

# An Integrated Approach Towards Stock Price Prediction using LSTM Algorithm

1<sup>st</sup> Kummari Vikas

Department Of CSE

Vardhaman College Of Engineering

Hyderabad, India

kummarivikas.11@gmail.com

2<sup>nd</sup> Kamal Kotgire

Department Of CSE

Vardhaman College Of Engineering

Hyderabad, India

kotgire58@gmail.com

3<sup>rd</sup> Bayyapu Sai Prudhvi Reddy

Department Of CSE

Vardhaman College Of Engineering

Hyderabad, India

saiprudhvi369@gmail.com

4<sup>th</sup> Akula Sai Krishna Teja

Department Of CSE

Vardhaman College Of Engineering

Hyderabad, India

krishnalalith999@gmail.com

5<sup>th</sup> H Venkateswara Reddy

Department Of CSE

Vardhaman College Of Engineering

Hyderabad, India

h.venkateswarareddy@vardhaman.org

6<sup>th</sup> Shobarani Salvadi

Department Of Information Technology

Chaitanya Bharathi Institute of

Technology

Hyderabad, India

reddymallashobarani2@gmail.com

**Abstract**— Investing in stock markets is always a bit complex and unpredictable for numerous reasons. This study is made by an extensive study of various machine learning and deep learning algorithms which would help reducing the risk in predicting the stock prices. As part of the study, Tehran Stock Exchange is considered in the sectors of non-metallic minerals, basic metals, and finance for testing new ideas. Various algorithms such as extreme Gradient Boost (XGBoost), Support Vector Machine, Random Forest, Decision Tree, K-Nearest Neighbors, Naive Bayes, Logistic Regression and Artificial Neural Network (ANN) are explored, compared, and analyzed in this work. Apart from the above-mentioned algorithms, this research study has used two powerful deep learning techniques, they are Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). This research study will implement around ten technical indicators that have been gathered over a decade. Stock trading values are utilized to calculate the indicators whereas, binary data is used to convert the indicators. Each prediction model is evaluated by three metrics based on the input way. For continuous data, the evaluation results depict that RNN, and LSTM overcome other prediction models to a greater extent. These deep learning are equally good at evaluating binary data, but this difference hinders due to a considerable enhancement in the model's performance.

**Keywords**—*Recurrent Neural Network, Long Short-Term Memory, Artificial Neural Network, Stock Price.*

## I. INTRODUCTION

Forecasting the stock prices accurately has always been a critical issue for a majority of economic and financial experts. The fluctuating and unpredictable nature of this field made the study even harder [1]. The companies which purchase stocks that are expected to rise and sell them whose stock price is going to fall is the key takeaway in this study. There are two basic steps involved in predicting stock market, first being the fundamental information based on the position of company in market and second, the expenses, annual increase rates which are used for crucial analysis.

Past years analysis such as charts and patterns of how the stock price is varying, it makes the future prediction easier and more accurate [2]. As everyone else, even data scientists face similar difficulties while making predictions of stock market. Due to the riskiness and uncertainty in stock market field, investors face two major problems: one being the political climate in a particular country and second being the interest, opinion of public over stocks in that nation [3].

With careful management, the data regarding stock prices can be used in prediction of future stock values and indexes. The prediction of share market can highly get influenced by the deep learning models and social media sentimental analysis [4] [5].

These techniques are created to automatically learn and analyse from large piles of data. There is no hard and fast rule that can predict the future of a stock because it depends not only on the past analysis but also the pulse of public's view over it. [6] RNN and LSTM are the two deep learning methods applied in this work to predict the nature of stock prices [7]. To feed the models, ten technical indicators are used and further to evaluate the impact of pre-processing, two different ways are used: one being continuous data and second being binary data.

The format of the paper is as follows: The introduction is discussed in Section I. The summary of prior initiatives and associated work on stock prediction, existing system and flowchart is presented in Section II. Section III discusses the proposed approach. Section IV details the implementation of the project which includes the pre-processing and front-end server. section V briefs about the experimental results of the work and finally the conclusion and future scope are covered as part of Section VI.

## II. LITERATURE SURVEY

### A. Stock Market

Over a period, there has been a sharp increase in the number of people doing communications, trading, and transactions of assets in stock market. This became

possible due to spread of financial and economic knowledge that people gained over the course of time. Stock markets act as power houses of economy, they have both the power to rise and dip the economy of a nation. Due to the increased number of investors in the market, it is obvious that there is a chance of anticipation in the fair prices of the stock.

Generally, every nation has a controlling and monitoring body that can oversee the atmosphere of stock market. In India the stock market is regulated by Security Exchange Board of India (SEBI) [8], similarly the US has the Financial Industry Regulatory Authority, which handles the market along with the Securities and Exchange Commission (SEC).

The stock market brings together millions of market players who seek to buy or sell shares, maintaining fair pricing procedures and transaction transparency. Modern computerized stock exchanges operate purely electronically, in contrast to historical stock markets that issued and traded actual share certificates made of paper. In other words, stock markets offer a secure and controlled setting where market participants can securely exchange securities such as stocks and other approved financial instruments with little to no chance of financial loss. The stock markets function as primary and secondary markets in accordance with the regulator's recommendations.

There are multiple studies which have been done in that past to analyze and predict stock prices. [9] The study made by Anand details about comparison of models using various neural networks that are pre-trained. Another study made by Xing, proposes how the Natural Language Processing (NLP) can be employed in analyzing stock prices for future [10]. A Logistic Regression with LSTM model-based study is done by Hari Krishnan details about the prediction of bitcoin prices in future using dataset normalization [11].

### B. Existing System

The main model used to forecast stock prices employs a technique called sentiment analysis, where they consider four nations from the list of developed, emerging, and underdeveloped economies, namely the United States, Hong Kong, Turkey, and Pakistan, and explore the impact of various significant events that occurred between 2012 and 2016 on stock markets.

These tweets of the residents that belong to these nations are served as the basis for Twitter datasets that were generated and used. The dataset used to calculate the sentiment component is made up of 11.42 million tweets. [12] This model uses deep learning algorithms like NLP, SSS, etc.

### C. Flow Chart

As mentioned in the below figure, the flow chart for this study starts with choosing dataset named "Stock Market Analysis + Prediction using LSTM" from Kaggle which is a compilation of data sources from Apple, Amazon, BTC, NTFX, DPZ, S&P, Tesla historical stock prices having more than 6000 rows.

The second step being the cleaning of unnecessary data such as removal of null values by using various pre-defined functions, eliminating repeated records, unwanted columns by using drop function, and finally filling the

missing values using the central tendency measures such as mean, median or mode of the data so that it can become easy to make predictions accurately.

Third step after processing data is to split entire data into train, test data as 80:20 ratio and create models with various machine learning, deep learning algorithms to check their accuracy, compare the metrics.

The final step is to choose the algorithm that has the best metrics in terms of least errors and best accuracy.

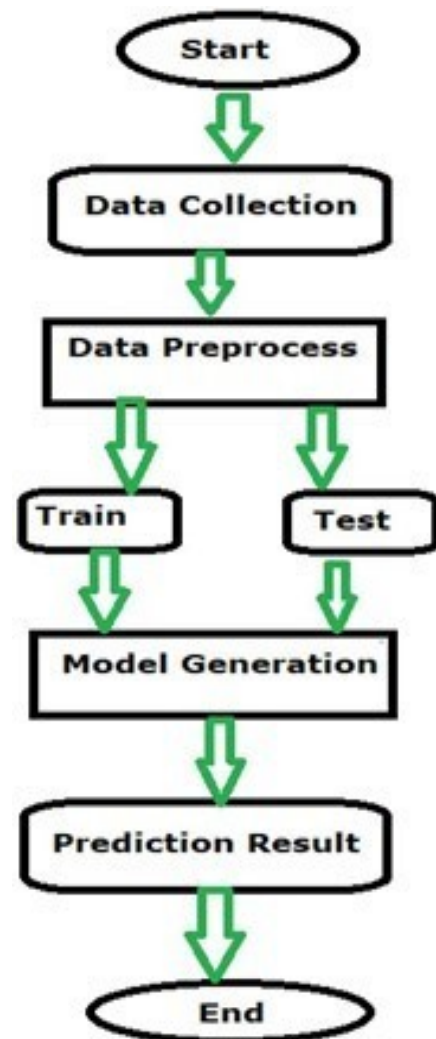


Fig. 1. Flow Chart

## III. PROPOSED METHODOLOGY

The current study revolves around the usage of Recurrent Neural Network and Long-Short Term Memory algorithm whose learning rate varies between 1.0 and 0.001 and has feedback connections in it [13]. The usage of 10 technical indicators is implemented in this work, that are collected over a period of ten odd years. The creation of indicators is made by taking into consideration of stock prices data for past years. The attributes such as dates and open prices were taken into consideration as key attributes. The data thus collected is being turned into binary form (categorical data is converted to either 0 or 1 to make the model accurate) so that can be used easily for the project.

The prediction model implements one of the three metrics while analyzing each input.

#### IV. IMPLEMENTATION

##### A. Pre-Processing

Data Pre-processing is a step of cleaning or processing the random data of the dataset, such that the filtered data can be easily used for the project. While dealing with piles of data in any work, it is quite common that the data used is unorganized and clumsy. It may contain null values, missing values and unwanted data. Hence it is important to remove unnecessary data by dropping the columns that are not useful. The data cleaning and formatting shall be one of the crucial steps while dealing with data related projects. Data which is uncleaned can have noisy data and which can hinder the performance and accuracy of the project. Data pre-processing in this project included removing null values, deletion of unwanted columns, filling up the missed values with central tendency such as mean, median or mode. Plotting of graphs using matplotlib library is done to study how the data in dataset is working.

##### B. Front End Server

The front-end part for this project is made with the help of flask server, which is a server software widely known for processing HTTP requests that are posed through various networks such as private Local Area Network, secured Wide Area Network or through World Wide Web. This server consists of multiple computers clubbed and working effectively to maintain software application while browsing [14]. The execution of flask application must be done with no errors for which the development needs to be flawless. The perfect running of flask application is attained when the connections of client side and deployed side are seamless with flask application. Once the connections are correctly matched then the server that runs and processes HTTP requests will be deployed. The flask which we use in this work is a modern day web application framework that makes use of Jinja template engine and web server gateway interface tools, which is a server-side interface allowing web applications related to python to function [15]. The flask server is incomplete without Web server gateway interface, because a simple server cannot understand the execution of python files and it does not support such applications. Hence during the early 1990's an module of apache is written for the running of python applications. Though the module is not a standard one, it makes easier than the previous when dealing with python scripts in an application. This makes the accomplishing of flask server that helps our front end design easier and simpler to navigate.

#### V. EXPERIMENTAL RESULTS

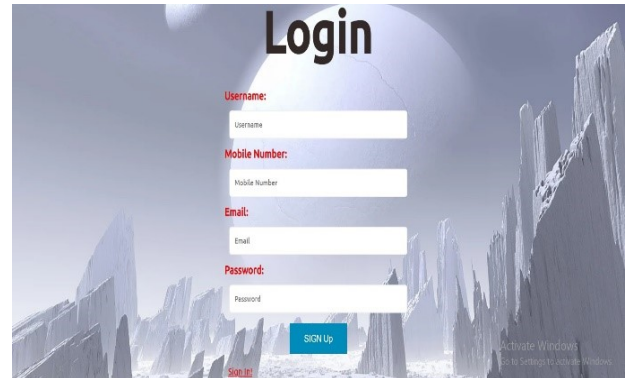


Fig. 2. Login Page

The above figure depicts about pictorial representation of the login page of front-end portion of this project.

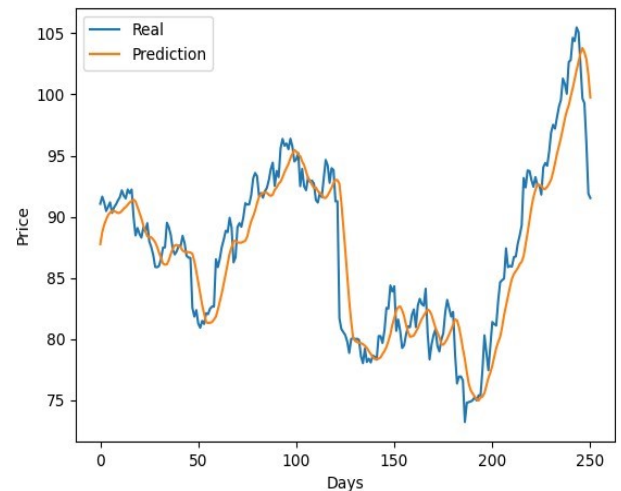


Fig. 3. Prediction Example

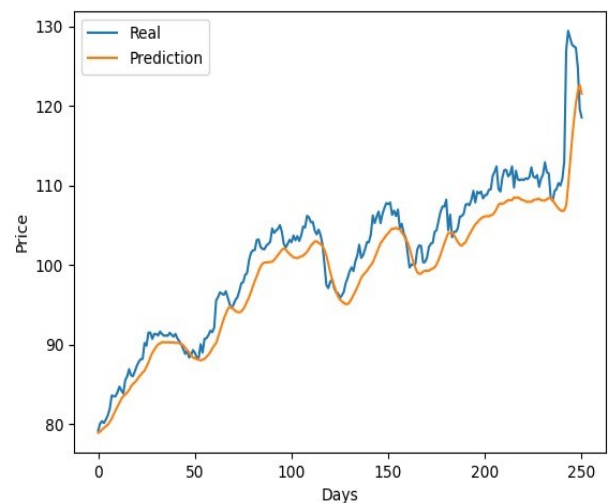


Fig. 4. VAR Stock Predicted

The above figures detail about graphical representation of the prediction results and real values of stock price as the number of days progress for a particular company's stock.

## VI. CONCLUSION

The usage of various machine learning and deep learning methods and their respective behaviors is closely examined in this study. The data is very diverse and is an accumulation for the past ten years stock price records consisting of groups such as the Tehran stock exchange including non-metallic minerals, basic metals, petroleum products and finance industry. The usage of ten odd technical feature is making a significant impact in the behavior of the data. In addition to nine machine learning methods (ADABOOST, XGBOOST, SVC, Naive Bayes, Decision Tree, KNN, Random Forest, ANN, and Logistic Regression), we introduced RNN and LSTM as predictors. The usage of two methods, one continuous data and second, binary data as an input value to the models is made and the implementation of three different classification metrics is done for assessing its performance. This research overall shows that using these techniques the performance of the model is increased significantly.

### A. Future Scope

To have a deeper idea of how the market fluctuates about price variance for a specified stock, we focus to involve sentiment research from various social media into our program in the coming future. This can be accomplished through the discussions happening in Twitter, Facebook and other social media sites with their APIs integrating into our application. This would be an advantage since it can provide users with plethora of parameters and improving accuracy of the model to a greater extent.

## REFERENCES

- [1] I. Pamar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 574-576, doi: 10.1109/ICSCCC.2018.8703332
- [2] 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), pp. 322-327, 2016.
- [3] K. Raza, "Prediction of Stock Market performance by using machine learning techniques", 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), pp. 1-1, 2017.
- [4] K. A. Surya Rajeswar, P. Ramalingam and T. SudalaiMuthu, "Stock Price Prediction Using social media," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675721.
- [5] K. Khare, O. Darekar, P. Gupta and V. Z. Attar, "Short term stock price prediction using deep learning," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017, pp. 482-486, doi: 10.1109/RTEICT.2017.8256643.
- [6] K. J. H. E, M. S. Jacob and D. R., "Stock Price Prediction Based on LSTM Deep Learning Model," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021, pp. 1-4, doi: 10.1109/ICSCAN53069.2021.9526491.

- [7] S. B. Islam, M. M. Hasan and M. M. Khan, "Prediction of Stock Market Using Recurrent Neural Network," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2021, pp. 0479-0483, doi: 10.1109/IEMCON53756.2021.9623206.
- [8] S. Sridhar, S. Mootha and S. Subramanian, "Detection of Market Manipulation using Ensemble Neural Networks," 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020, pp. 1-8, doi: 10.1109/ISCV49265.2020.9204330.
- [9] Anand, C. "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 02 (2021): 122-134.
- [10] F. Z. Xing, E. Cambria and R. E. Welsch, "Natural language based financial forecasting: a survey", Artif. Intell. Rev., vol. 50, no. 1, pp. 49-73, 2018.
- [11] Andi, Hari Krishnan. "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model." Journal of Soft Computing Paradigm 3, no. 3 (2021): 205-217.
- [12] N. N. Reddy, V. K. B P and N. E., "Prediction Stock Price Using Sentimental Analysis Through Twitter Data", UTC from IEEE Xplore, 2020.
- [13] M. A. Istiaque Sunny, M. M. S. Maswood and A. G. Alharbi, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 87-92, doi: 10.1109/NILES50944.2020.9257950.
- [14] M. R. Mufid, A. Basofi, M. U. H. Al Rasyid, I. F. Rochimansyah and A. rokhim, "Design an MVC Model using Python for Flask Framework Development," 2019 International Electronics Symposium (IES), 2019, pp. 214-219, doi: 10.1109/ELECSYM.2019.8901656.
- [15] P. Vogel, T. Klooster, V. Andrikopoulos and M. Lungu, "A Low-Effort Analytics Platform for Visualizing Evolving Flask-Based Python Web Services," 2017 IEEE Working Conference on Software Visualization (VISOFT), 2017, pp. 109-113, doi: 10.1109/VISOFT.2017.13.