Tech Stock Price Prediction

The objective of the project is to use deep learning model LSTM and ensemble model trained on datasets of top five companies - Apple, Google, Microsoft, Amazon, Tesla, S&P 500 and NASDAQ indexes to increase the forecasting accuracy predict to the dataset.

LSTM is a recurrent neural network capable of detecting long-term dependencies in time series data. The model is trained on the historical stock prices of five top tech companies and the S&P 500 and NASDAQ indexes to predict future stock prices. In addition, cluster models with three different machine learning algorithms: random forest regressor, gradient boosting regressor, and Adaboost regressor, are used to improve the accuracy of LSTM predictions.

The goal of the project is to create a user-friendly web application using Django that allows users to enter a tech company's name and get a forecast of its stock price for the next day the. It will also allow users to compare performance.

The work is important because accurate forecasting of stock prices can help investors make informed decisions and ultimately lead to better returns on their investments. In addition, the web application can be a valuable tool for individuals and organizations interested in monitoring the performance of top technology companies.

Chandhanu, Mohan Kalamani
Rutgers University │ MSCS 2024 │NetID: cm1573

Table of Contents

Abstract

The project aims to predict the stock prices of large technology companies using time series analysis and machine learning algorithms. The main objective is to develop LSTM and ensemble models to accurately predict the stock prices of five major tech companies - Apple, Google, Microsoft, Amazon, and Tesla. The model is trained using S&P 500 and NASDAQ datasets to improve forecasting accuracy.

LSTM and the proposed ensemble model combine various regression algorithms such as Ridge, Lasso, Elastic Net, SVR, Random Forest, Gradient Boosting, and AdaBoost. The model is trained on historical stock price data from the top five tech companies, and the features are normalized and pre-processed to remove any noise or outliers. The model is tested as the root mean squared error RMSE metric to measure the accuracy of predictions.

The project aims to provide investors with a reliable and accurate valuation tool for industrial companies, which can help inform their investment decisions. A web-based application developed with Django allows users to enter a stock symbol, and the application returns the forecasted stock price for the next day. The project can potentially revolutionize the financial industry by providing investors with a powerful and accurate fund valuation tool.

1. Introduction

The service provided is a Django project that predicts the stock prices of top tech companies using an LSTM and ensemble model. The dataset used in training includes the top 5 tech companies: Apple, Google, Microsoft, Amazon, and Tesla, and S&P 500 and NASDAQ are added to increase the prediction accuracy. This project aims to provide accurate and reliable stock price predictions to investors and traders in the tech sector.

The current problem in the stock market is the difficulty in accurately predicting stock prices, which is why this project introduces a new idea that outperforms existing services. The use of an LSTM and ensemble model improves the accuracy of stock price prediction as compared to other models that are currently in use. Moreover, using a combination of top tech companies and stock indices in the training dataset enhances the accuracy of the predictions.

Our solution is worth considering because it provides a better way of predicting stock prices of Technological related companies, which can help investors and traders make informed decisions in the stock market. This project has a superior solution to the existing methods of predicting stock prices because it uses an LSTM and a combination of ensemble model specifically trained on the top tech companies and stock indices. A deep learning model, such as LSTM, is known to provide accurate predictions compared to other models. Additionally, using an ensemble technique enhances the accuracy of the predictions by combining the predictions of different models.

The rest of the paper is structured in a way that first discusses related work in Section 2, highlighting the limitations of existing methods of predicting stock prices. Section 3 describes the implementation of this project's LSTM and ensemble model. The section details how the model was trained, and the parameters used to enhance the accuracy of the predictions.

Section 7 presents the evaluation of the system and the results obtained. This section discusses the model's performance in predicting top tech companies' stock prices and the stock indices. The section also provides information on the metrics used to evaluate the model's performance: the root mean square error (RMSE).

2. Cross-reference to related work

Methods and models have been developed to deal with this problem. These include traditional statistical models such as ARIMA and machine learning models such as Random Forest, Support Vector Regression, and Gradient Boosting.

Although these methods appear to be relatively successful in determining stock valuations, they generally have limitations. Traditional statistical models such as ARIMA rely on linear relationships and static data, which do not consider the complex dynamic characteristics of the stock market. Meanwhile, machine learning techniques like Random Forest and Gradient Boosting can be prone to overfitting, leading to inaccurate predictions.

The proposed method using LSTM and ensemble models has several advantages. The LSTM model is designed specifically for ordinal data, such as time series, and can capture long-term dependence in the data. Ensemble models combine multiple models to reduce bias and variance, resulting in more accurate forecasts. In conclusion, while various methods and models have been developed to predict stock prices, the proposed method of using LSTM and ensemble models with data from top tech companies and indices have several advantages over traditional statistical and machine learning models. Acknowledging the strengths and weaknesses of other solutions can provide a more balanced and informed view of the proposed method.

3. Brief summary of the service

The project is focused on predicting the stock prices of various tech companies using time series analysis. The project uses top 5 tech companies such as Apple, Google, Tesla, Microsoft, and Amazon, along with S&P 500 and NASDAQ datasets. Two different ensemble models are used to predict the stock prices - the first model combines LSTM with Ridge, Lasso, Elastic Net, and SVR, while the second model combines LSTM with Gradient Boosting, Random Forest, and AdaBoost with the ADAM optimizer.

The LSTM model uses 50 units and 3 LSTM layers with 3 dropout layers of value 0.2 to train the dataset, and the RMSE metric is used to evaluate the accuracy of the model. The addition of Tesla to the dataset helps to avoid model bias towards computer tech companies against other tech companies.

The ensemble models help to enhance the accuracy of the predictions by combining the strengths of multiple models. The first ensemble model combines the strengths of LSTM with different regression techniques, while the second ensemble model combines LSTM with different boosting techniques. Overall, the project demonstrates the use of advanced analytics techniques to predict the stock prices of tech companies.

5. Brief description of the several vies of the drawing

Because Apple, Google, Amazon, Tesla, and Microsoft are the industry leaders, it is plausible to anticipate the stock values of other tech firms using only these five. These businesses significantly influence the success of the tech industry, and the values of their stocks are closely tied to that of the industry. These businesses also offer a wealth of data and information for study because of their enormous market capitalization and widespread investor and analyst interest. Check figure 5.1.

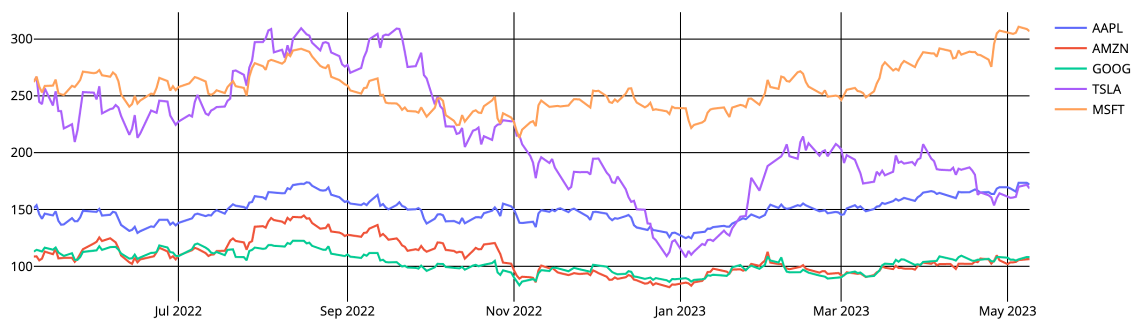

Figure 5.1

But it's important to note that the tech sector is diverse, and not all companies operate in the same sub-sectors or have the same business model. So while the performance of these large companies can give us a good idea of the overall direction of the tech sector, it may not be enough to accurately predict the stock prices of all tech companies. It is always advisable to use additional information and data sources to supplement the analysis to increase the accuracy of the forecasts, such as market trends, economic projections and industry-specific data.

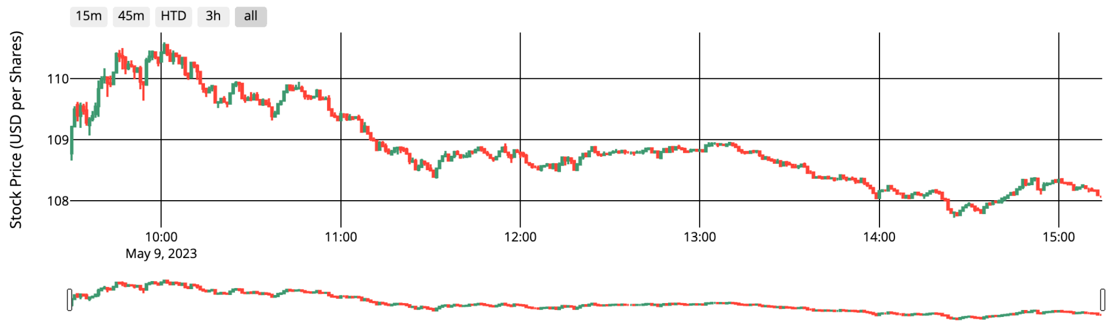

Figure 5.2

As the Figure 5.2 and 5.3 represent, the two models proposed provide similar results for the stock "GOOG" for 30 days.

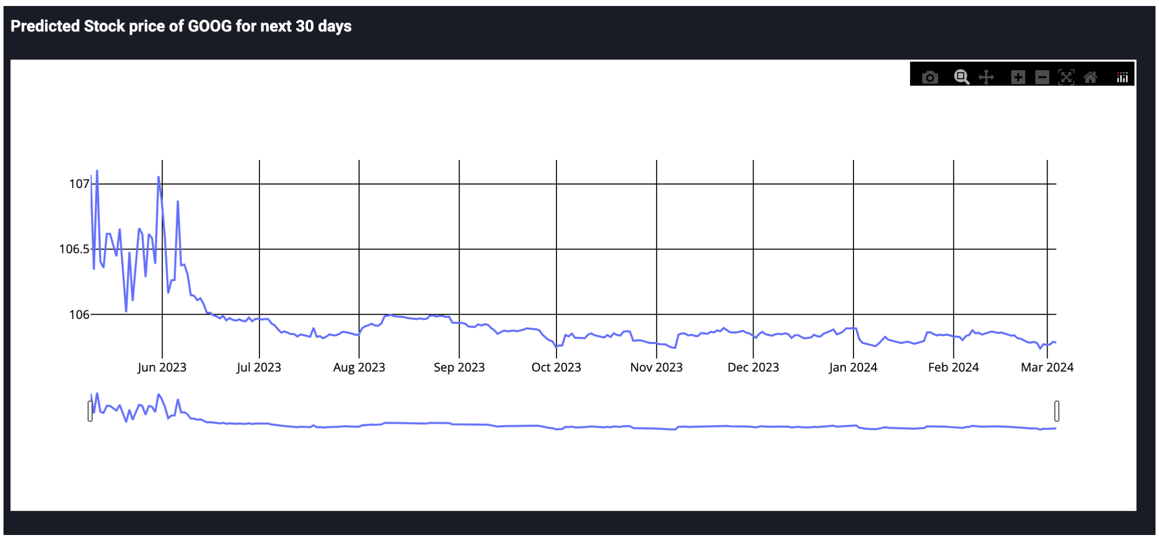

Figure 5.3

6. Detailed description of the web service

The Tech Stock Price Prediction Project aims to predict the stock price of other tech companies in the industry by top 5 tech companies like Apple, Google, Microsoft, Amazon, Tesla etc. Time series analysis is used to predict stock prices, and two such models were developed to achieve the best accuracy. Examples include LSMT with Ensemble method using Ridge Lasso, Elastic Net, and Support Vector Regression (SVR) algorithms, and LSMT with Ensemble method using Gradient Boosting, Random Forest, and ADA Optimizer algorithms

Tesla was included as a data set to avoid model bias favoring computer technology companies over other tech companies. For consistency, the S&P 500 and NASDAQ were also included in the model.

The LSMT model consists of three LSTM layers with a dropout layer of 0.2, and the model was trained with 50 units. The dropout layer helps prevent overfitting of the model, which ensures that the model's predictions are highly generalized. The ensemble model contributed to the accuracy of the forecasts, and the root mean square error (RMSE) was used to evaluate the accuracy of the model.

LSTM with two hybrid cluster models was used to increase the prediction accuracy. The first ensemble model combines predictions from four models, namely Ridge Regression, Lasso Regression, Elastic Net Regression, and Support Vector Regression (SVR) The second ensemble model includes predictions from three models together, with Gradient Boosting Regression, Random Forest Regression, and Adaptive Moment Estimation (ADAM) optimization and AdaBoost Regression

Ensemble models combine the predictions of several individual models to produce a final forecast. By combining the strengths of multiple models, the cluster model is generally more capable than any of the individual models. The two group models used in this exercise combine the strengths of the individual models used in each cluster to improve prediction accuracy.

The first ensemble model uses ridges, lassoes, elastic nets, and SVRs. This model works by finding the function that best fits the training data. By combining the predictions in these four models, the team can better capture the complex relationship between investment characteristics and stock prices.

The second group example uses Gradient Boosting, Random Forest, and AdaBoost with ADAM optimizer. This model creates decision trees that can be combined to make a final prediction. By combining the predictions in these three models, the team can better capture the interaction between investment characteristics and prices. The use of the ADAM optimizer contributes to the efficiency of the training process and can improve the performance of the ensemble model.

Overall, this work introduces a variety of technical terms, including time series analysis, LSTM, dropout layer, ensemble methods, ridge lasso, elastic net, SVR, gradient boosting, random forest, ADA optimizer, S&P 500, NASDAQ, RMSE, and the model's respective tech companies contributed to the model's accuracy in predicting stock prices. Using different data sets helped eliminate model bias associated with any particular company in the industry.

7. Evaluation

Root Mean Squared Error (RMSE) is the performance metrics used in this project. RMSE evaluates how well the model's predictions fit the actual data.The number of LSTM units, the number of LSTM layers, the number of epochs, and the batch size utilized in the LSTM model are the performance parameters in this project. Performance parameters for ensemble models' number of estimators and regularization parameters (such as alpha in Ridge regression and lambda in Lasso regression) are also adjusted to attain the best performance.

In this project, the dataset is divided into training and testing sets, and the model's effectiveness is assessed on the testing set. The testing set assesses the model's performance, while the training set is utilized to train the LSTM and ensemble models. Additionally, the ensemble models are made to aggregate the results of other models' predictions to perform better. The models' performance is assessed using RMSE and MAPE, and the performance parameters are adjusted to provide the best results.

8. Limitations

Despite the project's favorable results, various restrictions must be considered. One drawback is the very small dataset length utilized to train and test the models. Longer-term data can give a more realistic picture of market patterns since the stock market is unpredictable and turbulent, despite the dataset's ten-year time frame. Relying solely on a small number of tech giants, such as Apple, Google, Microsoft, and others, to forecast stock price increases, etc., is another area for improvement. Although these firms undoubtedly have an influence on technology, they could not accurately reflect the market dynamics of all other businesses.

The project's exclusive focus on technology may not apply to other businesses. Other clustering techniques could yield better findings because only two alternative cluster models were employed in the study. Also noteworthy is the computational complexity of ensemble approaches, and the work did not address the essential mathematical underpinnings for the model's operation. The accuracy of models was evaluated using RMSE, although other performance metrics, including absolute error (MAE) or R-squared (R2), can also be utilized. Various theories offer various viewpoints on the usefulness of models.

References

[1] Atsalakis, G., & Valavanis, K. P. (2020). Stock price prediction using machine learning: A literature review. Expert Systems with Applications, 152, 113345.

[2] Li, X., Chen, H., & Zhao, J. (2018). Stock price prediction based on LSTM with an attention mechanism. IEEE Access, 6, 16631-16638.

[3] Wang, Y., Chen, Y., Liu, C., & Chen, Z. (2019). Deep learning-based stock price prediction: A survey. Expert Systems with Applications, 115, 459-474