International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 10, Issue 5, May-2025, http://ijmec.com/, ISSN: 2456-4265

Insightforge: Forecasting Market Movements Using Machine Learning Algorithms

[1]Mohammed Omer Khan, [2]Mohammed Hunain Ali Khan, [3]Rayan Bin Ahmed Al Rubaki, [4]Mohammed Rahmat Ali

[1,2,3]B.E Students, Department of CSE, ISLEC, India. [4]Assistant Professor, Department of CSE, ISLEC, India.

ABSTRACT

The prediction of a stock market direction may serve as an early recommendation system for short-term investors and as an early financial distress warning system for long-term shareholders. Forecasting accuracy is the most important factor in selecting any forecasting methods. Research efforts in improving the accuracy of forecasting models have been increasing since the last decade. The appropriate stock selections that are suitable for investment is a very difficult task. The key factor for each investor is to earn maximum profits on their investments.

ecause of dependency on various factors, the stock prices are dynamic, highly noisy, and nonlinear time series data. The stock market prediction has always caught the attention of many analysts and researchers. Predicting stock prices is a challenging problem in itself because of the number of variables which are involved. In the short term, the market behaves like a voting machine but in the longer term, it acts like a weighing machine and hence there is scope for predicting the market movements for a longer timeframe. Application of machine learning techniques and other algorithms for stock price analysis and forecasting is an area that shows great promise. In this project, we first provide a concise review of stock markets, we then focus on some of the research achievements in stock analysis and prediction. We discuss technical, fundamental, short- and long-term approaches used for stock.

This project aims to shed light on the process of web scraping, emphasizing its importance in the new 'Big Data' era with an illustrative application of such methods in financial markets. The work essentially focuses on different scraping methodologies that can be used to obtain large quantities of heterogeneous data in real-time. Automatization of data extraction systems is one of the main objectives pursued in this work, immediately followed by the development of a framework for predictive modeling. applying neural networks and deep learning methods to the data obtained through web scraping. Anyone with nary an idea on the stock market will have a tool that may assist them in the quicks and that is day trading. The economy of our country is not a toy for huge corporations to fiddle with. It is the livelihood of many and retribution against the corporations is due with the first step being the mob short selling of GameStop by Redditors. These methods are applied on 5 years of data retrieved from Yahoo Finance. The results will be used to analyze and predict price.

1. INTRODUCTION

The ultimate goal of our application is to serve retail investors as a third-party investment tool that uses machine learning to help them navigate in the fast-changing stock market. The project aims to introduce and democratize the latest machine learning technologies that is Neural Networks for retail investors. No prediction is 100% accurate.

ISSN: 2456-4265
IJMEC 2025

Therefore, the upper bound and lower bound of the stock prices will be displayed to illustrate the trading range the investors should be looking at. This application serves as a supplementary quantitative tool for investors to see the market at a different perspective with the help of technology. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price. The technical objectives will be implemented in Python. The system must be able to access a list of historical prices. It must calculate the estimated price of stock based on the historical data. It must also provide an instantaneous visualization of the market index

2.

Figure 1:InsightForge: Forecasting Market Movements Using Machine Learning Algorithms

LITERATURE SURVEY STOCK PRICE PREDICTIONS

The art of forecasting the stock prices has been a difficult task for many of the researchers and analysts. In fact, investors are highly interested in the research area of stock price prediction. For a good and successful investment, many investors are keen in knowing the future situation of the stock market. Good and effective prediction systems for stock market help traders, investors, and analyst by providing supportive information like the future direction of the stock market. In this work, we present a recurrent neural network (RNN) and Long Short-Term Memory (LSTM) approach to predict stock market indices. The initial focus of our literature survey was to explore generic online learning algorithms and see if they could be adapted to our use case i.e., working on real-time stock price data. These included Online AUC Maximization, Online Transfer Learning, and Online Feature Selection. However, as we were unable to find any potential adaptation of these for stock price prediction, we then decided to look at the existing systems, analyze the major drawbacks of the same, and see if we could improve upon them. We zeroed in on the correlation between stock data (in the form of dynamic, long-term temporal dependencies between stock prices) as the key issue that we wished to solve. A brief search of generic solutions to the above problem led us to RNN's and LSTM. We decided to use an LSTM neural network to perform stock predictions. We concluded our literature survey by looking at how LSTM can be used to predict the close prices of tech giants like Apple, Microsoft, IBM, Facebook.

NEURAL NETWORKS

A neural network attempts to learn a function that maps the input features to the output predictions, serving as a universal function approximator. It consists of a network of neurons, each of which represents a weighted sum of inputs. Outputs from neurons are fit into activation functions which introduce non-linearity to the system, and then passed to some other neurons. In a typical dense feedforward neural network, the network consists of layers of neurons stacked together, with neurons between individual layers fully connected. Optimization of neural networks is usually done through backpropagation with gradient descent, which essentially propagates the error from the output layer back to the input layer, while computing the gradient of the error against each parameter in the process.

RECURRENT NEURAL NETWORKS

Recurrent neural network is a type of neural network where connections between neurons allow temporal, sequential information to be stored and processed in the network. One typical architecture is formed by feeding the output of the current unit back to the input with a time delay so that the network can use the information in processing the next input. Various techniques have been developed over the years to train such type of network. One of the popular approaches is backpropagation through time (BPTT), whose central idea is to unroll the recurrent network into a feedforward network, where each layer represents a timestep.

The term "recurrent neural network" is used indiscriminately to refer to two broad classes of networks with a similar general structure, where one is finite impulse and the other is infinite impulse. Both classes of networks exhibit temporal dynamic behavior. A finite impulse recurrent network is a directed acyclic graph that can be

unrolled and replaced with a strictly feedforward neural network, while an infinite impulse recurrent network is a directed cyclic graph that cannot be unrolled.

LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. LSTMs are explicitly designed to avoid the longterm dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

All recurrent neural networks have the form of a chain of repeating modules of neural networks. In standard RNNs, this repeating module will have a very simple structure.

LSTMs also have this chain-like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.

In the above diagram, each line carries an entire vector, from the output of one node to the inputs of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denotes its content being copied and the copies going to different locations.

An LSTM has four "gates": Forget, Remember, Learn and Use (or output) It also has three inputs: long-term memory, short-term memory, and E. (E is some training example/new data)

2-REQUIREMENTS ANALYSIS

Technical Feasibility

Feasibility Study Simply put, stock market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The stock market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases. So, there is a factor which causes people to buy and sell. It has more to do with emotion than logic. Because emotion is unpredictable, stock market movements will be unpredictable. Although, we've tried incorporating factor with the most weight in predicting stocks it is invariably futile to be accurate but, it is not so hopeless when predicting trends instead of exact numbers and that's what this project aims to do with the help of LSTM.

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements. These requirements are listed below:

Functional Requirements

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- The system should be able to generate an approximate share price.
- The system should collect accurate data from the NEPSE website in consistent manner. Non-Functional Requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The nonfunctional requirements are essentially based on the performance, information, economy,

control and security efficiency and services. Based on these the non-functional requirements are as follows:

- The system should provide better accuracy.
- The system should have simple interface for users to use.
- To perform efficiently in short amount of time.

3-SYSTEM DESIGN

SYSTEM ARCHITECTURE

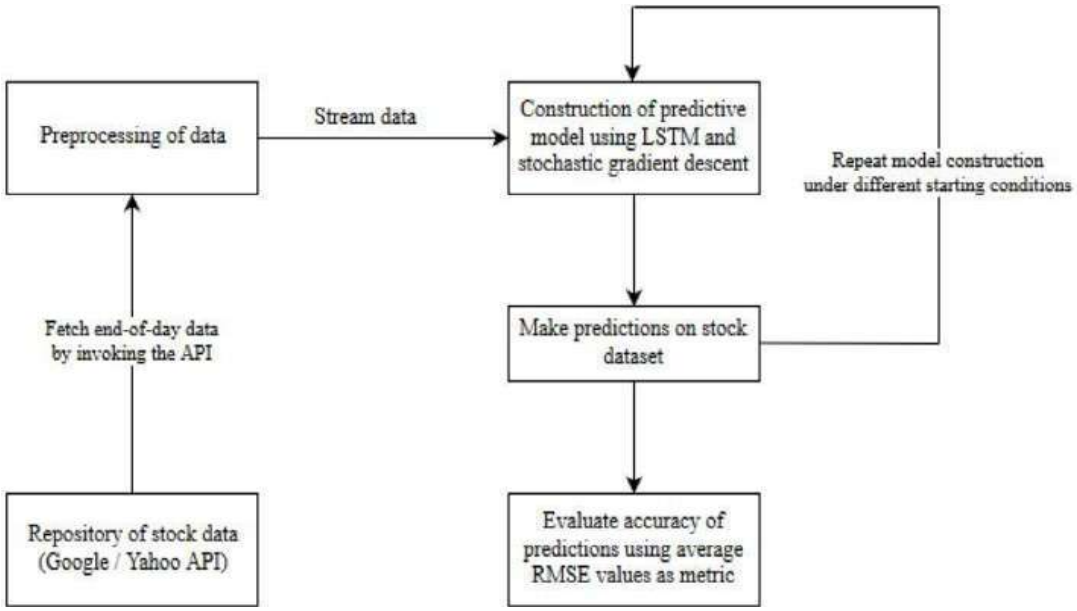


Figure 5: System Architecture

FLOW DIAGRAM

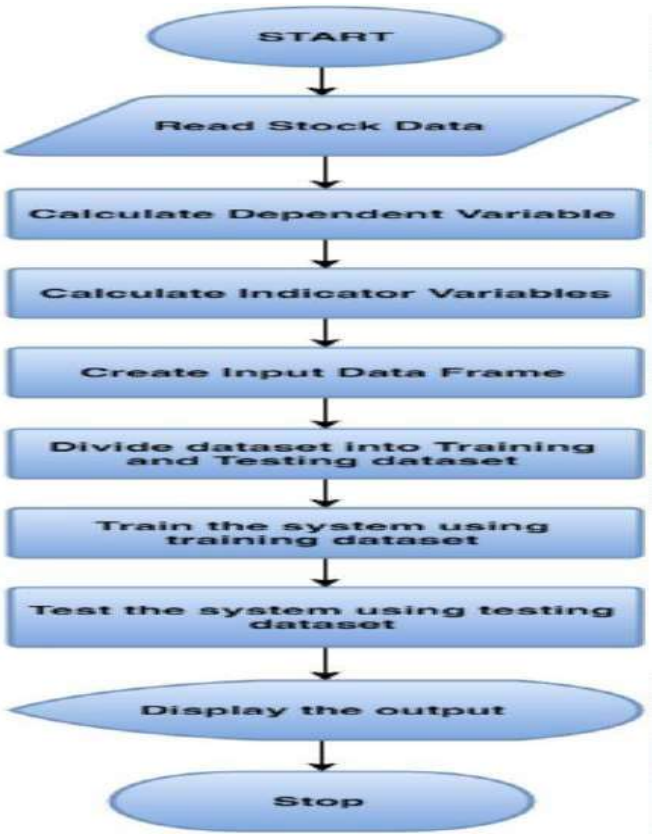


Figure 6: Flow Diagram

4-METHODOLOGY – IMPLEMENTATION

RESEARCH IMPLEMENTATION

Stock Price Data Collection

Data is collected from Yahoo Finance. It offers up to 20 years of daily stock price information on S&P500 stocks. A Python script is written to retrieve stock prices of different stocks automatically. The retrieved stock prices are stored as .csv files in a local folder during development and testing. In deployment, the downloaded stock price data will be transformed into a Graph which is displayed on screen. **Data Preprocessing**

Python scripts written to transform the raw stock prices (.csv files) into feature vectors, for training, predicting and testing respectively. The scripts take the input options and the raw stock prices as inputs and produce the correct features by building the lookback arrays and the moving averages. It concatenates the features into the final feature vectors, which will be passed to the model for training or testing. The 3 scripts share common operations in building a dataset except the output size and the range of dates to build from, so common functions are written to centralize the logic instead of repeating the same indexcalculation-intensive work across functions. NumPy and Pandas are used to build the datasets.

"Numpy" is a library that provides effective ndimensional array data structures as well as functions for array manipulations. It is frequently used for machine learning tasks because it is much for performant than Python lists, as NumPy arrays are implemented as densely packed lists, instead of a dynamic array where the elements are not stored contiguously.

"Pandas" is a popular framework for preprocessing time series data. It has various utilities for reading

raw input files such as .csv and transforming time series data to the correct format. Pandas uses NumPy as the underlying data structure, so it is very convenient to interoperate between the two **Model**

A model base class is used as a common interface for all machine learning models. All models then have their own model class, specifying modelspecific details like methods to build the model, train the model, use the model and save the model. To decouple model configurations from software code to provide flexibility and robustness and save engineering effort as mentioned in 2.2.2, each model is defined by a JSON object, which specifies the model's architecture and hyperparameters with model options and the model inputs with input options. A corresponding model can then be created by passing the object to the model class constructor. The model options specify which machine learning model to use, and the hyperparameters for the model like the number of hidden layers, the number of hidden units, activation functions used, as well as optimization algorithms and loss functions. Some example model options are in Appendix A. Apart from model configurations, the input can also vary, as there are many possible features that could be added to or removed from the feature vectors. The input options specify the features input that a model should expect, like the number of previous stock prices as features and different moving averages. The input options are related to a model in terms of the input format. All neural networks built in Keras requires the input tensor shape for layer shape inference during model building, a Python function is written to calculate the input shape for a given input option.

5-SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components. Sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually

satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6-ANALYSIS AND OUTPUT

ANALYSIS

The factors that are taken into account for change in the closing price of a particular company are: General Index, Price difference, highest value, lowest value, share volume and closing price. We performed analysis on obtained data to establish relation between our output parameters and the selected factors.

Our Predicted model is not 100% accurate but with the fluctuations it is nearer to accurate model which can be very useful for prediction of future closing prices for companies.

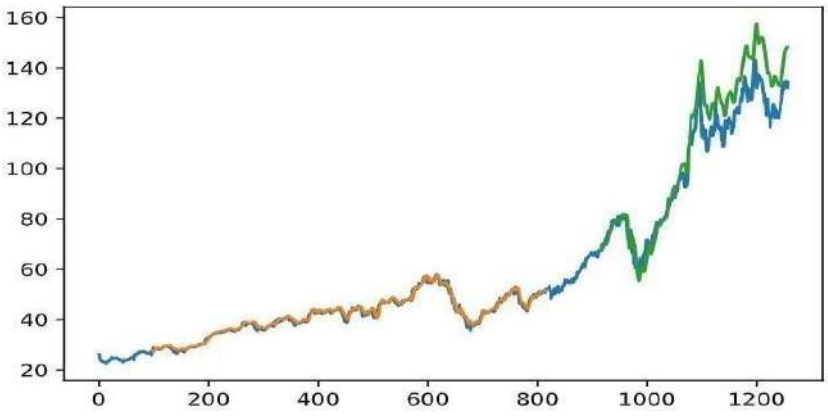


Figure 1: Forecasting Market Movements

OUTPUT

After collection of data the future share price is predicted using Artificial Neural Network model LSTM. The value is then compared the next day with the actual value. The results and deviations of three random companies namely APPLE, FACEBOOK, TESLA and MICROSOFT are illustrated in graphical form below.

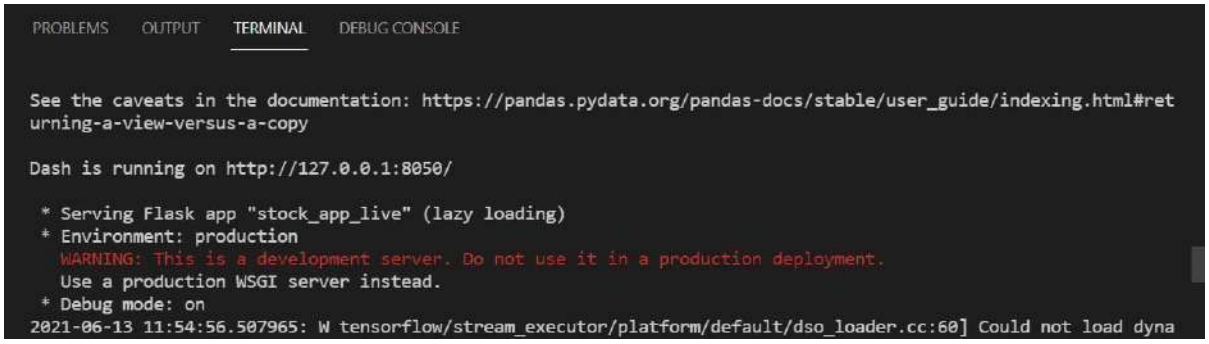


Figure 2: Execution Terminal

Following the link opens a website with 5 tabs:

1. Apple Stock Data
2. Tesla Stock Data
3. Microsoft Stock Data
4. Facebook Stock Data
5. Comparison of Stock Data

1. Apple Stock Data with Prediction

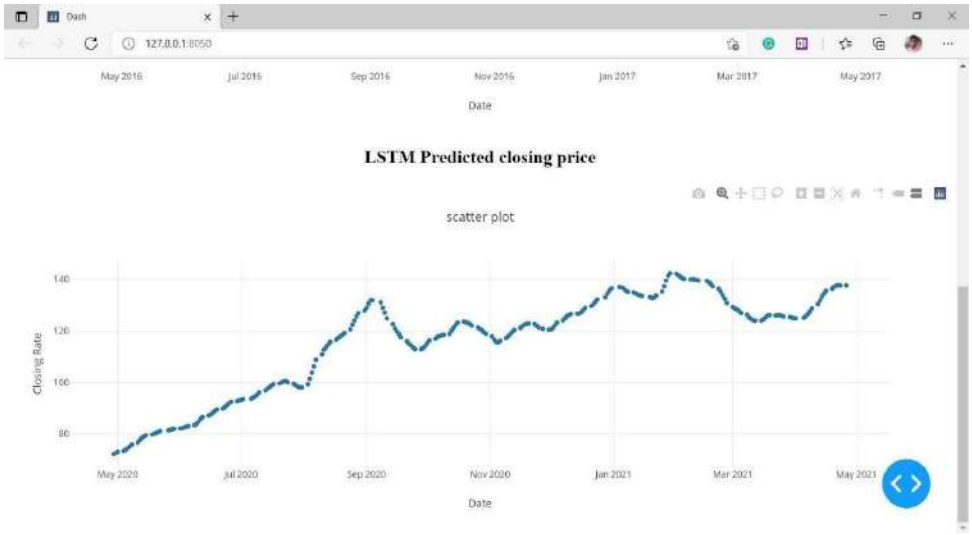


Figure 3: Actual vs LSTM predicted stock data of Apple.

2. Tesla Stock Data with prediction

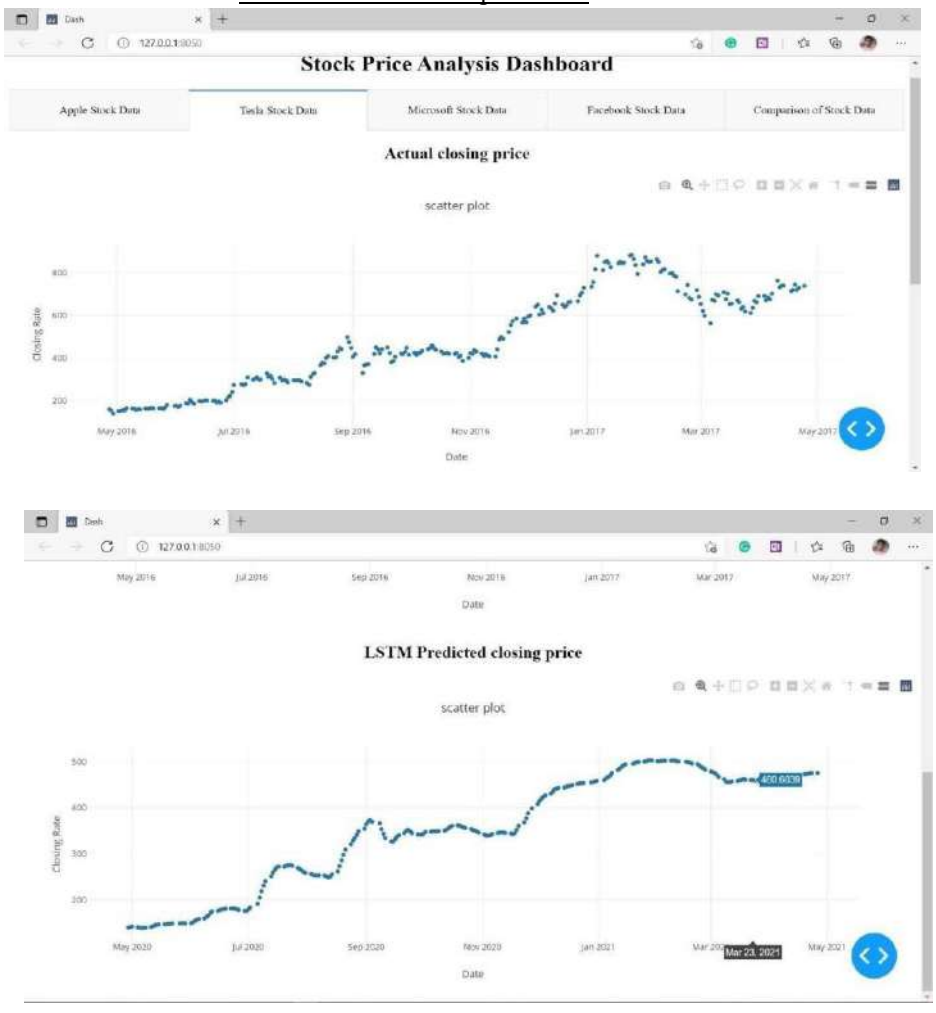


Figure 4: Actual vs LSTM predicted stock data of Tesla.

3. Microsoft Stock Data

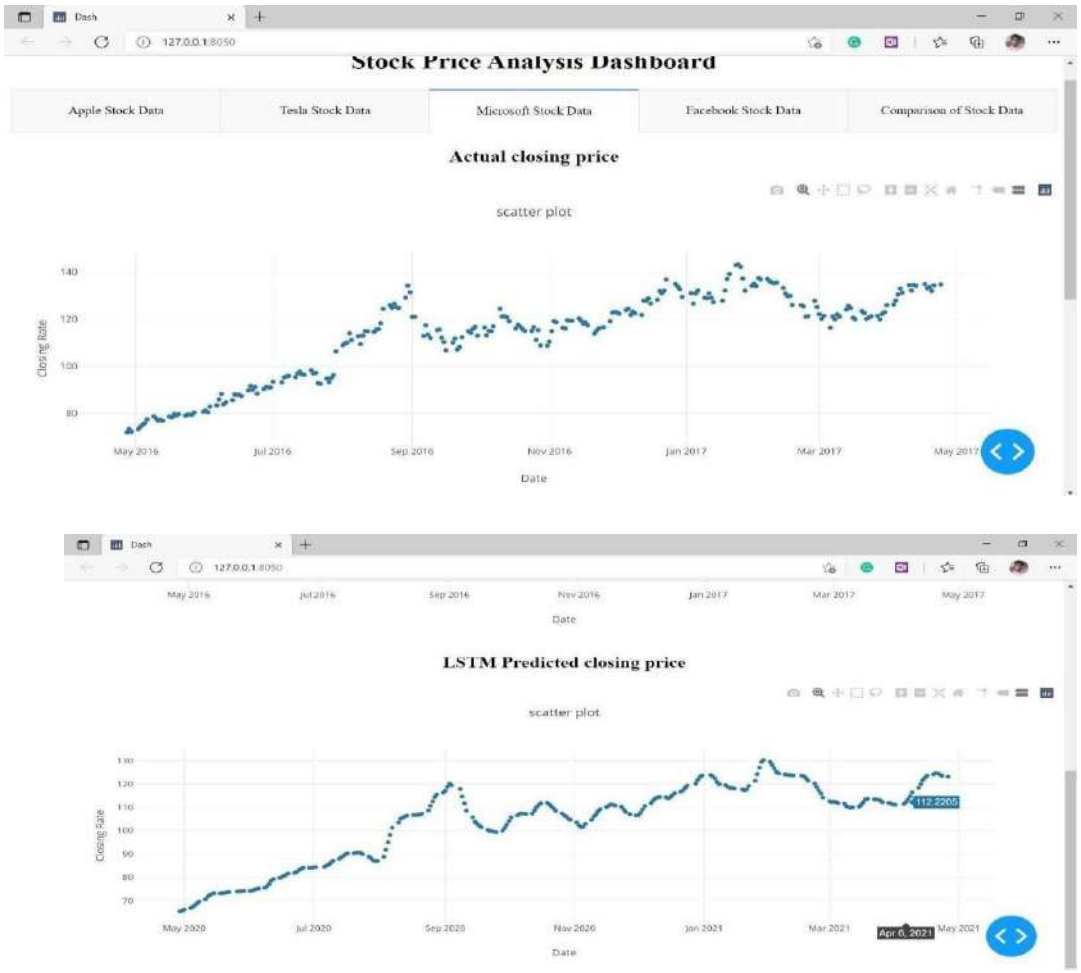
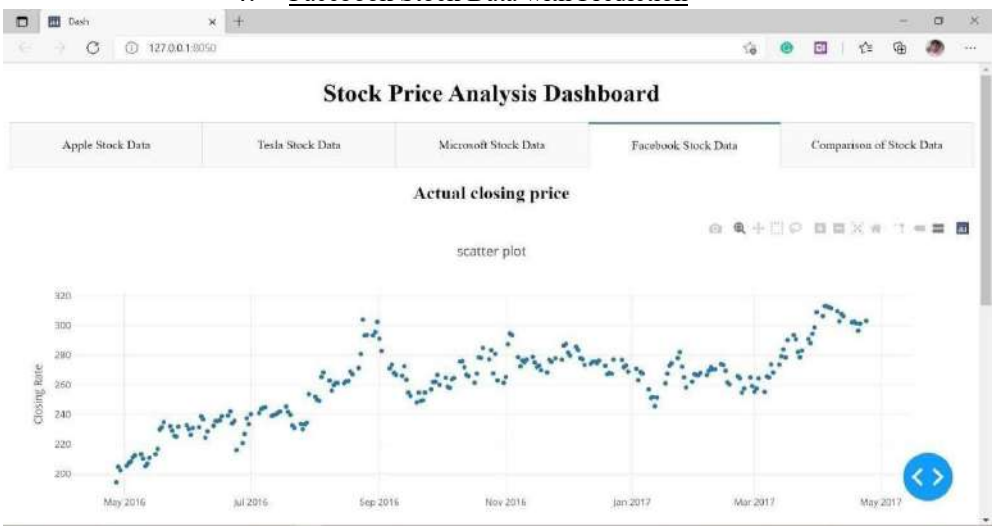


Figure 5: Actual vs LSTM predicted stock data of Microsoft.

4. Facebook Stock Data with Prediction



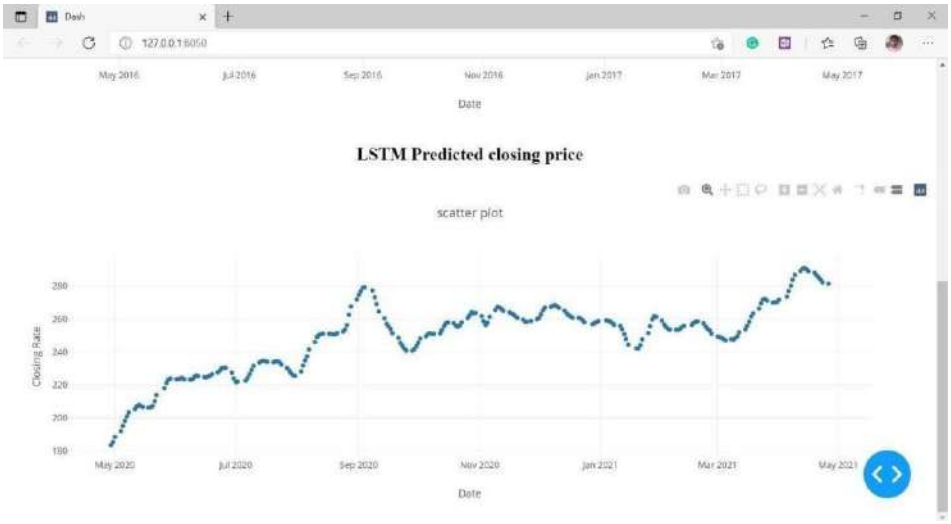


Figure 6: Actual vs LSTM predicted stock data of Facebook.

5. Comparison of stocks with each other:

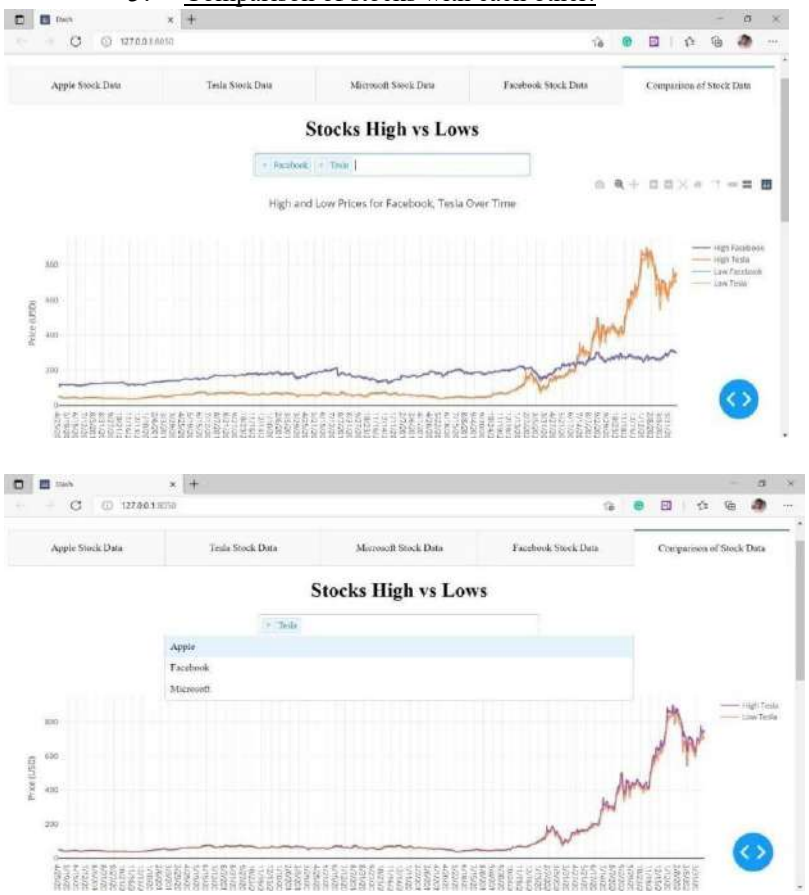


Figure 17: Stocks Comparision

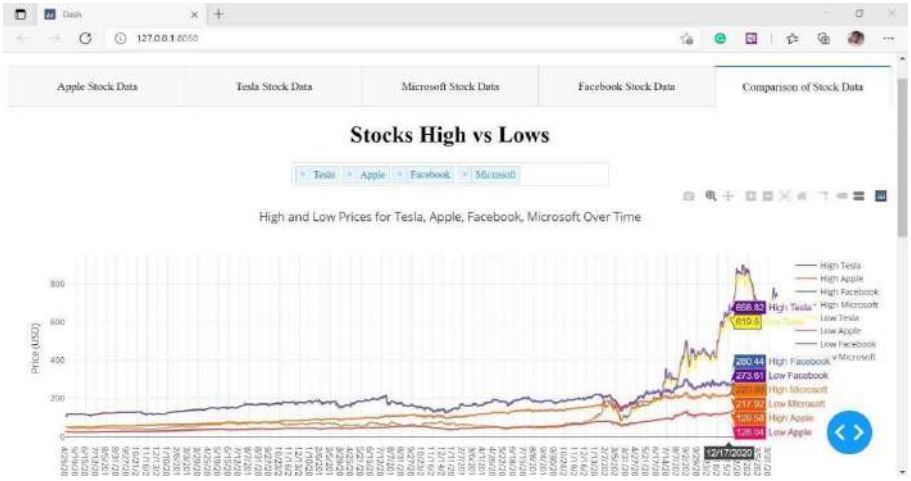


Figure 18: Stocks' High vs Low

7-CONCLUSION AND FUTURE WORKS

CONCLUSION

We implement the application of Artificial Neural Network to the task of stock market prediction and ANN model and salient feature. Our initial analysis show significant correlation between different input parameter. The result obtained in both the cases was fairly accurate. As is indent from fig 8.2.1, 8.2.2 the prediction is fairly accurate unless there is huge and sudden variation in the actual data. On other hand, this also proves the hypotheses that stock market are actually unpredictable. After the phase of prediction and analysis, the result will be displayed to users in the form web pages

FUTURE WORKS

Future scope of improvement

- Potential improvement can be made to our data collection and analysis method.
- Future research can be done with possible improvement such as more refined data and more accurate algorithm.
- Implementation of discussion forums and economic news portal including other sector apart from hydropower and going in national level.

REFERENCES

- https://www.researchgate.net/publication/33134588 3_Stock_Market_Prediction_Using_Machine_Learning
- https://www.researchgate.net/publication/34087437 6_Stock_market_analysis_using _candlestick_regression_and_market_trend_prediction_CKRM
- M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
- K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), 2018, pp. 151156, doi: 10.1109/IACS.2018.8355458

- Forecasting directional movements of stock prices for intraday trading using LSTM and random forests by Pushpendu Ghosh , Ariel Neufeld , Jajati Keshari Sahoo
- S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018, pp. 2787-2790, doi: 10.1109/ICIEA.2018.8398183.
- Alex Sherstinsky, Fundamentals of Recurrent Neural Network (RNN) and Long Short-TermMemory (LSTM) network, https://doi.org/10.1016/j.physd.2019.132306.
- C. R. Madhuri, M. Chinta and V.V.N.V.P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA," 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9202042.C. R. Madhuri, M. Chinta and V.V.N.V.P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA,"
- Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," *Routing Performance Analysis of Infrastructure- less Wireless Networks with Intermediate Bottleneck Nodes*", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023
- Md. Zainlabuddin, "*Wearable sensor-based edge computing framework for cardiac arrhythmia detection and acute stroke prediction*", Journal of Sensor, Volume2023.
- Md. Zainlabuddin, "*Security Enhancement in Data Propagation for Wireless Network*", Journal of Sensor, ISSN: 2237-0722 Vol. 11 No. 4 (2021).
- Dr MD Zainlabuddin, "*CLUSTER BASED MOBILITY MANAGEMENT ALGORITHMS FOR WIRELESS MESH NETWORKS*", Journal of Research Administration, ISSN:1539-1590 | E-ISSN:2573-7104 , Vol. 5 No. 2, (2023)
- Vaishnavi Lakadaram, " Content Management of Website Using Full Stack Technologies", Industrial Engineering Journal, ISSN: 0970-2555 Volume 15 Issue 11 October 2022
- Dr. Mohammed Abdul Bari,Arul Raj Natraj Rajgopal, Dr.P. Swetha ," *Analysing AWSDevOps CI/CD Serverless Pipeline Lambda Function's Throughput in Relation to Other Solution*", International Journal of Intelligent Systems and Applications in Engineering , JISAE, ISSN:2147-6799, Nov 2023, 12(4s), 519–526
- Ijteba Sultana, Mohd Abdul Bari and Sanjay," *Impact of Intermediate per Nodes on the QoS Provision in Wireless Infrastructure less Networks*", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021
- M.A.Bari, Sunjay Kalkal, Shahanawaj Ahamad," *A Comparative Study and Performance Analysis of Routing Algorithms*", in 3rd International Conference ICCIDM, Springer -

978- 981-10-3874-7_3 Dec (2016)

- Mohammed Rahmat Ali,: BIOMETRIC: AN eAUTHENTICATION SYSTEM TRENDS AND FUTURE APLLICATION", International Journal of Scientific Research in Engineering (IJSRE), Volume1, Issue 7, July 2017
- Mohammed Rahmat Ali,: BYOD. A systematic approach for analyzing and visualizing the type of data and information breaches with cyber security", NEUROQUANTOLOGY, Volume20, Issue 15, November 2022
- Mohammed Rahmat Ali, Computer Forensics -An Introduction of New Face to the Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321 8169-453 – 456, Volume: 5 Issue: 7
- Mohammed Rahmat Ali, Digital Forensics and Artificial Intelligence A Study, International Journal of Innovative Science and Research Technology, ISSN:2456-2165, Volume: 5 Issue:12.
- Mohammed Rahmat Ali, Usage of Technology in Small and Medium Scale Business, International Journal of Advanced Research in Science & Technology (IJARST), ISSN:2581-9429, Volume: 7 Issue:1, July 2020.
- Mohammed Rahmat Ali, Internet of Things (IOT) Basics - An Introduction to the New Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-32-36, Volume: 5 Issue: 10
- Mohammed Rahmat Ali, Internet of things (IOT) and information retrieval: an introduction, International Journal of Engineering and Innovative Technology (IJEIT), ISSN: 2277-3754, Volume: 7 Issue: 4, October 2017.
- Mohammed Rahmat Ali, How Internet of Things (IOT) Will Affect the Future - A Study, International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-424874 – 77, Volume: 3 Issue: 10, October 2017.
- Mohammed Rahmat Ali, ECO Friendly Advancements in computer Science Engineering and Technology, International Journal on Scientific Research in Engineering(IJSRE), Volume: 1 Issue: 1, January 2017
- Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay, "*Routing Quality of Service for Multipath Manets, International Journal of Intelligent Systems and Applications in Engineering"*, JISAE, ISSN:2147-6799, 2024, 12(5s), 08–16;
- Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
- Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review ", VOL 2021: ISSUE 08 IS SN

: 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct : 021

- Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh,
Syed Safi Uddin, Dr. Mohammed Abdul Bari,
"Saas Product Comparison and Reviews Using Nlp", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022

- Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali," Smartphone Security and Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS)
; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-
6

- .A.Bari& Shahanawaj Ahamad, "Managing
Knowledge in Development of Agile Software", in International Journal of Advanced Computer Science & Applications (IJACSA), ISSN: 21565570, Vol: 2, No: 4, pp: 72-76, New York, U.S.A., April 2011

- Imreena Ali (Ph.D), Naila Fathima, Prof. P.V.Sudha ,"Deep Learning for Large-Scale
Traffic-Sign Detection and Recognition", Journal of Chemical Health Risks, ISSN:2251-6727/ JCHR (2023) 13(3), 1238-1253

- Imreena, Mohammed Ahmed Hussain, Mohammed Waseem Akram" An Automatic Advisor for
Refactoring Software Clones Based on Machine Learning", Mathematical Statistician and Engineering ApplicationsVol. 72 No. 1 (2023)

- Mrs Imreena Ali Rubeena,Qudsiya Fatima
Fatimunisa "Pay as You Decrypt Using FEPOD Scheme and Blockchain", Mathematical Statistician and Engineering Applications: https://doi.org/10.17762/msea.v72i1.2369 Vol. 72 No. 1 (2023)

- Imreena Ali , Vishnuvardhan, B.Sudhakar," Proficient Caching Intended For Virtual Machines
In Cloud Computing", International Journal Of Reviews On Recent Electronics And Computer Science
, ISSN 2321-5461,IJRRECS/October 2013/Volume-1/Issue-6/1481-1486

- Heena Yasmin, A Systematic Approach for
Authentic and Integrity of Dissemination Data in Networks by Using Secure DiDrip, INTERNATIONAL JOURNAL OF PROFESSIONAL ENGINEERING STUDIES, Volume VI /Issue 5 / SEP 2016

- Heena Yasmin, Cyber-Attack Detection in a Network, Mathematical Statistician and
Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)
Heena Yasmin, Emerging Continuous Integration Continuous Delivery (CI/CD) For Small Teams, Mathematical Statistician and Engineering Appli