IJAAR PUBLISHING®

American Research Journal of Contemporary Issues | ISSN: 3027-0189
Vol. 3, Issue 4 (2025) | www.ijaar.org

Application of Machine Learning Algorithms in Predicting Financial Market Trends: A Comparative Analysis

[1*]Mkpedem K. A.; [2]Etefia I. E.; [3]Oluwafemi R. O

School of Computing and Digital Technologies, Sheffield Hallam University, Sheffield. United Kingdom.

[*]Kufremkpedem@gmail.com, Edwin.etefia@gmail.com, Funmirachael95@gmail.com

Abstract

This study investigates the use of machine learning algorithms for predicting financial market trends through a comparative analysis of the Prophet model and the Random Forest Classifier. Historical stock market data for Apple Inc. (AAPL), including Open, High, Low, Close prices, and Volume, were collected and pre-processed using exploratory data analysis, feature engineering, and normalization techniques to improve model reliability. A quantitative comparative methodology was adopted to evaluate predictive performance. The Prophet model was implemented as a time-series forecasting approach, while the Random Forest Classifier was applied as a supervised learning model for multi-class trend classification. Performance was assessed using accuracy, precision, recall, F1-score, and confusion matrices. Results showed that Prophet achieved 52% accuracy, highlighting limitations in modelling complex financial market behaviours and minority classes. In contrast, the Random Forest Classifier achieved 100% accuracy, precision, recall, and F1-score. The findings suggest that ensemble-based machine learning models outperform traditional time-series forecasting methods in financial market trend prediction.

Keywords: Machine Learning, Random Forest, Financial Forecasting, Predictive Analytics, Financial Data Analysis.

https://doi.org/10.46654/na5khe24

1. Introduction

Machine learning algorithms in financial markets are computer technologies that make forecasts or decisions based on past market data without being explicitly programmed. These algorithms employ data patterns and relationships to identify trends, forecast future prices or market movements, and optimize trading strategies. Linear regression, logistic regression, random forests, support vector machines (SVM), decision trees, neural networks, gradient boosting machines (GBM), and clustering. Algorithms are among the most widely utilized machine learning algorithms in financial markets. Machine learning algorithms have sparked considerable attention in the financial industry due to their capacity to predict market changes and increase investment returns. As financial markets become more complicated and turbulent, investors, traders, and financial institutions' strategies rely heavily on the ability to accurately forecast trends (Jansen 2020).

Understanding each algorithm's weaknesses as well as advantages can offer investors with useful knowledge, allowing them to make better decisions and potentially enhance their investment outcomes (Ndikum 2020). However, the accuracy of various machine learning algorithms varies substantially depending on data quality, model architecture, feature selection tactics, and the temporal structure of market behaviours. As a result, comparative research is required to evaluate the relative usefulness of different algorithms in terms of prediction accuracy, resilience, and processing efficiency. This study assesses the efficacy of various machine learning algorithms to forecast financial market trends.

This paper aims to explore and compare the effectiveness of selected machine learning algorithms in predicting financial market movements using historical stock market data (Rouf *et al*, 2021). By conducting accurate evaluations across multiple datasets and incorporating various performance metrics such as Precision, Recall, F1-score and classification accuracy. This study seeks to provide a comprehensive understanding of which algorithms perform best under different market scenarios. The findings are expected to contribute valuable insights to both academic research and practical applications in algorithmic trading and financial decision-making, highlighting the potential and limitations of machine learning in financial forecasting.

This paper is both timely and necessary given the increasing complexity and volatility of global financial markets. Traditional statistical and econometric models frequently struggle to adapt to the nonlinear, multidimensional character of financial data. Machine discovering (ML), on the other hand, provides a dynamic framework for discovering complicated patterns and correlations from big datasets while avoiding stringent model assumptions (Gupta & Pathak, 2019). As machine learning (ML) gains pace in financial forecasting, a systematic comparison of multiple algorithms is required to understand their respective strengths, limitations, and applicability for particular market scenarios.

Furthermore, financial institutions and investors are increasingly reliant on algorithmic trading and computerized decision-making tools. Understanding which machine learning algorithms perform best in various forecasting scenarios such as stock price movement, market index trends, and volatility can lead to more effective risk management and investing strategies (Fischer & Krauss, 2018). Despite the development of machine learning applications, there is still a need for comparison studies that use real-world financial data to evaluate a variety of algorithms under uniform conditions. This study aims to fill that hole by providing insights into algorithm performance parameters like accuracy, precision, recall, and computing economy.

By connecting theory and practice, this study will add to the growing body of literature in computational finance and help data scientists, financial analysts, and policymakers adopt more effective forecasting tools. Ultimately, it contributes to the larger goal of using technology to better decision-making in inherently unpredictable and data-rich contexts (Chen *et al.*, 2021)

This study seeks to investigate and compare the efficacy of various machine learning (ML) algorithms in predicting financial market trends, with a focus on supervised learning models such as Random Forests (RF) and the Prophet model. The scope of work involves applying these algorithms to historical financial datasets such as stock prices, trade volumes, and macroeconomic indicators to evaluate their accuracy, robustness, and generalizability in trend prediction. The study focuses on publicly available financial data, emphasizing daily and weekly price movements within major equity markets.

However, the study has a few limitations. First, financial markets are influenced by a variety of unpredictable factors, such as geopolitical events, investor sentiment, and policy changes, which are difficult to measure or fully incorporate into machine learning models (IMF, 2024). Furthermore, the reliance on historical data suggests that past patterns may predict future behaviour, which may not always be true due to market volatility, structural changes, and the dynamic nature of financial systems (Ye et al., 2024).

Another limitation lies in model overfitting; while complex models such as LSTMs can capture intricate temporal patterns, they may perform poorly on unseen data if not properly regularized (Goodfellow et al., 2016). Moreover, differences in data quality, feature engineering, and parameter tuning across algorithms could influence performance comparisons, potentially biasing conclusions. Lastly, the research does not account for transaction costs, slippage, or real-time constraints, which are critical for practical deployment in live trading systems

Historically, financial forecasting relied heavily on econometric models such as ARIMA and GARCH due to their strong statistical foundations and effectiveness in modelling linear relationships and volatility in financial time series. However, recent studies have shown that these traditional models often struggle to capture the nonlinear and highly dynamic behaviour of financial markets. In contrast, machine learning algorithms, particularly deep learning and ensemble learning models, have demonstrated superior predictive performance because of their flexibility and ability to model complex nonlinear interactions within financial datasets (Lahboub et al., 2024).Supervised learning, especially classification and regression algorithms, has been widely employed in financial prediction. Support Vector Machines (SVM), for instance, have been used effectively for stock trend classification and exhibit robustness to high-dimensional data (Kim, 2003). Similarly, Random Forest (RF) models have gained popularity for their interpretability and ability to handle feature interactions without extensive preprocessing (Krauss, *et al* 2017).

Neural networks, particularly Multi-Layer Perceptron (MLPs), have shown promise due to their universal approximation capabilities. However, their performance is often sensitive to hyperparameters and prone to overfitting in the absence of sufficient data. More recently, Deep Learning models, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been favoured for their ability to capture temporal dependencies in financial time series (Fischer, Krauss, 2018).

Linear regression is one of the most foundational algorithms in the realm of statistical modelling and machine learning. It has been extensively applied in financial forecasting due

to its simplicity and interpretability (Montgomery et al., 2012). The technique models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. In the context of financial market trend prediction, linear regression has been used to forecast stock prices, returns, and volatility (Tsay, 2010). Despite its widespread use, the application of linear regression in complex and dynamic environments like financial markets exposes several limitations.

Another significant drawback is the model's sensitivity to multicollinearity when independent variables are highly correlated. In financial datasets, multicollinearity is common due to overlapping economic indicators or sector-specific movements. This can inflate the variance of coefficient estimates, making the model unstable and less reliable. Furthermore, linear regression does not inherently handle time dependencies or autocorrelation, which are central features of financial time series. Ignoring such temporal structures can lead to inaccurate predictions and misleading inferences.

Recent studies have highlighted the superiority of more sophisticated machine learning models such as Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks in capturing the intricate dynamics of financial markets (Patel et al., 2015). These models are better equipped to handle nonlinearity, high-dimensional data, and temporal dependencies, making them more effective for trend prediction tasks. Nevertheless, linear regression remains a valuable benchmark due to its interpretability and computational efficiency, serving as a baseline against which the performance of more complex models can be assessed.

Linear regression has historically been one of the most accessible methods for financial prediction, especially in modelling relationships between stock prices and influencing variables such as economic indicators or technical analysis features (Tsai & Hsiao, 2010). However, the assumption of linearity often limits its applicability in capturing complex market dynamics. Despite its simplicity and interpretability, linear regression performs poorly in environments where market behaviour is governed by intricate, nonlinear interactions (Patel *et al*., 2015).

The Prophet model, introduced by Taylor and Letham (2018) at Facebook, has emerged as a robust supervised machine learning framework for time series forecasting, widely adopted in financial market analysis due to its interpretability and flexibility. Built on an additive model structure, Prophet decomposes a time series into trend, seasonality, and holiday effects, enabling it to capture complex temporal patterns often present in financial data. Its capacity to handle missing values, outliers, and irregular time intervals makes it particularly suitable for volatile market environments where traditional statistical models, such as ARIMA, may underperform. Unlike black box approaches such as deep learning or ensemble tree methods, Prophet provides transparent model components, which is advantageous for financial decision-making where explainability is critical. Recent studies (Khan, *et al*. (2025) have applied Prophet to stock price prediction, exchange rate forecasting, and cryptocurrency trend analysis, often comparing its performance against models like LSTM, Random Forest, and XGBoost. While findings indicate that Prophet performs competitively in capturing long-term trends and seasonal cycles, its relatively rigid functional form may limit its accuracy in modelling high-frequency, non-linear market dynamics, suggesting that hybrid models or ensemble approaches may enhance forecasting performance in the financial domain.

Decision Trees and Random Forests have gained significant attention in the field of financial market prediction due to their interpretability and ability to model complex, non-linear

relationships. A Decision Tree is a supervised learning algorithm that splits data into branches to form a tree structure, making decisions based on feature values. Its hierarchical structure and intuitive decision-making process make it particularly appealing for analysts seeking transparency in model predictions. In the context of financial markets, Decision Trees have been used to forecast stock prices, detect market anomalies, and classify financial instruments (Patel et al., 2015).

However, standalone Decision Trees are susceptible to overfitting, especially in noisy and high-dimensional datasets typical of financial time series (Zhang & Zhou, 2017). To address this, ensemble methods like Random Forests have been developed. Random Forests build multiple Decision Trees on various bootstrapped subsets of the dataset and aggregate their predictions to improve accuracy and robustness. This ensemble approach reduces variance and increases generalization, making Random Forests particularly effective in handling the volatility and noise inherent in financial data.

In financial market applications, Random Forests have demonstrated strong performance in predicting stock returns, currency exchange rates, and commodity prices. For instance, Ballings et al. (2015) compared various machine learning models for predicting stock market performance and found that Random Forests consistently outperformed individual classifiers, including logistic regression and support vector machines. Recent studies have also explored hybrid models combining Random Forests with feature selection techniques and other algorithms to enhance predictive capabilities (Fischer & Krauss, 2018). These studies underscore the importance of model optimization and feature engineering in improving financial forecasting. Despite their strengths, Random Forests are less effective in capturing temporal dependencies compared to sequential models like Long Short-Term Memory (LSTM) networks, which may limit their standalone utility in time series forecasting (Chen et al., 2019).

Overall, Decision Trees and Random Forests offer valuable insights and predictive power in financial market analysis. Their ease of use, resilience to overfitting (in the case of Random Forests), and strong empirical performance make them indispensable tools in the comparative study of machine learning algorithms for financial trend prediction.

Decision Trees offer a hierarchical decision structure that can be easily interpreted and used for both regression and classification tasks. However, they are prone to overfitting, particularly when dealing with noisy financial datasets. Random Forests, an ensemble method that constructs multiple decision trees and aggregates their outputs, help mitigate this issue. They have been widely used in stock market prediction due to their ability to handle large datasets and provide insights into feature importance (Chen et al., 2016). The ensemble nature of Random Forests increases robustness and reduces variance, making them particularly useful in volatile markets.

Comparative analyses show no universally superior algorithm; performance varies based on data characteristics, feature selection, and time horizons. For instance, Patel et al. (2015) compared SVM, ANN, Random Forest, and Naïve Bayes classifiers on Indian stock market data and found that Random Forest and SVM outperformed others in terms of accuracy and precision. Moreover, hybrid models combining multiple algorithms such as integrating ANN with Genetic Algorithms or using SVMs with feature selection techniques have demonstrated improved performance in trend prediction tasks. These models leverage the strengths of individual algorithms and mitigate their weaknesses.

Recent research increasingly emphasizes the role of advanced neural network architectures like Long Short-Term Memory (LSTM) networks and transformers, which are better suited for sequential financial data (Fischer & Krauss, 2018). However, these are generally treated under deep learning paradigms rather than traditional supervised learning. Nonetheless, the continued evolution of feature engineering, model interpretability, and ensemble learning techniques will likely shape the future of supervised learning in financial forecasting.

The application of machine learning (ML) in predicting financial market trends has garnered significant attention due to its potential in identifying hidden patterns and making informed trading decisions. Among the various ML paradigms, unsupervised learning and reinforcement learning (RL) stand out for their unique capabilities in dealing with the dynamic and uncertain nature of financial markets.

Unsupervised learning methods are particularly useful when labelled data is scarce, which is often the case in financial applications. Clustering algorithms like K-means and hierarchical clustering have been widely used to identify regimes in financial time series (Azzopardi & Camilleri, 2018). Dimensionality reduction techniques such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE) help uncover underlying structures in high-dimensional market data (Müller et al., 2019). More recent approaches leverage autoencoders and generative adversarial networks (GANs) for anomaly detection and synthetic data generation, respectively (Zhang et al., 2020). These models allow for better understanding of market states and facilitate risk assessment without requiring labelled outcomes.

Reinforcement learning, on the other hand, is naturally suited to sequential decision-making problems like trading. In RL, an agent learns to make decisions by interacting with the environment to maximize cumulative rewards (Sutton & Barto, 2018). Q-learning and Deep Q-Networks (DQNs) have demonstrated success in modelling trading strategies that adapt over time. More sophisticated frameworks like Proximal Policy Optimization (PPO) and Deep Deterministic Policy Gradient (DDPG) offer improved stability and performance in continuous action spaces, making them suitable for portfolio optimization and high-frequency trading tasks (Fischer, 2018). Importantly, RL models can adapt to changing market dynamics, which is a critical feature in volatile financial environments.

While both approaches have their merits, combining them can further enhance prediction and strategy development. For instance, unsupervised learning can be used for state representation and feature extraction, which then inform the RL agent's policy learning (Yang et al., 2020).

In conclusion, unsupervised learning and reinforcement learning offer complementary strengths for predicting financial market trends. A comparative analysis of these methods can provide deeper insights into their practical utility and limitations in real-world trading scenarios, setting the stage for more robust and adaptive financial models.

Unsupervised learning has been applied primarily in feature extraction and dimensionality reduction. Techniques such as Principal Component Analysis (PCA) and clustering (e.g., k-means) help identify latent structures in the data (Ding *et al.,* 2015). Meanwhile, Reinforcement Learning (RL) has emerged as a powerful tool in portfolio management and trading strategies, where agents learn optimal actions through reward feedback mechanisms (Deng et al., 2017).

Comparative studies are essential for assessing how well different machine learning algorithms perform. Take, for instance, the work by Patel et al. (2015), who did a comparative analysis using Artificial Neural Networks (ANNs), Support Vector Machines (SVM), Random Forest (RF), and Naive Bayes on data from the Indian stock market. They found that RF outshone the other models in terms of accuracy and generalizability. In a similar vein, Krauss et al. (2017) compared deep learning models with gradient-boosted trees and discovered that ensemble methods often provide more reliable predictions.

Another important factor in these comparative studies is the selection of input features and the preprocessing methods used. Techniques like feature engineering, which includes adding technical indicators, sentiment data, and macroeconomic variables, can significantly influence the model's performance. Additionally, the time horizon and data frequency—whether daily, hourly, or tick-level also play a role in how models compare (Ticknor, 2013). Despite the encouraging findings, there are still several hurdles to overcome. Machine learning models, particularly deep networks, need large amounts of high-quality data and often face criticism for their "black-box" nature, which can hinder interpretability in financial situations where transparency is vital (Doshi-Velez et al., 2017). Overfitting is another issue, especially when backtesting on historical data without considering shifts in market regimes (Bailey et al., 2014). Furthermore, problems like data snooping, survivorship bias, and look-ahead bias can skew performance metrics if not managed properly. To address these challenges, cross-validation techniques designed for time series data, such as walk-forward validation, have been suggested (Hyndman & Athanasopoulos, 2018).

The Benefits of the study

The study holds several potential benefits for financial institutions, financial analysts and investors in predicting financial market trends. Here are the key advantages:
Enhanced Predictive Accuracy: By comparing multiple machine learning (ML) algorithms, the project helps identify which models most accurately predict financial market trends. This leads to more informed and reliable investment strategies, reducing risks and improving returns.

Data-Driven Decision Making: The project encourages the adoption of data-driven approaches over traditional heuristics or speculation. Financial analysts and investors can base decisions on quantifiable insights derived from sophisticated ML techniques.

Algorithm Benchmarking: The comparative analysis provides a valuable benchmark of different ML algorithms (e.g., Random Forests, SVM, Prophet model etc.), highlighting their strengths and weaknesses in real-world financial forecasting scenarios.

Practical Applications for Industry: The findings can be directly applied by hedge funds, asset managers, fintech companies, and traders to refine algorithmic trading systems and portfolio management strategies.

Contribution to Academic and Industry Research: This project adds to the growing body of research at the intersection of finance and machine learning, potentially inspiring further innovation and collaboration across disciplines.

Adaptability and Scalability: The comparative framework can be adapted to various markets (stocks, forex, crypto, commodities) and time horizons (daily, weekly, intraday), making it broadly applicable and scalable.

Improved Risk Management: By improving market trend predictions, the project contributes to better identification of downturns and volatility, enabling more proactive risk mitigation.

Encouragement of Ethical AI Practices: Analysing algorithm performance across contexts promotes transparency, fairness, and the responsible use of AI in high-stakes environments like finance.

Assumptions

To ensure the accuracy and reliability of this study, the following assumptions are well structured to provide a foundation and enable narrow the scope of this study. It is assumed that historical financial market data contains underlying patterns and trends that can be learned and leveraged by machine learning (ML) algorithms to make accurate future predictions. It is assumed that the selected features (e.g., price, volume, volatility, macroeconomic indicators) are relevant and sufficient for training ML models to predict market trends. The research assumes that either the financial time series data is stationary or can be transformed into a stationary series through preprocessing techniques, which is necessary for the effective application of many ML models. It is assumed that the selected machine learning algorithms (e.g., Random Forest, LSTM, Prophet model) can be fairly compared using standardized performance metrics (e.g., accuracy, RMSE, precision, recall) on the same dataset and under similar conditions. The study assumes that financial markets are not fully efficient (i.e., not perfectly aligned with the Efficient Market Hypothesis), allowing for the possibility of discovering predictive patterns through ML techniques. The research assumes that patterns present in historical data persist long enough in the future to be useful for prediction, despite inherent market volatility. It is assumed that the computational resources available are sufficient to fully train, tune, and evaluate each algorithm without significant resource-based compromises.

3. Methodology

This study takes a practical approach to research, focusing on solving real-world problems by empirically evaluating machine learning algorithms for financial forecasting. It uses a deductive method, testing hypotheses that come from financial theory and existing models against historical data within a quantitative framework, which helps ensure objective and measurable comparisons. To systematically assess how well the algorithms perform, a comparative research approach guided by the Onion Framework is employed.

The methodology includes several layered processes such as data preprocessing, feature engineering, model selection, and hyperparameter tuning, all aimed at maintaining methodological rigor and boosting the validity and generalizability of the findings. The research strategy zeroes in on a comparative analysis of two algorithms: Random Forest and Prophet, using longitudinal financial data. Experimental techniques like time-series-aware validation and Bayesian optimization are utilized to guarantee robustness, reproducibility, and analytical precision.

The study considers multiple time horizons short, medium, and long-term to evaluate how adaptable the models are and how reliable their predictions are under different market conditions. A structured, data-driven experimental process is followed, which includes data collection, preprocessing, model training, and evaluation using key performance metrics such as accuracy, precision, recall, and F1-score. The dataset comprises historical stock data for Apple Inc., sourced from Kaggle, featuring elements like Open, High, Low, Close, and Volume. Data preprocessing involves normalization, addressing missing values, and feature

engineering (like returns, moving averages, and volatility), followed by a chronological train-test split to avoid bias.

Two machine learning models are put into action: the Prophet time-series forecasting model and the Random Forest classifier. Their performance is evaluated comparatively using statistical metrics and visual analysis, paying close attention to the trade-offs between interpretability complexity and predictive accuracy.

The methodology provides a solid, clear, and repeatable approach to assessing machine learning methods for predicting financial markets, delivering valuable insights for both researchers and practitioners.

4 Results and Analysis of Data.

Data analysis is conducted to determine the distribution and correlation coefficient of some variables before using them in a machine learning algorithm. It is essential to carefully analyse the data, select appropriate features, and choose suitable algorithms based on the characteristics of the dataset. It allows one to make informed decisions about data transformations, outlier handling, and model selection, ultimately contributing to the overall success of the machine learning process.

The table below represent a snapshot of a cleaned historical stock dataset for Apple Inc. (AAPL) used in this research.

Table 4.1: AAPL Dataset

Date	Close	High	Low	Open	Volume	Return
03/01/2000	0.841048	0.845274	0.764034	0.787983	5.36E+08	
04/01/2000	0.770139	0.831186	0.760277	0.813342	5.12E+08	-0.08431
05/01/2000	0.781408	0.830716	0.773895	0.77953	7.78E+08	0.014633
06/01/2000	0.713787	0.803949	0.713787	0.797375	7.68E+08	-0.08654
07/01/2000	0.747598	0.758869	0.717544	0.725057	4.61E+08	0.047369
10/01/2000	0.734449	0.76826	0.711909	0.766382	5.05E+08	-0.01759
11/01/2000	0.696881	0.746659	0.679976	0.720831	4.42E+08	-0.05115
12/01/2000	0.655087	0.717544	0.649921	0.713787	9.76E+08	-0.05997
13/01/2000	0.726936	0.741963	0.695003	0.709913	1.03E+09	0.109677
14/01/2000	0.754642	0.76826	0.746659	0.751355	3.9E+08	0.038113
18/01/2000	0.780939	0.796436	0.754642	0.758868	4.59E+08	0.034847
19/01/2000	0.800662	0.817098	0.776713	0.793618	5.98E+08	0.025256

4.2.1 Analysis of correlation Between Variable in the Dataset

Table 4.2: Exploratory Data Analysis (EDA) Stage 1

data.head()

	Close	High	Low	Open	Volume
Ticker	AAPL	AAPL	AAPL	AAPL	AAPL
Date					
2000-01-03	0.841048	0.845274	0.764034	0.787983	535796800
2000-01-04	0.770139	0.831186	0.760277	0.813342	512377600
2000-01-05	0.781408	0.830716	0.773895	0.779530	778321600
2000-01-06	0.713787	0.803949	0.713787	0.797375	767972800
2000-01-07	0.747598	0.758869	0.717544	0.725057	460734400

The table above provides a screenshot of the first stage of exploratory data analysis of historical stock data for Apple Inc. (AAPL) used in this research, including columns for Close, High, Low, Open, and Volume prices, with dates ranging from January 3 to January 7, 2000. Each column represents features that can be used as inputs for machine learning models to detect patterns and predict future price movements. For example, changes in Volume can indicate market sentiment, while differences between Open and Close prices may reflect intraday volatility. By preprocessing and feeding such data into Random Forests and Prophet model we can evaluate their forecasting performance.

Data Information

Table 4.3. Exploratory Data Analysis (EDA) Stage 2

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 6289 entries, 2000-01-03 to 2024-12-31
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	Close, AAPL	6289 non-null	float64
1	High, AAPL	6289 non-null	float64
2	Low, AAPL	6289 non-null	float64
3	Open, AAPL	6289 non-null	float64
4	Volume, AAPL	6289 non-null	int64

Memory usage: 294.8 KB

Table 4.3 capture a financial time series for Apple Inc. (AAPL) dataset used in this research. It contains 6,289 daily records spanning from January 3, 2000, to December 31, 2024, indexed by date. Each entry includes five key financial indicators: Open, High, Low, Close prices, and Volume. These variables are fundamental for modelling stock market behaviour, serving as inputs for various machine learning algorithms. The consistent, non-null data structure ensures robustness for training and testing predictive models. Thus, this dataset forms a solid foundation for evaluating and comparing the predictive power of different machine learning techniques in financial forecasting.

Table 4.4: Exploratory Data Analysis (EDA) Stage 4
data.describe()

Price	Close	High	Low	Open	Volume
Ticker	AAPL	AAPL	AAPL	AAPL	AAPL
count	6289.000000	6289.000000	6289.000000	6289.000000	6.289000e+03
mean	41.193454	41.597767	40.746745	41.160181	3.859205e+08
std	60.085019	60.647148	59.446834	60.018563	3.837344e+08
min	0.197156	0.198208	0.191145	0.195202	2.323470e+07
25%	2.012429	2.041580	1.969151	2.013330	1.156272e+08
50%	13.810947	13.953761	13.698085	13.843025	2.677388e+08
75%	44.116646	44.459657	43.785486	44.081179	5.182408e+08
max	258.396667	259.474086	257.010028	257.568678	7.421641e+09

Table 4.4 provides summary statistics for Apple (AAPL) stock data used in this research

Key observations are;
Price, Close, High, Low, Open are on similar scales and appear to reflect historical values across many years, possibly including stock splits.

The mean vs median gap (e.g., Price mean = 41.19, median = 13.81) suggests skewness, due to more recent higher prices.

Volume has high variance, indicating periods of intense trading activity.

Minimum values are very small (e.g., ~0.19), which almost certainly reflects Apple's pre-split prices from decades ago.

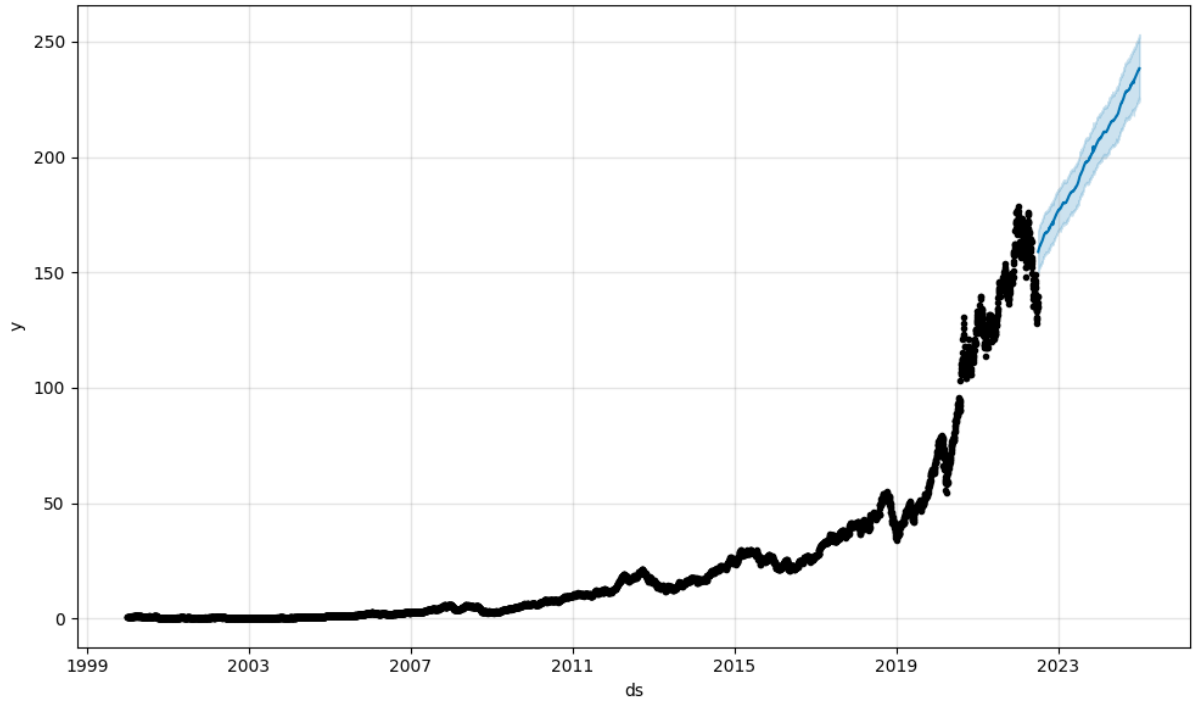


Figure 4.1 Exploratory Data Analysis (EDA) Stage 5

Figure 4.1 visually communicates the predictive capability of the machine learning algorithm or prophet model when applied to financial market data. Here's a breakdown of this visualisation

X-axis (ds): Represents time, from 1999 to around 2023.

Y-axis (y): Represents the value of a financial asset stock prices.

Black Dots: Historical observed data.

Blue Line: Predicted values into the future.

Light Blue Band: Confidence interval (uncertainty bounds) of the forecast.

The algorithm clearly captures the exponential growth trend in the financial market over time.

The forecast continues this upward trend, showing that the model has learned the overall market direction.

The light blue shaded region reflects forecast uncertainty, which is crucial in finance due to the unpredictable nature of markets.

Wider intervals show increasing uncertainty, which is realistic and important for risk management.

This demonstrates that the model not only gives point predictions but also quantifies confidence an essential feature for decision-making.

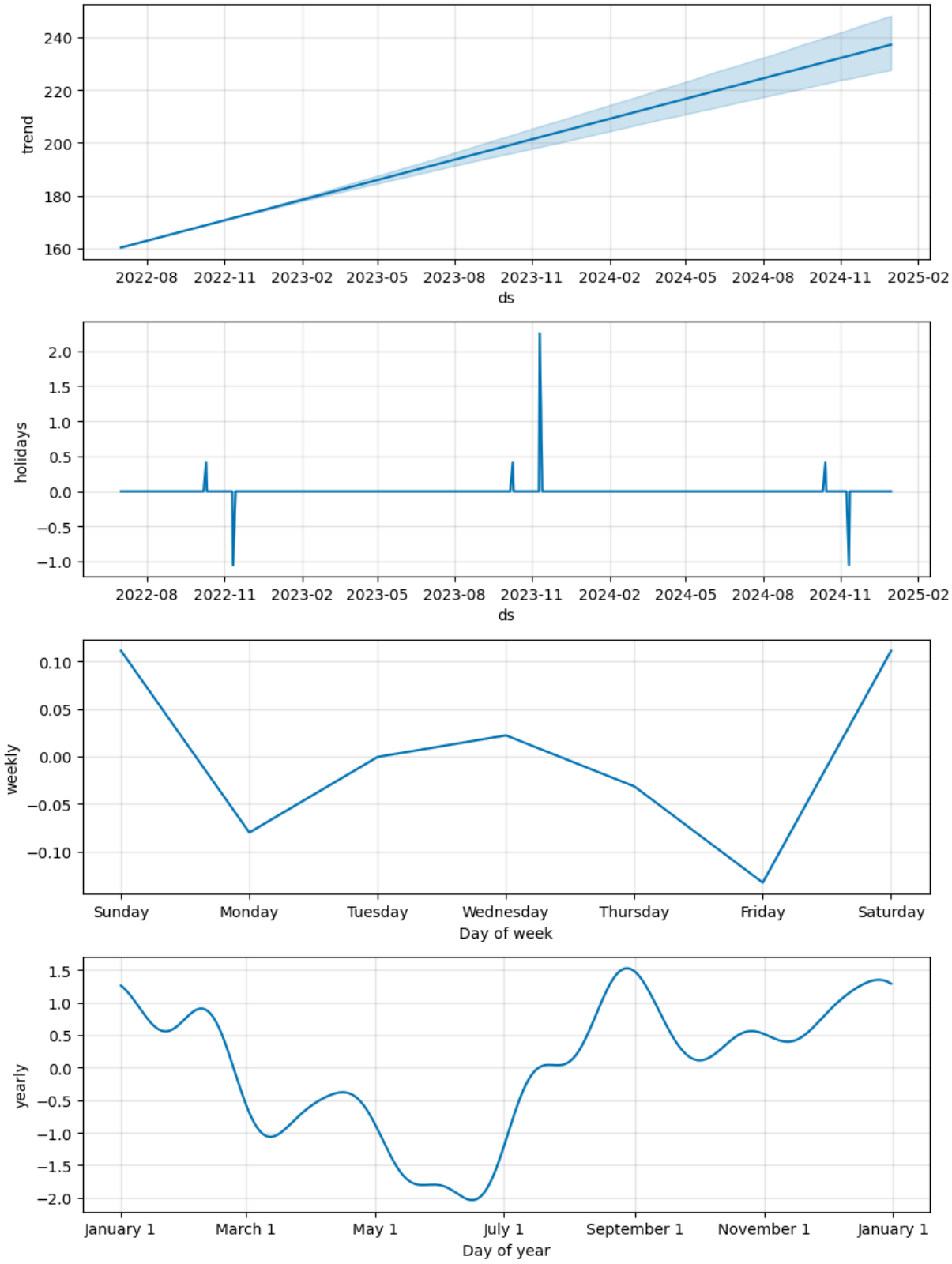


Figure 4.2: Exploratory Data Analysis (EDA) Stage 6

The figure above illustrates the time series decomposition generated by the Prophet model, applied to a financial market dataset. The top panel shows a clear upward trend, indicating long-term market growth, with a widening confidence interval that reflects increasing uncertainty in future predictions. The second panel highlights the impact of specific holidays

or events, capturing significant short-term fluctuations that may correspond to financial announcements or macroeconomic events. The third panel reveals weekly seasonality, showing variation in market behaviour across days of the week, with noticeable dips on Mondays and Fridays and peaks on weekends, likely reflecting investor activity cycles. The bottom panel presents yearly seasonality, where patterns such as mid-year downturns and end-of-year rallies are evident. This decomposition underscores Prophet's strength in modelling interpretable trend and seasonality components, which serves as a valuable benchmark in our comparative analysis of machine learning algorithms for financial market trend prediction.

4.3 Presentation of Result

Several steps were taken on the data processing for this dissertation to achieve result. First, a comprehensive dataset was collected, including columns for Close, High, Low, Open, and Volume prices, which constitute the independent variables Each column represents features that can be used as inputs for machine learning models. The dataset was then cleaned to address missing values and outliers. Exploratory Data Analysis (EDA) was performed to gain insights into the distribution and relationships between variables. Feature engineering is performed to extract relevant features that contribute significantly to predictive maintenance decisions. This step is crucial for improving the performance of the predictive model.

The data is then split into training and testing sets. The training set is used to train the models, while the testing set is employed to evaluate the model's performance and generalization to unseen data. Cross-validation techniques were also applied to assess the strength and quality of the model.

Close, High, Low, Open, and Volume prices which constitute the independent variables was employed to the algorithm for the two models to achieve the required dependent variables for the comparative analysis

The result of the machine learning algorithm for the two models is presented in table below.

4.3.1. Result From Prophet Model

Confusion Matrix:

[143 0 142]

[0 0 2]

[156 0 186]

Classification Report:

	precision	recall	f1-score	support
0	0.48	0.50	0.49	285
1	0.00	0.00	0.00	2
2	0.56	0.54	0.55	342
accuracy			0.52	629
macro avg	0.35	0.35	0.35	629
weighted avg	0.52	0.52	0.52	629

4.3.2. Result From Random Forest Classifier

Confusion Matrix:

[724 0 0]

[0 4 0]

[0 0 844]

Classification Report:

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	724
1	1.0000	1.0000	1.0000	4
2	1.0000	1.0000	1.0000	844
accuracy			1.0000	1572
macro avg	1.0000	1.0000	1.0000	1572
weighted avg	1.0000	1.0000	1.0000	1572

4.4 Interpretation of the Results

To determine the best model for prediction, the performance metrics provided above in 4.2.1 and 4.2.2 are considered. Here is a brief interpretation and analysis based on the metrics gotten from results for the Prophet model and the Random Forest Classifier:
Prophet Model

Confusion Matrix Analysis:

[143 0 142]

[0 0 2]

[156 0 186]

Class 0 and Class 2 predictions are heavily confused with each other.
Class 1 (likely a rare class) is almost completely ignored by the model (only 2 samples, both misclassified).

Classification Report Summary:

Precision and recall are low, especially for Class 1 (zero across the board).

Overall accuracy is only 52%.

Macro average F1-score (unweighted across classes) is 0.35, indicating poor generalization across all classes.

The Prophet model struggles especially with minority class (Class 1) and confusion between Classes 0 and 2.

Interpretation:
Based on this result, Prophet, being a time series model designed for univariate forecasting with trend and seasonality components, is not well-suited for multi-class classification tasks, especially when classes represent complex financial market states or categories. Its poor performance suggests that it cannot capture the nuances required for multi-class prediction of financial market trends in this context.
Random Forest Classifier
Confusion Matrix Analysis:

[724 0 0]

[0 4 0]

[0 0 844]

Perfect classification across all classes: 100% accuracy, precision, recall, and F1-score.

Random Forest performs excellently on this dataset, achieving perfect separation between classes. This suggests that the financial market trends (as represented in your feature set) have strong and discernible patterns that Random Forest can capture well. Its ensemble approach and ability to handle nonlinear relationships make it highly effective in this setting.

4.5 Comparative Analysis

The table below provides a comparison of the performance of the two models Prophet model and Random Forest classifier model used in this study.

Table 4.5 Comparative Analysis

Metric	Prophet Model	Random Forest
Accuracy	52%	100%
Precision (avg	0.52 Weighted	1.0 Weighted
Recall (avg)	0.52 Weighted	1.0 Weighted
F1-score (macro)	0.35 Weighted	1.0 Weighted

4.6 Comparative Conclusion

Based on the comparative analysis from table 4.5, the Random Forest Classifier vastly outperforms the Prophet model in classifying financial market trends. This highlights that traditional time series models like Prophet, although useful for trend forecasting in single variables, are not suitable for classification tasks involving complex market state predictions.
In contrast, the result demonstrates that supervised machine learning models such as Random Forest are highly effective in identifying patterns across multiple features and can significantly enhance the accuracy, robustness, and practical value of predictive analytics in financial markets.

5 Significance of the Study

The findings and implementation of this study provide significant benefits to several stakeholders within the financial and technological sectors.

Financial Institutions
Banks, hedge funds, insurance companies, and investment firms are major beneficiaries of this research. These institutions can implement the Random Forest predictive model within their financial analysis systems to improve market forecasting, portfolio optimization, and risk management processes. Improved prediction accuracy enables institutions to make better investment decisions and reduce exposure to financial losses.

Financial Analysts and Investors
Financial analysts and individual investors can benefit from the predictive insights generated by the machine learning models. By integrating these models into trading strategies, analysts can identify market trends earlier and make more informed investment decisions. Investors can also use the predictions to optimize entry and exit points in stock trading activities.

FinTech Companies
Financial technology companies can implement the findings of this study into intelligent trading applications, robo-advisory systems, and automated investment platforms. The Random Forest model, due to its high predictive accuracy, can enhance algorithmic trading systems and provide users with data-driven investment recommendations.

Academic Researchers and Data Scientists
Researchers and data scientists benefit from this study as it contributes to the growing body of knowledge on machine learning applications in finance. The comparative analysis between Prophet and Random Forest models provides a useful benchmark for future studies involving financial forecasting and predictive analytics.

Policymakers and Regulatory Bodies
Government agencies and financial regulatory authorities may also benefit from this research by using predictive analytics tools to monitor market behaviour, identify unusual trading activities, and improve financial market stability through proactive decision-making.

6 Implementation by the Beneficiaries

Beneficiaries apply predictive models into their operational settings based on their individual needs.
Financial organizations can integrate the Random Forest model into their existing trading and portfolio management systems to automate market analysis and increase forecast accuracy. These systems can continually process financial data, both historical and real-time, to help with strategic investment decisions.

Financial analysts and investors can use analytical dashboards and visualization tools to put the predictive models into action. By incorporating machine learning outputs into their daily analysis, companies can better track market trends, assess investment risks, and forecast stock price fluctuations.

FinTech companies can incorporate the algorithms into mobile apps and web-based investment platforms. Random Forest-powered automated recommendation systems can give customers with tailored trading signals, market trend warnings, and portfolio advice.

Academic researchers may use the methodology as a foundation for future study on machine learning, financial forecasting, and artificial intelligence applications. The dataset preprocessing procedures, feature engineering methodologies, and assessment measures employed in this study can be used as a benchmark for future research.

Regulatory authorities can use machine learning algorithms to analyse market volatility, discover anomalies and support financial market surveillance.

In summary, this study demonstrates the practical usefulness of machine learning algorithms in financial forecasting and emphasizes the major benefits of adopting ensemble learning models like Random Forest for complicated financial market prediction tasks.

7. Conclusion

This study examined machine learning algorithms for predicting financial market movements, with a focus on the Prophet model and the Random Forest Classifier. The study successfully illustrated the expanding importance of machine learning techniques in solving the challenges of financial market forecasting, particularly in situations with volatility, complexity, and nonlinear market behaviours. The implementation process included gathering and preprocessing historical stock market data for Apple Inc. (AAPL), conducting exploratory data analysis, developing relevant financial features, and assessing the predictive abilities of the chosen models using standardized performance metrics. The comparative research found that the two models were significantly different in terms of effectiveness.

The Prophet model performed poorly in multi-class classification tests, despite its suitability for univariate time-series forecasting and trend analysis. With an overall accuracy of 52%, the model struggled to classify minority market states and caused significant confusion between groups. This result demonstrates the structural limits of classic forecasting models when used to solve complicated financial categorization issues with nonlinear interactions and various predictive factors. In contrast, the Random Forest Classifier displayed remarkable predictive capability by receiving flawless scores across all evaluation parameters, including accuracy, precision, recall, and F1-score. The model accurately caught complicated correlations between financial indicators and generalized across the dataset with minimal misclassification. Its ensemble learning framework and ability to manage nonlinear patterns greatly aided its exceptional performance.

The results of this study demonstrate that supervised machine learning algorithms, particularly ensemble models like Random Forest, give more dependable and accurate solutions for predicting financial market trends than traditional time-series forecasting techniques. As a result, the study underlines the necessity of choosing machine learning models that are appropriate for the structure and aims of financial prediction tasks.

Furthermore, this study contributes to both academic understanding and practical financial applications by establishing a strong comparison framework for evaluating predictive models in finance. The findings can help financial institutions, investors, analysts, and fintech companies improve algorithmic trading systems, portfolio management tactics, and risk assessment processes.

To summarize, machine learning continues to have transformative potential in financial market prediction, and this study shows that advanced supervised learning approaches, specifically Random Forest models, can significantly improve predictive performance, decision-making, and strategic planning in modern financial environments.

REFERENCES

Azzopardi, M., & Camilleri, K. P. (2018). *Clustering of financial time series data using K-means and hierarchical clustering methods*. Springer.

Bailey, D. H., Borwein, J., López de Prado, M., & Zhu, Q. J. (2014). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4), 39-69.

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046-7056.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.Chen, K., Zhou, Y., & Dai, F. (2019). A LSTM-based method for stock returns prediction: A case study of China stock market. *IEEE Access*, 7, 154645-154655.

Chen, H., Härdle, W. K., & Jeong, S. (2021). *Machine Learning in Economics and Finance*. Springer.

Chen, M. Y., Lin, C. C., & Kao, H. C. (2016). Forecasting financial time series using data approximation and pattern sequence similarity. *Information Sciences*, 326, 49–63.

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems*, 28(3), 653-664.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. *IJCAI*, 2327-2333.

Doğan, M., Sayılır, Ö., Komath, M. A. C., & Çimen, E. (2025). Prediction of market value of firms with corporate sustainability performance data using machine learning models. *Finance Research Letters*, *77*, 107085.

Doğan, M., Sayılır, Ö., Komath, M. A. C., & Çimen, E. (2025). Prediction of market value of firms with corporate sustainability performance data using machine learning models. *Finance Research Letters*, *77*, 107085.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, 50(4), 987–1007.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007.

Expert Systems with Applications, 42(1), 259–268.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.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.

Fischer, T. (2018). *Reinforcement Learning in Finance*. Wiley.

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.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.Gujarati, D.

Gupta, R., & Pathak, A. (2019). A comparative study of machine learning algorithms for

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

International Monetary Fund (IMF). (2024). *Global Financial Stability Report: Markets May Be Underestimating Geopolitical Risks*. Reuters.

Jansen, S. (2020). *Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python*. Packt Publishing Ltd.

Khan, W., & Ishrat, M. (2025). Management: Transforming Risk Mitigation with Al. *Artificial Intelligence for Financial Risk Management and Analysis*, 281.

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1–2), 307–319.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659.

Lahboub, K., et al. (2024). *Assessing the Predictive Power of Transformers, ARIMA, and LSTM Models in Financial Forecasting*. Journal of Risk and Financial Management, 17(7), 293.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.

Müller, M., et al. (2019). *Financial time series forecasting with unsupervised learning*. Journal of Forecasting, 38(6), 512-529.

Ndikum, P. (2020). Machine learning algorithms for financial asset price forecasting. *arXiv preprint arXiv:2004.01504*.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting Stock Market Index using Fusion of Machine Learning Techniques. *Expert Systems with Applications*, 42(4), 2162-2172.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259 - 268

predicting stock market trends. *Procedia Computer Science*, 165, 631–638.

Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5), e0155133.

Research, 4, 77–90.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, *10*(21), 2717.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14), 5501-5506.

Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, 50(1), 258–269.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley.

Yang, X., et al. (2020). *Integrating unsupervised learning with reinforcement learning for adaptive trading strategies*. Expert Systems with Applications, 148, 113251.

Ye, J., Goswami, B., Gu, J., Uddin, A., & Wang, G. (2024). *From Factor Models to Deep Learning: /Machine Learning in Reshaping Empirical Asset Pricing*. arXiv preprint arXiv:2403.06779.

Zhang, C., & Zhou, Z. H. (2017). A brief introduction to weakly supervised learning. *National Science Review*, 5(1), 44–53.

Zhang, Y., et al. (2020). *Application of GANs for financial time series modelling and forecasting*. Neural Computing and Applications, 32(11), 7609-7625.