Stock Prediction using Genetic Algorithm and LSTM Neural Network

An Intelligent Computing project by
Ankit, Sakshi, David and Konark

Supervised by :

Dr. Pratik Chattopadhyay

Team :

Ankit   - 19075009
Sakshi   - 19075064
David   - 19075089
Konark   - 19075040

Contents

Chapter 1

What is Stock Market?

How is its data formatted?

Stock Market

Stock Market refers to public markets that exist for issuing, buying and selling stocks that trade on a stock exchange, which represents partial ownership claims on businesses. It consists of exchanges in which shares and other financial securities of publicly held companies are issued and traded.

Example of Stock Market Data:

Date	Open	Close	High	Low	Volume
01/12/2021	131.6500	132.4450	131.1200	132.420	28349859

Where,
    Date: The date for which the stock data is present
    Open: Opening price of the stock (here in $)
    Close: Closing price of the stock
    High: Maximum price of the stock in the day
    Low: Minimum price of the stock in the day
    Volume: Total number of stocks exchanged in the day

What are Genetic Algorithms?

Working of Genetic Algorithm

Genetic Algorithms

Genetic Algorithm is a population based probabilistic search and optimization technique that works based on the mechanics of natural genetics and evaluation.

Genetic algorithms are excellent for searching through large and complex data sets.

Genetic Algorithm Operators

1) Selection Operator: The idea is to give preference to the individuals with good fitness scores and allow them to pass their genes to successive generations.

2) Crossover Operator: This represents mating between individuals. Two individuals are selected using selection operator and crossover sites are chosen randomly. Then the genes at these crossover sites are exchanged thus creating a completely new individual (offspring).

Genetic Algorithm Operators

3) Mutation Operator: The key idea is to insert random genes in offspring to maintain the diversity in the population to avoid premature convergence. Mutation is done to explore other available solutions.

Working of Genetic Algorithm:

➔ The algorithm begins by creating a random initial population.
➔ The algorithm then creates a sequence of new populations. At each step, the algorithm uses the individuals in the current generation to create the next population.
➔ The algorithm stops when one of the stopping criteria is met.

Working of Genetic Algorithm:

→ To create the new population, the algorithm performs the following steps:
- Scores each member of the current population by computing its fitness value (fitness scores).
- Scales the raw fitness scores to convert them into a more usable range of values. These scaled values are called expectation values.
- Selects members, called parents, based on their expectation.

Working of Genetic Algorithm:

- Some of the individuals in the current population that have lower fitness are chosen as elite and passed to the next population.
- Children are produced from parents either by making random changes to a single parent, i.e., **mutation** or by combining the vector entries of a pair of parents, i.e., **crossover**.
- Replaces the current population with the children to form the next generation.

Chapter 3

What is LSTM?

Why are we using it?

LSTM

LSTM is a recurrent neural network (RNN) architecture that remembers values over arbitrary intervals. As we know that an RNN can be used to solve problems like auto completion, translation, NER or sentiment analysis. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration where Relative insensitivity to gap length gives an advantage to LSTM over alternative RNNs

How LSTMs Remember:

RNN cell takes in two inputs, output from the last hidden state and observation at time = t. Besides the hidden state, there is no information about the past to remember.
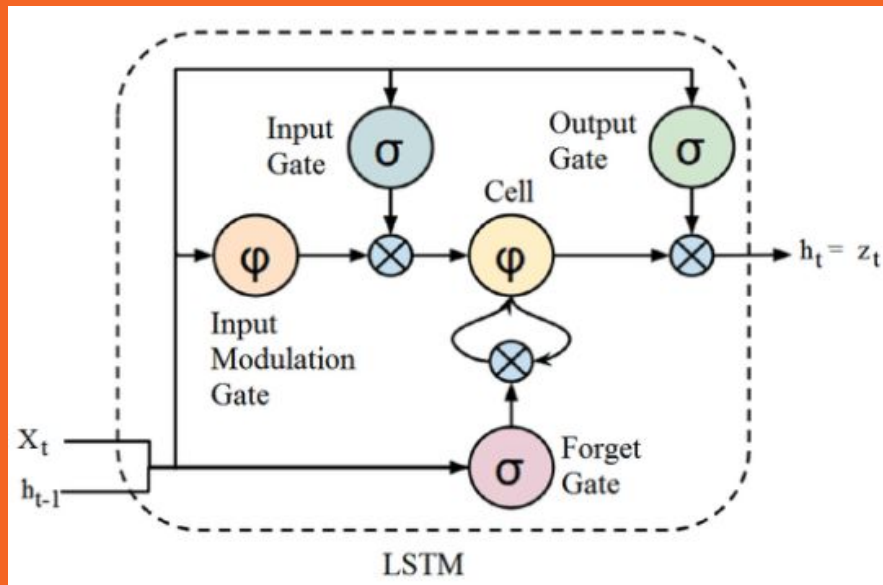
LSTM

How LSTMs Remember:

The **long-term memory** is usually called the **cell state**. Recursive nature of the cell is shown by the arrow. This allows information from previous intervals to be stored with in the LSTM cell. Cell state is modified by the forget gate placed below the cell state and also adjust by the input modulation gate.

The **remember vector** is usually called the **forget gate**.The output of the forget gate tells the cell state which information to forget by multiplying 0 to a position in the matrix. If the output of the forget gate is 1, the information is kept in the cell statesigmoid function is applied to the weighted input/observation and previous hidden state.

The **save vector** is usually called the **input gate**. These gates determine which information should enter the cell state / long-term memory.The **focus vector** is usually called the **output gate**. Out of all the possible values from the matrix, which should be moving forward to the next hidden state?

We are using LSTM because:

➔ LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.

➔ The main advantage is that since the model uses LSTM, the prediction of stock prices will be more accurate.

Chapter 4

Approach

And the way our website works

The steps involved in prediction of prices:

1. To build the stock price prediction model, we will use the **Alpha Vantage API** dataset. This data is maintained daily and is trustworthy.
2. The data fetched will then be stored in a file in CSV format which can be then read by our data.
3. Let's assume the user wants to predict the closing price of a particular stock.
4. Sort the dataset on **Date** and filter-out "Date" and "Close" columns
5. Normalize the new filtered dataset
6. Build and train the LSTM model
7. Save the LSTM model
8. Predict the closing prices of that stock using the generated model.

The workflow of the website:

1. Firstly, the user chooses whichever stock they want to analyse.
2. Then, the user is prompted and asked for a date range.
   (Out of which our model **trains on first 70% days**, and then **predicts for remaining days**.)
3. Using the **Alpha Vantage API**, we fetch the action stock prices for that stock between the start-date and end-date.
4. Using the **Keras' LSTM** neural network function, we predict the stock prices (steps mentioned in the last slide).
5. Using the **Plotly Module**, we created the plots of the stocks .
   Emphasis in the project has been on to ensure that user interface is very easy and allows users to analyse their favorite stock with ease.

Workflow of the website

Mr. Stocker

Stock Price Prediction

Stock

Agilent Technologies Inc. Common Stock

Search Type

Close

Next

Stock Price Prediction

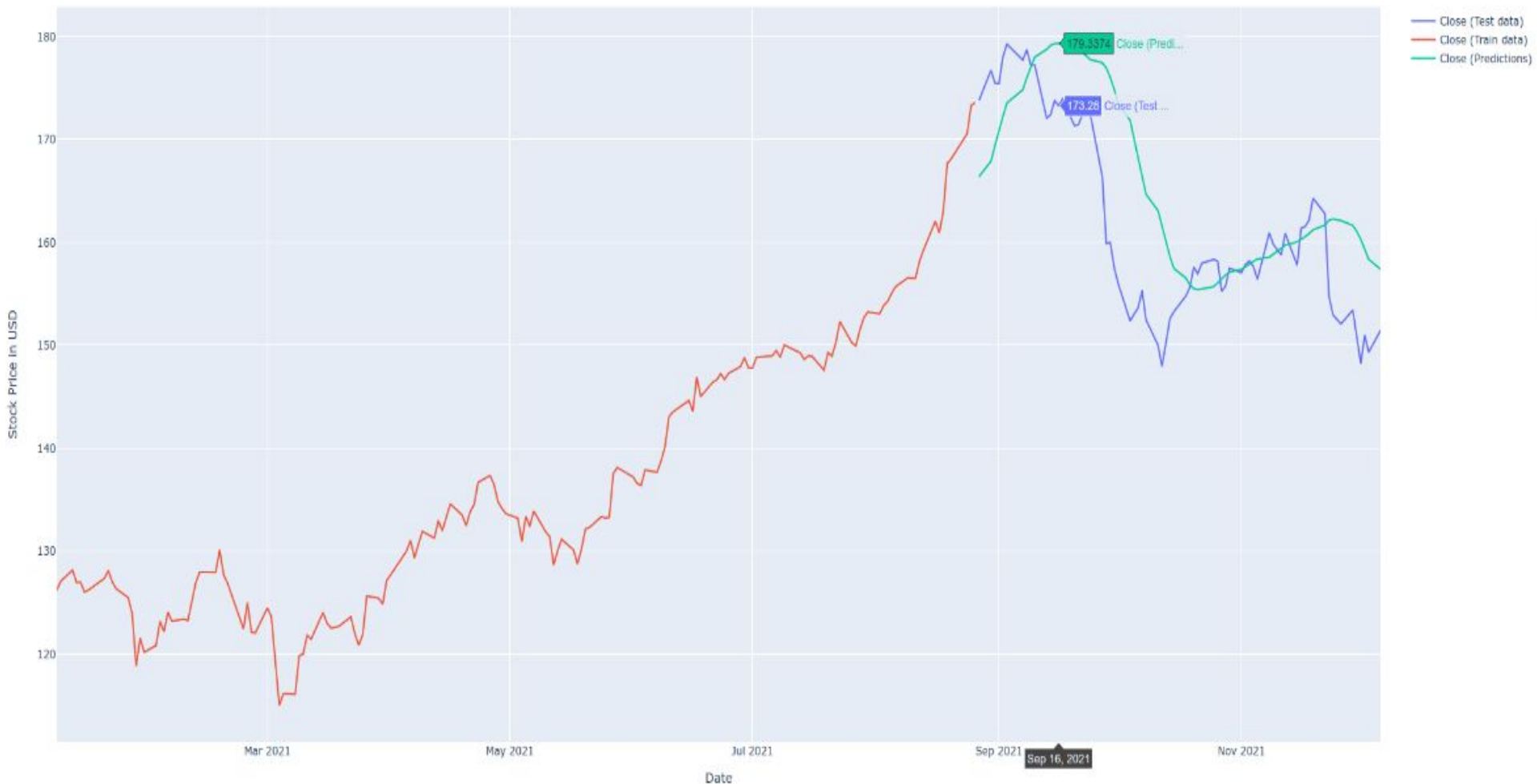## Agilent Technologies Inc. Common Stock
Plot Type: Close

Starting Date

dd/mm/yyyy

Ending Date

dd/mm/yyyy

Predict

Root-mean square error: 7.93

Chapter 5

Results and Plots

Root-mean square error: 0.66

Opening price of the stock for the given period with the **root mean square error** of 0.66

Closing price of the stock for the given period with the **root mean square error** of 0.76

Root-mean square error: 0.61

Minimum price of the stock for the given period with the **root mean square error** of 0.61

Root-mean square error: 0.61

Maximum price of the stock for the given period with the **root mean square error** of 0.61

Chapter 6

Conclusion and Future Work

How is this model helpful?

The issues faced:

Firstly, we tried from genetic algorithm, but the results obtained for stock prediction were **very inaccurate**. The model predicted a straight line but the actual curve was not a straight line.

Then, we had to switch to the LSTM technique to predict the prices of the stocks and we obtained more accurate results as compared to the results obtained from the model which uses genetic algorithm.

We conclude that:

The LSTM Neural Network has given better results than the genetic algorithm. This can be attributed to the fact that RNNs are used for the sequential problems and the stock market is one unique sequential problem where the activities of the previous day largely impact the activities of the next day.

Future Work:

➔ In future, we should try to use different techniques such as CNN, RNN etc. to compare the results obtained from them too and achieve the best possible result.

➔ The code with Genetic Algorithm did not give appropriate results. A ambitious future goal is to make appropriate changes into the algorithm and make it work.

➔ We tried hosting our web-app on the internet, but the free hosting services allow only upto 500 MB of hostage size while our package Bundle was nearly double the size. Hence, to host the web-app would be one of our future goals.

Chapter 7

Individual Contribution

David Garg

- Model using LSTM
- Fetching Data and Regeneration

Konark Sharma

- Django
- Mutation

Ankit Kumar

- Graph plotting using Plotify
- Model using GA and Selection

Sakshi Maheshwari

- Front-end
- Cross-over

Chapter 8

References

- https://www.analyticsvidhya.com/blog/2021/05/stock-price-prediction-and-forecasting-using-stacked-lstm/ (Stock Price Prediction and Forecasting using Stacked LSTM.)

- http://worldcomp-proceedings.com/proc/p2012/GEM4716.pdf (Stock price prediction using genetic algorithms)

- https://link.springer.com/article/10.1007/s13198-021-01209-5 (Integration of genetic algorithm with artificial neural network)

- https://www.datacamp.com/community/tutorials/lstm-python-stock-market (Stock Market Prediction using LSTM)

Thanks!

That's it from our side.

—