

Brought to you by Huma Tejaswi Akke,

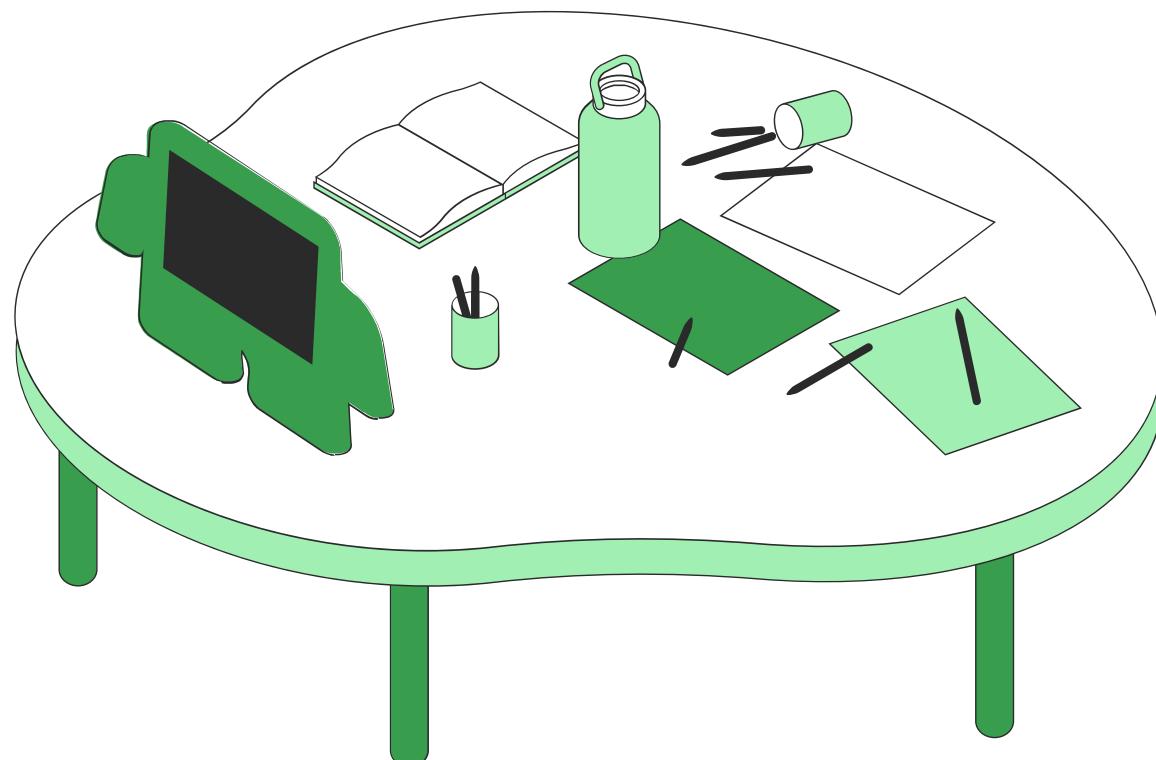
Stock Price prediction

UIC Honors Capstone



Agenda

What to Know



Background Information about Stock Market Prediction

Details on Report

Goals of Capstone

Domain questions

Visualizations of stock Exchange Data

Results

Background Information about Stock Market Prediction

- A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares.
- Stock price analysis has been a critical area of research and is one of the top applications of machine learning
- Why stocks can be so important to humans and why it started developing.

Details on Report

- Expertise proved that more than 90% of people who lost their jobs in the pandemic, started knowing and owning the stocks.
- This model will go through K Nearest Neighbour Algorithm and then go through the average moving formula using closing prices of different stocks.
- This model also highlights the personal views of the topic, theoretical concepts, mathematical formula used in support of dealing with KNN Algorithms and graphs from desired outputs.

Goals:

- The goal of this machine learning project is to predict the stock as accurately as possible by using the data set that consists of the information of stock over past years and also it helps to analyze the historics data sets.
- The main intention for the project is to know the usage of machine learning algorithms in the financial market system.
- Also to show that KNN gives more accurate results than other methods.
- Exploring data mining by mathematical calculation and graph visualization models.

Domain Questions

- (1) What is the task definition or setting? What are the input and output? As it is related to machine learning, what kind of data and annotations are required?
- (2) Why is solving this task important? Does it have any killer applications?
- (3) How will the performance be evaluated?

How to plot the Data (Implementation of data)

Examples of L1 form used in project :

- 1) Used for Linear regression (when stocks increased in a straight line).
- 2) Discrete Frequency Distributions.

$$d_1(I_1, I_2) = \sum |I_1^p - I_2^p|$$

Then, similarly I used L2 form (Euclidean distance) which took the majority of plotting decision in data.

$$d_2(I_1, I_2) = \sqrt{\sum_p (I_1^p - I_2^p)^2}$$

The stock data and the test data is mapped into a set of vectors. Each vector represents N dimension for each stock features. Then, a similarity metric such as Euclidean distance is computed to take a decision. A majority vote is performed among the selected k records to determine the class label and then assign it to the query.

The prediction of stock price is computed using KNN as follows:

- a) Determine number of nearest neighbors.
- b) Compute the distance between the training samples and the query.
- c) Sort all training records according to the distance values.
- d) Use a majority vote for the class labels of k nearest neighbors, and assign it as a prediction value of the query record.

(1) What is the task definition or setting? What are the input and output? As it is related to machine learning, what kind of data and annotations are required?

- Inputed the data in form of query and tested the input data by determining the distance L2 form
- for this model we are taking the following characteristics into consideration for training,
 - Open
 - High
 - Low
 - Close
 - Price
- T/F about the increase in price
- To implement we first have to handle the data from files to datasets, and then implement a KNN algorithm for it
-

```
df.rename(columns={'Adj Close': 'ticker'}, inplace=True)  
df.drop(['Open', 'High', 'Low', "Close", 'Volume'], axis=1, inplace=True)
```

‘High’ denotes the highest value of the day. ‘Low’ denotes the lowest. ‘Open’ is the opening Price and ‘Close’ is the closing for that Date. Now, sometimes close values are regulated by the companies. So the final value is the ‘Adj Close’ which is the same as ‘Close’ Value if the stock price is not regulated. ‘Volume’ is the amount of Stock of that company traded on that date.



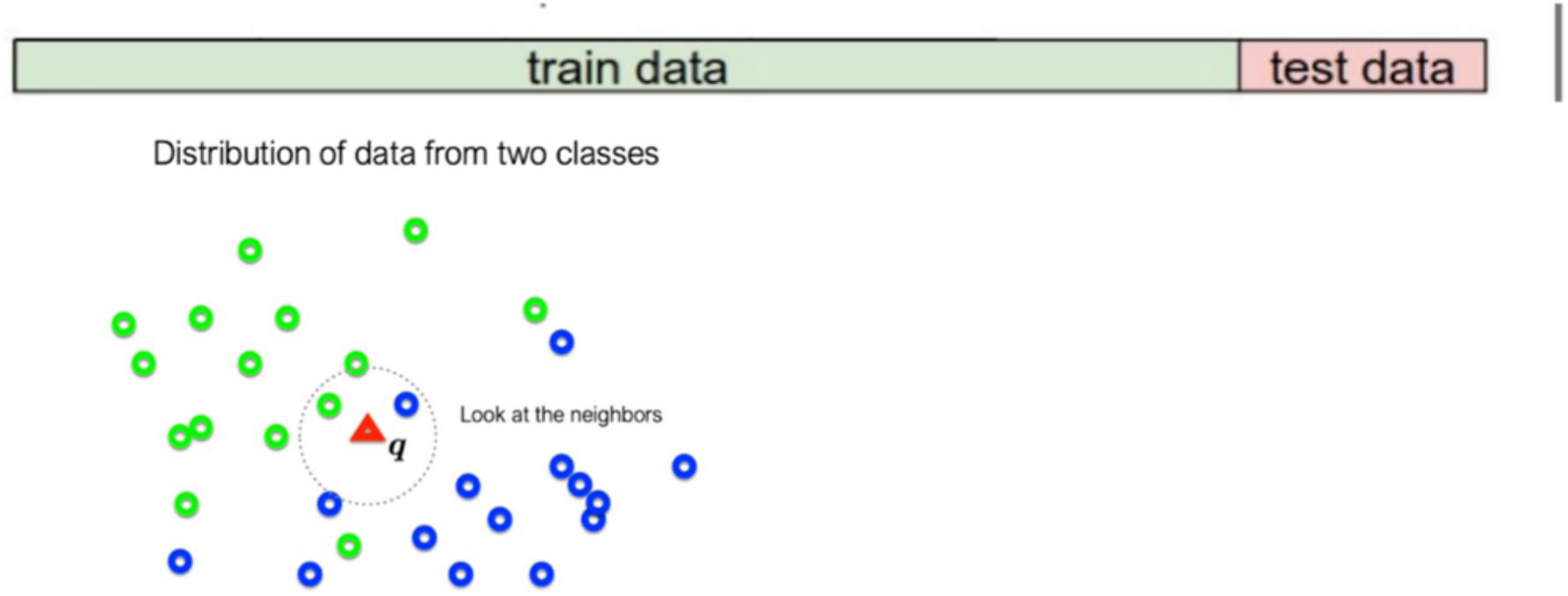
2) Why is solving this task important? Does it have any killer applications?

Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. There are other factors involved in the prediction, such as physical and psychological factors, rational and irrational behavior, and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy.

(3) How will the performance be evaluated?

The KNN approach is complicated while classifying the discrete values but the continuous values it was much easier then expected. When we use KNN, We divide the chosen data into two parts.

1. Test Data: This contains the values that are predicated in stock data sets with the help of algorithm
2. Training Data : divide the data into vectors and then distance from the test data to its nearest neighbour which can be calculated by using several techniques(we used Euclidean distance).



Before Evaluating the model we will train the data to avoid any look ahead bias using cross-validation and based on accuracy of the model will be evaluated.

Our Model accuracy rate is 72% which is was improved by 12%

The result of this program would be a graph showing how the stock has been behaving before and how it will look in the future, the results are not as accurate as I would like them to be, as there are a lot of factors that we are still not taking into consideration for this model.

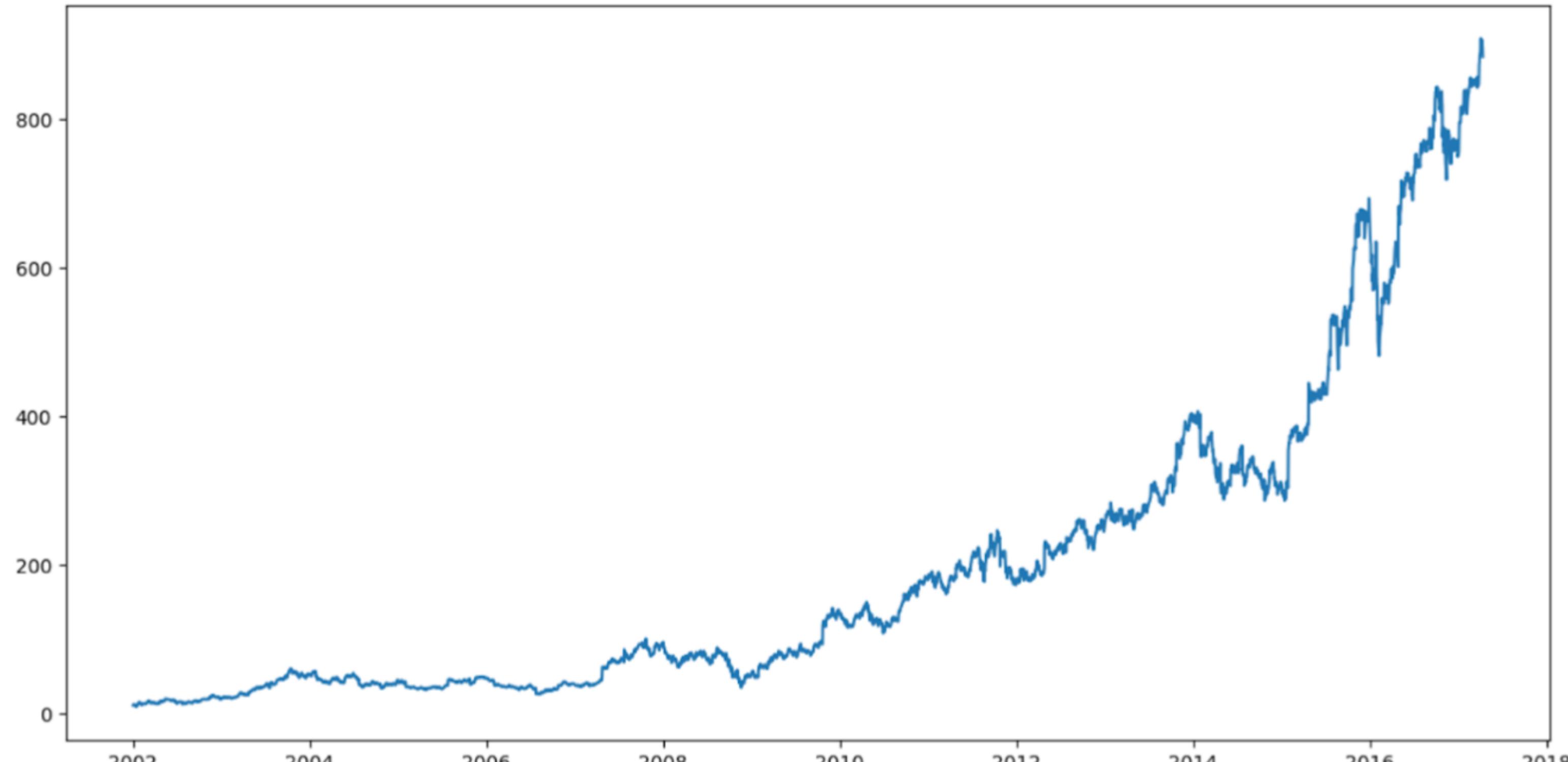
This are the two factors I would take into consideration to improve the accuracy rate.

LSTM : Long short Term Memory Network

For evaluation, thought of considering (Root Mean Squared Error)

some of the results

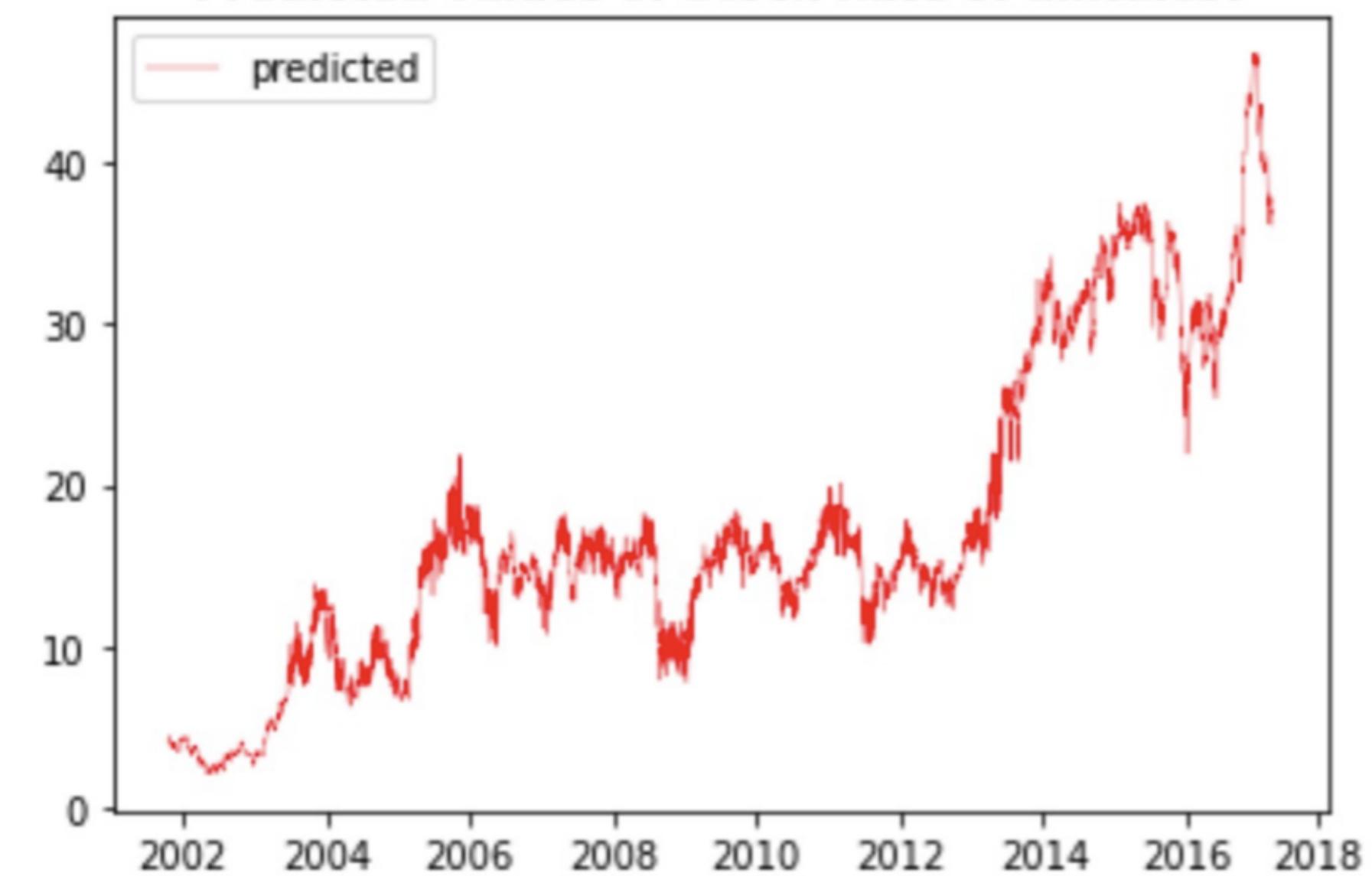
Korn Ferry and Uber

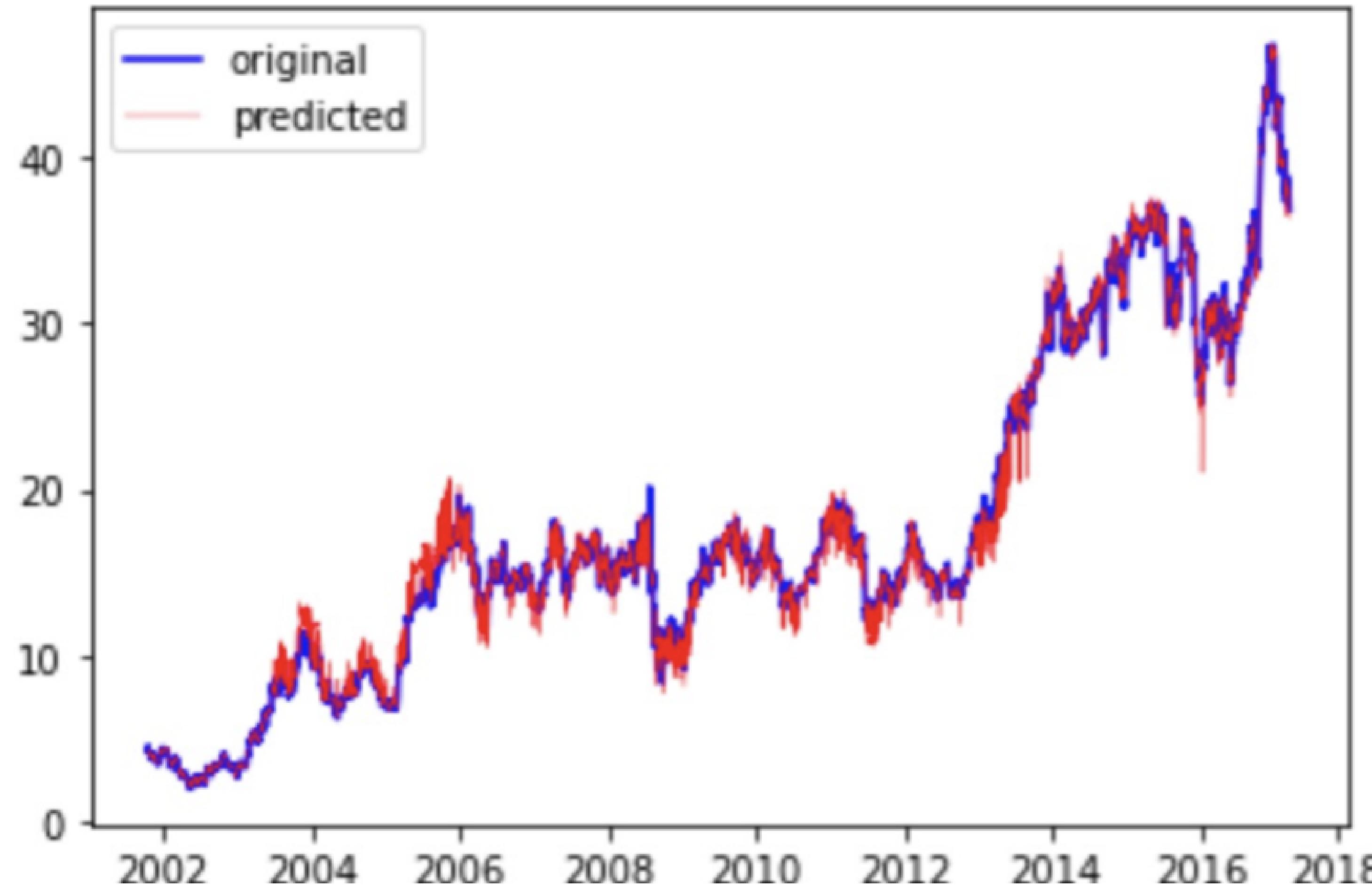


Orignal values of Stock Rate of amtd.csv



Predicted values of Stock Rate of amtd.csv

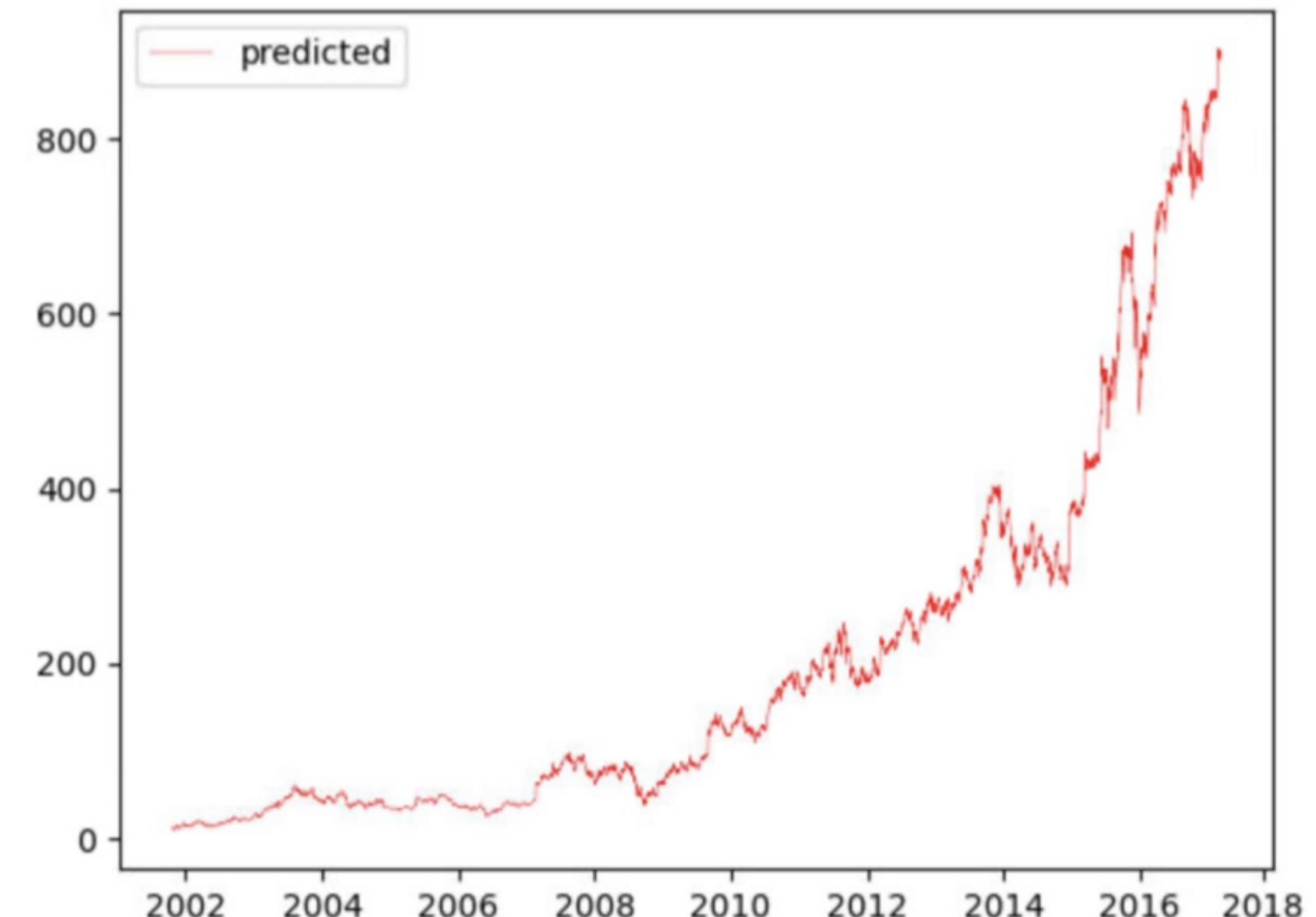




Orignal values of Stock Rate of amazon.csv



Predicted values of Stock Rate of amazon.csv



Predicted values of Stock Rate of amazon.csv

