T.C.

MARMARA UNİVERSITY

FACULTY of ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

Title of the Project

Stock Price Prediction

Group Members

Mustafa Sertaç Öztürk - 150116078

Tunahan Aydın -150114074

Enes Garip -150116034

Table Of Contents

Overview

Abstract

Estimation of stock price behavior is important for several reasons and for different stakeholders in the market.Many studies are trying to put forward theories to explain this phenomenon and more still have tried to use these theories in order to predict future changes in prices. [12]The growing linkages of national markets in currency,commodity and stock with world markets and the existence of common players, have given stock price behavior new property – that of its speedy transmissibility across markets.

Due to the extremely volatile nature of financial markets, it is commonly accepted that stock price prediction is a task full of challenge. However in order to make profits or understand the essence of equity market, numerous market participants[11] or researchers try to forecast stock price using various statistical, econometric or even neural network models. In this work, we find the predictive power of six neural network models,but we use only Lstm networks.

We use stocks data which are contains records about stock price of Google and Tesla.The dataset also contains a date-wise price of stock with open, close, high, and low prices along with volume traded on that day.There are 2194 tesla stock market datas and 253 Google markets datas.

This whole project has been done with ipython on kaggle.

SubTask

1) Import the Libraries.
2) Load the Training Dataset.
3) Use the Open Stock Price Column to Train Your Model.
4) Normalizing the Dataset.
5) Creating X_train and y_train Data Structures.
6) Reshape the Data.
7) Building the Model by Importing the Crucial Libraries and Adding Different Layers to LSTM.
8) Fitting the Model.
9) Extracting the Actual Stock Prices
10) Predicting the Values for Stock Prices.
11) Plotting the Actual and Predicted Prices for Google Stocks

Task Division

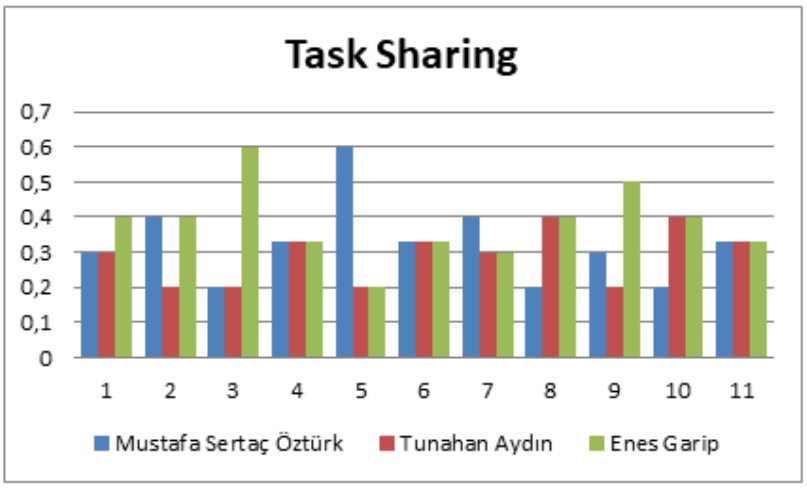


Figure 1

Project Schedule

Week 1 [Nov10-16]: Literature Search and Investigating Approaches

Week 2 [Nov17-23]: Literature Search and Investigating Approaches

Week 3 [Nov24-30]: Midterms

Midterms Report

Week 4 [Dec 01-07]: Exploratory Data Analysis and Data Statistical

Week 5 [Dec 08-14]: Implementing Feature Engineering Techniques

Week 6 [Dec15-21]: Implementing Feature Engineering Techniques

Week 7 [Dec22-28]: Implementing Lstm Networks Algorithm

Week 8 [Dec 29-04]: Designing and Running Several Experiments

Week 9 [Dec 05-11]: Testing and Checking

Week 10[Dec12-18]:Preparing Final Report and Presentation

I. Introduction

The term stock price refers to the current price that a share of stock is trading for on the market. Every public traded company, when its shares are issued, are given a price – an assignment of their value that ideally reflects the value of the company itself. [1]The price of a stock will go up and down in relation to a number of different factors.

Stock prices change everyday by market forces. By this we mean that share prices change because of supply and demand. If more people want to buy a stock (demand) than sell it (supply), then the price moves up. Conversely, if more people wanted to sell a stock than buy it, there would be greater supply than demand, and the price would fall.

Understanding supply and demand is easy. What is difficult to comprehend is what makes people like a particular stock and dislike another stock. This comes down to figuring out what news is positive for a company and what news is negative. There are many answers to this problem and just about any investor you ask has their own ideas and strategies.

We find stocks data which is contains records about stock price of Tesla and Google. [3] The dataset also contains a date-wise price of stock with open, close, high, and low prices along with volume traded as well as turnover on that day.May be we will use another stock dataset with multiple stocks like Apple, Microsoft, Facebook,Tata: Stocks Dataset.

We will use the Long Short-Term Memory(LSTM) method which networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems.

II. Literature Review

The financial market is quite volatile and experiences periods of contraction as well as expansion. The stock market, as a major financial market, is likewise highly volatile. [14]The stock market has the characteristics of high return which has attracted the majority of investors and high risk which puts pressure on investors to sell out at the wrong time. In order to reduce unnecessary losses and obtain higher trading profits, the investors usually except to predict the stock price trend. As a result, stock market forecasting has been a major research topic in the financial area and attracts the attention of investors. In the stock market, the factors affecting the rise and fall of stock prices are complex and diverse. It includes not only the impact of economic factors such as price indicator, circulation indicator, activity degree, and economic uncertainty but also the impact of noneconomic factors such as traders' expectations, traders' psychological factors, and political environment. Therefore, the prediction of stock price has always been a challenging task.

According to the efficient market hypothesis [1], the stock price can be predicted according to the data of historical stocks. Furthermore, in recent years, since the increasing computing power and the decreasing data storage costs, especially the rise and development of innovative technologies such as big data, machine learning, reinforcement learning, and other optimization technologies, researchers have developed various models for predicting stock prices. Machine learning has been widely used in the capital market and plays an indispensable role in predicting future stock prices based on historical data. Traditional stock price forecasting models are mainly linear models, including autoregressive integrated moving average (ARIMA) model [2], multiple linear regression model, and exponential smoothing model [3, 4]. However, those (autoregressive integrated moving average,multiple linear regression model, and exponential smoothing model) linear models play an important role in promoting the progress and development of stock forecasting. Stock prices are typically noisy, fluctuating, and nonparametric, resulting in nonlinear and nonstationary characteristics in the stock market. The standard linear prediction model is unable to produce reliable stock predictions. With the development of deep learning methods, nonlinear neural networks are increasingly employed to predict the stock price for their higher accuracy.

Figure 2

The artificial neural network (ANN) includes MP neural network and back propagation (BP) neural network. However, the structure of ANN model is too single and there are some problems: (1) over fitting leads to the weak ability of the model generalization, (2) local extremal leads to the decline of the prediction ability of the model, and (3) the gradient disappears or explodes due to the excessive weight of neurons in the optimization process, resulting in the failure of prediction. Therefore, relevant scholars introduce deep neural networks (DNN), including convolutional neural network (CNN), recurrent neural network (RNN), long-term and short-term memory neural network (LSTM), and gated recurrent neural network (GRU), to improve the problems existing in the ANN model, so as to improve the accuracy and efficiency of prediction.

CNN is a type of neural network that has been increasingly popular in recent years. A one-dimensional CNN is a neural network that is designed to analyse image data efficiently. CNN can read and automatically extract the most significant features from the original input data for learning. This method feeds the network observed time series value as input and uses

a multilayer network to predict the unobserved value. For example, Xu et al. [5] employed CNN to extract important stock features from stock market returns for forecasting stock market trends. Recurrent neural networks (RNN) such as long-term and short-term memory neural networks (LSTM) are another tool for predicting time series [6, 7]. LSTM accurately estimates time series data by using both the historical and the present stock data. In recent years, LSTM has been applied to stock market forecasting in different stock markets around the world. Chen et al. [8] used an LSTM model to predict China's Shanghai and Shenzhen stock markets. Li et al. [9] introduced the stock indicator with investor sentiment based on the LSTM model to predict the CS1300 index value, and the research results showed that the model was better than the support vector machine method in prediction accuracy. However, this model does not reduce the dimension of stock indicator. Jiawei and Murata [10] attempted to identify the influencing factors of stock market trend prediction through the LSTM model, which used a preprocessing algorithm to reduce the dimension of stock features and a sentiment analyzer to present financial news for stock trend prediction. However, only one dimension reduction method is used, and there is no comparison with other methods

III. METHODOLOGY

A. LSTM Networks

[15]Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people in following work.[1] They work tremendously well on a large variety of problems, and are now widely used**.**

Predicting the future is was once a thing of speculation and mystery. Thanks to human advancements, it has become a task only limited by the amount and depth of data.

And as we live in a society that continuously generates data at an exponential rate, this task of foresight is becoming more accessible.

The further you look into data driven predictions, the term LSTM is sure to rear it confusing head. As with many tech concepts, it is an acronym and it stands for Long Short Term Memory.

Simply stated, it is a Neural Network — a system of machine learning meant to emulate human learning patterns — that is able to "remember" previous data and conclusions, and use that to more accurately come to a final conclusion.

"... LSTM holds promise for any sequential processing task in which we suspect that a hierarchical decomposition may exist, but do not know in advance what this decomposition is."

— Felix A. Gers, et al., Learning to Forget: Continual Prediction with LSTM, 2000

LSTM is a type of Recurrent Neural Network in Deep Learning that has been specifically developed for the use of handling sequential prediction problems.

B.Lstm Formula And Cell

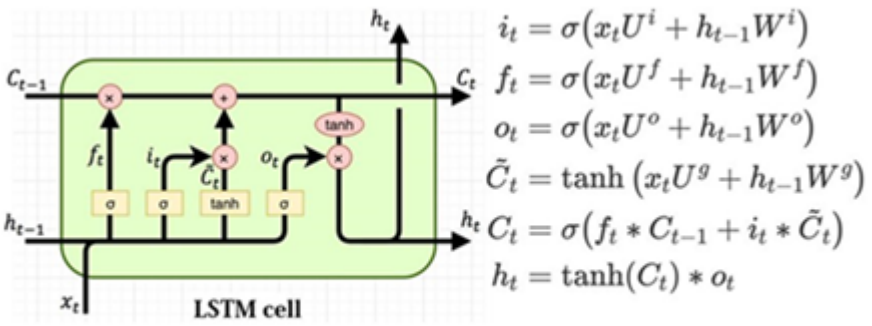


Figure 3

LSTMs work in a three-step process.

- The first step in LSTM is to decide which information to be omitted from the cell in that particular time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function.

- There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.

- The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate.

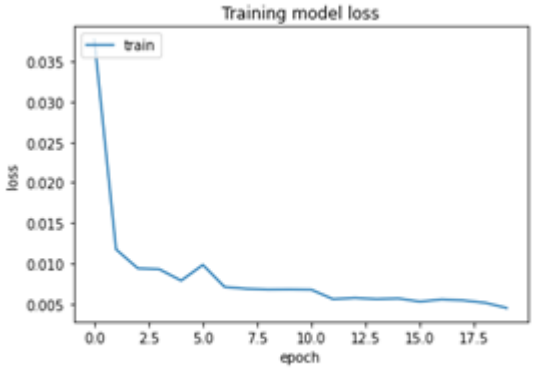


Figure 4

C. Dataset Information

We use stocks data which are contains records about stock price of Google and Tesla.The dataset also contains a date-wise price of stock with open, close, high, and low prices along with volume traded on that day.There are 2194 tesla stock market datas and 253 Google markets datas.

Figure 5

IV. REVISED PROJECT PLAN

Week 1 [Nov10-16]: Literature Search and Investigating Approaches

Week 2 [Nov17-23]: Literature Search and Investigating Approaches

Week 3 [Nov24-30]: Midterms

Midterms Report

Week 4 [Dec 01-07]: Exploratory Data Analysis and Data Statistical

Week 5 [Dec 08-14]: Implementing Feature Engineering Techniques

Week 6 [Dec15-21]: Implementing Feature Engineering Techniques

Week 7 [Dec22-28]: Implementing Lstm Networks Algorithm

Week 8 [Dec 29-04]: Designing and Running Several Experiments

Week 9 [Dec 05-11]: Testing and Checking

Week 10[Dec12-18]:Preparing Final Report and Presentation

V. Project Accomplishment:

In

Figure 6

this project, we implemented the lstm algorithm. We could do an analysis of how lstm works.We trained more than 2000 data of Tesla and 253 data of Google and predicted the future data.
We classified and normalized these data according to time.
We took a sample of a dataset to make stock price predictions using the LSTM model and visualized the predicted stock costs with actual stock costs.

```
In [20]: accuracy = (np.mean(predicted_price)/np.mean(y_test))*100

In [21]: print(accuracy)
         95.21829663482048

In [ ]:
```

When we calculate the mean of the actual stock prices and predicted prices, we observe that the accuracy is about 95 percent.

If we list it, we list it as follows

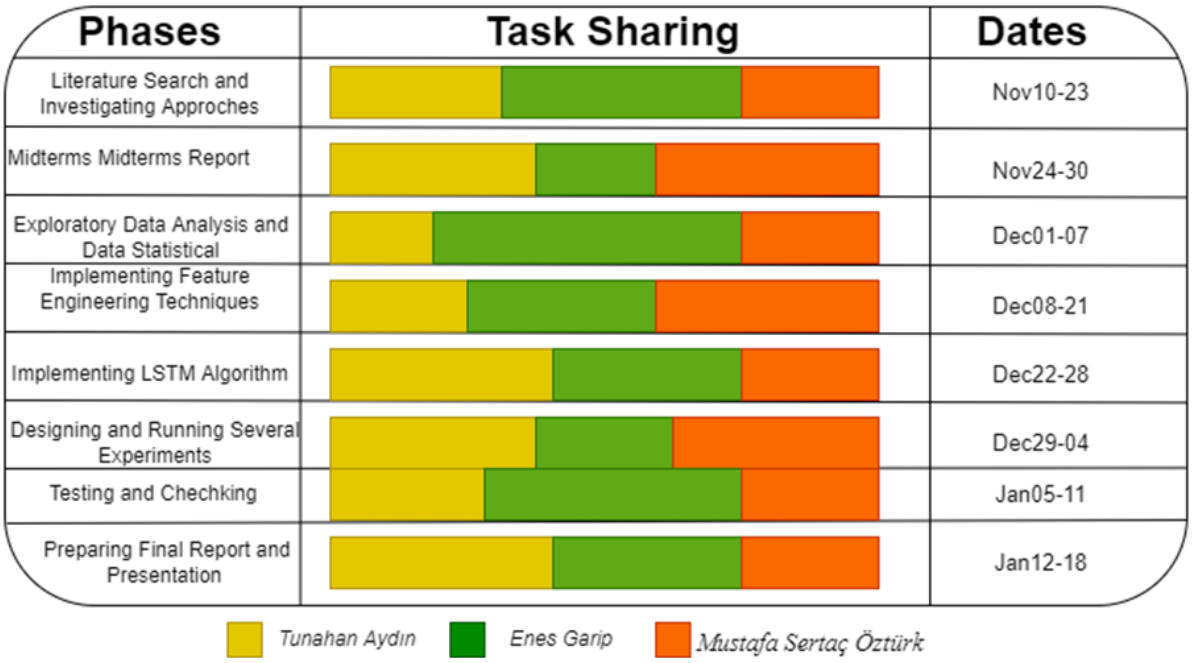
1) We found datasets for prediction.
2) We decided property of data.
3) We sorted dataset on date time and filter “Date” and “Close” columns.
4) We normalized the new filter dataset.
5) We analyzed Lstm networks.
6) We built and trained the Lstm Model.
7) We took a sample of a dataset to make stock price predictions using the LSTM model.
8) We visualized the predicted stock costs with actual stock costs.

VI. Summary

The most important thing I learned in this project is that there is a lot of information and open source code on the internet about machine learning.In our opinion, the most important thing in this project was to step into the field of deep learning.

We realized that very complex things can be implemented in a simple way thanks to libraries. We gain inspiration from several Kaggle notebooks as well. We implemented advanced preprocessing techniques specialized for dimensional datasets and multilabel classification methods. Also, we learned how to increase our results model optimizations.

Task Sharing between group members



VII. References

1. 1. E. F. Fama, “Efficient capital markets: a review of theory and empirical work,” *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.View at: Publisher Site | Google Scholar

2. 2. P.-F. Pai and C.-S. Lin, “A hybrid ARIMA and support vector machines model in stock price forecasting,” *Omega*, vol. 33, no. 6, pp. 497–505, 2005.View at: Publisher Site | Google Scholar

3. 3. X. B. Huang, C. F. Wang, Z. M. Fang, and C. L. Xiong, “Detection of Chinese stock information based on hidden Markov model,” *Syst. Eng.-Theory Pract*, vol. 32, no. 4, pp. 713–720, 2012.View at: Google Scholar

4. 4. J. J. Shi and T. Song, “Analysis of long-term fluctuation trends and influencing factors in stock market-based on spline-GARCH model,” *J. Appl. Stat. Manag*, vol. 34, no. 1, pp. 175–182, 2015.View at: Google Scholar

5. 5. B. Xu, D. Zhang, S. Zhang, H. Li, and H. Lin, “Stock market trend prediction using recurrent convolutional neural networks,” in *Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 166–177, Hohhot, China, August 2018.View at: Publisher Site | Google Scholar

6. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.View at: Publisher Site | Google Scholar

7. 7. H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.View at: Publisher Site | Google Scholar

8. 8. K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: a case study of China stock market,” in *Proceedings of the 2015 IEEE international conference on big data (big data)*, pp. 2823-2824, IEEE, Santa Clara, CA, USA, October 2015.View at: Publisher Site | Google Scholar

9. 9. J. Jiahong Li, H. Junjie Wu, and J. Wu, “Sentiment-aware stock market prediction: a deep learning method,” in *Proceedings of the 2017 international conference on service systems and service management*, pp. 1–6, IEEE, Dalian, China, June 2017.View at: Publisher Site | Google Scholar

10. 10 .X. Jiawei and T. Murata, “Stock market trend prediction with sentiment analysis based on LSTM neural network,” in *Proceedings of the International Multi-Conference of Engineers and Computer Scientists (IMECS)*, pp. 13–15, Hong Kong, China, March 2019.View at: Google Scholar

11. *Stock Prediction Based on Optimized LSTM and GRU Models*. (2021, September 29). Hindawi. Retrieved January 18, 2022, from https://www.hindawi.com/journals/sp/2021/4055281/

12. *Theories Explaining Stock Price Behavior*. (2018, November 1). ResearchGate. Retrieved January 18, 2022, from https://www.researchgate.net/publication/331210273_Theories_Explaining_Stock_Price_Behavior_A_Review_of_the_Literature

13. Y.-L.Z.Y.-G.S. (2018, May 1). *Neural networks for stock price prediction*. Research Gate. Retrieved January 18, 2022, from https://www.researchgate.net/publication/325439759_Neural_networks_for_stock_price_prediction

14. Gao, Y. (2021, September 29). *Stock Prediction Based on Optimized LSTM and GRU Models*. Hindawi. Retrieved January 18, 2022, from https://www.hindawi.com/journals/sp/2021/4055281/

15. Knight, A. (2022, January 7). *LSTM Neural Network: The Basic Concept - Towards Data Science*. Medium. Retrieved January 18, 2022, from https://towardsdatascience.com/lstm-neural-network-the-basic-concept-a9ba225616f7