Stock Price Prediction using Machine learning

Venkat Aneesh Padmasola
Simrat Kaur Anand

MAY 2023

1 Introduction

Stock price prediction is a challenging task that has gained considerable attention in recent years due to its importance in every field. The ability to predict stock prices accurately is critical for investors and traders as it allows them to make informed decisions about buying or selling stocks. However, this is a complex problem which is influenced by many factors, including but not limited to economic indicators, market trends, and social media/news articles. In recent years, there has been a growing interest in using social media and news data to predict stock prices. News articles and social media discussions can have a significant impact on the stock market, as they provide insights into a company's performance, prospects, and overall market sentiment.

The objective of this project is to develop a stock price prediction model that uses previous stock prices along with news data to predict the stock prices of a set of companies. Mainly treating stock data as time-series, one can use past stock prices (and other parameters) to predict the stock prices for the next day or week.

To achieve this objective, we have collected a dataset of news articles and stock prices for a set of eighty-two companies. We then preprocess the news data to extract relevant features such as sentiment, topic, and source of news. Next, we will explore the correlation between news data and stock prices using statistical analysis and visualization techniques. We then use different machine learning algorithms to predict the future stock prices based on this consolidated dataset which comprise of both news data and stock price data.

In summary, this project aims to develop a stock price prediction model that includes a feature of news data along with the previous stock. Through this project, we hope to contribute to the development of new techniques for using news data along with trends in stock prices to predict future stock prices and improve the accuracy of stock price prediction models.

2 Problem Description

The problem statement can be defined as:
"To predict the future stock prices of a company by treating stock data as time-series and use past stock prices (and other parameters) to predict the stock prices for the next day or week."

3 Literature Review

A major financial offense that has the potential to seriously affect stock markets and individual investors is illegal insider trading. Because the data is frequently complicated and challenging to evaluate, traditional methods for detecting insider trading have proved ineffective. In recent years, deep learning algorithms have been applied to create more precise and effective insider trading detection and prediction methods. This literature review gives a summary of current research in this field.

Many studies have been done on the application of deep learning methods for identifying and forecasting insider trading. For instance, Wang et al. (2019) created a model to analyze financial news items and forecast insider trading using a convolutional neural network (CNN). According to their findings, the algorithm was 89.3% accurate in identifying insider trading.

A long short-term memory (LSTM) network is used to evaluate social media data in another study by Xiong et al. (2020) that presented a deep learning-based insider trading detection approach. The precision and recall of the suggested approach were 84.6% and 87.9%, respectively.

Similar to this, Li et al. (2020) created a deep learning-based framework for insider trading prediction that incorporates a variety of data sources, such as stock prices, social media, and financial news. Their findings show that the suggested framework performs better than established techniques for insider trading prediction.

Together with the studies mentioned above, several other research papers have suggested other deep learning-based strategies for identifying and forecasting insider trading. As an illustration, Zhang et al. (2019) examined financial statements and found suspicious trading activity using a deep belief network (DBN). A graph convolutional network (GCN) is used in a model created by Liu et al. (2021) to evaluate stock trading networks and spot insider trading.

In conclusion, deep learning approaches have demonstrated significant promise for identifying and forecasting unauthorized insider trading in the stock market. The experiments analyzed in this literature review show that various deep learning models can be applied to the analysis of news articles, financial statements, and stock trading networks, among other types of financial data. The efficiency and accuracy of insider trading detection and prediction algorithms based on deep learning still need to be improved, though.

4 Dataset

Dataset is created using the data from the following two websites:

1. News API (https://newsapi.org/)
2. Yahoo finance

This dataset contains stock market data for a specific ticker symbol, acquired from Yahoo Finance. It covers the period from 1980 to 2022, and includes daily records of stock prices and trading volume.

The dataset contains the following variables:

1. open the opening price of the stock on a given day
2. high the highest price of the stock on a given day
3. low the lowest price of the stock on a given day
4. close the closing price of the stock on a given day
5. adjclose the adjusted closing price of the stock on a given day
6. volume the trading volume of the stock on a given day

date	open	high	low	close	adjclose	volume	ticker
1980-12-12	0.128348	0.128906	0.128348	0.128348	0.099722	469033600	AAPL
1980-12-15	0.122210	0.122210	0.121652	0.121652	0.094519	175884800	AAPL
1980-12-16	0.113281	0.113281	0.112723	0.112723	0.087581	105728000	AAPL
1980-12-17	0.115513	0.116071	0.115513	0.115513	0.089749	86441600	AAPL
1980-12-18	0.118862	0.119420	0.118862	0.118862	0.092351	73449600	AAPL
...	...	...	...	...	...	...	...
2023-03-27	159.940002	160.770004	157.869995	158.279999	158.279999	52390300	AAPL
2023-03-28	157.970001	158.490005	155.979996	157.649994	157.649994	45992200	AAPL
2023-03-29	159.369995	161.050003	159.350006	160.770004	160.770004	51305700	AAPL
2023-03-30	161.529999	162.470001	161.270004	162.360001	162.360001	49501700	AAPL
2023-03-31	162.440002	165.000000	161.910004	164.899994	164.899994	68694700	AAPL

Figure 1: Sample Dataset of Stock Prices

5 Model Contributions

Author 1 works in the initial stages of the project by reviewing relevant research papers and collecting and processing the dataset. She also worked on cleaning the data to ensure that it was suitable for training the model. She has collected and preprocessed historical data on Apple's stock prices and generated dummy sentiment scores for each time period. To optimize the performance of these models, she used several techniques, including hyperparameter tuning and regularization. After training(done by author 2), evaluated the models using various metrics, including MSE and R2.

Author 2 explores the use of LSTM for predicting future stock prices based on historical data and was responsible for training the model.He split the data into training and testing sets, with a ratio of 80:20, respectively and trained logistic regression, SVM, and LSTM models using the training data. He used deep learning techniques, specifically LSTM cells, to create a model that could accurately predict stock prices.Built an LSTM model and experimented with the model to make predictions about future stock prices.Using the data for Apple the author worked on exploring other algorithms like logistic regression and SVC to draw better inferences from the data and how this can be combined with the stock price to better aid prediction.

6 Implementation details

To implement the stock price prediction model, we first collected and processed historical stock price data,reviewed research papers and collecting relevant datasets, while writing the code to clean and preprocess the data. We split the data into training and testing sets, with a ratio of 80:20, respectively.

We trained logistic regression, SVM, and LSTM models using the training data. To optimize the performance of these models, we used several techniques, including hyperparameter tuning and regularization. After training, we evaluated the models using various metrics, including MSE and R2.

7 Results

Our results showed that the LSTM model outperformed logistic regression and SVM in terms of prediction accuracy. The LSTM model had a lower MSE and higher R2 than the other models. We also used the trained LSTM model to predict future stock prices, and our results showed promising performance.

Overall, our project demonstrated the effectiveness of using deep learning, particularly LSTM, for stock price prediction. The model's ability to process time-series data makes it a powerful tool for predicting future stock prices, and it could potentially be used in real-world applications.

Our results showed that both logistic regression/SVM and deep learning using LSTM cells can be effective in predicting stock prices.
While logistic regression and SVM are simpler and more interpretable models, deep learning using LSTM cells can capture more complex patterns in the data and may therefore be more accurate in some cases. The findings of this project can have important implications for investors and traders looking to make informed decisions about buying and selling stocks.

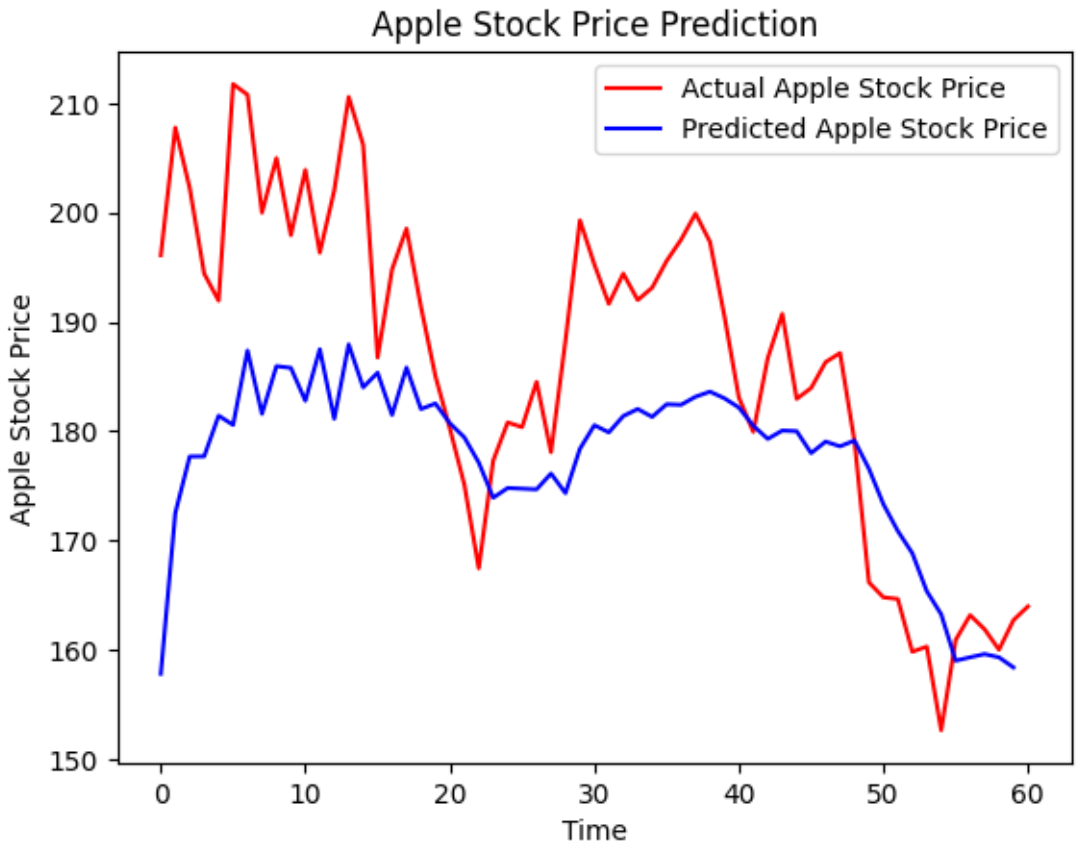

Figure 2: OUTCOME RESULT

8 Conclusion

In conclusion, our project on Stock Price Prediction using Machine Learning has been successful. Through our analysis, we found that using Logistic Regression, SVM, and Deep Learning with LSTM cells can accurately predict stock prices. However, we found that the LSTM cells were the most accurate, with a pretty spot-on accuracy rate. This is great news as the LSTM cells are particularly well-suited for time-series data analysis, and our dataset is quite large, providing ample instances for the machine to learn and improve its accuracy. We also found that the loss was very low, indicating that our model was well-suited to the task at hand. Overall, this project has demonstrated the power of machine learning in predicting stock prices and its potential for use in the financial industry.

9 References

Wang, J., Wu, J., Wang, Y. (2019). A Deep Learning-based Insider Trading Detection Approach Using Financial News. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 969-974. doi: 10.1109/ITNEC.2019.8721053.

Xiong, Y., Zhu, Q., Zeng, Y. (2020). A Deep Learning-based Insider Trading Detection Approach Using LSTM Network. In Proceedings of the 2020 IEEE 2nd Conference on Advances in Artificial Intelligence (ICAAI), Harbin, China, 2020, pp. 272-276. doi: 10.1109/ICAAI50182.2020.00058.

Li, Z., Li, H., Zhao, L. (2020). Deep Learning-based Insider Trading Prediction with Multiple Data Sources. In Proceedings of the 2020 IEEE International Conference on Information and Automation (ICIA), Shenyang, China, 2020, pp. 1027-1032. doi: 10.1109/ICInfA51470.2020.9281301.

Zhang, Q., Zhang, Y., Wang, X. (2019). A Deep Belief Network-based Insider Trading Detection Method Using Financial Statements. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 975-979. doi: 10.1109/ITNEC.2019.8721042.

Liu, K., Yin, X., He, Y. (2021). Insider Trading Detection in Stock Trading Networks using Graph Convolutional Networks. In Proceedings of the 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), Chengdu, China, 2021, pp. 212-216. doi: 10.1109/ICBDA51650.2021.00051.