DEPARTMENT OF MANAGEMENT SCIENCE

University of Strathclyde Business School

EXPLAINABLE AI DEEP REGRESSION MODEL FOR STOCK PRICE PREDICTION

Ikechi Loveday Ogbujah

202284532
MSc Data Analytics
Academic Supervisor: Prof Anil Fernando
Industrial Supervisor: Prof Anil Fernando
University of Strathclyde

30/08/2023

Declaration

Except where explicitly stated, all the work in this dissertation including any appendices is my own and was carried out by me during my MSc course as part of the group under the supervision of Prof Anil Fernando. It has not been submitted for assessment in any other context.

Signed:

Ikechi Loveday Ogbujah
Date: 30/08/2023

Summary

The Stock market is a multi-billion-pound industry globally, and a such accurately predicting the price of stocks presents an excellent opportunity to make healthy returns on investment trading this equity item. Over time there have been different attempts to achieve this feat from the efficient market hypothesis to the random walk model and lately a foray into the use of machine learning/artificial neural networks to improve the accuracy in this regard. For all its apparent superiority in predicting stock prices, machine learning/artificial learning introduction into the space has not achieved the tremendous lift as anticipated due largely to the lack of trust by its intended users i.e., stock analysts and market speculators given that it is often treated as a black box, this is because there is no information as to its internal decision mechanism leading to lack of transparency and interpretability.

In an attempt to bridge this trust gap, a body of knowledge know as explainable artificial intelligence or XAI for short was born, some of the earlier algorithms include sensitive analysis, layer-wise relevance propagation, and LIME (Local Interpretable Model-agnostic Explanation) all of which did not seem to fully explain the prediction mechanism and susceptible to one downside or the order until Lundberg and Lee (2017) proposed the SHAP (SHapley Additive exPlanations), which is a robust technique that consolidated six previous additive feature attribution methods (LIME, Deep Lift, Layer-wise relevance propagation, Shapley regression values, Shapley sampling values, Quantitative input influence).

This project was carried out by extracting S&P 500 historical data from yahoo finance via API, the data was further enriched by using feature engineering to develop technical data, pre-processing helped to reduce dimensionality by leveraging the PCA module. Three separate models were used to predict the stock price i.e., LSTM, GRU and a voting regressor ensemble containing the individual outputs of four models (XGBoost, LightGBoost, SVM and Gaussian process regressor all hyper parameter tuned with grid search to get the best combination for each model). Performance of each model was compared using 3 different metrics RMSE, MAE and MSE.

Finally, SHAP algorithm was applied in order to explain the decision-making mechanism of each model and how it rated the different features in order to arrive at its prediction in each instance, interestingly there was divergence in the features that were prioritized by the different models, Regular LSTM had low price, open price and high price as its top 3, PCA-LSTM model open price, high price and percentage change , GRU low price, high price and open price while the voting regressor prioritized open price, low price and percentage change, which was a surprise but exciting outcome of this project as it gives further credence to the pursuit of explainable AI algorithms given that two different configurations of the same model can make decisions in remarkably different ways and hoped that it will be found very useful by the industry subject matter experts.

Acknowledgement

Completing this dissertation would not have been possible without the invaluable assistance of several individuals, and I wish to express my sincere gratitude to each of them.

First and foremost, I extend my appreciation to Strathclyde Business School for granting me this enriching opportunity. I am also deeply thankful to almighty God for his mercies and to my beautiful wife Nnenna and my lovely children Chidinma, Chimaobi, Chizurum and Chizaram whose unwavering support and understanding were essential to the success of this work.

My gratitude extends to Prof Anil Fernando who wore the dual hat of Client and Supervisor, and whose guidance, and responsiveness to my queries played an instrumental role throughout the project. And to the project liaison Piyumi Perera, her efforts in facilitating feedback sessions with the supervisor significantly contributed to ensuring that the information flow and coordination was seamless.

Finally, I extend a heartfelt thank you to my team members Jaeseung Park and Shahan Khan whose collaboration were fantastic throughout the journey of this dissertation.

Ethics

This dissertation was written as part of a group work involving two other current students Jaeseung Park and Shahan khan under the supervision of Prof Anil Fernando. Permission was granted to collaborate in the development of the various models involved.

There are no ethical issues concerning the S&P 500 dataset as it was extracted from a publicly available database via the yahoo finance website and this information was declared as the ethics approval process completed on Myplace portal.

Table of Contents

List of Figures and Tables

Project Setting

Client Background

The department of management sciences is situated within the business school arm of University of Strathclyde and its mainly focused on the intersection of business and science. One of its critical objectives is empowering students with the right knowledge, technical competence and capability required to enable the evaluation of various business niches to understand their operations and pains points, and applying the most effective theories, tools, and techniques to design and develop fit for purpose solutions to meet those needs. The client, Prof Anil Fernando who is a highly respected subject matter expect doubled as the supervisor and provided useful guidance in the execution of the project.

One of such needs is the yearnings by users of automated decision-science-based products which are developed using artificial neural network architecture on the opaque nature of inner workings of such models and hence the low trust has hindered a fast-track transition from a largely manual/semimanual prediction systems to fully automated AI led online real-time nimble models.

Project background and Motivation

Academia and industry practitioners all agree that artificial neural networks can achieve significant traction if more sustained efforts can be geared towards unravelling the mystery often associated with understanding the inner workings of the models to identify the basis of its prediction in any field, this lack of transparency and interpretability has been the albatross standing between AI and increased adoption especially in critical sectors like medical research.

This dissertation is yet another contribution in the propagation of explainable AI body of work, the model is trained on a mature stock market (S&P), then used to predict a less mature market (KOSPI index in South Korea) to see how well the model generalizes with unseen data.

Project Plan

Figure 1: Project Plan

Task	Start	End	JUNE				JULY				AUGUST			
WEEK			1	2	3	4	1	2	3	4	1	2	3	4
Project Identification	01-Jun	04-Jun												
Submit Research Proposal	03-Jun	03-Jun												
Data Mining	02-Jun	05-Jun												
Literature Review	02-Jun	30-Jun												
Write and Finish Dissertation	02-Jun	30-Aug												
Development of Models	12-Jun	04-Aug												
Model&Graph Modification	13-Aug	16-Aug												
Submit Dissertation	30-Aug	30-Aug												

DEPARTMENT OF MANAGEMENT SCIENCE

University of Strathclyde Business School

EXPLAINABLE AI DEEP REGRESSION MODEL FOR STOCK PRICE PREDICTION

Ikechi Loveday Ogbujah

202284532
MSc Data Analytics
Academic Supervisor: Prof Anil Fernando
Industrial Supervisor: Prof Anil Fernando
University of Strathclyde
30/08/2023

Table of Contents

List of Figures and Tables

Executive Summary

Contained within this body of work is the development of three separate models (Voting regressor, GRU and LSTM) for predicting the daily stock price of S&P 500 and the use of SHAP module to interpret the internal decision mechanism used by the different models in arriving at the predicted prices.

This was achieved by importing data from yahoo finance via the use of dynamic API architecture ensuring the most up to date data from the website, six features were imported initially (open, low. high, close, volume and adj close) and further enriched using feature engineering techniques to develop technical indicators (Percentage change, MACD, Signal Line, Volatility, Middle Band, Std Dev, Upper Band, Lower Band) which are typically useful in the manual day to day prediction exercise by stock analysts in real life scenarios.

To reduce the influence of features with large magnitude on the outcome, MinMax scaler was used to keep the values to between 0 and 1, while Principal component analysis (PCA) was employed to reduce the dimensionality while retaining the important features. The data was then split into train and test datasets, this was important such that after building the different models, the former will be used to train the model and latter acts as a check to avoid under/over fitting ensuring that it generalizes well when deployed to live systems. Given the unique nature of the data i.e., time series, it was critical that the split considered the sequential nature of the datapoints while splitting in order not to lose the important relationship that exist between subsequent datapoints and hence for this reason, a special type of split module within the scikit library called “Timeseries split” had to be imported and utilized.

Finally, the pre-processed input was fed into the 4 models viz Regular stacked LSTM, PCA stacked LSTM, GRU (Gated recurrent unit) and voting regressor (an ensemble which was configured to use the output from four different algorithms – (XGBoost, LightGB, SVM, and Gaussian process regressor) as input to make its own superior prediction leveraging the strengths of individual algorithm while mitigating their weaknesses. Each model went into an iterative tuning approach to find the best hyper parameter combination to yield the best result, and comparison was achieved by using three popular metrics RMSE (root mean square error), MAE (mean absolute error) and MAPE (mean absolute percentage error) to measure the respective performances.

The results of each model was plotted using the SHAP module to show the features that were most important in the prediction, this last part is the climax of the entire dissertation effort since it is expected to improve the adoption rate by the intended user groups as it provided the much needed transparency on how the models made their prediction and allows for validation against the current decision models employed in everyday trading process.

Acknowledgment

Completing this dissertation would not have been possible without the invaluable assistance of several individuals, and I wish to express my sincere gratitude to each of them.

First and foremost, I extend my appreciation to Strathclyde Business School for granting me this enriching opportunity. I am also deeply thankful to almighty God for his mercies and to my beautiful wife Nnenna and my lovely children Chidinma, Chimaobi, Chizurum and Chizaram whose unwavering support and understanding were essential to the success of this work.

My gratitude extends to Prof Anil Fernando who wore the dual hat of Client and Supervisor, and whose guidance, and responsiveness to my queries played an instrumental role throughout the project. And to the project liaison Piyumi Perera, her efforts in facilitating feedback sessions with the supervisor significantly contributed to ensuring that the information flow and coordination was seamless.

Finally, I extend a heartfelt thank you to my team members Jaeseung Park and Shahan Khan whose collaboration were fantastic throughout the journey of this dissertation.

Introduction

Evolution of Stock Trading

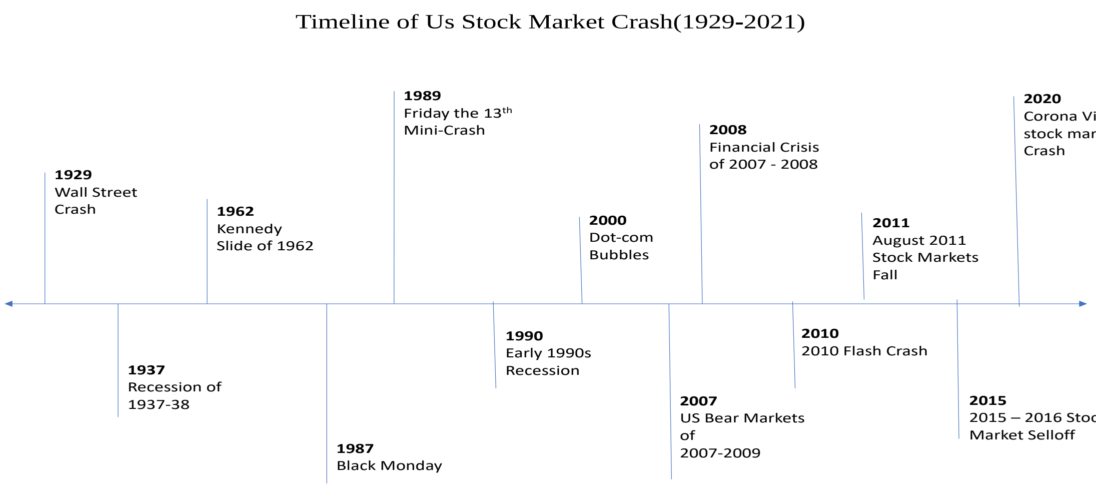
While the London Stock Exchange is recognized as the oldest in its current form, tracing back to recorded history, the origins of stocks can be found even further back in the 1300s (Mueller (2019), Smith et al. (2004) as cited in Beattie, A. (2023)). During this time, Venetian merchants began selling debt issues to lenders and individual investors, marking the earliest semblance of stock trading. In the 1500s, Belgium gained prominence as the first country to exclusively deal in promissory notes and bonds. Subsequently, by the 17th century, various East Indian companies in Europe, issuing stocks based on anticipated success from voyages to distant lands, primarily Asia. This led to a stock market boom and, ironically, the first recorded stock market crash.

Growth and Challenges in the Stock Market

Based on the most recent data on [1], a renowned financial statistics website, the domestic equity markets capitalization worldwide grew from 65.04 trillion U.S. dollars in 2013 to 98.5 trillion U.S. dollars in 2022. As of April 2023, the website quotes that the total market capitalization of domestic companies listed on the various stock exchanges worldwide was 108.23 trillion U.S. dollars, which is a growth of 66.40% in just under 10years [1].

On the opposite side of the spectrum, the diagram below shows that the US stock market has crashed 13 times in the last 100 years with devastating consequences for investors([9][10][11][12][13][14] as cited in Williams, W. (2023)), for instance the information on the said article titled "Timeline of US market crash" quoted the loss attributed to the dotcom bubble alone at over 5 trillion U.S dollars (Williams, W. (2023).

Figure 2: Timeline of US Stock Market Crash (1929 – 2021)



These data points underpin how critical it is to make the right prediction for individual stock prices and therefore if there was such a system that could predict prices, its value cannot be overemphasised as stock market analysts can then depend on it to conduct their daily trades which would lead to increased revenue.

Use of technology in stock prediction

The importance of accurate stock price prediction becomes evident when considering these data points. If a reliable prediction system existed, stock market analysts could rely on it for their daily trades, leading to increased revenue. Over time, various attempts have been made to use software technology for stock price prediction. Initially, basic supervised machine learning models like logistic regression, random forests, and support vector machines were developed using historical price data.

More recently, advancements in the machine learning field have incorporated concepts like artificial neural networks, which utilize technical data, financial news, and social media posts in developing the prediction model. This has significantly improved the accuracy and robustness of predictions by employing models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), specifically the Long Short-Term Memory (LSTM) approach within RNNs. Practitioners now experiment with hybrid approaches, combining different machine learning algorithms into an ensemble to enhance model robustness and accuracy.

The need for Explainable AI

Despite these advancements, the challenge of neural networks being treated as black boxes remains. This lack of transparency and the inability to precisely describe how machine learning models learn and make decisions hinder their adoption by market analysts in real-life situations. To address this, a discipline called explainable artificial intelligence (XAI) has emerged. XAI utilizes tools like Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) to overcome the trust deficit and improve transparency.

The Project's Objective

This project aims to investigate and highlight the powerful nature of SHAP algorithm to interpret the internal decision mechanism of machine learning models. To achieve this, data for S&P 500 index will be extracted from the yahoo finance website via API integration and be employed due to its industry maturity and sophistication for building and training the model, while to test the model's generalization, a less mature market, such as the KOSPI index in South Korea will be used. Four different models will be developed, and evaluated using three key performance error metrics i.e., RMSE, MAE & MAPE, the SHAP model will then be used to show the features prioritized by each model in making the prediction.

Chapter 2: Literature review

This section looks at previous work done around stock market prediction historically from mathematically models to application of computational models as well as the evolution of explainable AI and the different approaches available out there to achieve interpretability and transparency.

2.1 History of stock prediction models

One of the earliest theories in the equity market industry was efficient market hypothesis (EMH), which broadly defined means that prices "fully reflect" the total information available at that time. A more comprehensive explanation was offered recently by Malkiel (2003)

"The accepted view was that when information arises, the news spreads very quickly and is incorporated into the prices of securities without delay. Thus, neither technical analysis, which is the study of past stock prices in an attempt to predict future prices, nor even fundamental analysis, which is the analysis of financial information such as company earnings and asset values to help investors select "undervalued" stocks, would enable an investor to achieve returns greater than those that could be obtained by holding a randomly selected portfolio of individual stocks, at least not with comparable risk."

In trying to validate this theory empirically, Fama (1970) conducted a series of experiments which culminated in his famous random walk model which extended the former by implying that that successive changes are independent and identically distributed represented by the equation below.

$$f\ (rj,\ t+1|\Phi t) = f\ (rj,\ t+1) \qquad (1)$$

(Where f is the density function, rj is the price of the equity at time t+1 and Φt is the total information available at time t+1)

The outcome was strongly in favour of the efficient market theorem. However, Malkiel (2003) held a slightly different view suggesting that price irregularities and even predictable patterns can persist for short periods, concluding that markets cannot be perfectly efficient and that it is possible for the collective judgement of investors to be wrong.

Another very popular mathematical model is the famous "Black-Scholes options pricing model (BSOPM)" developed by Fisher and Myron (1973), options are loosely like an index and hence relevant to this study. This model is based on factors like volatility, risk free-rate, stock price, volatility of the stock price and time. Anwar (2008) provided an equation to show this relationship using a second order parabolic partial differentiation function noting that it is a powerful tool for equity option valuation, noting that volatility had the highest impact on the model and can often bias it.

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0$$

Where r = risk-free rate, V = value of the option w.r.t to time and stock price, σ = volatility = stock price, t = time

Mckenzie (2007) conducted an experiment to empirically examine the accuracy and statistical significance of BSOPM using the Australian index ASX as baseline while applying tools like maximum likelihood, logit and probit models and concluded that it was relatively accurate and statistically significant at 1% level when the factors are considered as a group.

However, Srivastava and Shastri (2020) in their study of the Indian market held a different view as they stated unequivocally in their study that BSOPM was ineffective as there was insignificant relationship between the theoretical and actual price of an option, instead providing an alternative narrative that traders often take advantage of arbitrage opportunities to forecast price for subsequent days.

2.2 Machine learning based prediction models.

2.2.1 Traditional Machine learning algorithm

Due to the advancement in the machine learning space, it has become increasingly popular to find sophisticated models developed to make a better prediction of stock prices. The earlier efforts used models like Support Vector Machine (SVM), Random Forest, ARIMA just to mention a few, while most recent approaches have tilted towards artificial neural networks like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM, a variant of RNN preferred for its unique memory characteristics). Obthong et al. (2020) conducted a survey of over 25 different machine learning algorithms useful for stock price prediction, highlighting their advantages and disadvantages, e.g., the study asserted that SVM can provide optimal global solution but is sensitive to outliers while ANN can learn from the non-linear nature of time series data but does have tendency to overfit, thereby helping anyone trying to choose a model to understand the limitations and make appropriate decisions.

Reddy (2018) developed a model based on SVM using 2 features each from the fundamental analysis and technical analysis respectively (price volatility, price momentum, sector volatility and sector momentum). Radial Basis Function (RBF) was the kernel used for hyperparameter tuning, RBF essentially a low band pass filter for smoothening the graph. The result showed high degree of accuracy while avoiding over fitting (it generalized well when exposed to unseen data). Another study by Gururaj et al. (2019) comparing the prediction from a linear regression model and SVM on stock data for Coca cola from 2018 to 2019 showed that the latter produced superior outcome to the former as the RMSE stood at 3.22 and 1.58 respectively.

Henrique et al. (2018) also used SVM for stock price prediction specifically targeting 3 markets- Brazil, China and USA, However unlike Reddy(2018), this study employed both the RBF as before as well as another kernel - linear kernel and adopted a different set of metrics for the technical analysis input (Simple moving average(SMA), weighted moving average(WMA), Relative strength Index(RSI) and Accumulation/Distribution Oscillation(ADO)). The goal of the experiment was to test the basis of the claim by EMH i.e., that it is impossible to achieve above market returns adjusted for level of risk consistently over time. The outcome was measured using the root mean square error metric which showed that the model had smaller error compared to the random walk model for some stocks meaning that the former had superior predictive power in those instances while the RBF showed a superior prediction to all stocks when the model was tweaked to be updated daily based on live data.

Rather than look at a single model, Kumar et al (2021) implemented an ensemble of 3 different algorithms consisting of the following components - deep learning, gradient boost machine (GBM) and distributed random forest (DRF) while the data extracted consisted of the historical data only with the following columns "Index (time in minutes)", "Date", "Time", "Open", "High", "Low" and "Close" only the Index and Close were used as input and target respectively. The result showed that the ensemble outperformed each of the individual results posting a 0.99 accuracy and RMSE of 0.1 respectively.

2.2.2 Artificial neural networks

Over the last couple of years, artificial neural networks have become increasingly popular in various sectors of the economy, the finance space and stock markets are no exception as we have seen varied novel solutions towards predicting the price of blue-chip stocks. These technologically driven innovations have come in various shapes and sizes ranging from the use of only historical data to more sophisticated algorithms like Natural language processing to evaluate information from the news media and sentiment analysis to improve overall accuracy.

One of the earliest efforts in this regard was made by Amilon (2003) where he conducted a study using a multilayer perceptive neutron model to predict the call option price of Swedish stock index data (1997 – 1999) and showed that it outperformed the black-scholes formula in both the pricing and hedging performance benchmarks. Chowdhury et al. performed a more comprehensive experiment leveraging the power of the RapidMiner solution software, a software that allows for simultaneous testing of different models including decision tree variants as well as other ensemble learning models and arrived at the same conclusion on the superiority of neural network models over BSOPM with the best accuracy achieved using the ensemble model. However, it is worth pointing out the fact since the study was carried out for frontier markets, it was assumed that the volatility and risk-free interest rate which are key critical components of the BSOPM will have to be kept constant given that options trading did not exist in the target market which is markedly different from the usual scenario where these facts were assumed to change over time.

Jin et al. (2019) designed a model using investor sentiment analysis (Convolution neural network was used for analysis) and historical data (decomposed using Empirical Modal Decomposition), both are fed into an LSTM algorithm, it was discovered that these two improvements outperformed a vanilla LSTM model proving that both the extraction of useful sequence from the time series and feeling of investors (EMH) are important. In quantitative terms, the RMSE and Accuracy where (8.71,0.6) and (3.19,0.701) for plain LSTM and improved LSTM respectively. Similar studies by Guo (2020), Mehtab and Sen (2021) also arrived at a similar conclusion on the significance of investor sentiment in predicting the price of a stock. On their part, Srijiranon et al. (2022) developed a hybrid model architecture using PCA-EMD-LSTM and leveraged FINBERT for the sentiment analysis which again resulted in a much higher performance compared to vanilla LSTM.

Nabipour et al. (2020) carried out an extensive empirical experiments to predict the price of stocks in the Iranian stock market for 1, 2, 5, 10, 15, 20, and 30 days ahead using 10 technical indicators as input and 9 different models including 3 artificial neural network based (RNN, LSTM, ANN) six decision tree based algorithms (Decision Tree, Bagging, Random Forest, Adaboost, Gradient Boosting, and XGBoost) the outcome showed LSTM had the most accurate result and fitting capability while Adaboost, Gradient Boosting and XGBoost ranked as the best tree based models. Nikou et al. (2019) also came to a similar conclusion on the superiority of LSTM model for stock price prediction after evaluating results for four models (RNN, RF, SVF, LSTM) using the iShares MSCI United Kingdom. Conversely, Khare et al. (2017) conducted a study with 3 technical indicators as inputs using two different models 1. Multi-layer feed forward neural network and 2. LSTM to compare the prediction of 10 selected stocks from the New York Stock Exchange (NYSE). It was discovered that the former surprisingly performed better than the latter for short term predictions (model was used to predict price for the next minute) contrary to expectation.

For their part, Shrivastav and Kumar (2022) chose to build an ensemble model comprising three algorithms, deep learning, gradient boost and distributed random forest. The idea behind this configuration is in recognition of the no model is perfect, hence it tries to take advantage of the different strengths of the models while avoiding their individual weaknesses and it is in fact the motivation for one of the models examined in this dissertation. This was further echoed by Zhao et al. (2021) in their analysis and evaluation of about 86 papers on stock prices/ forex price prediction wherein one of their conclusions was that hybrid models look very promising having showed positive signs and required more research in future.

2.2.3 Explainable Artificial Intelligence

Despite its superior prediction capability, ML based models were often treated as a black box given the difficulty in getting insights its internal mechanism, and therefore to improve trust and dependency, the field known as explainable artificial intelligence (XAI) was born (Adadi and Berrada (2018)). Samek et al. (2017) proposed two methods for achieving this functionality, the first was sensitive analysis which explains the prediction based the model's locally evaluated gradient, by assuming that most relevant input features are those that the out is most sensitive to which is then represented using a heatmap, while the second method called **layer-wise relevance propagation** performs a backward redistribution of relevance score to every each input variable from the output stage all the way to the very first input stage.

Ribeiro et al. (2016) equally introduced another technique called LIME (Local Interpretable Model-agnostic Explanation) which explains a model by approximating it locally within a model, in other words it provides a granular text-based explanation of a specific instance within a model and in at the same time have global perspective and relevance hence the term model-agnostic. For instance, both Bandi et al. (2021) and Gite et al (2021) used this to explain the results of their respective LSTM based model which used technical fundamentals and sentiments from financial news as input. Another example is Çelik et al. (2023) that leveraged the algorithm in an EMD-ANNRC-RF-LIME architecture to predict and explain the importance features used by the model to predict direction of stock index price across six different markets simultaneously.

Most of the models earlier introduced had one downside or the other making it difficult to fully rely on its interpretation, To address these issues, Lundberg and Lee (2017) proposed the SHAP (SHapley Additive exPlanations), It is a robust technique which consolidates six previous additive feature attribution methods (LIME, Deep Lift, Layer-wise relevance propagation, Shapley regression values, Shapley sampling values, Quantitative input influence) into one solution with improved computational performance and better consistency with human intuition that any of the previous methods. Mandeep et al. (2022) used SHAP to explain the outcome of various models they built as part of the study including SVM, Random Forest and several ANNs and demonstrated the importance of specific features in each of the models. Kumar et al. (2022) developed a model based on the DQN algorithm and used SHAP to explain a stock analyst actions while trading i.e., buy, hold, or sell the stock option.

Chapter 3: Methodology

This looks at the end-to-end process flow from data sourcing to detailed analysis of the results and recommendation while highlighting any feature engineering steps carried out to enrich the data. While there may be many methods out there, this dissertation will largely follow an adapted version of the process flow outlined by Jiang (2021), this is because it was the outcome of a survey that looked at the current state of play in the industry by analysing numerous articles and studies.

3.1 Data Acquisition and Extraction

The data for this project was extracted from the yahoo finance web page, which is open-source secondary data source, there were two ways to ingest this data, and both approaches are explained below.

The first which is the more traditional approach is to download the data into a csv file on the local machine and feed it into python using the pandas library command, while this ensures consistency and reliability of the dataset as it is insulated from any risk of corruption in case anything happens to the source database, it places a huge burden on the local drives storage resources and can be a big performance issue depending on the size of the data.

A second more modern approach is to connect to the website with the use of API (application programming interface), the merit of this method is that it is very nimble and efficient as the data is not stored in the local drive but only required anytime the model is initialized, however it suffers from the fact that any change in the data structure e.g. re-arrangement of columns will negatively impact the model. While it is not impossible to envisage the crystallization of this risk, the probability of it occurring is very minimal as the current format has become standardized over the years and has become the de-facto standard in stock market data presentation.

Hence the second method was adopted due to inherent advantages as stated above.

3.2 Data Cleansing

After extraction, the data will be examined for the presence or otherwise of missing values, skewness, correctness and consistence of data types and the necessary steps taken to correct any anomalies identified. For instance, missing data can be deleted if it constitutes less than 5% of the total data or filled with mean or median values, in this case there was no missing data.

3.3 Feature engineering

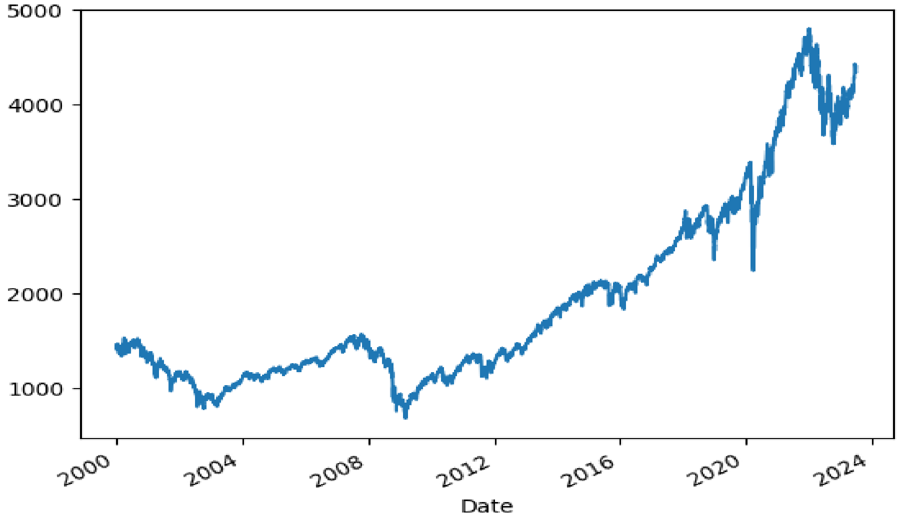
Feature engineering involves the enrichment of the original dataset by including columns that are derived from performing further mathematical operations like average, standardization and dimensionality reduction in order improve the input into the model. The following feature engineering steps were applied to the data.

Times series data has two very notable characteristics which adds to the complexity in building models to use the data to make predictions which can be seen in the figure below.

- Non-linearity: This is defined as the absence of direct relationship between an independent and dependent variable.

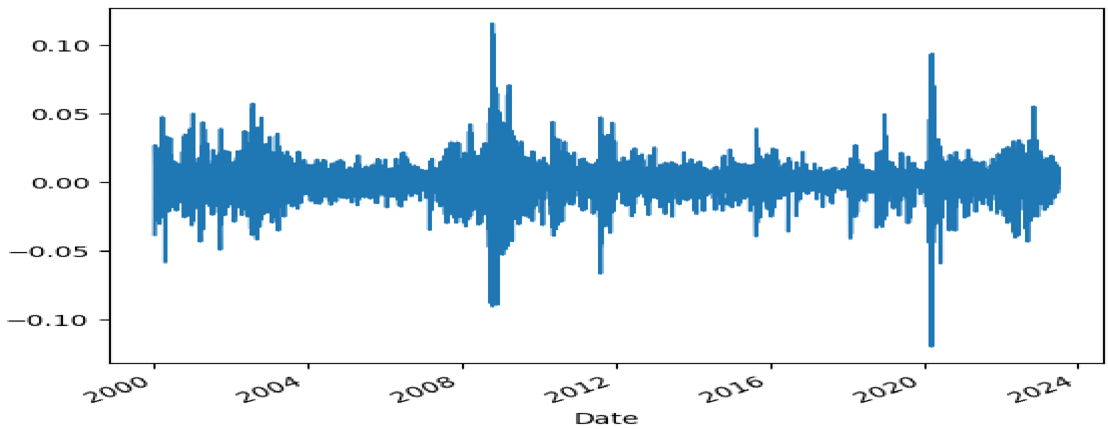
- Non-stationarity: Can be explained as the scenario where statistical properties of data like mean, variance change overtime.

Figure 3: S&P 500 Stock price trend (2000 – 2023)



To account for these conditions, a column was created to capture the percentage change in a bid to centre the data around zero as seen below.

Figure 4: Percentage Change in stock price of S&P 500(2000 – 2023)



It is a common practice by stock market analysts to use technical indicators like relative strength index, overlays, stochastic oscillator etc in addition to the historical data in making stock predictions. For this model, eight of such indicators were derived and added as input to the model (Percentage change, MACD, Signal Line, Volatility, Middle Band, Std Dev, Upper Band, Lower Band)

3.4 Pre-processing.

3.4.1 Splitting the data.

Building a machine learning model requires that the data is split into two, a larger set is used to train the model while a holdout data never seen by the model and designated as test data is used to evaluate its performance. An easy way to achieve this is by using the train test split module and indicate the split ratio, this has a short coming as the either side might not contain a sufficient representation of the different segments contained in the data, which is sometimes overcome by setting the stratify parameter to “y”. A more efficient practice is using something like the k-fold cross validation approach which breaks the data into different samples and selects one part as test set per time and continuing until it iteratively tests all the samples to get a more balanced output.

Unlike other data types, using the vanilla k-fold cross validation will not work for time series as this will not recognise the importance of the information related to its sequential nature, and for this case, a unique module known as time series split was developed by sci-kit learn which was deployed in this project.

3.4.2 Standardization

Different features in a data set can have varying magnitudes and this has the tendency to affect the performance of the model as the large numbers tend to influence the results, to mitigate this risk, pre-processing steps like standardization, normalization etc can be employed. In this model, standardization was achieved using the standard scaler from sci-kit learn module, this module subtracts the mean from each sample and divide the outcome by the standard deviation.

3.4.3 Dimensionality reduction

Summarizing a large dataset into a smaller set of features is useful as it allows for easy visualization and analysis, however it is difficult to manually select which columns to keep or exclude, this is where a tool like PCA comes in as you can either ask it to dynamically make this selection or you can supply the allowable variance explicitly and the module will evaluate the features and choose the features meeting this criteria. The automatic method was used for this project which resulted in PCA dropping the least important feature.

3.5 Models

In the early days of machine learning, data scientists typically built models with single machine learning algorithms, however, over time practitioners have unanimously agreed that machine learning models are not perfect, each one has its strengths and weakness, therefore it is now usual practice to combine various models in some way to improve the overall performance of the model.

For this project we built 4 different models, three from the recurrent neural network family which has shown to be good predictors of time series data (Regular LSTM, PCA LSTM, GRU) and a voting regressor ensemble which aggregates the output of four individual algorithms and uses same as input to make the final prediction.

3.5.1 Recurrent neural network

The first three models were chosen from the family of recurrent neural network (RNN) models.

I. Long short-term model (LSTM): Perhaps the most popular algorithm for handing sequential data e.g., timeseries due to its unique architecture which has three distinct parts (Srijiranon et al. (2022))
   a. Forget layer gate: this determines which information from the previous stage is forgotten or passed on to the subsequent memory block.
   b. Input layer gate: useful for decide if there is need to introduce new data to the current stage.
   c. Output layer: This determines the information to passed to the next hidden later and is dependent on both the forget gate and input gate.

These characteristics helps it to avoid the vanishing or exploding gradient problem and makes it handle sequential data much better than other models.

The Regular LSTM was configured as plain vanilla while PCA was employed for dimensionality reduction before passed into the PCA LSTM model hence the name.

II. Gated Recurrent Unit (GRU): This is a simplified version of the LSTM but with fewer parameters, it has two gates instead of 3 like in the case of LSTM.
   a. Reset gate: Determines how much of the information from the previous memory block should be remembered.
   b. Update gate: Determines the magnitude of the new information should be added to the current layer.

   It tends to have a faster training time because of the lesser number of features, but this does not guarantee a better performance compared to its more robust kin i.e., LSTM.

All three RNN based models where training using the sliding window method with look back period of 10days, and built using standard configuration consisting of input layer, hidden layers and out layer and hence does not require much introduction, however, it is expedient to dwell more on the 4th Model (Voting regressor) due to its complex nature and unique architecture.

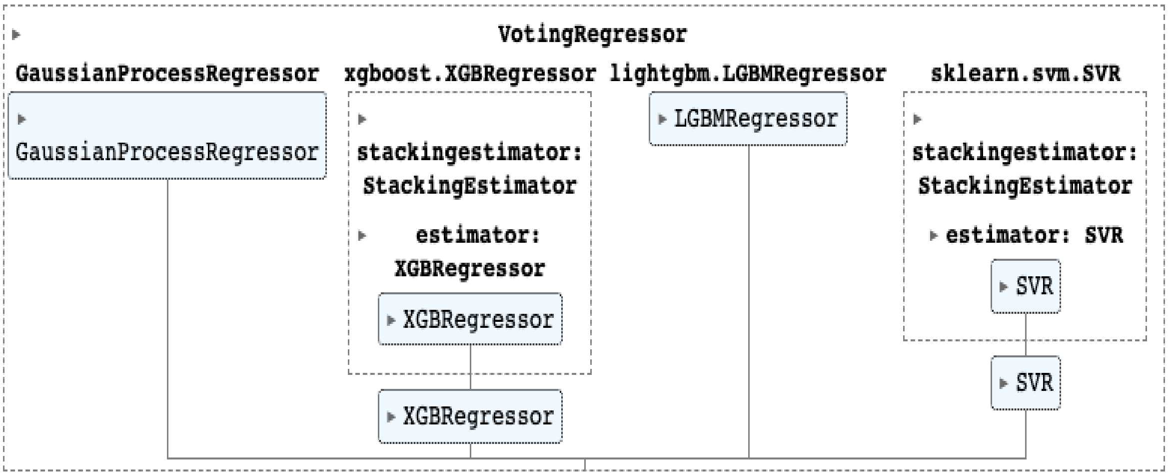
3.5.2 Voting regressor ensemble

A hybrid approach to modelling was adopted for the fourth model, specifically utilizing an ensemble model known as the voting regressor. This technique involves combining the outputs of several individual models to generate a final prediction. The ensemble model selected for this project integrates the following distinct models:

1. XGBoost: XGBoost is a robust gradient boosting algorithm that excels in handling complex datasets. It's known for its high performance and ability to capture intricate patterns in data.
2. Light GBM: Light Gradient Boosting Machine (LightGBM) is another gradient boosting framework that stands out for its efficiency and speed. It's particularly suitable for large datasets due to its optimized handling of categorical features.

3. Gaussian Process Regressor: The Gaussian Process Regressor is a probabilistic model that can capture uncertainty in predictions. It's often used in cases where the relationships between variables are not linear and exhibit complex patterns.
4. Support Vector Machine (SVM): SVM is a powerful classification and regression algorithm that works well in cases where the data is not linearly separable. It's effective in capturing complex decision boundaries.

Figure 5: Voting regressor ensemble



Creating an ensemble model of this nature requires careful parameter tuning to ensure optimal performance. The complexity of the model and the variety of algorithms used make manual hyperparameter tuning impractical. To address this challenge, the project employed the grid search module from the sci-kit learn library. Grid search automates the process of hyperparameter tuning by systematically searching through different combinations of parameters to find the optimal configuration for each algorithm.

The adoption of this hybrid model demonstrates a sophisticated approach to predictive modelling. By leveraging the strengths of multiple algorithms, the ensemble model can capture a broader range of patterns and make more accurate predictions. This project's emphasis on hyperparameter tuning further ensures that the model's performance is optimized. The output of the various models is used as input into the voting regressor model to generate the final prediction for the model, this method takes the strengths from the constituent models while mitigating the weakness and therefore able to predict a more superior and reliable outcome.

The models were train using S&P 500 data from yahoo finance and used to predict the south Korean KOSPI index to see how well it performs with unseen data.

3.5.3 Performance measurement

The performance of the models is measured using three error metrics RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error), these metrics measure the error difference between the predicted and actual values. The smaller the number, the better the performance of the model.

3.6 Model Interpretability

3.6.1 SHapley Additive exPlanations (SHAP)

This is the explainable component which tries to provide insight into the decision-making process of the model by showing the contribution of individual features thereby offering interpretability and transparency the model's decision-making process. It was selected because it is very robust having been developed as a consolidation of six different models, so it is an ensemble of sorts making it perhaps the most powerful explainer model out there.

4.0 Analysis

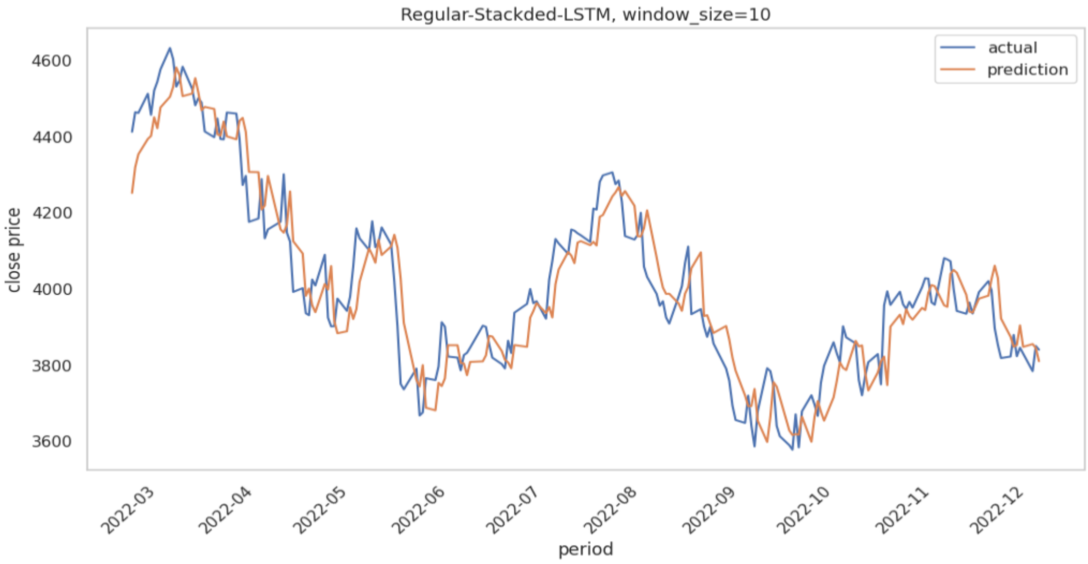
This section divides the analysis into two parts, the first looks at the performance of the different models across the three error metrics (RMSE,MAE,MAPE) which is the standard way of selecting the best model in the traditional machine learning domain, in order to show the importance of explainable artificial intelligence hereafter referred to as XAI for short, we look under the hood to see exactly how each of the models arrived at their decisions.

4.1 Model performance

4.1.1 Regular LSTM

While not exact, the regular LSTM model prediction closely mirrors the actual results and in fact has a few points where the model appears to have predicted the actual price.

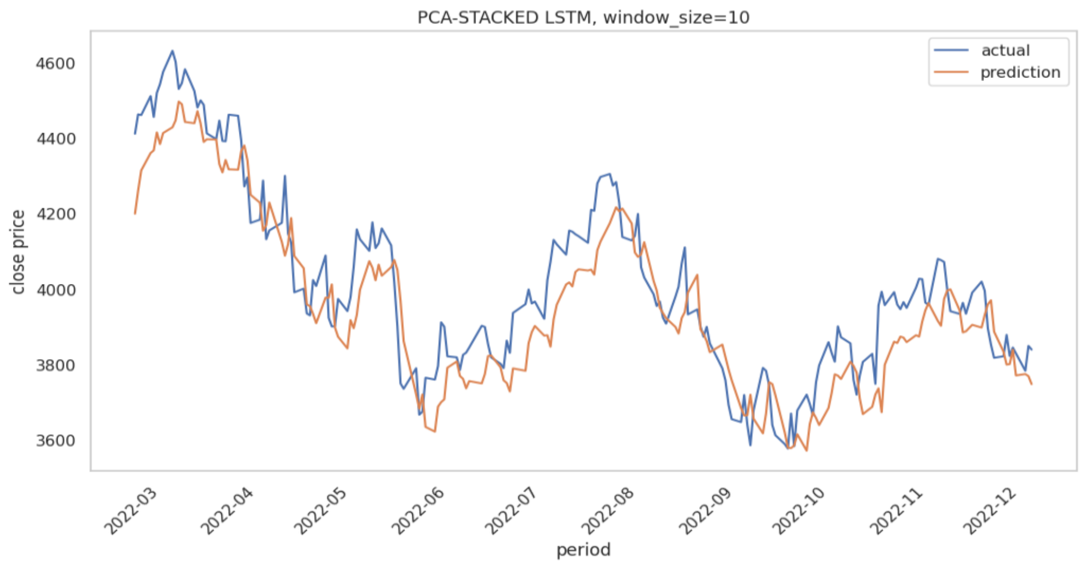
Figure 6: Regular Stacked LSTM Prediction vs Actual



4.1.2 PCA LSTM

Like the vanilla model, the PCA-LSTM model also predicts the trend of the actual price with a sawtooth graph, however, there appears to be two distinct patterns noticed with one type being a very tight fit between actual and prediction, while the other shows marked space between the two respectively, this is different when compared to the graph for the vanilla model as the latter has a fairly even distance between actual and predicted prices along the graph.

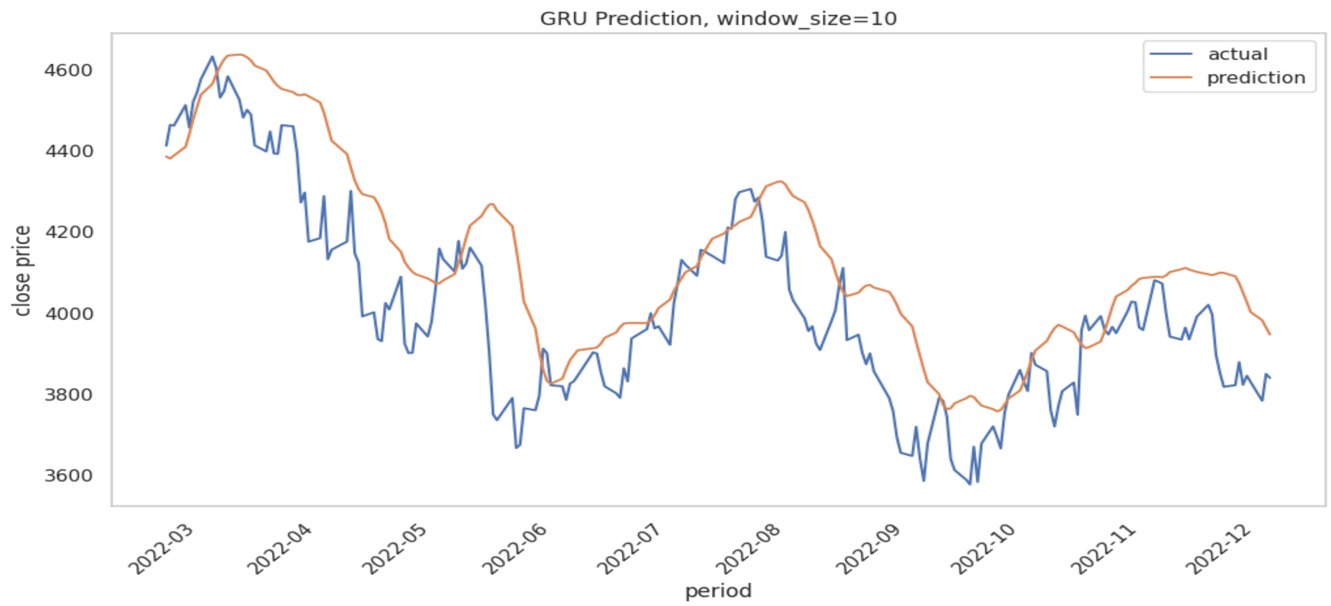
Figure 7: PCA Stacked LSTM Prediction vs Actual



4.1.3 GRU

Broadly speaking, the GRU prediction always follows the actual price trend, however the appears to predict a higher price that the actual in majority of the instances after starting off closest to the beginning of the chart. Furthermore, there is also a somewhat smoothing effect compared to the sawtooth graph exhibited by the actual price trend, perhaps suggesting that the model is not learning enough about the dataset.

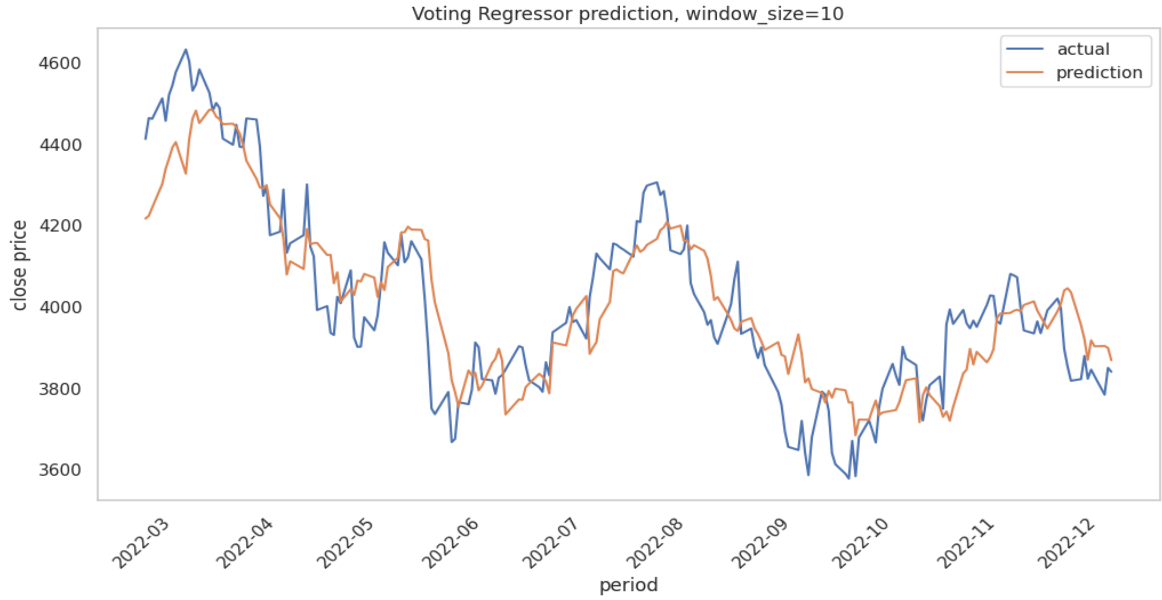
Figure 8: GRU Prediction vs Actual



4.1.4 Voting Regressor

The voting regressor model started with a wide gap at the beginning but seemed to have closely predicted the prices except for a few instances of significant divergence between both points converging towards the end.

Figure 9: Voting regressor Prediction vs Actual



The performance of the four models is summarised in the table below, Regular LSTM had the best performance across all indices followed by the PCA-Stacked LSTM model while voting regressor had the least, this is not obvious in the plots hence the need to provide a table to highlight same.

Table 1: Model Performance metrics

Model	RMSE	MAE	MAPE
PCA-Stacked-LSTM	107.548	89.929	0.022
Regular-Stacked-LSTM	89.402	73.284	0.018
GRU	171.715	133.826	0.034
Voting regressor	113.549	91.207	0.022

Figure 10: All Models Prediction vs Actual

4.2 XAI with SHAP interpretation

In attempting to provide transparency and improve trust, SHAP was used to show how each of the models, and interestingly this showed exciting and profound results because even in the case of the two LSTM variants, the approach was so remarkably different that it might have been otherwise inconceivable without the use of SHAP. On the other end of the spectrum, all 4 models had open price and high price in their top 3 features for predictions.

4.2.1 Regular-Stacked-LSTM

This model prioritized the low price as the most important closely followed by open, high, and middle band in that order while volatility, signal line, percentage change and standard deviation make up the bottom four. Most instances of the top features fall within the positive range, and on the other hand the bottom four looked largely undecided and clustered somewhat around the mean. Furthermore, the high feature values for MACD and percentage lie on the positive side of the of the model while this is reverse for volatility and signal line.

Figure 11: SHAP plot of Regular Stacked LSTM

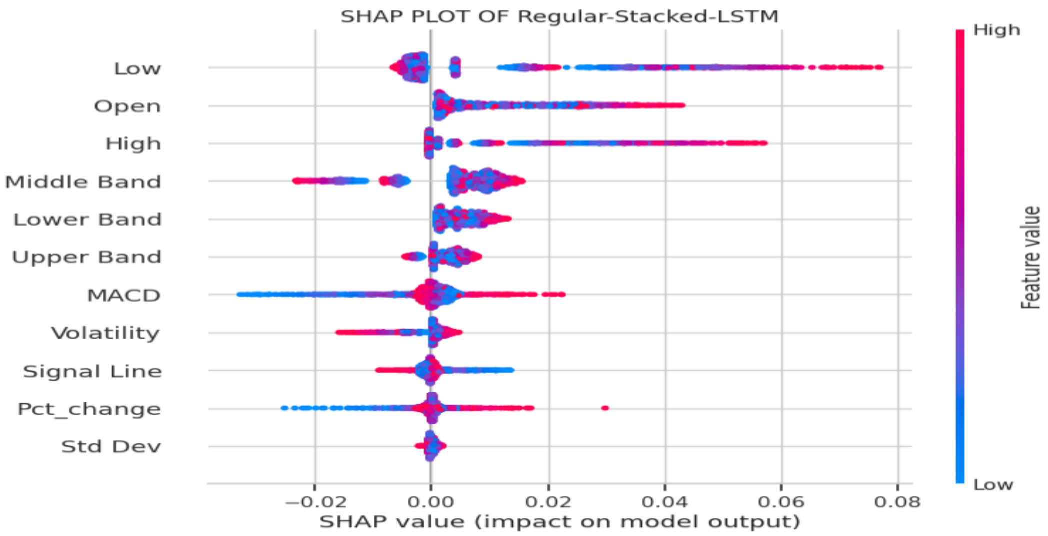
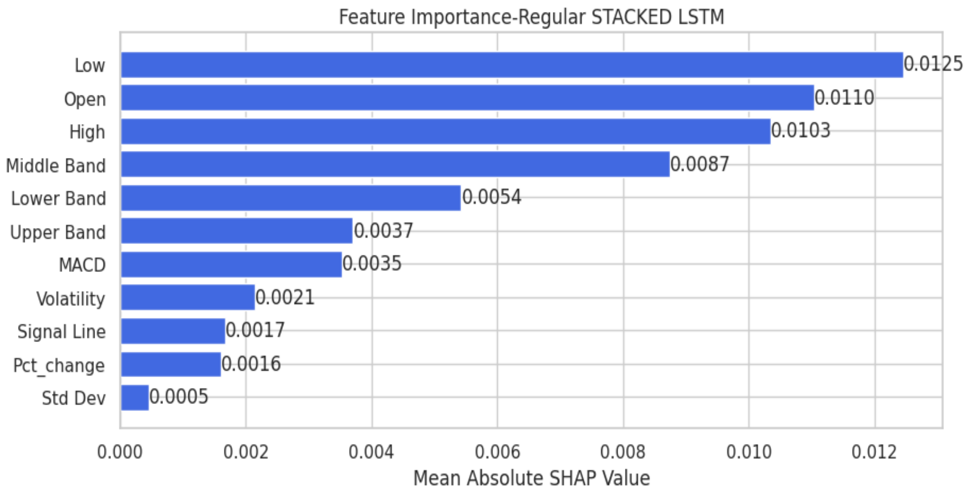


Figure 12: Feature Importance plot of Regular Stacked LSTM



4.2.2 PCA-Stacked-LSTM

With the PCA model, the four most importance features are the open price, high price, percentage change and signal line, the bottom four are low price, middle band, upper band, and lower band. This is completely different compared to the regular LSTM model not just in the features prioritized, but the magnitude of the weights difference between the first and second features, this model placed a weight on open price which is roughly 18 times that for the high price while almost completely disregarding the last 7 features. While all the instances for the open price lie in the positive side of graph, the case different for the high price, percentage change, signal line and volatility where the high feature value lie on the negative side and the low features on the positive side respectively.

Figure 13: SHAP plot of PCA-Stacked LSTM

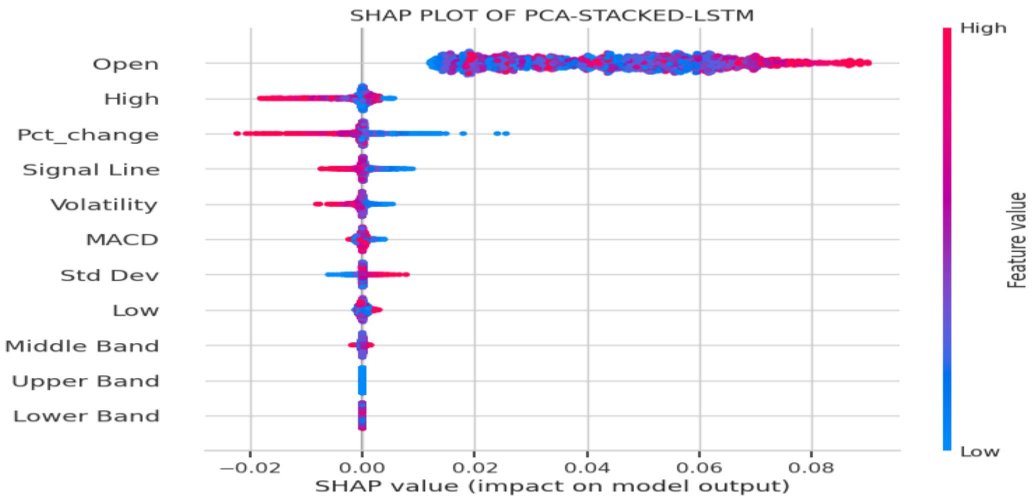
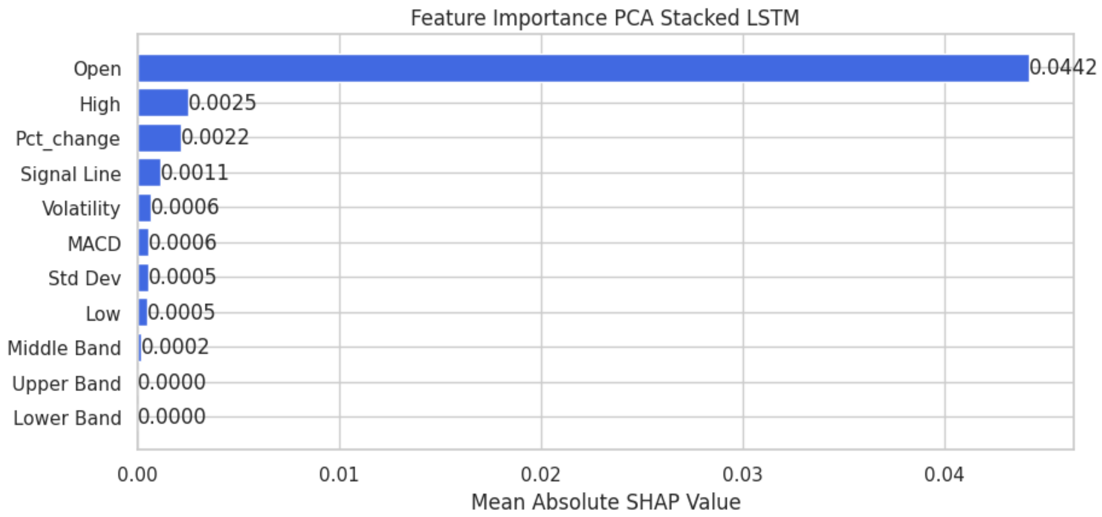


Figure 14: Feature importance plot of PCA-Stacked LSTM



4.2.3 GRU

The GRU model has low price, high price, open price, and lower band as its top four important features, it is however expedient to point out that it is the only model where all the instances of the top five features lie completely on the positive side of the graph. Like the regular LSTM, the magnitude of its top four features is very close. The signal line has its high feature values predominantly on the negative side and vice versa for MACD and percentage change respectively.

Figure 15: SHAP plot of GRU

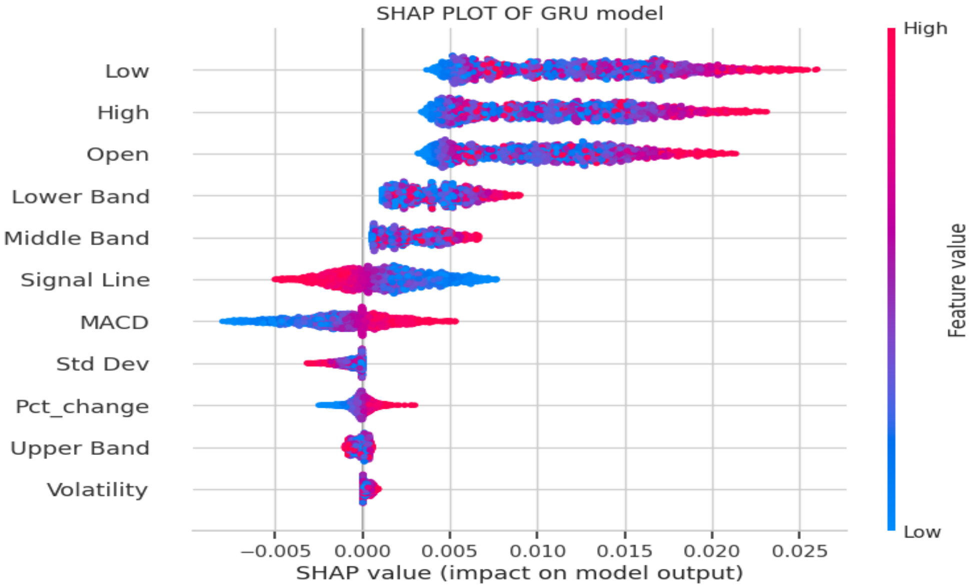
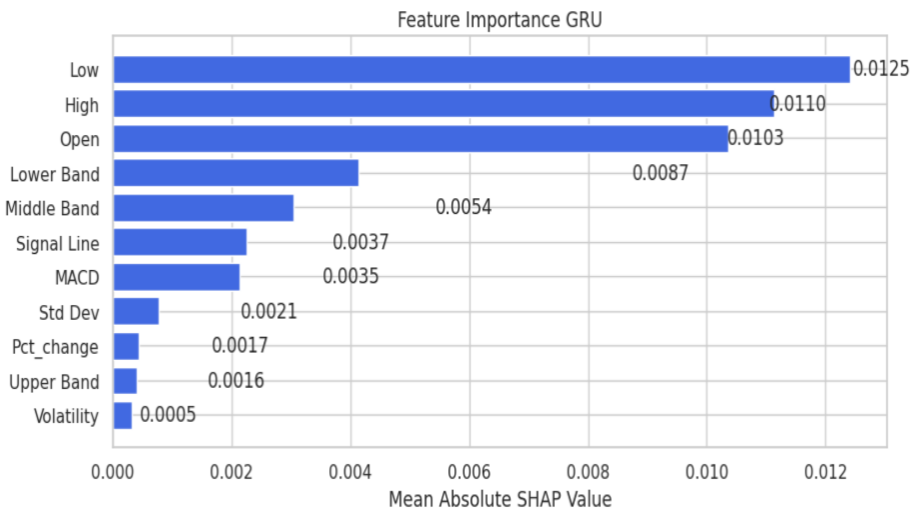


Figure 16: Feature Importance plot of GRU



4.2.4 Voting regressor

The voting regressor ensemble prioritized the open price, low price, percentage change and high price as its top four features while the bottom four are standard deviation, signal line, volatility, and lower band. It is important to highlight the fact that very similar to the PCA-LSTM model, the magnitude of the open price is so large compared to the second feature, it is roughly 19 times. Given that both models had undergone dimensionality reduction via PCA, it can be argued that this common characteristic is inherited from that operation. In a similar manner, the rest of the features are also clustered around the mean again consistent with the PCA LSTM model. But this is where the commonality ends as the ordinal arrangement of feature importance is different once you ignore the first feature which is the same in both models.

Figure 17: SHAP plot of Voting regressor

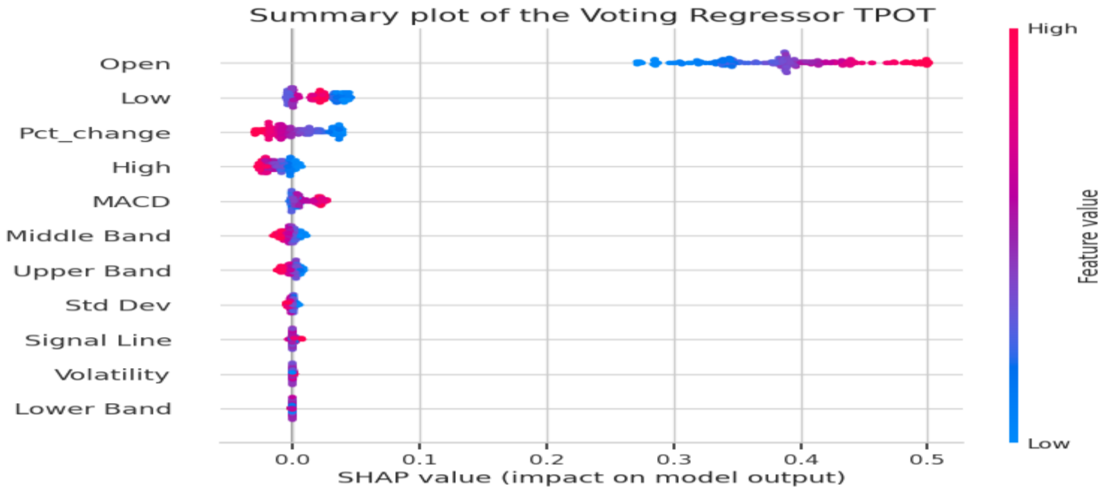
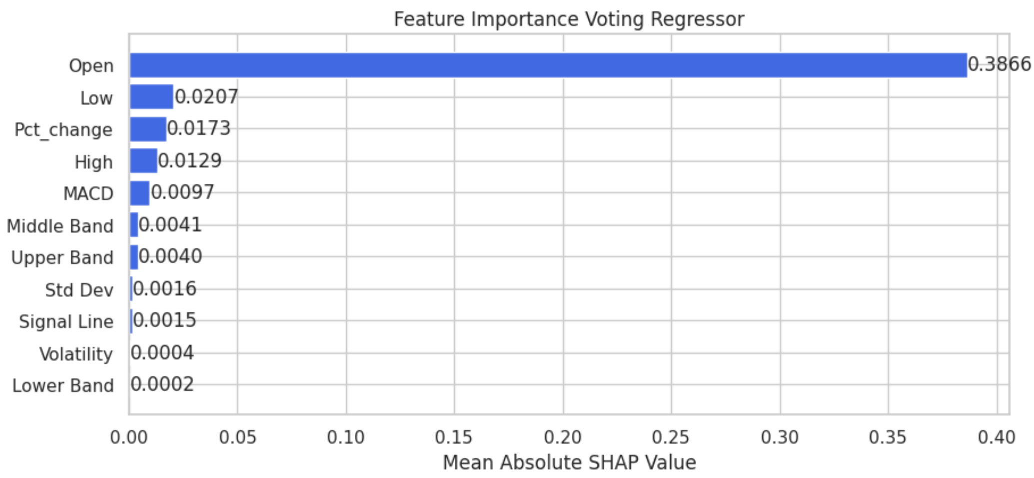


Figure 18: Feature importance plot for Voting regressor



4.2.5 KOSPI data results

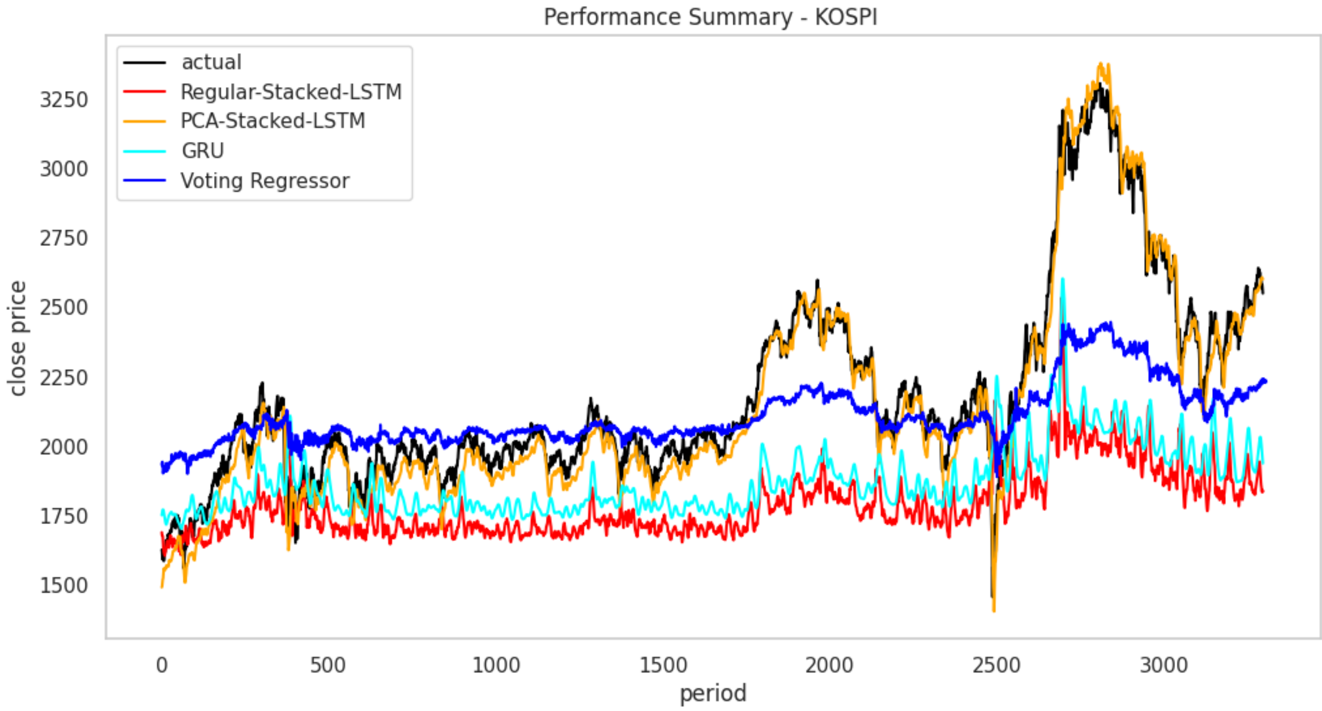
To see how the model performs with unseen data, we extracted data from a frontier market index South Korea (KOSPI) was chosen for this exercise and the period was from January 1, 2010, to June 30, 2023. The result below is very telling, whereas the training performance metrics showed the regular stacked LSTM model to be the best and GRU as the least, testing in real life revealed a completely different story, with PCA-LSTM being the most superior by a mile with the voting regressor ensemble as second while the regular stacked LSTM as the worst.

It can be inferred from this result that whilst in training mode, there was a level of overfitting by the regular stacked LSTM which could have led to the woeful performance with unseen data, furthermore, the surprise performance by the voting regressor can be attributed to the inherent advantages of the different component models which perhaps allowed for robustness in adapting better than the rest of the neural networks.

Table 2: Performance measure for KOSPI data

Model	RMSE	MAE	MAPE
PCA-Stacked-LSTM	71.250	59.679	0.028
Regular-Stacked-LSTM	480.291	402.935	0.172
GRU	415.982	329.096	0.139
Voting regressor	263.474	180.051	0.075

Figure 19: Performance plot for KOSPI



5.0 Recommendations

The primary objective of this study was to address a significant challenge in utilizing artificial intelligence for predictive purposes, particularly in critical real-world scenarios - the lack of transparency and interpretability of the results. This is often due to the complex nature of artificial neural networks, which are treated as "black boxes" making it challenging to understand the decision-making process within these models.

To demonstrate the potential of overcoming this challenge, four models were developed using the S&P 500 dataset obtained from Yahoo Finance. The dataset covered the time range from January 1, 2012, to December 31, 2022. The data was enriched by engineering eight technical indicators commonly used in stock analysis: Percentage Change, MACD, Signal Line, Volatility, Middle Band, Std Dev, Upper Band, and Lower Band.

Three models belonged to the recurrent neural network family, with two variants of LSTM models (one vanilla and one fed with PCA-transformed input) and a Gated Recurrent Unit (GRU) model. The fourth model was a voting regressor ensemble that utilized input from four different algorithms: XGBoost, LGBM, SVM, and Gaussian Process Regressor. The RNN variants employed a sliding window approach with a look-back period of 10 days. The voting regressor used PCA-transformed input without any look-back period. Due to the time series nature of the data, a time series split was employed to divide the dataset into training and test sets. The pre-processed input had a total of 11 features with the open, low and high price added to the 8 technical indicators.

All models were scaled using Min-Max scaling from scikit-learn, and the SHAP (SHapley Additive exPlanations) algorithm was applied to interpret the importance of features and their contributions to predictions. The model was then extended to a frontier market using the South Korean KOSPI index to see how well it generalised with unseen data, the outcome was very instructive as the regularly LSTM which performed best during training became the least with PCA-LSTM and Voting regressor occupying first and second positions respectively.

The application of eXplainable AI (XAI) was particularly impactful for the LSTM variants. SHAP analysis revealed starkly contrasting decision mechanisms in these models, which might not have been apparent without the SHAP algorithm's insights.

This study marks a significant step forward in encouraging the adoption of AI-driven solutions. By addressing transparency and interpretability concerns, it fosters trust among stakeholders and paves the way for collaborative efforts between technology experts and domain professionals. This synergy is expected to drive further innovation and possibilities in the AI field.

One intriguing outcome is the relatively low emphasis placed by the models on technical indicators for predictions, as evidenced by SHAP interpretations. This opens avenues for future research. Additionally, exploring sentiment analysis integration into the model could offer insights into how decision mechanisms react to sentiment-based inputs, potentially altering the prioritization of existing features such as technical indicators. This direction could uncover interesting insights into the interplay between quantitative and sentiment-based factors in prediction models.

References

1. Published by Statista Research Department and 6, J. (2023) Domestic equity market capitalization globally 2023, Statista. Available at: https://www.statista.com/statistics/274490/global-value-of-share-holdings-since-2000/ (Accessed: 21 June 2023).
2. Mueller, R.C. (2019) in The Venetian money market: Banks, panics, and the public debt, 1200-1500. Baltimore, MD: Johns Hopkins University Press, pp. 288–406.
3. Smith, B.M. and Smith, B.M. (2004) in A history of the global stock market: From ancient rome to Silicon Valley. Chicago: University of Chicago Press, pp. 12–45.
4. Beattie, A. (2023) The birth of stock exchanges, Investopedia. Available at: https://www.investopedia.com/articles/07/stock-exchange-history.asp (Accessed: 21June 2023).
5. Williams, W. (2023) Timeline of U.S. Stock Market crashes, Investopedia. Available at: https://www.investopedia.com/timeline-of-stock-market-crashes-5217820 (Accessed: 21 June 2023).
6. Srijiranon, K., Lertratanakham, Y. and Tanantong, T., (2022). A Hybrid Framework Using PCA, EMD and LSTM Methods for Stock Market Price Prediction with Sentiment Analysis. Applied Sciences, 12(21), p.10823.
7. Fama, E.F., (1965). The behavior of stock-market prices. *The journal of Business*, *38*(1), pp.34-105.
8. Greenwich Associates. (2019). The Future of Trading, refinitive. pp.1-12
9. Malkiel, B.G., (2003). The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1), pp.59-82.
10. Jiang, W. (2021) 'Applications of deep learning in stock market prediction: Recent progress', Expert Systems with Applications, 184, p. 115537. d
11. Gray, W.R. and Carlisle, T.E. (2013). Quantitative value: a practitioner's guide to automating intelligent investment and eliminating behavioural errors. Hoboken, N.J.: Wiley ; Chichester.
12. Amilon, H., (2003). A neural network versus Black–Scholes: a comparison of pricing and hedging performances. Journal of Forecasting, 22(4), pp.317-335.
13. Anwar, M.N. and Andallah, L.S., (2018). A study on numerical solution of Black-Scholes model. Journal of Mathematical Finance, 8(2), pp.372-381.
14. Srivastava, A. and Shastri, M., (2020). A study of Black–Scholes Model's applicability in Indian capital markets. *Paradigm*, *24*(1), pp.73-92.
15. McKenzie, S., Gerace, D. and Subedar, Z., (2007). An empirical investigation of the Black-Scholes model: evidence from the Australian Stock Exchange. *Australasian Accounting, Business and Finance Journal*, *1*(4), pp.71-82.
16. Reddy, V.K.S., (2018). Stock market prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 5(10), pp.1033-1035.
17. Gururaj, V., Shriya, V.R. and Ashwini, K., (2019). Stock market prediction using linear regression and support vector machines. *Int J Appl Eng Res*, *14*(8), pp.1931-1934.
18. Dingli, A. and Fournier, K.S., (2017). Financial time series forecasting-a deep learning approach. International Journal of Machine Learning and Computing, 7(5), pp.118-122.
19. Henrique, B.M., Sobreiro, V.A. and Kimura, H., (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of finance and data science*, *4*(3), pp.183-201.
20. Chowdhury, R., Mahdy, M.R.C., Alam, T.N., Al Quaderi, G.D. and Rahman, M.A., (2020). Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model. *Physica A: Statistical Mechanics and its Applications*, *555*, p.124444.
21. Shrivastav, L.K. and Kumar, R., (2022). An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction. *Journal of Information Technology Research (JITR)*, *15*(1), pp.1-19.
22. Hu, Z., Zhao, Y. and Khushi, M., (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, *4*(1), p.9.

23. Jin, Z., Yang, Y. and Liu, Y., (2020). Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 32, pp.9713-9729.
24. Adadi, A. and Berrada, M., (2020). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). IEEE Access, 6: 52138–52160, 2018. based on change detection with interpolation kernels. GeoInformatica, pp.1-21
25. Gite, S., Khatavkar, H., Srivastava, S., Maheshwari, P. and Pandey, N., (2021). Stock prices prediction from financial news articles using LSTM and XAI. In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020 (pp. 153-161). Springer Singapore.
26. Ribeiro, M.T., Singh, S. and Guestrin, C., (2016), August. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
27. Nieto Juscafresa, A., (2022). An introduction to explainable artificial intelligence with LIME and SHAP.
28. Lundberg, S.M. and Lee, S.I., (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
29. Çelik, T.B., İcan, Ö. and Bulut, E., (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. *Applied Soft Computing*, *132*, p.109876.
30. Kumar, S., Vishal, M. and Ravi, V., (2022). Explainable Reinforcement Learning on Financial Stock Trading using SHAP. *arXiv preprint arXiv:2208.08790*.
31. Bandi, H., Joshi, S., Bhagat, S. and Ambawade, D., (2021), June. Integrated Technical and Sentiment Analysis Tool for Market Index Movement Prediction, comprehensible using XAI. In 2021 International Conference on Communication information and Computing Technology (ICCICT) (pp. 1-8). IEEE.
32. Obthong, M. et al. (2020) 'A survey on machine learning for stock price prediction: Algorithms and techniques', Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business.
33. Khare, K. et al. (2017) 'Short term stock price prediction using Deep Learning', 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)
34. Sen, J. and Mehtab, S. (2021) A robust predictive model for stock price prediction using Deep Learning and Natural Language
35. Guo, Y. (2020) 'Stock price prediction based on LSTM neural network: The effectiveness of news sentiment analysis', 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)
36. Nabipour, M. et al. (2020) Deep learning for stock market prediction
37. Samek, W. and Müller, K.-R. (2019) 'Towards explainable artificial intelligence', Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, pp. 5–22.
38. Nikou, M., Mansourfar, G. and Bagherzadeh, J. (2019) 'Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms', Intelligent Systems in Accounting,Finance and Management, 26(4), pp. 164–174.
39. Ogbujah, I.L. (2023) *GitHub*. Available at: https://github.com/iykeogbujah/My-first-repo/blob/main/United_Whole_Model_(1)%20(2).ipynb (Accessed: 31 August 2023).