



Vaasan yliopisto
UNIVERSITY OF VAASA

Lauri Huutoniemi

**The Predictive Power of Machine Learning
Compared to Traditional Models in Stock Market
Forecasting**

School of Accounting and Finance
Bachelor's Thesis
Finance

Vaasa 2026

UNIVERSITY OF VAASA**School of Accounting and Finance****Author:** Lauri Huutoniemi**Title of the Thesis:** The Predictive Power of Machine Learning Compared to Traditional Models in Stock Market Forecasting**Degree:** Bachelor of Science in Economics and Business Administration**Major Subject:** Finance**Supervisor:** Veda Fatmy**Year of Graduation:** 2026 **Number of Pages:**31

ABSTRACT:

The purpose of this thesis is to investigate how machine learning can help stock market and how different models perform compared with traditional financial methods. The study also examines how data quality, feature selection, and market conditions influence forecasting accuracy. The motivation for this research arises from the limitations of traditional financial models. These models often assume stable relationships even though financial markets are complex and constantly changing.

The thesis reviews theoretical background related to market efficiency and behavioral finance, which help explain why predicting stock prices is difficult. Behavioral finance shows that investors do not always act rationally, and for example psychological factors can affect price movements. Because of this, traditional econometric models do not always fully describe real market behaviour.

Previous research shows that machine learning models can produce more accurate forecasts than traditional forecasting methods. For example, Random Forest, Support Vector Machines and deep learning models have shown promising results. However, result may vary depending on data quality and market situations. Also, practical forecasting use for the models have been discussed. With strengths and limitations, machine learning is showing to work better as a supportive tool rather than a complete replacement for traditional financial analysis.

Keywords: machine learning, stock prices, economic forecasts, artificial intelligence, investment activities, data acquisition

VAASAN YLIOPISTO**Laskentatoimen ja rahoituksen akateeminen yksikkö**

| | | |
|--------------------------|---|----------------------|
| Tekijä: | Lauri Huutoniemi | |
| Tutkielman nimi: | The Predictive Power of Machine Learning Compared to Traditional Models in Stock Market Forecasting | |
| Tutkinto: | Kauppätieteiden maisteri | |
| Oppiaine: | Rahoitus | |
| Työn ohjaaja: | Veda Fatmy | |
| Valmistumisvuosi: | 2026 | Sivumäärä: 31 |

TIIVISTELMÄ:

Tämän tutkielman tavoitteena on selvittää, miten koneoppimista voidaan hyödyntää osakemarkkinoiden ennustamisessa ja miten eri mallit toimivat verrattuna perinteisiin taloudellisiin ennustemenetelmiin. Lisäksi työssä tarkastellaan, miten datan laatu, muuttujien valinta sekä markkinaolosuhteet vaikuttavat ennusteiden tarkkuuteen. Tutkimuksen lähtökohtana ovat perinteisten rahoitusmallien rajoitteet, sillä ne olettavat usein markkinoiden käyttäytyvän vakaasti, vaikka todellisuudessa markkinat ovat monimutkaisia ja jatkuvasti muuttuvia.

Työssä käydään läpi markkinoiden tehokkuuteen ja käyttäytymistaloustieteeseen liittyvää teoriaa, jotka auttavat ymmärtämään, miksi osakekurssien ennustaminen on vaikeaa. Käyttäytymistaloustiede osoittaa, etteivät sijoittajat aina toimi täysin rationaalisesti, ja esimerkiksi psykologiset tekijät voivat vaikuttaa hintojen liikkeisiin. Tämän vuoksi perinteiset ekonometriset mallit eivät aina kuvaa markkinoiden todellista käyttäytymistä kovin hyvin.

Aiempi tutkimus osoittaa, että koneoppimismallit voivat monissa tapauksissa tuottaa tarkempia ennusteita kuin perinteiset menetelmät. Esimerkiksi Random Forest-, Support Vector Machine- ja syväoppimismallit ovat antaneet lupaavia tuloksia. Tulokset kuitenkin vaihtelevat riippuen käytetystä datasta ja markkinatilanteesta. Tutkimuksissa on myös korostettu mallien käytännön soveltamisen haasteita. Sekä vahvuuksista että rajoitteista huolimatta koneoppiminen näyttää toimivan paremmin ennustamisen tukena kuin täysin perinteisten analyysimenetelmien korvaajana.

AVAINSANAT: koneoppiminen, pörssikurssit, taloudelliset ennusteet, tekoäly, sijoitustoiminta, tiedonhankinta

Contents

| | | |
|-------|--|----|
| 1 | Introduction | 5 |
| 1.1 | Purpose of this thesis | 6 |
| 1.2 | Structure of the thesis | 7 |
| 2 | Theoretical Background | 8 |
| 2.1 | Efficient Market Hypothesis | 8 |
| 2.2 | Asset Pricing and Stock Market Predictability | 9 |
| 2.3 | Traditional Prediction Methods | 10 |
| 2.3.1 | Fundamental Analysis | 10 |
| 2.3.2 | Technical Analysis | 11 |
| 2.4 | The Role of Data in Stock Market Forecasting | 11 |
| 3 | Machine learning in stock market prediction | 11 |
| 3.1 | Introduction to Machine Learning in Finance | 13 |
| 3.2 | Key Machine Learning Techniques for Stock Prediction | 13 |
| 3.2.1 | Supervised Learning Models | 14 |
| 3.2.2 | Deep Learning Models | 15 |
| 3.2.3 | Reinforcement Learning in Trading Strategies | 15 |
| 3.3 | Strengths and Limitations of Machine Learning Models | 16 |
| 4 | Literature review | 18 |
| 4.1 | Supervised Learning Models | 18 |
| 4.2 | Random Forest and Ensemble Models in Stock Market Prediction | 20 |
| 4.3 | Deep Learning Models: LSTM and CNN | 22 |
| 4.4 | Synthesis of Findings | 23 |
| 5 | Conclusion | 25 |
| | References | 27 |

1 Introduction

Stock market forecasting has been with us ever since the first stock market opened. Everyday investors, analysts and researchers try to predict where stock prices are moving, but time has shown that predicting is not that easy. Financial markets bring together buyers and sellers to exchange assets, and prices are determined by supply and demand. Stock markets are also important for the economy because they enable capital allocation and wealth creation. Predicting stock market movements with the new technology should be easier, but perfect prediction method has not been found. If we consider traditional financial theories, for example the Efficient Market Hypothesis, predicting should be extremely difficult. All available information is assumed to be already in stock prices, and price movements follow a random walk.

However, it's clear that markets are not always perfectly balanced. Research on behavioral finance suggests that investor sentiment, psychological biases, and herd behavior may sometimes move prices away from their fundamental values (Shiller, 2003). This complicates forecasting. Traditional statistical models struggle in such situations because they assume rational behavior and stable linear relationships, assumptions that often fail when markets react emotionally or collectively. Market dynamics also change over time, which has made researchers search for forecasting methods that are more flexible.

This has given machine learning a chance. Recent advances in computing power and data availability have made machine learning a more attractive tool for analyzing financial data, especially when large volumes of information must be processed simultaneously. Machine learning refers to algorithms that learn patterns from historical data and use them to generate predictions (Jordan & Mitchell, 2015). The approach is increasingly data-driven. For example, deep learning models have been used in stock market forecasting, analyzing price movements, trading volumes, and other financial indicators to uncover patterns that simpler models may overlook.

In this thesis, I examine with existing literature, can machine learning solve previous challenges and can machine learning predict stock market movements better than traditional methods. I also go through literature that examines how data quality and market conditions affect prediction accuracy. Comparisons between models are made and also how they perform when they are used together.

1.1 Purpose of this thesis

The purpose of this thesis is to study the performance of machine learning in stock market forecasting and can it outperform traditional statistical methods. Traditional finance theory suggests that all available information is already in stock prices. This means that predicting future prices would be nearly impossible. Behavioral finance challenges this theory and shows that people's irrational behavior can push prices far from their fundamental values. If markets are not perfectly balanced, machine learning might find patterns and create predictions from those patterns.

Advances in computing power and the increasing availability of financial data have made it possible to apply machine learning models to large and complex datasets. Financial markets generate information from many different sources, and relationships between variables are often non-linear and change over time. Traditional statistical models may struggle in such environments because they typically assume more stable relationships. Machine learning models, particularly deep learning approaches, are designed to handle large datasets and to capture complex patterns that simpler models may miss. If such models are able to adjust when new information becomes available and when market conditions change, it is reasonable to expect that forecasting accuracy may improve compared with traditional financial models. On this basis, the first hypothesis of the thesis is formulated as follows:

H1: Machine learning models, particularly deep learning approaches, outperform traditional financial models in terms of predictive accuracy.

Second, even when machine learning models produce accurate forecasts, their performance is unlikely to remain stable under all market conditions. Financial markets evolve over time, and prediction accuracy therefore depends not only on data availability but also on how information is selected and processed. Forecasting performance typically declines when relevant information is missing or when data quality deteriorates, and models trained during relatively stable periods may struggle once market conditions become more volatile, as recent studies also suggest (Bao et al., 2025). Conditions matter. These challenges indicate that forecasting success is closely tied to the environment in which models are applied. For this reason, the thesis proposes a second hypothesis:

H2: The effectiveness of machine learning in stock market prediction is influenced by data quality, feature selection, and prevailing market conditions.

These hypotheses build on previous literature suggesting that traditional forecasting models face important limitations, while machine learning approaches may offer improvements in stock market prediction. This motivates the present study. The objective is therefore to examine whether machine learning models can help address long-standing challenges in financial forecasting.

1.2 Structure of the thesis

This thesis is divided into five chapters. The first chapter introduces the research hypotheses and outlines the purpose of the study. Chapter two presents the theoretical background, covering the Efficient Market Hypothesis, asset pricing theory, traditional forecasting methods, and the role of data in stock market prediction. Theory provides the foundation for later analysis. Chapter three focuses on machine learning models and explains how they are applied in financial forecasting. Chapter four reviews previous research comparing machine learning approaches with traditional forecasting methods. Finally, chapter five summarizes the main findings and concludes the thesis.

2 Theoretical Background

2.1 Efficient Market Hypothesis

Stock market prediction has long been considered challenging, and one explanation for this is provided by the Efficient Market Hypothesis. According to this view, stock prices already incorporate publicly available information because markets react quickly when new information emerges, making it difficult to consistently earn abnormal returns through forecasting or trading strategies (Durusu-Ciftci, Ispir & Kok, 2019). However, research has not reached a clear consensus on whether markets are always efficient. Stock price behaviour may change during economic crises, policy shifts, or other structural changes that influence financial markets.

Because new information constantly enters financial markets, price movements are often described as following a random walk, meaning that changes in prices are mainly driven by unexpected news rather than by patterns in past prices. Forecasting future prices therefore becomes highly challenging, since yesterday's movements rarely provide reliable guidance about tomorrow.

At the same time, actual market behavior does not always fully follow the Efficient Market Hypothesis. Investors sometimes allow emotions or behavioral biases to influence their decisions, which can temporarily push prices away from their fundamental values. In such situations, price movements may reflect collective investor behavior and changing market sentiment rather than new information alone (Shiller, 2003). A clear example occurred during the financial crisis of 2008, when panic and uncertainty strongly influenced trading decisions and markets overreacted, causing prices to deviate from fundamental values for extended periods (Inklaar, 2012). Episodes like this show that markets are not always perfectly efficient, and price movements may occasionally become predictable when certain behavioral patterns emerge.

2.2 Asset Pricing and Stock Market Predictability

Financial markets sometimes display patterns, suggesting that price movements are not entirely unpredictable. Well-known examples include momentum effects, mean reversion, and volatility clustering, which may allow short-term return predictability. These opportunities rarely last long. However, such patterns often weaken once they become widely used by investors, making their long-term usefulness uncertain. Lo's Adaptive Market Hypothesis (2004) offers one explanation for this by proposing that market efficiency changes as investors adapt and market conditions evolve. Similar challenges appear in modern forecasting models, as evidence from LSTM-based predictions shows that models may require time to adjust when markets experience structural changes (Villavicencio et al., 2026).

Traditional asset pricing theories look at predictability from another perspective. For example, Capital Asset Pricing Model (CAPM) link expected returns to the level of risk investors take. The model was first introduced in the 1960s and key contributors were Sharpe, Lintner and Mossin. In the simplest way, CAPM explains that investors expect higher returns only when they accept higher market risk. This return includes risk-free rate and an extra return that depends on how strongly the asset moves with the market:

$$E(R_i) = R_f + \beta_i [E(R_m) - R_f]$$

where $E(R_i)$ is the expected return of asset i , R_f is the risk-free rate, and $E(R_m)$ is the expected return of the market portfolio. β_i describes how sensitive the asset is to market movements, and if we take a closer look at the model, only this type of market risk is rewarded because company specific risk can be reduced with diversification. Still, in real markets prices react strongly to news and short-term price movements are very volatile, which makes short-term forecasting difficult if using traditional models. But behavioral finance could explain why prediction can be sometimes done. Like discussed earlier, investors are not always fully rational, and decisions can be influenced, for example by overconfidence. Prices can shift far away from their fundamental values for long periods

of time. These situations show that models assuming stable and fully rational behavior do not always capture real market dynamics. More flexible forecasting approaches may therefore perform better when market conditions change (Shiller, 2004).

Financial markets have also evolved alongside digitalization and growing data availability. Information now spreads faster, and investors have easier access to market data than before. Because of this, predictability can vary across time and environments. Forecasting success increasingly depends on how effectively large volumes of information can be processed and interpreted. At the same time, artificial intelligence is being applied not only to forecasting tasks but also to areas such as financial reporting, where it helps reduce information gaps between market participants (Bin-Nashwan et al., 2024).

2.3 Traditional Prediction Methods

Fundamental analysis and technical analysis are commonly seen as the two main traditional approaches used in stock market forecasting. While both aim to support investment decisions, they rely on very different types of information and methods when evaluating future price movements.

2.3.1 Fundamental Analysis

Investors often rely on company fundamentals when assessing whether a stock appears overvalued or undervalued. Rather than concentrating only on price movements, fundamental analysis focuses on financial statements, earnings performance, industry developments, and the wider economic environment surrounding a firm. Ratios such as Price-to-Earnings (P/E) and Price-to-Book (P/B) are frequently used to compare market prices with company earnings or book value, helping investors form expectations about potential investment value.

Company performance, however, does not depend solely on internal factors. Broader economic conditions also play an important role. Changes in inflation, interest rates, economic growth, or employment levels influence both business activity and investor sentiment, and these effects often appear in stock prices over time. For instance, rising inflation together with higher interest rates can slow economic activity and weaken stock market performance (Akram, 2009). Despite providing useful insights into a company's financial position, fundamental analysis has its limitations. It often struggles to explain short-term price fluctuations or episodes where investor behavior becomes irrational and prices temporarily deviate from fundamentals (Nti et al., 2020).

2.3.2 Technical Analysis

Many traders attempt to understand market movements by studying price charts rather than focusing on company fundamentals. This practice is commonly referred to as technical analysis. It relies on historical price data and trading volumes to identify patterns or trends that might reappear in future market movements. Indicators such as moving averages, Bollinger Bands, and the Relative Strength Index (RSI) are frequently used when making short-term trading decisions, especially when traders search for suitable entry and exit points. However, technical analysis also faces clear limitations. Because it mainly relies on historical price behavior, it may perform poorly during sudden market disruptions or unexpected economic events. In addition, broader economic conditions and company fundamentals are largely ignored, which reduces its usefulness for longer-term investment decisions (Akram, 2009).

2.4 The Role of Data in Stock Market Forecasting

Investors and researchers today have access to far more data than in the past when attempting to forecast stock market movements. Modern prediction models therefore rely heavily on large datasets, typically combining historical price information, trading volumes, and macroeconomic indicators. At the same time, new forms of data have become available. Sentiment indicators derived from news articles or social media

discussions, for instance, are increasingly used to approximate how investors perceive market conditions at a given time (Chong et al., 2017). Not all available information, however, can be used immediately. Large datasets often require cleaning, filtering, and careful integration before they become suitable for forecasting purposes, and differences in data preparation can significantly influence model performance (Kumbure et al., 2022). These challenges also explain why machine learning methods have attracted attention in financial forecasting. Such models are able to process large volumes of data and identify complex patterns, which may improve prediction accuracy. The following chapter introduces the main machine learning approaches together with their advantages and limitations.

3 Machine Learning in Stock Market Prediction

3.1 Introduction to Machine Learning in Finance

Recent developments in data processing and computational methods have increased interest in applying machine learning to stock market forecasting. Unlike many traditional statistical approaches, machine learning models can learn patterns from historical data and update their predictions as new information becomes available. This shift has been rapid. This is particularly relevant in financial markets, where price movements are influenced by many factors simultaneously, making prediction difficult for simpler linear models. For instance, approaches such as linear regression or standard fundamental models may fail to capture complex or non-linear relationships that appear in stock price dynamics. As a result, machine learning techniques have become increasingly common in trading strategies and financial forecasting studies (Hiransha et al., 2018).

Machine learning methods, however, also face important limitations. While some models can operate with modest computing resources, effective implementation often requires large datasets and considerable computational capacity. Prediction accuracy also depends strongly on data quality, and models may overfit or perform poorly when market conditions change abruptly. Consequently, even advanced machine learning systems cannot fully overcome the uncertainty inherent in financial markets, and forecasting remains a challenging task despite technological progress (Atsalakis & Valavanis, 2009).

3.2 Key Machine Learning Techniques for Stock Prediction

Machine learning includes a wide range of models, and their suitability for stock market prediction varies considerably. Some approaches perform better with certain types of data, while others adapt more effectively to changing market conditions or longer forecasting horizons. No single model works best in every situation. For this reason,

machine learning methods are commonly grouped into categories such as supervised learning, deep learning, and reinforcement learning, which are introduced in the following sections.

3.2.1 Supervised Learning Models

Supervised learning methods are among the most widely applied machine learning approaches in financial forecasting. These models rely on historical observations to identify patterns that may help predict future price movements. One commonly used example is the Support Vector Machine (SVM), which can be applied to both classification and regression tasks. SVM models are often used in stock market prediction because they can handle high-dimensional data and capture complex relationships between variables. In some cases, combining SVM with technical indicators has even outperformed traditional time-series models such as ARIMA (Patel et al., 2015).

Supervised learning in finance, however, is not limited to SVM models. Decision Tree methods, for instance, divide forecasting problems into smaller decision steps, which helps reveal which variables are most influential in predicting prices. A single tree may be unstable, but combining many trees leads to more reliable results. Random Forest models follow this idea by aggregating multiple trees to produce more accurate predictions while reducing overfitting, making them particularly useful when financial data is noisy or unstable (Khaidem et al., 2016). Another model that we slightly go through is Gradient Boosting Machines. This model is special and it's created to find mistakes these earlier models make and correct them. Popular gradient boosting machines are XGBoost and LightGBM. These are used in stock market prediction, and they have shown to have good accuracy, and they can handle large datasets efficiently (Chen & Guestrin, 2016).

3.2.2 Deep Learning Models

Financial data comes often is sequences where earlier events influence later price movements, which makes prediction more complicated than in many other fields. One of the reasons why deep learning models have become popular in stock market forecasting is that they are designed to handle large and complex sequential datasets. One of these models is Long Short-Term Memory (LSTM) network. It can retain information for longer periods of time and use this when later making predictions. This ability has made LSTMs have lower prediction errors and it can better capture time-series relationships, compared to more traditional models. (Fisher & Krauss, 2018).

Other deep learning models have also been adapted for financial forecasting, even though they were originally created for some other purpose. Convolutional Neural Networks (CNN) was originally created for image recognition, but now they are used to analyze stock price charts and trading patterns. The model tries to find long-term patterns in market data and used with technical indicators can achieve good prediction accuracy in stock market forecasting (Sezer and Ozbayoglu, 2018). Recently, deep learning approaches based on attention mechanisms have attracted growing interest in financial forecasting, as they allow models to capture long-term dependencies in data without relying on traditional recurrent network structures. Transformer-based models have been applied in tasks such as financial sentiment analysis, where news articles and company reports are used to support predictions of market movements. Combining technical market data with textual information through such models can improve forecasting performance, although results still depend strongly on data quality and prevailing market conditions (Nayak et al., 2024).

3.2.3 Reinforcement Learning in Trading Strategies

Reinforcement learning differs from many other machine learning approaches because models learn not only from historical data but also through interaction with their environment. Instead of simply identifying patterns, the system makes trading decisions

such as buying or selling and receives feedback in the form of gains or losses, which is then used to improve future actions. The objective is typically to develop strategies that maximize long-term returns rather than only predicting short-term price movements. A commonly used framework in this context is the Deep Q-Network (DQN), which evaluates possible actions and gradually refines decisions based on accumulated trading experience (Li, 2023).

Applying reinforcement learning in real financial markets remains challenging. Training often requires extensive simulated or historical data, and model performance may deteriorate when market conditions shift unexpectedly. For this reason, reinforcement learning is still more frequently used in research environments and experimental trading systems than in large-scale real-world trading applications.

3.3 Strengths and Limitations of Machine Learning Models

Traditional forecasting methods have been used in stock markets for decades, yet their limitations have encouraged the development of more flexible approaches such as machine learning. One important advantage of machine learning models is their ability to process large volumes of data originating from multiple sources. In addition to historical prices and trading volumes, models can incorporate newer inputs such as news sentiment or social media discussions. These approaches are also capable of capturing complex and non-linear relationships that simpler forecasting methods often miss. As a result, prediction performance may improve, and models can continuously update as new data becomes available (Fischer & Krauss, 2018).

Despite these strengths, machine learning models also face important limitations. Prediction results can become unreliable when data is incomplete or biased, and many models are vulnerable to overfitting, meaning that performance may deteriorate once market conditions change (Fischer & Krauss, 2018). Financial markets are also affected by unexpected events, which remain difficult to anticipate even for advanced forecasting systems because such shocks rarely follow predictable patterns.

Nevertheless, machine learning continues to evolve within financial research. Hybrid approaches and new techniques are being developed, which may further improve forecasting performance in the future. The next chapter reviews empirical studies that examine how these models perform compared with more traditional forecasting methods.

4 Literature review

In recent years, research on machine learning in stock market forecasting has expanded as more researchers try to understand whether modern prediction methods can outperform traditional approaches. Traditional forecasting models like linear regression or ARIMA assume that relationships between variables remain relatively stable over time. However, financial markets often behave differently. Volatility changes over time, structural breaks appear, and investor decisions are not always rational. This is why researchers have started to study if machine learning can make better predictions. However, reviewing this literature is not entirely straightforward, since the field is still relatively new, and only a limited number of studies have been made. This can be seen in the small number of core papers used in this review.

4.1 Supervised Learning Models

One commonly used machine learning approach in stock market forecasting is supervised learning. Akhtar et al. (2022) examined how different machine learning models performed when applied to stock market data, and several forecasting approaches were tested. Their study compared models including Support Vector Machines (SVM), with results showing promising outcomes as SVM achieved prediction accuracy close to 80%. Compared to traditional statistical models, machine learning approaches generally produced more reliable forecasting results. These findings support the first hypothesis while also highlighting the importance of data quality and preprocessing in determining model performance. This effect is worth noting. Differences between algorithms further suggest that forecasting accuracy depends strongly on how models and data are handled, which in turn provides support for the second hypothesis. In addition to SVM, the study also evaluated other machine learning approaches, including Random Forest models.

Similar findings can be found in a study made by Illa, Parvathala and Sharma (2021). They compared SVM with Random Forest and used historical data but also information derived from news sources. This methodology combined numerical market trends with investor sentiment. In this study SVM performed strongly, with accuracy close to 90%, and it demonstrated that transforming textual information into numerical features using techniques such as unigram and bigram representations contributes to prediction accuracy. Clear comparison to more traditional models wasn't made but results clearly showed that machine learning can predict with high accuracy. However, support to the second hypothesis can be found, with authors highlighting that model performance is sensitive to parameter selection and feature construction. (Illa, Parvathala and Sharma, 2021).

Support Vector Machines have been also compared to deep learning models. Study made by Chen et al. (2023) compared SVM to Long Short-Term Memory (LSTM) networks. Results gave advantage to LSTM if only predictive accuracy is considered. However, SVM performed more consistently when the amount of training data is limited, which further supports our second hypothesis. This is important in real investing situations, because large and clean datasets might not be available.

One common topic in recent literature is model interpretation. Celik, Ican and Bulut (2023) and Kumar et al. (2024) point out that SVM models remain difficult to interpret compared to simpler linear models. It's always better to understand where something is coming from rather than just seeing the results. The same goes with these models. It's important to understand how these models create their trading signals, not only whether it works statistically. This is relevant to institutional investors and regulators, who often require transparency in decision making. For this reason, explainable AI tools should be used with SVM to make forecasting more transparent and easier to use in real investment decisions.

Taken together, these studies indicate that supervised learning models often outperform traditional statistical approaches in stock market forecasting, which supports the first hypothesis. The picture is not entirely clear, however. Direct comparisons between machine learning and traditional models remain limited, as many studies instead compare different machine learning techniques with each other. SVM models, in particular, tend to perform well when markets exhibit non-linear or unstable behavior, although several studies also emphasize that prediction accuracy alone does not automatically translate into better investment performance. Successful trading decisions additionally require risk management and broader economic understanding.

4.2 Random Forest and Ensemble Models in Stock Market Prediction

Beyond single-model approaches, recent research has increasingly focused on ensemble methods that combine multiple learners. Ensemble methods aim to improve forecasting reliability by combining predictions from multiple decision trees rather than relying on a single estimated relationship. In Random Forest models, each tree captures slightly different aspects of the data, which reduces sensitivity to outliers, missing observations, and sudden disturbances. This improves robustness. Illa et al. (2021), for instance, show that Random Forest produces more stable forecasts than both SVM and traditional econometric models when datasets contain noise or incomplete information. These findings provide partial support for the first hypothesis, although direct comparisons with traditional forecasting methods remain limited. The study also reinforces earlier observations regarding the importance of data quality in model performance.

Kumar et al. (2014) compared Random Forest forecasting with more traditional regression-based approaches in stock price prediction. Instead of analyzing each financial indicator separately, the study examined how variables interact when forecasting price movements. Their results indicate that Random Forest performs particularly well when relationships between variables are complex or non-linear. Market conditions matter here. For instance, momentum signals may behave differently when volatility or trading volume changes, and Random Forest models appear

better able to capture these shifts than simpler models. The authors further note that ensemble approaches remain more stable when markets experience structural changes or when datasets contain noise, situations that frequently occur in real financial markets. This suggests that forecasting accuracy depends not only on model choice but also on how interactions within the data are captured. Similar conclusions are reported by Huang et al. (2026), who show that ensemble and tree-based methods maintain more stable performance when market conditions deteriorate or data quality weakens. Consequently, forecasting errors may be reduced in unstable environments when such models are applied, highlighting the importance of data quality and changing market conditions in financial forecasting.

As discussed earlier, interpretability remains important in financial applications. Celik et al. (2023) show that Random Forest models allow researchers to identify which variables influence predictions the most. This makes practical use easier. As a result, such models are more suitable in institutional environments where investment decisions often need to be clearly explained.

Overall, recent literature often positions Random Forest between traditional econometric approaches and more complex deep learning models. In many cases, it outperforms classical methods and remains more robust than models such as SVM when market data becomes noisy. Deep learning techniques may occasionally achieve higher prediction accuracy, yet they are typically more difficult to interpret and require greater computational resources. Practical usability therefore becomes important. For this reason, many investors prefer models that are easier to understand and implement, even if they do not always deliver the highest possible statistical performance. This observation connects to the first hypothesis by showing that, although prediction accuracy matters, usability and model stability often play an equally important role in real investment decisions.

4.3 Deep Learning Models: LSTM and CNN

Recent literature highlights deep learning as an increasingly important approach in modern stock market forecasting (Bao et al., 2025). Some studies report that deep learning models often achieve higher forecasting accuracy than traditional prediction methods, particularly when markets become volatile or when patterns evolve over time. A major reason for this performance is the use of Long Short-Term Memory (LSTM) networks, which are designed to process sequential data. Financial markets naturally follow such patterns, as trends and investor behavior often develop gradually and past events may continue to influence future price movements. Najem et al. (2024) compared several forecasting approaches and found that LSTM networks achieved the lowest prediction errors across multiple major stock indices. The study further showed that LSTM models can capture behavioral effects such as gradual market overreaction followed by later correction. Markets do not always behave efficiently. Nevertheless, forecasting results still depend strongly on data quality and changing market conditions, meaning that applying these models in real-time trading environments remains more complex than research results sometimes suggest.

Comparable findings have been reported in several recent studies. Hiransha et al. (2018) compared LSTM and CNN models with the traditional ARIMA approach using stock market data from India and the United States, finding that both deep learning models generally outperformed ARIMA in forecasting tasks. The advantage became more visible over longer forecasting horizons. The study also reported that forecasting errors decreased noticeably when neural networks replaced traditional statistical models. Similar evidence is provided by Bansal et al. (2022), who show that LSTM models performed well during periods of high market volatility, although forecasting performance remained sensitive to data preprocessing and parameter selection. Together, these results support the first hypothesis by suggesting that deep learning models can outperform traditional forecasting approaches, while still depending strongly on data quality and model configuration.

.

Deep learning in finance extends beyond time-series forecasting. Convolutional Neural Networks (CNNs), originally developed for image recognition, have increasingly been applied to financial prediction tasks where visual price patterns and chart structures can also be analysed. Sezer and Ozbayoglu (2018), for example, examined how such models can learn patterns directly from stock price charts. Their results indicate that visual structures commonly used in technical analysis may contain useful predictive information. Different modelling approaches can therefore complement each other.

These examples illustrate that deep learning models offer clear advantages in financial forecasting, but important limitations remain. Large datasets and substantial computing resources are often required, and model predictions may be difficult to interpret in practical decision-making. Performance alone is not sufficient. Institutional investors must therefore balance forecasting accuracy with operational risk and regulatory requirements. High accuracy observed in research settings does not automatically translate into reliable performance in live trading environments, where data is noisy and market conditions can change rapidly. Consequently, forecasting performance depends strongly on data quality, feature selection, and prevailing market conditions, which aligns with the second hypothesis. For this reason, deep learning models should be evaluated based not only on statistical accuracy but also on how reliably they perform in real market environments.

4.4 Synthesis of Findings

Taken together, the reviewed studies reveal several consistent patterns in stock market forecasting research. Machine learning models frequently outperform traditional statistical approaches, particularly in markets where price movements are non-linear and investor sentiment contributes to higher volatility. Classical asset pricing models, in contrast, usually assume more stable and mostly linear relationships between variables, which partly explains why they struggle to capture complex market dynamics. Market

behaviour is rarely stable. This also highlights a connection between machine learning results and behavioral finance, since asset prices are often shaped by psychological and institutional factors that traditional econometric models do not fully incorporate.

Despite these advances, it remains difficult to identify a single machine learning model that performs best in all situations. Model performance varies across forecasting environments, data availability, and market conditions. Support Vector Machines, for instance, often perform well in short-term forecasting tasks or when datasets are limited or noisy (Patel et al., 2015). Random Forest models, in turn, frequently offer a practical balance between prediction accuracy and interpretability, which is particularly important for institutional investors who must justify decisions based on model outputs (Illa et al., 2021). No model dominates in every case. Deep learning approaches such as LSTM and CNN models have often achieved the highest forecasting accuracy, yet their practical use may be constrained by large data requirements, computational costs, and limited interpretability (Hiransha et al., 2018).

For market participants, the reviewed evidence indicates that machine learning currently functions more as a supporting tool than as a complete replacement for traditional financial models. Full replacement remains unlikely. Model success depends strongly on data quality and prevailing market conditions, but also on how effectively practical challenges such as overfitting and operational risks are managed. In this sense, machine learning does not replace financial theory, rather, it relaxes some of its assumptions and provides more flexible tools that can adapt to changing market behaviour.

5 Conclusion

This thesis examined whether machine learning models can outperform traditional statistical methods in stock market prediction and whether combining different models can further improve forecasting results. Particular attention was also given to how data quality, feature selection, and changing market conditions influence forecasting performance. Predicting markets remains difficult. The theoretical discussion linked these challenges to the Efficient Market Hypothesis and behavioral finance, showing how traditional forecasting models often assume stable relationships even though financial markets evolve continuously over time.

The findings of the reviewed literature suggest that machine learning models are often more accurate at forecasting prices than traditional models. In particular, ensemble methods and deep learning models such as LSTM networks appear to frequently improve accuracy. Therefore, the first hypothesis can be confirmed, with the caveat that this is still a developing area of study, and forecasting may depend on several characteristics of the sample data, including markets and forecasting periods. Accuracy may also be affected by structural breaks, volatility changes, and other behavioral shifts in investor behavior. The practical usefulness of these models may be limited. Noisy data and overfitting are some of the problems that can reduce the interpretability and applicability of such models in more general use (Fischer & Krauss, 2018). Moreover, the reviewed studies suggests that no single machine learning model works equally well in all conditions. This confirms the second hypothesis of the thesis. Nevertheless, the reviewed findings taken together support the adaptive or behavioral market theories over those purporting efficient markets.

Future research could extend forecasting analysis across different asset classes and geographic markets, particularly under varying economic conditions. Greater attention is also needed on model interpretability and on how forecasting systems can be integrated into practical investment decision processes. Bridging research and practice

remains a challenge. Addressing these questions may help reduce the gap between academic forecasting results and real financial applications.

Overall, this thesis indicates that machine learning can improve stock market forecasting, although mainly under specific conditions. Model performance depends strongly on data quality, model selection, and the surrounding market environment. Machine learning therefore does not replace traditional forecasting methods; instead, combining these approaches may offer more reliable forecasting outcomes in the future. Financial markets will continue to change. As forecasting technologies evolve alongside them, further research and practical experimentation will remain essential.

References

- Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University – Science*, 34, 101940. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1018364722001215>
- Akram, Q. F. (2009). Commodity prices, interest rates, and the dollar. *Energy Economics*, 31(6), 838-851. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0140988309000917>
- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932-5941. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0957417408004417>
- Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock Market Prediction with High Accuracy using Machine Learning Techniques. *Procedia Computer Science*, 215, 247–265. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1877050922020993>
- Bao, W., Cao, Y., Yang, Y., Che, H., Huang, J., & Wen, S. (2025). *Data-driven stock forecasting models based on neural networks: A review*. *Information Fusion*, 113, 102616. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1566253524003944>
- Bin-Nashwan, S. A., Li, J. Z., Jiang, H., Bajary, A. R., & Ma'aji, M. M. (2024). Does AI adoption redefine financial reporting accuracy, auditing efficiency, and information asymmetry? An integrated model of TOE-TAM-RDT and big data

- governance. *Computers in Human Behavior Reports*, 100572. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S2451958824002057>
- Çelik, T. B., İcan, Ö., & Bulut, E. (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. *Applied Soft Computing*, 132, 109876. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1568494622009255>
- Chen, J., Wen, Y., Nanekaran, Y. A., Suzauddola, M. D., Chen, W., & Zhang, D. (2023). *Machine learning techniques for stock price prediction and graphic signal recognition*. *Engineering Applications of Artificial Intelligence*, 121, 106038. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0952197623002221>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://medial-earlysign.github.io/MR_Wiki/attachments/5537821/5537823.pdf
- Chong, E., Han, C. & Park, F.C., (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert systems with applications*, 83, 187-205. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0957417417302750>
- Durusu-Ciftci, D. Ispir M. & Kok, D. (2019). *Do stock markets follow a random walk? New evidence for an old question*. *International Review of Economics & Finance*, 64, 165–175. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1059056018310979>
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2),

654-669. https://libkey.io/libraries/3310/articles/165797794/full-text-file?utm_source=api_2884&allow_speedbump=true

Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science*, 132, 1351-1362. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1877050918307828>

Huang, X., Dang, Y., & Bhuiyan, M. Z. A. (2026). *Multi-objective portfolio optimization for stock return prediction using machine learning*. *Expert Systems with Applications*, 219, 119623. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0957417425032877>

Illa, P. K., Parvathala, B., & Sharma, A. K. (2022). Stock price prediction methodology using random forest algorithm and support vector machine. *Materials Today: Proceedings*, 56, 1776–1782. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S2214785321069674>

Inklaar, R., & Yang, J. (2012). The impact of financial crises and tolerance for uncertainty. *Journal of Development Economics*, 97(2), 466–480. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0304387811000563>

Jordan, M.I., Mitchell, T.M., (2015). Machine learning: Trends, perspectives, and prospects. *Science*. 349, 255-260. <https://www.science.org/doi/10.1126/science.aaa8415>

Khaidem, L., Saha, S., & Dey, S. (2016). Predicting the direction of stock market prices using random forest. <https://arxiv.org/pdf/1605.00003>

Kumar, P., Hota, L., Tikkiwal, V. A., & Kumar, A. (2024). Analysing Forecasting of Stock Prices: An Explainable AI approach. *Procedia Computer Science*, 235, 2009–

2016. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1877050924008664>

Kumbure, 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. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0957417422001452>

Li, J., Wang, X., Ahmad, S., Huang, X., & Khan, Y. A. (2023). Optimization of investment strategies through machine learning. *Heliyon*, 9(5), e16155. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S2405844023033625>

Najem, R., Bahnasse, A., & Talea, M. (2024). Toward an Enhanced Stock Market Forecasting with Machine Learning and Deep Learning Models. *Procedia Computer Science*, 241, 97–103. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1877050924017253>

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007-3057. <https://link-springer-com.proxy.uwasa.fi/article/10.1007/s10462-019-09754-z>

Nayak, G. H. H., Alam, M. W., Avinash, G., Ranjan Kumar, R. R., & Ray, M. (2024). Transformer-based deep learning architecture for time series forecasting. *Software Impacts*, 22, 100716. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S2665963824001040>

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. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S0957417414006551>

Sezer, O.B., Ozbayoglu, A.M., (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach.

Applied Soft Computing, 70, 525-538. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S1568494618302151>

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, 17(1), 83-104. https://www.aeaweb.org/articles?id=10.1257%2F089533003321164967&source=post_page-----

Villavicencio, J., Miranda, O., & Villanueva, E. (2026). The adaptive markets hypothesis through the lens of machine learning. *Heliyon*, 12(1), e34567. <https://www-sciencedirect-com.proxy.uwasa.fi/science/article/pii/S2405844025026842>