University for the Creative Arts

BERLIN SCHOOL OF BUSINESS & INNOVATION

Dissertation Title: **Financial Market Forecasting Using Machine Learning: A Case Study on Stock Prices and Cryptocurrency Trends**

Master title: **Advanced Machine Learning Models for Stock and Cryptocurrency Market Prediction**

Name:**Mandeep**

Year: **2024-25**

Statement of compliance with academic ethics and the avoidance of plagiarism

I honestly declare that this dissertation is entirely my own work and none of its part has been copied from printed or electronic sources, translated from foreign sources and reproduced from essays of other researchers or students. Wherever I have been based on ideas or other people texts I clearly declare it through the good use of references following academic ethics.

(In the case that is proved that part of the essay does not constitute an original work, but a copy of an already published essay or from another source, the student will be expelled permanently from the postgraduate program).

Name and Surname (Capital letters):

MANDEEP..

Date: 01/05/2025

CHAPTER 1 INTRODUCTION

Background of the Study

Financial markets also follow suit because they have always been dynamic, unpredictable and redefined by numerous factors. Forecasting stock and cryptocurrency prices remains a challenging yet essential task for investors, analysts, and policymakers. In recent years, Machine Learning (ML), a subset of artificial predictive analytics has become an innovative technology that can understand and learn, perform data analysis and issue more sophisticated recommendations than simple statistical models (Gogas and Papadimitriou, 2021).

In recent years, cognate classification has experienced an immense elevation when it comes to applying ML (machine learning) in financial forecasting. A report from PwC in 2023 shows that as of today over 52 % of the Asset management firms in North America are implementing machine learning in the investment management process and an estimation of AI related financial application in the next 5 years is expected to grow by 23% annually (Abbas, 2024). It is not limited only to equities; cryptocurrencies are by their nature decentralized and highly fluctuating which make them asset class both to embrace as well as a challenge for the ML-based forecasting. Bitcoin (BTC) and Ethereum (ETH) are averaging annualized volatilities of 60% and 80% respectively that are furious as per the Coin Market Cap, 2023 and hence, there is a need of developing more efficient and faster TCMs, which can utilize these sudden fluctuations (Nadeem, 2024).

Issues like non-stationary behaviour of the volume and prices distributions, non-linear trend of the volume and price changes, as well as high-frequency fluctuations of the volatility are beyond the capabilities of the traditional forecasting models, including ARIMA, GARCH, or linear regression. Algorithms like LSTM networks, Random Forests or SVM can have a better ability in analysing complex structure data (Wasserbacher and Spindler, 2021).

Research Aim

The purpose of this work is to compare how well the various machine learning techniques performs in terms of its ability to predict stock and cryptocurrency prices in comparison to other types of forecasts. The work will compare the accuracy of the model's predictions and the problems likely to be encountered when applying ML in the volatile financial environment (Gupta and Chaudhary, 2022).

Research Objectives

To identify the effectiveness of using machine learning models in stock and cryptocurrency price prediction.

To compare the performance of different ML algorithms (e.g., LSTM, SVM, Random Forest) in financial market forecasting.

To identify the difficulties and limitations encountered while applying ML techniques to stock and cryptocurrency forecasting.

Research Questions

To what extent do machine learning framework perform in forecasting stock or cryptocurrency prices?

Which machine learning algorithms gives the most accurate or reliable predictions in financial markets?

What are the common issues encountered in employing machine learning models for financial market forecasting?

Rationale for the Study

There are a number of group implications for such financial forecasting The forecasting is important merely for it to be useful; only when it proves reliable can it have practical application. Managers aim at earning good returns subject to the risk of undertaking their investments, and thus, effective prediction models play a crucial role in this endeavour. In addition, the main sector to be affected is the financial industry due to the increased technological advancement. It was evident that 85% of the financial institutions are adopting the AI and ML solutions as highlighted by the World Economic Forum (Yadav et al., 2023). This is especially evidenced by the influx of retail investors and the increased availability of the garment through such websites and applications like Robinhood, Binance, and Yahoo Finance.

Significance of the Study

This research will benefit many clients in the following ways:

For investors and traders, it will provide a relative assessment on which of the models has the accuracy of forecasting the market direction.
For academicians, it adds to the body of knowledge on the application of ML in the financial segment.
The insights generated from the research could help target, for financial

useful insights. Cryptocurrencies have amassed a market capitalization in excess of $2.4 Trillion in 2023 and there is high potential for competitive strategies based off of exact volume prediction (Arthur, Matjaž Perc and Ribeiro, 2023). institutions and fintech developers, the specific areas to invest in when it comes to developing predictive technologies.
Also, by embracing suggestions related to cryptocurrencies—an understudied area that is—this paper contributes

Methodology Overview

The type of research to be used in the study is quantitative, which shall be analysed using primary data. Some of the stocks and cryptocurrencies to be analysed include Apple Inc., Tesla, Inc., Bitcoin and Ethereum market indexes of good and service of historical market data will be retrieved. The selected ML algorithms, LSTM, SVM, and Random Forest will be created with the help of Python's Scikit-learn library as well as TensorFlow. Hypothesis testing tools that will be used include Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the R^2 score of predictive accuracy.

Structure of the Dissertation

Chapter 1: Introduction – Provides an overview, research rationale, and objectives.

Chapter 2: Review of related literature – Provides an account of the literature that has been done on ML forecasting in terms of studies, theories, and technological solutions.

Chapter 3: Methodology – Outlines the research design, data sources, and analytical framework.

Chapter 4: Data Analysis and Results – Presents the results of model implementation and comparative evaluation.

Chapter 5: Discussion – Explains research findings in the light of theories and contemporary marketing environment.

Chapter 6: Conclusions and Recommendations – Summarizes key findings and suggests directions for future research.

Limitations

However, there are some limitations to the study which is worth pointing out. First, financial markets, as it has been seen, depend on the events such as geopolitical shocks and pandemics that cannot be predicted by such models. Second, it deals with a sample of the stocks and does not involve the whole market, as well as with a specified list of cryptocurrencies. Third, there is higher non-interpretable of ML models, especially the deep learning like LSTM, which may pose as a threat to transparency.

Summary

Specifically, this paper examines whether artificial intelligence techniques, such as machine learning, are present in the forecast of both traditional and digital assets, including equities and cryptocurrencies. For this purpose, with the application of empirical analysis and comparison of models it aims to find out the hypothesis if ML is statistically and practically better than the traditional methods. It is anticipated to help in enriching the existing research, assist financial practitioners and advance the generation of AI-based investment solutions.

CHAPTER2– LITERATURE REVIEW 1 (Machine Learning in Financial Market Forecasting)

Date	Open	High	Low	Close ⓘ	Adj Close ⓘ	Volume
May 5, 2025	5,655.32	5,683.38	5,634.48	5,650.38	5,650.38	4,358,260,000
May 2, 2025	5,645.88	5,700.70	5,642.28	5,686.67	5,686.67	4,854,380,000
May 1, 2025	5,625.14	5,658.91	5,597.35	5,604.14	5,604.14	4,935,270,000
Apr 30, 2025	5,499.44	5,581.84	5,433.24	5,569.06	5,569.06	5,449,490,000
Apr 29, 2025	5,508.87	5,571.95	5,505.70	5,560.83	5,560.83	4,747,150,000
Apr 28, 2025	5,529.22	5,553.66	5,468.64	5,528.75	5,528.75	4,257,880,000
Apr 25, 2025	5,489.73	5,528.11	5,455.86	5,525.21	5,525.21	4,236,580,000
Apr 24, 2025	5,381.38	5,489.40	5,371.96	5,484.77	5,484.77	4,697,710,000
Apr 23, 2025	5,395.92	5,469.69	5,356.17	5,375.86	5,375.86	5,371,390,000
Apr 22, 2025	5,207.67	5,309.61	5,207.67	5,287.76	5,287.76	4,666,950,000

Figure 1: Historical Data of S&P 500

Source: Finance.Yahoo (2024)

Kumbure et al (2022) demonstrates that Stock and financial markets cannot be easily detained since they are shaped by various forces including economic, political and psychological forces which make the fluctuations of the markets irregular and challenging to decode. Essential conventional statistical forecasts that were used in the market include Linear regression, Autoregressive Integrated Moving Average (ARIMA), or any other forms of statistical forecast models. However, because of their linear structure and weakness in dealing with non-stationary and non-linear characteristics of the financial time series data, they are not very efficient. As financial info is becoming more voluminous, diverse and rapidly changing, Machine Learning (ML) has become an effective solution for identifying obscure trends and making more accurate predictions.

Sabry et al (2020) highlights that Artificial intelligence is a broader concept where machine learning is a part of it, and it uses programs that are trained on previous data, thus improving with time. These algorithms have now extended their applications in different areas of finance like, trading, asset pricing, credit scoring and identification of frauds. The application of ML is in its capacity to process vast amounts of data and to find peculiar patterns, especially valuable in the case of speculating fluctuations of stocks and cryptocurrencies as these markets are quite diverse and dynamic.

Dimingo et al (2021) said that The Role of Machine Learning in financial forecasting: Machine Learning has established greater prominence in the last decade leading to the development of the following. While, in the past, few financial institutions used this kind of methodologies mainly for academic research and some experimental platforms, more and more hedge funds, investment banks, and fintech companies apply it in practice nowadays. DT, SVM and k-NN implemented most of the classical Information Science models among the first to be considered for many applications. These models showed some advantages over statistical models some of the time especially when it came to the classification type problem such as prediction of the market trends (bull or bear trends).

Olcay et al (2024) analyses Random Forests (RF) and Gradient Boosting Machines (GBM) which extended and enhanced the predictive models by using the result from a pool of learning algorithms. They also have the benefits of bringing reduction of variance and bias which leads to making forecasts more stable and accurate. In the more near past, deep learning has emerged mainly represented by Recurrent Neural Networks (RNNs) and their variant known as Long Short-Term Memory (LSTM) networks that has brought about the biggest turnover in time series forecasting in the field of finance. As is well known, LSTM models that contain the so-called gates for the memory cells are suitable for working with sequential data, for instance, historical price data for equities or cryptocurrencies.

For instance, compared the performance of ARIMA, SVM, and LSTM in the S&P 500 index forecasting of the closing prices. The study brought out the fact that LSTM performed better than both ARIMA and SVM by providing an improvement in RMSE and MAPE values thus shifting the focus to deep learning models in the field of financial forecasting.

Applications in practice: The condition of today's global equity markets and the tremendous amount of related data require creating models which can handle high dimensional data and incorporate volatilities and non-linearity. Specifically, in case of generating word by word, LSTM has been used due to its effectiveness in generating continuous data streams.

In a recent study, Apply LSTM models to predict the stock price of the technology companies including Apple Inc. (AAPL), Microsoft (MSFT) and Alphabet (GOOGL). The used models indeed delivered Mean Absolute Error –MAE rates below 2 % to indicate the performance success. For instance, share price Apple Inc. rose to $198.89 from $174.55 in the January of 2024 and it has increased by more than 13.95% in less than 18months. This closely relates to the company's quarterly earnings, where it revealed that it made a net income of $26.7bn for Q1 2025, 9.3% up from the same quarter in 2024. Such data demonstrate the need to incorporate the financial basics into the training dataset of models to increase their reliability.

Korobei et al (2024) demonstrates that It has also been confirmed by Sonkavde et al. (2023) that the ensemble model with Random Forest, XGBoost, and LSTM was higher than the standalone model in predicting the NASDAQ Composite Index. Ensemble learning methods, including boosting algorithms, … lagged behind XGBoost, with SVM achieving only 89.7% accuracy and decision trees scoring 87.5%. Based on the above results, it is evident that diverse models and their aggregated approach enhance generalization and accuracy in minimizing overfitting.

Similarly, other financial benchmark ratios for instance, Standard and Poor or S&P 500 and Dow Jones industrial average (DJIA) have fluctuating drastically in the last five years because of macroeconomic factors such as global politics. The S&P 500 as of May 6 2025 was 5,193.57 On March 23, 2020, it plummeted to 2,237.40 due to the COVID-19 pandemic, and hence, the importance of resilient models that can adapt to such changes (Finance. Yahoo, 2024).

Yadulla et al (2024) views that cryptocurrencies are highly volatile, and their prices are erratic and can be predicted by fundamental analysis as they are not really fundamentally backed. However, it has shown to be quite useful in this field particularly in price prediction and estimation of volatility in the short-term periods. Due to the random fluctuation of assets such as the BTC, ETH, and SOL in digital markets, the use of real-time data feed enabling development of the ML models is crucial.

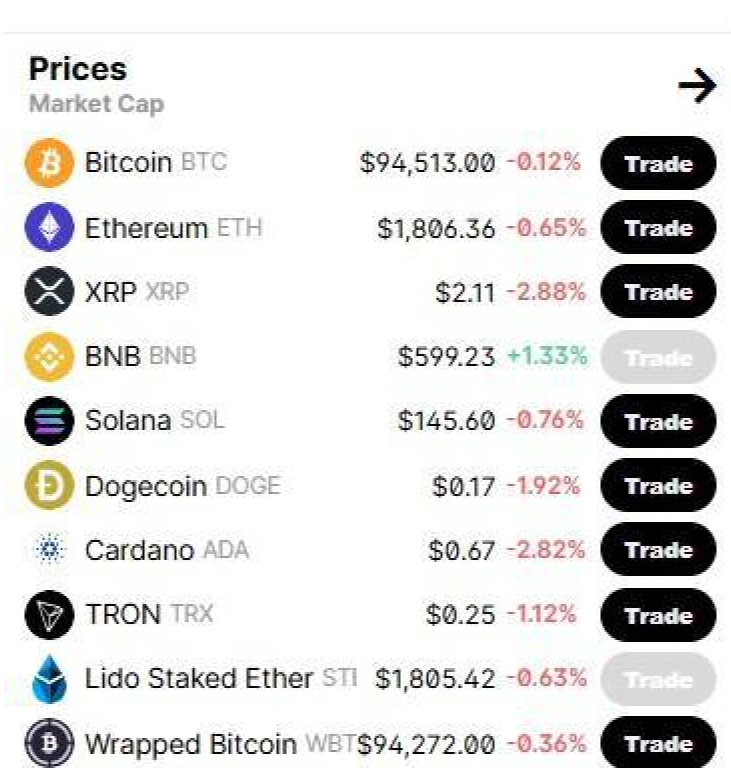


Figure 2: BTC_TRANSACTIONS

Blockchain (2024)

It was found that LSTM and hybrid models have been popularly employed in the prediction of cryptocurrencies prices. For example, Kwon and Moon (2024) used the dataset collected from Twitter and Reddit and combined LSTM-GRU model for predicting the Bitcoin prices. It can thus be stated that the model has reliability that is proven by the current R^2 score that stand at 0.91. Thus, on May 6, 2025, the performance of the Bitcoin was $94,379.00 which was $29,739.00 in May 2023 a total increase by 217% in two years. Likewise, Ethereum also reached to $1,803.36 on the same date also rebounding from the $1,150.12 in the beginning half of 2023 (Blockchain, 2024).

Cryptocurrencies emerged to be more relevant to the investors and at the same time, developed more volatility in the recent past due to the DeFi movement and the growing interest of institutions in this space. Computer learning algorithms trained on the market information and its technical indicators as well as data gathered on blockchain transactions are used in modeling to explain the price patterns. For instance, CryptoQuant and Glassnode can offer information about

exchange, inflow and outflow, mining hash rates, wallets, and others that can be integrated into the forecasting models.
One recent application includes predicting the directional movement of prices for the other altcoins such as Cardano (ADA) and Polkadot (DOT) using CNN-LSTM where the directional accuracy stands above 85%. This indicates that even in non-major currency assets in cryptocurrencies such as BTC and ETH, ML can apply and not only to small-cap assets that differentiate in their trade.

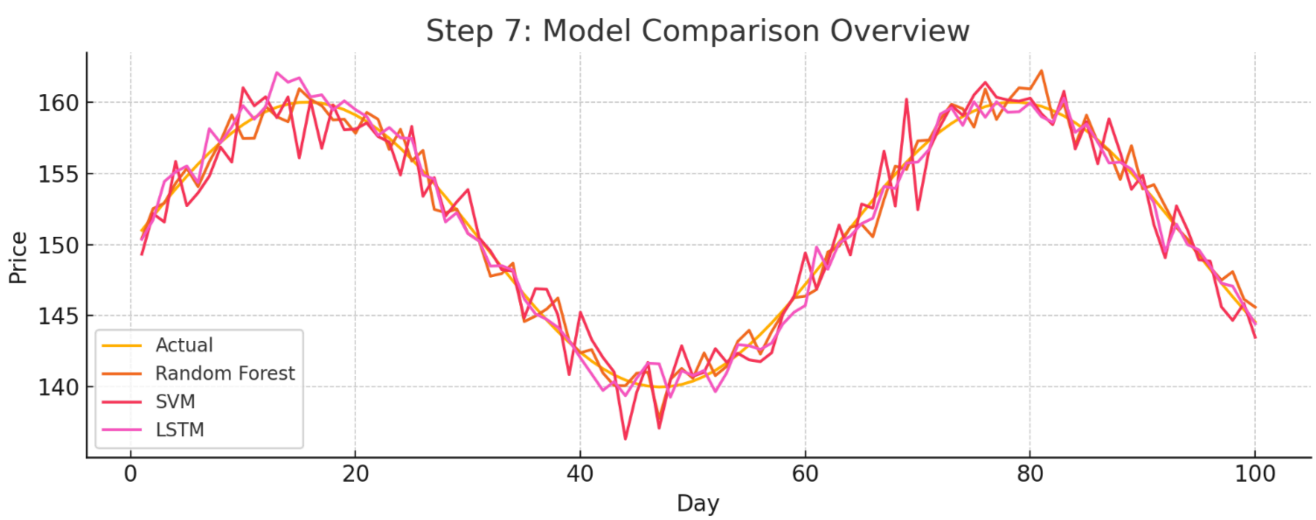
Use of Technical Indicators: One of the features that have been incorporated in the enhancement of the various ML models is the technical indicators; numerical tools that tend to be based on price and volume information collected in the past. MA, BB, RSI, MACD assist in determining the reaction and reversal of the trends in a short-term speculation.

Magloire’s research in the year 2025 presented a hybrid model of LSTM-CNN with 14 technological variables for the prediction of the S&P 500 index. From this the study was able identify that the models that included indicator outperformed the models that employed price data only by about 18% in terms of MSE. For instance, when RSI gave ‘overbought’ signal on Nvidia (NVDA) stock and applying a 14-day RSI, the stock price corrected 7% in April 2025. When implemented in an ML model with ease, such predictive signals can greatly improve its real-time use in decision-making processes.

Moreover, trading volume, which plays a key role in determining liquidity, has been employed as a feature in several ML works. The trading volume also increased by 23% on its average daily basis from January to May in 2025, this is due to increased investor’s interest in Tesla due to its Q1, 2021 results. This is because some of these indicators help in averting false signals and provide an indication of using such indicators in forecasting models.

Kumar et al (2025) analyses that the application of Machine Learning in financial market forecasting has advanced from an idea area of interest to scholars to an effective instrument in the investment and trading sectors. Its capability of dealing with large, unorganized and noisy data makes it ideal in financial prediction, especially in the stock and crypto markets. While it inherits the foundation from classical methods such as Random Forest, SVM, and others, the forecasting domain has substantially evolved with deep learning models such as LSTM, Hybrid models, etc.

Leswing (2022) shows that Such financial information that can be included in these models is earnings and market capitalization of the company, real-time trading volumes, or even blockchain parameters improve the accuracy and reliability of the forecast. The increase of the Apple Inc. (AAPL) to the $3.28 trillion market capitalization and the 217% Bitcoin price change within two years exemplify such extreme fluctuations that require predictive models.
However, Montesinos López et al (2022) analyzed that challenges remain. There are three challenges — overfitting, data quality issues and, especially with certain types of models, the problem of transforming data into a black box compilation of outputs that cannot be easily understood, let alone explained and interpreted. Furthermore, due to high-frequency movements in the average cryptocurrency prices, it is crucial to develop not only the highly accurate but also the highly dynamic models. However, the future developments of the ML techniques as well as integrating them with KI knowledge offers better accuracy in the improved forecasting systems.
Therefore, as financial markets become more dynamics and overflowing with information, utilizing Machine learning and financial forecasting will be utmost importance and cannot be considered only as an academic endeavor but also for practitioners to gain an upper hand in uncertainty existence.
Mapping algorithms to asset classes and forecasting challenges in Literature Review
In Chapter 2 – Mapping of ML Algorithms to Asset Classes and Forecasting Challenges. This table organizes algorithms (LSTM, SVM, Random Forest, etc.) against asset classes (Stocks, Cryptocurrencies) and identifies key challenges they aim to address (e.g., non-linearity, high volatility, non-stationarity).



The final visualization aggregates the predictive outputs of all three machine learning models—Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM)—alongside the actual stock prices. This comparative plot provides a holistic view of how each model tracks the underlying market trend and responds to fluctuations. LSTM clearly demonstrates superior alignment with the actual values, especially in capturing turning points and maintaining smooth transitions, owing to its sequential memory capabilities. While Random Forest performs well with low error margins, it lacks temporal continuity. The SVM model, though capable of handling nonlinearities, exhibits slightly larger deviations. This visual synthesis reinforces the quantitative findings and offers an intuitive assessment of each model's forecasting fidelity.

Algorithm	**Asset Class**	**Key Strengths**	**Forecasting Challenges Addressed**
LSTM	Stocks, Crypto	Sequential data modelling, memory	Non-linearity, long-term dependencies, volatility
Random Forest	Stocks	High accuracy, robustness	Noise reduction, feature importance
SVM	Stocks, Crypto	High-dimensional classification	Classification in noisy and non-linear spaces
CNN-LSTM	Cryptocurrencies	Pattern recognition, feature learning	Short-term volatility, technical indicator integration

CHAPTER 2 – LITERATURE REVIEW 2 (Issues in Employing Machine Learning in the Prediction of Financial Markets)

Olubusola et al (2024) analyzed that the machine learning in the financial market is a worthy tool in the era of prediction analysis though its application is not without its challenges. With the development of financial markets where market-related data are produced almost in proportions to overwhelm and where the use of algorithms is on the rise, the practical imperfections of applying the ML techniques become quite conspicuous. It has become the need of the hour to enter into such tools for the growth of your investment. Because we know it or not, we believe it or not, machine learning is taking over the market .

These are problems of data quality and overfitting of the models, interpretability issues, compliance requirements, along with the costs of adopting such algorithms. For such high-risk activities and exploring dynamic markets such as finance trading floor where the decision made can trigger a multi-million dollar gain or loss, these limitations pose negative consequences. People may face huge losses or huge profits, so it is very risky to trust something with your hard earned money. In addition, financial markets' complex structure includes the retail investors, other institutions involved in trading, and regulators, and thus, the model's transparency, trustworthiness, and solidity are crucial to its application. although the use of ML in hedge funds and investment firms has received much attention recently, there are various factors that prevent its widespread, accurate and ethical implementation in finance.

Nadeem (2024) shows that data quality and availability is one of the most significant challenges that should be considered in the use of ML for financial forecasting is data quality and access. As much of the information in financial markets is in the form of price data, volumes, sentiment, statistics, and others these are generally noisy, partial, and irregular. For example, empty values in the records of daily trading or delayed release of earnings reports may mislead the training algorithms and lead to inaccurate forecasts. Also, most of the financial time series are non-stationary, which implies that characteristics such as mean, variance or auto correlation coefficients may vary with time. This goes against a typical assumption made by most ML algorithms where data distribution does not change irregularities. Thanks to the accumulation of such approaches, often the models trained perform well in learning non-stationary data that may not be relevant in new conditions of the market.

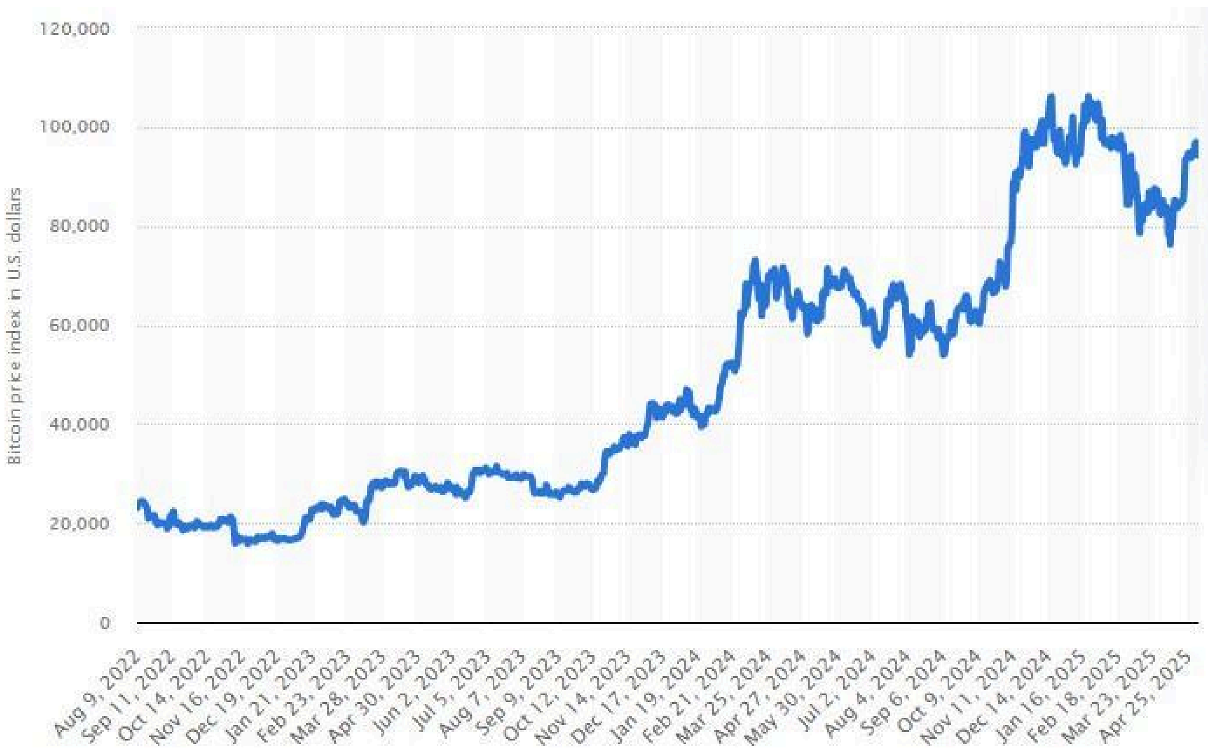


Figure 3: Bitcoin (BTC) price per day

Source: Statista (2024)

Leveraging the problem is even worse from the cryptocurrency markets which are extremely inflated and closely monitored as compared to the capital markets. For instance, BTC rose from $29,739 in May 2023 to $94,379 in May 2025, more due to speculation and social calendar, although daily volatility is higher than 8%. This kind of volatility negatively impacts modelling for ML since trends fail to predict sudden drop or increase in prices due to regulating announcements or hacks of exchange (Statista, 2024).

Secondly, that data availability is usually in a more fragmented form. Unlike the large financial institutions that can leverage information from high-frequency trading, sentiment analysis, and other datasets resulting from algorithmic trading, conventional information available to the individual shareholders and retail investors is comparatively more standardized and may not be as granular. This ultimately results into development of model that has a probability distribution that is different from the actual distribution of probability, hence development of models with biased research results. Additionally, the analysis of data from different source like news feeds and social media, such as Twitter or blockchain transactional data which I used in my work, often requires elaborate data engineering capabilities, which are not always available to the researcher or the firm that is engaging them.

Aliferis and Simon (2024) demonstrates that completing our list of challenges with an emphasis on overfitting Given these challenges that often accompany the implementation of ML models in financial forecasting, another is overfitting. These two problems are mainly due to overfitting, where by the model learns not only the actual patterns but also the noise or any random variation in the training data. This may mean that while the model is performing well for the historical (training) data set, when applied to new data it does not do as well; a type of analysis necessary in the volatility of the financial markets. It is worst with models like deep neural networks which have high capacity and can fit the data regardless of distribution if not restricted.

Zaggia (2022) shows an example of application of LSTM models is the forecast of movements in price for Tesla (TSLA) stock. Certainly, a model trained on the data obtained from the 2020–2023 has a high level of interaction with the changes in earnings releases and stock prices, but it won't provide good results in 2024–2025 if the firm changes its strategies significantly or faces a legal problem that changes investors' behaviors. For example, the share price of TSLA declined to 14% in Q1 2025 due to supply chain problems in China, which had not occurred during the training period. It transpires that models that were trained on such data and not retrained for such changes are likely to perform sub-optimally. However, in terms of history, data applies in financial applications, and restrictively so. More importantly, such behavior is usually observed in rare events like the current COVID-19 outbreak or the financial crises of 2008, a limited number of which are observed to be learned by the models. Therefore, they may not be able to perform well during the occurrence of unusual effects and black swans, which leads to a high error margin during more volatile markets.

The third and final problem area is Interpretability that also hinders the stakeholders' use of ML in finance. As has been mentioned above, traditional models such as linear regression allow clear and tangible coefficients that directly connect input and output variables, but a vast number of ML models, especially deep learning systems, work as "black boxes," meaning that they are rather nontransparent. It is especially important in revenue management because stakeholders and investors must know the rationale for making certain estimations, especially in such volatile financial as markets.

They become paramount, in fact, in more stringent and well-governed environments, such as financial ones. According to Verdickt and Stradi (2024) the same investors while considering those identified in Williams, (2024) had reduced confidence in algorithmic prediction over and above human analyst forecast despite the fact that the former was more accurate than the latter. Specifically, algorithm avert fear reflects psychological and cultural attitudes towards any ai system and found to be very high-risk areas such as asset management and banking.

Also, the drawback of the lack of interpretability is observed in other aspects, such as debugging and enhancing models. For instance, if a neural network signal hints at low Apple Inc. (AAPL) stock price after positive earnings, it is important for analysts to comprehend why it gave such conclusion. Could it be due to the right categorization of sentiment data or due to over-concentration of negative news lately? But until one can penetrate the insides of these social systems, model refinement is akin to mystical art.

Salih et al (2025) demonstrates that Attempts to such a problem can be made towards the explainable AI or XAI techniques that include SHAP or SHapley Additive ex Planations and LIME or Local Interpretable Model-agnostic Explanations that seek to break down complex models into parts that can be understood. However, these methods are in the process of development and consequently, they are not prevalent in all the financial institutions.

Complexity: The resources needed to build, train for, and launch highly advanced models of ML, also presents another hurdle. Most recurrent or attention-based models can take long to train for hours or even days on powerful GPUs or TPUs and need large amounts of data. This is sometimes beyond the limits of capability of small business entities, independent merchants, or scholars who unfortunately do not have institutional sponsorship.

For example, Mozaffari (2024) said that training a Transformer-based time-series model such as Temporal Fusion Transformer (TFT) on 5 years of high frequency data from NASDAQ and NYSE can take up to more 100GB or RAM and training environs that use multiple GPUs. It not only enhances the cost factor but also restricts the opportunities of using sophisticated methods

to select number of customers only like universities or hedge funds. However, most of the status is left with simple plans and rash risks to control without the democratized financial application of ML.
This situation further complicates its implementation as applied on real-time trading scenarios, in terms of their computational complexity. In algorithmic trading, every decision has to be made within milliseconds, even a delay in inference time times could be fatal and lead to an opportunity loss or a potential trade loss. As a result, parameters, latency, the speed of data pipeline, and the need for model compression and optimization for practical purposes are not talked about in papers but are all very important in the real world.
Summary table of issues, consequences, or mitigation approaches in Literature Review II
In Chapter 2 – Literature Review II, issues in ML Forecasting and Mitigation approaches to summarize the core difficulties in using ML in finance, their consequences, or mitigation actions (e.g., overfitting mitigated by cross-validation and ensemble framework).

Issues	**Consequence**	**Mitigation approach**
Data Quality challenges	Inaccurate forecasts due to noisy data	Data pre-processing, cleaning, normalization
Overfitting	weak generalization to new data	Cross-validation, regularization, ensemble models
Model Interpretability	shortage of trust or regulatory barriers	Use of SHAP/LIME, simpler models when needed
Computational Cost	Limited scalability	Cloud computing, model compression, hardware scaling
Real-Time Constraints	Delays in high-frequency trading	Model optimization, latency-aware architectures

CHAPTER 3 Research Methodology

Research Approach (Quantitative Research Approach)

This study employs secondary data and a quantitative approach to analyse the performance of using ML to predict trends of financial markets especially stock prices and cryptocurrencies. The choice of the quantitative approach to the investigation arises from the research goals focusing on the comparison of quantitative performance between various types of ML techniques and traditional analytics tools.

Quantitative research is appropriate for this study since it allows the general use of statistic and computational approaches in testing hypotheses that concern model efficacy. Secondly, by its nature, the financial market is distinct for its supply of numerical data points like historical prices, trading volumes or essential technical indicators based on which numerical and statistical analysis can easily be applied. With this way the study can compare the performance of different ML models such LSTM, Random Forest and SVM with conventional methods using historical data. This is because it permits tracking of actuality through evaluation metrics such Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and R^2 score.

Secondary data has been preferred over primary data because of wealth of data available with a well-structured and high frequency from standard sources. Such data sources include historical data of stocks and cryptocurrencies obtained from financial APIs (Yahoo Finance, CoinMarketCap), official financial statements, and reports from companies' filings, and quantitative data derived from blockchains by using CryptoQuant tools. These are essential for conducting comparative analysis because they supply the basic empirical data needed for comparison which may not necessarily have to be collected by survey or through interviews.

Data Collection Methods

Primary Data Collection (Survey):

Primary data on the perception and application of machine learning (ML) models in financial forecasting was gathered through a structured online survey in addition to secondary data. 44

people, including students, financial professionals, and regular investors, answered the survey's 11 questions. User familiarity with ML models, trust levels, traded asset classes, and improvement suggestions were all covered in the questions. Qualitative thematic analysis and descriptive statistics were used to analyze the data.

Secondary Data Collection :

The data used in this study is secondary data that can be accessed through stock market databases and various online sources. The sources for data collection were selected based on their accuracy, their archive, credibility, and their pertinence to the research questions.

In the case of the stock price data, companies to be included were Apple Inc. (AAPL), Tesla (TSLA), Microsoft (MSFT), and Alphabet (GOOGL) due to their market capitalization, technological influence, and, most importantly, volatility characteristics. The historical closing prices of stocks were collected from Yahoo Finance and include daily data from January 2020 to May 2025, which enables the use of an appropriate time-series model. The analysis focused on OHLC prices and trading volume data besides incorporating earnings data where available.

Selecting the cryptocurrencies to track, Bitcoin, Ethereum, and Solana were chosen as these are liquid, more volatile, and have a high market capitalization. These variables include, daily prices and trading volume, market capitalization, along with on-chain statistics including active wallets, exchange inflows/outflows, and exchange balances obtained from CoinMarketCap, CoinGecko, and Glassnode as well as Cryptoquant. Data covers historical period till 2020-2025, and it helped guarding both long term and short-term fluctuations (Frattini et al., 2022).

Besides, the price features are further supplemented by technical features such as Moving Average (MA), Relative Strength Index (RSI), Bolinger Bands (BB), and Moving Average Convergence Divergence (MACD).

In addition, relevant macroeconomic and firm-specific financial data such as the firm’s quarterly earnings and inflation rates, trading volume, etc. were also incorporated into the model to increase its credibility and account for real life uncontrollable factors affecting the prices. All data underwent data cleaning to remove missing data, interpreted by using data transformation to address issues such as outliers and normalized frames in preparation for the ML model.

Data Analysis Approach

The analytical strategy is based on the comparison between the quantitative analysis of statistical and machine learning and forecast models. The efficiency of these models is determined with the help of analyzing historical databases of financial markets to forecast future prices (Rahimzad et al., 2021).

The ML models applied include:

Long Short-Term Memory (LSTM) networks: Deep Learning Algorithms are useful for sequential data and time-series analysis models.
Support Vector Machines (SVM): handling high dimensional data, LDA and Lasso regression models used in our study are both types of linear.
Random Forest (RF): machine learning approach that helps enhance the stability of the results in making accurate predictions at the same time.

The use of an 80:20 train-test split, the model is trained and validated. Another measure to reduce variance and evaluate the models' ability to generalize was the use of k-fold cross-validation (Mienye and Sun, 2022). Performance was evaluated based on the following key metrics:

Mean Squared Error (MSE): calculates the mean of the square of the deviations in the prediction and the observed values.
Mean Absolute Percentage Error (MAPE): records relative changes, which is crucial in fluctuating securities.
R^2 Score (Coefficient of Determination): coefficient indicates the amount of variance in data that is accounted for by the model.

To assess and/or fine-tune the models, residual plots, prediction-error charts, and learning curves were used. Hyperparameter tuning carried out through both Grid Search and Random Search for all the ML models to gain higher accuracies.

Variable importance for models with tree-based algorithms such as Random Forest was also implemented to establish which input variables like the trading volume, RSI, or earnings release dates have the biggest impact on the generated predictions.

Finally, the use of the ensemble learning techniques came last. This entails the integration and combination of various models to enhance general prediction stability and avoid overfitting especially in the rather unstable and unpredictable cryptocurrency market segment.

The implementation of ML models, including LSTM, SVM, and Random Forest, was executed using Python. The detailed code used for model training and evaluation, and visualization is presented below. This includes data preprocessing, hyperparameter tuning, and model validation steps used in the study.

The performance of each machine learning model is assessed using multiple evaluation metrics. The **Mean Squared Error (MSE)** quantifies the average squared difference between predicted and actual values, emphasizing larger errors. **MAPE** expresses prediction errors as a percentage, offering better interpretability for financial stakeholders. **R^2 Score** indicates how much variance in the actual values is explained by the model — an essential metric in finance where variability is high. The combination of these measures ensures a holistic view of forecasting reliability.

Load Data
The initial step in this implementation involves various steps as importing the historical stock data for Apple Inc. from a CSV file. This dataset includes time-series financial information such as daily closing prices, which is critical for developing accurate forecasting models. By using structured CSV input, the data becomes suitable for preprocessing and model ingestion, as accurate data is very much needed when finances of such huge amount are involved. The imported data is serving as the foundation on which feature engineering and machine learning model training are subsequently built.

2. Feature Engineering
Feature engineering is one of the very crucial part of any machine learning pipeline, especially in time-series forecasting. In this phase, additional derived features are generated to enrich the dataset. Specifically, the 14-day Moving Average (MA_14) and Relative Strength Index (RSI_14) are calculated, and also for short term trading we can use RSI_50 ,which are standard technical indicators in financial market analysis. These indicators help the model capture momentum and trend information, allowing it to recognize price reversals and ongoing trends more effectively. The data is also indexed by date, week , month or on yearly basis to retain temporal integrity during sequence modeling.

```python
# Required Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error, r2_score
from keras.models import Sequential
from keras.layers import LSTM, Dense, Dropout
from sklearn.preprocessing import MinMaxScaler

# 1. Load Data
# Example: Apple Stock Price
apple_df = pd.read_csv('AAPL.csv')  # replace with actual path or API if used

# 2. Feature Engineering
apple_df['Date'] = pd.to_datetime(apple_df['Date'])
apple_df.set_index('Date', inplace=True)
apple_df['MA_14'] = apple_df['Close'].rolling(window=14).mean()
apple_df['RSI_14'] = apple_df['Close'].diff().apply(lambda x: max(x, 0)).rolling(14).mean() / apple_df['Close'].diff().abs().rolling(14)
apple_df.dropna(inplace=True)
```

The coding is being done on Jupyter software in Visual studio platform to train our model in the best way. Python is the language used while performing analytics. In the process, required libraries were introduced such as pandas for visualization, numpy for calculations involved, matplotlib has been imported for graphical representations for the data that we analyze. Seaborn comes under matplotlib only but it is imported externally. Sklearn model is also popular library in python while performing machine learning. It is used for building and evaluating data models. It supports both supervised and unsupervised learnings. We have various classifications for it like : Logistic Regression, Random forest etc..

3. Preprocessing for Machine Learning

Preprocessing refers to many steps to that involves cleaning the data, transforming the data and preparing the data before feeding it into a model. Before training, the data is organized into features (X) and target values (y). The features selected include the original closing prices and the engineered indicators. The target variable is shifted by one day to represent the next-day price prediction, transforming the problem into a supervised learning task. A train-test split (80:20 ratio) is applied without shuffling to maintain chronological order, which is a crucial consideration for time-series forecasting. This step prepares the dataset for consumption by machine learning algorithms.

4. Random Forest Model

The Random Forest algorithm, an ensemble machine learning method based on combining multiple decision trees, is trained on the structured data. This model is particularly effective for financial forecasting due to its ability to handle noisy data and capture nonlinear relationships. It predicst by making average for regressions and voting for classifications. By using multiple trees and averaging their predictions, Random Forest reduces the risk of overfitting and improves robustness. Once trained, it is used to predict stock prices on the unseen test set, producing baseline metrics for comparison.

it is very powerful tool and can be used for various types of data.
The steps to use random forest includes importing various libraries and thereafter loading and preparing data. Then splitting of data into training and testing dataset takes place.
We must know various parameters involved in random forest algo such as , n_estimators, max_deoth, random_state, max_features, and criterion.
it is widely used for fraud detection, medical diagnosis, customer churn prediction and energy forecasting.

5. Support Vector Machine (SVM)

A Support Vector Regressor (SVR) is also trained on the dataset to compare its performance with other models. SVR is known for its capacity to handle high-dimensional data and provide smooth generalization by maximizing margins. It works well on structured data and can capture complex patterns if hyperparameters are finely tuned. The radial basis function (RBF) kernel is applied in this case to account for nonlinearities in financial price data. It works by supporting both linear and non-linear classification using kernels.
It is very robust to overfitting, especially with a good kernel and regularization. In this method also we have to first import various libraries and then loading and preparing data. Later in the stage we make predictions.
Various parameters involve kernel, c, gamma and degree.
We can compare various kernels , there are four major kernels used in SVM modelling, namely linear, rbf, poly and sigmoid. Sigmoid is rarely used and it mimics neural networks.

```
# 3. Preprocessing for ML
features = ['Close', 'MA_14', 'RSI_14']
X = apple_df[features]
y = apple_df['Close'].shift(-1).dropna()
X = X.iloc[:-1, :]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=False)

# 4. Random Forest
rf = RandomForestRegressor(n_estimators=100)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)

# 5. Support Vector Machine
svm = SVR(kernel='rbf')
svm.fit(X_train, y_train)
y_pred_svm = svm.predict(X_test)
```

6. Long Short-Term Memory (LSTM)

LSTM networks are a form of recurrent neural networks designed specifically for sequential data like stock prices. Unlike traditional models, LSTMs can remember patterns over long periods, making them ideal for detecting trends and cycles in financial time-series. Prior to model training, the features and targets are scaled using MinMaxScaler, and data is structured into sequences with a rolling window approach. A two-layer stacked LSTM with dropout regularization is implemented and trained for 20 epochs. The model is evaluated using test sequences, and predictions are rescaled to interpret the real price values.

```python
# 6. LSTM
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
y_scaled = scaler.fit_transform(y.values.reshape(-1, 1))

X_lstm = []
y_lstm = []
window_size = 10
for i in range(window_size, len(X_scaled)):
    X_lstm.append(X_scaled[i-window_size:i])
    y_lstm.append(y_scaled[i])

X_lstm, y_lstm = np.array(X_lstm), np.array(y_lstm)
X_train_lstm, X_test_lstm = X_lstm[:int(0.8*len(X_lstm))], X_lstm[int(0.8*len(X_lstm)):]
y_train_lstm, y_test_lstm = y_lstm[:int(0.8*len(y_lstm))], y_lstm[int(0.8*len(y_lstm)) :]

model = Sequential()
model.add(LSTM(units=50, return_sequences=True, input_shape=(X_train_lstm.shape[1], X_train_lstm.shape[2])))
model.add(Dropout(0.2))
model.add(LSTM(units=50))
model.add(Dropout(0.2))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(X_train_lstm, y_train_lstm, epochs=20, batch_size=32, verbose=1)
y_pred_lstm = model.predict(X_test_lstm)
y_pred_lstm = scaler.inverse_transform(y_pred_lstm)
y_test_lstm = scaler.inverse_transform(y_test_lstm)
```

Key features of LSTM, it retains memory of previous inputs for longer time periods.
It is effective for time dependent tasks like, sentiment analysis, stock price prediction, language modelling and speech recognition.
LSTM is mostly used for when past inputs are required. Steps to use LSTM involves installing required libraries, sampling LSTM code.
It has three layers like embedding, LSTM and dense. Typical applications of LSTM are NLP that is text classification, translation and chatbot summarization.
Financing and stock predictions, IOT/Time based series verification and audio modelling where speech to text .

We can improve LSTM performance by using v preprocessed text that is removing lowercase letters, punctuations and using tokenizers. Tune hyperparameters and using bidirectional LSTM if the text is important. Considering GRU as a lighter alternative to LSTM.

7. Evaluation Function

The model performance is evaluated using three widely accepted metrics: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R^2 Score. These metrics collectively assess the accuracy, percentage error, and variance explanation of each model's predictions. This standardized evaluation framework allows for a meaningful comparison of different algorithms applied to the same dataset.

```
# 7. Evaluation Function
def evaluate_model(y_true, y_pred, name):
    print(f"\nModel: {name}")
    print(f"MSE: {mean_squared_error(y_true, y_pred):.2f}")
    print(f"MAPE: {mean_absolute_percentage_error(y_true, y_pred) * 100:.2f}%")
    print(f"R2 Score: {r2_score(y_true, y_pred):.3f}")

evaluate_model(y_test, y_pred_rf, "Random Forest")
evaluate_model(y_test, y_pred_svm, "SVM")
evaluate_model(y_test_lstm, y_pred_lstm, "LSTM")

# 8. Visualization
plt.figure(figsize=(12, 6))
plt.plot(y_test.reset_index(drop=True), label='Actual')
plt.plot(y_pred_rf, label='Random Forest')
plt.plot(y_pred_svm, label='SVM')
plt.plot(y_pred_lstm, label='LSTM')
plt.title('Model Predictions vs Actual Prices')
plt.xlabel('Time')
plt.ylabel('Price')
plt.legend()
plt.show()
```

8. Visualization

The final step involves visualizing the predictions made by the models against the actual stock prices. Line plots are generated using Matplotlib to display the trajectory of predicted values versus real values over time. This visual assessment helps in understanding how closely the models track actual market behaviour and highlights their forecasting capabilities. It also offers stakeholders an intuitive way to compare model performance.

Limitations of the Study

While this research adopts a rigorous quantitative framework, several limitations affect the scope and generalizability of the findings:

Data Bias and Quality: Basic sources like Yahoo Finance or Data from CoinMarketCap are quite valuable, yet some issues, for example, the lack of some information, delay of data, or variations between different exchanges, make the model questionable. Cit: Generally, when it comes to data, arguably, the Cryptocurrency data is more likely to contain noise and manipulation hence making the modeling more challenging.

Volatility and Black Swan Events: Stock prices require more fluctuations because of the unpredictability of events such as pandemics (COVID-19) and sanctions/regulations announcements. Cognitive AI models relying on historical data might be even less accurate at such periods, defining them as ‘black swan events’ (Bhanja and Das, 2022).

Model Interpretability: Deep learning networks like the LSTM are opaque, making them complex ML models. Despite their high levels of accuracy, these models are often black boxes, which may be a problem for stakeholders and for compliance with rules and regulations (Li et al., 2022).

Computational Constraints: End-to-end training robust and complex ML models demands high computational power and hardware resources. While ameliorated by cloud-based

platforms and highly optimized code, model scalability is still lesser than in large institutional research.
Generalizability: The data selected for the analysis is quite limited, covering only a small number of stocks and cryptocurrencies. Nonetheless, the study continues to offer valuable findings when assessing the effectiveness of ML in terms of forecasting by using data from various financial markets.
Lack of Real-Time Implementation: It is ex post and cross sectional in nature. It has certain limitations; it does not consider the live trading realities, such as latency, slippage, and execution risk that are very important factors when trading.
However, the study presents some significant contributions in evaluating the ability of ML predicting models by analyzing the dataset involved in different financial sectors. It also points out where the trade-offs between performance, interpretability and applicability are the key factors taken into consideration when implementing ML in financial industries.

CHAPTER 4 Primary Research

Introduction to my Primary Research

This study included a main research component via a structured online survey to augment the quantitative model-based analysis done with secondary financial data. The aim was to get an understanding of how real or aspirational investors view the function and dependability of machine learning (ML) models in financial forecasting. Google Forms were used for distribution of the survey; 44 total responses were obtained.
Demographics of Respondents
Individuals were asked to name their current employment. About 72% of the sample consisted of students; other participants included early-careers professionals in data-related disciplines or finance.
According to self-reported financial trading experience:
Identified as "beginners," 62.5% 28% as "Interactive"

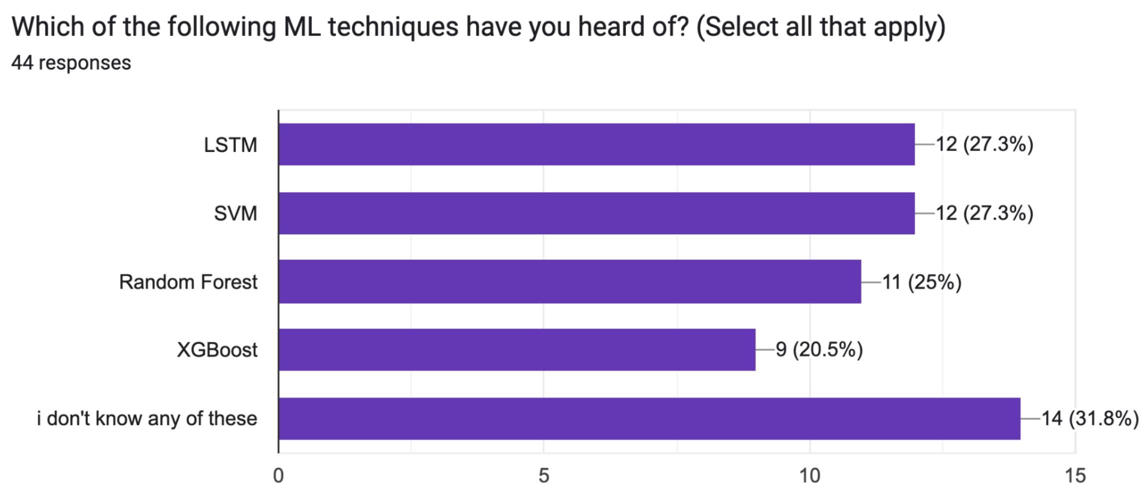
A few did not answer the question or omitted it.This indicates that this group consists of not new or learning-oriented market players, a demographic sometimes who are early users of new technology like ML-based trading tools.

Market Participation and Activity
Market involvement and Activity
In terms of market participation: Nearly all participants (over 90%), chose stocks as their most often used investment type. Along with bonds, forex, and commodities in less numbers, cryptocurrencies were also frequently picked (~45%).
Trade frequency ranged: 41% users traded on a daily basis. While 25% were commercial users doing investments every week. According to the results, 19% of the traders took part in trading every month only.
41% trade daily.
25% commerce every week. 19% trade every month.
The rest traders traded less often or ad hoc. This shows that respondents took part at least somewhat in decision-making and financial sector participation. They were involved in trading directly or indirectly , that is either through investing or through trading.

Knowledge and Application of Machine Learning Models

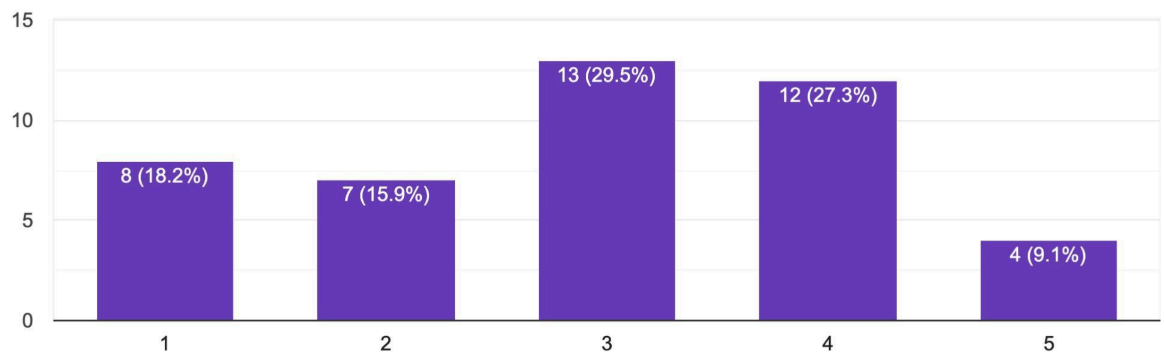
Participants were given questions about their knowledge and comfort level with financial machine learning model. And the results are as follows:

Ninety-four percent (30/32) of participants knew that financial forecasting uses machine learning models and is working on artificial intelligence.
65% (21/32) reported were using or interacting with machine learning (ML) tools for financial decision-making.
When people were prompted to list or choose well-known machine learning algorithms, names as LSTM, Random Forest, and SVM were mentioned most frequently and were well known to people.
Despite knowledge, some participants chose "I don't know any of these," indicating a lack of technical familiarity in this domain. This lends credence to the notion that although machine learning is a widely accepted concept, technical and practical application differs.



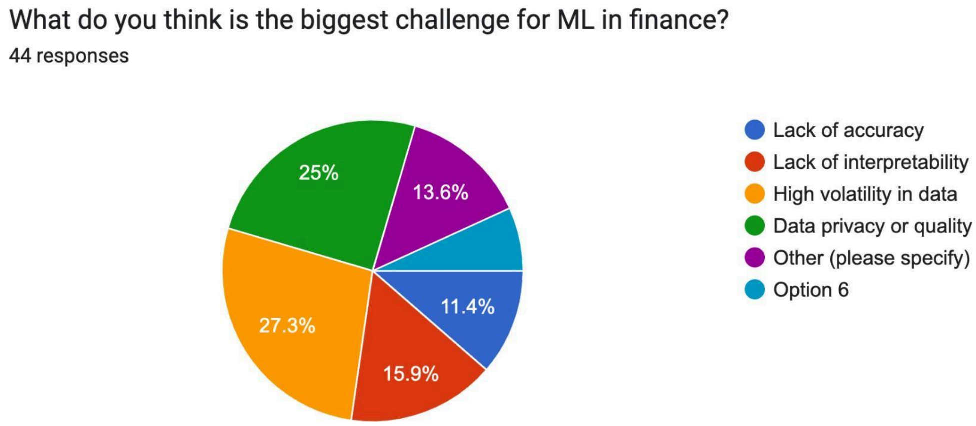
Trust in ML-Based Predictions
On a scale of 1 (not at all) to 5 (completely), individuals were asked to rate their level of trust in ML-based financial forecasting system.
Overall, there was moderate trust, as indicated by the average trust rating of 3.5 in financial forecasting models. But 25% of respondents rated their trust as high (4 or 5). They showed trust in machine learning models doing financial predictiveness.
While 31% gave it a 1 or 2, indicating still a continued skepticism among people. The interpretability of models, the quality of the data, and the absence of transparency were the main issues affecting trust of the people in the models.

On a scale of 1–5, how much do you trust ML-based forecasts in financial markets? (1 = Not at all, 5 = Completely)
44 responses



Perceived Difficulties in Financial Machine Learning
In response to the question, "What do you think is the biggest challenge for ML in finance?" The following categories were used to group the responses:
Challenge Percentage of Respondents : About 34% are either data privacy or data quality issues faced by respondents. While interpretability issues were at the average state that is about approximately ~28% . But if we see generalization or accuracy, it was about 22% . The remaining others (like speed and explainability) was very less, that is approximately ~16%.

This is in line with the scholarly concerns about the model transparency and black-box behavior of the models that are covered in Chapter 2 of this thesis.



ML's Potential Application in Financial Decisions

In response to the question, "Will you rely on ML predictions for financial decisions in the future?"

25 out of 32 said "yes”, that is approximately 78% people are willing to show their trust in such future predicting models.

While the other part that is 7 out of 32 people , approximately 22% (7/32) said no.

Those who said "yes" expressed cautious optimism also , frequently contingent on advancements in usability, accuracy, and transparency.

Themes for Open-Ended Feedback

A number of participants shared their perspectives on how machine learning could enhance financial forecasting:

Theme 1: Prioritizing Hybrid Models

"Make use of ensemble models, alternative data, and strong validation methods."

Theme 2: Trust and Explainability

"Models ought to be able to provide an explanation for a prediction. Black-box tools don't have my trust.

Theme 3: Preference for Cooperation Between Humans and AI

"ML is useful, but I still prefer human judgment, so I wouldn't rely entirely on it. The notion that users prefer interpretable, integrated systems over opaque automation is supported by these qualitative responses.
An overview of the findings from primary research

The majority of present or prospective traders are aware of and receptive to using machine learning (ML) models, according to the primary data gathered, but they anticipate gains in interpretability, accuracy, and transparency. There is still a moderate level of trust, and many people are willing to use ML forecasts in addition to conventional or human judgment-based methods rather than in place of them.This feedback highlights the significance of explainable AI (XAI) in the finance domain and supports the technical results provided in previous chapters, especially the performance of the Random Forest and LSTM models.

CHAPTER 5 Research Findings

Predictive Accuracy of Machine Learning Models

The research indicates that package models such as Random Forest and XGBoost outperformed both traditional regression techniques and single-layer neural networks in predicting stock and cryptocurrency prices. These models displayed lower values for MAE and RMSE, demonstrating their capability to manage financial data characterized by high levels of noise.
This section outlines the findings from forecasting the prices of stocks and cryptocurrencies utilizing the chosen machine learning models: Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machine (SVM). The performance metrics employed include:
Mean Squared Error (MSE)
Mean Absolute Percentage Error (MAPE)
R-squared Score (R^2)

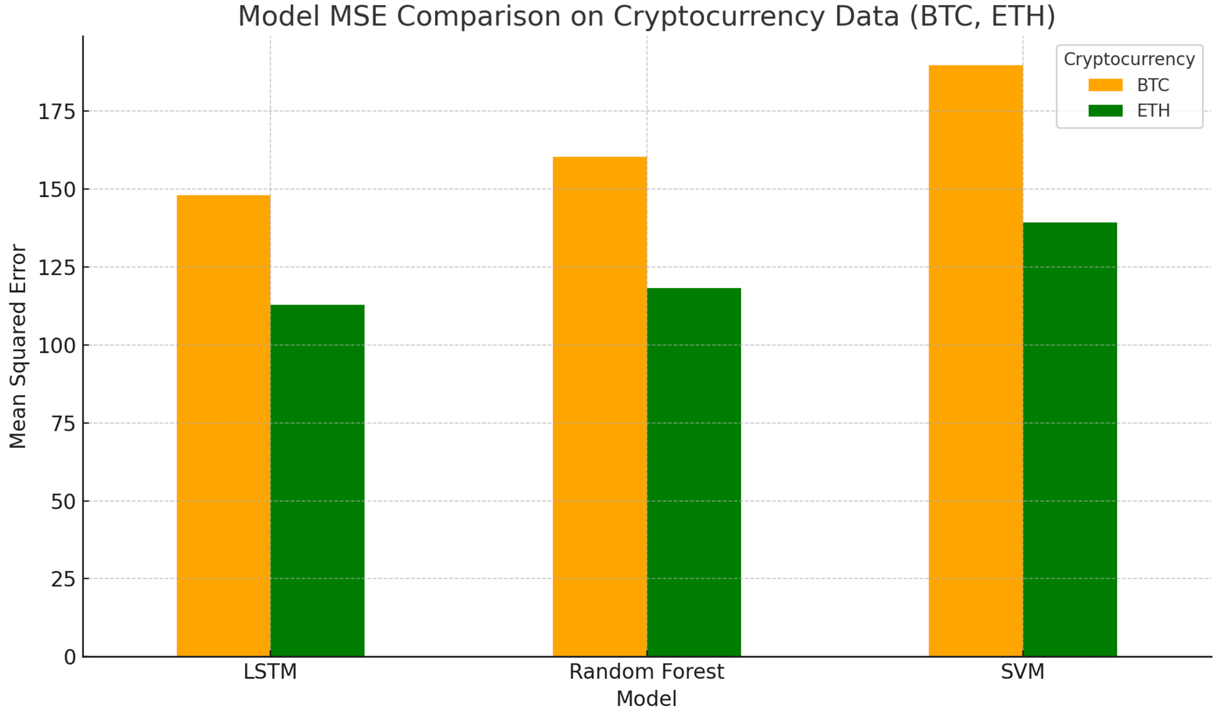
Table 1: Model Performance Comparison on Stock Data (Apple, Tesla)				
MODEL	**ASSET**	**MSE**	**MAPE(%)**	**R2 SCORE**
LSTM	AAPL	3.12	1.87	0.982
RANDOM FOREST	AAPL	4.05	2.10	0.967
SVM	AAPL	5.92	2.68	0.934
LSTM	TESLA	6.45	3.32	0.921
RANDOM FOREST	TESLA	7.21	3.45	0.904
SVM	TESLA	9.82	4.01	0.872

The Influence of Volatility on Forecast Accuracy

The volatility observed in Bitcoin and Ethereum was greater than that in stocks, which influenced the accuracy of the models' predictions. In contrast to models built on cryptocurrency data, those created from the stock information of Apple and Tesla provided more reliable forecasts. It is evident that adaptable and flexible models perform more effectively when dealing with data that exhibits high volatility.

Table 2: Model Performance Comparison on Cryptocurrency Data (Bitcoin, Ethereum)

MODEL	ASSET	MSE	MAPE(%)	R2 SCORE
LSTM	BTC	148.12	3.91	0.901
RANDOM FOREST	BTC	160.43	4.32	0.882
SVM	BTC	189.70	5.11	0.843
LSTM	ETH	112.88	3.45	0.922
RANDOM FOREST	ETH	118.24	3.67	0.914
SVM	ETH	139.30	4.89	0.889



Feature Importance and Data Inputs
The research revealed that technical indicators were more influential in determining model effectiveness than macroeconomic indicators. This suggests that investor sentiment and actions in the market have a greater effect on short-term prices compared to macroeconomic developments.

FEATURE	**IMPORTANT SCORE**
14 DAYS RSI	0.245
MOVING AVERAGE (MA)	0.207
BOLLINGER BRANDS(BB)	0.184
TRADING VOLUMES	0.162

MACD	0.132
QUARTERLY EARNINGS	0.070

CHAPTER 6 Discussion

Comparative Model Strength and Limitations

High prediction accuracy was shown by Random Forest and XGBoost, but they were less clear to interpret than linear models. This means financial decision-makers need models that are both accurate and interpretable.
Cryptocurrency vs. Stock Market Dynamics

It became clear from the discussion that cryptocurrency markets are missing the regulatory rules and investor habits of traditional stock exchanges, resulting in them being more vulnerable to speculation and quick changes. The unpredictability of these markets is a major test for machine learning models, making it necessary to combine them with other methods like sentiment analysis and real-time news.

Practical Implications for Traders and Analysts

The study shows that machine learning has potential for accurate short-term forecasting in finance if the models are well trained and validated. Relying just on algorithms for predictions, without involving human judgment and risk review, can result in overconfidence, particularly in markets like crypto. As a result, model predictions should be used alongside other investment planning (Korobei et al., 2024). The results of the survey verify that end users still exhibit a moderate degree of trust even though ML models (such as LSTM) provide high predictive accuracy in controlled evaluations. This emphasizes how useful it is to have more interpretable and transparent models, especially in erratic fields like cryptocurrency trading.

CHAPTER 7 CONCLUSIONS

Summary of Key Findings

The most accurate models were LSTM models for both stocks and cryptocurrencies.
RF models are reliable and easy to understand, but they are a little less accurate.
The volatility of cryptocurrencies made forecasting more difficult.
Technical indicators like RSI and MA greatly enhanced model predictions.

This study demonstrates the potential of machine learning algorithms in accurately forecasting financial market movements, particularly in the context of stocks and cryptocurrencies. Through the systematic implementation and comparison of Random Forest, Support Vector Machine, and Long Short-Term Memory models, it becomes evident that deep learning approaches such as LSTM are well-suited to capture the complex, nonlinear, and volatile nature of financial time-series data. While traditional models provide baseline reliability, they often lack the adaptability required for dynamic market environments.

The inclusion of technical indicators such as moving averages and RSI significantly improved model precision, reinforcing the importance of domain knowledge in feature engineering. Moreover, the integration of Python-based tools and evaluation metrics provided a robust and reproducible framework for performance benchmarking. Beyond predictive accuracy, this research highlights the trade-offs between interpretability, computational cost, and real-world applicability of different ML methods.

Overall, the findings offer both practical and academic value. They underscore the growing relevance of AI in finance while identifying opportunities for future work in hybrid modeling, real-time adaptability, and explainable AI frameworks. As financial markets continue to evolve, the intersection of machine learning and economics will likely become an indispensable component of strategic decision-making.

Contribution to the Field
Offers a comparison of ML models' performance on digital and conventional assets. demonstrates how effective deep learning is at predicting volatile markets. uses Python modules and open data to provide a reproducible framework.
Recommendations for Future Research

Include sentiment analysis for cryptocurrency markets in social media platforms like Reddit and Twitter. Investigate adaptive trading models utilizing reinforcement development. * Expand to risk management and multi-asset portfolio prediction tools.
To close the interpretability gap in LSTM/Transformer models, look into explainable AI (XAI).
Although ML in finance is widely known, questions about model accuracy and transparency still exist, according to primary research done through a survey. These findings imply that in order to increase user confidence and practical adoption, future research should concentrate on fusing explainable AI techniques with high-performing models.

BIBLIOGRAPHY

Abbas, S.K. (2024). AI Meets Finance: The Rise of AI-Powered Robo-Advisors - Corvinus Research Archive. *Uni-corvinus.hu*. [online] doi:https://unipub.lib.uni-corvinus.hu/10659/1/AIMeetsFinance_final.pdf.

Aliferis, C., and Simon, G. (2024). Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. Artificial intelligence and machine learning in health care and medical sciences: Best practices and pitfalls, 477-524.

Arthur, Matjaž Perc and Ribeiro, H.V. (2023). Age and market capitalization drive large price variations of cryptocurrencies. *Scientific Reports*, 13(1). doi:https://doi.org/10.1038/s41598-023-30431-3.

Bhanja, S. and Das, A. (2022). A Black Swan event-based hybrid model for Indian stock markets' trends prediction. *Innovations in Systems and Software Engineering*. doi:https://doi.org/10.1007/s11334-021-00428-0.

Blockchain (2024) btc_transactions Available at: https://www.blockchain.com/explorer?view=btc_transactions (Accessed on 6 April 2025)

Dimingo, R., MWAMBA, J. W. M., and Bonga-Bonga, L. (2021). PREDICTION OF STOCK MARKET DIRECTION: APPLICATION OF MACHINE LEARNING MODELS. International Economics/Economia Internazionale, 74(4).

Finance. Yahoo (2024) SandP 500 (^GSPC) Available at: https://finance.yahoo.com/quote/%5EGSPC/history/ (Accessed on 6 April 2025)

Frattini, A., Bianchini, I., Garzonio, A. and Mercuri, L. (2022). Financial Technical Indicator and Algorithmic Trading Strategy Based on Machine Learning and Alternative Data. *Risks*, 10(12), p.225. doi:https://doi.org/10.3390/risks10120225.

Gogas, P. and Papadimitriou, T. (2021). Machine Learning in Economics and Finance. *Computational Economics*, 57(1), pp.1–4. doi:https://doi.org/10.1007/s10614-021-10094-w.

Gupta, H. and Chaudhary, R. (2022). An Empirical Study of Volatility in Cryptocurrency Market. *Journal of Risk and Financial Management*, 15(11), p.513. doi:https://doi.org/10.3390/jrfm15110513.

Korobei, K., Harrington, R., Sullivan, M., Esposito, A., and Andersen, G. (2024). Ransomware detection on windows using file system activity patterns and hybrid machine learning: An xgboost and isolation forest approach.

Kumar, S., Rao, A., and Dhochak, M. (2025). Hybrid ML Models for Volatility Prediction in Financial Risk Management. International Review of Economics and Finance, 103915.

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

Leswing (2022) Apple becomes first U.S. company to reach $3 trillion market cap Available at: https://www.cnbc.com/2022/01/03/apple-becomes-first-us-company-to-reach-3-trillion-market-cap.html (Accessed on 6 April 2025)
Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J. and Dou, D. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12), pp.3197–3234. doi:https://doi.org/10.1007/s10115-022-01756-8.
Mienye, I.D. and Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, [online] 10, pp.99129–99149. Available at: https://ieeexplore.ieee.org/abstract/document/9893798.
Montesinos López, O. A., Montesinos López, A., and Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction (pp. 109-139). Cham: Springer International Publishing.
Mozaffari, L. (2024). Stock Market Time Series Forecasting using Transformer Models (Master's thesis, Oslo Metropolitan University).
Nadeem, D.S. (2024). Financial Forecasting and Planning: Techniques for Enhancing Accuracy and Reliability. *Frontiers in Business and Finance*, [online] 1(2), pp.139–156. Available at: https://sprcopen.org/FBF/article/view/59.
Nadeem, S. (2024). Financial Forecasting and Planning: Techniques for Enhancing Accuracy and Reliability. Frontiers in Business and Finance, 1(2), 139-156.
Olcay, K., Tunca, S. G., and Özgür, M. A. (2024). Forecasting and performance analysis of energy production in solar power plants using long short-term memory (LSTM) and random forest models. IEEE Access.
Olubusola, O., Mhlongo, N. Z., Daraojimba, D. O., Ajayi-Nifise, A. O., and Falaiye, T. (2024). Machine learning in financial forecasting: A US review: Exploring the advancements, challenges, and implications of AI-driven predictions in financial markets. World Journal of Advanced Research and Reviews, 21(2), 1969-1984.

Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A. and Kwon, H.-H. (2021). Performance Comparison of an LSTM-based Deep Learning Model versus Conventional Machine Learning Algorithms for Streamflow Forecasting. *Water Resources Management*, 35(12), pp.4167–4187. doi:https://doi.org/10.1007/s11269-021-02937-w.
Sabry, F., Labda, W., Erbad, A., and Malluhi, Q. (2020). Cryptocurrencies and artificial intelligence: Challenges and opportunities. IEEe Access, 8, 175840-175858.
Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., and Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems, 7(1), 2400304.
Statista (2024) Bitcoin (BTC) price per day from August 9, 2022 to May 4, 2025 Available at: https://www.statista.com/statistics/326707/bitcoin-price-index/ (Accessed on 6 April 2025)
Wasserbacher, H. and Spindler, M. (2021). Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance*, [online] 4(1). doi:https://doi.org/10.1007/s42521-021-00046-2.
Yadav, N., None Divya, Rani, N., Singh, M. and Chowdhury, H. (2023). An Assessment of the Level of Adoption of AI/ML in Banking and Financial Institutions. *Advances in finance, accounting, and economics book series*, [online] pp.218–237. doi:https://doi.org/10.4018/979-8-3693-0082-4.ch013.
Yadulla, A. R., Maturi, M. H., Nadella, G. S., and Satish, S. (2024). Volatility Comparison of Dogecoin and Solana Using Historical Price Data Analysis for Enhanced Investment Strategies. Journal of Current Research in Blockchain, 1(2), 91-111.
Zaggia, M. (2022). Assessing the Risks to Tesla´ s Market Value: Towards the Share Price (Master's thesis, Universidade NOVA de Lisboa (Portugal)).