Chapter 1

INTRODUCTION

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed.

Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many timeseries prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. In this project, LSTM model is used to predict the stock price

1.1 MOTIVATION FOR WORK

Businesses primarily run over customer's satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances thereby resolving them leads to customer satisfaction as well as trustworthiness of an organization. Hence there is a necessity of an un biased automated system to classify customer reviews regarding any problem. In today's environment where we’re justifiably suffering from data overload (although this does not mean better or deeper insights), companies might have mountains of customer feedback collected; but for mere humans, it’s still impossible to analyse it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. You know you need insights to inform your decision making and you know that you’re lacking them, but don’t know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

1.2 PROBLEM STATEMENT

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory (LSTM)

Chapter 2

PERSPECTIVE

"What other people think" has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which needs to be walked all over in order to attain proper outcomes out of them. In this survey we analysed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

2.1 EXISTING METHODS

2.1.1 STOCK MARKET PREDICTION USING MACHINE LEARNING

The research work done by V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

2.1.2 FORECASTING THE STOCK MARKET USING ARTIFICIAL INTELLIGENCE TECHNIQUES

The research work done by Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering. The weak form of Efficient Market hypothesis (EMH) states that it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools.

2.1.3 THE STOCK MARKET AND INVESTMENT

The research work done by Manh Ha Duong Boriss Siliverstovs. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, the Netherlands and the United Kingdom. Increasing integration of European financial markets is likely to result in even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. Indeed, our vector autoregressive models suggest that the positive correlation between changes equity prices and investment is, in general, significant. Hence, monetary authorities should monitor reactions of share prices to monetary policy and their effects on the business cycle.

2.2 DEEP LEARNING

Although the various forecasting methods mentioned above can achieve a rough forecast of future stock price changes, they cannot achieve accurate forecasts. This is because my country's stock market is still in an immature state of development, and many factors such as national economic conditions, macroeconomic policies, and investors' psychological expectations in the short term will affect stock prices to some extent. Therefore, in future forecasts, various factors should also be considered comprehensively, such as the fundamentals of the operating enterprise, technical indicators, etc., in order to achieve the investment goal of maximum profit or avoidance of maximum risk.

The traditional methods to forecast stock include qualitative econometric methods and machine learning methods. The stock price series can be regarded as complex and time series with much nonlinearity, so using qualitative econometric models cannot achieve the higher forecasting ability. In the machine learning algorithm, due to the unique structure and learning mechanism of neural network, domestic and foreign scholars have gradually increased the research on using it to predict stock prices and trends. In recent years, with the continuous development of deep learning, deep neural network has gradually been applied to the fields of image, speech and finance. It can extract high-level abstract features from a large amount of original data without relying on prior knowledge, and has stronger learning ability and generalization ability. Especially the LSTM neural network, which is a kind of cyclic neural network in the deep learning algorithm, has a special gate structure, and has the characteristics of good selectivity, memory and internal influence of time series, which can process financial data sequences more effectively. Stock forecasts offer new ideas.

Chapter 3

DATASET AND IMPLEMENTATION

3.1 DATASET DETAILS

The dataset consists of the stock historical data from the **Yahoo Finance ,** It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. It also offers some online tools for personal finance management. In addition to posting partner content from other web sites, it posts original stories by its team of staff journalists. Its main task is to be a finance manager for you and provides financial reports for you to use. It is used for finance management and is a medium-size encyclopaedia-type website which provides information about financial news. In short, Yahoo finance is a financial news and management platform owned by Yahoo.

3.2 TOOL & TECHNOLOGIES

3.2.1 PYTHON

The language of select for this project was Python. This was a straight forward call for many reasons.

1. Python as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Google. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.

2. Python is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast.

3. Python is a language as forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial papers should be required and verified. Using python this step is sometimes fairly trivial. However, Python is not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation

did not clearly state the return type of a method, this can’t lead without a lot of trials and error testing otherwise happen in a powerfully written language.

This is a problem that produces learning to use a replacement Python package or library more difficult than it otherwise may be.

3.2.2 NUMPY

Numpy is python package which provide scientific and higher level mathematical abstractions wrapped in python. It is [20] the core library for scientific computing, that contains a provide tools for integrating C, strong n-dimensional array object, C++ etc. It is also useful in random number capability, linear algebra etc. Numpy’s array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy additionally provides basic numerical routines, like tools for locating Eigenvectors

3.2.3 PANDAS

Pandas is a Python library used for working with data sets.It has functions for analyzing, cleaning, exploring, and manipulating data.The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008. Pandas allows us to analyze big data and make conclusions based on statistical theories.Pandas can clean messy data sets, and make them readable and relevant.Relevant data is very important in data science.

Pandas gives you answers about the data. Like:

- Is there a correlation between two or more columns
- What is average value
- Max value
- Min value

3.2.4 M1ATPLOTLIB

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Installation : Windows, Linux and macOS distributions have matplotlib and most of its dependencies as wheel packages.

3.2.5 SCIKIT LEARN

Scikit-learn could be a free machine learning library for Python. It features numerous classification, clustering and regression algorithms like random forests, k-neighbours, support vector machine, and it furthermore supports Python scientific and numerical libraries like SciPy and NumPy. In Python Scikit-learn is specifically written, with the core algorithms written in Cython to get the performance. Support vector machines are enforced by a Cython wrapper around LIBSVM .

3.2.6 TENSORFLOW

In the TensorFlow has an open source software library for numerical computation using data flow graphs. Inside the graph nodes represent mathematical formulae, the edges of graph represent the multidimensional knowledge arrays (tensors) communicated between them. The versatile architecture permits to deploy the computation to at least one or many GPUs or CPUs in a desktop, mobile device, servers with a single API. TensorFlow was firstly developing by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well. Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platforms together with iOS and Android

3.2.7 KERAS

Keras is a high-level neural networks API, it is written in Python and also capable of running on top of the Theano, CNTK, or. TensorFlow. It was developed with attention on enabling quick experimentation. having the ability to travel from plan to result with the smallest amount doable delay is key to doing great research.Keras permits for straightforward and quick prototyping (through user-friendliness, modularity, and extensibility). Supports each recurrent networks and convolutional networks, also as combinations of the 2. Runs seamlessly on GPU and CPU. The library contains numerous implementations of generally used neural network building blocks like optimizers, activation functions, layers, objectives and a number of tools to create operating with text and image data easier. The code is hosted on GitHub, and community support forums embody the GitHub issues page, a Gitter channel and a Slack channel.

3.2.8 COMPILER OPTION

Anaconda is free premium open-source distribution of the R and Python programming languages for scientific computing, predictive analytics, and large-scale process that aim is to modify package managing and deployment. Package versions unit managed by the package management system conda.

3.2.9. JUPITER NOTEBOOK

The Jupyter Notebook is an open-source web application that enables to making and sharing documents that contain visualizations, narrative text, live code and equations. Uses include: data , data visualization, data transformation, statistical modelling, machine learning, numerical simulation, data cleaning and much more.

3.2.10 STREAMLIT

Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers. Data scientists or machine learning engineers are not web developers and they're not interested in spending weeks learning to use these frameworks to build web apps. Instead, they want a tool that is easier to learn and to use, as long as it can display data and collect needed parameters for modeling. Streamlit allows you to create a stunning-looking application with only a few lines of code

Chapter 4

METHODOLOGY

4.1 PROPOSED SYSTEMS

The prediction methods can be roughly divided into two categories, statistical methods and artificial intelligence methods. Statistical methods include logistic regression model, ARCH model, etc. Artificial intelligence methods include multi-layer perceptron, convolutional neural network, naive Bayes network, back propagation network, single-layer LSTM, support vector machine, recurrent neural network, etc. They used Long short-term memory network (LSTM).

Long short-term memory network:

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

Working of LSTM:

LSTM is a special network structure with three "gate" structures. Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate.

The experimental data in this paper are the actual historical data downloaded from the Internet. Three data sets were used in the experiments. It is needed to find an optimization algorithm that requires less resources and has faster convergence speed.

- Used Long Short-term Memory (LSTM) with embedded layer and the LSTM neural network with automatic encoder.

- LSTM is used instead of RNN to avoid exploding and vanishing gradients.

- In this project python is used to train the model, MATLAB is used to reduce dimensions of the input. MySQL is used as a dataset to store and retrieve data.

- The historical stock data table contains the information of opening price, the highest price, lowest price, closing price, transaction date, volume and so on.

- The accuracy of this LSTM model used in this project is 57%.

4.2 LSTM ARCHITECTURE

Forget Gate:

A forget gate is responsible for removing information from the cell state.

- The information that is no longer required for the LSTM to understand things or the information that is of less importance is removed via multiplication of a filter.

- This is required for optimizing the performance of the LSTM network.

- This gate takes in two inputs; h_t-1 and x_t. h_t-1 is the hidden state from the previous cell or the output of the previous cell and x_t is the input at that particular time step.

Input Gate:

Regulating what values need to be added to the cell state by involving a sigmoid function. This is basically very similar to the forget gate and acts as a filter for all the information from hi-1 and x_t. 2. Creating a vector containing all possible values that can be added (as perceived from h_t-1 and x_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1. 3. Multiplying the value of the regulatory filter (the sigmoid gate) to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation.

Output Gate:

The functioning of an output gate can again be broken down to three steps: • Creating a vector after applying tanh function to the cell state, thereby scaling the values to the range -1 to +1. • Making a filter using the values of h_t-1 and x_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function. • Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as a output and also to the hidden state of the next cell

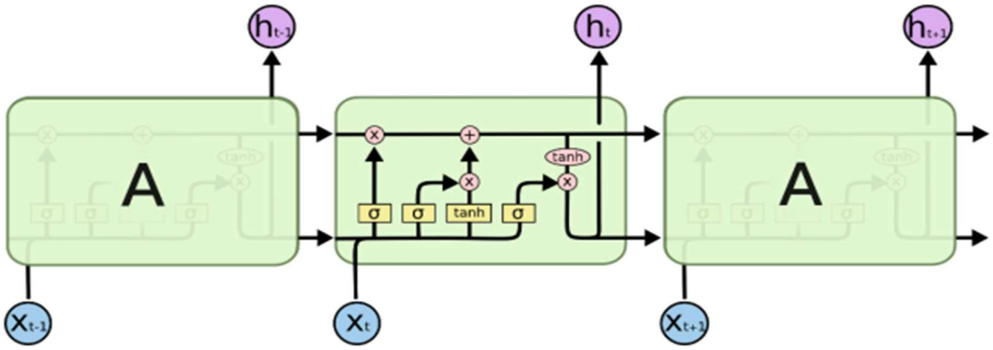


Fig 4.1 LSTM Architecture

- **Mean Absolute Error (MAE)**

Mean Absolute Error (MAE) is the most basic evaluation method, and its expression is as follows

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

- **Mean Square Error (MSE)**

The mean squared error (Mean Squared Error) expression is as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

The value of MSE is inversely proportional to the accuracy of the model. The larger the MSE, the worse the prediction effect of the model.

- **Root Mean Square Error (RMSE)**

Root Mean Square Error (Root Mean Square Error) can be used to calculate the deviation between the observed value and the true value. Because the average index is non-robust, this makes the average error very sensitive to outliers. The expression is as follows:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

4.3 SYSTEM ARCHITECTURE

1) PREPROCESSING OF DATA

Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance, and is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects.

Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running any analysis. Often, data preprocessing is the most important phase of a machine learning project, especially in computational biology.[3] If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Examples of data preprocessing include cleaning, instance selection, normalization, one hot encoding, transformation, feature extraction and selection, etc. The product of data preprocessing is the final training set.

Data preprocessing may affect the way in which outcomes of the final data processing can be interpreted. This aspect should be carefully considered when interpretation of the results is a key point, such in the multivariate processing of chemical data

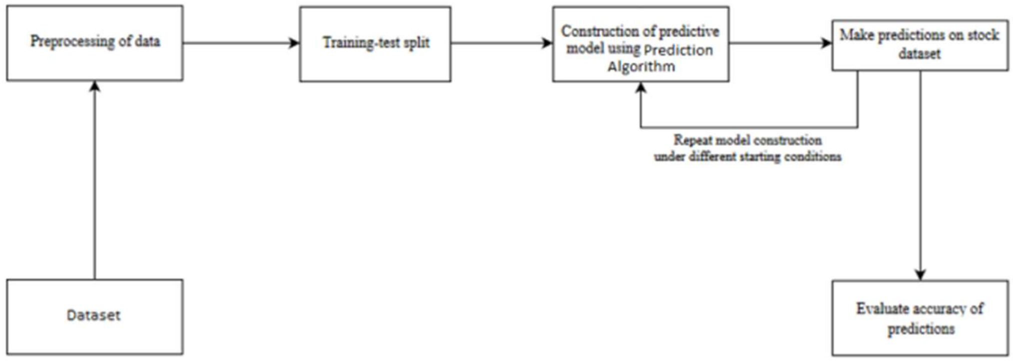

Preprocessing of data
Training-test split
Construction of predictive model using Prediction Algorithm
Make predictions on stock dataset
Repeat model construction under different starting conditions
Dataset
Evaluate accuracy of predictions

Fig. 4.2: Pre-processing of data

Chapter 5

SYSTEM REQUIREMENTS

5.1 HARDWARE REQUIREMENTS:

- RAM: 4 GB
- Storage: 512 GB
- CPU: 2 GHz or faster
- Architecture: 32-bit or 64-bit

5.2 SOFTWARE REQUIREMENTS:

Python

Python is a popular programing language for machine learning and data analysis. Install python and ensure it is properly configured on your system.

Python Libraries:

Install the necessary Python libraries for data manipulation, visualization, and machine learning. The key libraries you will need include:

- NumPy: For numerical computations and array operations.
- Pandas: For data manipulation and analysis.
- Matplotlib: For data visualization.
- Scikit-learn: For machine learning algorithms and preprocessing.
- TensorFlow or Keras: For building and training the LSTM model. Keras is a high-level

Jupyter Notebook or IDE:

Choose an Integrated Development Environment (IDE) or a Jupyter Notebook for coding. Popular options include Jupyter Notebook, JupyterLab, PyCharm, and Visual Studio Code. Install and set up your preferred environment.

5.3 FUNCTIONAL REQUIREMENTS

Functional requirements describe what the software should do (the functions). Think about the core operations. Because the “functions” are established before development, functional requirements should be written in the future tense. In developing the software for Stock Price Prediction, some of the functional requirements could include:

- The software shall accept the tw_spydata_raw.csv dataset as input

- The software should shall do pre-processing (like verifying for missing data values) on input for model training.

- The software shall use LSTM ARCHITECTURE as main component of the software.

- It processes the given input data by producing the most possible outcomes of a CLOSING STOCK PRICE.

5.4 NON-FUNCTIONAL REQUIREMENTS

Product properties

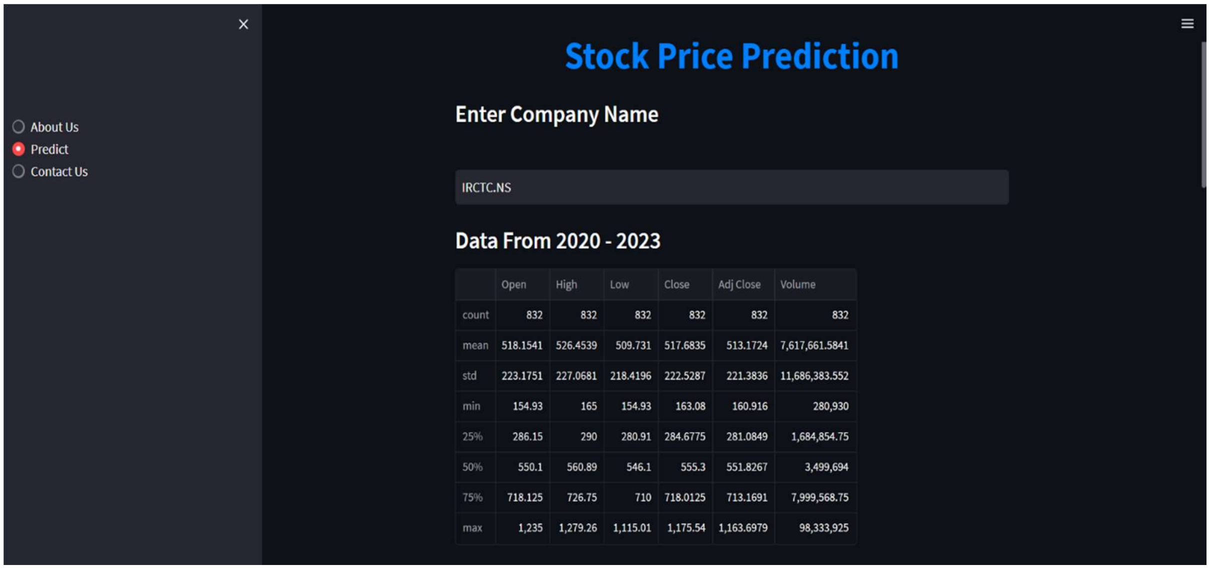
- Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of stock prediction software, for any kind of stock trader and other stakeholders in stock market.

- Efficiency: maintaining the possible highest accuracy in the closing stock prices in shortest time with available data.

Chapter 6

SCREENSHOTS

Overall View



Stock Price Prediction

Enter Company Name

IRCTC.NS

Data From 2020 - 2023

	Open	High	Low	Close	Adj Close	Volume
count	832	832	832	832	832	832
mean	518.1541	526.4539	509.731	517.6835	513.1724	7,617,661.5841
std	223.1751	227.0681	218.4196	222.5287	221.3836	11,686,383.552
min	154.93	165	154.93	163.08	160.916	280,930
25%	286.15	290	280.91	284.6775	281.0849	1,684,854.75
50%	550.1	560.89	546.1	555.3	551.8267	3,499,694
75%	718.125	726.75	710	718.0125	713.1691	7,999,568.75
max	1,235	1,279.26	1,115.01	1,175.54	1,163.6979	98,333,925

Fig 6.1 Overall View

Closing price vs Time chart

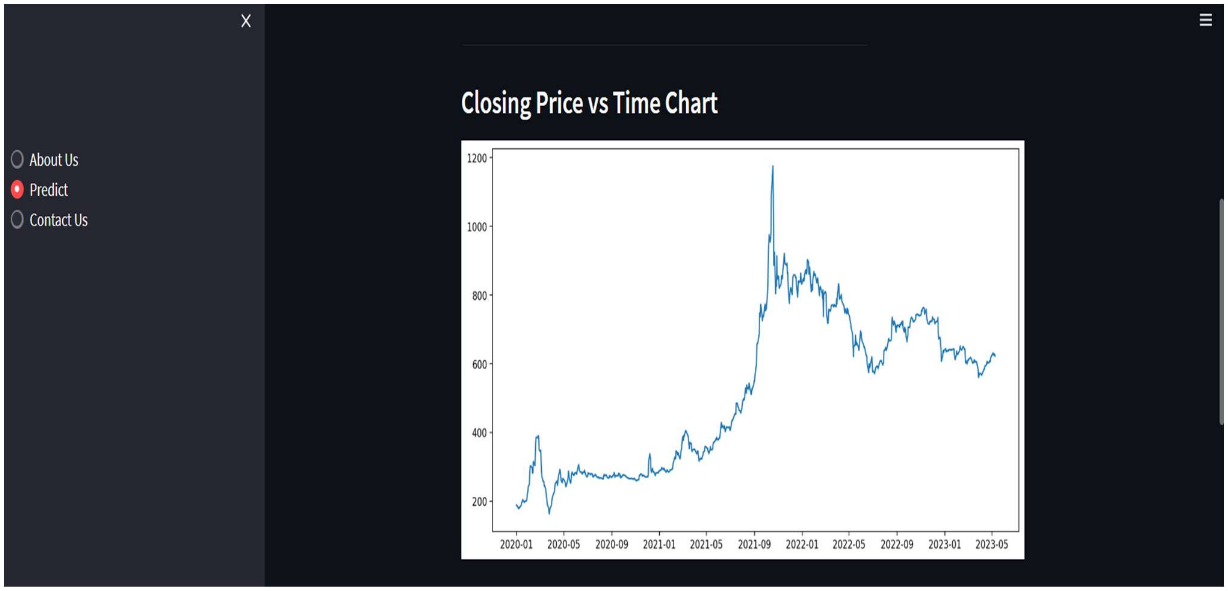


Fig 6.2 Closing price vs Time chart

Closing Price vs Time chart for 100MA

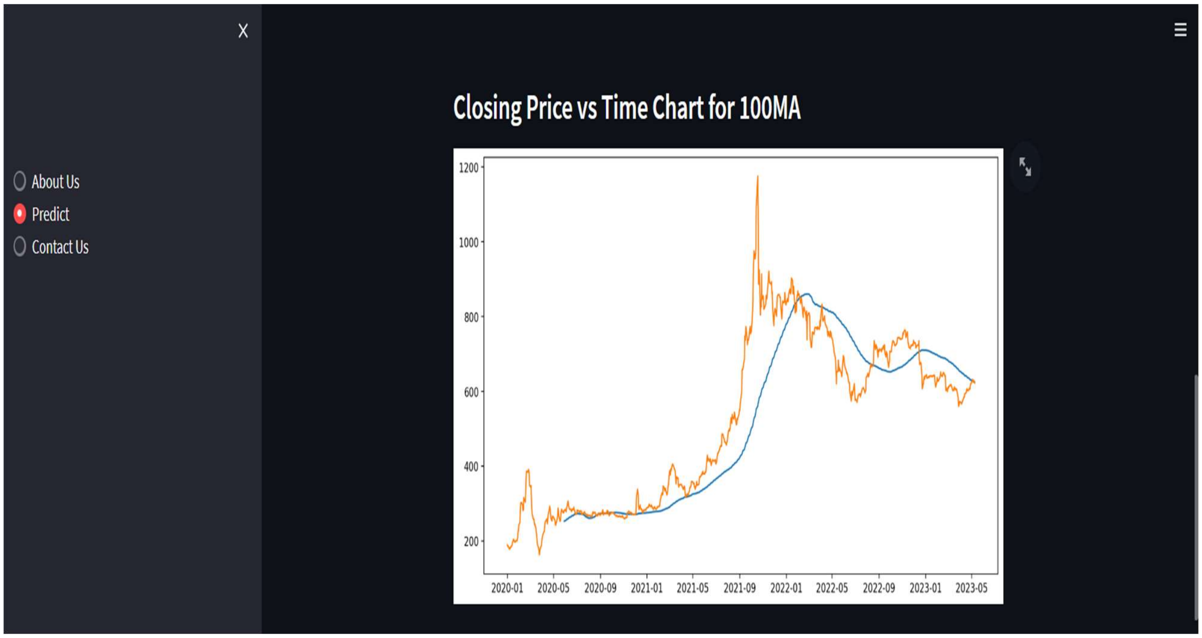


Fig 6.3 Closing Price vs Time chart for 100MA

Closing Price vs Time chart for 100MA and 200MA

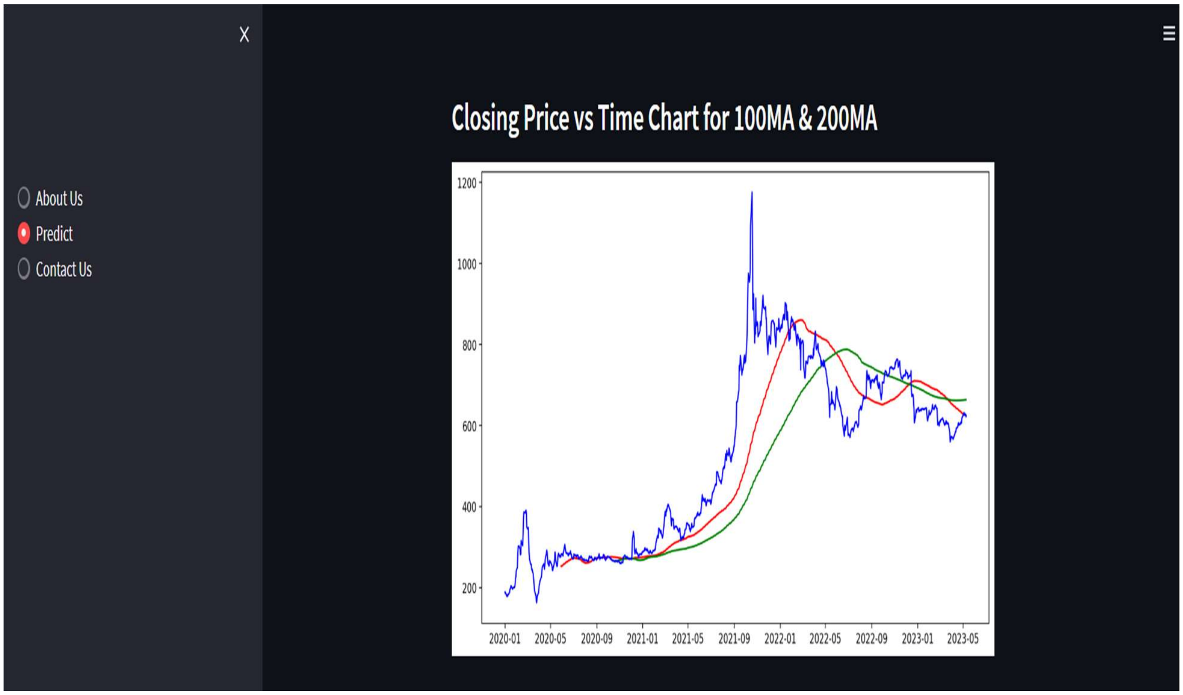


Fig 6.4 Closing Price vs Time chart for 100MA and 200MA

Predicted Price vs Original price

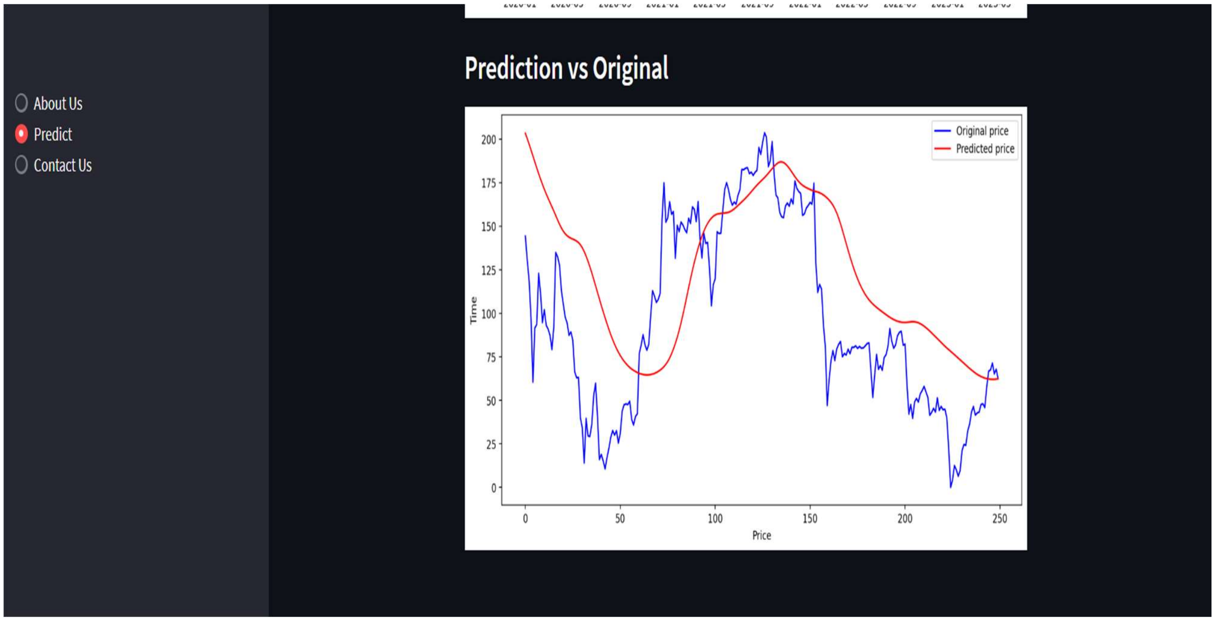
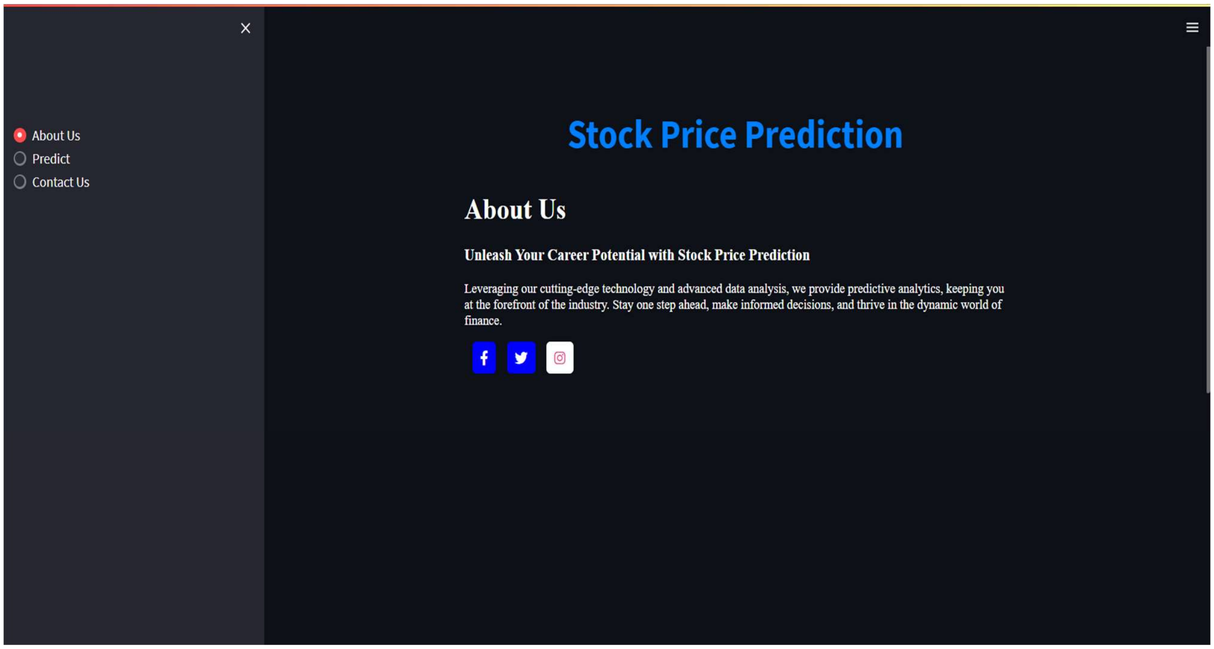


Fig 6.5 Predicted Price vs Original price

About Us



Stock Price Prediction

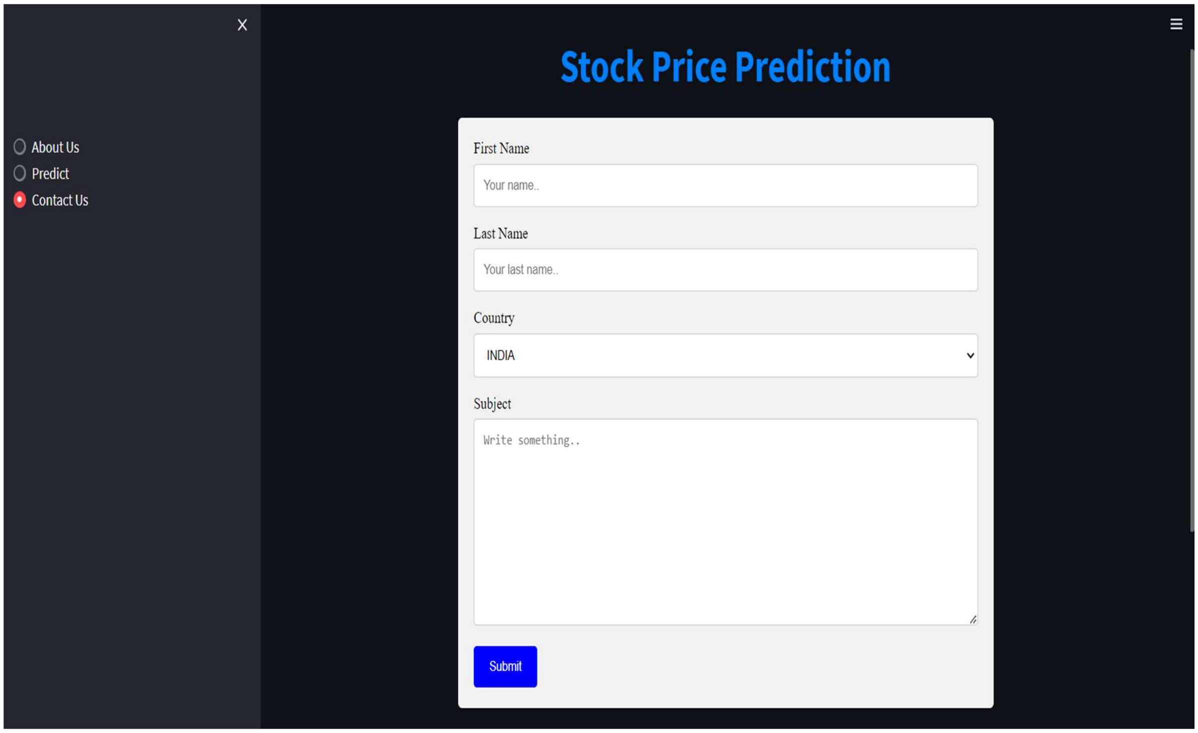
About Us

Unleash Your Career Potential with Stock Price Prediction

Leveraging our cutting-edge technology and advanced data analysis, we provide predictive analytics, keeping you at the forefront of the industry. Stay one step ahead, make informed decisions, and thrive in the dynamic world of finance.

Fig 6.6 About Us

Contact Us



Stock Price Prediction

First Name

Your name..

Last Name

Your last name..

Country

INDIA

Subject

Write something..

Fig 6.7 Contact Us

CONCLUSION

The importance of the stock market to a country's economy will make the types of stock price forecasting methods continue to develop and grow, and will continue to be derived from the development of other disciplines. In the development process of the follow-up forecasting method, it is necessary to continuously explore and deeply study the characteristics of the stock market, so as to make the model closer to reality, expand the applicability of the method, and obtain better forecasting accuracy.

The multi-feature input LSTM model not only takes into account the influence of external factors, but also can process non-linear data, and its prediction performance is better. Through the result of the prediction, we can see that the prediction result of the mixed model is the best.

REFERENCES

[1] Stock Price Prediction Using LSTM on Indian Share Market by Achyut Ghosh, Soumik Bose1, Giridhar Maji, Narayan C. Debnath, Soumya Sen

[2] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, “Predicting Stock Prices Using LSTM” in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.

[3] V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.

[4] Manh Ha Duong Boriss Siliverstovs June 2006 - The Stock Market and Investment

[5] M. Nabipour Faculty of Mechanical Engineering, Tarbiat Modares University, 14115-143 Tehran, Iran; Mojtaba.nabipour@modares.ac.ir - Deep Learning for Stock Market prediction.

[6] Lavanya Ra SRM Institute of Science and Technology | SRM · Department of Computer Science - Stock Market Prediction.