

Stock Market Prediction using Sentimental Analysis

Team Predictors

Gaurav Mahadik¹, Pratik Parekh¹, Sakshat Surve¹ and Sumedh Joglekar¹

Department of Computer Science

¹University of North Carolina at Charlotte



1. Abstract

The Stock Market Prediction has been an active interest in the world for quite a few years now. Stocks are majorly based on human behavior or reactions to certain changes. Nowadays, most of the activity of humans can be tracked on social media, And any events that occur affect the human behavior and in turn the stock market. Thus, by extracting patterns and analyzing this data we try to predict the stocks of the certain company.

We are doing this using Machine Learning algorithms based on Neural Networks like Artificial Neural Networks, Long Term Short Term Memory and Genetic Algorithm with Long Term Short Term Memory. Data Extraction is done from various sources such as 'BusinessTimes.com' for news related data, Twitter, and WorldTradinData for stock price data.

2. Motivation

- Predicting the best time of buying or selling is one of the most difficult tasks. An accurate prediction can help investors to acquire more opportunities of gaining profit in the stock exchange.

- Shifts in sentiment on social media and news websites have been shown to correlate with shifts in the stock market.

- In this project, we have predicted the stock market trend using two standard measures, namely, technical analysis and fundamental analysis.

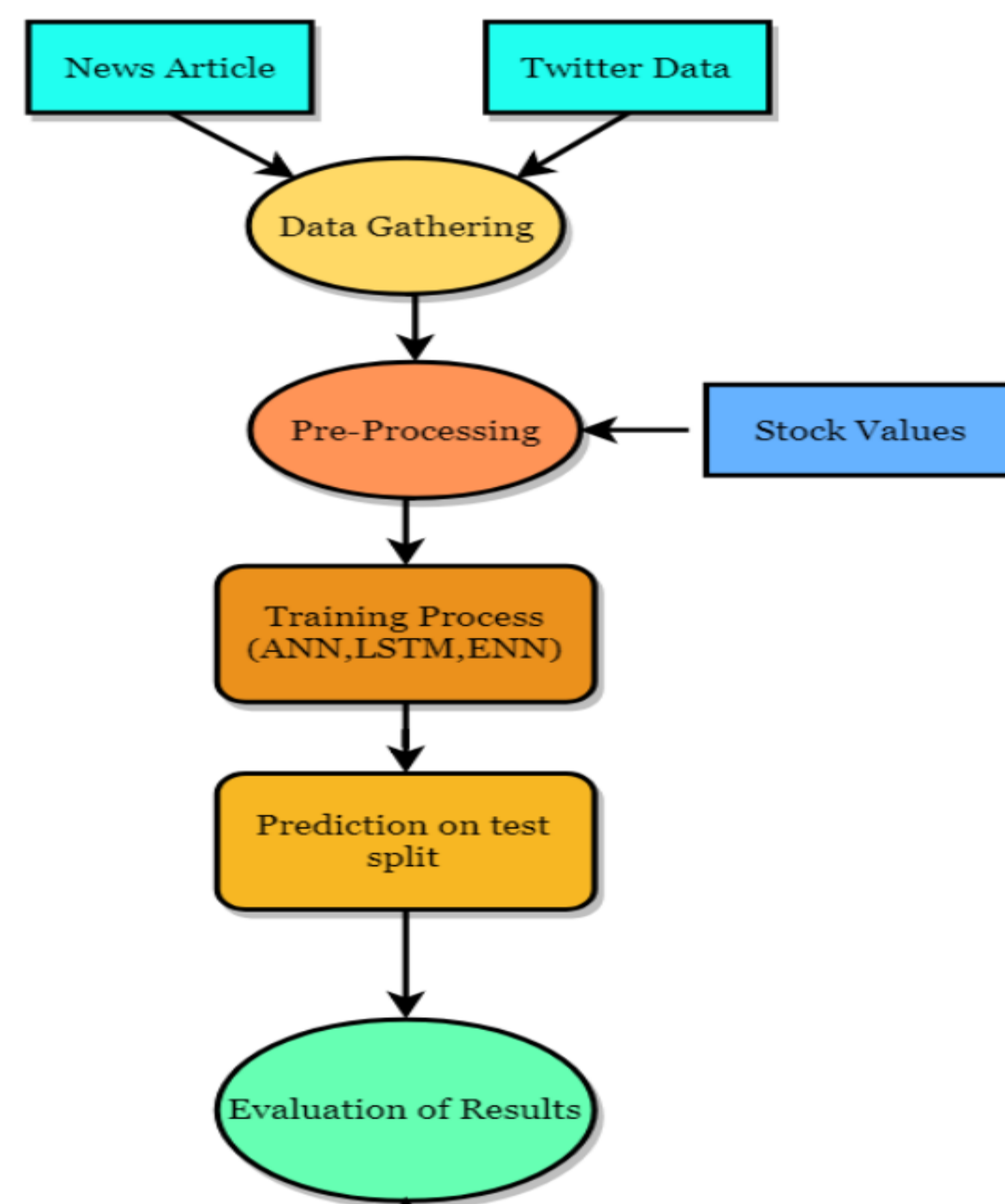
3. Research Challenge

- As the data of sentiment score is calculated using news as well as twitter data, we were facing the difficulty while finding the dataset which will give us the data having a common time frame.

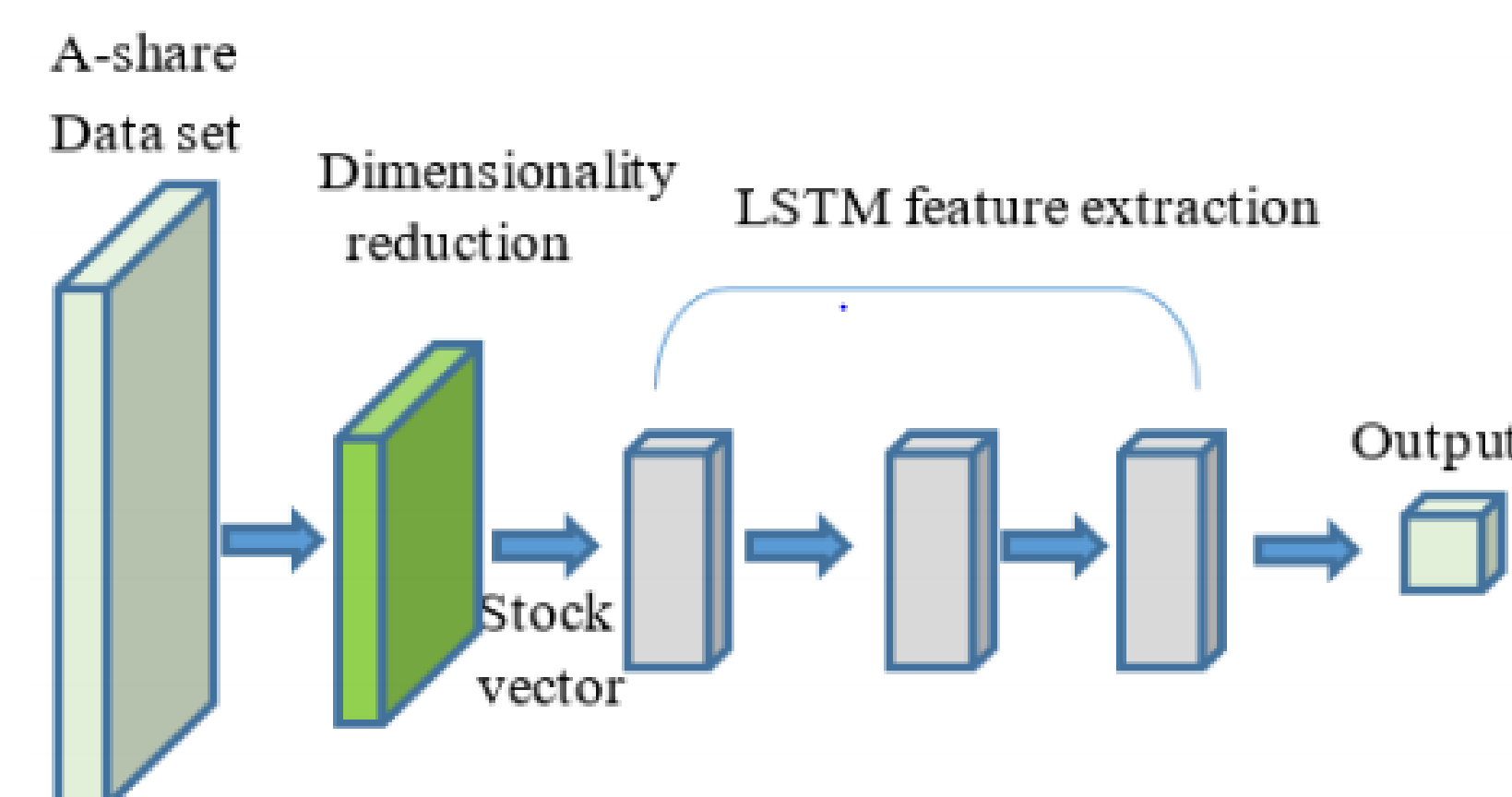
- Understanding of different models, their limitations and selecting the model which is suitable for our project was a difficult task.

- Performance of the model will be evaluated using Root Mean Squared Error (RMSE) and confusion matrix. Based on the results, we will tune our training model by varying hyperparameters to achieve maximum accuracy.

3. Overview



In our project we have collected data from Twitter and News. This data is then pre-processed, after which Semantic Analysis is done on it and a positive, negative, neutral and compound score is calculated. Then this data, with the information of stocks is passed to our training model. The model consists of LSTM, ANN and ENN. After training the model it is tested of the initially split test data and the results are then evaluated.



4. Approach

1. Data Cleaning.

- Performed web crawling using a python script and 'BeautifulSoup' library to collect news from 'businesstimes.com' and 'tweepy' library to collect Twitter data.
- After acquiring headline and article, the data related to a specific company whose stock price we are predicting, we pass the article data to 'Sentiment Intensity Analyzer' (SIA) tool.
- SIA is a part of 'nltk.sentiment.vader' library, it returns neg(negative), pos(positive), neu(neutral), compound(aggregate score) model scores.

2. Model

Initially, a basic artificial neural network is built to predict the stock price movement. In our neural network, the data flows forward to the output continuously without any feedback. Using 'Keras', a four-layer neural network model for predicting the closing price of the stock shares is developed. The input nodes consist of technical variable, while the output layer provides the predicted result based on sentiments variable and true value (actual closing price). To improve the prediction result, a Long Short Term Memory (LSTM) neural network is created. Implemented LSTM model stores previous stock price and makes the future prediction considering this stored information.

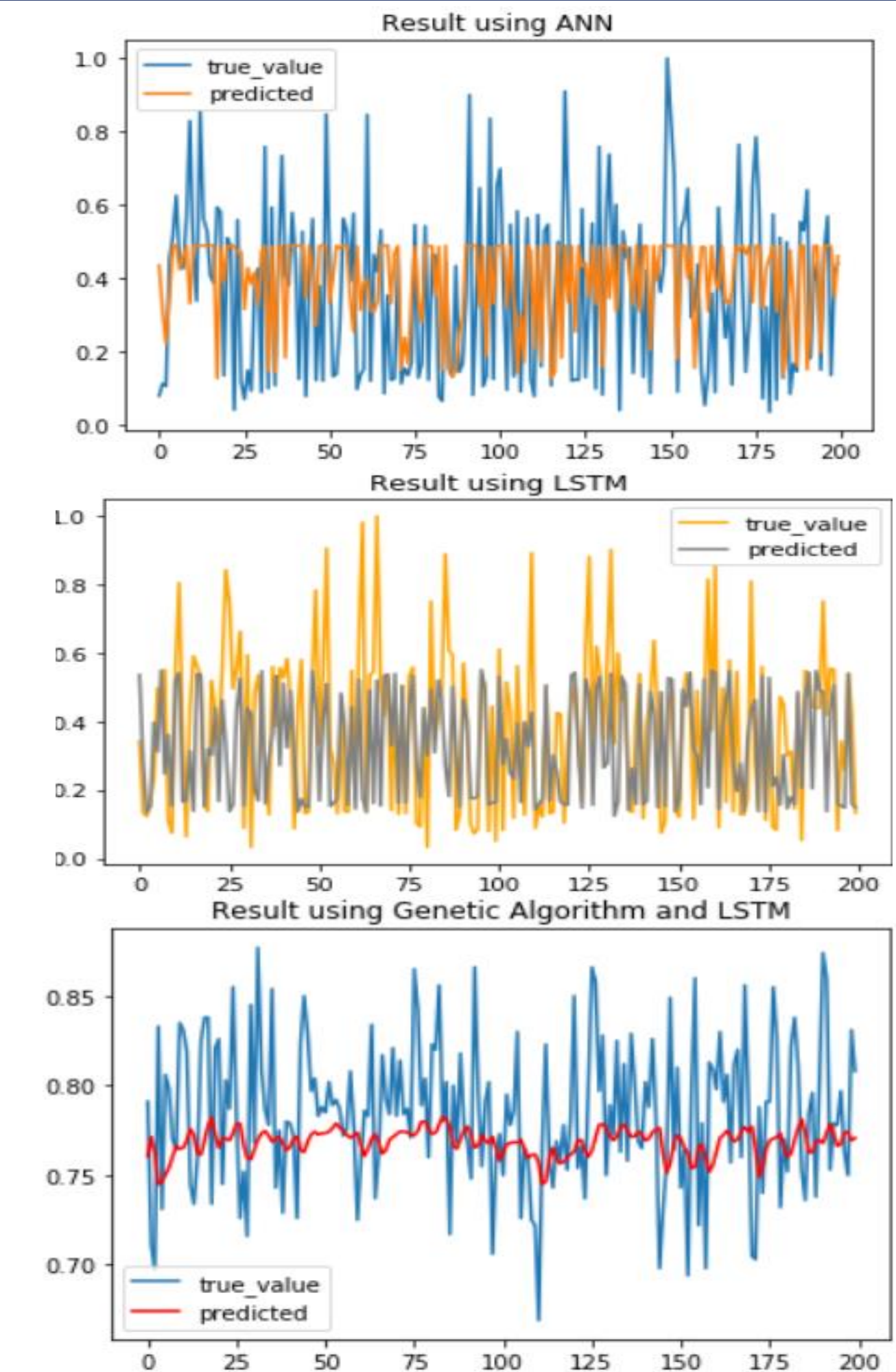
LSTM + GA

The hyperparameters of the network such as weight, bias, learning rate, etc, play an important role in the prediction of the target variable, a genetic algorithm is induced into the LSTM model to optimize the model and select best hyper-parameters for the given model and training data set. Genetic Algorithm computes the fitness score of the initial random population (weights and bias) to select the hyperparameters over which crossover and mutation are performed till the population converges and we get the fittest hyperparameters.

3. Evaluation

- Model's performance is evaluated by implementing the algorithm on test split.
- RMSE score is calculated for each model and compared the results.
- Accuracy of the GA+LSTM model is maximum hence it can be concluded that GA is finding the best parameter correctly.

5. Evaluation



- RMSE value for implemented algorithms:-
 - Simple Artificial Neural Network (ANN) – 0.22081
 - Long /Short Term Memory (LSTM)– 0.21037
 - Evolutionary Neural Network(ENN)- 0.04156

6. References

- Mittal, A. Goel, A. (2011). Stock Prediction Using Twitter Sentiment Analysis. URL <https://medium.com/neuralspace/bayesian-neural-network-series-post-1-need-for-bayesian-networks-e209e66b70b2>.
- Chen, S. He, H. (2018). Stock Prediction Using Convolutional Neural Network. URL https://www.researchgate.net/publication/328754813_StockPrediction_Using_Convolutional_Neural_Network
- Makrehchi, M . Shah, S. Liao, W. (2013). Stock Prediction Using Event-Based Sentiment Analysis. Atlanta, USA. URL <https://ieeexplore.ieee.org/document/6690034>.
- Sphinx. (2018). NLTK Project . URL <https://www.nltk.org/api/nltk.sentiment.html>.

7. Acknowledgements

Dr. Minwoo Jake Lee

Department of Computer Science,
University of North Carolina at Charlotte.