

Machine Learning Applications in Predicting Stock Returns: A Scientific and Literary Approach to Modern Methods

Atallah BENTIRECHE^{1,*}

¹University of Laghouat (Algeria), at.bentireche@lagh-univ.dz

Received date: 27/03/2026;

Revised date: 27/06/2026 ;

Publication date: 30/06/2026

Summary: The global financial landscape is undergoing a structural shift driven by digital transformation and the integration of high-velocity big data. Traditional econometric models are increasingly supplemented or replaced by machine learning (ML) architectures that can process non-linear relationships. Despite the proliferation of AI, economic actors suffer from "bounded rationality," where cognitive and informational constraints hinder optimal decision-making. Furthermore, market volatility and the "black box" nature of advanced algorithms create risks for institutional and retail investors. This study aims to analyze the efficacy of modern ML methods—such as Random Forest, XGBoost, and Deep Neural Networks—in predicting stock returns and identifying market anomalies while evaluating their impact on financial market efficiency. Drawing on a multi-dimensional bibliometric analysis and comparative empirical studies across diverse markets (US, China, South Africa, and Algeria), the research evaluates performance through metrics like Accuracy, F1 Score, and the Matthews Correlation Coefficient (MCC). Findings indicate that ensemble methods like Random Forest and gradient-boosted models consistently outperform linear regressions by 5–15% in predictive accuracy. Additionally, deep learning models achieve up to 93.8% accuracy in stock classification by incorporating financial ratios from 10-K reports. This research bridges the gap between behavioral economics (bounded rationality) and computational finance, proposing a hybrid framework for "dynamic efficiency" in modern markets.

Keywords: Machine Learning, Stock Returns, Big Data Analytics, Algorithmic Trading, Financial Transparency, Bounded Rationality.

* Corresponding author, Atallah BENTIRECHE, at.bentireche@lagh-univ.dz

I- Introduction :

Background :

The evolution of Artificial Intelligence (AI) from a theoretical concept in the mid-twentieth century to a cornerstone of modern innovation has reshaped organizational practices (Aissat, 2026, p. 44). In the financial sector, AI applications have seen a remarkable increase since 2017, driven by the explosion of computing power and the availability of large-scale unstructured datasets (MERMI & KHACEF, 2025, pp. 18-22).

Research Gap:

While traditional rating and forecasting methodologies have relied on subjective judgment and linear assumptions, they often fail to capture the complex, non-linear interactions inherent in financial data (GBADEBO, 2025, p. 02). Current literature shows a disconnect between advanced technical tools and their strategic alignment within institutional frameworks, leading to a "fragmented" scholarly landscape (Smaali & Boubaa, 2025, p. 193).

Problem Statement:

Economic actors face information asymmetry and cognitive biases that limit their ability to achieve the "perfect rationality" assumed in classical mode (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). This study addresses the core question: To what extent do AI-driven tools—specifically ensemble and deep learning models—enhance the accuracy of stock return predictions and contribute to financial market efficiency?

Objectives:

- (1) To clarify the conceptual framework of AI in finance.
- (2) To evaluate the predictive power of ML classifiers across different economic sectors.
- (3) To examine the role of AI in reducing information asymmetry and identifying fraudulent financial practices.

Research Hypotheses:

- (H1) Machine learning techniques represent non-linear relationships better than traditional models, leading to superior predictive accuracy (SAAD BEKHOUCHE et al., 2025, p. 85) .
- (H2) The integration of big data analytics and AI significantly enhances financial risk management by providing proactive early-warning signals (SAAD BEKHOUCHE et al., 2025, p. 85)

Significance:

This study is critical for investors, regulators, and policymakers in emerging and developed markets alike, as it provides evidence-based recommendations for integrating "explainable" AI into lending and investment practices.

Brief Methodology:

This paper utilizes an interdisciplinary approach, synthesizing bibliometric data, empirical case studies of corporate entities (e.g., Apple Inc., ACS Holding), and comparative performance analysis of statistical vs. algorithmic models.

Structure :

The paper follows a logical progression from historical theory to technical implementation, followed by global case studies and a concluding synthesis of future trends.

Section I: Theoretical Foundations and the Epistemological Evolution of Financial AI

The historical genesis and conceptual architecture of Artificial Intelligence (AI) represent a multidisciplinary convergence of computer science, mathematics, logic, and cognitive science, aimed at creating systems capable of simulating human-like intelligent behavior (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). The formal inception of the field is historically anchored to the 1956 Dartmouth Summer Research Project on Artificial Intelligence, organized by pioneers such as John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). During this foundational era, researchers operated under the optimistic conviction that every facet of learning and intelligence could be articulated with such precision that a machine could be engineered to simulate it (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). This early paradigm, often designated as "Classical AI," prioritized symbolic reasoning and rule-based logic to solve mathematical theorems, play complex games like chess, and address structured logical puzzles (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). However, these inaugural systems were profoundly constrained by the limited computational hardware of the mid-twentieth century and an inherent inability to manage the stochastic uncertainty and nuanced complexities of real-world environments (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). Conceptually, AI is defined through various scholarly lenses: as a scientific discipline focused on designing machines for reasoning and problem-solving (McCarthy, 2007, as cited in Aissat, 2025, p. 45); as the study of autonomous agents that perceive their environment and act to maximize goal achievement (Russell & Norvig, 2021, as cited in Aissat, 2025, p. 45); and as a suite of techniques enabling machines to learn from data and identify patterns with minimal human intervention (Alpaydin, 2020, as cited in Aissat, 2025, p. 46). These evolving definitions underscore a shift from rigid symbolic logic to a flexible, data-driven framework that now underpins modern organizational and economic contexts (OECD, 2019, as cited in Aissat, 2025, p. 46).

The developmental trajectory of AI was characterized by periods of significant stagnation and intermittent "AI winters," driven primarily by severe funding shortages and a disconnect between theoretical aspirations and existing storage and processing capacities (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). These periods of decline—specifically from 1974 to 1980 and again from 1987 to 1993—stifled progress until the turn of the twenty-first century (Bahoo et al., 2024, as cited in Mermi & Khacef, 2025, p. 19). The subsequent revitalization of the field, particularly since 2017, has been fueled by an unprecedented explosion in computing power, the emergence of high-velocity big data, and a surge in venture capital investments for technological innovation (Mermi & Khacef, 2025, p. 19; Ernst et al., 2018, as cited in Mermi & Khacef, 2025, p. 19). Within the financial domain, the integration of AI technologies has transitioned from basic automation to an advanced analytical ecosystem where Machine Learning (ML) and Deep Learning (DL) models are deployed to manage vast, unstructured datasets that characterize modern global markets (Mermi & Khacef, 2025, p. 20; Djennad, 2025, p. 407). Big data in finance is fundamentally defined by five key dimensions: Volume (the massive scale of tick-by-tick trading data and macroeconomic indicators), Velocity (the necessity for real-time processing where milliseconds dictate profit or loss), Variety (the diverse formats of numerical, textual, and multimedia content), Veracity (the critical reliability

of input data), and Value (the extraction of actionable insights that shape investment strategies) (Iseal, Sheed., Joseph, Oluwaseyi, & Joseph, Shalom, 2025, as cited in Djennad, 2025, p. 409). This digital transformation has reshaped organizational agility, enabling financial institutions to respond dynamically to market volatility and align strategic goals with real-time computational capabilities (Smaali & Boubaa, 2025, p. 181).

From an epistemological perspective, the evolution of financial AI reflects a structural shift from traditional linear econometric methodologies to algorithmic reasoning capable of processing non-linear interactions. For decades, traditional models like simple linear regression served as the cornerstone of economic forecasting, yet they often failed to capture the intricate, multi-dimensional patterns inherent in complex financial data (Messaoudi, 2024, p. 683; Medini, 2024, p. 574). Classical regression techniques typically treat datasets as a single, static block and rely on the method of least squares to estimate standard relationships (Medini, 2024, p. 574). In contrast, Machine Learning models adopt a data-driven philosophy that partitions information into discrete training and testing sets to ensure predictive validity and mitigate biases (Medini, 2024, p. 586; Messaoudi, 2024, p. 691). These modern architectures, such as Random Forest (RF) and Gradient Boosting Machines (GBM), are designed to mitigate overfitting and enhance model interpretability while autonomously improving performance through iterative learning cycles (Messaoudi, 2024, p. 684; Djennad, 2025, p. 379). Unlike early expert systems that relied on rigid, rule-based knowledge bases that were difficult to update, contemporary ML algorithms can analyze millions of data points simultaneously to uncover hidden trends that would remain inaccessible through manual quantification (Aissat, 2025, p. 48; Medini, 2024, p. 582). This evolution represents a transition toward "Dynamic Efficiency," where the intrinsic value of a market participant is defined not just by the acquisition of information, but by the algorithmic capacity to process it with maximum accuracy and speed (Nechachda, 2026, p. 43).

The theoretical integration of AI in economic decision-making directly addresses the fundamental problem of "Bounded Rationality" as conceptualized by Herbert Simon. Simon challenged the classical economic assumption of "perfect rationality," arguing that human actors are inherently constrained by cognitive, informational, financial, and temporal limitations (Kehri et al., 2024, p. 27; Kehri et al., 2024, p. 28). Informational constraints involve delays in data disclosure and the high costs of acquisition, while cognitive constraints refer to biases—such as overconfidence or loss aversion—that impair objective judgment (Kehri et al., 2024, p. 28). AI agents compensate for these human deficiencies by processing massive datasets objectively and identifying complex patterns at scales that surpass biological cognitive limits (Kehri et al., 2024, p. 31; Aissat, 2025, p. 43). This human-AI synergy is described as "Cognitive Complementarity," where AI systems automate repetitive, tedious tasks, thereby freeing human decision-makers to focus on high-value activities requiring creativity, intuition, and strategic foresight (Kehri et al., 2024, p. 34; Djennad, 2025, p. 425). However, this duality is not without risk; over-reliance on automated systems can lead to "Disempowerment," where individuals blindly follow algorithmic recommendations without critical discernment, potentially resulting in an atrophy of human cognitive skills over time (Kehri et al., 2024, p. 38, 39). Therefore, the modern financial landscape requires a balanced approach where AI serves to reinforce rationality rather than replace it, ensuring

that decision-making remains ethically grounded and contextually aware (Kehri et al., 2024, p. 43; Aissat, 2025, p. 171).

Finally, the evolution of financial AI has profound implications for the Efficient Market Hypothesis (EMH), transforming it into a dynamic framework consistent with the digital age. In its classical form, EMH posits that asset prices fully reflect all available information, appearing in weak, semi-strong, and strong forms (Nechachda, 2026, p. 42). However, in an era dominated by high-frequency trading (HFT) and alternative data, the decisive factor in market efficiency is the ability to process information rather than simply obtain it (Nechachda, 2026, p. 43). Artificial Intelligence has fundamentally reshaped price discovery mechanisms by reducing information asymmetry—the disparity in data access and accuracy that leads to market distortions (Nechachda, 2026, p. 44). Global evidence indicates that markets employing AI and ML tools adjust prices 35% faster to unexpected events than those relying on traditional methods (Bank for International Settlements, 2024, as cited in Nechachda, 2026, p. 43). Technologies such as Sentiment Analysis, which scan millions of social media posts and news articles in seconds, have reduced market reaction times from 30 minutes to under 5 minutes (FINRA, 2023, as cited in Nechachda, 2026, p. 45). Furthermore, AI-enhanced Regulatory Technology (RegTech) now allows for real-time monitoring of transactions to detect insider trading and market manipulation, thereby strengthening the semi-strong form of efficiency and fostering institutional trust (Nechachda, 2026, p. 49). This evolution confirms that while AI enhances market transparency, it also introduces systemic risks like "Flash Crashes," requiring a proactive regulatory approach to ensure that technological innovation contributes to inclusive and sustainable financial development (Nechachda, 2026, p. 43; Gbadebo, 2025, p. 12).

Section II: Comparative Performance Analysis of Machine Learning Classifiers and Ensemble Methods

The structural transition from traditional econometric frameworks to algorithmic forecasting represents a fundamental shift in the epistemological approach to financial market analysis. For decades, simple linear models and ordinary least squares (OLS) regressions served as the primary tools for estimating relationships between variables, yet these classical methods are increasingly scrutinized for their inability to handle the "structural turbulence" and non-linear complexities inherent in high-velocity financial data (Medini, 2024, p. 574; Saad Bekhouche et al., 2025, p. 1002). Empirical evaluations consistently indicate that machine learning (ML) architectures provide a more robust mechanism for stock return prediction because they do not rely on the rigid assumption of normality or the treatment of data as a single, static block (Medini, 2024, p. 574; Bekaddour, 2025, p. 750). In a comprehensive study of listed companies on the US market, Bekaddour (2025, p. 750) demonstrated that neural network architectures outperformed logistic regression across every performance metric, achieving a classification accuracy rate of 93.8% for stock price trends. This superiority is largely attributed to the data-driven philosophy of ML, which partitions information into training and experimental sets—typically an 80/20 or 70/30 split—to validate the model's predictive reliability and reduce the risk of systemic bias (Medini, 2024, p. 586; Bekaddour, 2025, p. 785). Furthermore, while traditional models struggle with high-dimensional interactions, ML classifiers excel at capturing the nuanced interactions between disparate financial ratios such as asset turnover, debt-to-equity, and dividend yields (Bekaddour, 2025, p. 751;

Gbadebo, 2025, p. 7). This methodological shift is not merely a technical upgrade but a necessary evolution to address the "bounded rationality" of economic actors, providing them with objective, high-speed analytical tools that transcend human cognitive and temporal limitations (Kehri et al., 2024, p. 28; Nechachda, 2026, p. 43).

Ensemble methods, particularly Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), have emerged as the most consistently high-performing architectures for predicting market directions and macroeconomic growth. Random Forest operates on the principle of "bagging," constructing a multitude of independent decision trees and aggregating their outputs through a majority voting or averaging mechanism to bolster model robustness and mitigate overfitting (Messaoudi, 2024, p. 686; Gbadebo, 2025, p. 24). In the South African context, Gbadebo (2025, p. 9) found that ensemble classifiers like RF and gradient boosting consistently outperformed logistic regression by margins of 5–15% in predictive accuracy. Similarly, Messaoudi (2024, p. 724) highlighted the exceptional performance of RF in capturing intricate economic patterns, noting an R-squared value of 0.9444 in forecasting GDP growth, which far surpassed the capabilities of traditional Vector Autoregressive (VAR) models. On the other hand, XGBoost utilizes "boosting" logic, where each sequential tree aims to correct the residual errors of its predecessor, creating a highly efficient system that reduces model complexity through strict regularization parameters like Gamma (Rouaba, 2024, p. 1171, 1182). Rouaba (2024, p. 1185) utilized this algorithm to predict Apple Inc. stock closing prices, revealing a high predictive ability based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. These ensemble architectures are particularly valuable in volatile emerging markets because they can effectively process noisy predictors and handle missing values without extensive manual preprocessing (Messaoudi, 2024, p. 687; Gbadebo, 2025, p. 34).

Deep learning models and multi-layered neural networks represent the advanced frontier of financial prediction, particularly when dealing with unstructured data and "breakout" stock identification. These architectures, such as Deep Neural Networks (DNN) and Multilayer Perceptrons (MLP), are designed to simulate the human brain's hierarchical processing, utilizing hidden layers to learn complex patterns within financial statement indicators (Bekaddour, 2025, p. 786; Ansary, 2022, p. 366). Ansary (2022, p. 367) conducted a comparative analysis for identifying breakout stocks—those whose prices deviate significantly from support or resistance levels—and found that MLP and Support Vector Machines (SVM) were the leading models, while simple Decision Trees yielded the poorest outcomes. The effectiveness of DNNs is often optimized through the implementation of the Adam Solver and ReLU (Rectified Linear Unit) activation functions, which enhance the model's ability to interpret non-linear trends and classify stock directions with high precision (Bekaddour, 2025, p. 786). For instance, Bekaddour (2025, p. 794) reported that the DNN model achieved a sensitivity of 93.7% and a precision of 93.7% when analyzing 10-K filing reports, demonstrating that these models can effectively distinguish between "uptrend" and "downtrend" classifications. However, the "black-box" nature of these deep architectures remains a significant challenge for regulatory compliance, leading to a growing demand for "Explainable AI" (XAI) frameworks that clarify the specific financial ratios, such as Return on Capital Employed (ROCE) or dividend yields, that drive a model's final prediction (Gbadebo, 2025, p. 3; Bekaddour, 2025, p. 805).

The performance of Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) highlights the importance of kernel-based and neighborhood-based learning in financial anomaly detection and volatility analysis. SVM identifies an "optimal separating hyperplane" that maximizes the margin between different classes, making it highly effective for datasets with overlapping distributions or high-dimensional features (Djennad, 2025, p. 394; Gbadebo, 2025, p. 23). Gbadebo (2025, p. 31) noted that SVM performed optimally when utilizing a Radial Basis Function (RBF) kernel with a penalty parameter of $C=10$, suggesting that non-linear separability is crucial for accurate credit risk and stock classification. In contrast, KNN relies on similarity metrics to classify new data points based on their closeness to historical neighbors, often smoothing out local fluctuations through a higher number of neighbors, such as the nine-neighbor threshold identified as optimal for South African firm data (Gbadebo, 2025, p. 32). Comparative studies during extreme events, such as the COVID-19 pandemic, have shown that these models can adapt to external shocks when integrated with pandemic-specific indicators (Bouriche & Benbouziane, 2022, p. 647). Bouriche and Benbouziane (2022, p. 671) found that while RF generally outperformed others, SVM and KNN achieved respectable accuracy levels (69–71% F1 Score) in predicting e-commerce stock movements in the US and Chinese markets. These findings underscore that while SVM is robust for complex decision boundaries, KNN serves as an intuitive but sometimes less precise alternative in highly volatile environments (Gbadebo, 2025, p. 38; Bouriche & Benbouziane, 2022, p. 673).

Finally, the efficacy of all machine learning classifiers is significantly enhanced through the application of feature selection techniques and rigorous hyperparameter optimization. The predictive power of a model is not solely determined by its architecture but by the quality of the input features, such as technical indicators (SMA, RSI, MACD) and macroeconomic variables (Bouriche & Benbouziane, 2022, p. 658; Rouaba, 2024, p. 1177). Bouriche and Benbouziane (2022, p. 666) demonstrated that utilizing "Random Forest Feature Importance" to isolate the most contributing variables led to a notable improvement in accuracy and F1 Score metrics across RF, SVM, and KNN classifiers. Furthermore, the calibration of hyperparameters—such as learning rate, max tree depth, and number of estimators—is essential to prevent overfitting and ensure that the model generalizes well to unseen data (Zhang, 2022, as cited in Rouaba, 2024, p. 1182; Gbadebo, 2025, p. 30). For example, XGBoost performed best with a low learning rate of 0.01 and shallow trees to ensure gradual, stable learning (Gbadebo, 2025, p. 35). Interestingly, while advanced models often outperform traditional baselines, post-hoc statistical tests like Tukey HSD sometimes reveal that the performance gap between well-tuned models can be marginal, suggesting that the choice of an algorithm should be guided by contextual factors such as data availability, computational cost, and the need for regulatory transparency (Gbadebo, 2025, p. 41, 45). Ultimately, the synthesis of feature engineering and algorithmic rigor creates a "dynamic efficiency" that empowers financial institutions to manage risk and predict returns with unprecedented precision (Nechachda, 2026, p. 43; Saad Bekhouche et al., 2025, p. 1001).

Section III: The Synergy of Big Data Analytics and Algorithmic Trading Architectures

The global financial ecosystem is currently navigating a period of profound structural turbulence, characterized by the rapid acceleration of digital innovation and a fundamental transformation in risk dynamics (Saad Bekhouche et al., 2025, p. 85). At the center of this evolution is the integration of big data analytics, which refers to the sophisticated technical process of

examining massive, high-velocity datasets to uncover hidden patterns, market trends, and complex correlations that remain invisible to traditional analytical methods (Djennad, 2025, p. 404). Digital transformation within this context transcends mere technological adoption; it serves as a strategic enabler that redefines organizational agility, allowing financial institutions to respond dynamically to market volatility while aligning their long-term strategic objectives with real-time computational capabilities (Smaali & Boubaa, 2025, p. 181). Big data in finance is uniquely defined by five critical dimensions: Volume, representing the petabytes of tick-by-tick trading data and historical records; Velocity, the necessity for near-instantaneous processing where milliseconds dictate the boundary between profit and loss; Variety, encompassing numerical market data, textual news reports, and even multimedia content; Veracity, which focuses on the absolute reliability of the data to ensure informed decision-making; and Value, the ultimate extraction of actionable insights that shape institutional investment strategies (Djennad, 2025, p. 409). This data-driven paradigm addresses the inherent "bounded rationality" of human economic actors—those cognitive, temporal, and informational constraints that Herbert Simon argued prevent individuals from achieving perfect market optimization (Kehri et al., 2024, p. 28). By leveraging artificial intelligence to process millions of data points simultaneously, organizations can now mitigate the impact of information asymmetry, reducing price distortions and creating a more transparent environment where asset prices more accurately reflect their underlying intrinsic value (Nechachda, 2026, p. 44). Furthermore, the transition toward a data-centric architecture has fostered the rise of "Dynamic Efficiency," a theoretical framework suggesting that market effectiveness in the digital age depends not merely on the acquisition of information, but on the algorithmic capacity to process it with maximum precision and speed (Nechachda, 2026, p. 43).

Algorithmic trading, frequently referred to as automated or "black-box" trading, represents the practical implementation of these data-driven architectures, utilizing computer programs to execute trades based on predefined, mathematically rigorous instructions without the need for direct human intervention (Djennad, 2025, p. 412). The core architecture of an effective algorithmic trading system is composed of several interdependent components, beginning with the market data feed, which provides the foundation of real-time price and volume information (Djennad, 2025, p. 420). This data flows into the strategy logic, the system's analytical core where complex instructions—such as trend-following crossovers or mean-reversion signals—are processed to generate buy or sell decisions (Djennad, 2025, p. 420). Once a decision is reached, the execution system carries out the trade, minimizing slippage and optimizing transaction costs, while a dedicated risk management module ensures the system adheres to strict parameters, such as limiting the maximum exposure per trade (Djennad, 2025, p. 421). The integration of machine learning into these architectures has revolutionized traditional strategies; for instance, supervised learning algorithms are now used to predict the short-term price movements of financial instruments by learning from historical patterns, while reinforcement learning agents optimize profits through a trial-and-error process that adapts to shifting market states (Djennad, 2025, p. 432). This automation eliminates the emotional and psychological biases, such as overconfidence or loss aversion, that frequently impair the judgment of human traders during periods of high market stress (Djennad, 2025, p. 415; Kehri et al., 2024, p. 28). Consequently, algorithmic trading has become a dominant force in global markets, with high-frequency trading (HFT) alone accounting for approximately 60–

70% of the total trading volume on major exchanges like the New York Stock Exchange and NASDAQ (Nechachda, 2026, p. 46). By automating the entire trading lifecycle, from signal generation to order refinement, these architectures provide institutional investors with a "disciplined and consistent approach" that is capable of managing the immense complexity of modern, data-saturated financial environments (Djennad, 2025, p. 416).

High-frequency trading (HFT) stands as the most technologically advanced manifestation of algorithmic architectures, relying on deep learning and reinforcement learning algorithms to execute thousands of orders in fractions of a second (Nechachda, 2026, p. 46). The operational efficacy of HFT has significantly influenced market microstructure, contributing to a documented 22% increase in market liquidity and a notable 25–30% reduction in bid-ask spreads within advanced markets (Nechachda, 2026, p. 46, 48). These improvements in liquidity facilitate smoother price discovery mechanisms and reduce the overall cost of capital for market participants, thereby supporting the semi-strong form of the Efficient Market Hypothesis (EMH), which posits that prices should incorporate all publicly available information (Nechachda, 2026, p. 40, 42). In the U.S. experience, the expansion of AI-driven HFT has improved information incorporation speed by 66%, reducing the reaction time to unexpected market events from 30 minutes to under 5 minutes (Nechachda, 2026, p. 45, 48). Furthermore, advanced HFT systems utilize Multi-Layer Perceptrons (MLP) and Convolutional Neural Networks (CNN) to extract features from "alternative data," such as satellite imagery of retail parking lots or mobile money transaction records, to predict economic shifts before they are reflected in official reports (Gbadebo, 2025, p. 3; Djennad, 2025, p. 397). This synergy between computational speed and alternative data sources provides a unique "alpha" or competitive advantage, allowing firms to identify profitable anomalies in a crowded marketplace (Djennad, 2025, p. 427). However, the reliance on high-velocity automated systems also introduces systemic risks, such as the potential for "flash crashes" where collective algorithmic reactions to a single signal can trigger massive, unjustified price swings in seconds (Nechachda, 2026, p. 43). To manage these vulnerabilities, institutions are increasingly adopting "Smart Order Routing" (SOR) technologies, which use AI to compare prices and liquidity depth across multiple venues simultaneously, ensuring the best possible execution while minimizing market impact and enhancing informational fairness for all investors (Nechachda, 2026, p. 46).

The predictive landscape of algorithmic trading has been further enriched by the application of Natural Language Processing (NLP) and sentiment analysis to evaluate the vast streams of unstructured data generated by news feeds and social media (Nechachda, 2026, p. 45). By analyzing millions of online expressions within seconds, deep learning models like FinBERT can interpret investor sentiment and integrate these findings into price-prediction models, improving short-term forecasting accuracy by an average of 15% (Nechachda, 2026, p. 42). This capability is essential for identifying "breakout stocks"—those securities whose prices deviate significantly from established support or resistance levels accompanied by high volume—allowing traders to capitalize on high-margin opportunities that traditional fundamental analysis might miss (Ansary, 2022, p. 351). Empirical studies indicate that Multilayer Perceptron (MLP) and Support Vector Machines (SVM) are particularly effective in identifying these breakout points, providing a robust basis for investment during periods of extreme volatility (Ansary, 2022, p. 367). In emerging markets like South Africa and China, the integration of sentiment analysis and behavioral data has also been

used to enhance credit risk assessment, allowing platforms to assign credit scores to hundreds of millions of unbanked individuals based on their digital footprints rather than traditional collateral (Gbadebo, 2025, p. 3; Nechachda, 2026, p. 51). This "democratization of information" addresses the informational constraints identified in Simon's theory of bounded rationality, providing individual investors with analytical tools that were once the exclusive domain of large financial institutions (Kehri et al., 2024, p. 31; Nechachda, 2026, p. 44). However, as these models grow more complex, they often function as "black boxes," creating a demand for Explainable AI (XAI) frameworks to clarify the specific features—such as dividend yields or debt-to-equity ratios—that drive a particular trading signal (Gbadebo, 2025, p. 3; Bekaddour, 2025, p. 806). The move toward XAI ensures that while algorithms provide the speed and accuracy necessary for modern trading, human oversight remains a critical component for maintaining ethical standards and regulatory compliance (Kehri et al., 2024, p. 42; Ali, 2025, p. 492).

As we look toward the future of financial markets, the industry is witnessing a decisive shift from simple automation toward fully autonomous trading systems governed by deep reinforcement learning (Nechachda, 2026, p. 46). These autonomous systems achieve approximately 8–10% higher annual returns than conventional models by dynamically adjusting their strategies in response to evolving market conditions, while simultaneously maintaining volatility levels that are 30% lower (Zhao et al., 2024, as cited in Nechachda, 2026, p. 46). This transition signifies the emergence of a self-regulating, data-driven ecosystem where AI agents not only execute trades but also stabilize prices during periods of panic by rapidly correcting mispricing patterns (Nechachda, 2026, p. 46). To safeguard this new frontier, regulatory bodies are implementing Regulatory Technology (RegTech), utilizing machine learning to monitor market behavior in real time and detect illicit activities such as money laundering or insider trading with high precision (Nechachda, 2026, p. 49). In Europe, RegTech initiatives have already demonstrated a 37% reduction in compliance costs and a significant improvement in the accuracy of detecting market manipulation (ESMA, 2024, as cited in Nechachda, 2026, p. 50). This convergence of big data analytics, algorithmic trading, and RegTech confirms that AI is no longer a technical choice but a strategic necessity for institutional stability and market transparency (Nechachda, 2026, p. 53). Ultimately, the synergy between human judgment and algorithmic precision creates a cognitive ecosystem where "Cognitive Complementarity" allows human decision-makers to focus on strategic creativity while AI manages the immense volume of the digital financial landscape (Kehri et al., 2024, p. 34, 40). To sustain this progress, it is recommended that financial institutions invest in high-performance computing infrastructure and specialized training programs to bridge the skills gap between traditional financial expertise and advanced data science (Nechachda, 2026, p. 54; Saad Bekhouche et al., 2025, p. 1069).

Section IV: Explainable AI (XAI) and Forensic Auditing in the Mitigation of Creative Accounting

The pervasive challenge of creative accounting represents a sophisticated attempt to exploit ambiguities in financial standards to present a distorted, albeit technically legal, image of corporate health, which fundamentally erodes stakeholder trust and market stability (Ali, 2025, p. 885). These practices encompass a variety of manipulative mechanisms, including revenue manipulation—where the timing of recognition is adjusted to align with analyst forecasts—as well as expense

deferral and the use of special purpose entities (SPEs) for off-balance-sheet financing to conceal liabilities (Ali, 2025, p. 892, 899). Traditional forensic auditing techniques, which often rely on human judgment and static financial metrics like the Beneish M-score, have proven increasingly inadequate for identifying the high-dimensional, non-linear patterns characteristic of modern financial fraud (Ali, 2025, p. 907; Eghdouri & Bouras, 2025, p. 246). The integration of big data analytics into the accounting profession addresses this gap by enabling the processing of vast, heterogeneous datasets that can uncover hidden trends and suspicious activities signaling earnings manipulation (Eghdouri & Bouras, 2025, p. 247, 332). Within complex organizational structures like ACS Holding, which manages a diverse portfolio across multiple industries, the ability of data analytics to aggregate information from disparate sources provides a more comprehensive view of financial integrity, thereby reducing the risk that subtle discrepancies will be overlooked during an audit (Eghdouri & Bouras, 2025, p. 243, 271). Consequently, the adoption of these technologies marks the transition from a descriptive and historical auditing approach toward a proactive, data-driven framework capable of safeguarding transparency in volatile global financial systems (Eghdouri & Bouras, 2025, p. 250; Ali, 2025, p. 888).

The technological vanguard of anomaly detection in forensic auditing is currently dominated by unsupervised machine learning architectures such as Isolation Forest, One-Class Support Vector Machines (SVM), and Autoencoders, which collectively provide a robust defense against fraudulent activities (Ali, 2025, p. 886, 909). Isolation Forest operates on the unique principle that anomalous data points are "few and different," utilizing random decision-tree partitions to isolate these rare events in significantly fewer steps than normal points (Ali, 2025, p. 910, 912). This makes the algorithm exceptionally efficient for large-scale financial monitoring, where identifying irregular revenue patterns or unusual spikes in expense classifications is critical (Ali, 2025, p. 916). Similarly, One-Class SVM defines a decision boundary around normal data and flags anything falling outside this margin as a potential irregularity, while Autoencoders—a form of neural network—learn to compress and then reconstruct "normal" data patterns (Ali, 2025, p. 921, 930). When these models encounter manipulated data, they generate high reconstruction errors, effectively serving as an early-warning signal for auditors (Ali, 2025, p. 931, 935). In empirical applications, these models have demonstrated detection accuracies exceeding 90%, representing a significant improvement of 18–22% over traditional ratio-based forensic methods (Ali, 2025, p. 887; Eghdouri & Bouras, 2025, p. 297). This shift toward algorithmic oversight is essential for institutions like CPA Bank, where the combined dimensions of AI—specifically machine learning and intelligent agents—have been found to explain 68.6% of the variance in the effectiveness of financial risk management (Saad Bekhouche et al., 2025, p. 1000, 1061).

Beyond purely quantitative indicators, natural language processing (NLP) has introduced a linguistic dimension to forensic auditing by systematically evaluating management commentary in 10-K filing reports and annual disclosures (Ali, 2025, p. 887; Bekaddour, 2025, p. 750). Techniques such as sentiment analysis and topic modeling allow auditors to detect discrepancies between the narrative tone of a report and the actual financial performance of a company (Ali, 2025, p. 951, 961). For instance, deep learning models like FinBERT can identify "red flags" such as overly optimistic management language in the face of declining profitability or abrupt thematic shifts in risk disclosures that may indicate a deliberate effort to obscure financial vulnerabilities (Ali, 2025,

p. 887, 953). Furthermore, readability and complexity analysis evaluate whether management is using technical jargon or excessive verbosity to intentionally mislead stakeholders about the company's true financial condition (Ali, 2025, p. 958, 971). By integrating these textual insights with quantitative indicators such as asset turnover, debt ratios, and dividend yields, machine learning models can achieve a holistic evaluation of corporate integrity (Bekaddour, 2025, p. 751; Ali, 2025, p. 963). This multi-dimensional approach is crucial for modern financial analysts, as it empowers them to select companies with higher earnings per share while mitigating exposure to firms engaged in creative accounting practices (Bekaddour, 2025, p. 805, 809).

Despite the high accuracy of these advanced models, their "black-box" nature presents a significant barrier to regulatory adoption and institutional trust, necessitating the development of Explainable AI (XAI) frameworks (Ali, 2025, p. 888, 938). Auditors and regulators must be able to understand why a specific transaction or report was flagged as an anomaly to justify subsequent legal or remedial actions (Gbadebo, 2025, p. 3; Ali, 2025, p. 942). Frameworks such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are used to provide transparency by identifying which specific features—such as Return On Capital Employed (ROCE) or cash liquidity ratios—had the greatest influence on the model's prediction (Gbadebo, 2025, p. 13; Bekaddour, 2025, p. 751). This clarity is essential for reconciling the trade-off between the high predictive performance of deep architectures and the strict interpretability requirements of financial governance (Gbadebo, 2025, p. 14; Ali, 2025, p. 975). Without explainable models, decision-makers risk either "disempowerment" where they follow algorithmic recommendations blindly—or overconfidence, both of which weaken the ultimate rationality of the decision-making process (Kehri et al., 2024, p. 38; Ali, 2025, p. 888). Therefore, the modern forensic auditing toolkit must prioritize XAI solutions that provide a "right to explanation," ensuring that technological innovation remains ethically grounded and aligned with regulatory standards like Basel III (Nechachda, 2026, p. 41; Ali, 2025, p. 942).

As we look toward the future of financial oversight, the convergence of AI-driven auditing with Regulatory Technology (RegTech) is paving the way for real-time monitoring systems that can stabilize global markets (Nechachda, 2026, p. 49; Ali, 2025, p. 888). Advanced projects like ESMA's "RADAR" and "AMLbot AI" utilize machine learning to analyze market data from hundreds of institutions simultaneously, detecting abnormal trading patterns and hidden relationships between accounts with over 90% accuracy (Nechachda, 2026, p. 1130). These systems help mitigate systemic risks and price manipulation, thereby supporting the strong form of the Efficient Market Hypothesis by ensuring that prices rapidly reflect even semi-private information (Nechachda, 2026, p. 49, 1129). However, successful integration requires a holistic approach that includes high-performance computing infrastructure, robust data governance, and specialized training for risk managers (Nechachda, 2026, p. 54; Eghdouri & Bouras, 2025, p. 331). Organizations must also overcome barriers such as high costs and resistance to change by fostering an interdisciplinary culture where financial experts, data scientists, and IT specialists collaborate to refine predictive models (Eghdouri & Bouras, 2025, p. 245, 337; Saad Bekhouche et al., 2025, p. 1069). Ultimately, the synergy between computational rigor and domain expertise ensures that AI serves as a powerful instrument for enhancing transparency, reducing financial fraud, and fostering

inclusive and sustainable growth in the digital financial era (Ali, 2025, p. 969; Nechachda, 2026, p. 40, 53).

II. Conclusion

This study concludes that machine learning represents a transformative innovation in financial forecasting, consistently outperforming traditional econometric models. Ensemble methods like Random Forest and gradient-boosted architectures (XGBoost) provide high accuracy in predicting both stock directions and macroeconomic growth (Gbadebo, 2025, p. 3; Messaoudi, 2024, p. 735). The integration of AI tools addresses the "bounded rationality" of economic actors by processing vast, heterogeneous datasets to reduce information asymmetry (Kehri et al., 2024, p. 27; Nechachda, 2026, p. 43). However, the duality of AI implies that while efficiency is gained, risks of "algorithmic bias" and "flash crashes" must be managed through RegTech and XAI (Nechachda, 2026, p. 49; Ali, 2025, p. 488). This research provides a cross-disciplinary framework linking computational rigor with behavioral economics, validating the move toward "dynamic market efficiency" (Nechachda, 2026, p. 43).

Recommendations:

(1) Financial institutions should invest in high-performance computing infrastructure and cloud platforms to support real-time data analysis (Nechachda, 2026, p. 54).

(2) Policymakers should establish "regulatory sandboxes" to test AI tools in controlled environments (Nechachda, 2026, p. 54).

(3) Organizations must prioritize training programs to bridge the "skills gap" between financial experts and data scientists (SAAD BEKHOUCHE et al., 2025, p. 1069; Gbadebo, 2025, p. 12) .

Limitations:

The study relies heavily on historical data, which may not always account for unprecedented structural breaks (e.g., global pandemics), and many advanced models still lack full interpretability (Messaoudi, 2024, p. 729).

Future Research Directions:

Subsequent studies should explore hybrid models that combine neural networks with blockchain-enabled transparency and examine the long-term socioeconomic impacts of AI-driven automation on the financial workforce (Ali, 2025, p. 492; Eghdouri & Bouras, 2025, p. 341).

III. Reference List

1. Aissat, M. (2026). Artificial Intelligence: Concept, Origin, and Evolution. *Scientific Journal of Economic Dialogue*, 04(02), 43-53.
2. GBADEBO, A. (2025). Advancing Corporate Credit Risk Assessment in Emerging Markets: A Comparative Analysis of Machine Learning Classifiers in South Africa. *International journal of economic performance*, 08(02), 01-14.
3. MERMI, M., & KHACEF, D. (2025). Artificial Intelligence Applications in Finance: A bibliometric Analysis. *Economics Researcher's Journal* , 12(01), 17-27.
4. Smaali, D., & Boubaa, A. (2025). Digital Transformation and the Steering of Strategic Decisions: A Bibliometric Analysis. *Journal of Economics and Human Development*, 16(01), 171-200.
5. Ali, T. (2025). The detection of Creative Accounting in the financial reports Using Machine Learning Models. *Journal of Advanced Economic Research*, 10(01), 476–495.

6. Ansary, M. S. (2022). Breakout Stocks Identification using Machine Learning Approaches. Ahsanullah University of Science and Technology, 349–373.
7. Bekaddour, F. Z. (2025). Stock Price Classification Using Financial Indicators from 10-K Reports: A Comparative Study between Neural Networks (NN) and Logistic Regression (LR). Journal of Finance, Investment and Sustainable Development, 10(01), 224–243.
8. Djennad, M. (2025). Machine Learning and Big Data in the Financial Sector: Applications, Algorithmic Trading Integration, Challenges, and Emerging Opportunities. Setif 1 University, 374–463.
9. Eghdouri, C., & Bouras, F. (2025). Big Data and Data Analytics in Accounting: A New Era for the Detection and Prevention of Earnings Management -A case study ACS Holding-. University of Blida2, 242–348.
10. Kehri, S., Ouotmane, Ahmed., & Lehad, R. (2024). The artificial intelligence: a solution to the limited rationality of economic actors? Knowledge of Aggregates Magazine, 10(01), 26–44.
11. Messaoudi, M. (2024). Random Forest Modeling for Forecasting Economic Growth: A Machine Learning Approach in the Case of Algeria (1964-2022). Revue Abaad Iktissadia, 681–748.
12. Nechachda, C. (2026). The Role of Artificial Intelligence Tools in Enhancing Financial Market Efficiency: A Study of Global Models and Experiences. Creativity Review, 16(01), 40–55.
13. Nouria, B., & Mohamed, B. (2022). Predicting the Direction of E-Commerce Stock Prices during COVID 19 Using Machine Learning. Journal of Economic Integration, 10(01), 634–680.
14. Rouaba, M. (2024). Using the XGBoost algorithm to predict the closing prices of Apple Inc. stock. University of Tiaret, 1161–1191.
15. SAAD BEKHOUCHE, H., AISSAOUI, S., & ZEID, D. (2025). The Impact of Big Data Analytics Using Artificial Intelligence on Financial Risk Management: An Applied Study on the National Popular Credit Bank (CPA). Journal of Administrative and Financial Sciences, 09(02), 84–99.

How to cite this article by the APA method:

Atallah BENTIRECHE (2026), Machine Learning Applications in Predicting Stock Returns: A Scientific and Literary Approach to Modern Methods, Economic Development Review, Volume 11 (Number 01), Algeria: University of Eloued, pp.355-368.

