

STOCK MARKET PREDICTION USING SENTIMENT ANALYSIS AND DEEP LEARNING

TIJO THOMAS
TRESA POTHEN
GLADDEN RUMAO

UNDER THE GUIDANCE OF PROF SUNIL CHAUDHARI

INTRODUCTION

Stock market is a place where buying and selling of shares happen for publicly listed companies. Stock exchange is the mediator that allows buying and selling of stocks.



Modern data mining techniques have led to the development of sentiment analysis, an algorithmic approach for detecting the predominant sentiment about a product or company using social media data.

Stock sentiment alone cannot always predict changes in share prices, but when combined with tools such as technical analysis applying Machine Learning, a better understanding can be gained to determine possible scenarios.

Literature review



1

Li, S., Wang applied sentiment analysis by using twitter data for predicting success rate of movie. For this purpose movies were classified as Flop, Hit, and Average. The tweets from 2009 to 2013 duration were extracted, and each tweet was classified as positive, negative, neutral and irrelevant. Lingpipe sentiment analyzer was used to test the sentiments, and result showed that movie prediction accuracy of the developed system was 64.4 % better than conventional system

2

Molla made sentiment analysis for user opinions about different Samsung products using different twitter official accounts of Samsung Company. For visualizing the result of user's opinion data visualization tool such as NodeXL was used for the social network graph. Future work was proposed that it may be focus on the location management of each tweet and inclusion of emotions.



Tao Xing and Yuan Sun et al. suggested a model which considers the historical equity share price of a company price and applies RNN (Recurrent) technique called Long Short Term Memory (LSTM). The proposed approach considers available historical data of a share and it applies prediction on a particular feature. The features of shares are Opening price, day High, day Low, previous



Xi Zhang¹ et al. suggested that the stock markets play critical roles in modern society's economic operations. It is also mentioned in the paper that the analysis can be performed on the data that is retrieved from a legitimate source and proposed a methodology in which we can utilize multiple source of information to predict the stock values.



Understanding the market

Problem Statement



- > To accurately predict the trend of a given stock across a given period of time.
- > Use different deep learning models available and compare them in graphical analysis

Web Scraping using BeautifulSoup




How does Web Scraping work ?



- When we scrape the web, we write code that sends a request to the server that's hosting the page we specified. The server will return the source code — HTML, mostly — for the page (or pages) we requested.
- So far, we're essentially doing the same thing a web browser does — sending a server request with a specific URL and asking the server to return the code for that page.
- But unlike a web browser, our web scraping code won't interpret the page's source code and display the page visually. Instead, we'll write some custom code that filters through the page's source code looking for specific elements we've specified, and extracting whatever content we've instructed it to extract.

What is sentimental analysis in the stock market?



Stock sentiment analysis can be used to determine investors' opinions of a specific stock or asset. Sentiment may at times hint at future price action. This is also an example of how trading psychology can affect a market, assisting as a forecasting tool to determine possible future price changes in a particular asset.

There are various factors that influence stock sentiment, which include news (economic, political and industry related) and social media. These factors help influence stock sentiment as they impact stock market volatility, trading volume and company earnings.

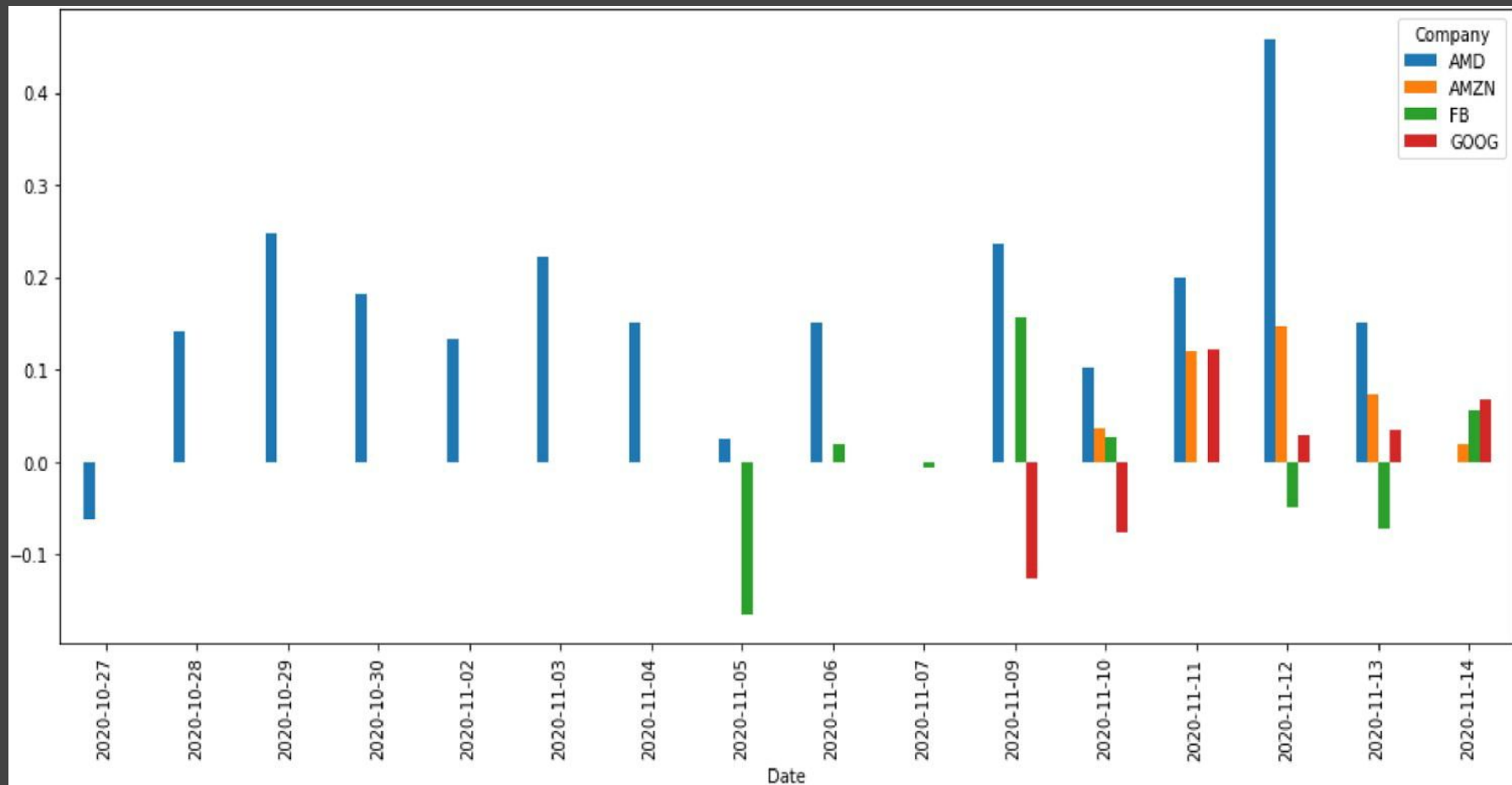
Sentiment Analysis using VADER



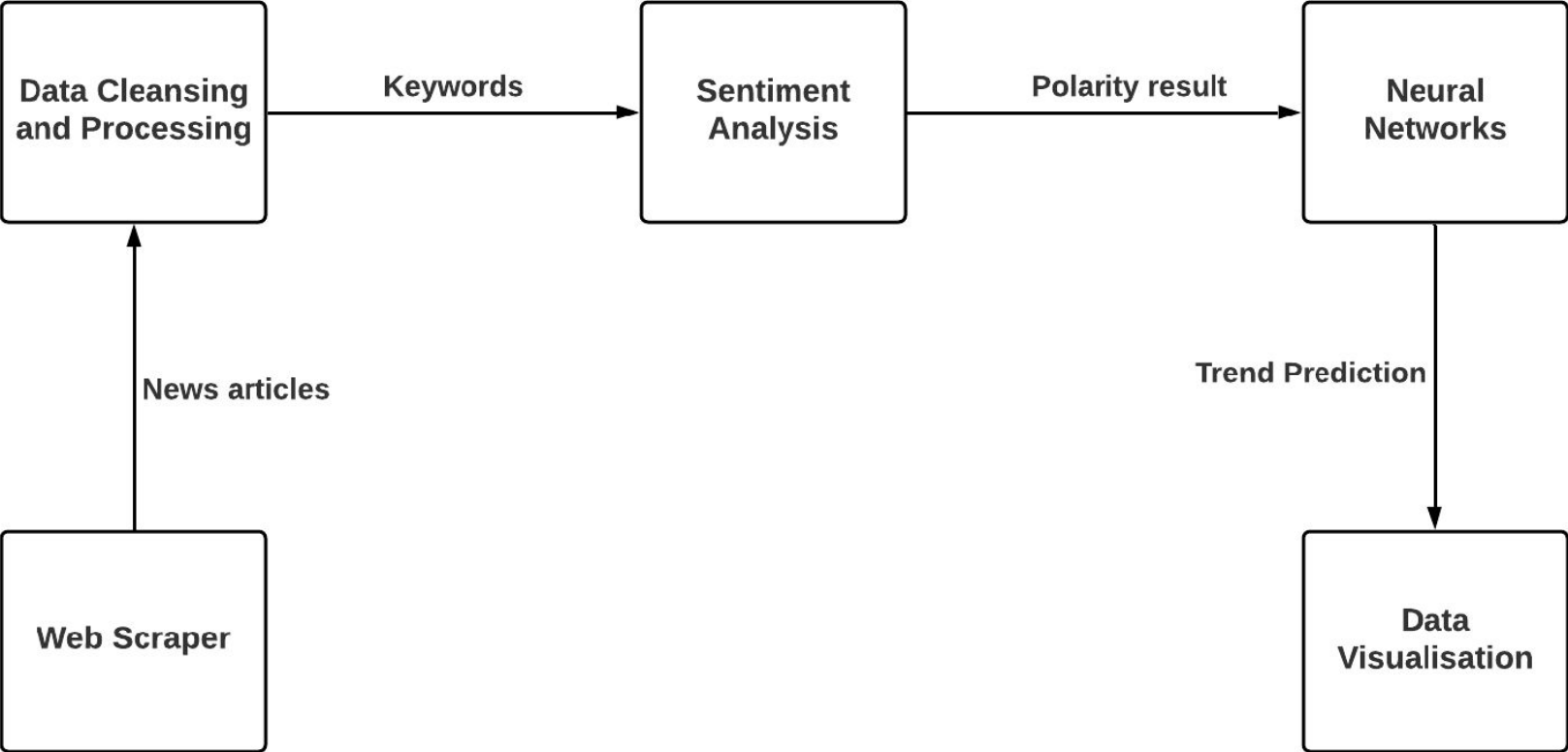
Sentiment Analysis using VADER



- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.
- It is used for sentiment analysis of text which has both the polarities i.e. positive/negative. VADER is used to quantify how much of positive or negative emotion the text has and also the intensity of emotion.



System Overview



Methodology

Algorithms / Technologies/software specifications

Predictions will be made using three algorithms: SVM, LSTM, Linear Regression. We aim to make a Web App that will combine the predicted prices of the next seven days for any given stock under NASDAQ or NSE as input by the user with the sentiment analysis of tweets to give recommendation whether the price is going to rise or fall.

- **Three Approaches for this problem**

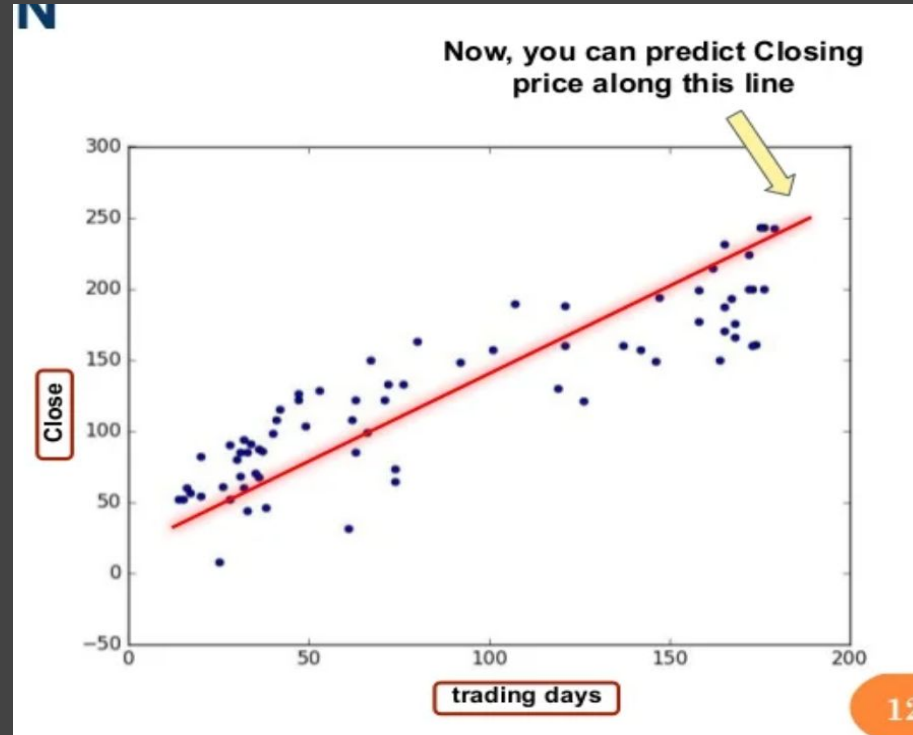
Linear Regression

Support Vector Machine

RNN + LSTM

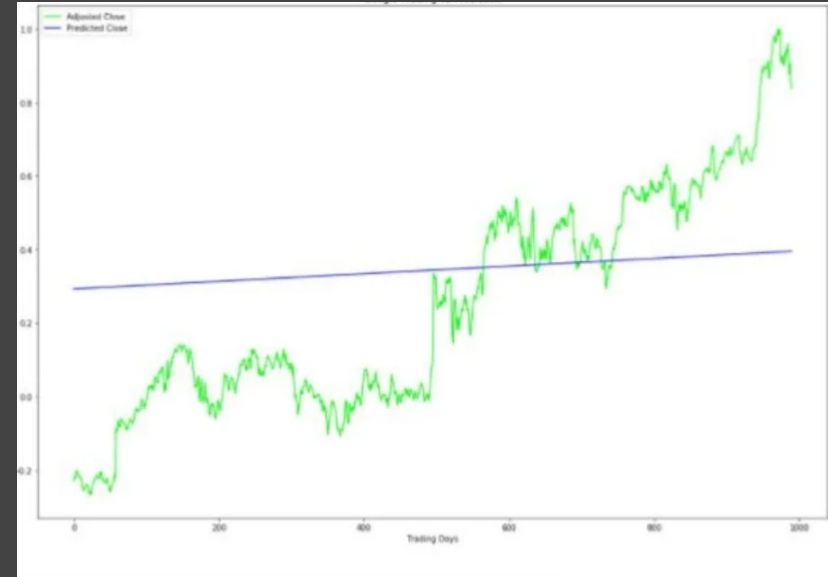
Linear Regression

Linear Regression is an approach for predictive modeling to showcase the relationship between a scalar dependent variable 'Y' (in our case we have the close attribute) and one or more independent variables X (Trading day attribute)



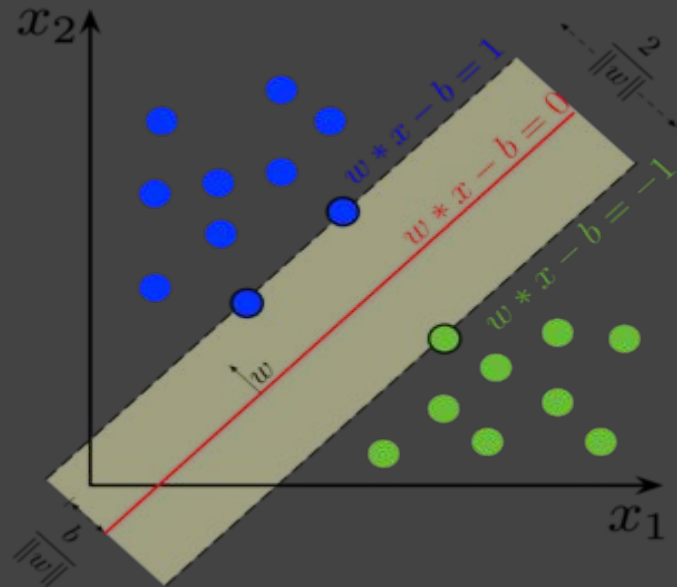
A linear model was later fit to this graph and displayed and observations were made. The values of MSE, RMSE, MAE, MAPE and R were obtained and are shown below

Method	Result
RMSE	3.22
MAE	2.53
MSE	10.37
R-Squared	0.73



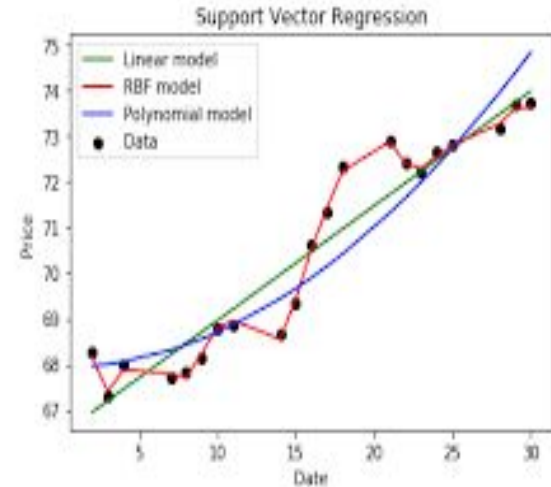
Support Vector Machine

Support Vector Machines (SVMs) are used for classification. The goal of an SVM is to define a boundary line between the 2 classes on a graph. We can think of this as "splitting" the data in the best possible way. This boundary line is called a hyperplane.



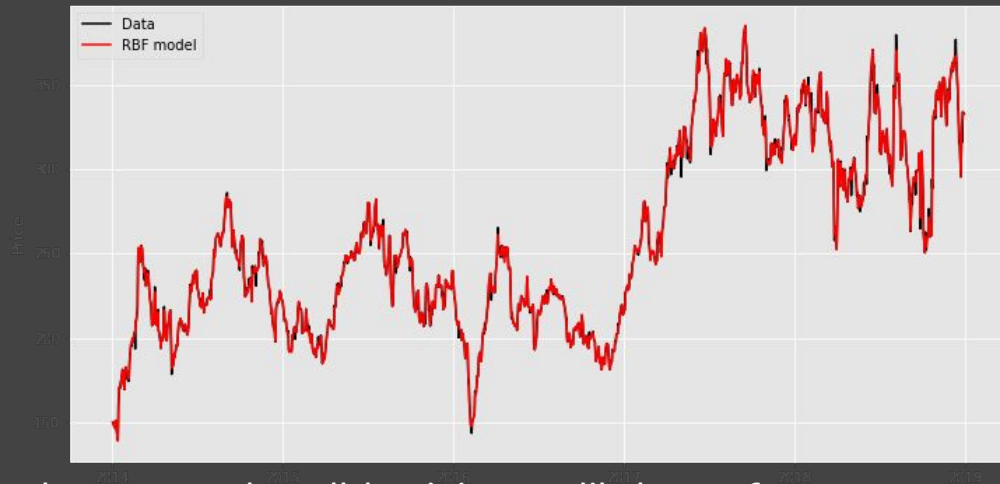
The exact same steps were performed for Prediction using SVM with the only change being the calling of the svm() function. It is therefore, conclusive that SVM performs better than LR.

Method	Result
RMSE	1.58
MAE	1.33
MSE	2.51
R-Squared	0.93



Support Vector Regression (Research paper results)

Using regression it minimizes the cost function using something like gradient descent. With SVM it tries to try to draw a hyperplane between 2 different classes. So SVR is the combo of the 2, to minimize the error within a certain threshold .

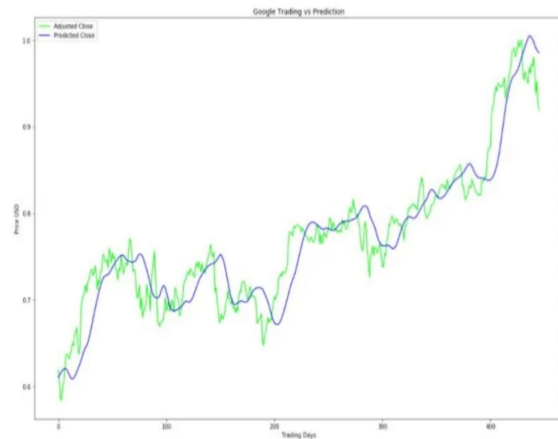


This model fits the data extremely well, but it is most likely overfit

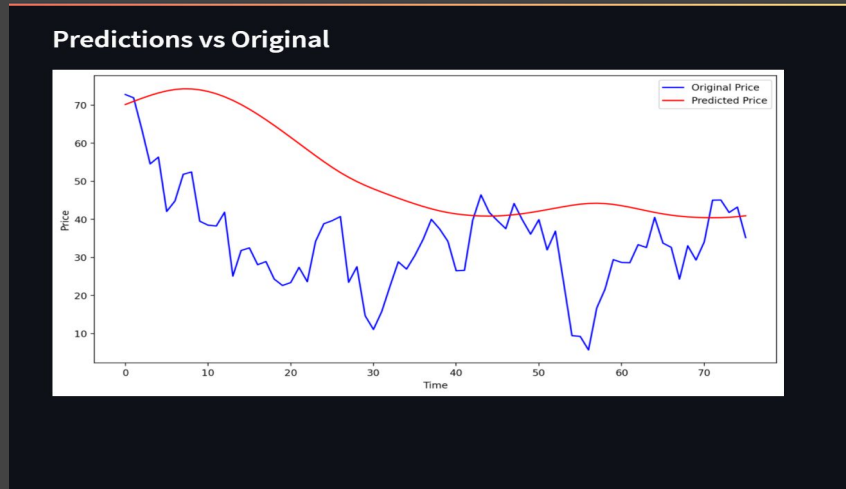
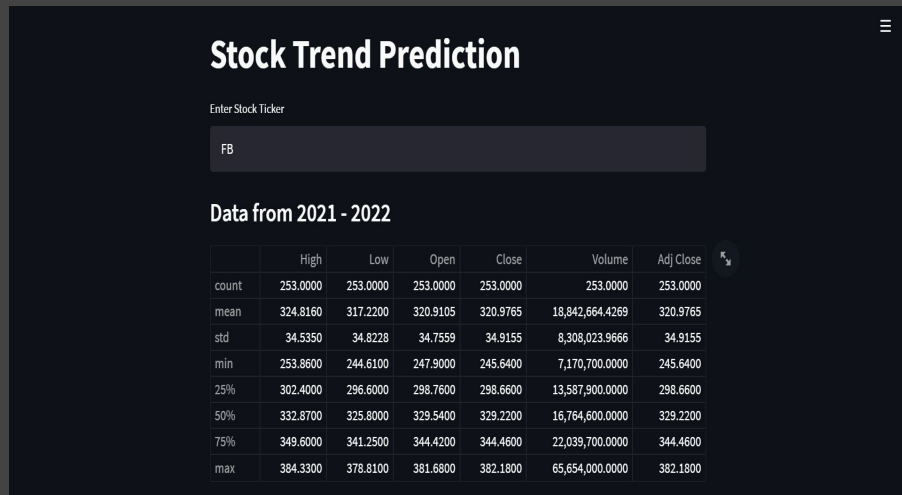
RNN + LSTM

LSTM stands for Long Short Term Memory. Recurrent neural networks suffer from the vanishing gradient problem. LSTMs are an advanced version of recurrent neural networks. LSTMs are a type of RNN with gates inside of each LSTM cell. These gates inside LSTM cells help the LSTM decide what data is important to be remembered and what data can be forgotten even on long series of data. Hence we have chosen LSTM model to be our final model for our project.

RECURRENT NEURAL NETWORK + LSTM



Screenshots



After comparing the results of different models , the streamlit website displays graphical results (actual vs predicted) (vs 100 ma) (vs 200 ma) of the stock ticker entered by the user. The project used LSTM model for prediction. The data is called through yahoo finance.

References



- [1] Diederik P. Kingma, Jimmy Lei Ba “Adam: A Method for Stochastic Optimization“ ICLR Conference, 2015.
- [2] S. Nnewi - PREDICTING THE NIGERIAN STOCK MARKET USING ARTIFICIAL NEURAL NETWORK
- [3] H.B.Kekre, Hitesh Makhija, Pallavi N.Halarnkar. Prediction of Closing Price of Stock Using Artificial Neural Network
- [4] Zabir Haider Khan. Price Prediction of Share Market using Artificial Neural Network (ANN)



Thank you.

