multiple variables, with an emphasis on their respective predictive capabilities in volatile macro-environment settings. Particular attention is given to the selection of model variables for anticipating the macroeconomic conditions.

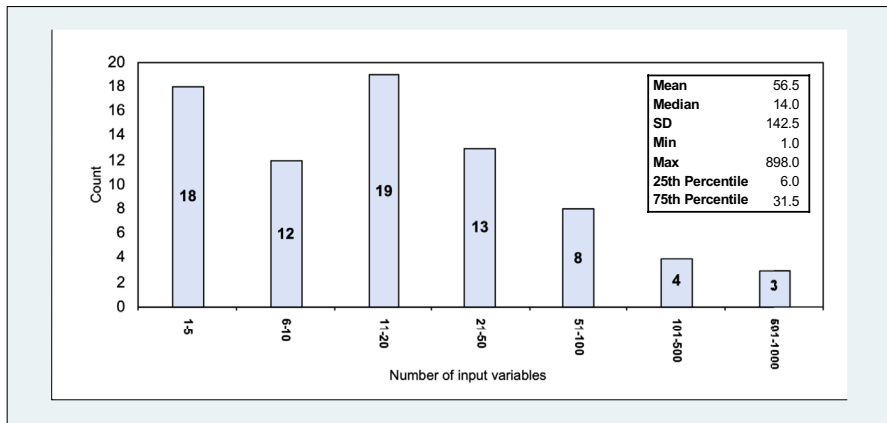


Fig. 6 Number of input variables. Note: If a study contains multiple datasets, each one is accounted for in the evaluation, which may result in a single study contributing more than one set of input variables. Figure 6 is based on the datasets of the empirical studies

In fact, a large number of the reviewed studies in the literature sample ($n=49$) either partially or entirely used empirical models which include multiple input variables (multivariate). Within the context of demand prediction, numerous papers highlight the advantage of multivariate methods compared to univariate approaches (e.g., Abolghasemi et al. 2020; Claveria et al. 2020; Punia and Shankar 2022). This preference primarily seems to be attributable to the inherent constraints of univariate methods that rely solely on a single input variable, such as historic demand behaviour, to extrapolate future developments. Punia and Shankar (2022) measure in their empirical investigations superior performance by employing the same prediction model with rather than without contextual variables. Consequently, Abolghasemi et al. (2020) conclude that specific univariate methods “work only well when the future is similar to the past [...] [and] might fail to forecast well if demand time series is subject to volatility.” Similarly, Hasheminejad et al. (2022) underscore that traditional, rather univariate prediction methods “generally do not work when the market is constantly fluctuating.” This perspective aligns with the understanding that demand is typically influenced by a multitude of factors (Ma and Fildes 2020), thereby implying that the integration of external factors holds the potential to make a number of uncertainties in future demand predictions explainable, thus reducing them (Ghodusi et al. 2019). The empirical studies in the literature review further support the importance of a comprehensive integration of input variables. An analysis of the studies in this review reveals that researchers consider a median of 14 input variables for their modelling, with the maximum observed in the large customer dataset used by Wang et al. (2019) with a total of 898 input variables (see Fig. 6).

Regarding the specific context of this research, Wang (2022) differentiates the problem of selecting appropriate model variables into two distinct issues: (1) determining whether and how the variables exhibit temporal causality (time lags)

and (2) identifying the most effective predictors. Concerning the first issue (1), the author classifies potential input variables based on their time dependencies into three categories: leading, coincident, and lagging variables. These categories indicate whether a potential input is expected to change before, after, or simultaneously with a variation in the demand output. When predicting in the context of broader economic conditions, Wang (2022) highlights the significance of the first one and posits that a “leading indicator helps [...] predict future changes before the outcome of the economy begins to go up or down [and can thus] used as an alert signal.” Poza and Monge (2023) concur with this view, emphasising the advantage of leading variables in not only anticipating trends but also discerning turning points in the economic market environment.

For the practical identification of relevant demand dependencies, the reviewed papers contain mentions of the Autocorrelation Function or the Partial Autocorrelation Function ($n=5$) as a relevant approach to identify them before modelling. These methods indicate the correlation of a potential demand series with its delayed variants, respectively their residuals; hence, they provide guidance for determining the lag structure within the data (see further Alsahref et al. 2022; Bukhari et al. 2020; Contreras-Masse et al. 2022; Kmiecik and Zangana 2022; Meisenbacher et al. 2022). Other methods are also employed by researchers, such as Wang (2022) and Wu et al. (2022), using Akaike and Schwarz Information Criteria, while some authors do not explicitly indicate their procedure for determining the lag structure or consider performing this immediately within modelling (e.g., Gürses-Tran and Monti 2022; Liu et al. 2021).

To gather data encompassing such leading variables for demand prediction, several studies suggest relevant approaches (e.g., Ghonghadze and Lux 2012; Petropoulos and Siakoulis 2021), such as using future-oriented sentiments or expectations to anticipate the macroeconomic situation. Since they are not directly observable, two major approaches can be identified in the literature sample to obtain further information about this: a more traditional approach that uses economic indicators, especially based on sentiment data from surveys, and a relatively modern approach relying on search engine trends. Economic indicators can be derived for instance from economic tendency surveys that address questions such as how individuals perceive their own financial situation or the economy in general. Using survey responses is seen to be a valid source of information for this purpose, as: “(a) they are based on the knowledge of agents that de facto operate in the market, (b) they contain information on a wide variety of economic variables, and (c) they are available prior to the publication of official data” (Claveria et al. 2020). Respondents are asked to evaluate whether such economic-related developments are likely to develop positively, remain unchanged, or develop negatively in the future. From the responses of a survey, balances can be computed by contrasting positive and negative responses, excluding neutral ones, and dividing by the total number of respondents (Balance (B) = $\frac{\text{Positive response (n+)} - \text{Negative response (n-)}}{\text{number of participants (n}_{\text{total}})}$) (Ghonghadze and Lux 2012;

Claveria et al. 2020). The economic sentiment indicators derived are considered as “key for monitoring the current state of the economy and providing forward

looking information” (Claveria et al. 2020), and Meisenbacher et al. (2022) observe that sales data in particular exhibit a frequent correlation with such economic indicators.

In terms of the specific empirical application of particular indicators, no universally adopted indicators could be identified in the literature sample. This suggests that the indicators for economic sentiment can also be ambiguous, their usage tends to be complex, and none of them is consistently applicable across all demand prediction scenarios in marketing, making it important to critically evaluate these indicators on a case-by-case basis. However, there are certain studies that refer to official indicators associated with the European Union, for example in the context of Joint Harmonised EU Programme of Business and Consumer Surveys. Claveria et al. (2020) discuss the corresponding consumer confidence and industry confidence indicators, where the former reflects expectations regarding employment, export, or production and the latter refers to general economic conditions or specific demand decisions in the next 12 months. In addition, Ghonghadze and Lux (2012) deal with economic sentiment indicators based on survey data from the European Commission, focusing on specific sectors such as construction, consumer, manufacturing, retail and services. Furthermore, the paper by Petropoulos and Siakoulis (2021) is particularly relevant for capturing volatility for this research background and deals with the volatility index VIX. This index is based on the fluctuations of the S&P 500 Index, representing the 500 largest listed companies in the U.S., and reflects its expected fluctuation in the coming month (see further Lehrer et al. 2021). In their study, Petropoulos and Siakoulis (2021) also apply this S&P 500 index to operationalise a “crisis event”, which they describe as a scenario that sees it declining by more than 8% over a 3-month period. This definition provides an objective measure of significant market volatility and can thus contribute to its incorporation into quantitative ML predictions.

In relation to collection of relevant determinants of demand, many authors in the literature sample also recommend alternative indicators over survey data, particularly emphasising advanced search engine trends (e.g., Punia and Shankar 2022; Tsao et al. 2022; Tudor 2022). Poza and Monge (2023) suggest the use of search queries, which can show a correlation with key economic indicators to provide predictive insights about future directions and Tudor (2022) further substantiates this, summarising within the present research context that “the inclusion of GT [Google Trend] information offers significant benefits in the form of improved forecasting performance. Nonetheless, GT [Google Trend] data has also been acknowledged in previous research as a leading indicator for key variables of interest.” In light of existing research, Ghoddsi et al. (2019) highlight the potential of internet sentiment for volatility-related predictions and consider ML techniques to be a powerful tool for this purpose. In addition to other studies, Ryu et al. (2020) observe: “Prediction of economic activities by using social network data or internet search data ahead actual activities has been reported in the stock market, marketing and tourism.” For this, the focus is typically on the popularity of specific keywords, reflected in a search engine trend index that indicates the amount of queries in certain geographical areas either in absolute numbers or normalised terms (Punia and Shankar 2022; Tudor 2022). As Punia and Shankar (2022) note, search queries originating not only

from Google but also from other platforms such as YouTube, and reflected in a specific index, can also indicate a positive relationship in demand predictions.

Furthermore, an important observation to mention is that the literature also covers a broad spectrum of other predictive variables for demand. They encompass various dimensions such as consumer behaviour, demographics, weather, advertisement measures, product data, or seasonal and calendar effects.² The price variable is in this context a further critical factor in academic discourse around the investigated research question: “The impact of price changes should not be ignored while designing algorithms for predicting customer choice” (Chen et al. 2021). Beyond its inherent relevance, price can also indirectly reflect macroeconomic conditions like inflationary environments, potentially possessing predictive value for broader economic circumstances. Menhaj and Kavooosi-Kalashami (2022) emphasise the complexity of the price variable in this context, noting that price changes can occur due to macroeconomic conditions but also due to seasonal, cyclical, or trend components. Such price fluctuations may, in turn, lead to varying predictive implications, further highlighting the need for sophisticated understanding and application of the price variable in demand predictions during macroeconomic volatility.

After considering temporal causality and especially leading indicators, issue (2) described by Wang (2022) is relevant, which involves identifying the most effective final set of potential predictors for the model. The reason for focusing on the most important variables is driven, among other reasons, by the need to reduce computational time and complexity of the predictive model (Castillo et al. 2017; Quintero et al. 2022). Such large datasets are also reflected in the literature sample, where notable examples include Martínez-de-Albéniz et al. (2020), who process 1.74 trillion pieces of clickstream data from an online flash sales retailer, and Bi et al. (2022), who tackle 165 million sales transactions. Furthermore, regarding the vast number of potential variables, the study by Wu and Li (2021) initially consider 625 potential input variables from the financial sector for predicting customer churn, and Hasheminejad (2022) starts with 313 input variables.

To address this issue, various methods are applied to find an optimal subset of features that provides the best predictive results for the model. This task is often performed before the actual modelling but is sometimes also integrated into the modelling process itself (Meisenbacher et al. 2022; Quintero et al. 2022). Castillo et al. (2017) differentiate between two primary methods of processing features: feature reduction and extraction. The aim of feature reduction is to reduce the total number of input variables to the most critical ones, while feature extraction transforms the existing features and potentially creates new or adapted ones (Castillo et al. 2017).

² Examples for consumer behaviour include click rate or payment history and for demographics gender or education level (e.g., Ballestar et al. 2019; Chen 2022; Esmeli et al. 2021; Puterman et al. 2020). For weather, examples are temperature or wind speed (e.g., Liu et al. 2021; Punia and Shankar 2022). Advertisement data encompasses campaigns or discounts (e.g., Punia and Shankar 2022; Wang et al. 2019). Examples of product data are article categories and attributes (e.g., Kharfan et al. 2021). Seasonal and calendar effects refer to aspects such as holidays and weekdays (e.g., Albrecht et al. 2021; Ma and Fildes 2020; Raizada and Saini 2021).

For reducing the set of variables to the most relevant demand features, a number of supervised learning procedures from traditional statistical analysis can be identified in the literature sample as being relevant for this purpose. These consist of regression methods ($n=7$), including not only simple linear regression but also other variants of regression, such as the Least Absolute Shrinkage and Selection Operator (LASSO) method (e.g., Meisenbacher et al. 2022; Puterman et al. 2020). Correlation analysis also plays an important role ($n=7$), for example measured by Pearson's correlation coefficient (e.g., Gürses-Tran and Monti 2022; Liu et al. 2021; Shaikh et al. 2022). In addition to purely quantitative approaches, there are also studies employing expert knowledge, like that obtained via surveys or judgments from specialists in relevant academic or professional fields or by researchers themselves ($n=6$), with the aim of selecting important features (Castillo et al. 2017; Gürses-Tran and Monti 2022; Hasheminejad et al. 2022; Khandani et al. 2010; Milošević et al. 2017; Tsao et al. 2022).

For feature extraction, in contrast, unsupervised methods have a more important role, including principal component analysis, examined in multiple studies ($n=7$) within the reviewed literature (e.g., Quintero et al. 2022; Wu and Li 2021), and factor analysis, used in two cases (Poza and Monge 2023; Shaikh et al. 2022). Furthermore, there are also procedures that go beyond purely quantitative analysis. For example, the input data can also be transformed to align with established theoretical constructs or conceptual frameworks, for instance from disciplines like economics or marketing. This can also be observed in studies ($n=5$) in the literature sample, such as by transforming variables into concept versions of the Recency, Frequency and Monetary value (RFM) regarding consumers' demand to model their behaviour (e.g., Chashmi et al. 2021; Wu and Li 2021; Xie 2020).

4.3.2 Modelling

This section examines the modelling phase, focusing on selecting and generating the most efficient models for predicting in a volatile macroeconomic environment. Therefore, the section begins with an investigation of the general relevance and suitability of ML algorithms within this specific research context. This is followed by an analysis of individual models, evaluating their usage over time and identifying the models that show the best results in the literature sample. Additionally, different approaches for effective model selection and training are discussed to account for macroeconomic volatility.

The modelling section is initially intended to investigate the general relevance and suitability of ML for coping with volatility in demand prediction. In the majority of the studies analysed in the literature sample, ML models are found to perform superiorly and are therefore generally considered an important prediction approach in complex macroeconomic contexts. For example, the research by Ma and Fildes (2020) demonstrates that ML modelling (Gradient Boosting Trees) outperforms more traditional methods (e.g., Exponential Smoothing, ARIMA, Naïve, Theta) as it is able to improve the anticipation of non-linear patterns in diversified datasets. The study by Albrecht et al. (2021) also illustrates that ML modelling (Random Forest) has a superior performance compared to traditional forecasting methods (e.g., STL

decomposition with ARIMA or ETS), which the authors largely attribute to its ability to anticipate complex interactions in marketing-related datasets. Another example of the application of ML in dynamic market conditions is provided by the study of Punia and Shankar (2022), who present a combined model of two ML algorithms (LSTM, Random Forest). It shows superior predictive performance for short-, medium-, and long-term demand prediction compared to traditional approaches (e.g., ARIMA, regression) and most effectively incorporates a large number of variables—from promotions to social media and weather conditions to regional economic factors.

However, in some of the studies analysed, typical ML methods also prove not to be superior. Abolghasemi et al. (2020) note: “simple statistical models can outperform some of the sophisticated ML and statistical models.” In the study by Lotfi et al. (2023), the developed fractional calculus model—a generalized diffusion model called GDMR—performs more accurately than typical ML models in capturing the dynamics of repeated purchases in product lifecycles. In the context of the volatility induced by the COVID-19 pandemic, Tudor (2022) examines the demand for video conferencing solutions and finds that the traditional statistical method of exponential smoothing performs best compared to classical ML algorithms (e.g., Artificial Neural Networks). Studies also show that ML alone can be inferior to a combination with other methods. This is illustrated by the example of Purohit et al. (2021), whose agricultural price predictions yield the best results using ML (e.g., Support Vector Machines) in multiplicative combination with statistical time series models (e.g., Exponential Smoothing State Space Model).

Consequently, when faced with volatile market environments and complex and non-linear demand dynamics, conventional prediction techniques can reach their limits and ML solutions provide a valuable contribution. Although there are a significant number of studies in the literature sample that attribute superior performance to ML, there is no general predominance. Others—whether traditional statistical, time series, econometric, or alternative mathematical-quantitative models—can also be superior in this context, necessitating individual testing of the modelling algorithms.

When it comes to selecting a suitable set of algorithms for the prediction problem, there is a wide range of available ML learning models to consider. Regarding the overarching learning style used in this context, a large majority of the studies ($n=54$) use either fully or partially supervised ML for empirical modelling. In contrast, other learning styles, such as unsupervised or semi-supervised learning, are each less commonly used (≤ 5) within the empirical core models. For a detailed overview of the individual models used, in Fig. 7 the number of model usages is shown across the empirical studies over time. Given the limited number of publications before 2019 and after 2022 in the literature sample, with less than three per year, this analysis is focused on this period, where important developments are evident. Furthermore, beyond identifying the models commonly employed in studies, it is essential to highlight those that exhibit the best overall results and hence are recommended for continued application in this context. To facilitate this, in Fig. 8, a graphical representation is presented of the frequency at which specific main ML algorithms are recognised by researchers in the empirical studies for their superior performance. In Fig. 8, a distinction is highlighted in terms of the type of prediction

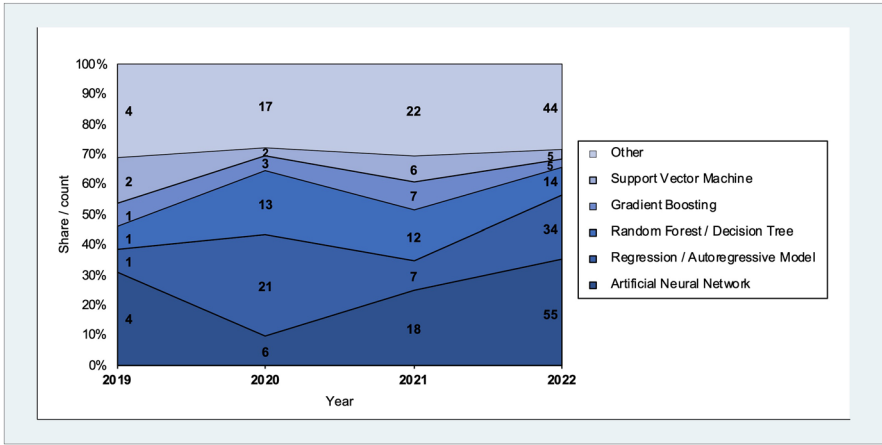


Fig. 7 Timeline of model usages

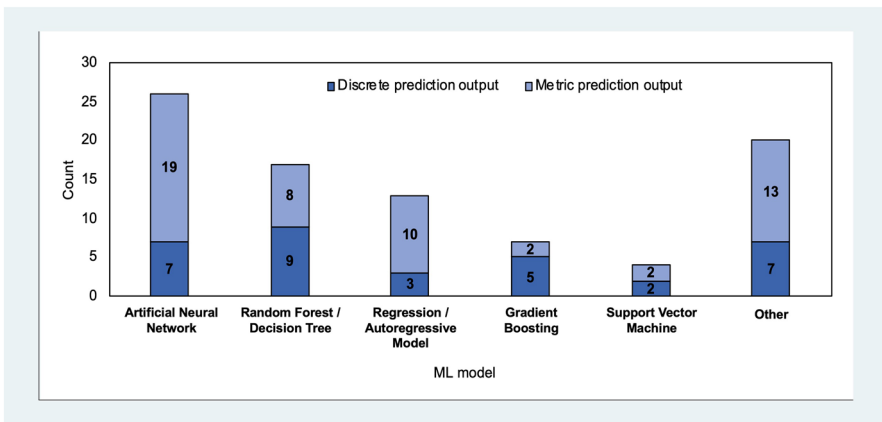


Fig. 8 Preferred model by prediction output type. Note on Figs. 7 and 8: A study can both use and recommend multiple models—each model is then counted individually. Studies that do not have the conceptual aim of using or evaluating models, such as reviews, are not considered in the figures. A model is assigned to the “other” group if its usage count is fewer than 10; if the usage count is higher, the model name is explicitly indicated. Figure 7 is based on studies from 2019–2022. Figure 8 has no time period limit

output, since the selection of ML algorithm can depend on this factor, indicating overall that metric prediction results are generally considered more frequently than discrete ones.

Additionally, it is important to note that Figs. 7 and 8 present different models distinctly, even if they are part of a combined approach where multiple algorithms are used synergistically to optimise the overall predictive outcome (Ghoddusi et al. 2019; Kharfan et al. 2021; Wang et al. 2019). A primary objective behind this approach is to mitigate the influence of occasional suboptimal demand predictions.

By integrating multiple models, especially in macroeconomic environments characterised by pronounced volatility or outlier-rich data scenarios, researchers aim to achieve improved stability and resilience for predictions compared to those delivered by standalone models (Esmeli et al. 2021; Kharfan et al. 2021). The observation of Kharfan et al. (2021) that “combining different techniques tend to perform better than a single method” resonates with broader scientific research, as an analysis of the literature sample demonstrates that in around half of the studies best performance is achieved by combination of multiple models (e.g., Ahmed et al. 2022; Bukhari et al. 2020; Punia and Shankar 2022; Purohit et al. 2021; Yang and Chang 2020).

As illustrated in Fig. 7, Artificial Neural Networks are identified as the predominant model regarding usages. A strong increase in their utilisation since 2020 is observable, with a remarkable usage level ($n=55$) in 2022. Among the different outcome types analysed, Artificial Neural Networks are consistently selected as the preferred ML algorithm by the researchers in the studies. Specifically, they exhibit particularly pronounced efficacy for problems with metric prediction output—including time series forecasting—as illustrated in Fig. 8, with them being preferred 19 times. One fundamental reason for this preference might be that Artificial Neural Networks, categorised as non-parametric ML models, are capable of generating non-linear and complex relationships (Tudor 2022). Further endorsing their adaptability, Liu et al. (2021) emphasise the diverse range of network architectures that have emerged over time, each tailored to address specific challenges or to mitigate limitations of existing architectures. To this end, the pie chart in Fig. 9 provides deeper insight into which specific architectures were used for the Artificial Neural Networks, as highlighted in the studies as the preferred model. Considering the various possible model architectures of Artificial Neural Networks, this visualisation displays only those demonstrating superior performance in the empirical studies. The original architecture of Artificial Neural Networks, Multilayer Perceptron, is 8 times recommended as being best performing. Moreover, Recurrent Neural Networks, predominantly based on the Long Short-Term Memory variant, emerge as the foremost high-performing Artificial Neural Networks in the literature, being 11 times preferred (e.g., Liu et al. 2021; Pessanha and Soares 2021; Punia and Shankar 2022). Their dominance can be attributed to their capacity to effectively process time-dependent relationships, a characteristic enhanced by the design that facilitates sequential feedback between inputs and outputs (Yang and Chang 2020; Gharibshah and Zhu 2021). Convolutional Neural Networks are highlighted 3 times as being particularly appropriate since they demonstrate adeptness in handling a wide range of different input features (Yang and Chang 2020). Other architectural variants, either infrequently employed or without explicit categorisation in the papers, are consolidated in this review under the label “Others/not specified”. This includes Adaptive Network-based Fuzzy Inference System (Hasheminejad et al. 2022), Extrem Learning Machine (Kmieciak and Zangana 2022), Conditional Generative Adversarial Networks (Lin et al. 2022), Recursive Neural Networks (Chashmi et al. 2021), or Autoencoder Neural Networks (Quintero et al. 2022).

Following Artificial Neural Networks, Random Forests and Decision Trees are the preferred algorithms in 17 cases. The split between discrete and metric

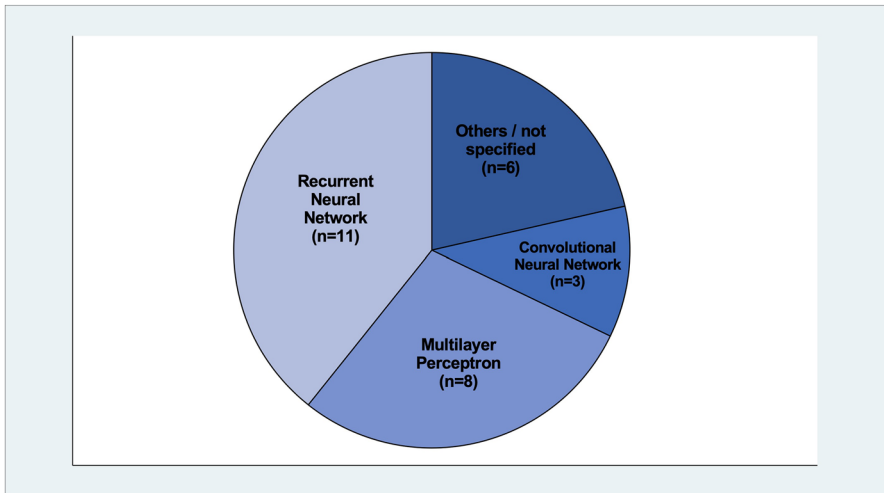


Fig. 9 Architectures in studies with Artificial Neural Networks as the preferred model

prediction output is in this case more balanced than for Artificial Neural Networks. In absolute numbers, this algorithm shows in the studies the highest rate of recommendations based on superior performance for discrete prediction output ($n=9$) (e.g., Bohanec et al. 2017; Esmeli et al. 2021; Puterman et al. 2020)—even ahead of Artificial Neural Networks. A Decision Tree can be characterised as a binary tree-like flowchart with internal nodes that continuously branch into two-child nodes (Wang et al. 2019). Building on this, a Random Forest is an algorithm where multiple Decision Trees are constructed using training samples drawn through bootstrap sampling, aptly named for the ensemble of trees it entails (Bohanec et al. 2017).

In addition, Regression and Autoregressive Models are recommended 13 times in the literature sample due to their superior performance (e.g., Abolghasemi et al. 2020; Ryu et al. 2020). This suggests that even those models that are often considered as being more basic are able to achieve remarkable results (Abolghasemi et al. 2020). For example, models in the category of Regression and Autoregressive Models show the highest usage count with 21 times in the year 2020—a year marked by high macroeconomic volatility due to the first pandemic outbreak of COVID-19 in many countries (Kmieciak and Zangana 2022). One reason cited by researchers in the literature sample for this is the benefit of starting with simpler linear or univariate methods in order to utilise them for initial insights as a general benchmark, which must then be at least surpassed by more complex ML methods (see further Albrecht et al. 2021; Ma and Fildes 2020; Milošević et al. 2017). The considerable share of the usage of these models from 2020 onwards, such as different versions of the autoregressive model ARIMA for time series forecasting, can be understood as an indicator of an evolving modelling landscape tending towards more predictive robustness. Due to the fact that under volatile macroeconomic conditions certain predictors may have behaved differently in the past and can hence provide misleading information for the future, models like univariate ones could provide stable

guidance (Ahmed et al. 2022; Gürses-Tran and Monti 2022; Hasheminejad et al. 2022). However, despite this, they may struggle to simultaneously establish covariate-based as well as (non-linear) temporal dependencies when predicting demand patterns, the usually lower computational requirements, popularity, ease of use, and fast development phase should be seen together with the above as further reasons (Alsahref et al. 2022; Makridakis et al. 2022). As suggested by Meisenbacher et al. (2022), such simpler models could therefore enrich the methodological diversity to validate advanced ML models for comparison purposes and contribute with their demand predictions to the overall robustness of the results.

Furthermore, Gradient Boosting is recognised as a significant category of algorithms belonging to the broader concept of combined methods, which involve a specific synergistic integration of multiple models, leading to so-called ensemble models. These are meta-algorithms that employ strategies aimed at improving model performance, including reducing bias and variance, through techniques such as boosting (Geiler et al. 2022). Since 2020, Gradient Boosting algorithms are used to a relevant extent in the literature sample, potentially owing to their relatively recent advancements (e.g., Ghodduzi et al. 2019; Wang et al. 2019). Moreover, Gradient Boosting is recommended by researchers in the literature review in multiple instances as the preferred model ($n=7$). Geiler et al. (2022) highlight one particular algorithm, stating, “the most widely used implementation for boosting is XGBoost, a computationally efficient implementation of stochastic gradient boosting.” Their perspective finds validation in the literature sample, underscored by a number of studies that mention the superiority of using this specific algorithm (Contreras-Masse et al. 2022; Geiler et al. 2022; Petropoulos and Siakoulis 2021; Wu and Li 2021).

The remaining categories are Support Vector Machines and “Other”. Support Vector Machines are a kernel-based ML model designed to build an optimal separating hyperplane (Geiler et al. 2022), and are used at a similar level as Gradient Boosting (Fig. 7)—but at a much lower level identified as preferred model (Fig. 8). “Other” includes a broad spectrum, for example unsupervised ML methods such as K-Means Clustering, which are regularly used in combination with supervised ML models.

After the discussion of the generally used and the specifically recommended models, the next step is to identify from this selection those models best suited to the specific problem and to finally conduct modelling with them for further marketing application purposes. In the reviewed studies, it is generally suggested that, even within this particular research context, the majority of studies follow an already widely accepted procedure in ML training.³ However, the literature sample also

³ After the selection of the model, the literature sample suggests that, even in this specific context, most studies adhere to a accepted process. Hence, the practical implementation of the model is mainly performed using programming languages such as Python or R (Claveria et al. 2020; Kmiecik and Zangana 2022). In addition, the dataset is partitioned to provide different subsets for training and testing the model (Kozak et al. 2021; Raizada and Saini 2021). A common approach is that the training data is further subdivided in order to adjust the hyperparameters—so-called validation—which is a process to improve those model parameter settings that remain unchanged during the training (Alsahref et al. 2022).

reveals several specialised approaches that can be particularly useful for tailoring the ML modelling process to the specific requirements of demand prediction in a macroeconomically volatile environment. For instance, Abolghasemi et al. (2020) provide guidance appropriate for the research purpose of predicting the behaviour of demanders under macroeconomic volatility. They recommend utilising the Coefficient of Variation as a key measure to quantify volatility and as a valid measure that illustrates uncertainties in the demand data in order to determine the most appropriate prediction models based on it. Consequently, the Coefficient of Variation can facilitate a comparative analysis of relative variations between different demand series. Since the Coefficient of Variation is calculated by dividing the standard deviation of demand by its mean, the resulting ratio is also scale-independent. This approach might prove especially advantageous in scenarios where multiple data series are prevalent, such as in marketing where demand prediction for various products is common. The authors suggest that here model selection should be based on the respective level of volatility: “demands with different CoV [Coefficients of Variation] require different types of a forecasting model that suits their characteristics” (Abolghasemi et al. 2020). In terms of practical application details, they emphasise that there is not yet a general consensus on thresholds that would imply the applicability of specific categories for model selection with the Coefficient of Variation. For this reason, Abolghasemi et al. (2020) propose a three-tier classification of volatility, suggesting a Coefficient of Variation of <0.5 as low volatility, of $0.5\text{--}1.0$ as moderate volatility, and of >1 as high volatility and recommend distinct prediction models for each of them. These thresholds of Abolghasemi et al. (2020) can offer initial guidance but, however, also require further empirical validation. The way of proceeding with the Coefficient of Variation is also considered by other authors such as Makridakis et al. (2022), who refer to Syntetos and Boylan (2005) and Syntetos et al. (2010), where the Coefficient of Variation is used to classify data regarding volatility for guiding model selection. Consequently, the Coefficient of Variation metric can be identified in the literature sample as a tool to effectively facilitate model selection amidst shifting macroeconomic conditions in marketing.

In terms of handling modelling data, Ma and Fildes (2020) propose a strategy of not only determining the best model parameters during training but also learning the optimal window width of the data to algorithmically select the most appropriate range from the dataset that effectively captures the local dynamics. In this way, volatile demand trends may be captured more accurately by using windows to focus on a time period that best represents the anticipated scenario and neglecting disparate periods that could cause misleading predictions. This addresses the challenge highlighted by Khandani et al. (2010) of models and measures that are not sufficiently adaptive and therefore “adjust only slowly over time and are relatively insensitive to changes in market conditions.” This is in sharp contrast to the rapid shifts seen in actual demanders’ behaviour triggered by fluctuating macroeconomic conditions, according to the authors. Yang and Chang (2020) further enrich the research discourse around this topic and emphasise the importance of continuously updating models with the most recent data. Hence, they suggest the use of a rolling prediction technique and explain that “in this scheme, a fixed window was used, and the value was updated in the fixed window with each newly predicted value. This process

involved removing historical data and adding future data so that the fixed window always retained the same amount of time-series data” (Yang and Chang 2020).

Another approach which is highlighted for model training is the concept of “pooling.” Pooling is a strategy of leveraging multiple analogue data series as a collective group to enable ML models to effectively learn common patterns (Frees and Miller 2004; Ma and Fildes 2020). Building on the foundational research of Duncan et al. (2001), Ma and Fildes (2020) underscore the potential relevance of it and explain to “use pooling where there is high volatility or outliers.” They argue that this approach facilitates more robust predictions by reducing the variance of the model compared to individual predictions, allowing for more effective adaptation to structural changes (also Menhaj and Kavooosi-Kalashami 2022). This method is therefore relevant to test in the present context for model training, but it should also be further validated for its applicability, for instance with very different marketing data or in regard to the risk of overfitting.

A final approach to handling the complex problem of volatile demand series during model training involves decomposition, as recommended by Abolghasemi et al. (2020). In environments characterized by volatility and subject to multiple macroeconomic influences, demand series in marketing can exhibit fluctuating patterns. Abolghasemi et al. (2020) propose dividing volatile demand into two major, manageable components: a baseline demand component, which reflects a rather regular pattern, and a more volatile demand component, which may be influenced by marketing promotions or other factors such as market shifts and, consequently, shows more volatile patterns. Each of these components can then be modelled independently and predicted based on their specific volatility degrees. To illustrate the practical application, the authors note: “we develop a hybrid model that first decomposes demand into main parts and forecast them separately, and then combines them to forecast the demand” (Abolghasemi et al. 2020).

4.3.3 Postprocessing

The postprocessing stage centres around the final model evaluation—such as in marketing business practice—showcasing from the literature sample two approaches: a narrower and a broader one. Furthermore, in the following section, a spotlight is also placed on the limitations and challenges inherent to ML models, ranging from data inconsistencies and generalisability boundaries to interpretability and transitioning from theory to real-world applications. Given the background of predicting demand in volatile economic environments, this section involves the accentuation of the importance of postprocessing activities in ensuring that the derived implications are both substantiated and relevant in a macroeconomic volatile environment.

An initial task that occurs at this stage is the model evaluation. Two distinct perspectives regarding model evaluation can be identified through the analysis of the review sample. The first perspective is a narrower one, predominantly bases the overall judgement of the model on a quantitative assessment of its predictive performance (“narrower statistically-oriented evaluation”). This typically involves the use of benchmark methods to appraise model performance and a comparison is made regarding their prediction errors through statistical testing methodologies to uncover

any significant differences between models. This approach also provides a potential means for recognising and accommodating uncertainties within data conditions in the research context. The second perspective is a broader one, extending beyond a purely quantitative assessment to also include the integration of a business-oriented viewpoint (“broader business-oriented evaluation”). This perspective accentuates the business value generated by the demand predictions and involves consideration of its effect on organisational operations and resources. This perspective specifically aligns with the practical marketing context of the research and thus emphasises the wider implications associated with the consequences of predictions (Gürses-Tran and Monti 2022; Kmiecik and Zangana 2022).

Considering the first approach, which is a narrower statistically-oriented evaluation, the emphasis lies on measuring predictive performance through statistical error. As outlined by Alsharif et al. (2022), the objective here is “to find models with the highest accuracy when trained and tested on the dataset.” To quantitatively measure the predictive performance, various error metrics are employed in the literature sample. All of those involve a comparison of the actual value y_i with the predicted value \hat{y}_i (e.g., Ghodusi et al. 2019). Gürses-Tran and Monti (2022), referring to the foundational work of Hyndman and Athanasopoulos (2018), highlight the relevance of choosing appropriate error metrics, declaring that “a robust and reliable forecast model can only be achieved when the metrics for scoring and selection are distinct.” Punia and Shankar (2022) add that relying on a single error metric might not provide comprehensive information for model evaluation—potentially lacking in particular under conditions of macroeconomic volatility. This opinion is broadly supported in the literature sample, with a predominant majority of studies incorporating multiple metrics, averaging 3.1 per study (see Fig. 10). Indeed, only in a minority of the studies ($n=11$) is the predictive performance evaluated using just one error metric. Endorsing the importance of different metrics, Ma and Fildes (2020) recommend a set of error metrics to fulfil a range of requirements, such as robustness to potential outliers, ease of interpretation, and scale-independent metric outcomes.

The narrower statistically-oriented evaluation of tasks with metric prediction output commonly involves error metrics averaging the difference across all predictions, potentially squaring or rooting them, and expressing the results as absolute or percentage values. As indicated in Fig. 11, the three most frequently cited metric within the reviewed literature are the Root Mean Square Error (RMSE = $\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}}$; $n=20$), the Mean Absolute Error (MAE = $\frac{\sum|y_i - \hat{y}_i|}{n}$; $n=16$) and the Mean Absolute Percentage Error (MAPE = $(\frac{100}{n}) * \sum|(\frac{y_i - \hat{y}_i}{y_i})|$; $n=11$). In terms of metric prediction outputs and the associated error metrics, Castillo et al. (2017) provide relevant insights in their empirical study on predicting sales of newly published books, underscoring that the impact of predictive error can vary depending on its direction. In their case, overpredictions lead to increased inventory costs and financial losses, while underpredictions solely result in another print run, suggesting that in this case, the economic impact of overpredictions significantly outweighs the cost of underpredictions. Consequently, when selecting error metrics, consideration of the specific circumstances of the demand prediction problem is relevant because, as in this

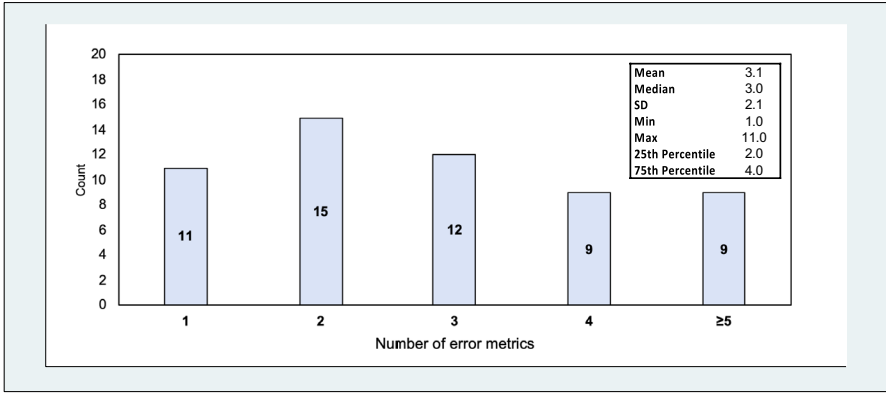


Fig. 10 Number of error metrics. Note: Evaluation of the datasets in the empirical studies

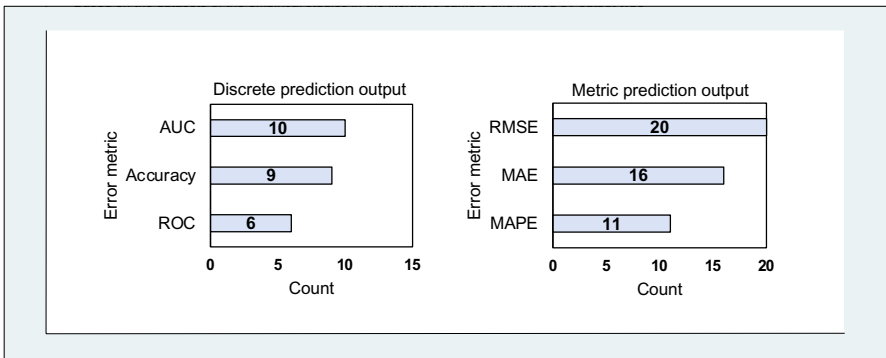


Fig. 11 Most frequently used error metrics. Note: Evaluated by output type and based on the empirical studies

example, commonly used metrics may not adequately account for such domain-specific implications of the prediction error.

For tasks with discrete prediction output, the ROC (Receiver Operating Characteristic) curve and the Area Under this Curve (AUC) have an important role in the literature sample. The ROC curve (n=6) plots the rate of true positives against the rate of false positives and then the AUC (n=10) shows the general performance of the ML model with regard to distinguishing between positive and negative classifications (Chen et al. 2021; Geiler et al. 2022). Regarding the fourth and fifth most frequent metrics for this output type, which are precision (true positive predictions among all positive predictions) and recall (true positive predictions among all actual positive predictions), the researchers Kozak et al. (2021) discuss further implications of their use. Reflecting on the balance between these two error metrics in response to changing economic conditions, Kozak et al. (2021) explain: “Considering the volatility of global value chains, [...], the need for flexible marketing planning is

stronger than ever [...] [requiring] to flexibly react to fast-changing conditions, e.g., prioritizing recall in churn targeting when the economic climate improves and prioritizing precision in difficult times.”

When interpreting these error metrics, it is suggested in the literature sample to constantly take into account the demand prediction horizon and its associated uncertainties, especially when dealing with volatile conditions. For example, from research conducted by Sorjamaa et al. (2007), it is implied that there is a need to evaluate prediction performance in relation to the prediction horizon. As cited by Ma and Fildes (2020), their work illustrates that demand predictions intended to provide an outlook for multiple future periods (multilevel predictions) pose a higher level of complexity than single-period demand predictions due to the accumulation of errors, which in turn increase uncertainty. This relationship between the prediction horizon, predictive performance and uncertainty is documented by various authors, including Kmiecik and Zangana (2022) or Wang et al. (2019). Additionally, Albrecht et al. (2021) note that “overall, the models’ performances worsen slightly with increasing lead time.” Given this complexity, Abolghasemi et al. (2020) argue for the definition of prediction intervals, which specify an upper and lower bound for the predictions and represent the range into which projected demand is likely to fall at a certain confidence level. They make the argument, that such intervals can enrich management understanding, stating: “Prediction intervals will provide managers with insights into the most appropriate choice of forecasting methods when the degree of uncertainty is taken into account” (Abolghasemi et al. 2020).

Regarding the second approach, the broader business-oriented evaluation, there is an emphasis on the importance of looking beyond the mere measurement of the performance of demand predictions. The contrast is highlighted by Gürses-Tran and Monti (2022), who distinguish between “forecast evaluation based on established forecast error statistics on one side, and the economic valuation of the applied forecast on the other side.” The necessity for those broader considerations is highlighted by Kmiecik and Zangana (2022), who argue that “while comparing the ex-post errors of different forecasts is a method of determining which is superior, it cannot alone justify the value it adds, it is rather a relative measurement.” Therefore, the authors emphasise the necessity of considering the added business value resulting from any improvement of demand predictions, which they describe as “forecast value added.” Within this assessment, they acknowledge various expenses related to model deployment, including financial, temporal and human resources (Kmiecik and Zangana 2022). While the predominant focus in the reviewed studies is on quantitative statistical error metrics, some studies are identified that offer such a more holistic assessment, integrating broader economic and business dimensions for a comprehensive model assessment. For example, Glaeser et al. (2019) look beyond traditional error metrics and evaluate their models based on its potential financial impact on the business. For this reason, they run simulations to compare the potential revenue impacts of using their algorithm to optimise retail locations. Similarly, Khandani et al. (2010) explore the practical implications of enhancing their credit risk prediction model, expecting a 6 to 25% reduction in total losses by identifying high-risk demanders earlier. In addition, Milošević et al. (2017) investigate the potential profit gain associated with their ML churn prevention approach, which

promises to increase profit twofold by providing a more accurate identification of demanders likely to churn.

An essential final step in the postprocessing phase arises from the general academic need to identify and understand the limitations associated with ML predictions for demand behaviour under volatile macroeconomic conditions. In order to provide a concise systematisation of these issues and restrictions, they are extracted from the 64 reviewed studies and systematically organised into four main areas of limitations: (1) data, (2) generalisability, (3) interpretability and (4) methodological and application limitations.

An important number of the identified limitations relate to the (1) data dimension, identified as a relevant constraint in a majority of the studies. These limitations include data inconsistencies such as missing data, omitted variables or unbalanced data due to the origin of the dataset or the way of data collection, where even small deviations in data quality can significantly affect model outcomes, as several authors in the sample indicate (Ghoddusi et al. 2019; Liu et al. 2021). Moreover, this category encompasses issues related to a limited sample size (e.g., Anand and Mishra 2022; Esmeli et al. 2021; Lin et al. 2022) or a restricted number of input variables (e.g., Khandani et al. 2010; Kozak et al. 2021; Raizada and Saini 2021), which can both affect the robustness and predictability of a model. Particularly in this research context, involving a rapidly changing macroeconomic environment where data may not fully reflect evolving demand structures, a regularly cited concern is the use of ML with data from narrowly limited time periods or relying excessively on historical demand data, which can burden the anticipation of more recent turbulent events. As Castillo et al. (2017) emphasise “the efficiency of these techniques strongly depends on the field of application and the correctness of the problem data.” Therefore, in volatile periods, a potential requirement resulting from this limitation is the necessity to incorporate sufficient demand data from such transformative or turbulent phases into model training to provide an accurate data foundation for generating reliable demand predictions about the future development in these times. This is addressed, for example, by Tudor (2022), who specifically splits the data, thereby allowing to incorporate data from the pandemic into the training datasets, enabling the models to adapt to the volatile period.

The next category, (2) generalisability, refers to limitations in validating the proposed models for a broader spectrum of demand prediction scenarios, research problems or domain areas that may arise from a very specific research perspective in the study (Kozak et al. 2021). For example, several authors limit their study results to a specific industry, such as retail (e.g., Chashmi et al. 2021; Glaeser et al. 2019; Martínez-de-Albéniz et al. 2020). Others limit their focus to a particular country (e.g., Castillo et al. 2017; Menhaj and Kavoosi-Kalashami 2022; Wu et al. 2022), a prediction item (e.g., Punia and Shankar 2022) or a particular company (Bi et al. 2022; Raizada and Saini 2021). A consequence of this is the general necessity to evaluate the broader applicability of specific research results in relation to the prevailing framework and the practical marketing prediction scenario. Furthermore, in the ML context, the scope for generalisability also includes the fact that not all existing algorithms—especially those that are particularly complex or recently developed—can be taken into account in the modelling or preprocessing stages. Factors

contributing to the limited number of ML models considered include not only the necessity for extensive computing time and resources (Chen et al. 2021) but also the frequent emergence of new models in this field. This is evidenced by the literature sample, as several studies introduce new models, such as the PLR-ALSTM-NN (Poly-linear regression with Long Short-Term Memory) by Ahmed et al. (2022), the ATLAS (Advanced Temporal Latent-factor Approach to Sales forecasting) by Bi et al. (2022) or the EHTS model (Ensemble of Boosted Hybrid of Deep Learning Models and Technical Analysis for Forecasting Stock Prices) by Kamara et al. (2022).

A further category addresses the limited (3) interpretability regarding the model or the relationships it establishes, which is addressed in a smaller but relevant number of studies. Consequently, predicting demand behaviour in a volatile macroeconomic environment involves looking not only at the predictive capabilities of ML but also at the requirements of stakeholders in terms of its understandability, explainability and transparency. As Gürses-Tran and Monti (2022) observe that “more complex models emerge that are less transparent or understandable for the decision makers. However [...] when forecast models lack interpretability [...], the trust and confidence of stakeholders can suffer when making decisions.” Kozak et al. (2021) echo this by stating: “If decision-makers are not able to interpret data properly or are not able to prepare decisions, big data analytics provides only little value.” From a methodological perspective, specific approaches to gain insights into the demand model can be found in the sample, like feature importance assessment, which estimates the contributions of individual features to improve the predictive model (Ma and Fildes 2020). Another technique is sensitivity analysis, where the changes in the output model are observed when specific inputs are adjusted (Bohanec et al. 2017). To this end, Bohanec et al. (2017) warn that “the explanations closely follow the prediction model; if the model is wrong or performs poorly, the explanations will reflect that.” Study examples for such evaluations can be found in Claveria et al. (2020), who analyse the impact of variables on their GDP growth prediction model and find that underestimated yet crucial survey variables are currently not included in official statistics. Another example can be found in Chen et al. (2021), who use ML in purchasing behaviour research and model customer preferences when considering the influence of sales prices. After modelling, the authors extract the price sensitivity for certain categories and find that products they see as essential to life, like eggs, milk and bread, indicate a low price sensitivity in the model. In contrast, categories such as canned food or ready-to-eat meals, which have a longer shelf life and can be easily stored, show a higher model price sensitivity. They claim that these insights can guide tailored promotions based on consumer price behaviour at the category or product level.

The final limitations encountered in the transition from study results to (4) real-world applications as well as notable methodological limitations. Regarding the first aspect, the contribution of Ballestar et al. (2019) from the literature sample, with the acknowledgement of the study of Lamberton and Stephen (2016), can be used to help characterise this problem. They critically discuss the “myopic” approach common in scientific analyses, which often tends to narrow

the actual scope of a much larger and multi-layered real-world problem to just a few facets. This can result, for example, from the inability of certain models to include all relevant factors that might influence outcomes in real-world scenarios or from the difficulty of measuring these effects—ranging from broader external variables such as certain prevailing macroeconomic conditions to more detailed information on consumers' demand behaviour (e.g., Bi et al. 2022; Khandani et al. 2010; Hasheminejad et al. 2022). However, Anand and Mishra (2022) also emphasise the inherent complexity of real-world applications and, in particular, the intricacies that arise from human behaviour, describing it as “nonlinear, non-parametric, irrational, and time-variant.” Furthermore, the necessity of continued practical verification is expressed as a concern regarding real-world operational contexts. As formulated by Kozak et al. (2021) “most of the decision-making techniques use deterministic machine learning (ML) techniques but unfortunately, they do not take into account the variety and volatility of decision-making situations and do not allow for a more flexible approach, i.e., adjusted to changing environmental conditions or changing management priorities.” Unexpected macroeconomic volatility can exacerbate this challenge, as, for example, the development of certain demand influencing variables might not be anticipated in advance at the time when the demand prediction is generated (Castillo et al. 2017). Apart from the issue of real-world applications, the second aspect to be addressed is that of methodological limitations, which cover a wide variety of specific problems. These can be, for example, methodological shortcomings on a technical level regarding the way the model and its parameters are generated in the specific context of ML—which can be the problem of overfitting the model, the partly manual search for model settings (hyperparameters) or the overall strong reliance on the principle of trial-and-error. All of these can significantly complicate the task of placing the ML model and the relationships it contains on a solid theoretical foundation (Claveria et al. 2020; Ghoddusi et al. 2019). Coveney et al. (2016) underscore this, stating: “We argue that it is vital to use theory as a guide to experimental design for maximal efficiency of data collection and to produce reliable predictive models and conceptual knowledge,” and Bi et al. (2022) conclude: “Incorporating economic theory considerations into machine learning models can provide significant additional advantages, and, thus, constitutes a promising direction for future work.”

4.4 Limitations

Also the research conducted in this study to develop the systematic literature review is bound by certain limitations. Outlining them helps to contextualise this study's outcome and also offers an opportunity to build on this foundation in further research. Firstly, a limitation arises from the selection of the Web of Science-database for this literature review, based on the research study of Gusenbauer and Haddaway (2019), as its scope is restricted to the journals indexed therein. This implies that formats outside journal publications, such as conference proceedings or published books (Kepes et al. 2012; Vlačić et al. 2021), as well as non-academic

journals or grey literature (Rejeb et al. 2020), are not specifically covered. Therefore, as the literature sample is focused on journal publications to ensure comparability of the data through strict journal review processes, there is a risk of excluding some current research findings. Furthermore, Ghoddsi et al. (2019) highlight an unavoidable but nevertheless notable restriction for interdisciplinary research involving computer science by pointing to unpublished developments of ML algorithms that are only used for commercial business purposes and are therefore not accessible via studies. In general, however, the interdisciplinary research orientation in this literature review allows for the formation of an overview across disciplines that goes beyond the mainstream of a particular field and that provides a differentiated and more holistic new perspective (Ghoddsi et al. 2019; Rejeb et al. 2020).

As a second limitation of this paper, it should be mentioned that the precisely defined search string not only involves the risk of false positives, thus including non-relevant studies (Linnenluecke et al. 2019; Xiao and Watson 2019), but also a risk of false negatives through excluding papers that fall outside the search string but that may provide relevant knowledge (Durugbo and Al-Balushi 2022). Another consequence of the search string used is that the identified studies largely focus on specific industries such as consumer goods and retail ($n=22$) as well as economics and financial management ($n=13$). Also, the inclusion of only English language studies could ignore relevant findings from non-English-language studies, as the analysis in this paper highlights relevant geographical research centres, for example in Asia ($n=100$) or in large parts of Europe ($n=86$), that are non-native English speaking.

Third, the definition of the inclusion and exclusion criteria still allows researchers a certain leeway, which can affect interrater reliability (Clark et al. 2021; Xiao and Watson 2019). However, the review protocol with an objective list of criteria clearly limits individual subjectivity and thus ensures the general replicability of the previously presented research findings (Vlačić et al. 2021).

5 Conclusion

This systematic literature review summarises recent research evidence on the effective application of ML in times of macroeconomic volatility, particularly focusing on demand prediction. The need to synthesise this specific research knowledge is motivated by the current period, marked by pronounced global uncertainty and turbulence (see Durst et al. 2022; Durugbo and Al-Balushi 2022). Here, accurate demand prediction is crucial for organisational functions like marketing (Kozak et al. 2021; Seyedan and Mafakheri 2020). Therefore, this thematic area is assessed by analysing a sample of 64 journal articles from an initial pool of 2877 papers, which were selected using a catalogue of criteria specifically developed as part of the review protocol (Xiao and Watson 2019). Using an integrative approach with a hybrid methodology (Trapp 2012), relevant knowledge is synthesised from the final sample (Sect. 4), thereby contributing to closing the gap identified in previous literature reviews. This systematic literature review thus contributes to the ongoing academic discourse on the effective use of ML in application practice, integrating

multidisciplinary insights from various streams within computer science and management science.

At the beginning of this systematic literature review, an analysis of the characteristics and abstracts of the articles (Sects. 4.1 and 4.2) reveals that literature since 2010 approximates an exponential trend curve in publications, which demonstrates evidence of a growing scientific interest in this interdisciplinary research topic. A geographical concentration of academic discussion is found, with Asia accounting for the largest share of publications, followed by Europe and North America, with the U.S. and China as the leading countries. The journal “Expert Systems with Applications” and authors such as Hyndman and Breiman deserve special mention, as they are identified as essential references in this complex field.

In the further course, this review paper also provides a more detailed perspective with a synthesis of the full texts, chronologically outlining the ML application process (Sect. 4.3). The analyses in the preprocessing phase (Sect. 4.3.1) indicate that over 75% of the empirical studies involve the use of multivariate models, which emphasises the need to consider multiple factors to capture the multifactorial nature of demand behaviour in a volatile environment. Additionally, forward-looking variables are identified as important input variables for demand predictions, offering early signals for macroeconomic trends and turning points. Examples cited include market participants’ expectations and sentiments via economic indicators or search trends. In addition, the research regarding the modelling phase (Sect. 4.3.2) initially reveals that a majority of the studies illustrate an outperformance of ML over other prediction methods (e.g., Ma and Fildes 2020; Albrecht et al. 2021). This implies the general suitability of ML for modelling complex demand patterns and interdependencies in dynamic market conditions. However, the research also identifies studies where alternative quantitative models surpass ML (e.g., Abolghasemi et al. 2020; Lotfi et al. 2023), indicating that ML models do not possess universal superiority in this context. Furthermore, the analysis delivers an assessment of the usage frequency of different models in the empirical studies, as well as the frequency with which models are explicitly recommended by researchers due to their superior predictive performance. There, it becomes evident that supervised ML algorithms are prevalent in approximately 90% of the studies. Furthermore, around half of the research papers demonstrate the outperformance of combining multiple models via hybrid or ensemble modelling, underscoring their significance for resilient demand prediction in complex data environments. The analysis also highlights specific ML algorithms, recommending their inclusion for modelling in this context. Around 40% of the studies report that Artificial Neural Networks, either fully or partially, contribute to the best-performing model, with Recurrent Neural Networks particularly standing out for their ability to capture time-dependent and non-linear relationships. The analysis identifies further key ML algorithms in this context—specifically Random Forests/Decision Trees, and Gradient Boosting—as particularly effective, each demonstrating superior performance in over 10% of the studies. Additional modelling strategies to account for demand volatility are also discussed, such as employing the Coefficient of Variation, demand decomposition, pooling, or specific time windows. Finally, in the postprocessing phase (Sect. 4.3.3), evaluation

strategies are identified that account not only for statistical performance but also for the broader business-oriented implications of ML demand predictions. Moreover, from the 64 studies, four main types of limitations are synthesised to consider when implementing ML demand prediction under macroeconomic volatility: (1) data, (2) generalisability, (3) interpretability, as well as (4) methodological and application-related limitations.

Overall, based on the research findings, an evidence-based guideline can be provided to reinforce business areas such as marketing under the prevailing conditions. Future research directions can be derived from the conceptualisation of the field as presented in this systematic literature review (Sect. 4.2). As determined in this paper by a correspondence analysis of the extracted abstract terms, the scientific field can be categorised into five focus research areas: (F₁) ML methods and their application context, (F₂) Economic factors, (F₃) Time series forecasting, (F₄) Neural network predictions, and (F₅) Customer frameworks. Within these areas, further in-depth investigation is encouraged. Moreover, additional conceptual work is important, such as developing a structured framework to provide state-of-the-art directives for applying ML in demand prediction (Alsharif et al. 2022; Durugbo and Al-Balushi 2022; Ghoddusi et al. 2019). Research opportunities also stem from expanding upon the key insights through further empirical investigations.

In a broader sense, this paper encourages the fostering of more integrated approaches that harmonise rigorous algorithmic methods with practical business objectives. It emphasises the need for a particularly holistic perspective to consider interdisciplinary interactions between research areas, as seen, for example, especially between the disciplines of computer science and management science. The comprehensive overview resulting from this study is intended to provide a platform for future developments in this rapidly evolving field. By introducing a new view and a deeper systematisation of the research area, this systematic literature review can act as a robust scientific foundation for further academic and practical endeavours.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11301-024-00447-8>.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability As a supporting document, an Excel file with a detailed overview of all papers incorporated in the literature sample is provided with the manuscript submission. This file complements the information provided in this systematic literature review, which already includes specifics regarding the data collection, such as the literature sources, search string, and selection criteria. For additional analyses and information regarding the research paper, the corresponding author can be contacted. We ensure that all data supporting this study's results are available in the article, supplementary documents, or through the corresponding author, upholding transparency and full accessibility to the scientific community.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abolghasemi M, Beh E, Tarr G, Gerlach R (2020) Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. *Comput Ind Eng* 142:106380. <https://doi.org/10.1016/j.cie.2020.106380>
- Ahmed S, Chakraborty RK, Essam DL, Ding W (2022) Poly-linear regression with augmented long short term memory neural network: predicting time series data. *Inf Sci* 606:573–600. <https://doi.org/10.1016/j.ins.2022.05.078>
- Albrecht T, Rausch TM, Derra ND (2021) Call me maybe: methods and practical implementation of artificial intelligence in call center arrivals' forecasting. *J Bus Res* 123:267–278. <https://doi.org/10.1016/j.jbusres.2020.09.033>
- Alsharaf A, Aggarwal K, Sonia KM, Mishra A (2022) Review of ML and AutoML solutions to forecast time-series data. *Arch Comput Method Eng* 29:5297–5311. <https://doi.org/10.1007/s11831-022-09765-0>
- Anand S, Mishra K (2022) Identifying potential millennial customers for financial institutions using SVM. *J Financ Serv Mark* 27:335–345. <https://doi.org/10.1057/s41264-021-00128-7>
- Angus RW, Packard MD, Clark BB (2023) Distinguishing unpredictability from uncertainty in entrepreneurial action theory. *Small Bus Econ* 60:1147–1169. <https://doi.org/10.1007/s11187-022-00651-4>
- Aria M, Cuccurullo C (2023) Package 'bibliometrix'—comprehensive science mapping analysis. CRAN. <https://cran.r-project.org/web/packages/bibliometrix/bibliometrix.pdf>
- Aria M, Cuccurullo C (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* 11:959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Arunraj NS, Ahrens D (2015) A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *Int J Prod Econ* 170:321–335. <https://doi.org/10.1016/j.ijpe.2015.09.039>
- Ballestar MT, Grau-Carles P, Sáinz J (2019) Predicting customer quality in e-commerce social networks: a machine learning approach. *Rev Manag Sci* 13:589–603. <https://doi.org/10.1007/s11846-018-0316-x>
- Bi X, Adomavičius G, Li W, Qu A (2022) Improving sales forecasting accuracy: a tensor factorization approach with demand awareness. *INFORMS J Comput* 34:1644–1660. <https://doi.org/10.1287/ijoc.2021.1147>
- Black JE, Kueper JK, Williamson TS (2022) An introduction to machine learning for classification and prediction. *Fam Pract* 40:200–204. <https://doi.org/10.1093/fampra/cm104>
- Bohanec M, Borštnar MK, Robnik-Šikonja M (2017) Explaining machine learning models in sales predictions. *Expert Syst Appl* 71:416–428. <https://doi.org/10.1016/j.eswa.2016.11.010>
- Booth A, Sutton A, Papaioannou D (2016) *Systematic approaches to a successful literature review*, 2nd edn. Sage, Los Angeles
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Bukhari AH, Raja MAZ, Sulaiman M, Islam S, Shoaib M, Kumam P (2020) Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access* 8:71326–71338. <https://doi.org/10.1109/access.2020.2985763>
- Cariolle J, Goujon M (2013) Measuring macroeconomic instability: a critical survey illustrated with exports series. *J Econ Surv* 29:1–26. <https://doi.org/10.1111/joes.12036>
- Castillo PA, Mora AM, Faris H, Merelo JJ, Garcia-Sanchez P, Fernandez-Ares AJ, De las Cuevas P, Garcia-Arenas MI (2017) Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment. *Knowl-Based Syst* 115:133–151. <https://doi.org/10.1016/j.knysys.2016.10.019>

- Chashmi AJ, Rahmati V, Rezasoroush B, Alamoti MM, Askari M, Khalili FH (2021) Predicting customer turnover using recursive neural networks. *Wirel Commun Mob Comput* 2021:1–11. <https://doi.org/10.1155/2021/6623052>
- Chen M (2022) A MapReduce-based artificial neural network churn prediction for music streaming service. *Int J Comput Sci Netw Secur* 22:55–60. <https://doi.org/10.22937/IJCSNS.2022.22.1.9>
- Chen SS, Choubey B, Singh V (2021) A neural network based price sensitive recommender model to predict customer choices based on price effect. *J Retail Consum Serv* 61:102573. <https://doi.org/10.1016/j.jretconser.2021.102573>
- Clark WR, Clark LA, Raffo DM, Williams RI (2021) Extending Fisch and Block's (2018) tips for a systematic review in management and business literature. *Manag Rev Q* 71(1):215–231. <https://doi.org/10.1007/s11301-020-00184-8>
- Claveria O, Monte E, Torra S (2020) Economic forecasting with evolved confidence indicators. *Econ Model* 93:576–585. <https://doi.org/10.1016/j.econmod.2020.09.015>
- Conrad CA (2020) *Angewandte makroökonomie—eine praxisbezogene einföhrung*. Springer, Wiesbaden. <https://doi.org/10.1007/978-3-658-30055-5>
- Contreras-Masse R, Ochoa-Zezzatti A, Perez-Dominguez L, Naudascher K (2022) Data-driven enterprises for economic recovery after pandemic: a study case in consumer-packaged goods industry. *Int J Comb Optim Probl Inform* 13:33–39
- Coussement K, Van den Poel D (2008) Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. *Expert Syst Appl* 34:313–327. <https://doi.org/10.1016/j.eswa.2006.09.038>
- Coveney PV, Dougherty ER, Highfield RR (2016) Big data need big theory too. *Phil Trans R Soc A* 374:20160153. <https://doi.org/10.1098/rsta.2016.0153>
- Danaher PJ, Brodie RJ (2000) Understanding the characteristics of price elasticities for frequently purchased packaged goods. *J Mark Manag* 16:917–936. <https://doi.org/10.1362/026725700784683726>
- Denyer D, Tranfield D (2009) Producing a systematic review. In: Buchanan DA, Bryman A (eds) *The sage handbook of organizational research methods*. Sage Publications Ltd, London, pp 671–689
- Denyer D, Tranfield D, Van Aken JE (2008) Developing design propositions through research synthesis. *Organ Stud* 29:393–413. <https://doi.org/10.1177/0170840607088020>
- Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: an overview and guidelines. *J Bus Res* 133:285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Dörner A, Maisch M, Münchrath J (2023) Beben an den märkten. *Handelsblatt*, Düsseldorf, pp 6–8
- Duca JV, Muellbauer J, Murphy A (2010) Housing markets and the financial crisis of 2007–2009: lessons for the future. *J Financ Stabil* 6:203–217. <https://doi.org/10.1016/j.jfs.2010.05.002>
- Duncan GT, Gorr WL, Szczypula J (2001) Forecasting analogous time series. In: Armstrong JS (ed) *Principles of forecasting: a handbook for researchers and practitioners*. Springer, Boston, pp 195–213. https://doi.org/10.1007/978-0-306-47630-3_10
- Durst S, Foli S, Edvardsson IR (2022) A systematic literature review on knowledge management in SMEs: current trends and future directions. *Manag Rev Q* 74:263–288. <https://doi.org/10.1007/s11301-022-00299-0>
- Durugbo CM, Al-Balushi Z (2022) Supply chain management in times of crisis: a systematic review. *Manag Rev Q* 73:1179–1235. <https://doi.org/10.1007/s11301-022-00272-x>
- Efat MIA, Hájek P, Abedin MZ, Azad RU, Jaber MA, Aditya S, Hassan MK (2022) Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales. *Ann Oper Res* 22:1–32. <https://doi.org/10.1007/s10479-022-04838-6>
- El Shoubaki A, Block J, Lasch F (2021) The couple business as a unique form of business: a review of the empirical evidence. *Manag Rev Q* 72:115–147. <https://doi.org/10.1007/s11301-020-00206-5>
- Ernst J, Schmidt J, Beneken G (2020) *Grundkurs informatik: grundlagen und konzepte für die erfolgreiche IT-praxis eine umfassende, praxisorientierte einföhrung*, 7th edn. Springer, Wiesbaden
- Esmailzadeh P (2023) The role of ChatGPT in disrupting concepts, changing values, and challenging ethical norms: a qualitative study. *AI Ethics* 2023:1–14. <https://doi.org/10.1007/s43681-023-00338-w>
- Esmeli R, Bader-El-Den M, Abdullahi H (2021) Towards early purchase intention prediction in online session based retailing systems. *Electron Mark* 31:697–715. <https://doi.org/10.1007/s12525-020-00448-x>
- European Central Bank (2023) Eurosystem staff macroeconomic projections for the euro area-December 2023. https://www.ecb.europa.eu/pub/pdf/other/ecb.projections202312_eurosystemstaff-9a39ab5088.en.pdf

- Fink A (2014) Conducting research literature reviews: from the internet to paper. Sage Publications, Thousand Oaks, CA
- Fisch C, Block JH (2018) Six tips for your (systematic) literature review in business and management research. *Manag Rev Q* 68:103–106. <https://doi.org/10.1007/s11301-018-0142-x>
- Frees EW, Miller TW (2004) Sales forecasting using longitudinal data models. *Int J Forecast* 20:99–114. [https://doi.org/10.1016/s0169-2070\(03\)00005-0](https://doi.org/10.1016/s0169-2070(03)00005-0)
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Geiler L, Affeldt S, Nadif M (2022) A survey on machine learning methods for churn prediction. *Int J Data Sci Anal* 14:217–242. <https://doi.org/10.1007/s41060-022-00312-5>
- Gharibshah Z, Zhu X (2021) User response prediction in online advertising. *ACM Comput Surv* 54:1–43. <https://doi.org/10.1145/3446662>
- Ghoddusi H, Creamer GG, Rafizadeh N (2019) Machine learning in energy economics and finance: a review. *Energy Econ* 81:709–727. <https://doi.org/10.1016/j.eneco.2019.05.006>
- Ghonghadze J, Lux T (2012) Modelling the dynamics of EU economic sentiment indicators: an interaction-based approach. *Appl Econ* 44:3065–3088. <https://doi.org/10.1080/00036846.2011.570716>
- Glaeser CK, Fisher M, Su X (2019) Optimal retail location: empirical methodology and application to practice. *Manuf Serv Oper Manag* 21:86–102. <https://doi.org/10.1287/msom.2018.0759>
- Göcke L, Hülsebusch K, Menter M (2021) The legitimacy of corporate entrepreneurship: a structured literature review. *Manag Rev Q* 72:385–416. <https://doi.org/10.1007/s11301-021-00246-5>
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge, MA
- Gordon BR, Goldfarb A, Li Y (2013) Does price elasticity vary with economic growth? A cross-category analysis. *J Mark Res* 50:4–23. <https://doi.org/10.1509/jmr.11.0162>
- Gürses-Tran G, Monti A (2022) Advances in time series forecasting development for power systems' operation with MLOps. *Forecast* 4:501–524. <https://doi.org/10.3390/forecast4020028>
- Gusenbauer M, Haddaway NR (2019) Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res Synth Methods* 11:181–217. <https://doi.org/10.1002/jrsm.1378>
- Hasheminejad SA, Shabaab M, Javadinarab N (2022) Developing cluster-based adaptive network fuzzy inference system tuned by particle swarm optimization to forecast annual automotive sales: a case study in Iran market. *Int J Fuzzy Syst* 24:2719–2728. <https://doi.org/10.1007/s40815-022-01263-6>
- Homburg C (2020) Marketingmanagement: strategie-instrumente-umsetzung-unternehmensführung, 6th edn. Springer, Berlin
- Huang M-H, Rust RT (2021) A strategic framework for artificial intelligence in marketing. *J Acad Mark Sci* 49(1):30–50. <https://doi.org/10.1007/s11747-020-00749-9>
- Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. Otexts, Lexington, Kentucky
- Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *Int J Forecast* 22:679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Kamakura WA, Du RY (2011) How economic contractions and expansions affect expenditure patterns. *J Consum Res* 39:229–247. <https://doi.org/10.1086/662611>
- Kamara AF, Chen EH, Pan Z (2022) An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Inf Sci* 594:1–19. <https://doi.org/10.1016/j.ins.2022.02.015>
- Kaushal N, Kaurav RPS, Sivathanu B, Kaushik N (2023) Artificial intelligence and HRM: identifying future research Agenda using systematic literature review and bibliometric analysis. *Manag Rev Q* 73:455–493. <https://doi.org/10.1007/s11301-021-00249-2>
- Keding C (2021) Understanding the interplay of artificial intelligence and strategic management: four decades of research in review. *Manag Rev Q* 71:91–134. <https://doi.org/10.1007/s11301-020-00181-x>
- Kepes S, Banks GC, McDaniel MA, Whetzel DL (2012) Publication bias in the organizational sciences. *Organ Res Methods* 15:624–662. <https://doi.org/10.1177/1094428112452760>
- Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. *J Bank Financ* 34:2767–2787. <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- Kharfan M, Chan VWK, Firdolas Efendigil T (2021) A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Ann Oper Res* 303:159–174. <https://doi.org/10.1007/s10479-020-03666-w>

- Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering—technical report EBSE-2007-01. Keele University and Durham University, Keele and Durham
- Kmiecik M, Zangana H (2022) Supporting of manufacturing system based on demand forecasting tool. *Logforum* 18:35–50. <https://doi.org/10.17270/j.log.2022.637>
- Kozak J, Kania K, Juszczuk P, Mitrega M (2021) Swarm intelligence goal-oriented approach to data-driven innovation in customer churn management. *Int J Inf Manag* 60:102357. <https://doi.org/10.1016/j.ijinfomgt.2021.102357>
- Kuckertz A, Block J (2021) Reviewing systematic literature reviews: ten key questions and criteria for reviewers. *Manag Rev Q* 71:519–524. <https://doi.org/10.1007/s11301-021-00228-7>
- Lamberton C, Stephen AT (2016) A thematic exploration of digital, social media, and mobile marketing: research evolution from 2000 to 2015 and an Agenda for future inquiry. *J Mark* 80:146–172. <https://doi.org/10.1509/jm.15.0415>
- Lanquillon C (2019) Grundzüge des maschinellen lernens. In: Sacht S, Lanquillon C (eds) *Blockchain und maschinelles lernen—Wie das maschinelle lernen und die distributed ledger-technologie voneinander profitieren*. Springer, Berlin Heidelberg, pp 89–142
- Lehrer S, Xie T, Zhang X (2021) Social media sentiment, model uncertainty, and volatility forecasting. *Econ Model* 102:105556. <https://doi.org/10.1016/j.econmod.2021.105556>
- Lin HF, Lin J, Wang F (2022) An innovative machine learning model for supply chain management. *J Innov Knowl* 7:100276. <https://doi.org/10.1016/j.jik.2022.100276>
- Linnenluecke MK, Marrone M, Singh AK (2019) Conducting systematic literature reviews and bibliometric analyses. *Aust J Manag* 45:175–194. <https://doi.org/10.1177/0312896219877678>
- Liu C-H, Gu J-C, Yang M-T (2021) A simplified LSTM neural networks for one day-ahead solar power forecasting. *IEEE Access* 9:17174–17195. <https://doi.org/10.1109/access.2021.3053638>
- Loayza NV, Rancière R, Servén L, Ventura J (2007) Macroeconomic volatility and welfare in developing countries: an introduction. *World Bank Econ Rev* 21:343–357. <https://doi.org/10.1093/wber/lhm017>
- Lotfi A, Jiang Z, Lotfi A, Jain DC (2023) Estimating life cycle sales of technology products with frequent repeat purchases: a fractional calculus-based approach. *Inf Syst Res* 34:409–422. <https://doi.org/10.1287/isre.2022.1131>
- Ma S, Fildes R (2020) Forecasting third-party mobile payments with implications for customer flow prediction. *Int J Forecast* 36:739–760. <https://doi.org/10.1016/j.ijforecast.2019.08.012>
- Ma L, Sun B (2020) Machine learning and AI in marketing—connecting computing power to human insights. *Int J Res Mark* 37(3):481–504. <https://doi.org/10.1016/j.ijresmar.2020.04.005>
- Makridakis S, Spiliotis E, Assimakopoulos V (2022) The M5 competition: background, organization, and implementation. *Int J Forecast* 38:1325–1336. <https://doi.org/10.1016/j.ijforecast.2021.07.007>
- Martínez-de-Albéniz V, Planas A, Nasini S (2020) Using clickstream data to improve flash sales effectiveness. *Prod Oper Manag* 29:2508–2531. <https://doi.org/10.1111/poms.13238>
- Meisenbacher S, Turowski M, Phipps K, Rätz M, Müller D, Hagenmeyer V, Mikut R (2022) Review of automated time series forecasting pipelines. *Wiley Interdiscip Rev Data Min Knowl Discov* 12:1–42. <https://doi.org/10.1002/widm.1475>
- Menhaj M, Kavooosi-Kalashami M (2022) Developing a hybrid forecasting system for agricultural commodity prices (case study: Thailand rice free on board price). *Cienc Rural*. <https://doi.org/10.1590/0103-8478cr20201128>
- Milošević M, Živić N, Anđelković I (2017) Early churn prediction with personalized targeting in mobile social games. *Expert Syst Appl* 83:326–332. <https://doi.org/10.1016/j.eswa.2017.04.056>
- Ministry for Economic Affairs and Climate (2023) *Schlaglichter der wirtschaftspolitik—Februar 2023 monatsbericht*. https://www.bmwk.de/Redaktion/DE/Publikationen/Schlaglichter-der-Wirtschaftspolitik/schlaglichter-der-wirtschaftspolitik-02-2023.pdf?__blob=publicationFile&v=1
- Mitchell T (1997) *Machine learning*. McGraw-Hill, New York
- Mondello E (2022) *Corporate finance*. Springer, Wiesbaden. <https://doi.org/10.1007/978-3-658-34408-5>
- Murphy KP (2012) *Machine learning: a probabilistic perspective*. MIT Press, Cambridge
- Mustak M, Salminen J, Plé L, Wirtz J (2021) Artificial intelligence in marketing: topic modeling, scientometric analysis, and research agenda. *J Bus Res* 124:389–404. <https://doi.org/10.1016/j.jbusres.2020.10.044>
- Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH (2006) Defection detection: measuring and understanding the predictive accuracy of customer churn models. *J Mark Res* 43:204–211. <https://doi.org/10.1509/jmkr.43.2.204>

- Niedzela L (2022) Ansätze zur wirtschaftlichen bewertung von geschäftsprozess-compliance—eine systematische literaturanalyse. Springer, Wiesbaden. <https://doi.org/10.1007/978-3-658-36725-1>
- Pessanha GRG, Soares EA (2021) Just one post? Forecasts of daily sales of beauty and cosmetics retail companies based on the influence of social media. *Remark Rev Bras Mark* 20:241–266. <https://doi.org/10.5585/remark.v20i4.17914>
- Petropoulos A, Siakoulis V (2021) Can central bank speeches predict financial market turbulence? Evidence from an adaptive NLP sentiment index analysis using XGBoost machine learning technique. *Cent Bank Rev* 21:141–153. <https://doi.org/10.1016/j.cbrev.2021.12.002>
- Petticrew M, Roberts H (2006) *Systematic reviews in the social sciences: a practical guide*. Blackwell Publishing, Oxford
- Porter MF (1980) An algorithm for suffix stripping. *Program* 14:130–137. <https://doi.org/10.1108/eb046814>
- Poza C, Monge M (2023) Forecasting Spanish economic activity in times of COVID-19 by means of the RT-LEI and machine learning techniques. *Appl Econ Lett* 30:472–477. <https://doi.org/10.1080/13504851.2021.1994122>
- Punia S, Shankar S (2022) Predictive analytics for demand forecasting: a deep learning-based decision support system. *Knowl Based Syst* 258:109956. <https://doi.org/10.1016/j.knosys.2022.109956>
- Purohit SK, Panigrahi S, Sethy PK, Behera SK (2021) Time series forecasting of price of agricultural products using hybrid methods. *Appl Artif Intell* 35:1388–1406. <https://doi.org/10.1080/08839514.2021.1981659>
- Puterman E, Weiss J, Hives BA, Gemmill A, Karasek D, Mendes WB, Rehkopf DH (2020) Predicting mortality from 57 economic, behavioral, social, and psychological factors. *Proc Natl Acad Sci USA* 117:16273–16282. <https://doi.org/10.1073/pnas.1918455117>
- Quintero Y, Ardila D, Aguilar J, Cortés S (2022) Analysis of the socioeconomic impact due to COVID-19 using a deep clustering approach. *Appl Soft Comput* 129:109606. <https://doi.org/10.1016/j.asoc.2022.109606>
- Raizada S, Saini JR (2021) Comparative analysis of supervised machine learning techniques for sales forecasting. *Int J Adv Comput Sci Appl* 12:102–110. <https://doi.org/10.14569/ijacsa.2021.0121112>
- Raju YN, Acharya D (2020) Revisiting the volatility-growth relationship: some cross country evidence, 1978–2017. *Cogent Econ Financ* 8:1826655. <https://doi.org/10.1080/23222039.2020.1826655>
- Rejeb A, Keogh JG, Wamba SF, Treiblmaier H (2020) The potentials of augmented reality in supply chain management: a state-of-the-art review. *Manag Rev Q* 71:819–856. <https://doi.org/10.1007/s11301-020-00201-w>
- Rousseau DM (2006) Is there such a thing as “evidence-based management”? *Acad Manag Rev* 31:256–269. <https://doi.org/10.5465/amr.2006.20208679>
- Rowley J, Slack F (2004) Conducting a literature review. *Manag Res News* 27:31–39. <https://doi.org/10.1108/01409170410784185>
- Ryu G-A, Nasridinov A, Rah H, Yoo K-H (2020) Forecasts of the amount purchase pork meat by using structured and unstructured big data. *Agriculture* 10:21. <https://doi.org/10.3390/agriculture10010021>
- Sauer S (2019) *Moderne datenanalyse mit R*. Springer, Wiesbaden
- Schmid Y, Dowling M (2022) New work: new motivation? A comprehensive literature review on the impact of workplace technologies. *Manag Rev Q* 72:59–86. <https://doi.org/10.1007/s11301-020-00204-7>
- Seyedan M, Mafakheri F (2020) Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *J Big Data* 7:1–22. <https://doi.org/10.1186/s40537-020-00329-2>
- Shaikh AA, Lakshmi KS, Tongkachok K, Alanya-Beltrán J, Ramirez-Asís E, Pérez-Falcón J (2022) Empirical analysis in analysing the major factors of machine learning in enhancing the e-business through structural equation modelling (SEM) approach. *Int J Syst Assur Eng Manag* 13:681–689. <https://doi.org/10.1007/s13198-021-01590-1>
- Sievering O (2021) Inflation und geldpolitik. In: Drewello H, Kupferschmidt F, Sievering O (eds) *Markt und staat—eine anwendungsorientierte einföhrung in die allgemeine volkswirtschaftslehre*. Springer, Wiesbaden, pp 129–157. https://doi.org/10.1007/978-3-658-33096-5_6
- Sorjamaa A, Hao J, Reyhani N, Ji Y, Lendasse A (2007) Methodology for long-term prediction of time series. *Neurocomputing* 70:2861–2869. <https://doi.org/10.1016/j.neucom.2006.06.015>
- Syntetos AA, Boylan JE (2005) The accuracy of intermittent demand estimates. *Int J Forecast* 21:303–314. <https://doi.org/10.1016/j.ijforecast.2004.10.001>
- Syntetos AA, Nikolopoulos K, Boylan JE (2010) Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *Int J Forecast* 26:134–143. <https://doi.org/10.1016/j.ijforecast.2009.05.016>

- Taskan B, Junça-Silva A, Caetano A (2022) Clarifying the conceptual map of VUCA: a systematic review. *Int J Organ Anal* 30:196–217. <https://doi.org/10.1108/ijoa-02-2022-3136>
- Thommen JP, Achleitner AK, Gilbert DU, Hachmeister D, Kaiser G (2017) *Allgemeine betriebswirtschaftslehre: umfassende einföhrung aus managementorientierter Sicht*, 9th edn. Springer, Wiesbaden. <https://doi.org/10.1007/978-3-658-27246-3>
- Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br J Manag* 14:207–222. <https://doi.org/10.1111/1467-8551.00375>
- Trapp R (2012) *Konvergenz des rechnungswesens*. Springer, Wiesbaden. https://doi.org/10.1007/978-3-658-26431-4_6
- Troise C, Corvello V, Ghobadian A, O'Regan N (2022) How can SMEs successfully navigate VUCA environment: the role of agility in the digital transformation era. *Technol Forecast Soc Change* 174:121227. <https://doi.org/10.1016/j.techfore.2021.121227>
- Tsao Y-C, Chen Y-K, Chiu S-H, Lu J-C, Vu T-L (2022) An innovative demand forecasting approach for the server industry. *Technovation* 110:102371. <https://doi.org/10.1016/j.technovation.2021.102371>
- Tudor C (2022) The impact of the COVID-19 pandemic on the global web and video conferencing SaaS market. *Electronics* 11:2633. <https://doi.org/10.3390/electronics11162633>
- Verma S, Sharma R, Deb S, Maitra D (2021) Artificial intelligence in marketing: systematic review and future research direction. *Int J Inf Manag Data Insights* 1:100002. <https://doi.org/10.1016/j.jjime.2020.100002>
- Vlačić B, Corbo L, Costa e Silva S, Dabić M (2021) The evolving role of artificial intelligence in marketing: a review and research agenda. *J Bus Res* 128:187–203. <https://doi.org/10.1016/j.jbusres.2021.01.055>
- Walter A-T (2021) Organizational agility: ill-defined and somewhat confusing? A systematic literature review and conceptualization. *Manag Rev Q* 71:343–391. <https://doi.org/10.1007/s11301-020-00186-6>
- Wang C-H (2022) Considering economic indicators and dynamic channel interactions to conduct sales forecasting for retail sectors. *Comput Ind Eng* 165:107965. <https://doi.org/10.1016/j.cie.2022.107965>
- Wang QF, Xu M, Hussain A (2019) Large-scale ensemble model for customer churn prediction in search ads. *Cognit Comput* 11:262–270. <https://doi.org/10.1007/s12559-018-9608-3>
- Wu Z, Li Z (2021) Customer churn prediction for commercial banks using customer-value-weighted machine learning models. *J Credit Risk* 17:15–42. <https://doi.org/10.21314/jcr.2021.011>
- Wu C-F, Huang S-C, Chiou C-C, Chang T, Chen Y-C (2022) The relationship between economic growth and electricity consumption: bootstrap ARDL test with a fourier function and machine learning approach. *Comput Econ* 60:1197–1220. <https://doi.org/10.1007/s10614-021-10097-7>
- Xiao Y, Watson M (2019) Guidance on conducting a systematic literature review. *J Plan Educ Res* 39:93–112. <https://doi.org/10.1177/0739456x17723971>
- Xie S-M (2020) Comparative models in customer base analysis: parametric model and observation-driven model. *J Bus Econ Manag* 21:1731–1751. <https://doi.org/10.3846/jbem.2020.13194>
- Yang C-H, Chang P-Y (2020) Forecasting the demand for container throughput using a mixed-precision neural architecture based on CNN-LSTM. *Mathematics* 8:1784. <https://doi.org/10.3390/math8101784>
- Zhang XD (2020) *A matrix algebra approach to artificial intelligence*. Springer Nature Singapore Pte Ltd., Singapore. <https://doi.org/10.1007/978-981-15-2770-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.