IMPACT OF MACROECONOMIC FACTORS ON STOCK MARKET AND STOCK PRICE PREDICTION

Executive summary

A stock market is a type of marketplace that facilitates the buying and selling of publicly traded company stocks and other types of securities. It is a significant source of capital for businesses, and it presents an opportunity for individuals as well as institutions to make investments. In determining the performance of the stock market and stock prices, macroeconomic factors are essential. GDP, inflation, interest rates, employment, and political stability are a few of these elements. Investors can make better decisions about when to buy or sell stocks by having a solid understanding of how these macroeconomic factors may affect stock prices. There are different types of stock market analysis:

1. Technical analysis: This approach looks for patterns and trends in historical data and charts that might predict future market movements.

2. Fundamental analysis: In this technique, financial and economic information about a company, such as revenue and earnings, is examined to make predictions about the company's future performance.

3. Quantitative analysis (Time Series Analysis): This technique examines a lot of data and generates market forecasts using mathematical models and algorithms.

In this project, we are focusing on the Technical and Quantitative Analysis of Apple Inc. The dataset is taken from Yahoo Finance from 2010 January to 2022 November. Macroeconomic factors like GDP, Consumer Index Price (CPI) and the Unemployment rate were selected to find out which factor has the most influence on Apple Stock price.

This project gives an overview of the difficulty to predict the stock price accurately as it is influenced by a wide variety of factors, including economic, political, social, and psychological factors. The future is inherently uncertain, and there are many unknowns that can impact the stock market, such as natural disasters, political events, and technological breakthroughs. These factors can interact in unpredictable ways, making it difficult to predict market movements.

This project main purpose to focus on the following points:

- How do Apple's historical stock prices and financial results compare to those of other significant tech firms?
- How can technical indicators be used to forecast stock price movements for short-term trading? Which technical indicators have been found to be the most accurate at doing so?
- How do macroeconomic factors like GDP growth, the unemployment rate, and the consumer price index affect the stock prices of companies that are publicly traded?
- How can a model be created that can show how a macroeconomic variable affects the stock market and which macroeconomic factor has the biggest impact on the price of Apple stock?
- Which model is more effective at forecasting stock price information?

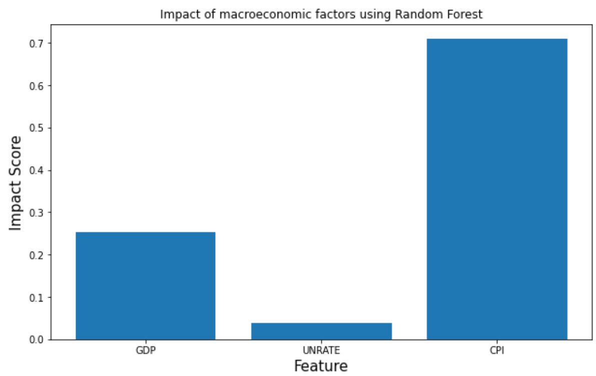
After fixing the research questions a lot of research papers were analysed. From the literature review, I gained insights to my problem statement:

- Developing a deeper comprehension of the impact of various macroeconomic factors on various stock markets
- Importance of technical indicators and which all technical indicators such as MACD, Bollinger Bands are effective at forecasting changes in stock prices for short-term trading.

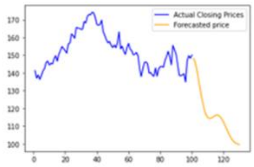
- Using macroeconomic variables to predict stock prices and effectively identify the most important factors influencing the stock market using Random Forest
- How LSTM is so good at predicting and forecasting stock prices using historical data.

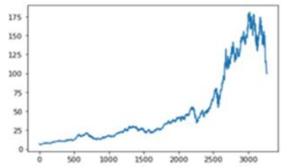
In 2016, Tesla had the second-highest stock price, behind Google and Apple, with Microsoft coming in top. By 2022, Microsoft was at the top. The scenario was similar for Tesla, Microsoft, and Google as it was for Apple during the epidemic. Tesla's stock price reached the highest level, then underwent a significant change and sided with Microsoft. Google's stock price had only slightly increased in the previous five years and experienced a big decline. Apple was not severely damaged by the pandemic, instead only needed four months to recover its losses. Apple has a solid base and great tactics for quickly recouping losses. Overall, the stock of Apple fell during the worst of the epidemic but mostly maintained a constant value, with a trend of continual rising. Microsoft’s stock is higher than Apple's, but Apple has a higher number of transactions. Tesla initially seems to be the opposite of Apple, focusing on revenue growth and not profits, which led to financial struggles in the past.

GDP, the industrial production index, the unemployment rate, Consumer Price Index, and stock return are all statistically related and correlation between them is found out by using Pearson correlation. The unemployment rate has a moderately negative correlation with Apple's stock price, whereas the country's Consumer Price Index has a strong positive correlation with the stock price. The nation's gross domestic product has a positive relationship with the country's stock price. Stock price movement was analysed using technical indicators such as Bollinger Bands, EMA (Exponential Moving Average) and analysed a trading strategy.



Using Random Forest, the CPI (Consumer Price Index) has the greatest influence on stock prices, with an influence of nearly 70%, followed by GDP, which has an influence of about 25%. The unemployment rate has the least effect on the value of Apple's stock. The findings indicate that the Consumer Price Index and the Gross Domestic Product are both reliable indicators of changes in Apple's stock price. Stock price prediction of Apple using Macroeconomic variables was predicted using Random Forest Regressor with a model score of 0.988 and a RMSE value of 8.164.





We first preprocessed the data by dividing it into training and testing sets and normalizing the prices. To enhance the model's performance, we also experimented with various architectural designs The Apple stock price was predicted using LSTMRNN for the next 30 days with a RMSE score of 23.30 for training set and 109.47 for testing set. The model shows that stock price movement is going down.

CONTENTS

1.INTRODUCTION

Stock markets are financial markets where investors can buy and sell shares of publicly traded companies. It is also referred to as an equity market or the securities market. Companies sell ownership stakes to investors to raise capital on the stock market. Companies can access the money they need to run and grow their businesses without taking on debt by listing shares for sale on the stock exchanges that make up the stock market. Companies are required to disclose information and give shareholders a say in how their businesses are run in return for the privilege of selling stock to the public. By exchanging their funds for shares on the stock market, investors gain profit because of companies use that money to invest in developing and expanding their businesses as the value of their stock increases over time, resulting in capital gains. As their profits increase, businesses also distribute dividends to their shareholders. Investors and researchers have long been interested in learning how to understand the shifting regularity of the stock market and forecast the trend of stock prices. Politics, economic factors, society, and the market all have an impact on how stock prices rise and fall. The stock market's trend forecast is directly related to the acquisition of profits for stock investors. The forecast's ability to effectively avoid risks increases with its accuracy. For publicly traded companies, the stock price not only reflects current operating conditions and expectations for future development but also serves as a crucial technical index for company analysis and research.

Stock price prediction is a difficult and complex task for businesses, investors, and equity traders. Stock markets are inherently chaotic, non-parametric, non-linear, noisy systems (Ahangar, Yahyazadehfar, & Pournaghshband, 2010). It makes it difficult to predict the price accurately and precisely in the future. Another challenging task in stock prediction is feature selection from the financial data, for which numerous methods have been proposed (Hoseinzade & Haratizadeh, 2019). Some researchers tend to only use technical indicators, while others prefer to use historical data (Bhandari et al., 2022). Due to the use of few features, the predictive model's performance might not be at its best. On the other hand, the model might be complex and challenging to understand if all the available features from the financial market are included. In addition, collinearity between several variables may cause the model performance to suffer. Making stock price predictions can be done using a variety of techniques, from straightforward technical analysis to sophisticated machine learning algorithms. By examining charts and statistical indicators of prior price movements, technical analysis is a technique for predicting future price movements. Its foundation is the notion that market trends, as depicted on charts and other technical indicators, can forecast future activity. Machine learning algorithms, on the other hand, are used to analyse vast amounts of data and make forecasts based on patterns and trends in that data. These algorithms can be used to forecast future stock prices after being trained on historical data.

Things were getting more interesting in the eighties because of the development of innovative analysis tools and techniques. For instance, the spreadsheet was invented to model financial performance, automated data collection became a reality, and improvements in computing power helped predictive models to analyse the data quickly and efficiently. Because of the availability of large-scale data, advancement in technology, and inherent problem associated with the classical time series models, researchers started to build models by unlocking the power of artificial neural networks and deep learning techniques in sequential data modelling and forecasting. These methods are capable of learning complex and non-linear relationships compared to traditional methods. They are more efficient in extracting the most important information from the given input variables.

Macroeconomic factors have a high influence on the stock market. Macroeconomic factors such as GDP, inflation rate, Consumer price index, and the Unemployment rate can affect the stock price of a company. The value of a company listed on a given country's stock market is thought to be based on the overall economic status quo and near-term prospects of that nation, just as the market value of a company is thought to depend on that company's current economic status and near-term prospects (Peiró, 2015).

Many deep learning architectures have been developed to address various problems and the inherent structure of datasets. In a basic feed-forward architecture information only flows in a forward direction. Since each input is dealt with separately, information from the previous step is not retained. Because a series of earlier events are required to predict future events when dealing with sequential data, these models are ineffective. Recurrent neural networks (RNNs) are designed to perform such tasks. The RNN architecture uses loops to preserve important information over time. The network internally exchanges data from one timestep to the next.

As a result, the RNN is better suited for time series applications like stock market forecasting, language translation, message/email auto-completion, and signal processing as well as sequential data modelling. The cost of error between the predicted values and the actual values from a labelled training dataset is calculated during the RNN's training process. Updating the weights and biases of the networks repeatedly until the lowest value is attained minimizes the error. A gradient, or the rate at which cost varies in relation to each parameter, is used in the training process. By iteratively adjusting the parameters, the gradient gives the error surface a direction to move in.

In this extended report, we will analyse the stock price of a company and draw upon theory and existing empirical studies to choose macroeconomic variables that affect that company's stock price and predict the stock price using different models.

2.CONTEXT AND BACKGROUND

Stock market analysis is the process of evaluating and interpreting financial data in order to make investment decisions. The goal of stock market analysis is to identify profitable investment opportunities by analysing the performance of individual stocks and the broader stock market. There are three main categories of stock market analysis and prediction as Fundamental Analysis, Technical Analysis, and Time Series Analysis (Quantitative analysis). Fundamental Analysis is used for Long Term investment. It's basically a research-based approach that looks at a company's financial performance, such as its revenue, earnings, and assets, to determine the company's intrinsic value and potential for growth. Fundamental analysis is mainly used for long-term investment. Technical analysis, on the other hand, is used for Short-term trading, Swing Trading, and Positional trading. Technical analysis is a method of evaluating securities by statistics generated by market activity, such as past prices and volume. Technical analysts do not attempt to measure a security's intrinsic value, but instead, use charts and other tools to identify patterns that can suggest future activity. Technical analysis is often used to identify trends and support and resistance levels, which can be useful in making trading decisions. It is important to note that technical analysis is not an exact science, and the results of technical analysis can vary widely. It can be used as conjunction with fundamental analysis, which involves analysing the inherent value of the security. Time series analysis is a statistical method used to analyse data points collected over a period of time. It is commonly used in the field of finance to analyse stock prices, but it can be applied to a wide variety of data sets, including economic indicators and meteorological data. Time series analysis can involve techniques such as smoothing, decomposition, and forecasting. Smoothing techniques are used to reduce the noise in the data and identify trends, while decomposition techniques can be used to isolate individual components of the time series data, such as trend, seasonality, and residuals. Forecasting techniques can be used to make predictions about future values of the time series based on past data using machine learning and deep learning models such as Random Forecast and LSTM.

In this project, we are focusing on Technical Analysis and Quantitative analysis of Apple stock price. Also, we are analyzing the impact of macroeconomic factors on the Apple Stock price. Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services. The company's hardware products include the iPhone smartphone, the iPad tablet computer, the Mac personal computer, the iPod portable media player, the Apple Watch smartwatch, the Apple TV digital media player, and the HomePod smart speaker. Apple's software includes the macOS and iOS operating systems, the iTunes media player, the Safari web browser, and the iLife and iWork creativity and productivity suites. Its online services include the iTunes Store, the iOS App Store, and Mac App Store, Apple Music, and iCloud. Apple Inc. has been publicly traded on the NASDAQ stock exchange since December 12, 1980. The company's initial public offering (IPO) was at $22 per share and the company sold 4.6 million shares to the public.

The macroeconomic variables that are we are used to find the impact on Apple stock price are:

GROSS DOMESTIC PRODUCT (GDP)

The standard measurement of the value added produced through the production of goods and services in a nation over a specific time is the Gross domestic product (GDP). GDP is the most significant indicator for measuring economic activity. Consequently, it also accounts for the revenue generated by that production, or the total amount spent on finished goods and services (fewer imports), and includes some nonmarket production, such as defense or education services provided by the government. When the GDP rises, the economy is expanding because more people are spending money, businesses may be growing, and prices for goods and services, salaries, and profits are rising.

UNEMPLOYMENT RATE (UNRATE)

The unemployment rate is calculated by taking the total number of workers in an economy and dividing that number by the total number of workers who are currently without jobs but are looking for work. Those individuals who have not actively sought employment in the preceding four weeks are excluded from this measure. It is essential to keep in mind that the rate is a measurement of the percentage of unemployed job seekers in the labour force, which is the sum of employed and unemployed persons. It does not measure the unemployment rate of the entire population.

CONSUMER PRICE INDEX (CPI)

The Consumer Price Index is a measurement of the overall change in prices paid by consumers over time based on a sample collection of products and services that is representative of those prices. The Consumer Price Index (CPI) is the inflation measure that is utilised the most frequently, and it is closely followed by policymakers, financial markets, businesses, and consumers. A related index that covers wage earners and clerical workers is used for cost-of-living adjustments to federal benefits, while the widely quoted CPI is derived from an index that encompasses 93 percent of the population of the United States. The change in housing rents is used to estimate the change in the cost of shelter overall, which includes the costs of owner-occupied housing, which account for almost a third of the CPI.

Then by using Machine learning models and deep learning models like Random Forest, and LSTM we predict the stock price of Apple.

3.RESEARCH QUESTIONS

Throughout this project, the focus will be mainly on answering the following questions.

- How do the historical stock prices and financial performance of Apple compared to those of other major tech companies?

- What technical indicators have been found to be most effective in predicting future stock market movements and how to use them to predict stock price movements for short-term trading?
- What is the relationship between macroeconomic variables, such as GDP growth, Unemployment rate, and consumer price index affecting stock prices of publicly traded companies?
- How to build a model capable of pointing out the impact of the macroeconomic variable on the stock market and which macroeconomic factor has the major impact on Apple stock price?
- Which model is better at predicting stock price data?

4. LITERATURE REVIEW

This section gives an overview of machine learning and its applications in the financial world and a brief understanding of the effect of macroeconomic variables on the stock market and stock market price movement.

There has been a constant debate on the predictability of stock returns due to its immensely complex, chaotic and dynamic environment. There are different how to analyse the stock market in order to get better predictions results. The Efficient-Market hypothesis, which states that an asset's current price always reflects all prior information instantly available for it, was first presented in (Malkiel & Fama, 1970).There is also the Random-walk hypothesis , which asserts that a stock's price changes independently of its past; in other words, the price of a stock will only change based on information that will become available tomorrow, regardless of the price of the stock today. These two hypotheses show that there are no reliable methods for forecasting stock prices. A random strategy can also outperform some of the most well-known technical trading strategies, such as Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI), according to a series of experiments conducted by (Biondo et al., 2013). Technical analysis is a method that is frequently used to model and forecast the stock market. It is based on historical data from the market, primarily price and volume. Following are some presumptions: Prices change according to trends, supply and demand changes can be seen on charts, supply and demand changes can cause trends to reverse, prices change based on trends, and patterns on charts tend to repeat (Parchaliuk et al., 1999). In other words, technical analysis does not consider any external factors, such as macroeconomic, social, or political ones.

GPD, the industrial production index, the unemployment rate, long-term interest rates, and stock return are all statistically related, according to the study by (Jareo and Negrut,2016) conducted in the United States using Pearson correlation. Additionally, in Kenya, Uganda, and Tanzania, stock returns and MVs were found to be significantly correlated (Laichena and Obwogi, 2015). Using the Ordinary Least Squares technique, (Asekome and Agbonkhese,2015) found a significant relationship between Nigeria's GDP and money supply stock returns and a negligible relationship between the exchange rate, capacity utilisation, and stock returns.

The random forest classifier (Breiman, 2001) is a cost-effective feature selection method that has been successfully applied to high-dimensional data, including that from microarrays (Jiang et al., 2004), time series, and even spectra. The random forest learner is an ensemble of learners that uses random decision trees to provide multiple measures of feature significance. Two metrics are presented here; one is inspired by statistical permutation tests, while the other is derived from random forest classifier training. There is a fair correlation between the two metrics (Strobl et al., 2008). The random forest classifier (Breiman, 2001) is a cost-effective feature selection method that has been successfully applied to high-dimensional data, including that from microarrays (Jiang et al., 2004), time series, and even spectra. The random forest learner is an ensemble of learners that uses random decision trees to provide multiple measures of feature significance. Two metrics are presented here; one is inspired by statistical permutation tests, while the other

is derived from random forest classifier training. There is a fair correlation between the two metrics (Strobl et al., 2008).

To achieve the goal of stock return analysis, deep learning can process the massive amounts of data in the financial market, improve data processing capabilities, and extract features from transaction signals (J, S. Z., et al., 2012). For instance, deep learning is used to learn multi-stage behaviour strategies, and stock trading is a sequential decision-making method (Yeh et al., 2011). This technique can determine the state's best return and reduce transaction costs. It therefore has the highest practicability in the investment sector. People use deep learning in the finance industry as a result of its outstanding performance in time series problems, natural language processing, and image classification. By extracting the high-dimensional input data, a deep neural network automatically finds the corresponding low-dimensional representation. Its main objective is to incorporate responders' deviation into the hierarchical neural network structure (Evermann, Rehse, and Fettke 2017). Deep learning therefore has strong feature extraction and perception capabilities. Recursive feedback connections between neuron cells create a directed cycle in recurrent neural networks. It can store and use historical data to make predictions about the future and can offer a plan for building a complex system's cognitive decision-making system.

To predict stock market trends, a lot of research has been done. LSTM and ARIMA models were the subject of numerous studies to determine whether they may benefit investors. The purpose was to examine the efficiency of RNN, in particular the LSTM network, in predicting whether the stock price will increase over the following 15 minutes. The model's application was evaluated based on how accurately it matched real-world data. To see if the model offered any advantages over conventional machine learning algorithms, it was assessed. Despite the fact that it can learn without using approaches for dimension reduction like feature selection. Although the accuracy was acceptable, the model may have been more trustworthy had the variance been lower . It was concluded that training with less data and more epochs produced outputs that were quite accurate by using the model described, which trained the neural network on different epochs (12, 25, 50, and 100). Both data sets' opening price trends can be tracked using the model. However, it is more erratic with larger dataset values (Moghar & Hamiche, 2020).The accuracy gained was over 95%,maing the model’s precision superior to existing models that anticipate many variables at once.

Persio et al. (Persio & Honchar, 2016) explored the proficiency and adequacy of introducing LSTM for financial time series forecasting. Akita et al. (Akita et al., 2016) combined data with information from journal articles to demonstrate how previous events affected the opening price of the stock market. Their proposed formula handled numerical information and printed data to the LSTM system to carry out accurate forecasting. The LSTM model was employed by Chen et al. to forecast China stock returns (Chen, Zhou, & Dai, 2015). The historical information was converted into 30-day sequences with ten learning features and 3-day learning rate labelling. According to the authors, the model's accuracy increased from 14.3% to 27.2% when compared to the random prediction technique. The LSTM network was used by Fischer and Krauss for the classification problem of forecasting directional movements for the Constituent stocks of S&P 500 from 1992 until 2015 (Fischer & Krauss, 2018). After further research, the authors came to a conclusion that LSTM network could effectively extract useful information from financial time series data. Random forests, conventional deep networks, and logistic regression are all outperformed by LSTM in terms of prediction accuracy and daily returns after transaction costs. Based on historical data from the S&P 500, Dow Jones Industrial Average (DJIA), and Hang Seng Index dataset, Qiu et al. constructed an LSTM-based model (Qiu, Wang, & Zhou, 2020).

Hence from the above research, the purpose of this project is to examine

1) the impact of macroeconomic variables affect a company’s stock price and

2) Use a macroeconomic variable to predict the stock price for the next quarter which most of the studies have neglected to explore.

3) Stock market analysis and use of technical indicators to analyse the stock price movement

4)Use the LSTM model for Forecasting.

The current study contributes to the knowledge as follows:

1) Gaining a deeper understanding of how various macroeconomic variables impact various stock markets

2) How technical indicators are good at predicting stock price movements for short-term trading

3)Using Random Forest to effectively find out the most significant variables that affect the stock market and prediction of the stock price using macroeconomic variables

4)How LSTM is so good at predicting and forecasting stock prices using historical data.

5.METHODOLOGY

5.1. Tools Used

The Python programming language is the primary tool we used to analyse the data and implement our models. Python is a widely used language for data analysis because it is simple and user-friendly. It has a wide range of libraries like NumPy, Pandas, Matplotlib, and Seaborn that are made specifically for visualization analysis as well as data analysis, visualization, and manipulation. Furthermore, Python is a fantastic choice for more in-depth data analysis thanks to its robust community support and comprehensive documentation. Using libraries like Stats models, Keras, Sklearn, and Tensorflow helped to support the process of building models and making predictions. Overall, Python is a useful tool for anyone working with data because of its adaptability and accessibility.

5.2Methods Used

Pearson correlation

A statistical indicator of the linear correlation between two variables is the Pearson correlation, also referred to as Pearson's correlation coefficient. Its value falls between -1 and 1, with -1 denoting a strong negative correlation, 0 denoting no correlation, and 1 denoting a strong positive correlation. It is calculated as the product of the two variables' standard deviations divided by the covariance of the two variables. The strength and direction of the linear relationship between two variables are determined using this method. Here the Pearson correlation is used to find the correlation between macroeconomic variables and Stock price.

The correlation coefficient value (r) is calculated as follows:

$$r = \frac{n \times (\sum X,Y) - (\sum(X) \times \sum(Y))}{\sqrt{(n \times \sum(X^2) - \sum(X^2)) \times (n \times \sum(Y^2) - \sum(Y^2)}}$$

where:

n: number of observations

X: First variable
Y: second variable

Technical Indicators

Technical Indicators are mathematical tools to do technical analysis,i.e to do mathematical calculation based on historic price and volume. There are so many technical indicators available.They are divided into different types on the basis of where they are plotted,when do they indicate,what they indicate. The technical indicators that are used in this project are:

1.EMA (Exponential Moving Average)

Exponential moving average gives more weight to recent data and less weight to older data. It is frequently used in technical analysis of financial time series data to amplify long-term trends and smooth out short-term fluctuations. Exponential moving average is much more a better version of a simple moving average that doesnt have SMAs lag.

You must enter a "smoothing constant" or "decay factor" value in order to calculate an exponential moving average. The amount that older data is given less weight depends on this value. A shorter time period is used to calculate the average when the smoothing constant is higher, which gives more weight to recent data and less weight to older data. On the other hand, a lower smoothing constant results in a longer time period over which the average is calculated, giving older data more weight and more recent data less weight.

The formula for calculating an exponential moving average is as follows:

EMA(t) = (Price(t) * K) + EMA(t-1) * (1 - K)

where:

EMA(t) is the exponential moving average at time t Price(t) is the price at time t EMA(t-1) is the exponential moving average at time t-1 K is the smoothing constant.

You can use a simple moving average as the starting value for EMA to calculate the exponential moving average for the first period (t=1) (t-1).

Because they react more quickly to data changes, exponential moving averages are useful for spotting trends and making predictions. They might not always be as dependable as other kinds of moving averages, though, as they can be more susceptible to short-term fluctuations.

2. Bollinger Bands

Bollinger Bands are a technical analysis tool that depicts the high and low of a financial instrument over a specific period by using a moving average and two bands drawn above and below it. Typically, the bands are two standard deviations above and below the moving average, and they change in size in response to how volatile the data is.

Bollinger Bands are frequently used to spot overbought and oversold market conditions, as well as to validate trends and alert traders to potential market reversals. An instrument's price may be deemed overbought and may be due for a downward correction if it is above the upper Bollinger Band. On the other hand, if the price is below the lower Bollinger Band, that could be a sign that it is oversold and is therefore overdue for an upward correction.

3. Momentum

Technically speaking, momentum gauges how strong a price trend is. It is predicated on the idea that prices typically move in the same direction over an extended period of time. There are several different ways to calculate momentum, but one popular approach is to divide the difference between the current price and the price from a predetermined number of periods ago. Using a 14-period momentum indicator, for instance, it can be computed as follows:

Momentum = (Current price - Price 14 periods ago) / Price 14 periods ago

The resulting value is then plotted on a chart, with values above zero indicating an upward trend and values below zero indicating a downward trend. Some traders use momentum as a standalone indicator to generate buy and sell signals, while others use it in conjunction with other indicators or chart patterns to confirm trend strength or identify potential trend reversals.

Standardisation

The MinMax Scaler is a method for standardising a dataset's feature set. The data is transformed such that all of the features are on a same scale, which is often between 0 and 1. This can be helpful for machine learning methods like decision trees or neural networks that are sensitive to the size of the input characteristics.Each feature's minimum and maximum values are first determined by the MinMax Scaler, which then scales the values so that they fall within a predetermined range (usually 0 to 1). The MinMax Scaler would scale the values so that they lie between 0 and 1, with 0 indicating the minimum value and 1 representing the maximum value, for instance, if the minimum value of a feature is 0 and the maximum value is 100.

The transformation's formula is:

X' = (X - Xmin) / (Xmax - Xmin)

where X is the feature's original value, Xmin is its minimum value, Xmax is its highest value, and X' is its altered value.

Feature Selection using Random Forest

Classification and Regression trees are utilized to forecast discrete or continuous variables. A random forest is an ensemble tool that combines these two types of decision trees (Dou et al., 2019). RF combines the effectiveness of various decision tree techniques to anticipate the value of a variable. In a decision tree, which resembles a flowchart, each internal node stands in for a feature (or attribute), each branch for a decision rule, and each leaf node for the result. In a random forest, each decision tree is trained on a unique subsample of data, selected at random using replacement, and using a randomly selected subset of characteristics.

Feature importance in a Random Forest is a gauge of the relative weighting of each feature (or variable) in the dataset. It is a method for figuring out which characteristics are most crucial for forecasting the outcome variable. The average impurity decreases across all decision trees in the forest, for each feature, and is examined to determine the feature importance score after training a Random Forest model. The more significant the feature is deemed to be, the higher the average impurity decrease. There are two main ways to compute feature importance scores in a Random Forest: Gini Importance and Permutation importance. Here we are using Gini Importance. Gini Importance is a measurement of the reduction in impurity that results when the dataset is divided based on a specific feature (measured by the Gini index). The Gini index calculates the likelihood that a randomly selected element from a set would be erroneously identified if it were randomly assigned a label based on how those labels are distributed within the subset. Random Forest develops different regression trees and takes averages of the input dataset(D). By using Gini Importance, the features of dataset(D) are ranked. Dataset (D) contains samples from k classes. The probability of samples belonging to class i at a given node can be denoted as p_i. Then the Gini impurity is calculated using this equation:

$$\text{Gini(D)} = 1 - \sum_{i=1}^{c} (p_i)^2$$

The highest impurity is found in the node with a uniform class distribution. When all records fall under the same class, the least amount of impurity is obtained. Train the Dataset with Random forest with all the k features. Use the Gini importance equation to rank the features. Calculate the RMSE and MAE values of the model and predict the values using the Random Forest regressor.

MAE (Root mean square value Error)

Mean absolute error (MAE) is the measurement of the discrepancy between predicted and actual values i. It is employed to assess how well a model performs for a continuous variable. The MAE calculation averages the absolute differences between the actual and predicted values.

MAE $=\frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)|^2$

RMSE (Root mean square Error)

RMSE is the square root of the average of the squared differences between the predicted and actual values. It is commonly used to measure the error of a regression model.

RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)}$

n: Total number of datasets

y: actual value

$\hat{y}_i$: predicted value

Random Forest Regression

Random Forest Regression is a type of machine learning algorithm that combines several random decision trees, each of which has been trained on a subset of data. The algorithm is more stable and less chaotic when many trees are used. The random forest regression approach, which performs well for big and most types of data, is a frequently used model.
In the Random forest, each tree is produced by the algorithm using a unique sample of input data. A different sample of characteristics is chosen for splitting at each node, and the trees proceed independently of one another. A single result, the prediction of the Random Forest, is produced by averaging the predictions from each of the trees.

After the importing the Dataset(D), the independent variable and dependent variable are separated. Here the independent variables are CPI, Unemployment and GDP. These independent variable are used to predict the dependent variable stock price (Adj close).The dataset is spitted into training set and testing set. The 20% of the data is the testing set and other remaining are the training set. Random Forest regression model is then used to fit the training data. After using Random forest regression, the RMSE value is calculated to determine the accuracy of the model.

LSTM

Hochreiter and Schmidhuber first proposed the LSTM in 1997, and it has since become incredibly popular, especially for use in problems involving time series prediction (Hochreiter & Schmidhuber, 1997). Mathematically speaking, a Recurrent neural network can be defined as a differentiable function that maps one kind of variable to another kind of variable. The most popular model is the LSTM, which is a modified RNN method that can be used to solve a wide range of issues. The neural network design of the LSTM addresses the issue of figuring out how to recollect data over time by incorporating gate units and memory cells. The cell states of the memory cells contain information about recently experienced events. Cell state is combined to determine the outcome as soon as information enters a memory cell, and it is then updated.

LSTM can choose to read it from right to fit or reset the cell using an explicit gating mechanism. The gates are:

1. **Input Gate** controls whether the memory cell is updated.
2. **Forget Gate** controls the memory cell is reset to Zero.
3. **Output Gate** controls whether the information of the current cell state is made visible.

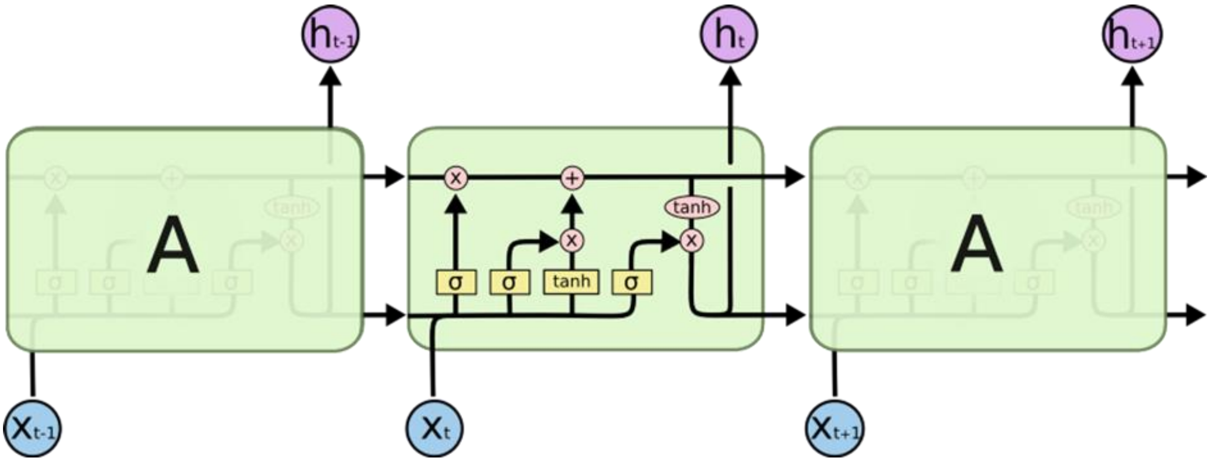


Fig1

Fig 1 shows a pictorial representation of the LSTM model. The number of nodes in a NN with a single hidden layer always depends on the dimension of the data, and the nodes in the input layer communicate with the hidden layer via connections known as "synapses." Every two-node relationship from (the input to the hidden layer) has a coefficient called weight that determines how signals are processed. After completing the learning process, the Artificial NN will have the best weights for each synapse. The process of learning is naturally a continuous adjustment of weights (Moghar & Hamiche, 2020).

The model will then be trained once this process is finished. Recurrent Neural Network (RNN) is a class of NN that uses earlier stages to learn data and predict future trends. It predicts future value based on a past sequence of observations. To forecast and guess future values, it is important to keep in mind the earlier stages of the data; in this case, the hidden layer serves as a repository for historical data from the sequential data. The method of using components of previous sequences to predict future data is referred to as recurrent. Long Short-Term Memory (LSTM) based on "memory line" proved to be very helpful in forecasting cases with long-time data because RNN cannot store long-time memory.

6.DATA AND ETHICAL IMPLICATIONS

6.1 Dataset

The historical stock price of Apple is taken from Yahoo Finance website from 2010 January through 2022 November, which includes the Open price, close price, Highest price and the lowest price in the day, low, Adjacent closing price, and Volume traded. Yahoo finance is one of the most well-known and widely used online resources for financial news and data. It also has a significant user base. Financial information such as news, data, and analysis can be found on Yahoo Finance. Yahoo is a Verizon Communications Inc. subsidiary. Stock quotes, market data, financial news, investment research, and portfolio management tools are just some of the resources available on the site. Users can also view historical stock prices, financial data on specific companies, financial news and analysis, and real-time stock market data. Apple stock traded on Nasdaq Global Select Market under the ticker symbol AAPL was selected. Nasdaq is a stock exchange located in the United States. It was established in 1971 and, after the New York Stock Exchange, is the second-largest stock exchange in the world by market capitalization (NYSE). Apple, Microsoft, and Amazon are just a few of the technology-based businesses that are listed on the Nasdaq, which is also known for its electronic trading platform. Exchange-traded funds (ETFs) and bonds are among the other securities in which trading is available. Because of its many users and high trading volume, Nasdaq offers its shareholders high liquidity. For stock analysis of Apple, the historical price of Apple competitors such as Google, Microsoft, and Tesla are taken for comparison.

Quarterly data on Macroeconomic variables such as GDP, Unemployment rate, and CPI is taken from FRED (Federal Reserve Economic Data) from 2010 January through 2022 November. FRED (Federal Reserve Economic Data) is a database of economic and financial data provided by the Federal Reserve Bank of St. Louis. Time series data on a variety of subjects, including GDP, inflation, interest rates, employment, and trade, are included in the data. The Bureau of Economic Analysis, the Bureau of Labour Statistics, and the Census Bureau are just a few of the government entities that provided the data. The FRED database is a free tool that is accessible to everyone. Economic experts, research scientists, scholars, and finance experts frequently use FRED to analyse and comprehend economic trends and conditions. Through the FRED website, users can access the data and download it in a few different formats, including CSV, Excel, and JSON. The website also offers graphing and data analysis tools, which make it simple to see and comprehend trends.

The adjacent close price is chosen from the historical dataset for analysis and prediction of the company. The closing price of a stock is adjusted in a calculation to produce the adjusted closing price. Compared to the closing price, it is more intricate and precise. Because external factors could have changed the true price, the adjustment made to the closing price represents the stock's true price. Adjusted closing prices have the primary benefit of simplifying the evaluation of stock performance. The adjusted closing price, in the first place, aids investors in understanding how much money they could have made by purchasing a particular asset. Second, investors can contrast the performance of two or more assets thanks to the adjusted closing price. Aside from the obvious problems with stock splits, undervaluing the profitability of value stocks and dividend growth stocks results from failing to take dividends into account. Comparing the long-term returns of various asset classes using the adjusted closing price is also crucial. Below are the following links:

- Yahoo Finance: https://uk.finance.yahoo.com/ where the historical data from Apple, Google, Microsoft, and Tesla are taken.
- FRED (Federal Reserve Economic Data): https://fred.stlouisfed.org/ where the GDP, CPI, and Unemployment rate data are taken.

6.2 Data pre-processing

The data preprocessing includes data Integration, data cleaning, and data transformation. Our dataset comes from a reliable publisher. There was no scenario when we had to deal with empty data. In the Exploratory data analysis part, the historical stock price of 4 big major tech companies Apple, Google, Tesla, and Microsoft taken from 2010 to 2022. For the LSTM model prediction, historical stock data of Apple from 2010 to 2022 were taken. Random forest, which was used to find the impact of macroeconomic variables on affecting stock market the macroeconomic variable such as GDP, Unemployment rate, and Consumer price index (CPI). The quarterly data of Macroeconomic variables were only available in FRED. To find the impact of macroeconomic variables affecting Apple stock price, we ensured that the same timeline of adjacent close price data aligned with the macroeconomic variable timeline. There were no categorical data used because everything was in numerical form as a time series dataset.

6.3 Ethics

Data ethics is the area of ethics that deals with issues relating to privacy, autonomy, and potential biases and harms that may result from the use of data, as well as the gathering, storing, using, and disseminating of data. In addition to developing policies and procedures to make sure that data is used in ways that are fair, just, and responsible, it also entails looking at moral concepts and values that ought to guide the handling and use of data. Our dataset is taken from reliable sources and is taken from Open Government Licensed sources (FRED). Yahoo Finance is a privately owned and operational news and data website which is a part of Yahoo and use government-provided data like financial report or statistics.

The dataset follows the UK ethics framework. The principles include:

Transparency: Businesses should be open and transparent about their data processing procedures and intended uses, and they should notify people of their rights and how to exercise them.

Fairness: Organizations must make sure that data is not utilized in a way that unfairly disadvantages or discriminates against any individuals or groups.

Accountability: Data should only be gathered and utilized for clear, explicit, and legal objectives, and should not be further processed in a way that is at odds with those aims. Data minimization: Organizations should just gather and hold onto the information required for their goals.

7.ANALYSIS AND RESULTS

7.1 Exploratory Data Analysis of Apple

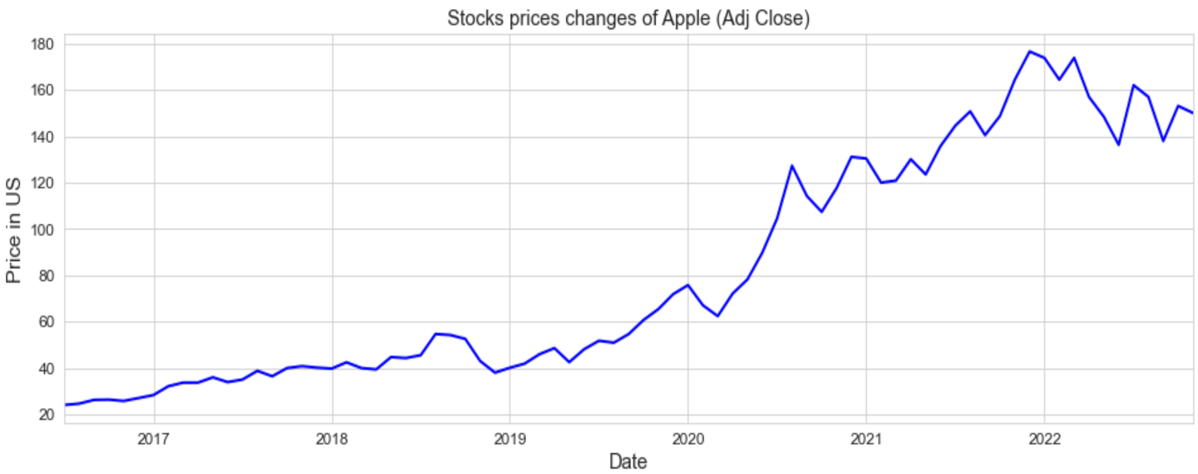


Fig 2

Fig. 2 displays Apple's stock price movements (adjacent close) from 2016 to 2022. According to these graphs, Apple stock has been increasing gradually and saw significant fluctuations following the 2020 pandemic. Although the epidemic has had a significant impact on Apple, we still need to examine the company's economic changes from a micro-perspective. The price began to drop at the onset of the outbreak. The price of the adjacent closing started to decline after this. The end of March saw the lowest stock price of the outbreak. Following that, it began to progressively rise until September 2020, when it peaked. After that, the price began to decline once more. The delay of new Apple goods like the iPhone 12, iPhone 12 Pro Max, etc. could be one of the causes of this decline. People may have raised their expectations for these new items and been waiting a long time, which is why Apple stock may have increased. This drive pushed people to purchase these items as soon as possible, which resulted in a significant rise in the price of Apple's shares. Additionally, Apple benefited from Covid-19 because of the significant rise in users of online Apple services like iCloud, Apple TV, Apple Music, Apple Books, etc. We can observe that in 2021, the apple industry significantly expanded in comparison to all prior years. At the end of 2022, Apple's stock price rose to a record high. The stock price may significantly decline after 2022.

One reason is that pressure on discretionary expenditure is increased by rising interest rates. Currently, many of the company's services and the upkeep of its current devices are fixed expenditures; purchasing new Apple products is unquestionably a discretionary spend. Additionally, Apple's international sales lose value when converted to U.S. dollars due to the strong U.S. dollar in comparison to other currencies. Apple's capacity to produce and market its products is being negatively impacted by China's slowing economy as well as production issues brought on by the pandemic.

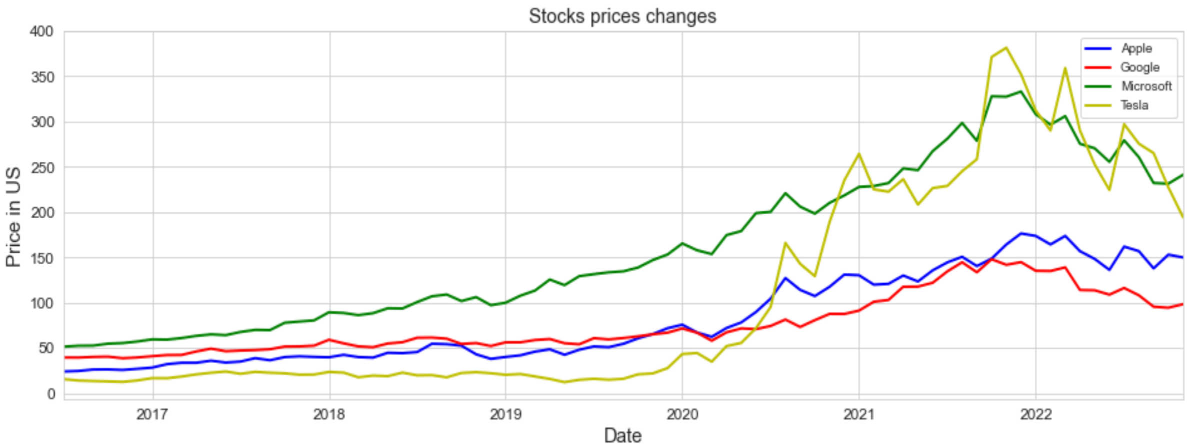


Fig 3

Figure 2 displays the changes in the stock prices of the four major technology companies, including Apple, Google, Microsoft, and Tesla. Tesla had the second-highest stock price in 2016, behind Google and Apple, with Microsoft coming in top. Microsoft was at the top of the list by the end of 2022. The scenario is the same for Tesla, Microsoft, and Google as it was for Apple during the epidemic. In the first few months of the outbreak, there was a significant fall from February to March, but it quickly recovered almost completely. Then, when compared to the other tech behemoths, Tesla's stock price reached the highest level. Then it underwent a significant change and sided with Microsoft. In comparison to other companies, Google's stock price had only slightly increased during the previous five years, and it had experienced a big decline. Overall, Apple was not severely damaged by the pandemic. Instead, it just needed four months to raise or even return the value of its stock. This is sufficient evidence that Apple has a solid base and great tactics for enabling them to quickly recoup its losses after suffering significant impacts. Most of the time, Apple's stock merely fell during the worst of the epidemic and then maintained a constant value. There was also a phenomenon of continual rising.

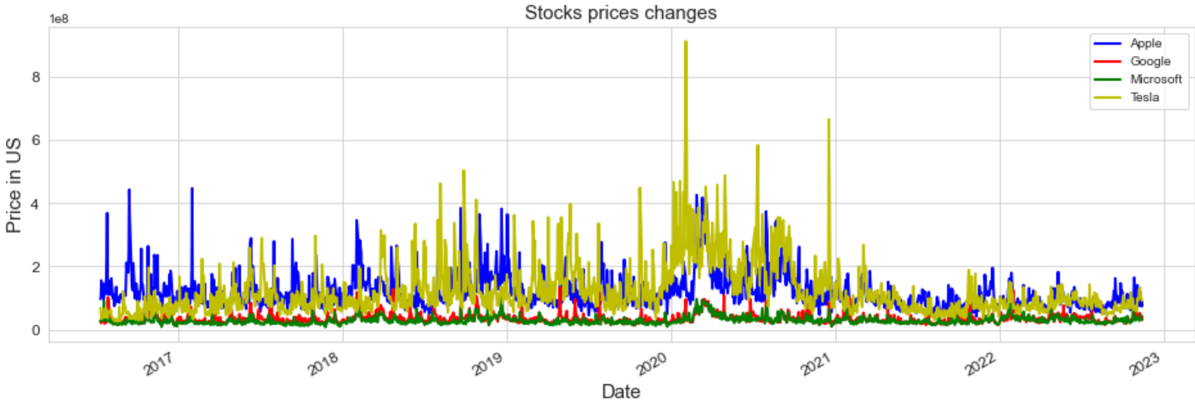


Fig 4

Fig 3 shows the Volume of stock prices traded during the last five years. Microsoft's stock price is much higher than Apple's, but when we compare their number of transactions, we find that Apple's number of transactions is many times greater than that of Microsoft emphasizing that most of the time trading volume of Apple is higher than Apple's. Tesla initially seems to be Apple's complete antithesis. The manufacturer of electric vehicles places a premium on revenue growth but places less emphasis on profits; a few years ago, it was burning through so much money that it was in danger of going out of business.

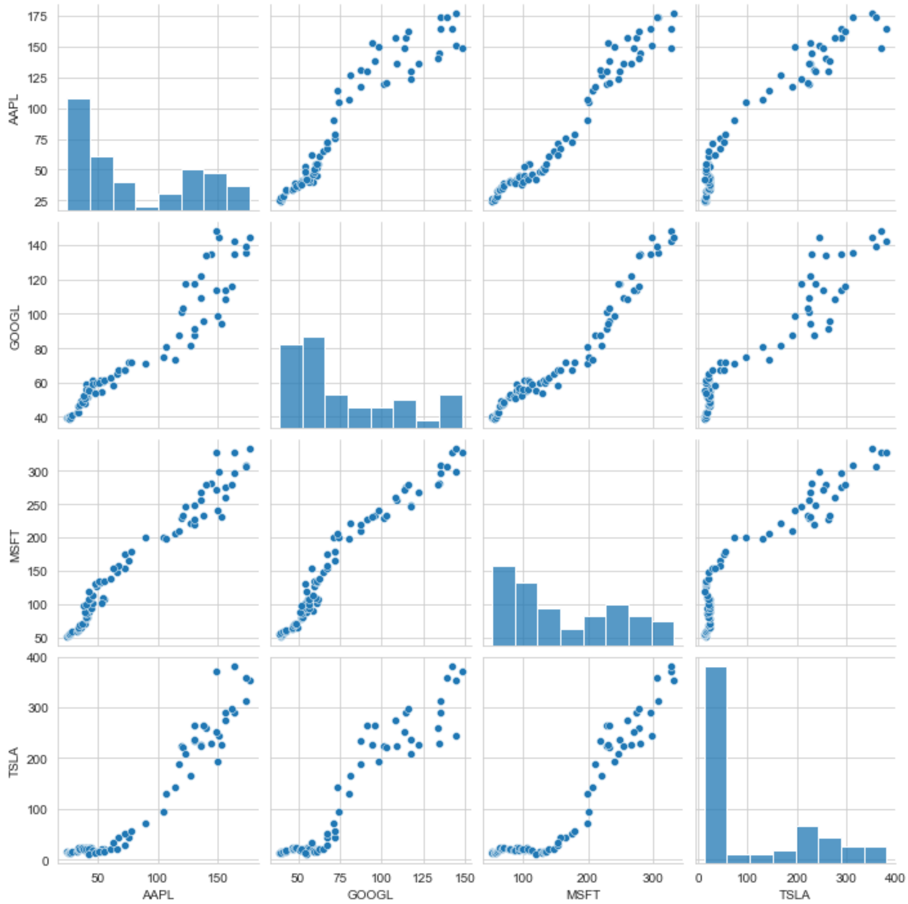


Fig 4

From the plots in Fig. 4, we can deduce that there is a large positive correlation between four big technology businesses, which is logical given that stock prices rise over time. This finding can be supported by the fact that the total average correlation is positive. This is one of the reasons why these companies usually operate in similar sectors or industries, which also means that they are governed by similar legal and economic frameworks. For instance, a large number of technology companies are involved in the production and marketing of technology-related goods and services, which are driven by factors such as consumer demand, competitive pressure, and technological advancement. In addition, the performance of these enterprises may be affected by broader market movements, such as shifts in interest rate levels or the state of the economy as a whole. The final factor that may have an effect on the stock prices of significant technology businesses is the emotion of investors as well as the activities of huge institutional investors. These investors may own sizeable stakes in a number of different tech companies. The strongest positive correlation compared with Apple stock prices are Apple-Microsoft, Apple-Tesla, and Apple-Google.

7.2 Stock Price movement using technical Indicators

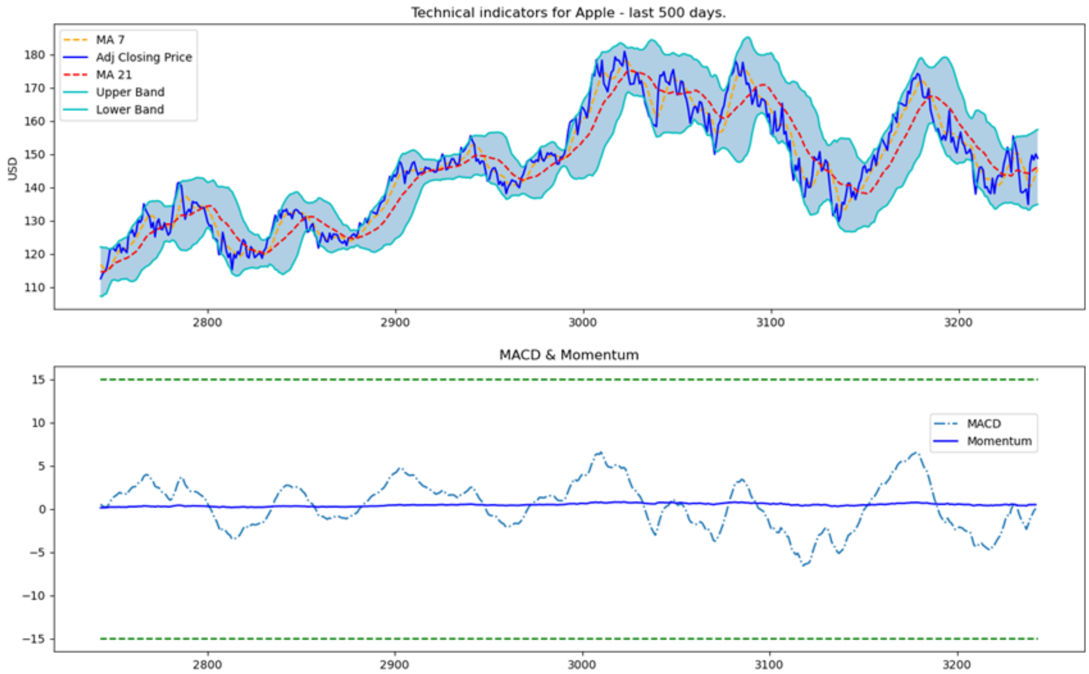


Fig 5

Fig 5 shows the stock price movement of Apple in the last 500 days using different technical indicators. For intraday trading, swing, and positional trading technical analysts predict future prices by analyzing historical data. Here the technical indicators used are Bollinger bands, MACD, and Momentum.

From the first plot, the stock price movement is analyzed using Bollinger bands. Here the Bollinger bands are composed of four lines. The upper band, lower band, 7 days moving average, and 21-day moving average. The middle band is used to determine the upper band, which is then increased by twice the daily standard deviation. The lower band is determined by subtracting the daily standard deviation twice from the middle band. The band widens when there is a price increase and narrows when there is a price decrease. When the stock price crosses the lower band repeatedly, it indicates an oversold signal, while when it crosses the lower band once, it indicates an overbought signal. The best trading strategy is when the stock price reaches the line of the lower band, buy the stocks, and when it reaches the limit o the upper band sells the stocks. According to the movement of the stock price, the price is going to increase, so it is better for the traders to hold the share. The threshold between the MACD and momentum is depicted in the second plot. As you can see, momentum provides the MACD's average value between peak values and the highest or lowest values. In my Analysis, the user should not sell the stock as the price is going up on the next day when and MACD line is increasing as well as the actual stock price is between the bands.

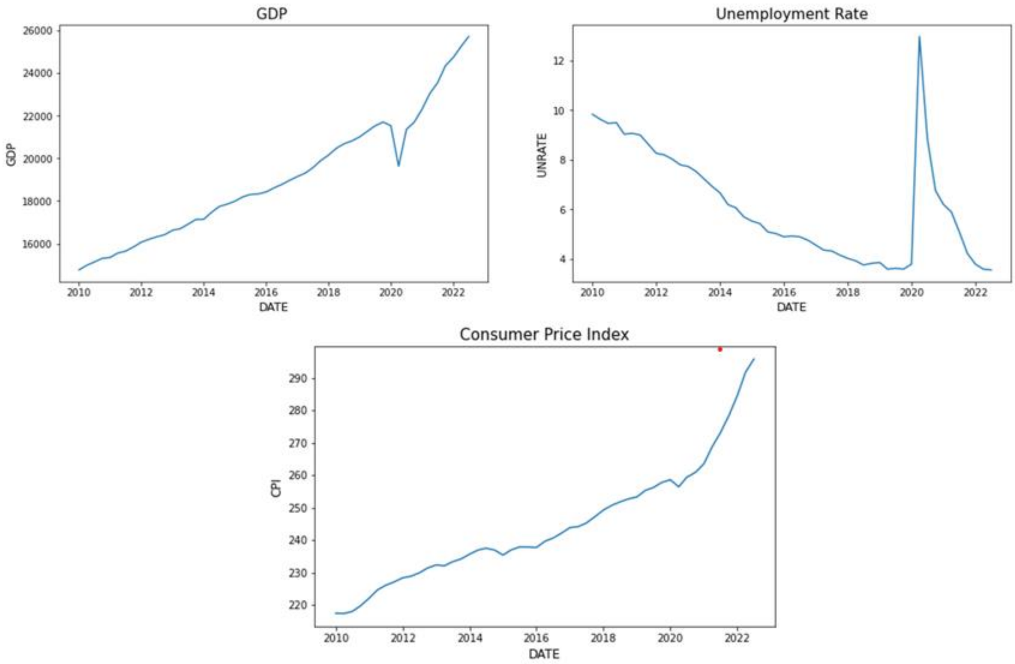


Fig 6

The United States GDP is depicted in Fig. 6. The GDP increased linearly between 2010 and the beginning of 2020. The impact of Covid 19 is then apparent in a huge way. Following that, the nation's economy recovered due to the stimulus measures taken by the US Government, low-interest rates, investment in infrastructure, and increase in manufacturing and production. These might be reasons that the GDP was boosted after the pandemic in a significant manner.

Beginning in 2010, the unemployment rate fell linearly until it peaked at the start of 2020. After 2020, it increased dramatically because of the economic impact. Businesses and industries were severely disrupted by the pandemic, which resulted in numerous layoffs and furloughs. Lockdowns and other restrictions forced many businesses to close or scale back operations, which decreased the demand for labour. Following the pandemic, restrictions were loosened, and the unemployment rate sharply declined.

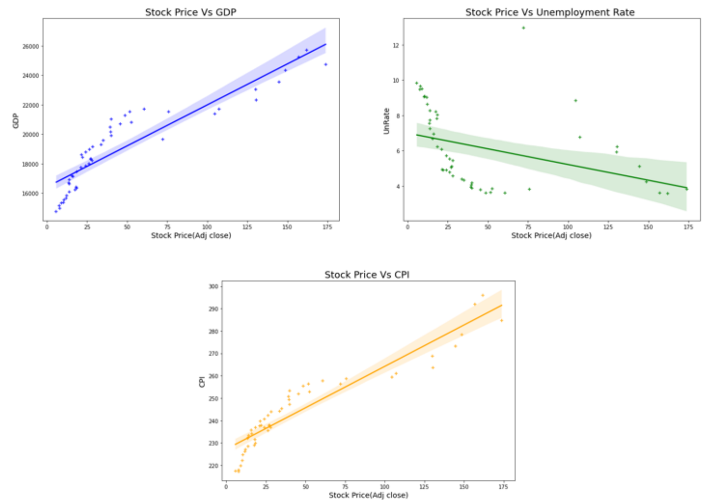


Fig 7

Fig 7 shows variation of stock price with macroeconomic factors GDP, Consumer price index and Unemployment rate.

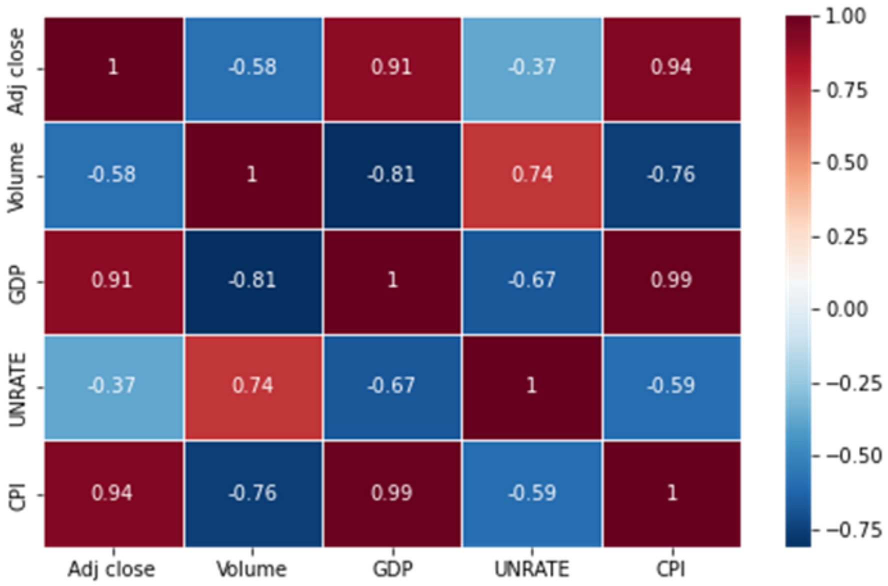


Fig 8

The heat map Pearson coefficient (Fig 8) illustrates the degree to which two variables measured on the same interval or ratio scale are associated with one another. The association between various macroeconomic parameters and stock price is displayed in the heat map.

The level of unemployment has a moderately negative correlation with Apple's stock price, while the Consumer Price Index of the country has a large positive association with the stock price. The nation's gross domestic product has been shown to have a positive link with the stock price of the nation.

7.3 Random Forest

	Feature	Impact Score
1	GDP	0.252116
2	Unemployment Rate	0.038809
3	CPI	0.709075

The table and the fig 9 reveal the most important factors that affect the price of Apple stock using Random Forest. According to the findings, the CPI (Consumer Price Index) is the factor that has the greatest impact on the stock price, with an influence of almost 70%, followed by GDP, which has an influence of approximately 25%. The rate of unemployment has the least impact on the value of Apple's stock. The results show that the Consumer Price Index and Gross Domestic Product are both reliable indicators of changes in the stock price of Apple.

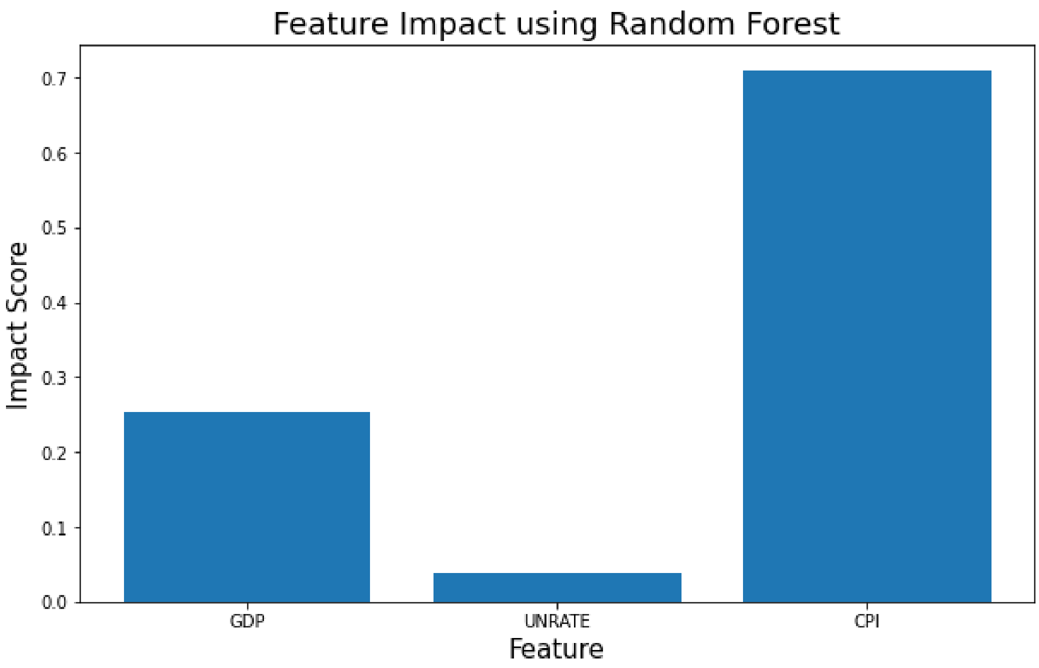


Fig 9

The table below the model score, MAE, and RMSE of the Random forest regressor. The model predicted the last-day value to be 158.35. The actual value was 162.015808.

Model Score	0.9887446785162478
MAE	5.768595295454532
RMSE	8.164310970825088

7.4 LSTM

Step 1: Standardisation

Following dataset loading, high variances are eliminated by utilising Minmax Scaler to normalise the data. This increases calculation speed and makes the model less susceptible to changes brought on by outlier data. The range of the data after normalisation ranges between 0 and 1.

Step 2: Fitting the model

We have defined a double-layer LSTM model. The output shape of the model is (None,100,50). The dataset is split into training and testing data with a split value of 0.65. The training set contains 2106 and the testing size contains 1135 values. The model is compiled and the optimizer used is adam optimization. Adam is a widely used optimization technique for neural networks and other deep learning models. It is a development of the popular stochastic gradient descent (SGD) algorithm for machine learning. The Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation benefits of two more versions of SGD are combined by the Adam optimizer (RMSProp). Adam employs an adjustable learning rate, which implies that rather than utilising a fixed learning rate for all parameters, it modifies the learning rate of each parameter in the model separately. As a result, the algorithm converges more quickly and effectively than it would otherwise.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 lstm_1 (LSTM)               (None, 100, 50)           20200

 lstm_2 (LSTM)               (None, 50)                20200

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0
_________________________________________________________________
```

Fig 10

Fig 10 shows the summary of the LSTM model.The input to the model is the adjacent closing price which we need to forecast.

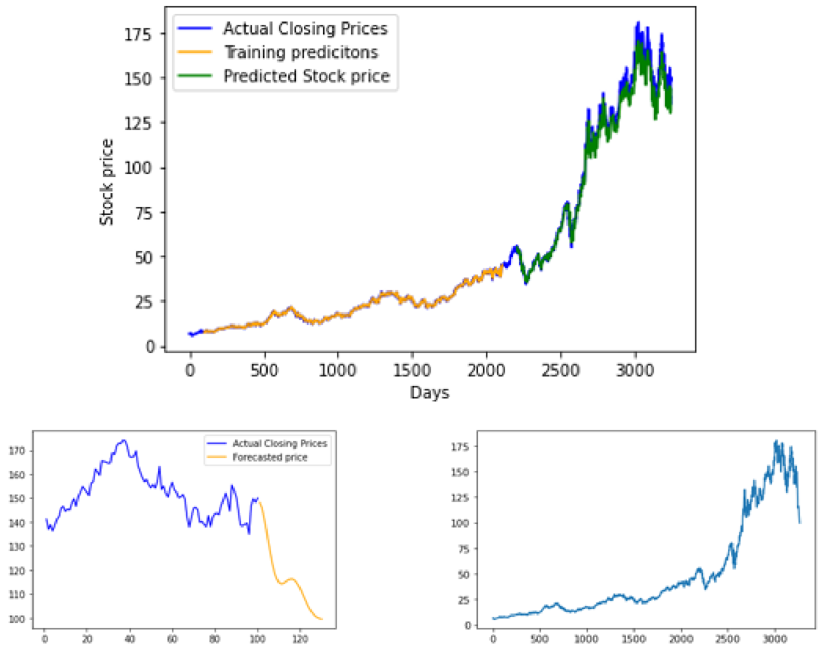


Fig11

Step 3: Forecasting

The above fig shows the prediction and forecasting of Apple stock price using LSTM.First chart shows the prediction of Apple Stock Price .The stock price is predicted for next 30 days and the results show that the stock price is going to decrease.The number of epochs for the model given is 30. The model's parameters are modified during each epoch to improve its ability to fit the data. A hyperparameter that can be changed to balance the trade-off between computational efficiency and model accuracy is the number of epochs.

Root mean square values for training set and testing set is shown in table 3.

Performance metrics for	RMSE
Training set	23.301865682606692
Testing set	109.47560175548688

Our models are more likely to be overfitting if the RMSE is small, whereas underfitting may be indicated by a very big RMSE. Within our range of dataset, our value is pretty good.

8.CONCLUSION & RECOMMENDATIONS

Stock price prediction is a complex task due to the chaotic and non-linear nature of the stock market. Researchers have proposed various methods for feature selection from financial data, such as using technical indicators or historical data. Collinearity between variables can also affect the performance of predictive models. Techniques for making stock price predictions include technical analysis, which examines charts and statistical indicators of prior price movements, and machine learning algorithms, which analyse vast amounts of data and make forecasts based on patterns and trends in that data. With the advancement of technology, researchers have started to build models using artificial neural networks and deep learning techniques, which are more efficient in extracting important information from the data.

In summary, stock market analysis is the process of evaluating and interpreting financial data in order to make investment decisions. There are three main categories of stock market analysis: Fundamental Analysis, Technical Analysis, and Time Series Analysis (Quantitative analysis)Here we have looked into Technical analysis and Quantitative analysis .Technical Analysis is used for Short-term trading, Swing Trading, and Positional trading, it involves using statistics generated by market activity, such as past prices and volume, to identify patterns that can suggest future activity. Time series analysis is a statistical method used to analyse data points collected over a period of time, it can be used to analyze stock prices and make predictions about future values using machine learning and deep learning models. In this project, the focus is on Technical Analysis and Quantitative analysis of Apple stock price, and the impact of macroeconomic factors such as Gross Domestic Product (GDP), Consumer Price Index (CPI), and Interest rate on the Apple Stock price. Apple Inc. is an American multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services, it has been publicly traded on the NASDAQ stock exchange since December 12, 1980.

This project looked into the use of technical indicators to forecast stock prices. Technical indicators are calculations based on a security's price and/or volume. These indicators are employed in pattern recognition and trading decision-making. In this project, we concentrated on Bollinger Bands, Exponential Moving Average (EMA), and Momentum, three widely used indicators. It's also crucial to remember that we shouldn't base your trading decisions solely on technical analysis. Technical indicators work best when combined with both the general market environment and fundamental analysis, which examines a company's financial and economic fundamentals. Additionally, since each indicator has strengths and weaknesses of its own, it is important to use a variety of indicators rather than relying on just one. Additionally, we only used a small number of indicators in this study; however, there are many other indicators available, and it would be intriguing to look into how well they perform in the future. This study showed that it is possible to predict stock prices using technical indicators, and it also highlights the need for more study in this area. Future research should investigate how to enhance the model's functionality and make it more resilient to changing market conditions since the use of technical indicators for stock price prediction is an active area of research.

Macroeconomic variables such as GDP, consumer price index and unemployment can have a significant impact on stock prices. As these variables can change rapidly, it is important for investors to keep a close eye on them, and to adjust their investment strategies accordingly. Random forest was used find the impact of the macroeconomic variables. Consumer price index had the most impact on apple stock price, followed by GDP. Unemployment rate had the least impact. The algorithm works by training multiple decision trees and combining their predictions to make a final prediction. Each tree is trained on a randomly selected subset of the data, and the feature importance is determined by the number of times a feature is used to split the data across all decision trees. One of the limitations of using Random Forest for feature selection is that it can be computationally expensive, especially when dealing with large datasets. Since our datasets is not that large, Random Forest works perfectly. Furthermore, it may be difficult to interpret the results of the feature selection process, especially when working with many features.

We investigated the use of Long Short-Term Memory (LSTM) models to predict stock prices. LSTMs are a type of recurrent neural network that are particularly well suited for handling time series data, such as stock prices. We first pre-processed the data by normalizing the prices and splitting it into training and testing sets. Then, we implemented an LSTM model and trained it on the training data. We also experimented with different architectures, such as adding multiple layers and dropout regularization, to improve the model's performance. The results of the study showed that the LSTM model was able to predict stock prices with a relatively high degree of accuracy. The model's performance was evaluated using metrics such as root mean square error. The model performed well on the testing set, indicating that it had good generalization ability. The model predicted that stock price of apple would fall in the next 30 days. It's important to note that the stock market is a highly dynamic and unpredictable environment, and even the best models will not be able to predict future prices with complete accuracy. Additionally, the results of this study are specific to the dataset and the parameters used, and different results may be obtained using different datasets or parameters.

Moreover, LSTM models are a powerful tool for stock price prediction and can be integrated with other techniques such as fundamental analysis, technical analysis, and news analysis to improve predictions. In the future, we can also try other advanced models such as Attention-based LSTM and Gated Recurrent Unit (GRU) to see if they perform better. Overall, this study demonstrated the feasibility of using LSTM models for stock price prediction, and it highlights the potential for further research in this area. The use of LSTM models for stock price prediction is an active area of research and future work should investigate how to improve the model's performance and make it more robust to different market conditions.

9. REFERENCES

1. Agarwal, M., Ullah Khan, A., & Shukla, P. K. (2019). Stock Price Prediction using Technical Indicators: A Predictive Model using Optimal Deep Learning. *International Journal of Recent Technology and Engineering*, *8*(2), 2297–2305. https://doi.org/10.35940/ijrte.b3048.078219
2. Ahangar, R. G., Yahyazadehfar, M., & Pournaghshband, H. (2010). The Comparison of Methods Artificial Neural Network with Linear Regression Using Specific Variables for Prediction Stock Price in Tehran Stock Exchange. *ArXiv:1003.1457 [Cs]*. https://arxiv.org/abs/1003.1457
3. Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). *Deep learning for stock prediction using numerical and textual information*. IEEE Xplore. https://doi.org/10.1109/ICIS.2016.7550882
4. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. C. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, 100320. https://doi.org/10.1016/j.mlwa.2022.100320
5. Biondo, A. E., Pluchino, A., Rapisarda, A., & Helbing, D. (2013). Reducing financial avalanches by random investments. *Physical Review E*, *88*(6). https://doi.org/10.1103/physreve.88.062814
6. Bollen, J., & Mao, H. (2011). Twitter Mood as a Stock Market Predictor. *Computer*, *44*(10), 91–94. https://doi.org/10.1109/mc.2011.323
7. Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web Search Queries Can Predict Stock Market Volumes. *PLoS ONE*, *7*(7), e40014. https://doi.org/10.1371/journal.pone.0040014
8. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324
9. Chen, K., Zhou, Y., & Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. *2015 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/bigdata.2015.7364089
10. Ding, G., & Qin, L. (2019). Study on the prediction of stock price based on the associated network model of LSTM. *International Journal of Machine Learning and Cybernetics*, *11*(6), 1307–1317. https://doi.org/10.1007/s13042-019-01041-1

11. Dou, J. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, *662*(662), 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221
12. Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, *29*(4), 927–940. https://doi.org/10.1016/j.eswa.2005.06.024
13. Evermann, J., Rehse, J.-R., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems*, *100*, 129–140. https://doi.org/10.1016/j.dss.2017.04.003
14. Fernando, J. (2021, October 5). *Correlation Coefficient Definition*. Investopedia. https://www.investopedia.com/terms/c/correlationcoefficient.asp
15. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054
16. Foelber, D. (2023, January 2). *3 Reasons Apple Stock Keeps Falling*. The Motley Fool. https://www.fool.com/investing/2023/01/02/3-reasons-why-apple-stock-keeps-falling/
17. Ghosh, P., Neufeld, A., & Sahoo, J. K. (2021). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*, 102280. https://doi.org/10.1016/j.frl.2021.102280
18. Hasan Shuvo, Md. N., Shahriar Maswood, M. M., & Alharbi, A. G. (2020, June 1). *LSRU: A Novel Deep Learning based Hybrid Method to Predict the Workload of Virtual Machines in Cloud Data Center*. IEEE Xplore. https://doi.org/10.1109/TENSYMP50017.2020.9230799
19. Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *06*(02), 107–116. https://doi.org/10.1142/s0218488598000094
20. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

21. Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, *129*, 273–285. https://doi.org/10.1016/j.eswa.2019.03.029
22. Istiake Sunny, Md. A., Maswood, M. M. S., & Alharbi, A. G. (2020). Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. https://doi.org/10.1109/niles50944.2020.9257950
23. Jareño, F., & Navarro, E. (2016). European Inflation and the Spanish Stock Market. *European Review*, *24*(4), 609–630. https://doi.org/10.1017/s1062798716000272
24. Jareño, F., & Negrut, L. (2015). US Stock Market And Macroeconomic Factors. *Journal of Applied Business Research (JABR)*, *32*(1), 325. https://doi.org/10.19030/jabr.v32i1.9541
25. Jiang, H., Deng, Y., Chen, H.-S., Tao, L., Sha, Q., Chen, J., Tsai, C.-J., & Zhang, S. (2004). *BMC Bioinformatics*, *5*(1), 81. https://doi.org/10.1186/1471-2105-5-81
26. Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311–5319. https://doi.org/10.1016/j.eswa.2010.10.027
27. Karmiani, D., Kazi, R., Nambisan, A., Shah, A., & Kamble, V. (2019). Comparison of Predictive Algorithms: Backpropagation, SVM, LSTM and Kalman Filter for Stock Market. *2019 Amity International Conference on Artificial Intelligence (AICAI)*. https://www.semanticscholar.org/paper/Comparison-of-Predictive-Algorithms%3A-SVM%2C-LSTM-and-Karmiani-Kazi/dc493907b904cde136f1191055d95a24b6c3fbf2
28. Laichena, K., & Obwogi, T. (2015). EFFECTS OF MACROECONOMIC VARIABLES ON STOCK RETURNS IN THE EAST AFRICAN COMMUNITY STOCK EXCHANGE MARKET. *International Journal of Education and Research*, *3*(10). http://ijern.com/journal/2015/October-2015/24.pdf
29. Malkiel, B. G., & Fama, E. F. (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

30. Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science, 170*, 1168–1173. https://doi.org/10.1016/j.procs.2020.03.049

31. Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn.2017.7966019

32. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction. *American Journal of Applied Sciences*, *16*(7), 200–212. https://doi.org/10.3844/ajassp.2019.200.212

33. OECD. (2022). *GDP and spending - Gross domestic product (GDP) - OECD Data*. OECD. https://data.oecd.org/gdp/gross-domestic-product-gdp.htm

34. Parchaliuk, D. L., Kirkpatrick, R. D., Simon, S. L., Agatep, R., & Gietz, R. D. (1999). Yeast two-hybrid system: part A: screen preparation. *Technical Tips Online*, *4*(1), 6–15. https://doi.org/10.1016/s1366-2120(08)70129-6

35. Park, S. H., Kim, B., Kang, C. M., Chung, C. C., & Choi, J. W. (2018, June 1). *Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture*. IEEE Xplore. https://doi.org/10.1109/IVS.2018.8500658

36. Persio, L., & Honchar, O. (2016). *Artificial Neural Networks architectures for stock price prediction: comparisons and applications*. https://core.ac.uk/download/pdf/217561207.pdf

37. Qiao, R., Chen, W., & Qiao, Y. (2022). Prediction of stock return by LSTM neural network. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2151159

38. Radhakrishnan, P. (2018, January 13). *Introduction to Recurrent Neural Network*. Medium. https://towardsdatascience.com/introduction-to-recurrent-neural-network-27202c3945f3

39. *Random Forest Algorithm: A Complete Guide | Built In*. (n.d.). Builtin.com. https://builtin.com/data-science/random-forest-algorithm#feature

40. Reader, T. C. (2021, December 2). *Random Forest Regression Explained with Implementation in Python*. Medium. https://medium.com/@theclickreader/random-forest-regression-explained-with-implementation-in-python-3dad88caf165#:~:text=What%20is%20Random%20Forest%20Regression

41. Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. *Procedia Computer Science*, *135*, 89–98. https://doi.org/10.1016/j.procs.2018.08.153

42. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1). https://doi.org/10.1186/1471-2105-9-307

43. Sun, W. (2021). *The Changes of Apple's Stock Price During the Pandemic*.

44. *Tesla, Microsoft, Google or Amazon: Which is emerging as megacap rival to Apple?* (2022, September 27). Mint. https://www.livemint.com/market/stock-market-news/tesla-microsoft-google-or-amazon-which-share-is-emerging-as-megacap-rival-to-apple-11664247600895.html

45. Tretina, K. (2021, April 22). *How Does The Stock Market Work?* Forbes Advisor. https://www.forbes.com/advisor/investing/what-is-the-stock-market/

46. Yeh, C.-Y., Huang, C.-W., & Lee, S.-J. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications*, *38*(3), 2177–2186. https://doi.org/10.1016/j.eswa.2010.08.004

10.APPENDIX

All the coding and datasets related to this project are uploaded in the DATAFILE.

Business Proposal

AIM

Stock Market Research Analysis and Impact of Macroeconomic Factors on Stock Market and prediction of the stock market using LSTM (Long Short-Term Memory).

OBJECTIVE

The company taken for the Research Purpose is Apple Inc.

- Stock Market Research Analysis of Apple.
- Technical Analysis of Apple stock Price
- Impact of macroeconomic factors such as GDP, Unemployment rate on Stock market
- Prediction of Apple stock market price using the LSTM method.

Data Set

The historical data of Apple Inc (Nasdaq, AAPL) from 2010 January to 2022 November is obtained from Yahoo Finance.

Yahoo! Finance is a media property that is part of the Yahoo! Network. In addition to stock quotes, news articles, financial reports, and original content, it offers financial news, data, and commentary. Additionally, it provides a few online tools for managing personal finances.

Quarterly data Macroeconomic factors such as GDP, Unemployment rate are taken from FRED (FEDERAL ECONOMIC RESERVE DATA). The Federal Reserve Bank of St. Louis maintains a database of economic and financial information called FRED (Federal Reserve Economic Data). The data include time series information on a range of topics, including GDP, inflation, interest rates, employment, and trade. Several government agencies, including the Census Bureau, the Bureau of Economic Analysis, and the Bureau of Labour Statistics, contributed the data.

About The Company

Apple Inc. is an American multinational technology company headquartered in Cupertino, California, United States. Apple is the largest technology company by revenue (totaling US$365.8 billion in 2021) and, as of June 2022, is the world's biggest company by market capitalization, the fourth-largest personal computer vendor by unit sales, and the second-largest mobile phone manufacturer. It is one of the Big Five American information technology companies, alongside Alphabet, Amazon, Meta, and Microsoft. Apple became the first publicly traded U.S. company to be valued at over $1 trillion in August 2018, then $2 trillion in August 2020, and most recently $3 trillion in January 2022. The company receives criticism regarding the labour practices of its contractors, its environmental practices, and its business ethics, including anti-competitive practices and materials sourcing. Nevertheless, the company has a large following and enjoys a high level of brand loyalty. It is ranked as one of the world's most valuable brands.

Literature review

There is ongoing debate about the ability to predict stock returns given the complexity, chaos, and constant change in the stock market. Different methods have been proposed for analyzing the stock market to improve predictions. The Efficient-Market Hypothesis, which argues that an asset's current price always reflects all information that is publicly available, was first introduced by Malkiel and Fama in 1970.

The study by Jareo and Negrut (2016) utilising Pearson correlation in the United States found statistical relationships between GPD, the industrial output index, the unemployment rate, long-term interest rates, and stock return. Additionally, it was discovered that stock returns and MVs highly connected in Kenya, Uganda, and Tanzania (Laichena and Obwogi, 2015). By employing the Ordinary Least Squares method, Asekome and Agbonkhese (2015) discovered a weak association between the exchange rate, capacity utilisation, and stock returns and a substantial relationship between Nigeria's GDP and money supply stock returns.

In order to forecast financial time series, Persio et al. (Persio & Honchar, 2016) investigated the effectiveness and suitability of LSTM introduction. Akita et al. (Akita et al., 2016) used data and details from journal papers to show how past events impacted the stock market's opening price. In order to perform precise forecasting, their suggested formula handled numerical data and printed input to the LSTM system. Chen et al. used the LSTM model to predict China stock returns (Chen, Zhou, & Dai, 2015). The historical data was transformed into thirty-day sequences with 10 learning features and three-day learning rate labelling.

Analysis outcome

- Apple Stock Market Research Analysis.
- The impact of macroeconomic variables like GDP and the unemployment rate on the stock market.
- The LSTM approach for predicting Apple stock market price and prediction graph of stock of 30 days.