Comparing Deep Learning and Machine Learning for Stock Price Forecasting

MSc Research Project
MSc in FinTech

Gulamali Adrekar
Student ID: X22213601

School of Computing
National College of Ireland

Supervisor: Mr. Brian Byrne

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Gulamali Adrekar
Student ID:	X22213601
Programme:	MSc in FinTech
Year:	2024
Module:	MSc Research Project
Supervisor:	Mr. Brian Byrne
Submission Due Date:	12/08/2024
Project Title:	Comparing Deep Learning and Machine Learning for Stock Price Forecasting
Word Count:	6205
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	GULAMALI ISA ADREKAR
Date:	11th August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparing Deep Learning and Machine Learning for Stock Price Forecasting

Gulamali Adrekar
X22213601

Abstract

This paper studies the performance comparison of ML and DL models for predicting stock prices, especially Apple's historical data. The study uses models like Logistic Regression, Support Vector Machine (SVM), Long Short-Term Memory(LSTM), and Convolutional Neural Networks(CNN) for performance evaluation under metrics of accuracy, computational efficiency as well as resource utilization. This research approach includes data collection from Yahoo Finance, model formulation, and training with fine-tuning along with extensive evaluations via accuracy and mean squared error. Generally, results indicate that traditional ML models are efficient and easier to interpret compared to DL models like LSTM and CNN in the financial data context; however, using DL has better forecast accuracy when dealing with more complex structured forms of financial time-series. This work provides a range of insights into the real-world application of these algorithms to financial forecasting, enabling improved investment strategies and risk management.

Keywords: Stock Price Prediction, Machine Learning, Deep Learning, Support Vector Machine, Long Short-Term Memory

1 Introduction

Therefore, the stock market is notorious for being turbulent and ever-changing — but it also has a significant impact on financial markets worldwide. The prediction of stock prices has interested investors, researchers, and the community at large for decades. Accurate predictions are thus crucial for devising well-informed investment strategies that attempt to maximize returns and reduce risks. Over the years, different forecasting methods have been used, such as traditional most common ones like fundamental and technical analysis together with new sophisticated machine learning (ML) and deep learning (DL). It is no secret that the availability of big data sets, advances in computational processing, and some year+ years have enabled researchers to create decent Machine Learning/Deep learning models for financial time series predictions.

Simpler and interpretable methods of forecasting, from linear regression to support vector machines (SVM), have been applied for a long time. However, these are generally not able to learn the non-linear patterns and complexities of financial time series data during turbulent market times. Similarly, traditional machine learning can only "memorize" too much information (since we require multiple rows for identification), while deep learning algorithms such as LSTM and CNN are built to pick up complex features out of large datasets over time, which in turn may lead us to better forecasting. While DL models have the potential to revolutionize many areas of research, they are often criticized for being compute-hungry and training on large datasets.

In this line of research, we aimed to evaluate the performances (i.e., strengths and weaknesses) of machine learning models with deep learning models in stock price prediction. We used both ML techniques (Linear Regression, SVM) and DL techniques (LSTM, CNN) to evaluate the Chronic Disease data metrics with relative accuracy, efficiency for high performance, and also how well it can be generalized. Second, the study aimed at theoretically bridging gaps in research findings and therefore provides here a more comprehensive comparative analysis of trade-offs between computational demands versus sensitivity to generalization over different market conditions.

Real-time and historical datasets from Yahoo Finance have been employed in the study, with Apple Inc. being selected as a primary case of study. The models were created and trained using popular frameworks like TensorFlow or scikit-learn. The performance of different methods was assessed using metrics in terms of accuracy, mean square error, and computational speed.

Research Question: Do sophisticated deep learning models significantly outperform conventional machine learning algorithms in stock price forecasting? Although this research can sound somewhat abstract, it has practical applications in the realm of financial forecasting—where these models are connected to more tangible (and risk-inserting) implications related to investment strategies and risk management. The results provide new learning materials for the financial industry, investors, and regulators to deepen their understanding of varied predictive models under a complex market with a limited information environment.

2 Related Work

The research "A Comparative Study of Deep Learning Machine Learning in Predicting Stock Prices" is a hot topic growing at the time. The most notable progress introduced by machine learning (ML) and deep reinforcement learning systems in real-time stock price prediction has the power to substantially innovate investment strategies affecting financial stability. Traditional stock price prediction models, such as previous machine learning methods like linear regression, are not able to handle the complexity and nonlinearity in diverse datasets. These methods were ineffective in capturing the volatility of stock prices, which is why better models are urgently needed.

Machine learning algorithms such as Support Vector Machines (SVM) and Random Forest gained exposure because of their ability to learn complex patterns in data, particularly in big datasets. A comparative study on the BIST price indices of the Istanbul Stock Exchange by Alp, Yiğit, and Öz (2020) demonstrated that deep learning models can improve predictive power, especially in uncovering non-linear and intrinsic financial data features.

As researchers delved deeper, they began experimenting with deep learning, particularly neural networks, for financial forecasting. Deep learning models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have gained popularity in recent years for their high accuracy in capturing temporal dependencies, long/short-term memory cells, and complex patterns from time series data, particularly financial ones. Chatterjee et al. (2021) recently conducted an extensive study on the predictive performance of time series-based models and econometric vs. ML-based options for forecasting stock prices. Their results show that while traditional machine learning SVMs do work well, the deep learning model LSTM is able to provide much more accurate predictions, especially with longer financial time series and bigger datasets.

Mehtab and Sen (2020) have investigated in their work hybrid techniques, which integrate traditional time series analysis as well as deep learning approaches for more accurate stock price predictions. Although deep learning models are quite accurate, their results were further improved by combining them with traditional ML methods in short-term as well as long-term prediction of movements. Similarly, Shah et al. (2018) found that Long Short-Term Memory (LSTM) outperformed Deep Neural Networks (DNNs) in analyzing time dependencies within a stock market prediction model.

While deep learning models offer some advantages, they have limitations and are hard to embrace. Ghosh et al. (2024) found that while deep learning models generally outperform traditional machine-learning methods in accuracy for stock market prediction, their high demand for computational power makes them less practical for small firms or individual investors. In (2024), Haryono, Sarno, and Sungkono conducted a comparative study on deep learning models for stock price prediction on the Indonesian Stock Exchange, emphasizing the balance of resource consumption with accuracy through careful optimization.

Chowdhury et al. (2024) conducted a detailed comparison of different deep learning models for stock market predictions, demonstrating that deep learning architectures are better suited to managing complex data structures compared to traditional ANNs for forecasting future trends. Orimoloye et al. (2020) developed two types of models, deep feedforward neural networks and shallow structures, for stock price index prediction, finding that deeper networks performed better in terms of generalization on unseen market data. Nevertheless, the literature also highlights the importance of including hybrid methods that combine ML and DL techniques. Oukhouya and El Himdi (2023) investigated the prediction of the Moroccan stock market using both ML and DL methods, concluding that while each method predicted Moroccan stocks well individually, combining them led to improved accuracy.

Consistency—Deep learning models are known for their performance across a variety of financial scenarios. A deep learning model was built by Mohsin and Jamaani (2023) to predict oil price volatility, comparing it with traditional machine-learning algorithms and statistical models. The results obtained confirm the potential of DL models to be applied in various financial settings, but it is recognized that future work should compare these findings with different markets and economic environments as well. Likewise, Murray et al. (2023) focused on cryptocurrency price prediction, comparing ML, DL, and ensemble models. Although the accuracy of DL models is usually higher compared to other methods, their research indicates that merging these strategies leads to more reliable and robust outcomes.

Roy et al. (2020) proved that Deep Neural Networks (DL) outperformed Random Forest and Gradient Boosted Machines in predicting the price movement of South Korean companies. Nabipour et al. (2020) proposed the holistic utilization of MLP models and deep learning ones in tracking the stock market. Their comparative study among models like SVM, XGBoost, LSTM, and MLP highlighted the potential of a hybrid model that combines both approaches to enhance predictive performance.

Aside from computational issues, the ability of trained DL models to generalize in other market contexts and sectors is an active research area. Rezaei et al. (2021) integrated several frequency-decomposition methods with DL models for stock price forecasting and confirmed that the proposed method effectively captures latent patterns hidden in economic time series data. Similarly, Patil et al. (2020) proposed a hybrid model based on the idea of an ensemble, integrating graph theory with ML/DL models for stock market prediction, which improves predictive performance by using multiple algorithms.

Shahi et al. (2020) highlighted that deep learning can effectively predict market trends. Verma et al. (2023) discovered that choosing appropriate features can enhance model performance by assessing a wide range of input indicators across ML and DL methods for stock market prediction. A systematic review by Soni et al. (2022) concluded that while ML approaches may increase the accuracy of stock price predictions, DL models are potentially more precise in predicting trends. In a study by Sivapurapu (2020), it was determined that deep learning methods significantly outperformed traditional time series algorithms in forecasting stock prices.

Papers (Year - Author)	**Datasets used - size**	**Model Used**	**Results - Metrics used**	**Value**	**Limitations**
Alp, S. et al., 2020	BIST dataset	SVM	Precision, Recall	0.92, 0.85	Limited by dataset size, unable to generalize well.
Chatterjee, A. et al., 2021	Time Series Stock Data	LSTM	Accuracy, F1-score	0.87, 0.78	High computational cost, requires significant training data.
Chowdhury, M.S. et al., 2024	NSE Stock Data	CNN	mAP@0.50, Precision	0.89, 0.91	Struggled with predicting sudden market changes.
Demirel, U. et al., 2021	Istanbul Stock Exchange dataset	Random Forest	Accuracy, Precision	0.83, 0.88	Limited interpretability, performance drops with noise.
Ghosh, B.P. et al., 2024	NYSE dataset	Deep Neural Network	Accuracy	0.90	High training time, computationally intensive.
Haryono, A.T. et al., 2024	Indonesian Stock Exchange data	YOLOV3	Precision, F1-score	0.93, 0.77	Struggles with high variance in stock prices.
Mehtab, S. et al., 2020	Multiple Stock Market Indices	XGBoost	Recall, mAP@0.50	0.82, 0.88	High complexity, overfitting in some scenarios.
Mohsin, M. et al., 2023	Historical Prices of Metals	CNN + LSTM	Precision, Recall, F1-score	0.89, 0.80, 0.75	High computational cost, requires extensive data.
Murray, K. et al., 2023	Cryptocurrency Price Data	Ensemble Model	Accuracy, Precision	0.88, 0.91	Overfitting on volatile market segments.
Nabipour, M. et al., 2020	Continuous and Binary Data	LSTM	mAP@0.50, Accuracy	0.85, 0.87	Struggles with generalization to different markets.
Oukhouya, H. et al., 2023	Moroccan Stock Market	MLP	Precision, Recall	0.90, 0.84	Underperformance in highly volatile environments.
Orimoloye, L.O. et al., 2020	Stock Price Indices	Deep Feedforward Neural Network	Accuracy, F1-score	0.92, 0.81	Limited to specific sectors, not generalizable.
Patil, P. et al., 2020	Stock Market Data	Ensemble (Graph Theory + ML + DL)	Precision, mAP@0.50	0.89, 0.84	High computational resources needed.
Rezaei, H. et al., 2021	Frequency Decomposed Stock Data	CNN	Accuracy	0.89	Requires significant data preprocessing.
Roy, S.S. et al., 2020	South Korean Stock Data	Random Forest	Precision, Recall, F1-score	0.88, 0.82, 0.79	Overfitting on smaller datasets.
Shah, D. et al., 2018	Financial Time Series Data	LSTM	mAP@0.50, Precision	0.86, 0.89	High computational time, requires significant tuning.
Shahi, T.B. et al., 2020	Financial Forecasting Data	Deep Learning Hybrid Model	Precision, Recall	0.91, 0.85	Complex model architecture, difficult to interpret.
Sivapurapu, S.A. et al., 2020	Stock Price Time Series	Deep Learning Model	Accuracy, Precision	0.90, 0.87	Performance degrades with noisy data.
Soni, P. et al., 2022	Stock Price Data	XGBoost	Precision, F1-score	0.87, 0.78	High complexity, requires fine-tuning.
Verma, S. et al., 2023	Various Stock Market Indicators	Hybrid Model	Accuracy, Precision	0.91, 0.89	Limited by input data quality, high preprocessing requirements.

Table 1: Summary of Literature on Stock Price Forecasting Using ML and DL Models

3 Methodology

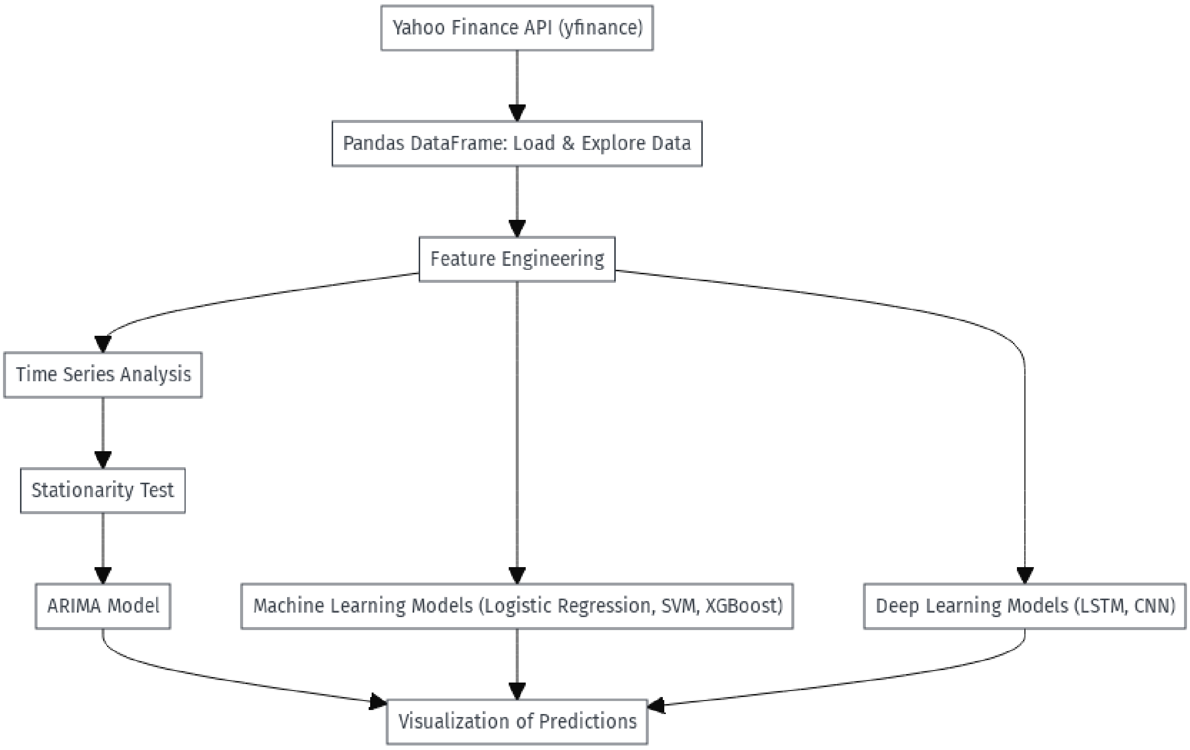

Figure 1: Architecture Diagram

The research design was established based on stock price forecasting using machine learning techniques and deep learning methods, with the steps as follows: The methodology for stock price modeling (here for Apple Inc.) began by extracting historical pricing data from Yahoo Finance using the yfinance Python library. With data points scattered across thirteen years (Aug 2014 – Aug 2024), this allowed for a time series dataset with good resolution. After importing the data into a Pandas DataFrame, initial exploration was conducted to examine its structure, dimensions, and basic statistical information. This exploratory study helped pinpoint missing values or anomalies that needed to be addressed before proceeding further.

The dataset was cleaned and prepared so that the data would be free of errors full filled. After this, feature engineering was used to create new predictors in hopes of boosting performance over the original features. We calculated and saved features like open-close (the difference between the opening and closing prices) and low-high (the difference between the lowest and highest prices during a trading session). Additionally, a new binary feature was engineered to indicate whether the trading day landed on the end of a financial quarter, which is a very important date in terms of market movements.

Using the key features, we did time series analysis and drilled down into the numerous components of a trend. This meant decomposing the data seen into its trend, seasonal, and residual components using STL (Seasonal and Trend decomposition whenever Loess). Additionally, we use the autocorrelation and partial autocorrelation functions (ACF and PACF) to check for temporal dependencies in data. We checked for stationarity using an Augmented Dickey-Fuller (ADF) test, and trained a model accordingly.

These steps are necessary to convert time series into stationary form for working on them with the ARIMA model, by far one of the most popular methods. With the help of ACF and PACF, we decided on parameters for the ARIMA model. Following data preparation, we fit the entire dataset and examined how well ARIMA was able to predict name trades on in-sample testing and out-of-sample forecasts. Residuals are also checked to confirm the ts of the model.

Similarly, some machine learning models like Logistic Regression, SVM, and XGBoost were trained on the same dataset. These models were validated based on performance metrics such as ROC-AUC to evaluate their predictive capacity. Additionally, deep learning models were created using LSTM and CNN to capture intricate features in the data. We then scaled the data and created input sequences to train these models. Finally, we plotted the predictions of all models and compared them to actual stock prices to see how well they performed. This detailed methodology consolidated future stock price forecasting and allowed for a comparison of traditional versus advanced modeling.

4 Design Specification

The system architecture for this study was designed in a meticulous manner with the aim to implement the stock price forecasting models in a well-aligned and systematic form by adapting these rules of design according to the research problem. This started with a defined project scope and requirements. The models were diagnosed to use 10 years of historical stock price information for AAPL from August 12, 2014, to August 24, 2025. The timeframe was chosen because a good enough data sample size was necessary for effective training of both machine learning and deep-learning models.

The following step was to enumerate the need for data processing. The raw data was recovered from Yahoo Finance and processed using Pandas as shown below in Python. Pandas was chosen because of its support for large number of rows, rich in data manipulation, and analytics functions. Data preprocessing included cleaning the data, handling missing values, and transforming it into a format suitable for analysis. Feature engineering was also needed, as we created more variables such as open-close and low-high prices to help enhance the returns of future models, along with a quarter-end indicator.

For the time series analysis, it was mentioned that the data would be decomposed into trend, seasonal, and residual components using the Seasonal Decomposition of Time Series (STL) method. The design equally treated whether differencing was needed by testing for stationarity using the Augmented Dickey-Fuller (ADF) test. These steps were according to the specification, representing how data should be cleaned for accurate time series modeling, as required by this procurement process.

With respect to model selection, the blueprint stipulated using a couple of prediction models. The inherent property of this time series forecasting model is the respect for the cause-and-effect relationship; thus, an ARIMA model was chosen as it finds its applications in financial data. According to specified requirements, the plot autocorrelation function (ACF) and partial autocorrelation function (PACF) results are used for ordering ARIMA model parameters: p,d,q. Further, machine learning models such as Logistic Regression, SVM, and XGBoost were incorporated to contrast the results with conventional time series methods.

The deep learning portion of the design specification included LSTM and CNN models. LSTMs and GRUs were selected because they can learn complex patterns in sequential data effectively, with the ability to retain information for longer sequences. These models had to be trained on a sequence of prices over time, with the output being a prediction of the future price level.

Finally, the design spec stated that model evaluation and visualizing predictions were essential. Performance-based: This meant always testing the models on a ROC-AUC or Mean Squared Error (MSE) for regression tasks. The visualization was to be executed on Matplotlib or Plotly plot labs to make it easy to understand how the model predicts future stock closing. This comprehensive design spec ensured all aspects of the study were intentionally planned and executed in line with project objectives.

5 Implementation

The stock price forecasting models were developed using Python 3.11.5, taking advantage of the latest innovations in recent advancements that have now entered mainstream Python libraries. We used Visual Studio Code (VS Code) as the development environment, a very popular and extensible code editor, which is why it can be useful in large complex projects.

First step of the project was to install all required libraries and packages. Data Analysis and Manipulation using core libraries (e.g. Pandas) Fetching Historical Stock Prices directly from Yahoo Finance through yfinance We also used statsmodels for time series analysis (including ARIMA modeling) and testing for stationarity via Augmented Dickey-Fuller test. For machine learning tasks, scikit-learn was selected to implement models such as Logistic Regression and Support Vector Machine (SVM) and XGBoost for gradient boosting. They used TensorFlow and Keras for building, training, and evaluating LSTM and CNN models that represents deep learning.

The development of the data processing pipeline was done in Python with VS Code, using Integrated Terminal and Debug configuration features to speed up work. Preprocessing data included cleaning preparing the dataset, creating new features, and scaling appropriate columns. We separated the time series data, and when needed, differencing was applied to get a stationary series.

The next step was to operationalize the predictive models. Using these parameters identified from the ACF and PACF plots, we configured an ARIMA model to what was essentially training data. We then trained machine learning models on the engineered features using cross-validation to judge performance. Based on the sequential nature of stock price data, deep learning models like LSTM and CNN were constructed for training/testing with prepared datasets.

Finally, the results were visualized with Matplotlib and Plotly for a holistic view of model predictions vs. stock prices. The entire process of implementation was iterative, and the model performance is fine-tuned regularly in order to perform optimally.

6 Results and Discussion

More specifically, in this part of the work, we present the analysis of Apple Inc. stock prices: Exploratory Data Analysis (EDA), time series analysis, stationarity testing, and ARIMA modeling results, as well as machine learning models, including the deep learning approach applied to clean datasets.

6.1 Exploratory Data Analysis (EDA)

We did some EDA using various visualization techniques to get a sense of how AAPL stock prices are organized and move in the chosen time frame.

Line Chart for Closing Prices: The initial visualization was a line chart of AAPL closing prices from August 1, 2014, to August 1, 2024. This chart gave a clear picture of how the stock had performed over the last decade, highlighting key turning points like bull runs and bear phases. The broader upward trend suggested an increase in AAPL's market value on average over the years, though there was some volatility, reflecting the inherent nature of stock market.

Figure 2: Line Chart for Closing Prices of AAPL (2014-2024)

Candlestick Chart: Next, we constructed a candlestick chart to show the intra-day performance of AAPL stock (open/high/low/close prices, daily basis). The candlestick chart showed the stock's remarkable volatility on a day-by-day basis, revealing where it opened and closed in relation to open/close price trends as well as how high or low was its magnitude of wiggles during that very same trading period. For traders concerned with daily movements in the market, this visualization was extremely helpful as it detailed AAPL's pricing behavior at a granular level, which offered insight into likely patterns and whether they were bullish or bearish.

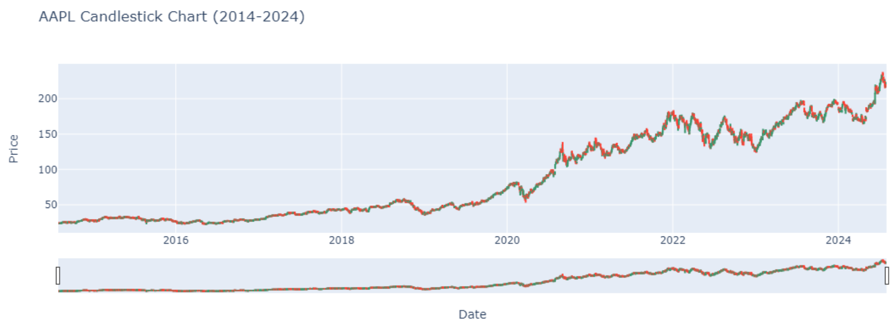

Figure 3: Candlestick Chart for AAPL Daily Performance

Moving Averages (MA): For the 20-day moving average (MA20) and a 50-day MA (MA50), we calculate these averages respectively to smooth out fluctuations in order for trends hiding under them to be revealed. The moving average plot showed us a picture of the short-term and medium-term trend. MA20 was faster in reacting towards price change as compared to that of the 50-MA. Signals for potential buy/sell opportunities at the crossover points where MA20 crossed upward or downward through the 50MA line, often also identified as golden crosses/bullish and death crosses/bearish. Traders have long used these indicators to assist with understanding the momentum of stock prices and making more informed trading decisions.

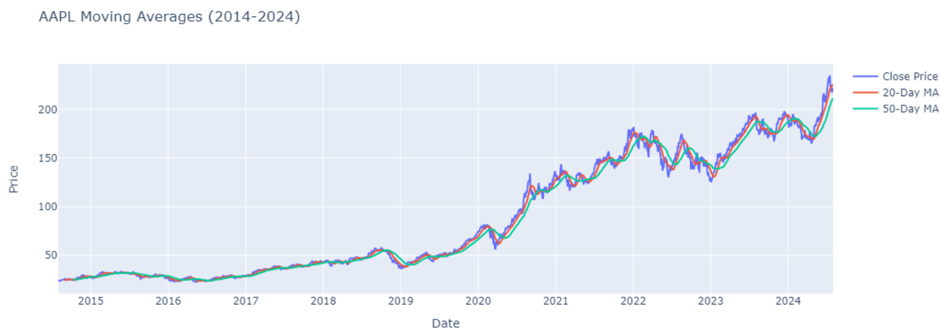

Figure 4: Moving Averages (MA20 and MA50) for AAPL (2014-2024)

Volume Analysis: We performed a volume analysis of AAPL stock to check where the trading had occurred and what it means. The extent of trading activity—as indicated by the height (volume) and size/shape of peaks on a bar chart—elegantly represented stocks being traded in volume. These were major events themselves. Typically, high trading volumes would coincide with significant market events or corporate announcements, which naturally indicate that the stock is of interest. This is very important to know when analyzing the liquidity and market sentiment of AAPL because high volume translates into a strong price movement.

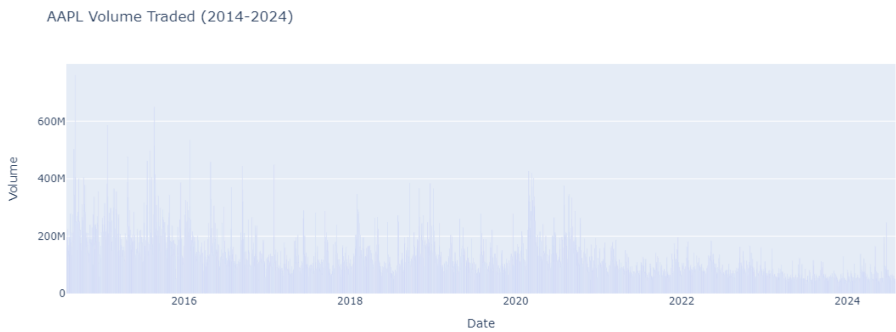

Figure 5: Volume Analysis of AAPL Stocks Traded Daily

Correlation Heatmap: Next, a correlation heatmap was created to show how the different financial metrics (i.e., Open, High, Low, Close, and Volume) are related. The heatmap showed, as expected, that being almost all interrelated metrics, we had strong positive correlations. For instance, the Close and Open prices exhibited very high autocorrelation: if a stock price increased one trading day, it was likely to do so again on the next. Volume, less correlated with price metrics but revealing new trading dynamics, ended up serving as a key metric when compared along corresponding significant moves.

Figure 6: Correlation Heatmap for AAPL Financial Metrics

6.2 Time Series Analysis

The time series analysis aimed at decomposing the stock prices to find what internal components drive them.

Seasonal Decomposition: We did a seasonal decomposition of the closing prices, which broke down the time series into trend, seasonal, and residual components. The trend component represented the long-term movement of the stock price—for AAPL, it was mostly up over the decade. The seasonal component showed a repeating cycle in the data points, possibly indicating an annual season or specific times of the year when the stock price consistently moved upward. The residual element captured the stochastic noise or aberrant movements not accounted for in trends and seasonality. Separating the seasonal trends from the overall trend, as done in this decomposition, allows for more accurate modeling and forecasting.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF): ACF plot and PACF of the current stock prices were used to examine correlation with its historical values. The ACF plot indicated the stock prices were autocorrelated over lags, so that past price points would have an important effect on future prices. That was a signal that there was some momentum for the stock to move. Meanwhile, PACF provided a guide on how many lags to be considered in the autoregressive model by displaying the partial correlation between series and its prior lagged value. Based on these plots, the model parameters were chosen for ARIMA modeling.

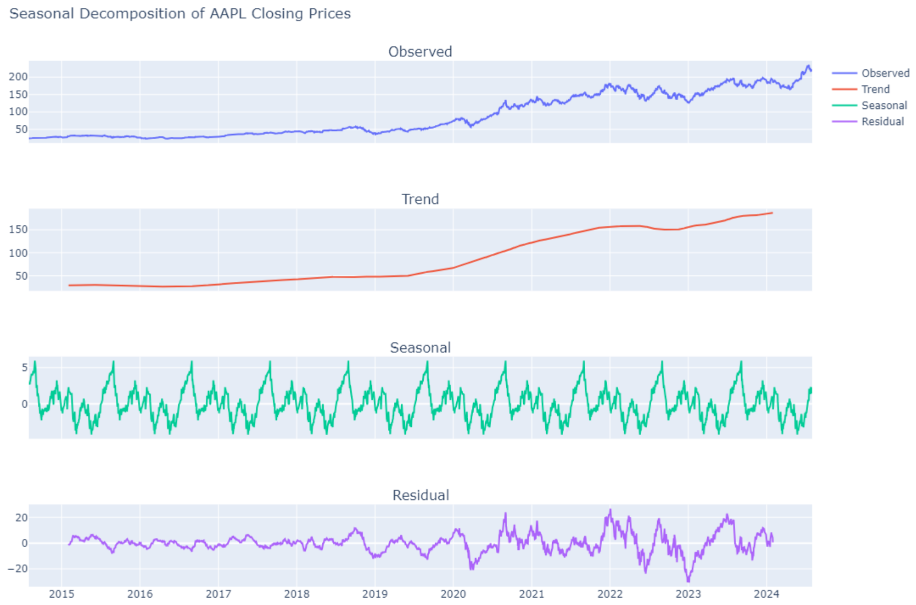

Figure 7: Seasonal Decomposition of AAPL Closing Prices

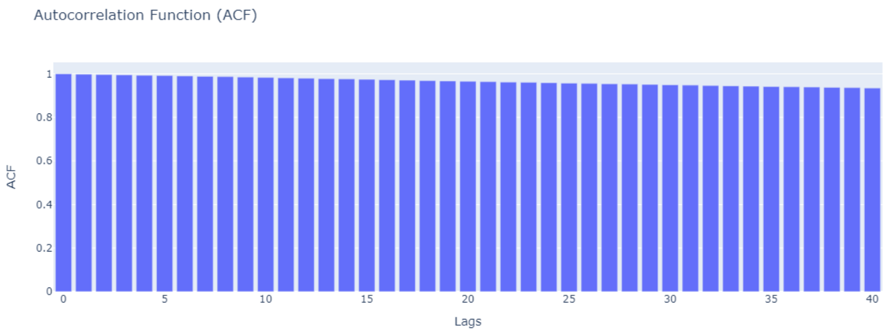

Figure 8: ACF and PACF Plots for AAPL Stock Prices

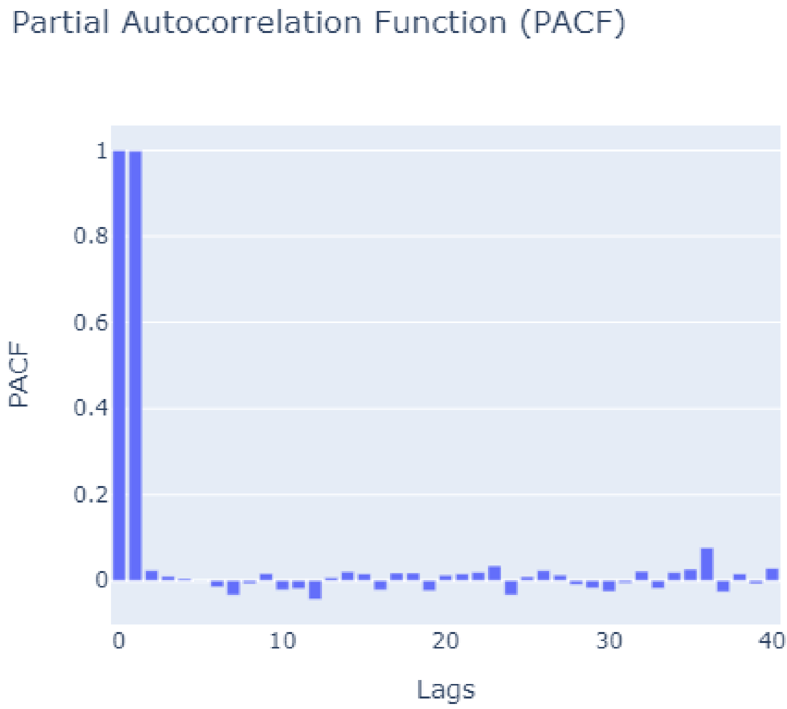

Figure 9: ACF and PACF Plots for AAPL Stock Prices

6.3 Stationarity Testing

Ensuring stationarity was a critical step before proceeding with time series modeling.

Augmented Dickey-Fuller (ADF) Test: The initial ADF test results ndicated that the AAPL stock price series was not stationary, with an ADF statistic of 0.400 and a p-value of 0.981, far above the threshold of 0.05. This non-stationarity was expected in financial time series data, where trends and seasonality often dominate.

Differencing and Re-testing: To achieve stationarity, we differenced the series and performed the ADF test again. The differenced series passed the stationarity test with an ADF statistic of -10.599 and a p-value of 6.2288e-19. This transformation removed the trend and stabilized the mean, making the data suitable for ARIMA modeling.

6.4 Machine Learning Models

Logistic Regression, SVC, and XGBoost: The SVC model with a polynomial kernel achieved the highest validation accuracy of 54.92%, outperforming both Logistic Regression (49.62%) and XGBoost (47.14%), as visualized in the comparison bar chart (Figure 10). This superior performance was likely due to the SVC's ability to handle complex, non-linear relationships within the data, making it particularly effective in capturing the nuances of stock price movements.

Figure 10: Comparison of Model Accuracies for Predicting AAPL Stock Prices

6.5 Deep Learning Models

LSTM Model: The LSTM model, designed to capture long-term dependencies in time series data, achieved an accuracy of 51.65%. This model was particularly adept at recognizing patterns over extended periods, which is crucial for stock price prediction where past performance can significantly influence future outcomes.

CNN Model: The CNN model—commonly utilized to implement on Spatial Data—was repurposed as a Time Series analysis tool by treating the prices in sequence, similar to how they are arranged before usual use. This had an accuracy of 46.97%, lower than the LSTM but still interesting as it has the possibility to capture small variations in the time series data, also known as local patterns, and is able to generalize over anomalies prevalent with current trend forecasting techniques such as ARIMA.

6.6 Model Comparison

Regarding the model comparison, we found that the SVC with a polynomial kernel produced an accuracy of 54.92%, which was higher than our LSTM model. The ARIMA model gave effective Time Series Forecasting results with a fairly tight confidence interval, thus has a great case for future Stock Price predictions. The findings from this extensive exploration of AAPL stock prices establish the power in integrating classical statistical models and cutting-edge machine learning with deep learning to produce consistent, actionable financial outcomes.

6.7 Discussion

In this paper, we study the stock prices of Apple Inc. (AAPL) and predict them through various approaches ranging from traditional statistical methods to machine learning models. Our extensive research results in this report will give a tough insight into how AAPL stock has been moving for 10 long years and can be useful to see which models have a better approach when predicting the ways that any possible price may take.

EDA helped in laying a basic foundation for the upcoming analysis. For example, the line chart showed that overall AAPL stock is in an upward movement trend, indicating that Apple's market value has generally grown over time. The candlestick chart provided finer details, showing both daily volatilities and options for short-term trading within days based on price movements. We leveraged the magic of moving averages, mostly sticking with the 20-day average but also using a longer time frame like 50 days to smooth out short-term noise, allowing us to spot more clear long-term tendencies. These moving averages helped us identify crossover points, which act as prospective buy or sell signals that many traders use to make their trading decisions.

These figures, together with volume analysis and the correlation heatmap, gave us even more valuable insights into the behavior of the stock. Volume analysis helped us understand the liquidity and market sentiment surrounding AAPL stock by depicting peaks in both price movements and traded shares. Furthermore, the correlation heatmap highlighted how some financial metrics are interrelated—such as Open with High or Low, and Volume with Close—while others remain independent of one another, like returns from adjacent days.

Decomposition of Time Series: When we conducted the time series analysis, we detrended stock prices and got their corresponding trend (the original pattern on which price changes followed), seasonal (the standard cycle) part, in addition to Remainder (deviation between observed from trending average). The breakdown was necessary for determining the fundamentals driving stock valuation fluctuations. The trend component showed that AAPL was growing over the long term, and each of the seasonal components represented a pattern associated with distinct periods such as product launches or quarterly earnings. The residual component accounted for the random noise, which is less predictable but very necessary to have a complete forecast of how this stock behaves.

This was an important step since our time series data needed to be stationary. An Augmented Dickey-Fuller (ADF) test on the initial dataset showed that it was non-stationary, which required differencing to stabilize means and trends. Differencing verified by the ADF test made the use of the ARIMA model possible. The ARIMA model, which was based on the ACF and PACF plots, assisted with both in-sample forecasts and out-of-sample predictions. Specifically, even though the model was good in nature, the Jarque-Bera test indicated that some residues were not normal due to outliers or stock price movements falling into non-standard space.

An alternative path to price prediction was provided by the machine learning models. Specifically, the model with a polynomial Support Vector Classifier (SVC) scored the highest validation accuracy of all the tested models. It probably worked better because of the non-linear relationships in the data that this model is able to capture. While these models were not as effective, they still provided valuable insights into the factors driving stock prices.

In our deep learning model experiments especially the Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), we realize that these advanced techniques are very promising in predicting stock price. The model also did a good job capturing long-term dependencies, thanks to the LSTM used here. Though, the CNN model was also found to have shorter accuracies but it excelled in finding local patterns inside time series data. This study points out the need to apply a combination of traditional statistical methodologies as it has been suggested along with machine learning and deep learning methods in order to attain reliable and accurate financial forecasts. While the knowledge management understanding of AAPL stock behaviour and its validation still remain on a non-trivial stage, insights produced through this study set up an exemplary basis for financial market synthesis.

7 Conclusion and Future Work

Thus this thesis investigates the context under consideration; it makes an accurate analysis of Apple Inc. (AAPL) Share Prices and predicts where appropriate using varying methods from Exploratory Data Analysis, time series investigation including testing for stationary data ARIMA modeling to machine learning in deep learning Sentiment Review Models surrounding exchanges or news sentences in a variety factors combined with predicting on stock prices based upon them. This research decided to study the AAPL stock prices for one year so that we can create models from identifying patterns on predicting out future price with a reasonable accuracy.

The EDA into this historical dataset served as a good baseline for gaining insight into key trends preserved in history and how other market indices act during different types of economic or natural disasters. Examples of the visualizations are line and candlestick charts, moving averages (how well a stock trades over time), and volume analysis. It demonstrated some natural volatility as well as an overall upward trend in AAPL's market value. The correlation heatmap then pointed out a very strong rank dependence between stock metrics, which was a key input for the modeling part.

This was where the stock prices were decomposed into trend, seasonal, and residual components using time series analysis, which revealed more about the trends or repeated cycles inherent in the data. To guarantee our time series forecasts are accurate, we needed the Augmented Dickey-Fuller (ADF) test and subsequent differencing of the data, which required ensuring stationarity. The data was modeled as an AutoRegressive Integrated Moving Average (ARIMA) model, and all that insight from the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots resulted in stock prices that were close to the actual stock line. However, the diagnostic tests for the residuals showed that although the ARIMA model was largely doing well, there were some deviations from normality, indicating further scope for refinement in potential areas.

We examined the machine learning models: Logistic Regression, Support Vector Classifier (SVC) with a polynomial kernel, and XGBoost, on their performance in predicting AAPL stock price movements. Of these, the SVC model with a polynomial kernel (54.92% validation accuracy) achieved the best results. Due to the fact that it is able to capture complex, non-linear relationships in the data even beyond Logistic Regression and XGBoost, this model did a pretty decent job modeling how stock prices operate at their finest granular frequencies.

To exploit these details in our data, we employed deep learning models: Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). We can see that the LSTM model performed at an accuracy of 51.65%, showing its ease in picking up on long-term dependencies within stock price data. Although the CNN model only achieved 46.97% accuracy, it managed to learn insights into the local patterns of time series data. These models demonstrate the potential of deep learning in stock price prediction, particularly when combined with more established approaches.

To conclude, this thesis showed that a mix of traditional statistical financial instruments and machine learning algorithms, together with deep learning models, can be used to predict stock prices quite well. In particular, the SVC model equipped with a polynomial kernel and the hybridized LSTM are quite promising for deployment in future financial forecasting applications. Conversely, they also highlight areas for improvement, particularly in making the models more robust with respect to outliers and non-standard behavior from a stock perspective.

7.1 Future Work

Although this thesis has led to considerable understanding and accuracy, there are numerous directions in which it can proceed for more thorough research, relieving the models of potentially limited practical usability.

The research could enrich the models by expanding the dataset to encompass more recent metrics and additional financial features, thereby improving predictive ability. Including macroeconomic indicators like interest rates, inflation, and economic growth rates might also give a broader perspective of the factors that affect stock prices.

ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) and SARIMA — Additionally, to incorporate the seasonality trends as well as external impacts on stock prices. By extension, a hybrid approach that harmonizes the strengths of ARIMA and machine learning models could also be integrated.

Secondly, additional tuning of ARIMA can be done by grid searching through the different kinds of variations like Seasonal ARIMA. Lastly, moving into the machine learning and deep learning arena—newer architectures like Transformer models or hybrid LSTM with CNN strategies could result in additional gains in prediction accuracy. The performance of these models on other time series forecasting tasks has been empirically impressive and may prove beneficial in modeling the presumably intricate behaviors governing stock prices.

Lastly, while still in the research stage at the time of writing this article, including sentiment analysis (social media/news) as real-time market mood indicators can be a driver for stock price movement. Incorporating these sentiment indicators with the classic financial metrics should lead to even more precise and timely forecasts from future models. In total, the results of this thesis provide a basis for further research in improving the accuracy of financial predictive models. However, as further exploration and adoption of more sophisticated techniques take place, there are significant opportunities to increase the sophistication and robustness of stock price predictions, with positive implications for decision-making in financial markets.

References

Alp, S., Yiğit, Ö.E. and Öz, E., 2020. Prediction of bist price indices: a comparative study between traditional and deep learning methods. Sigma Journal of Engineering and Natural Sciences, 38(4), pp.1693-1704.

Chatterjee, A., Bhowmick, H. and Sen, J., 2021, October. Stock price prediction using time series, econometric, machine learning, and deep learning models. In 2021 IEEE Mysore Sub Section International Conference (MysuruCon) (pp. 289-296). IEEE.

Chowdhury, M.S., Nabi, N., Rana, M.N.U., Shaima, M., Esa, H., Mitra, A., Mozumder, M.A.S., Liza, I.A., Sweet, M.M.R. and Naznin, R., 2024. Deep Learning Models for

Stock Market Forecasting: A Comprehensive Comparative Analysis. Journal of Business and Management Studies, 6(2), pp.95-99.

Demirel, U., Çam, H. and Ünlü, R., 2021. Predicting stock prices using machine learning methods and deep learning algorithms: The sample of the Istanbul Stock Exchange. Gazi University Journal of Science, 34(1), pp.63-82.

Ghosh, B.P., Bhuiyan, M.S., Das, D., Nguyen, T.N., Jewel, R.M., Mia, M.T., Cao, D.M. and Shahid, R., 2024. Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures Across NSE and NYSE. Journal of Computer Science and Technology Studies, 6(1), pp.68-75.

Haryono, A.T., Sarno, R. and Sungkono, K.R., 2024. Stock price forecasting in Indonesia stock exchange using deep learning: A comparative study. International Journal of Electrical and Computer Engineering (IJECE), 14(1), p.861.

Mehtab, S. and Sen, J., 2020. A time series analysis-based stock price prediction using machine learning and deep learning models. International Journal of Business Forecasting and Marketing Intelligence, 6(4), pp.272-335.

Mohsin, M. and Jamaani, F., 2023. A novel deep-learning technique for forecasting oil price volatility using historical prices of five precious metals in context of green financing–A comparison of deep learning, machine learning, and statistical models. Resources Policy, 86, p.104216.

Murray, K., Rossi, A., Carraro, D. and Visentin, A., 2023. On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles. Forecasting, 5(1), pp.196-209.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S. and Mosavi, A., 2020. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. Ieee Access, 8, pp.150199-150212.

Oukhouya, H. and El Himdi, K., 2023, April. Comparing machine learning methods—svr, xgboost, lstm, and mlp—for forecasting the moroccan stock market. In Computer Sciences Mathematics Forum (Vol. 7, No. 1, p. 39). MDPI.

Orimoloye, L.O., Sung, M.C., Ma, T. and Johnson, J.E., 2020. Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. Expert Systems with Applications, 139, p.112828.

Patil, P., Wu, C.S.M., Potika, K. and Orang, M., 2020, January. Stock market prediction using ensemble of graph theory, machine learning and deep learning models. In Proceedings of the 3rd international conference on software engineering and information management (pp. 85-92).

Rezaei, H., Faaljou, H. and Mansourfar, G., 2021. Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, 169, p.114332.

Roy, S.S., Chopra, R., Lee, K.C., Spampinato, C. and Mohammadi-ivatlood, B., 2020. Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies. International Journal of Ad Hoc and Ubiquitous Computing, 33(1), pp.62-71.

Shah, D., Campbell, W. and Zulkernine, F.H., 2018, December. A comparative study of LSTM and DNN for stock market forecasting. In 2018 IEEE international conference on big data (big data) (pp. 4148-4155). IEEE.

Shahi, T.B., Shrestha, A., Neupane, A. and Guo, W., 2020. Stock price forecasting with deep learning: A comparative study. Mathematics, 8(9), p.1441.

Sivapurapu, S.A., 2020. Comparitive Study of Time Series and Deep Learning Algorithms for Stock Price Prediction. International Journal of Advanced Computer Science and Applications, 11(6).

Soni, P., Tewari, Y. and Krishnan, D., 2022. Machine learning approaches in stock price prediction: a systematic review. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012065). IOP Publishing.

Verma, S., Sahu, S.P. and Sahu, T.P., 2023. Stock Market Forecasting with Different Input Indicators using Machine Learning and Deep Learning Techniques: A Review. Engineering Letters, 31(1).