AKURASI Jurnal Studi Akuntansi dan Keuangan
p-ISSN : 2720-9067
e-ISSN : 2685-1059
Program Studi Magister Akuntansi
Fakultas Ekonomi dan Bisnis
Universitas Mataram

FINANCIAL FORECASTING AND MACHINE LEARNING: A BIBLIOMETRIC ANALYSIS OF GLOBAL RESEARCH TRENDS

Iwan Kurniawan[1], Nugraha[2], Maya Sari[3]
Faculty of Economics and Business Education, Universitas Pendidikan Indonesia, Bandung, Indonesia[123]
[1]Corresponding author: iwan.kurniawan@upi.edu

ARTICLE INFORMATION

Article history:
Received: 25/10/2025
First Revision: 25/02/2026
Accepted:20/04/2026
Available online: 30/06/2026

ABSTRAK

Pesatnya pertumbuhan aplikasi kecerdasan buatan (artificial intelligence) dalam bidang keuangan telah menghasilkan literatur yang luas, namun tinjauan komprehensif mengenai struktur intelektualnya masih terbatas. Penelitian ini bertujuan untuk memetakan dan menganalisis tren riset global dalam peramalan keuangan (financial forecasting) dan prediksi harga saham menggunakan machine learning antara tahun 2015 dan 2025. Dengan menggunakan pendekatan bibliometrik, sebanyak 197 artikel terindeks Scopus dianalisis melalui paket Bibliometrix di R Studio dengan mengikuti kerangka kerja PRISMA. Analisis ini mencakup performa publikasi, jaringan kolaborasi antarpenulis (co-authorship), serta evolusi tematik. Hasil penelitian menunjukkan tingkat pertumbuhan publikasi tahunan sebesar 13,98% dengan rata-rata 16,16 sitasi per dokumen. "Forecasting" muncul sebagai tema penelitian sentral yang terhubung erat dengan "machine learning", "financial markets", dan "LSTM." Kolaborasi internasional mencakup 32,99% dari total publikasi, dengan Tiongkok, India, dan Amerika Serikat sebagai kontributor utama. Evolusi tematik menunjukkan adanya pergeseran dari pendekatan ekonometrika tradisional menuju model prediksi berbasis kecerdasan buatan dan deep learning. Studi ini berkontribusi dengan menyediakan peta intelektual yang komprehensif mengenai riset peramalan keuangan berbasis AI serta mengidentifikasi arah penelitian masa depan bagi para akademisi, praktisi, dan pembuat kebijakan.

Kata Kunci: Bibliometrik, financial forecasting, machine learning, deep learning, stock market.

ABSTRACT

The rapid growth of artificial intelligence applications in finance has generated a large body of literature. However, a comprehensive overview of its intellectual structure remains limited. To address this gap, this study aims to map and analyze global research trends in financial forecasting and stock price prediction using machine learning between 2015 and 2025. Using a bibliometric approach, 197 Scopus-indexed articles were analyzed through the Bibliometrix package in R Studio following the PRISMA framework. The analysis includes publication performance, co-authorship collaboration networks, and thematic evolution. The results indicate an annual publication growth rate of 13.98% with an average of 16.16 citations per document. "Forecasting" emerges as the central research theme, closely connected with "machine learning," "financial markets," and "LSTM."

International collaboration accounts for 32.99%, with China, India, and the United States as the leading contributors. Thematic evolution shows a shift from traditional econometric approaches toward artificial intelligence and deep learning–based prediction models. This study contributes by providing a comprehensive intellectual map of AI-driven financial forecasting research and identifying future research directions for scholars, practitioners, and policymakers.

Keywords: *Bibliometric, financial forecasting, machine learning, deep learning, stock market*

1. Introduction

The rapid advancement of digital technology over the past two decades has fundamentally transformed how researchers and practitioners understand, model, and predict financial phenomena. Open data availability, advances in computational power, and the emergence of artificial intelligence (AI) methods have shifted the paradigm of financial research from deterministic, theory-driven approaches toward adaptive, nonlinear, and data-driven methodologies (Makridakis et al., 2020). This transition has not only enhanced forecasting accuracy but has also expanded the scope of analysis to encompass increasingly complex, dynamic, and globally integrated financial markets, driven by recent advances in artificial intelligence and deep learning applications in finance (Sholapurapu, 2025; Li et al., 2023).

In the context of capital markets, stock price forecasting has emerged as a highly compelling area of research, both theoretically and practically. The accurate stock price prediction is crucial for informed investment decision-making, effective risk management, and the formulation of macroeconomic policy (Xiaolin Zhang & Tan, 2018). In the accounting and financial reporting context, forecasting models are increasingly employed to predict firm performance, assess financial distress, and support earnings management detection. Traditional forecasting models such as the Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) have been widely applied. However, they demonstrate limitations in capturing nonlinear relationships and extreme volatility commonly observed in financial markets (Marisetty, 2024). These shortcomings have catalyzed a paradigm shift toward machine learning and deep learning models, including Long Short-Term Memory (LSTM), Support Vector Machines (SVM), and Random Forests, which have the capacity to discern complex patterns from large-scale financial data.

This methodological transformation signifies the integration of classical statistical forecasting with sophisticated computational techniques capable of processing heterogeneous data, including both numerical and textual information. The incorporation of machine learning has enhanced financial research through innovations such as sentiment analysis, Natural Language Processing (NLP), and anomaly detection, all of which contribute to augmenting market efficiency (Chen et al., 2023). The intersection between artificial intelligence and financial economics has led to the emergence of data-driven finance, emphasizing algorithmic learning and big data analytics as the foundation of contemporary financial prediction.

With the proliferation of scientific publications in this domain, it becomes essential to understand the evolution of research trends, collaborations, and thematic focuses on a global scale. Bibliometric analysis offers a robust framework for mapping the intellectual structure and identifying the developmental trajectory of a discipline by examining publications, citations, and keyword networks (Aria & Cuccurullo, 2017). This approach facilitates the identification of core themes, influential authors, and collaborative networks that define the intellectual and methodological landscape of machine learning-based financial forecasting.

Over the last decade, the intersection between finance and machine learning has experienced remarkable growth. Based on the Scopus database records, there has been an annual increase of 13.98% in publications on “financial forecasting” and “stock price prediction” since 2015. Leading the research output are countries like China, India, and the United States, with the United Kingdom, South Korea, and Australia following closely. This trend highlights the globalization of financial forecasting research and the interdisciplinary connections with economics, data science, and information technology.

Beyond methodological factors, external events such as the COVID-19 pandemic have also significantly influenced the research direction. The emergence of topics like “volatility forecasting” and “market uncertainty” between 2020 and 2022 demonstrates how global crises can drive research innovation by focusing on financial system resilience and adaptive modeling under uncertainty (Meher et al., 2023). Similarly, the growing interest in cryptocurrency forecasting signifies a shift from traditional asset prediction to exploring digital and decentralized financial markets.

Although research on machine learning applications in financial forecasting is rapidly expanding, there are still considerable theoretical and empirical gaps in understanding the evolution of this field over time. Most existing studies focus heavily on algorithmic performance and predictive accuracy, while giving less attention to synthesizing the intellectual framework of the field. More specifically, only a few studies offer a thorough account of thematic development, international collaboration, and methodological convergence, leading to an incomplete grasp of how machine learning–based forecasting has advanced conceptually and methodologically within the broader context of financial science. In this context, bibliometric analysis plays a crucial role in overcoming these limitations by providing systematic, quantitative insights into the structure and global interconnectedness of scientific knowledge in the field (Cobo et al., 2011; Donthu et al., 2021; Dzigbede & Pathak, 2020).

Moreover, prior bibliometric studies on finance and artificial intelligence tend to address either financial technology or machine learning applications in general, rather than specifically exploring the intersection of financial forecasting and stock price prediction. This oversight results in a significant research gap in identifying the key themes, emerging research directions, and patterns of global collaboration that influence the development of predictive finance.

In order to fill this gap, this study presents an extensive bibliometric mapping of financial forecasting research. It integrates performance analysis, science mapping, and thematic evolution analysis, utilizing Scopus-indexed publications from 2015 to 2025. Through this approach, the study offers a systematic understanding of the shift in predictive financial research from traditional econometric models to artificial intelligence-driven methodologies.

2. Theoretical Framework

The evolution of financial forecasting has long been grounded in classical economic theories, such as the Efficient Market Hypothesis (EMH), which asserts that asset prices fully reflect all available information, thereby limiting the possibility of consistently achieving excess returns (Fama, 1970). However, empirical evidence from behavioral finance and adaptive market theories reveals that markets are not always perfectly efficient. Investor biases, information asymmetry, and irrational behavior can create predictable patterns and anomalies that undermine the EMH (Lo, 2004; Barberis et al., 1998). These findings have encouraged researchers to explore forecasting methods capable of modeling complex, nonlinear, and dynamic relationships in financial data. The increasing unpredictability of financial markets, particularly during crisis periods, has consequently redirected focus toward data-driven and computationally intelligent approaches to improve predictive performance.

Historically, financial forecasting research has been dominated by traditional econometric models like ARIMA and GARCH. ARIMA models capture linear temporal dependencies, while GARCH models are used to analyze volatility clustering and leverage effects (Engle, 1982; Bollerslev, 1986). Despite offering interpretable and statistically grounded insights, these models fall short in capturing nonlinear and high-dimensional interactions present in modern financial data (Ghysels et al., 2007). This limitation has prompted the adoption of machine learning (ML) and deep learning (DL) approaches, which can flexibly approximate complex relationships between input features and target variables without relying on strict distributional assumptions (Gu et al., 2020; Fischer & Krauss, 2018).

In recent years, machine learning has demonstrated remarkable potential in enhancing forecasting accuracy across various financial domains, including stock price prediction, volatility estimation, and risk management (Henrique et al., 2019). Techniques such as support vector machines (SVM), random forests, and gradient boosting trees have shown superior adaptability to nonlinear structures compared to traditional regression-based models (Hastie et al., 2009). Moreover, deep learning architectures, particularly Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), have been extensively employed to capture temporal dependencies and extract meaningful patterns from large-scale time-series data (Sezer et al., 2020). These approaches have proven effective in volatile and high-frequency environments where traditional models fail to generalize effectively.

Beyond model sophistication, researchers increasingly explore hybrid and ensemble frameworks that integrate the interpretability of economic models with the flexibility of ML and DL. Such approaches capitalize on the complementary strengths of both paradigms by modeling linear dependencies through statistical modeling while capturing nonlinear residual structures using deep neural networks (Makridakis et al., 2020; Fang & Wang, 2024). Empirical evidence suggests that hybrid models often outperform single-method approaches, particularly in volatile market conditions where adaptive learning and effective feature integration are crucial (Tian et al., 2022). The hybridization trend highlights an emerging paradigm in which artificial intelligence augments traditional forecasting theory, bridging the gap between statistical inference and computational prediction.

Another transformative development in financial forecasting research is the incorporation of Natural Language Processing (NLP) and textual sentiment analysis. Unstructured textual information from news articles, financial reports, and social media provides valuable signals about investor sentiment and market expectations. Early studies found that media tone and sentiment significantly affect market movements (Tetlock, 2007; Bollen et al., 2011). More recent advances in NLP, particularly transformer-based models like FinBERT, have enhanced the ability to quantify sentiment, uncertainty, and attention effects, thereby improving predictive accuracy. This integration of textual and numerical data marks a shift toward multi-source and real-time forecasting frameworks that capture behavioral and informational frictions in financial markets more effectively.

In parallel, the literature acknowledges the importance of market microstructure and asset-class heterogeneity in shaping forecasting performance. Factors such as order flow, bid–ask spreads, and market depth determine short-term price formation and thus affect the efficacy of predictive models (Hasbrouck, 2004). Empirical results suggest that ML models incorporating microstructure data tend to perform better in intraday and high-frequency settings, while macroeconomic and textual features contribute more substantially in long-horizon predictions (Sirignano, 2019; Gu et al., 2020). Furthermore, deep learning and sentiment-enriched models are particularly effective in volatile and sentiment-driven markets such as cryptocurrencies (Corbet et al., 2019; Jiang et al., 2017). Collectively, these findings underscore the need for adaptive modeling frameworks that account for asset-specific characteristics, data granularity, and forecasting horizons.

Despite the expanding body of research on machine learning applications in financial forecasting, a comprehensive understanding of the field's intellectual structure and thematic evolution remains limited. While numerous studies focus on model performance comparison and algorithmic innovation, relatively few attempt to systematically map the scientific landscape of this domain using bibliometric methods. Bibliometric analysis enables quantitative examination of research productivity, co-authorship networks, and conceptual development over time (Aria & Cuccurullo, 2017; Donthu et al., 2021). Although recent bibliometric works indicate a rapid growth in publications at the intersection of artificial intelligence and finance, particularly dominated by research from China, India, and the United States (V. Vo et al., 2022), there remains insufficient insight into how key research streams, such as forecasting, machine learning, volatility modeling, and sentiment analysis interact to shape emerging research directions.

Furthermore, although previous studies have demonstrated the effectiveness of machine learning techniques in improving financial forecasting accuracy, the literature remains fragmented across disciplines such as finance, computer science, and data analytics. This fragmentation obscures the intellectual structure of the field and limits understanding of the relationships among its evolving research themes. Therefore, a bibliometric approach becomes essential to systematically synthesize existing knowledge, identify dominant research clusters, and trace the developmental trajectory of machine learning applications in financial forecasting. Such mapping is crucial for informing future interdisciplinary studies and supporting the sustainable development of data-driven financial science.

3. Research Method

This study adopts a bibliometric research design aimed at systematically mapping the development, structure, and thematic evolution of studies on financial forecasting and stock price prediction using machine learning between 2015 and 2025. Bibliometric analysis is appropriate for this purpose as it facilitates a quantitative and visual examination of scientific production, co-authorship patterns, and intellectual connections within a specified research field (Zupic & Čater, 2015; Moral-Muñoz et al., 2020). Unlike traditional literature reviews that rely on narrative synthesis, bibliometric methods allow for the identification of emerging research fronts, core authors, and conceptual relationships through data-driven mapping techniques (Aria & Cuccurullo, 2017).

The research data were retrieved from the Scopus database, recognized for its comprehensive coverage of peer-reviewed journals and detailed metadata conducive to bibliometric analysis. Scopus was selected due to its extensive indexing of publications in the fields of economics, business, and computer science, thereby ensuring that the data reflect the interdisciplinary nature of financial forecasting research. The data collection process employed a Boolean search query specifically crafted to retrieve studies focusing on forecasting and stock price prediction. The Boolean query employed is as follows:

(TITLE-ABS-KEY (Financial Forecasting) AND TITLE-ABS-KEY (stock prices)) AND PUBYEAR > 2014 AND PUBYEAR < 2027 AND (LIMIT-TO (SUBJAREA, "ECON") OR LIMIT-TO (SUBJAREA, "BUSI")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (OA, "all"))

The initial search produced 5,470 records, which were subsequently filtered following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). Duplicate entries, inaccessible records, and irrelevant publications were excluded during the screening process. Titles and abstracts were carefully reviewed to ensure the inclusion of studies explicitly addressing financial forecasting and stock price prediction using computational or AI-based approaches. Following full-text screening, a total of 197 articles met the inclusion criteria and were included in the final analysis.

The inclusion criteria required that the documents: (1) be published in peer-reviewed journals between 2015 and 2025; (2) be written in English; (3) focus on financial forecasting or stock price prediction using machine learning or deep learning models; and (4) fall within the economics, finance, or business subject categories. Materials such as conference papers, book chapters, reviews, and non-English publications were excluded. This rigorous selection process ensured the validity, consistency, and analytical reliability of the final dataset for bibliometric evaluation.

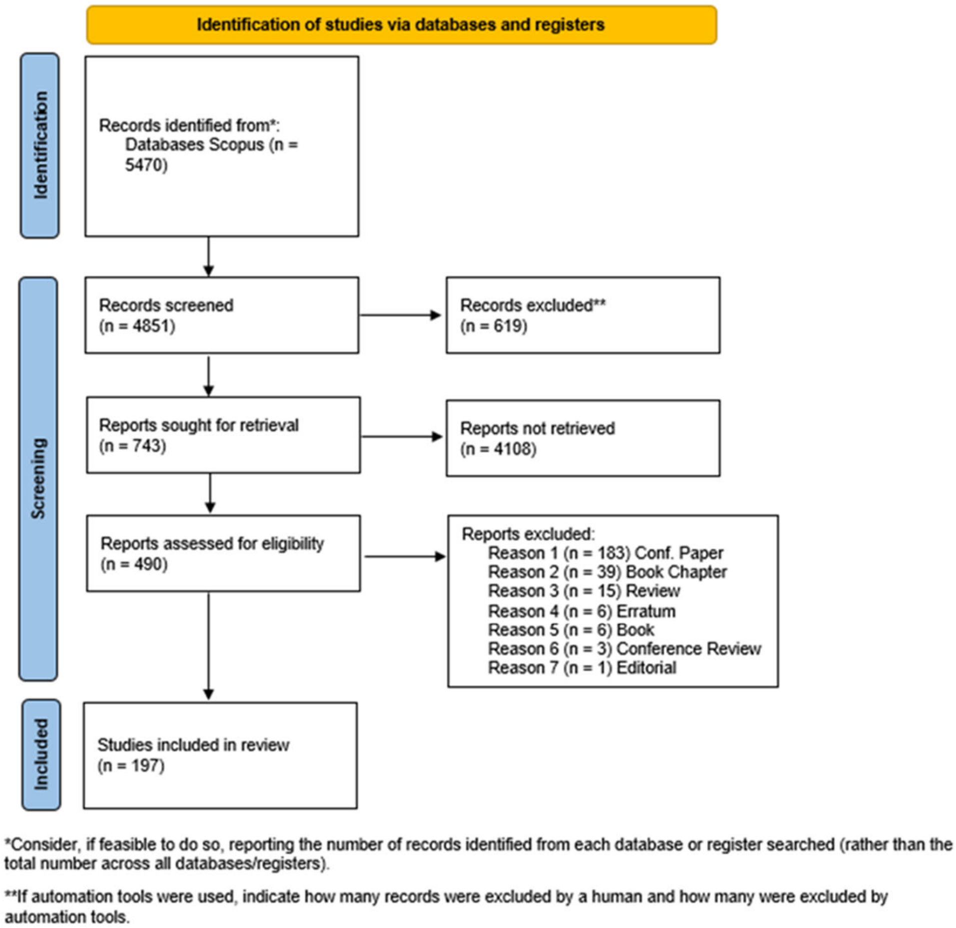

Figure 1. PRISMA Flow Diagram of the Research Screening Process
Source: Authors' elaboration

Data processing and analysis were conducted using R Studio (version 4.4.1) and the Bibliometrix package developed by Aria & Cuccurullo (2017). The analysis was systematically divided into three primary stages: (1) performance analysis, which examines productivity indicators such as annual publication growth, citation trends, and authorship patterns; (2) science mapping, which visualizes the intellectual and social structure of the field through co-citation, co-authorship, and keyword co-occurrence networks; and (3) thematic evolution analysis, which identifies the emergence, development, and decline of key research themes over time. The Biblioshiny web interface of Bibliometrix was also used to generate graphical representations of performance metrics, collaboration networks, and thematic maps.

To ensure methodological transparency, all stages of the analysis were documented using R Markdown files that recorded code scripts, data-cleaning steps, and parameter settings. Thematic clusters were identified using the Walktrap community detection algorithm, which groups related keywords into cohesive subfields based on their co-occurrence patterns. The positioning of themes was evaluated according to Callon's centrality and density measures, allowing for the classification of themes into four categories: motor themes (well-developed and central), niche themes (specialized but peripheral), emerging or declining themes (developing or fading), and basic themes (fundamental but underdeveloped).

Furthermore, the study applied Lotka's Law to examine author productivity and Bradford's Law to identify core journals constituting the knowledge base of the research domain (Cobo et al., 2011). To visualize keyword evolution and co-occurrence dynamics, a longitudinal analysis was conducted using time-sliced datasets divided into three periods: 2015–2020, 2021–2024, and 2025. This approach facilitated a comparative examination of shifts in thematic priorities over time, thereby reflecting the methodological and conceptual growth of the field.

In addition, ethical and reproducibility standards were rigorously adhered to throughout the research process. All data used were obtained from publicly accessible Scopus records, ensuring compliance with academic data usage guidelines. No personal or confidential information was involved. To promote reproducibility, all bibliometric procedures, from data retrieval to visualization, were documented as replicable R scripts and are available upon reasonable request from the corresponding author.

Through this methodological framework, the study provides a robust, transparent, and replicable mapping of the evolution of financial forecasting research in response to technological advances, external shocks, and growing interdisciplinarity. The bibliometric approach employed yields not only descriptive but also diagnostic value, enabling the identification of emerging research trajectories and opportunities for future inquiry within the domain of financial data analytics and predictive modeling. This approach is particularly valuable for accounting and finance researchers seeking to understand how predictive analytics and artificial intelligence are applied to financial reporting and performance forecasting.

4. Results and Discussion

The bibliometric analysis elucidates the dynamic progression of research on financial forecasting and stock price prediction using machine learning during the period 2015–2025. The findings indicate that the field has experienced significant expansion, characterized by increasing publication output, enhanced international collaboration, and the emergence of novel research themes indicative of technological and methodological advancements.

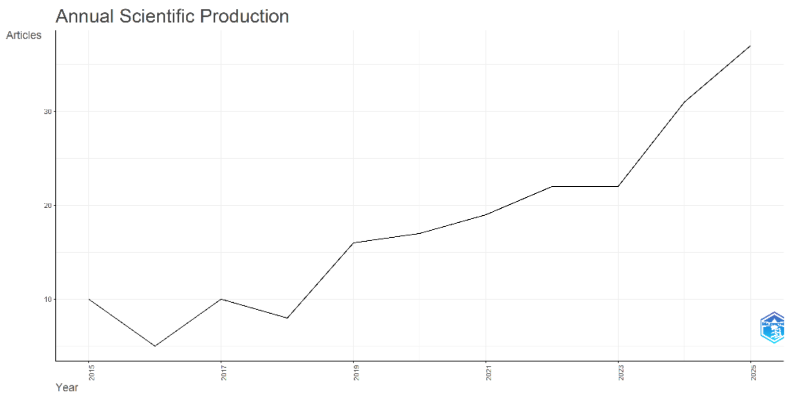

Figure 2. Annual scientific production.
Source: Authors' elaboration

The temporal growth trend demonstrates a consistent increase in scholarly output, characterized by a remarkable annual growth rate of 13.98 percent. During the period from 2015 to 2017, the number of publications remained relatively low, reflecting the early stage of research adoption in this domain. However, the period from 2018 to 2020 signifies the commencement of rapid development as machine learning techniques, particularly deep learning architectures such as LSTM and SVM, began to dominate the forecasting literature. This acceleration persisted beyond 2021, culminating in 2025 with over 35 publications annually. This upward trajectory underscores the field's growing academic significance and its responsiveness to both technological innovation and external global events.

Table 1. General Information About The Data

Description	Results
Main Information About Data	
Timespan	2015:2025
Sources (Journals, Books, etc.)	120
Documents	197
Annual Growth Rate %	13.98
Document Average Age	3.45
Average citations per doc	16.16
References	0
Document Contents	
Keywords Plus (ID)	508
Author's Keywords (DE)	783
Authors	
Authors	564
Authors of single-authored docs	19
Authors Collaboration	
Single-authored docs	19
Co-Authors per Doc	3.01
International co-authorships %	32.99
Document Types	
article	197

In examining research coverage and collaboration, a total of 197 articles from 120 different journals, authored by 564 researchers, were analyzed. The dataset indicates an average of 3.01 authors per document, confirming the highly collaborative nature of this research area. Only 19 papers were written by single authors. This indicates that financial forecasting has evolved into a team-oriented research domain that benefits from cross-disciplinary collaboration. The international co-authorship rate of 32.99% demonstrates the global interconnectedness of this field, suggesting that the application of machine learning in finance has surpassed regional boundaries, establishing itself as a genuinely international research endeavor.

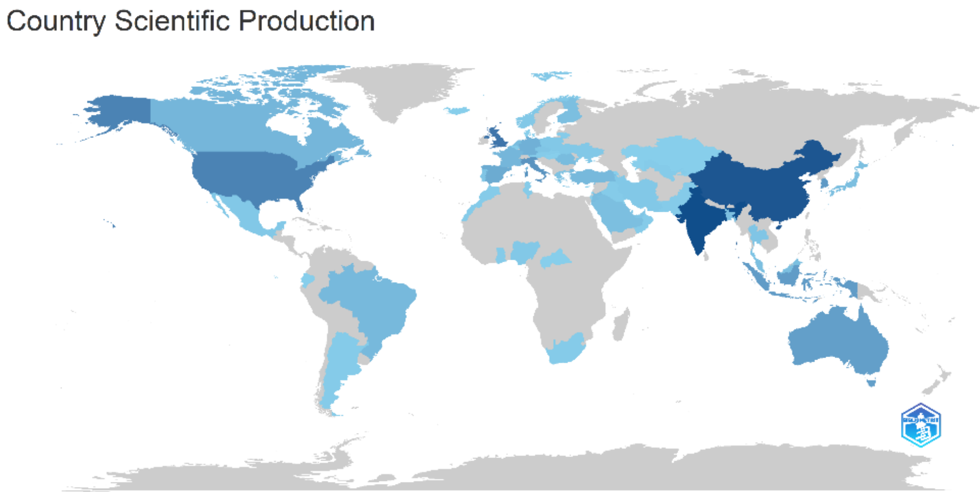

Figure 3. Most significant countries.
Source: Authors' elaboration.

The geographical distribution of scholarly publications highlights the concentration of research leadership within a few major economies. China and India emerge as the most prolific contributors, closely followed by the United States, the United Kingdom, and several European countries such as Germany, France, and Italy. This dominance is indicative of substantial national investments in AI research infrastructure and financial technology, consistent with previous bibliometric findings that identify China, India, and the United States as leaders in AI-based finance publications (Prasetyo et al., 2024). Meanwhile, emerging economies such as Indonesia, Brazil, and South Africa are beginning to demonstrate participation, indicating a diffusion of research interest into developing regions. The distribution pattern confirms that the knowledge base of financial forecasting is increasingly globalized, facilitated by diversified networks of collaboration and data-sharing practices.

Table 2. Articles with the highest number of citations.

Author	Paper	Total Citations (TC)	TC per Year
Feng et al. (2019)	Temporal Relational Ranking for Stock Prediction	378	54.00
Selvamuthu et al. (2019)	Indian stock market prediction using artificial neural networks on tick data	208	29.71
Xi Zhang et al. (2018)	Improving stock market prediction via heterogeneous information fusion	158	19.75
Atkins et al. (2018)	Financial news predicts stock market volatility better than close price	129	16.13
Ji et al. (2021)	A stock price prediction method based on deep learning technology	123	24.60
Ntakaris et al. (2018)	Benchmark dataset for mid-price forecasting of limit order book data with	102	12.75

	machine learning methods		
Shynkevich et al. (2016)	Forecasting movements of healthcare stock prices based on different categories of news articles using multiple kernel learning	76	7.60
Gupta & Wohar, (2017)	Forecasting oil and stock returns with a Qual VAR using over 150 years of data	66	7.33
Carriero et al. (2015)	Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility	65	5.91
Hung (2019)	Return and volatility spillover across equity markets between China and Southeast Asian countries	63	9.00

Source: Authors' elaboration.

Feng et al. (2019), as published in ACM Transactions on Information Systems, are recognized as the most globally cited authors, with a total of 378 citations. They are followed by Selvamuthu et al. (2019) in Financial Innovation and Xi Zhang et al. (2018) in Knowledge-Based Systems. These authors' works are foundational, focusing on the integration of deep learning and artificial intelligence into financial forecasting models. Their influence underscores the significance of research combining knowledge-based systems, AI techniques, and financial prediction in shaping the methodological core of the field.

Regarding author impact and citation influence, the analysis identifies Li B, Misra BB, and Nayak SC as the most influential contributors based on their bibliometric indicators, including the h-index, g-index, and total citations. Misra BB and Nayak SC each recorded 113 citations with an h-index of 3 and an m-index of 0.375, indicating their consistent scholarly impact since 2018. Li B, active since 2016, has accumulated 40 citations, reflecting moderate yet sustained contributions to the field. In addition, Chen S, who began publishing in 2019, demonstrates emerging potential with an h-index of 2 and 25 citations, suggesting early but meaningful engagement in this research domain.

Table 3. Most Relevant Authors

Author	h_index	g_index	m_index	TC	NP	PY_start
Li B	3	3	0.3	40	3	2016
Misra BB	3	3	0.375	113	3	2018
Nayak SC	3	3	0.375	113	3	2018
Chen S	2	2	0.286	25	2	2019
Fiszeder P	2	2	0.5	43	2	2022
Gupta R	2	2	0.222	69	2	2017
Kim H	2	2	0.25	51	2	2018
Lee J	2	2	0.286	12	2	2019
Liu Y	2	2	0.286	380	2	2019
Moon K	2	2	0.25	51	2	2018

Source: Authors' elaboration.

Overall, the findings highlight a growing and dynamic network of researchers contributing to the advancement of financial forecasting and stock price prediction using machine learning between 2015 and 2025. The observed citation patterns and author-level metrics illustrate both the steady rise of emerging scholars and the consolidation of a global research agenda focused on integrating machine learning into financial forecasting. This growing author base underscores the transition of the field from an emerging niche into a more established and impactful research frontier in the broader landscape of financial and economic sciences.

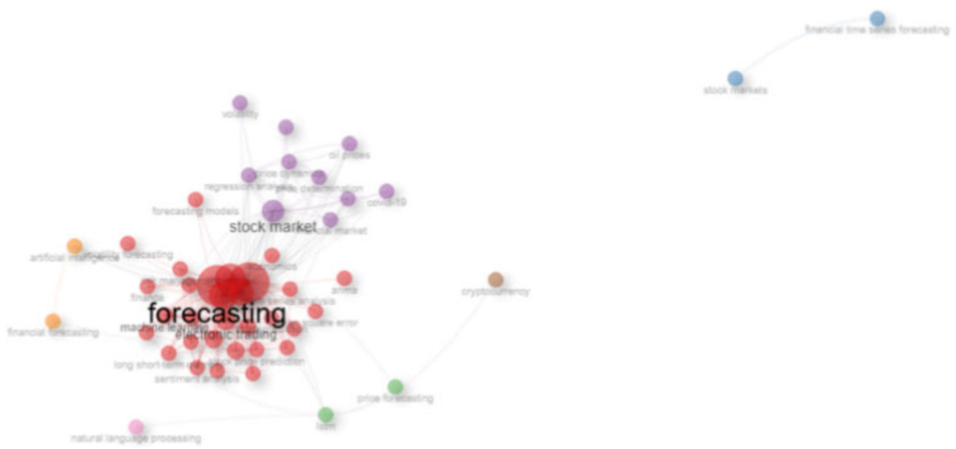

Figure 4. Co-occurrence Keywords Analysis
Source: Authors' elaboration

The network analysis of keywords provides valuable insight into the intellectual structure of the research domain. The co-occurrence map indicates that "forecasting" remains the central and highly interconnected term, linked strongly with "machine learning," "stock market," "financial markets," and "electronic trading." Using the Walktrap community detection algorithm, several distinct thematic clusters were identified. The largest cluster (red) represents the core forecasting community, encompassing topics such as time series analysis, LSTM, deep learning, and neural networks, thereby signifying the dominance of computational intelligence in predictive finance. The purple cluster centers on the stock market dynamics and related determinants, including volatility, regression analysis, and COVID-19. This reflects the role of external shocks and market behavior in modeling efforts. The green cluster is primarily associated with LSTM-based price forecasting, while the orange cluster centers on financial forecasting and artificial intelligence, illustrating the increasing integration of traditional statistical methods and AI-driven approaches.

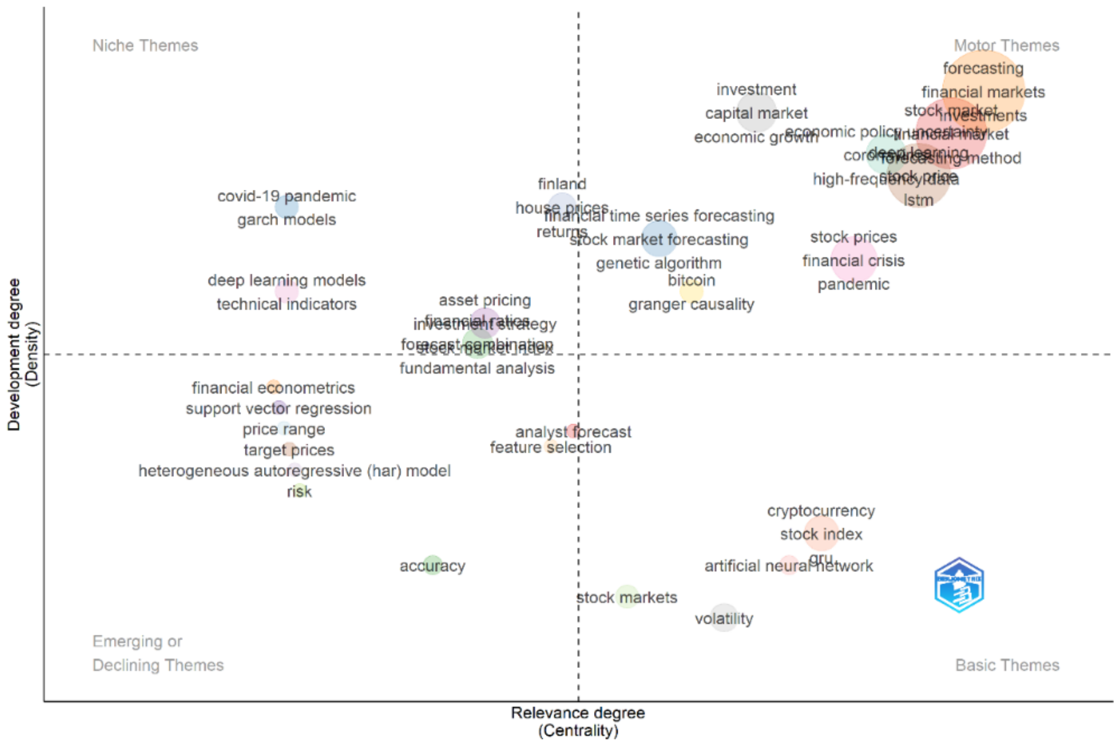

Figure 5. Thematic map

Source: Authors' elaboration

Table 4. Interpretation of Thematic Clusters

Theme	Key Keywords	Interpretation
Motor	forecasting, financial markets, stock markets, LSTM	Core research theme focusing on predictive modeling of financial assets
Niche	GARCH, deep learning, investment strategy	Computational intelligence techniques used for nonlinear financial prediction
Basic	cryptocurrency, volatility, GRU, neural network	Traditional econometric approaches for modeling financial risk
Emerging	support vector regression, financial econometrics	New research directions integrating alternative data sources

Thematic mapping results further classify research directions based on their centrality and density to provide a structured view of the field's development. The motor themes (upper-right quadrant) include "forecasting," "financial markets," "stock markets," and "LSTM", representing the most developed and central research areas. In contrast, the niche themes (upper-left quadrant) comprise topics such as "GARCH models," "investment strategy," and "deep learning," which are specialized but relatively isolated from the broader network. The basic themes (lower-right quadrant) include "volatility," "cryptocurrency," "GRU," and "neural network," indicating fundamental but still developing areas of study. Meanwhile, emerging or declining themes (lower-left quadrant) such as "financial econometrics" and "support vector regression" may reflect either diminishing scholarly attention or nascent topics with potential further exploration in the future. This quadrantal structure demonstrates that the central role of forecasting and machine learning remains as the intellectual core of the field, while also highlighting the evolution of emerging research areas.

Figure 6. Evolution of themes from 2015 to 2025.
Source: Authors' elaboration

Longitudinal analysis of thematic evolution across three time periods, 2015–2020, 2021–2024, and 2025, demonstrates the transformation in research focus over time. During 2015–2020, the dominant themes revolved around forecasting, financial markets, and investments, with econometric and volatility-based methods serving as the analytical foundation. In the subsequent period of 2021–2024, the thematic scope expanded significantly with the incorporation of machine learning, deep learning, convolutional neural networks (CNNs), and topics such as commerce and decision-making. This indicates a shift towards more data-driven and application-oriented domains. The emergence of COVID-19 and cryptocurrency during this phase underscores the responsiveness of the research community to real-world events and market innovations. By 2025, forecasting and financial markets persisted as the core themes, while deep learning emerged as a fundamental yet rapidly growing foundation, and cryptocurrency appeared as an emerging yet potentially volatile topic. This pattern confirms the continuous adaptation of financial forecasting research to technological and contextual developments, echoing recent evidence that the field continues to expand toward hybrid and cross-disciplinary forecasting models (V. Vo et al., 2022).

Based on the bibliometric findings, several promising directions for future research can be identified. The thematic mapping and longitudinal analysis reveal not only the current structure of the field but also its potential trajectory.

First, the increasing prominence of machine learning and deep learning techniques, particularly LSTM and neural networks, suggests that future research should prioritize enhancing model interpretability and explainability. Even though predictive accuracy has significantly improved, the "black-box" nature of deep learning models remains a critical limitation, especially in financial decision-making contexts that require transparency and accountability. Therefore, the development of explainable artificial intelligence (XAI) in finance emerges as a critical research priority.

Second, the emergence of themes such as sentiment analysis, natural language processing, and cryptocurrency forecasting signifies a shift toward integrating alternative and unstructured data sources. Future studies are encouraged to explore multimodal forecasting frameworks that integrate numerical financial data with textual information derived from news, social media, and financial reports. This integration can potentially improve predictive performance and better capture market dynamics driven by investor sentiment.

Third, the thematic evolution analysis highlights a transition from traditional econometric models toward hybrid approaches that combine statistical and machine learning techniques. Future research should further investigate hybrid and ensemble models that capitalize on the strengths of both paradigms, particularly under volatile and uncertain market conditions. Comparative and benchmarking studies evaluating the performance of econometric, machine learning, and hybrid models across different market regimes would yield valuable insights for both academia and industry.

Fourth, the global collaboration patterns identified in this study suggest opportunities for expanding research participation in emerging economies. Future research should examine how contextual factors such as market structure, regulatory environments, and data availability influence the performance of AI-based forecasting models in developing countries. This is particularly relevant for regions such as Southeast Asia, including Indonesia, where financial data science remains at an early stage of development.

Finally, the bibliometric results indicate that most existing studies focus on stock markets, while other financial domains such as corporate finance, financial reporting, and risk management remain underexplored. Future research can extend the application of machine learning forecasting models to these areas, thereby broadening the scope of predictive analytics in accounting and finance. Overall, these future research directions highlight the need for interdisciplinary, data-driven, and context-sensitive approaches to advance both the theoretical and practical development of financial forecasting in the era of artificial intelligence.

5. Conclusion, Implications, and Limitations

This bibliometric study offers an extensive overview of the intellectual framework, methodological advancements, and global research trends in the field of financial forecasting and stock price prediction using machine learning from 2015 to 2025. The findings confirm that forecasting remains a pivotal and ever-evolving research theme, supported by the increasing adoption of artificial intelligence and machine learning techniques. Over the past decade, this research domain has transformed from relying on traditional econometric models, such as ARIMA and GARCH, into an AI-driven discipline dominated by deep learning architectures, including LSTM and CNN. Such a shift reflects a broader transformation in financial research, where computational intelligence is increasingly leveraged to model nonlinearity, volatility, and behavioral aspects of financial markets.

The results further underscore the globalization of financial forecasting research, with China, India, and the United States emerging as predominant contributors. This geographical diversification indicates the democratization of knowledge production and the diffusion of AI expertise across regions. The increasing emphasis on interdisciplinary collaboration, evidenced by an average of three co-authors per publication and a 33% rate of international partnerships, underscores the field's integrative nature, bridging economics, computer science, and data analytics. In addition, external shocks such as the COVID-19 pandemic have prompted thematic diversification, as reflected in the emergence of topics such as volatility forecasting, sentiment analysis, and cryptocurrency prediction. These developments indicate that financial forecasting research not only adapts to technological innovation but also responds dynamically to global economic and social challenges.

Theoretically, this study contributes to the understanding of the evolution of financial forecasting as an interdisciplinary science. The persistent dominance of "forecasting" as a central theme demonstrates its theoretical significance, linking traditional financial theories such as the Efficient Market Hypothesis (EMH) with adaptive and probabilistic models rooted in artificial intelligence. The increasing prominence of machine learning and deep learning methods suggests a growing transition toward data-driven financial approaches that complement traditional financial theory, thereby reinforcing the hybridization of econometric reasoning and explainable AI frameworks (Fang & Wang, 2024). Furthermore, the integration of Bayesian reasoning, neural networks, and hybrid approaches reflects an epistemological convergence between statistical inference and computational intelligence, thereby expanding the theoretical foundation of predictive finance.

From a practical standpoint, the findings emphasize the transformative potential of AI-driven forecasting tools in supporting financial decision-making, investment management, and policy development. The increasing adoption of machine learning and deep learning algorithms boosts predictive accuracy and facilitates the identification of concealed market patterns. This directly affects financial institutions, investors, and regulators seeking to improve portfolio optimization, risk mitigation, and market monitoring. The emergence of topics such as cryptocurrency forecasting and sentiment analysis further demonstrates the growing influence of alternative data sources such as social media, financial news, and digital transactions in shaping predictive financial analytics. Policymakers are therefore encouraged to consider the implications of algorithmic decision-making for market transparency, ethical governance, and financial stability. Enhanced collaboration between academia, industry, and regulators is essential to ensure that AI applications in forecasting are not only efficient but also transparent, accountable, and equitable, aligning with emerging discussions on ethical and responsible AI adoption in financial decision-making (Raghuvanshi, 2025).

The study identifies several promising directions for future research. First, greater emphasis should be placed on developing explainable AI (XAI) frameworks to improve the interpretability, transparency, and accountability of forecasting models, particularly in regulated financial environments. Second, researchers should explore hybrid modeling approaches that combine the robustness and theoretical grounding of econometric techniques with the adaptability and learning capacity of neural networks, enabling a balance between transparency and accuracy. Third, longitudinal and cross-market comparative studies across emerging and developed markets could deepen understanding of how contextual factors influence the performance of AI-based forecasting systems. Fourth, the incorporation of natural language processing (NLP) and sentiment-driven features should be expanded to capture behavioral dimensions of market movements. Finally, extending bibliometric analyses by incorporating additional databases such as Web of Science and Dimensions would enhance the generalizability of results and allow a more complete understanding of the knowledge diffusion and intellectual structure of financial forecasting research.

In summary, this study reaffirms that forecasting remains the intellectual and methodological cornerstone of modern financial research. The integration of machine learning and artificial intelligence represents not merely a technical enhancement but a fundamental transformation of financial modeling and theory-building. By combining computational sophistication with theoretical insight, future research can advance the development of transparent, adaptive, and data-driven forecasting systems that enhance decision-making and contribute to the long-term sustainability of global financial markets. Within the accounting discipline, the adoption of AI-enhanced forecasting techniques supports evidence-based financial reporting and auditing practices, offering predictive insights that strengthen governance and reduce information asymmetry.

For policymakers and financial institutions, understanding these global research patterns can inform strategic policy formulation and foster the advancement of AI-enabled financial innovation. Insights from this bibliometric study can support the development of data-driven financial ecosystems, particularly in emerging economies such as Indonesia, where evidence-based forecasting is essential for sustainable economic planning and regulatory decision-making.

Future studies may extend this analysis by incorporating advanced bibliometric techniques, such as citation-based clustering, co-citation network mapping, or topic modeling, to explore deeper intellectual linkages across disciplines. Integrating bibliometric findings with text mining techniques or econometric validation would enhance the analytical robustness and practical applicability of future financial forecasting research.

Despite its contributions, this study has several limitations. First, the analysis relies solely on the Scopus database, which may exclude relevant publications indexed in other databases such as Web of Science or Dimensions. Second, bibliometric analysis focuses on publication metadata and citation relationships rather than evaluating the empirical performance of forecasting models. Therefore, future research may integrate bibliometric mapping with systematic literature reviews or meta-analytic approaches to provide more comprehensive insights into methodological effectiveness and theoretical development.

Acknowledgements

The authors would like to express their gratitude to Mr. Nugraha and Mrs. Maya for their invaluable guidance, supervision, and support throughout the research process and the completion of this study.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Data Availability Statement

The data supporting the findings of this study are available from www.scopus.com.

Author Contributions

Iwan Kurniawan conducted the research, performed data analysis, and drafted the manuscript.

Nugraha and Maya Sari served as supervisors, providing conceptual guidance, reviewing the analysis, and contributing to the final revision of the manuscript.

ORCID iDs

First Author: https://orcid.org/0000-0002-5464-2784
Second Author: https://orcid.org/0000-0002-6041-3574
Third Author: https://orcid.org/0000-0002-1512-3017

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/https://doi.org/10.1016/j.joi.2017.08.007

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, *4*(2), 120–137. https://doi.org/https://doi.org/10.1016/j.jfds.2018.02.002

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343. https://doi.org/https://doi.org/10.1016/S0304-405X(98)00027-0

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/https://doi.org/10.1016/j.jocs.2010.12.007

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/https://doi.org/10.1016/0304-4076(86)90063-1

Carriero, A., Clark, T. E., & Marcellino, M. (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *178*(4), 837–862. https://doi.org/https://doi.org/10.1111/rssa.12092

Chen, W., Hussain, W., Cauteruccio, F., & Zhang, X. (2023). Deep learning for financial time series prediction: A state-of-the-art review of standalone and hybrid models.

CMES-Computer Modeling in Engineering and Sciences, *139*(1), 187–224. https://doi.org/10.32604/cmes.2023.031388

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146–166. https://doi.org/https://doi.org/10.1016/j.joi.2010.10.002

Corbet, S., Lucey, B. M., & Yarovaya, L. (2019). The financial market effects of cryptocurrency energy usage. *Available at SSRN 3412194*, 1–13. https://doi.org/http://dx.doi.org/10.2139/ssrn.3412194

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/https://doi.org/10.1016/j.jbusres.2021.04.070

Dzigbede, K. D., & Pathak, R. (2020). COVID-19 economic shocks and fiscal policy options for Ghana. *Journal of Public Budgeting, Accounting & Financial Management*, *32*(5), 903–917. https://doi.org/https://doi.org/10.1108/JPBAFM-07-2020-0127

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987–1007. https://doi.org/https://doi.org/10.2307/1912773

Fama, E. F. (1970). Efficient capital markets. *Journal of Finance*, *25*(2), 383–417.

Fang, Z., & Wang, S. (2024). Boosting financial market prediction accuracy with deep learning and big data: Introducing the CCL model. *Journal of Organizational and End User Computing (JOEUC)*, *36*(1), 1–25. https://doi.org/10.4018/JOEUC.358454

Feng, F., He, X., Wang, X., Luo, C., Liu, Y., & Chua, T.-S. (2019). Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)*, *37*(2), 1–30. https://doi.org/https://doi.org/10.1145/3309547

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/https://doi.org/10.1016/j.ejor.2017.11.054

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90. https://doi.org/https://doi.org/10.1080/07474930600972467

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/https://doi.org/10.1093/rfs/hhaa009

Gupta, R., & Wohar, M. (2017). Forecasting oil and stock returns with a Qual VAR using over 150 years off data. *Energy Economics*, *62*, 181–186. https://doi.org/https://doi.org/10.1016/j.eneco.2017.01.001

Hasbrouck, J. (2004). Empirical market microstructure. *Economic and Statistical*

Perspectives on the Dynamics of Trade in Securities Markets.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *An introduction to statistical learning*.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. https://doi.org/https://doi.org/10.1016/j.eswa.2019.01.012

Hung, N. T. (2019). Return and volatility spillover across equity markets between China and Southeast Asian countries. *Journal of Economics, Finance and Administrative Science*, *24*(47), 66–81. https://doi.org/https://doi.org/10.1108/JEFAS-10-2018-0106

Ji, X., Wang, J., & Yan, Z. (2021). A stock price prediction method based on deep learning technology. *International Journal of Crowd Science*, *5*(1), 55–72. https://doi.org/10.1108/IJCS-05-2020-0012

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *ArXiv Preprint ArXiv:1706.10059*, 1–31. https://doi.org/https://doi.org/10.48550/arXiv.1706.10059

Li, T., Wang, L., & Xu, W. (2025). Financial forecasting and decision-making models based on intelligent algorithms and big data. *International Conference on Big Data Analytics for Cyber-Physical System in Smart City*, 69–78. https://doi.org/https://doi.org/10.1007/978-981-96-0208-7_7

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, Forthcoming*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.04.014

Marisetty, N. (2024). Prediction of popular global stock indexes volatility by using ARCH/GARCH models. *GARCH Models (July 24, 2024)*, 1–19. https://doi.org/http://dx.doi.org/10.2139/ssrn.4904475

Meher, B. K., Puntambekar, G. L., Birau, R., Hawaldar, I. T., Spulbar, C., & Simion, M. L. (2023). Comparative investment decisions in emerging textile and FinTech industries in India using GARCH models with high-frequency data. *Industria Textila*, *74*(6), 741–752. https://doi.org/10.35530/IT.074.06.202311

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de La Información*, *29*(1), 1–20. https://doi.org/10.3145/epi.2020.ene.03

Ntakaris, A., Magris, M., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2018). Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. *Journal of Forecasting*, *37*(8), 852–866. https://doi.org/https://doi.org/10.1002/for.2543

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*, 1–36. https://doi.org/https://doi.org/10.1136/bmj.n71

Raghuvanshi, A. (2025). The ethical dimensions of AI in financial decision-making: Balancing innovation and equity. *Journal of Computer Science and Technology Studies*, *7*(5), 220–227. https://doi.org/10.32996/jcsts.2025.7.5.28

Selvamuthu, D., Kumar, V., & Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation*, *5*(16), 1–12. https://doi.org/https://doi.org/10.1186/s40854-019-0131-7

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, *90*, 1-63. https://doi.org/10.48550/arXiv.1911.13288

Sholapurapu, P. K. (2025). AI-driven financial Forecasting: Enhancing predictive accuracy in volatile markets. *SSRN Electronic Journal*, *15*(2), 1282–1291. https://doi.org/10.2139/ssrn.5331686

Shynkevich, Y., McGinnity, T. M., Coleman, S. A., & Belatreche, A. (2016). Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning. *Decision Support Systems*, *85*, 74–83. https://doi.org/https://doi.org/10.1016/j.dss.2016.03.001

Sirignano, J. A. (2019). Deep learning for limit order books. *Quantitative Finance*, *19*(4), 549–570. https://doi.org/https://doi.org/10.1080/14697688.2018.1546053

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tian, M., Li, H., Huang, J., Liang, J., Bu, W., & Chen, B. (2022). Credit risk models using rule-based methods and machine-learning algorithms. *Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence*, 203–209. https://doi.org/https://doi.org/10.1145/3577530.357758

V. Vo, H., Nguyen, D. H., Nguyen, T. T., Nguyen, H. N., & Nguyen, D. V. (2022). Leveraging AI-driven realtime intrusion detection by using wgan and xgboost. *Proceedings of the 11th International Symposium on Information and Communication Technology*, 208–215. https://doi.org/https://doi.org/10.1145/3568562.3568660

Zhang, Xiaolin, & Tan, Y. (2018). Deep stock ranker: A LSTM neural network model for stock selection. *International Conference on Data Mining and Big Data*, 614–623. https://doi.org/https://doi.org/10.1007/978-3-319-93803-5_58

Zhang, Xi, Zhang, Y., Wang, S., Yao, Y., Fang, B., & Yu, P. S. (2018). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems*, *143*, 236–247. https://doi.org/https://doi.org/10.1016/j.knosys.2017.12.025

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/https://doi.org/10.1177/1094428114562629