Stock Price Prediction using LSTM Model

Team: Ojas Mandlecha, Nimisha Salve. **Project Mentor TA:** Ramya Ramalingam

1) Introduction

We propose to predict the stock price of a specific company listed on the stock exchange market on day k as a function of its closing price as well as market and technical indicators of the previous 3 days (i.e k-3 days).

We will import the dataset of a company from Yahoo Finance for the period March 25, 2020, to March 26, 2021. This data would include the closing price for the particular dates. We will also add more data such as the market and technical indicators. This data is divided into three parts: the first 80% would be the training dataset, the next 10% would be the validation data, and the remaining 10% would be the test data.

We will measure the mean squared error and mean absolute error for predicting the test data set. But we will minimize the mean squared error. First we will try to predict the stock price on just one feature which is 'closing price' and then try to include other market and technical indicators such as moving averages (simple and exponential moving averages), market index, and interest rate as input vectors for our LSTM model and check by what percent the error in prediction was reduced. In the end we will choose a model and features which will give us the highest accuracy.

Motivation:

People usually do not have time or capacity to understand the stock market. Our model understands and takes into consideration relevant and useful market indicators as features to predict the stock price and reduce the risk as much as possible. This model can help any layman to grow their money, by making smart investments. But the model will not have 100% accuracy, and looking at a high return on investment prediction, people might invest heavily causing them to lose their money. Investors are urged to do their own research and think carefully about their risk tolerance under varied market circumstances instead of only following the model predictions.

2) How We Have Addressed Feedback From the Proposal Evaluations

We understood from our mentor that it will take more effort and time to understand LSTM to implement its model to predict stock price. Thus, we have started to learn the LSTM model through various video lectures, mooc courses and research papers. During the mentor interactions, our mentor suggested certain changes as to how we should frame our contributions and go about the flow of the project. These suggestions were taken into consideration and we have edited the document accordingly.

3) Prior Work We are Closely Building From

A. Jon Cavallie Mester "Using LSTM Neural Networks To Predict Daily Stock Returns",VT2021. The author has used daily stock trading data and an LSTM training model at predicting daily returns for 60 stocks from the OMX30 and Nasdaq-100 indices.

4) What We are Contributing

1. **Contribution(s) in Application/Data: We would predict the stock price not just on closing price but also technical indicators such as moving averages (simple and exponential moving averages), market index, and interest rate.**
2. **Contribution(s) in Algorithm: The program will be written in Pytorch. The prior work which we are referring to considers the past 1 day for the prediction but we would make use of the past 3 days.**
3. **Contribution(s) in Analysis: We will analyze the model by using different combinations of these features and choose a model corresponding to the set of features which will give the highest accuracy.**

5) Detailed Description of Each Proposed Contribution, Progress Towards It, and Any Difficulties Encountered So Far

5.1 Methods

1) Data

There are 4 types of data that are needed to build the dataset for this project - moving average, exponential moving average, market index and historical stock price information.
The historical data and market index are much easier to acquire, and we pulled it from the Yahoo Finance Python API (yfinance)[I]. These tickers in our case are:
- Reliance Industries Ltd. (RELIANCE.NS)
- Nifty 50 (^NSEI)

The moving average and exponential moving average is calculated on the basis of the historical data.
In addition to this, the API also takes in the start and the end dates, and returns a dataframe for all the stock price data available between these 2 dates. In our case,
- Start date: 2020-03-25 (March 25, 2020)
- End date: 2021-03-26 (March 26, 2021)

2) Algorithm

Pseudocode for LSTM

```
def LSTM_cell (prev_ct, prev_ ht, input):
        combine = prev_ht + input
        f_t  = forget_layer(combine)
        candidate = candidate_layer (combine)
        i_t = input_layer(combine)
        C_t = prev_ct * f_t + candidate * i_t
        o_t = output_layer(combine)
        h_t = o_t * tanh(C_t)
        return h_t, C_t
c_t = [0,0,0]
h_t = [0,0,0]
```

for input in inputs:
    C_t, h_t = LSTM_cell (c_t, h_t, input)

The time period moving average and exponential average is considered to be 15 days.
Currently this prediction is made only using stock price data and we have not included any other features.

3)Analysis

While experimenting for different time periods, it was found that when the time period was large, the model tried to overfit and caused high variance. Also, the stock price of a company during establishment is not significant to predict today's stock price. Therefore, to overcome this, the time period was selected as 1 year. Training and predicting stock price on different sets of features and analyzing it will be the next step.

5.2 Experiments and Results

Our model will make use of many features such as closing price, moving averages (simple and exponential moving averages), market index, and interest rate etc. It will tell us which set of features gives us highest accuracy or lowest mean square error. First, we will just implement a model which would predict the stock price based on just the closing price. This will also be our baseline model.
We have imported the dataset of 'Reliance Industries' from Yahoo Finance for the period March 25, 2020, to March 26, 2021. This data would include the closing price for the particular dates. This data is divided into three parts: the first 80% would be the training dataset, the next 10% would be the validation data, and the remaining 10% would be the test data. For predicting the stock price on just closing price, we have used ReLU as activation function, Adam as optimizer and kept the learning rate = 0.0005 with epochs = 1000. The RMSE was 33.3393.

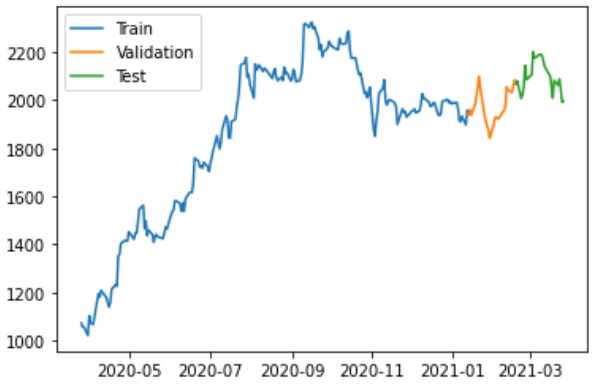


Fig 1. Division of training, testing and validation data

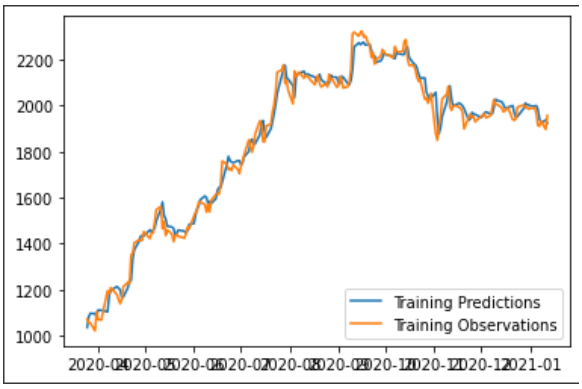


Fig 2. Training the model and predicting using test data

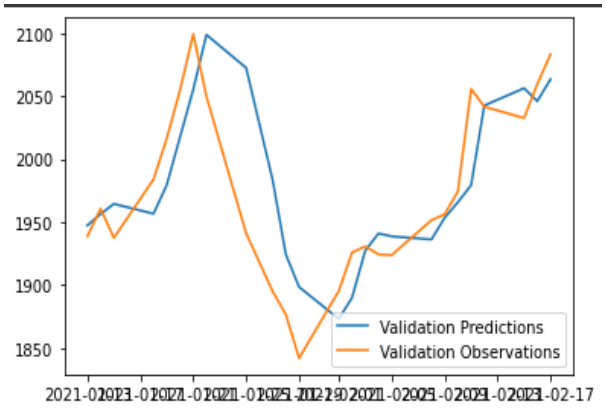


Fig 3. Validation Observations vs Validation Prediction

Fig 4. Testing Observations vs Testing Prediction

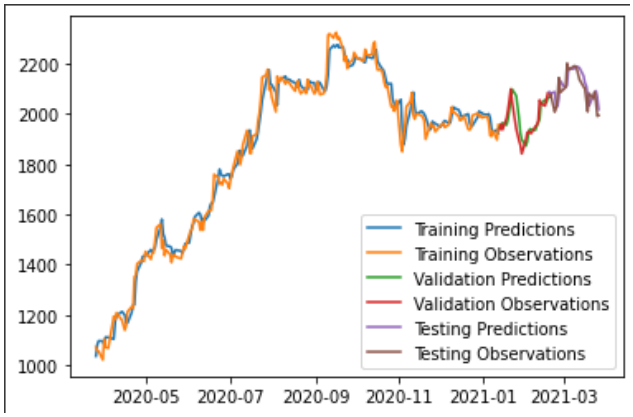


Fig 5. All the actual observations vs predictions from the model.

6) Risk Mitigation Plan

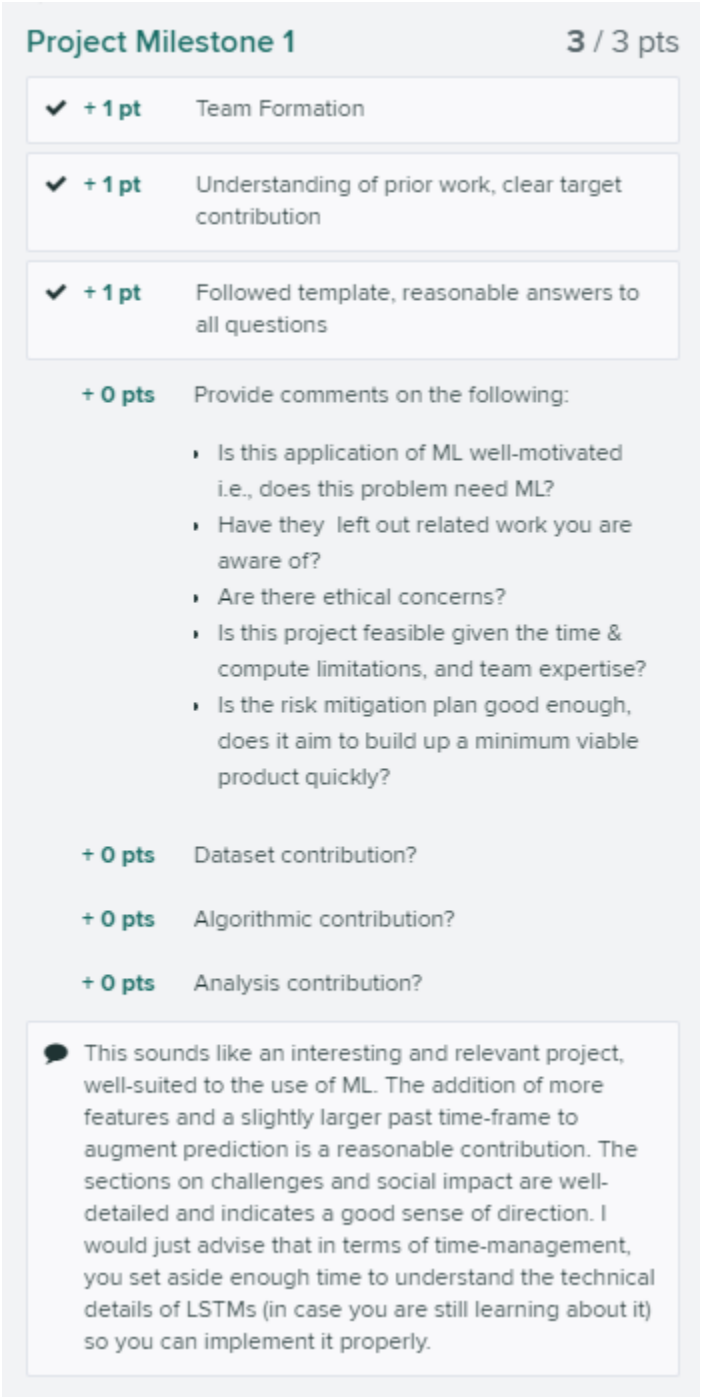
LSTM models tend to overfit the training dataset. Stock markets are very volatile and have non-linear behavior. Similarly, very old data points would have less significance in predicting future values in comparison with recent data points. So training the model for a particular company from the starting date of its listing would overfit the model and might give us a wrong prediction. Choosing the optimum time period will help us minimize this. For now, we have chosen the time period as 1 year. Another challenge would be that LSTM generally does not extrapolate well outside the range of values it is trained on unlike other regression models. To avoid this, we can either train a stock price with volatility or select a test data having similar trends of the training set. Using ensembling techniques along with LSTM might also help.

A minimum viable project would be to predict the stock price on just closing price, which will give us some results about our model. This model makes use of simple data (only dates and closing price), less number of features. This will also be our base code or model. We will add other features to this base model and analyze the above mentioned things. Even if our model fails or generates a huge error which has high chances of it, as it is a stock price prediction which does not depend on just one factor, we will try to analyze the reasons for failure and include it in our report.

(Exempted from page limit) Other Prior Work / References (apart from Sec 3) that are cited in the text:

- Ko CR, Chang HT. LSTM-based sentiment analysis for stock price forecast. PeerJ Comput Sci. 2021 Mar 11;7:e408. doi: 10.7717/peerj-cs.408. PMID: 33817050; PMCID: PMC7959635.
  In this paper, the stock price was predicted using NLP and LSTM and improved the RSME by 12.05%.
  Website:
- Yongqiong Zhu, “Stock price prediction using the RNN model”, Journal of Physics 2020. Here, the author has made use of Recurrent Neural Networks to predict Apple's stock price using the opening price in TensorFlow 2.0.
  The past 10 years of data have been used for analysis and over 95% accuracy has been achieved

(Exempted from page limit) Attach your proposal here, as a series of screenshots from Gradescope, starting with a screenshot of your main evaluation tab, and then screenshots of each page, including pdf comments. For example:



Project Milestone 1 **3** / 3 pts

✔ **+ 1 pt** Team Formation

✔ **+ 1 pt** Understanding of prior work, clear target contribution

✔ **+ 1 pt** Followed template, reasonable answers to all questions

+ 0 pts Provide comments on the following:

- Is this application of ML well-motivated i.e., does this problem need ML?
- Have they left out related work you are aware of?
- Are there ethical concerns?
- Is this project feasible given the time & compute limitations, and team expertise?
- Is the risk mitigation plan good enough, does it aim to build up a minimum viable product quickly?

+ 0 pts Dataset contribution?

+ 0 pts Algorithmic contribution?

+ 0 pts Analysis contribution?

This sounds like an interesting and relevant project, well-suited to the use of ML. The addition of more features and a slightly larger past time-frame to augment prediction is a reasonable contribution. The sections on challenges and social impact are well-detailed and indicates a good sense of direction. I would just advise that in terms of time-management, you set aside enough time to understand the technical details of LSTMs (in case you are still learning about it) so you can implement it properly.

Title: Stock market prediction model using LSTM
Team: Ojas Mandlecha (39506117) (CIS 5190), Nimisha Salve (35786696) (CIS 5190)

Task T: We propose to predict the stock price of a specific company listed on the stock exchange market on day k as a function of its closing price and market and technical indicators of the previous 3 days (i.e k-3 days).

Experience E: We will import the dataset of a company from Yahoo Finance for the period January 1, 2021, to December 31, 2021. This data would include the closing price for the particular dates. We will also add more data such as the market and technical indicators. This data is divided into three parts: the first 70% would be the training dataset, the next 15% would be the validation data, and the remaining 15% would be the test data.

Performance metrics P: We will measure the mean squared error and mean absolute error for predicting the test data set. But we will minimize the mean squared error.

Prior work:

1. Jon Cavallie Mester "Using LSTM Neural Networks To Predict Daily Stock Returns",VT2021. The author has used daily stock trading data to let an LSTM train model at predicting daily returns for 60 stocks from the OMX30 and Nasdaq-100 indices.
2. Ko CR, Chang HT. LSTM-based sentiment analysis for stock price forecast. PeerJ Comput Sci. 2021 Mar 11;7:e408. doi: 10.7717/peerj-cs.408. PMID: 33817050; PMCID: PMC7959635. In this paper, the stock price was predicted using NLP and LSTM and improved the RSME by 12.05%.
3. Yongqiong Zhu, "Stock price prediction using the RNN model", Journal of Physics 2020. Here, the author has made use of Recurrent Neural Networks to predict Apple's stock price using the opening price in TensorFlow 2.0. The past 10 years of data have been used for analysis and over 95% accuracy has been achieved.

Nature of main proposed contribution(s): Contribution in Algorithm: Most of the existing algorithms predict the stock price of a company based on just the previous day's closing price. In addition to this, we would also include other features like technical indicators such as moving averages (simple and exponential moving averages), market index, and interest rate as input vectors in LSTM to predict the stock price. This algorithm would be written in PyTorch instead of TensorFlow.

Why we care: Ojas loves stock market trading and has 2 years of experience in it and he would like to use this model for his own personal use as well. Nimisha likes this domain very much and she thinks it will be fun to build a model which has new data.

Which parts of the curriculum from this class do you expect to apply?: After learning about the basics and foundations of Neural Networks in class, we would use this knowledge to understand the LSTM model and implement it in this project.

Expected challenges and risk mitigation: LSTM models tend to overfit the training dataset. Stock markets are very volatile and have non-linear behavior. Similarly, very old data points would have less significance in predicting future values in comparison with recent data points. So training the model for a particular company from the starting date of its listing would overfit the model and might give us a wrong prediction. Choosing the optimum time period will help us minimize this. For now, we have chosen the time period as 1 year. Another challenge would be that LSTM generally does not extrapolate well outside the range of values it is trained on unlike other regression models. To avoid this, we can either pick a stock with varying training values that is a high standard deviation or pick test data that lies in the ranges of train data so the stock is not much volatile. Using ensembling techniques along with LSTM might also help.

Ethical considerations and broader social impact: As this project predicts the stock price of a company, this can emotionally affect people's decisions. Not everyone has time to learn the basics of the stock market, but investing is a necessity. So this model can help even an uneducated person to grow their money, savings, and investments for their own better future. This model will also help people from getting scammed by fake predictions from unreliable sources, such as friends, relatives, etc. But, as each coin has two sides, this also has a bad social impact. The model will not have 100% accuracy, and looking at a high return on investment prediction, people might invest heavily which can be subject to market risk and lose their money. This would have adverse effects if they invested their savings. Therefore, this model should not be the only factor considered when making an investment. Investors are urged to do due research and think carefully about their risk tolerance under varied market circumstances.

(Exempted from page limit) Supplementary Materials if any (but not guaranteed to be considered during evaluation):