

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.

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.

(Supplementary materials if any, are exempted from 2pg limit, but not guaranteed to be considered during evaluation)