BONAFIDE CERTIFICATE

Certified that the project report titled "**PREDICTION OF STOCK PRICE USING REGRESSION MODEL**" is the bonafide work of **Ms. SHAHANA BEGUM** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature of Student

Ms. SHAHANA BEGUM M

Roll No: 2130MCS0054

Register No: 75721200058

Signature of Guide

DR. PRADEEP S

Assistant Professor

Centre for Distance Education

Anna university

Chennai 600 025

CERTIFICATE OF VIVA-VOCE EXAMINATION

This is to certify that **Ms. SHAHANA BEGUM M (Roll No:** 2130MCS0054**; Register No:** 75721200058**)** has been subjected to Viva-voce-Examination on…… at the Study Centre **Centre for Distance Education, Anna University**.

Internal Examiner	External Examiner
Name :	Name :
Designation :	Designation :
Address :	Address :

Coordinator Study Centre

ACKNOWLEDGEMENT

First and foremost, I would like to thank Dr. P. NIRMAL KUMAR, Professor and Additional Director, Centre for Distance Education, Anna University, Chennai for his constant motivation throughout the course.

I am most grateful to Dr. S. PRADEEP, Assistant Professor, Centre for Distance Education, Anna University, Chennai for his guidance and constant supervision as well as for providing necessary information throughout the completion of this project. I also thank my review committee member Mrs. N. SARASWATHI, Assistant Professor, Centre for Distance Education, Anna University, Chennai for her moral support.

I am also grateful to my parents, friends and colleagues for their timely support and constant words of encouragement. Last but not the least; I extend my thanks to The Almighty.

(Ms. SHAHANA BEGUM M)

ABSTRACT

The prediction of stock prices is a challenging and important task in financial markets. Accurate forecasting of stock prices can provide valuable insights to investors and help them make informed decisions. This paper focuses on the prediction of Tata and Reliance stock prices using regression models.

Regression analysis is employed as a powerful tool to model the relationship between the historical stock prices and relevant factors that influence stock price movements. The paper explores various Regression models, including Linear Regression, Random Forest Regression, and ARD Automatic Relevance Determination Regression, to predict the future stock prices of Tata and Reliance. Data pertaining to the historical stock prices of Tata and Reliance, as well as other relevant factors such as market indices, news sentiment, and economic indicators, are collected and used as input variables for the regression models. Feature engineering techniques may be applied to extract meaningful features and improve the predictive performance of the models.

சுருக்கம்

நிதிச் சந்தைகளில் பங்கு விலைகளைக் கணிப்பது சவாலான மற்றும் முக்கியமான பணியாகும். பங்கு விலைகளின் துல்லியமான முன்னறிவிப்பு முதலீட்டாளர்களுக்கு மதிப்புமிக்க நுண்ணறிவுகளை வழங்குவதோடு, தகவலறிந்த முடிவுகளை எடுக்க அவர்களுக்கு உதவும். பின்னடைவு மாடல்களைப் பயன்படுத்தி டாடா மற்றும் ரிலையன்ஸ் பங்குகளின் விலைகளை கணிப்பதில் இந்த கட்டுரை கவனம் செலுத்துகிறது.

பின்னடைவு பகுப்பாய்வு வரலாற்று பங்கு விலைகள் மற்றும் பங்கு விலை நகர்வுகளை பாதிக்கும் தொடர்புடைய காரணிகளுக்கு இடையிலான உறவை மாதிரியாக்க ஒரு சக்திவாய்ந்த கருவியாக பயன்படுத்தப்படுகிறது. டாடா மற்றும் ரிலையன்ஸ் நிறுவனங்களின் எதிர்கால பங்கு விலைகளை கணிக்க, லீனியர் ரிக்ரஷன், ரேண்டம் ஃபாரஸ்ட் ரிக்ரஷன் மற்றும் ஏஆர்டி ஆட்டோமேட்டிக் ரிலீவன்ஸ் டெடர்மினேஷன் ரிக்ரஷன் உள்ளிட்ட பல்வேறு பின்னடைவு மாதிரிகளை இந்த தாள் ஆராய்கிறது. டாடா மற்றும் ரிலையன்ஸின் வரலாற்று பங்கு விலைகள் தொடர்பான தரவு, அத்துடன் சந்தை குறியீடுகள், செய்தி உணர்வு மற்றும் பொருளாதார குறிகாட்டிகள் போன்ற பிற தொடர்புடைய காரணிகள் சேகரிக்கப்பட்டு, பின்னடைவு மாதிரிகளுக்கான உள்ளீட்டு மாறிகளாகப் பயன்படுத்தப்படுகின்றன. அர்த்தமுள்ள அம்சங்களைப் பிரித்தெடுக்கவும், மாதிரிகளின் முன்கணிப்பு செயல்திறனை மேம்படுத்தவும் அம்ச பொறியியல் நுட்பங்கள் பயன்படுத்தப்படலாம்.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATION

ARD - AUTOMATIC RELEVANCE DETERMINATION

ER - ENTITY RELATIONSHIP

LSTMS- LONG-SHORT TERM MEMORY

GAN - GENERATIVE ADVERSARIAL NETWORK

DJIA - DOW JONES INDUSTRIAL AVERAGE

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

Predicting stock prices is a complex and crucial endeavour in the world of finance and investment. This introduction will provide an overview of using regression models to forecast the stock prices of two prominent Indian companies, Tata and Reliance. Tata and Reliance are among the largest conglomerates in India, with diversified business interests, making their stock prices sensitive to a myriad of factors, including market conditions, company performance, industry trends, and macroeconomic indicators.

In this context, regression models serve as valuable tools to understand and quantify the relationship between the stock prices of these companies and a set of independent variables. Common types of regression models employed in stock price prediction include linear regression, ARD regression, Random Forest regression. These models are designed to capture various levels of complexity in the relationships between stock prices and their determinants. However, it's essential to emphasize that stock price prediction is inherently uncertain and influenced by many unpredictable factors. Consequently, these regression models should be used in conjunction with other forms of analysis and should not be the sole basis for investment decisions. To embark on this journey, one must first collect and preprocess historical stock price data along with relevant independent variables, followed by model training and evaluation using metrics like Mean Squared Error, Root Mean Squared Error, and R-squared to assess the model's predictive performance.

1.2 LITERATURE SURVEY:

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that

may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis. A summary is a recap of important information about the source, but a synthesis is a re-organization or reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them.

Title: *Research on Amazon's stock price forecasting based on arbitrage pricing model based on big data*

Author: *Haocheng Du*

Year:*2022*

The generation of big data is based on the network data generated when people use Internet information systems to interact. Big data can reflect the general laws of specific fields and industries, provide more accurate references for decision makers and managers, and provide people with better Data services. Arbitrage pricing model has long been widely quoted by scholars as an alternative theory to capital asset pricing model, which is used to make a regression analysis on Amazon's stock price in this study. In our study, we aim to construct an arbitrage pricing model to make a regression analysis on Amazon's stock price, which is demonstrated to have a higher prediction accuracy and better fitting degree compared with the self-coding network. First, six relevant indicators are selected to conduct establishment of arbitrage pricing model. Then, a self-coding neural network is constructed to estimate the influence coefficients of each factor on Amazon's stock price, which are compared with the results obtained by regression analysis.

Title: *Time Series Prediction for Stock Price and Opioid Incident Location*

Author: *Kevin Thomas*

Year: *2019*

Time series forecasting is the prediction of future data after analysing the past data for temporal trends. This work investigates two fields of time series forecasting in the form of Stock Data

Prediction and the Opioid Incident Prediction. In this thesis, the Stock Data Prediction Problem investigates methods which could predict the trends in the NYSE and NASDAQ stock markets for ten different companies, nine of which are part of the Dow Jones Industrial Average (DJIA). A novel deep learning model which uses a Generative Adversarial Network (GAN) is used to predict future data and the results are compared with the existing regression techniques like Linear, Huber, and Ridge regression and neural network models such as Long-Short Term Memory (LSTMs) models.

1.3 EXISTING SYSTEM:

Stock trend prediction is a hot issue in the Fintech field. Effective stock profiling is challenging due to highly non-stationary dynamics and complex interplays. Existing methods usually regard each stock independently or detect simplistic homogeneous structures. Practically, stock correlation originates from diverse aspects and underlying relationship signals are implicit in comprehensive graphs. Besides, RNNs are extensively used to simulate stock volatility while inadequate in capturing fine-granular patterns across local time snippets. To this end, in this paper we propose HATR-I, a Hierarchical Adaptive Temporal-Relational Interaction model to characterize and predict stock evolutions. Specifically, we grasp short- and long-term transition regularities of stock dynamics based on cascaded dilated convolutions and gating paths. By formulating different views of domain adjacency graphs into a unified multiplex network with edge attributes, we inject node- and semantic level dual attention to refine the propagation of inter-stock collaborative information. Particularly, the stock pair matching is proceeding along each time-stage rather than until final compressed representations, meanwhile significant feature points and scales are identified considering the effect of time attenuation.

DEMERITS:

- Possible lack of comparative analysis
- Complexity and usability.
- They did not do a deployment.
- Limited scalability.

1.4 PROPOSED SYSTEM:

The proposed system aims to predict the stock prices of Tata and Reliance using regression models. The system utilizes historical stock price data and relevant factors to build regression models that can forecast future stock prices. The following components are included in the proposed system. Historical stock price data for Tata and Reliance, along with other relevant factors such as market indices, news sentiment, and economic indicators, will be collected. This data serves as the input for training and testing the regression models. The collected data may undergo pre-processing steps to handle outliers, missing values, and inconsistencies. Techniques such as outlier detection, imputation methods, and data normalization will be applied to ensure the quality and consistency of the dataset. Various regression models, including linear regression, polynomial regression, and support vector regression, will be implemented to predict the stock prices of Tata and Reliance. These models will utilize the pre-processed data and selected features to learn the underlying relationships between the input variables and stock prices. The proposed system provides a reliable and automated approach for predicting the stock prices of Tata and Reliance using regression models.

MERITS

- We compared more than two algorithms to get a better accuracy level.
- We build a user-friendly web application.
- We improved the accuracy level and performance level.
- We implemented Machine Learning properly.

1.5 OBJECTIVE AND SCOPE

The goal is to create a machine learning model for predicting the price of Tata and Reliance stock that might eventually replace the supervised machine learning regression models that can be updated by predicting results with the highest degree of accuracy by comparing supervised algorithms.

Customers can Purchase a variety of goods and services from Tata.com, Inc. Products from both its own third-party vendors and those purchased by it for resale are available through its stores. Additionally, it produces and retails electronic equipment. The Main scope of the project is to finding the accuracy, Minimize the error rate and getting result from the deployment.

1.6 ORGANIZATION OF THE REPORT

Chapter 1 - This chapter gives a brief introduction about the overview of the entire project, literature survey, existing & proposed system and the objective & scope of the project.

Chapter 2 - This chapter gives a brief introduction about the overall description, product perspective, product functions, user characteristics, operating environment, constraints and system requirements and the tools and technologies involved in the development of the project.

Chapter 3 - This chapter explains the overall system architecture. It explains how a user interacts with the system, how intercommunication happens between various services within the system and the various testing methods used for testing the platform.

Chapter 4 - This chapter gives a detailed explanation of how each module was developed and the various demonstrated images captured of the project.

Chapter 5 - This chapter summarizes the whole project, and talks about the future enhancements that can be done in the platform.

CHAPTER 2

SOFTWARE REQUIREMENT SPECIFICATION

2.1 OVERVIEW DESCRIPTION

The main purpose of Requirement Specifications is to describe in a precise manner all the capabilities that will be provided by the Software Application "PREDICTION OF STOCK PRICE USING REGRESSION MODEL" which deals with Linear algorithm, Random Forest algorithm and ARD algorithm for predicting the future stock price. The data collection and preprocessing describe the data sources and methods for collecting and preprocessing the data for the prediction task using data cleaning and transformation techniques, such as handling missing values, outliers, or scaling the data. It also utilizes the graphical tools, such as line plots, scatter plots, histograms, etc., to visualize the data. It also defines the metrics, such as mean squared error, mean absolute error etc., that would be used to evaluate the accuracy and performance of each model on the testing set.

2.1.1 PRODUCT PERSPECTIVE

The system or the project will use regression models, which are statistical techniques that analyse the relationship between a dependent variable (stock price) and one or more independent variables (such as market trends, company performance, economic indicators, etc). The system or the project will find a mathematical function that best fits the data and can be used to estimate the stock price for any given value of the independent variables.

2.1.2 PRODUCT FUNCTION

The system functions can be described as follows

User Interface Design: user should login into their account using username and password. New user should register their details to create the username and password.

Data upload: User can upload their personal data in the application. This method allow user to predict closing price of a stock. Once uploaded, the stock price can be viewed easily.

Stock to be predicted: Registered users need to choose the required stock for which closing price need to be predicted.

User view: The user can view the user list registered in the application.

2.1.3 USER CHARACTERISTICS:

User: The user is the one who uploads personal details in the application. The user has to select the required stock to predict the future stock price.

2.1.4 OPERATING ENVIRONMENT:

System requirements are the required specifications a device must have in order to use certain hardware or software. The hardware and software specification required for the application are listed below.

Hardware requirements:

The hardware specifications used for the development of the application is described in

Processor	Pentium IV/III
Hard disk	minimum 80 GB
RAM	minimum 2 GB

Software Requirements:

The software specifications used for the development of the application is described in

Front End	HTML, CSS
Back End	DJANGO
Operating System	Windows
IDE	Anaconda with Jupyter Notebook

2.1.5 CONSTRAINS:

The system or the project will only use the closing price of the stock as the dependent variable for the prediction task, which may not reflect the intraday fluctuations or volatility of the stock price.

2.2 SPECIFIC REQUIREMENTS

The following subsections of Requirement Specifications should facilitate in providing the entire overview of the "PREDICTION OF STOCK PRICE USING REGRESSION MODEL" under development. This document aims at defining the overall software requirements for developers. Efforts have been made to define the requirements of the prediction system exhaustively and accurately.

2.2.1 EXTERNAL INTERFACE REQUIREMENTS:

Google Chrome

Google Chrome is a cross-platform web browser developed by Google. It was first released in 2008 for Microsoft Windows built with free software components from Apple Web Kit and Mozilla Firefox. It was later ported to Linux, macOS, iOS, and Android where it is the default browser built into the OS. The browser is also the main component of Chrome OS, where it serves as the platform for web applications.

Most of Chrome's source code comes from Google's free and open-source software project Chromium, but Chrome is licensed as proprietary freeware. Web Kit was the original rendering engine, but Google eventually forked it to create the Blink engine; all Chrome variants except iOS now use Blink.

As of July 2021, Stat Counter estimates that Chrome has a 65% worldwide browser market share (after peaking at 72.38% in November 2018) on personal computers (PC), is most used on tablets (has surpassed Safari), and is also dominant on smartphones, and at 63.59% across all platforms combined. Because of this success, Google has expanded the Chrome brand name to other products: Chrome OS, Chromecast, Chromebook, Chromebit, Chromebox, and Chromebase.

2.2.2 SYSTEM FEATURES

This system needs to have certain features to enable its round the clock use by multiple stakeholders with similar but not the same requirements. Hence, the set system is designed to have the following elements: -

High availability - The system should be available under 24x7 with an acceptable down- time.

Security - Only authorized member should be able to access the system as deals with personal and medical data of the patients.

Usability - Its user interface should be highly responsive, easy to learn and operate and require less-time to onboard health workers and diagnostic staff.

Reliability - As this system deals with life and emergencies of the patients, it is critical that every component of the system functions as expected all the time. This will be proper testing and continuous monitoring post deployment to achieve the goal.

2.2.3 UML:

UML, short for Unified Modelling Language, is a standardized modelling language consisting of an integrated set of diagrams, developed to help system and software developers for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modelling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems. The UML is a very important part in the software development process and uses mostly graphical notations to express the design of software projects. Using the UML helps project teams communicate, explore potential designs, and validate the architectural design of the software.

2.2.3.1 USE CASE DIAGRAM

Use case diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analysed the functionalities are captured in use cases. So, it can say that uses cases are nothing but the system functionalities written in an organized manner.

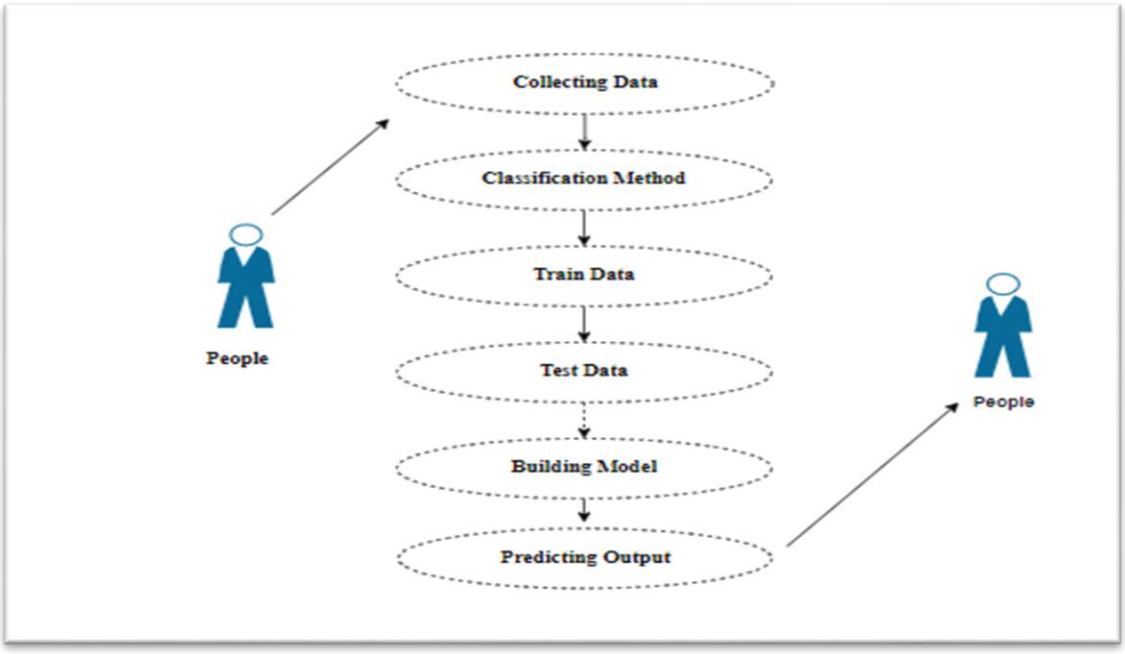


Figure 2.1 Use Case Diagram

2.2.3.2 Class Diagram:

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. So a collection of class diagrams represent the whole system. The name of the class diagram should be meaningful to describe the aspect of the system. Each element and their relationships should be identified in advance Responsibility (attributes and methods) of each class should be clearly identified for each class minimum number of properties should be specified and because, unnecessary properties will make the diagram complicated. Use notes whenever required to describe some aspect of the diagram and at the end of the drawing it should be understandable to the developer/coder. Finally, before making the final version, the diagram should be drawn on plain paper and rework as many times as possible to make it correct.

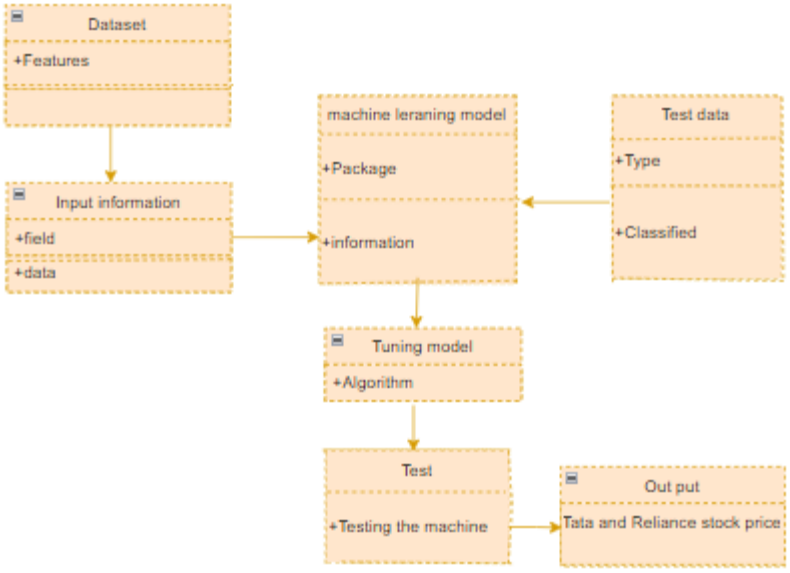


Figure 2.2 Class Diagram

2.2.3.3 SEQUENCE DIAGRAM

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modelling, which focuses on identifying the behaviour within your system. Other dynamic modelling techniques include activity diagramming, communication diagramming, timing diagramming, and interaction overview diagramming. Sequence diagrams, along with class diagrams and physical data models are in my opinion the most important design-level models for modern business application development

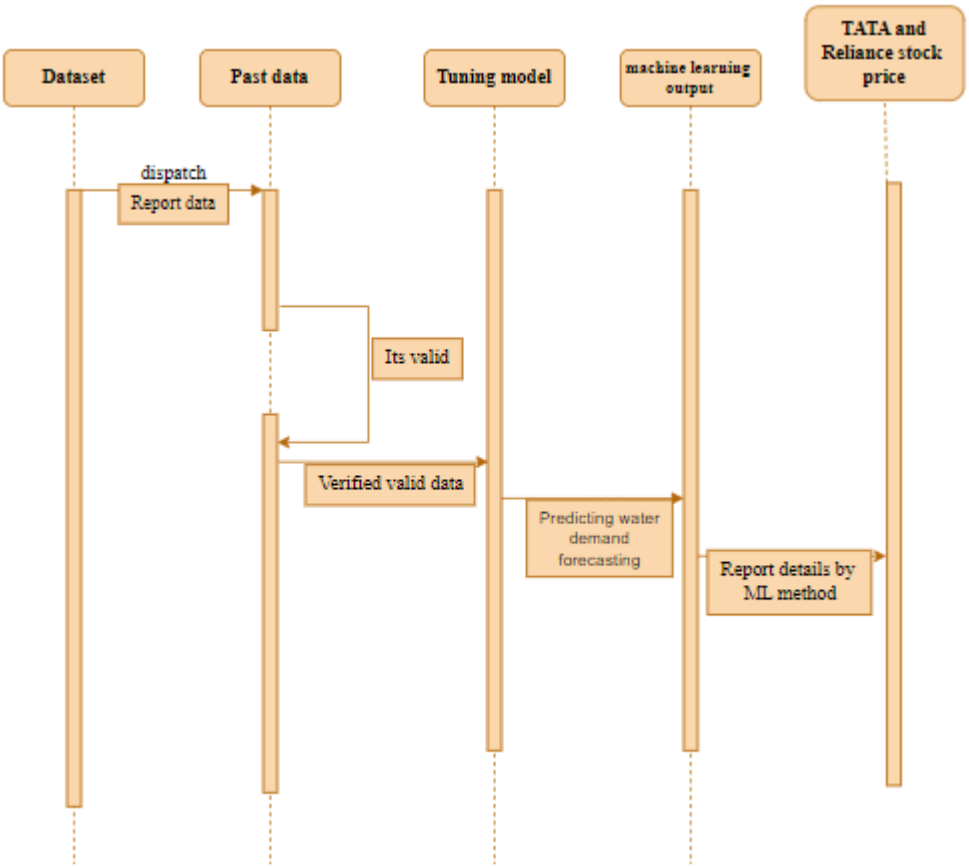


Figure 2.3 Sequence Diagram

2.2.3.4 WORK FLOW DIAGRAM

A workflow diagram (also known as a workflow) provides a graphic overview of the business process. Using standardized symbols and shapes, the workflow shows step by step how your work is completed from start to finish.

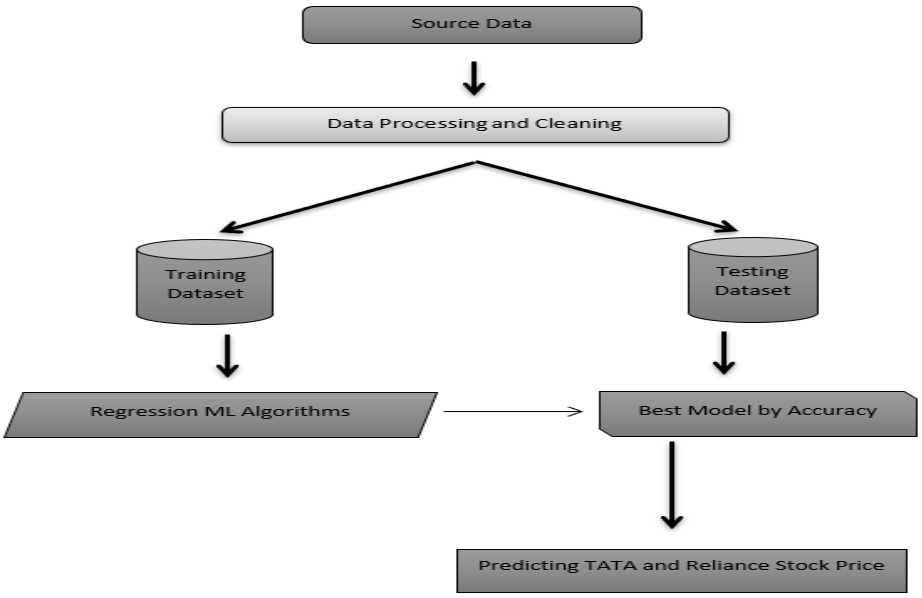


Figure 2.4 Work Flow Diagram

2.2.3.5 ENTITY RELATIONSHIP DIAGRAM

An entity relationship diagram (ERD), also known as an entity relationship model, is a graphical representation of an information system that depicts the relationships among people, objects, places, concepts or events within that system. An ERD is a data modeling technique that can help define business processes and be used as the foundation for a relational database. Entity relationship diagrams provide a visual starting point for database design that can also be used to help determine information system requirements throughout an organization. After a relational database is rolled out, an ERD can still serve as a referral point, should any debugging or business process re-engineering be needed later.

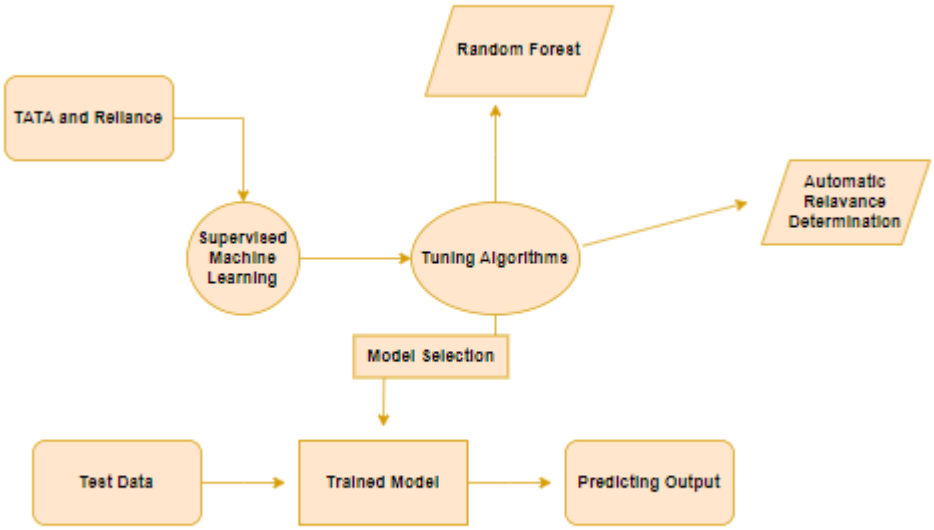


Figure 2.5 Entity Relationship Diagram

2.2.3.6 ACTIVITY DIAGRAM

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in activity diagram is the message part. It does not show any message flow from one activity to another. Activity diagram is some time considered as the flow chart. Although the diagrams looks like a flow chart but it is not. It shows different flow like parallel, branched, concurrent

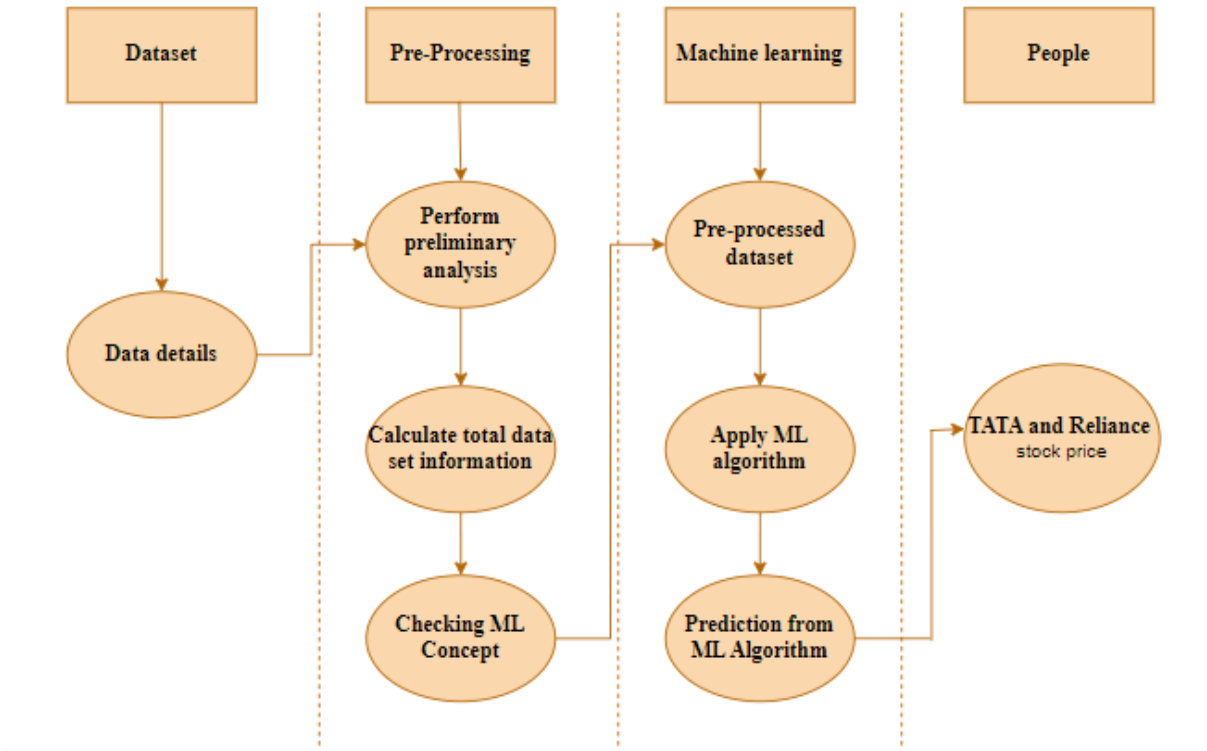


Figure 2.6 Activity Diagram

2.2.3.7 DATA FLOW DIAGRAM

Machine learning needs data gathering have lot of past data. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can't be used directly. It's used to pre-process then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

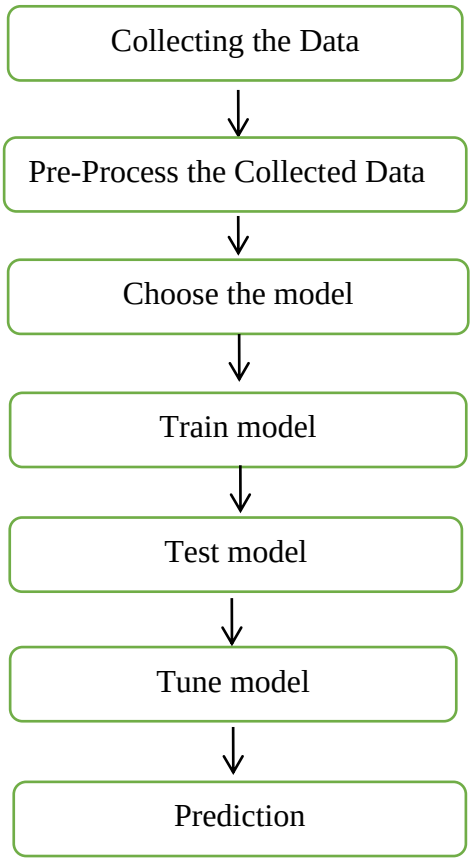


Figure 2.7 Process of dataflow diagram

2.2.4 SOFTWARE QUALITY ATTRIBUTES

Reliability – Reliability is the probability and percentage of the software performing without failure for a specific number of uses or amount of time.

Availability - This feature defines the amount of time the system is running, the time it takes to repair a fault, and the time between lapses. We plan a system to be available 24x7.

Maintainability - This feature indicates the average time and ease and rapidity with which a system can be restored after a failure.

Security - Security measures ensure your software's safety against espionage or sabotage. As prediction system contains lot of sensitivity data especially personal records of clients.

Data integrity - Data integrity refers to maintaining and assuring data accuracy and consistency over its entire lifecycle. If this factor is corrupted, data is lost due to a database error.

Usability - This feature concerns the users; it indicates how effectively they can learn and use a system.

CHAPTER 3

SYSTEM DESIGN AND TEST PLAN

3.1 DECOMPOSITION DESCRIPTION

The data set is pre processed by undergoing certain data cleaning techniques to remove the errors from dataset. Data visualization can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. The algorithms are implemented to create a machine learning "model". The deployment of the required model can be implemented for prediction of future stock price using a new data.

3.1.1 DATA PREPROCESSING

Importing the library packages with loading given dataset. To analysing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning models and procedures that you can use to make the best use of validation and test datasets when evaluating your models. Data cleaning / preparing by rename the given dataset and drop the column etc. to analyse the uni- variate, bi-variate and multi-variate process. The steps and techniques for data cleaning will vary from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making.

3.1.2 DATA ANALYSIS OF VISUALIZATION

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance.

Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. Being able to quickly visualize of data samples and others is an important skill both in applied statistics and in applied machine learning. It will discover the many types of plots

that you will need to know when visualizing data in Python and how to use them to better understand your own data.

- How to chart time series data with line plots and categorical quantities with bar charts.
- How to summarize data distributions with histograms and box plots.

3.1.3 AUTOMATIC RELEVANCE DETERMINATION

Automatic Relevance Determination (ARD) is a Bayesian regression algorithm designed to automate the process of feature selection in regression models. With a focus on handling datasets with numerous input features, ARD operates within the framework of Bayesian linear regression. It introduces the concept of feature relevance by associating each input feature with a binary variable, signifying its importance in the model. The ARD algorithm is particularly notable for its ability to automatically estimate hyperparameters, including precision terms for feature relevance and coefficients, typically using methods like Expectation-Maximization or Variational Inference. By analyzing the Bernoulli parameters for feature relevance, ARD effectively prunes irrelevant features, streamlining the model and reducing overfitting. As a result, ARD not only simplifies the model but also provides a posterior distribution over coefficients, accounting for uncertainty and aiding in feature selection for improved predictive accuracy and interpretability. While ARD offers benefits like automated feature selection and uncertainty quantification, it can be computationally intensive, and the choice of prior distributions can influence its performance. In summary, ARD is a powerful tool for enhancing the robustness and interpretability of regression models, particularly in scenarios with high-dimensional data.

3.1.4 IMPLEMENTING RANDOM FOREST REGRESSION

Random Forest is a powerful ensemble machine learning algorithm that combines the predictive strength of multiple decision trees to make accurate and robust predictions. At its core, it uses decision trees as building blocks, but the ensemble approach sets it apart. The "Random" aspect comes from the introduction of randomness in two key ways: bootstrapping and feature selection. When constructing each decision tree in the forest, a random sample of the training data, with replacement, is drawn. This diversity among trees helps mitigate overfitting. Additionally, for each tree, a random subset of input features is considered at each

node to determine the best split, ensuring that no single feature dominates the model. The final prediction is made by aggregating the individual tree predictions through majority voting (in classification tasks) or averaging (in regression tasks). Notably, Random Forest provides a robust estimate of its own performance through the out-of-bag (OOB) error, making it an appealing choice for practitioners. This method tests data points that were not included in the training set for each tree, offering an unbiased estimate of model accuracy. Random Forest is renowned for its high accuracy, robustness, feature importance measurement, and ability to handle missing data, making it a widely used and trusted tool in various domains, from finance to healthcare, where accuracy, resilience, and interpretability are paramount. Nevertheless, its challenge lies in model interpretability and the computational resources required when working with numerous features or trees. In summary, Random Forest excels in enhancing predictive performance and reducing the risk of overfitting, making it a valuable asset in the machine learning toolkit for a wide array of applications.

3.1.5 IMPLEMENTING LINEAR REGRESSION

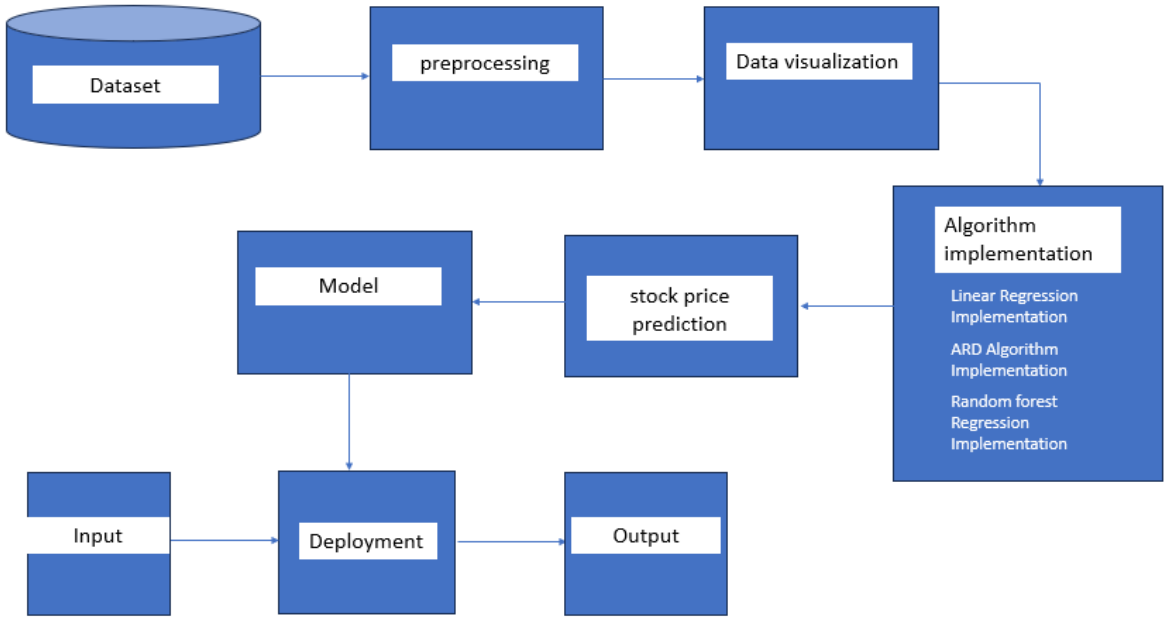
Linear regression is a statistical technique that tries to find a linear relationship between a dependent variable (such as stock price) and one or more independent variables (such as time, volume, indicators, etc.). It can be used to forecast future stock prices based on historical data and patterns. However, linear regression has some limitations and assumptions that may not hold true in the real world, such as linearity, normality, homoscedasticity, and independence of errors. Therefore, linear regression may not be able to capture the complexity and uncertainty of the stock market.

3.2 DEPENDENCY DESCRIPTION

The successful implementation of the prediction system is highly dependent on effective and accurate requirements gatherings. Moreover, the continuously evolving scenario of predicting requires that the design is kept up-to-date and implemented as soon as possible. This system can be installed in the client environment and this will help in the maintenance of the system in future. People are inherently resistant to change, and computers have been known to facilitate change. The success of prediction system also depends on easy-to-use user interface that can enable fast learning and adoption of the system by its end users.

3.2.1 SYSTEM DESIGN

A system design for prediction of stock price project is a way of describing the architecture and components of a system that can perform the task of predicting the future prices of stocks based on historical and current data. A system design can help to plan, develop, test, and deploy the system, as well as to communicate the system's functionality and requirements to the stakeholders.



3.1 FIGURE SYSTEM ARCHITECTURE

3.2.2 DATABASE DESIGN

Below are some of the sample database designs exists in the prediction application acting as an input.

Table 3.2 Tata Dataset

TATA DATASET:

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	##########	122.8	122.8	119.82	120.3325	88.08827	30646000
3	##########	121.2375	123.75	120.625	123.345	90.29355	24465208
4	##########	123.3125	123.75	122	123.5125	90.41612	21194656
5	##########	123.75	124.375	122.95	123.4875	90.39782	19935544
6	##########	123.7375	125.575	123.25	124.2075	90.9249	21356352
7	##########	125.75	137.5	123.795	124.7325	91.30921	9869856
8	##########	129.9875	129.9875	124.1125	124.3575	91.03472	9038672
9	##########	129.375	129.375	124.375	124.45	91.10244	5772232
10	##########	124.5	125.1	123.8875	124.2125	90.92857	6593984
11	##########	124.625	124.7375	122.3075	122.495	89.67131	7947184
12	##########	123.75	123.75	122	123.6	90.48019	6415072
13	##########	123.875	126.9375	123.875	125.4375	91.82535	20904912
14	##########	125.625	127.35	125.37	126.9575	92.93802	15335472
15	##########	127.25	127.5	125.15	125.7625	92.06323	11988288
16	##########	125.875	126.875	125.3125	126.0575	92.27919	7358224
17	##########	126.5	128.7375	126.2875	128.0625	93.74693	14627896
18	##########	129.05	129.6125	127.52	127.8625	93.60053	8552224
19	##########	128.1075	131.0625	127.75	130.725	95.69602	13897080
20	##########	131.05	132	129.6375	131.5075	96.26882	15371584
21	##########	131	131.125	128.375	128.7825	94.27399	15819680
22	##########	128.6375	130	128.25	128.6	94.14043	8083208
23	##########	128.4375	129.0625	126.875	127.595	93.40469	8229512
24	##########	127	128.425	126.5625	126.945	92.92889	6339416
25	##########	127.4075	129.125	127.02	128.875	94.34172	11737176
26	##########	129.45	130.1125	127.5825	128.3875	93.98488	12521432
27	##########	128.625	132.0575	128.295	130.9625	95.86981	10968888
28	##########	132.125	135.125	131.8075	134.5625	98.5052	15129824
29	##########	135.0375	136.125	134.4375	135.0875	98.8895	14341288
30	##########	135.125	137.85	134.3825	135.2	98.9719	17670664
31	##########	135.375	137.5	135.375	137.1375	100.3902	10080128
32	##########	137.25	138.875	136.875	137.2575	100.478	8558920
33	##########	137.5	138.9375	135.625	136.1625	99.67648	9332056

Table 3.3 Reliance Dataset

RELIANCE DATASET:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover	Trades	Deliverab	%Deliverb
2	#########	RELIANCE	EQ	204.65	205	206.1	203.65		205.75	205.26	3717450	7.63E+13			
3	#########	RELIANCE	EQ	205.75	205.25	206.25	202.65		204.15	204.13	6024650	1.23E+14			
4	#########	RELIANCE	EQ	204.15	207.5	216.95	205.25		205.7	207.04	7473500	1.55E+14			
5	#########	RELIANCE	EQ	205.7	203.75	204.4	201.05		203.8	202.47	7744000	1.57E+14			
6	#########	RELIANCE	EQ	203.8	203	203	200.65		202.4	202.05	5952000	1.20E+14			
7	#########	RELIANCE	EQ	202.4	202	202.15	192.3		194.3	197.57	6675550	1.32E+14			
8	#########	RELIANCE	EQ	194.3	188.55	192	182.25		188.7	186.15	13880150	2.58E+14			
9	#########	RELIANCE	EQ	188.7	185	189.45	183		183.85	186.24	9875700	1.84E+14			
10	#########	RELIANCE	EQ	183.85	182	190.5	181.75		189.6	185.82	14625600	2.72E+14			
11	#########	RELIANCE	EQ	189.6	191	192.45	186.35		187.6	188.95	13377500	2.53E+14			
12	#########	RELIANCE	EQ	187.6	187.05	188	184.6		185.65	186.15	7139950	1.33E+14			
13	#########	RELIANCE	EQ	185.65	185	187	184.4		186	185.71	7756400	1.44E+14			
14	#########	RELIANCE	EQ	186	185	185.25	177.05		178.6	181.66	10697700	1.94E+14			
15	#########	RELIANCE	EQ	178.6	174.05	176.9	171.8		173.65	174.42	13478000	2.35E+14			
16	#########	RELIANCE	EQ	173.65	171.25	173.4	169.45		170.6	171.23	11586900	1.98E+14			
17	#########	RELIANCE	EQ	170.6	171.15	175.75	171		173.5	173.8	12089550	2.10E+14			
18	#########	RELIANCE	EQ	173.5	174.1	174.85	157		166.15	170.68	12671700	2.16E+14			
19	#########	RELIANCE	EQ	166.15	164.7	165.7	158		160.45	161.39	12538400	2.02E+14			
20	#########	RELIANCE	EQ	160.45	160.1	160.8	152.75		153.7	157.11	12883650	2.02E+14			
21	#########	RELIANCE	EQ	153.7	153	167.25	152.05		165.8	160.64	23310600	3.74E+14			
22	#########	RELIANCE	EQ	165.8	171	189	168.5		180.7	174.76	37914350	6.63E+14			
23	#########	RELIANCE	EQ	180.7	179	185.75	172.25		176.75	179.43	26118950	4.69E+14			
24	#########	RELIANCE	EQ	176.75	175	185	174.5		183.85	180.21	22865050	4.12E+14			
25	#########	RELIANCE	EQ	183.85	188.75	206	188.75		203.8	196.45	29231000	5.74E+14			
26	#########	RELIANCE	EQ	203.8	215	225.2	208.55		220.75	216.42	43646350	9.45E+14			
27	#########	RELIANCE	EQ	220.75	226.05	230	197.6		202.25	218.04	44354450	9.67E+14			
28	#########	RELIANCE	EQ	202.25	191.05	208	189.05		206	201.29	31970000	6.44E+14			
29	#########	RELIANCE	EQ	206	208.1	227.65	206		224	214.12	29335550	6.28E+14			
30	#########	RELIANCE	EQ	224	232	242	224.1		237	233.49	35504550	8.29E+14			
31	#########	RELIANCE	EQ	237	231	235	219		222.15	228.02	34689050	7.91E+14			
32	#########	RELIANCE	EQ	222.15	220	244.15	218.1		235.5	225.63	36382650	8.21E+14			
33	#########	RELIANCE	EQ	235.5	241.5	244.95	233.85		235.65	238.52	22223700	5.30E+14			

3.3. 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 tests. Each test type addresses a specific testing requirement.

TYPES OF TESTS

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 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.

FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centred on the following items:

Valid Input: identified classes of valid input must be accepted.

Invalid Input: identified classes of invalid input must be rejected.

Functions: identified functions must be exercised.

Output: identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify

Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

Table 3.1 Test Case register

TEST CASE NO	MODULE	TEST CASE SUMMARY	EXPECTED OUTCOME	ACTUAL OUTCOME	RESULT
1	**Register**	To verify that the user id able to register the details in the	User should be able to register to	User is able to register to the system successfully	**PASS**

		system successfully.	the system successfully.		

Table 3.2 Test Case login

TEST CASE NO	MODULE	TEST CASE SUMMARY	EXPECTED OUTCOME	ACTUAL OUTCOME	RESULT
1	**Login**	To verify that the user id able to login to the system successfully.	User should be able to login to the system successfully.	User is able to login to the system successfully	**PASS**

Table 3.3 Test Case deployment

TEST CASE NO	MODULE	TEST CASE SUMMARY	EXPECTED OUTCOME	ACTUAL OUTCOME	RESULT
1	**Deployment**	To verify that the user is able to get the predicted closing price	User should be able to get the predicted closing price	User is able to get the predicted closing price	**PASS**

		in the system successfully.	in the system successfully	in the system successfully	

Table 3.4 Test Case deployment

TEST CASE NO	MODULE	TEST CASE SUMMARY	EXPECTED OUTCOME	ACTUAL OUTCOME	RESULT
1	**Deployment**	To verify that the user id able to login to the system successfully.	User should be able to login to the system successfully.	User is able to login to the system successfully	**PASS**

CHAPTER 4

IMPLEMENTATION AND RESULTS

4.1 IMPLEMENTATION

System Implementation uses the structure created during system design and the results of system analysis to construct system elements that meet the stakeholder requirements and system requirements developed in the early life cycle phases. Also, this is crucial phase in the system development life cycle is successful implementation of the system design. Implementation describes covering the system design into operation; which includes all those activities that take place to convert from the old system to new system. The new system may be totally new, replacing an existing manual or automated system or it may be a major modification to an existing system. Proper implementation is essential to provide a reliable system to meet the organization requirements. Successful implementation may not guarantee improvement in the organization using the new system, but improper installation will prevent it.

The implementation stage involves the following tasks: -

o Careful planning;

o Investigation of system and constraints;

o Design of methods to achieve the changeover phase;

o Training of staff in the changeover phase;

o Evaluation of the changeover method

The method of implementation and the time scale to be adopted are found out initially. Next, the system is tested properly and the system is tested and the same time users are trained in the new procedures.

4.2 RESULTS

Post implementation and deployment, we carried out testing with the end users and found that the features where very suitable for the purpose of Prediction of stock price using regression model. Tests are conducted and all the results are evaluated. That is test results are compared with expected results. When erroneous data are uncovered, an error is implied and debugging commences. It met the testing and authentication workflow as set out by various authorities and any changes in the workflow was quickly adopted through making changes to configurable items in the system. This system was successfully used by User during the field testing.

Working of the system has captured in screenshot during its run demonstrates the successfully implemented as elaborated in the requirements specification. The features covered are listed below:

4.2.1 HOME PAGE

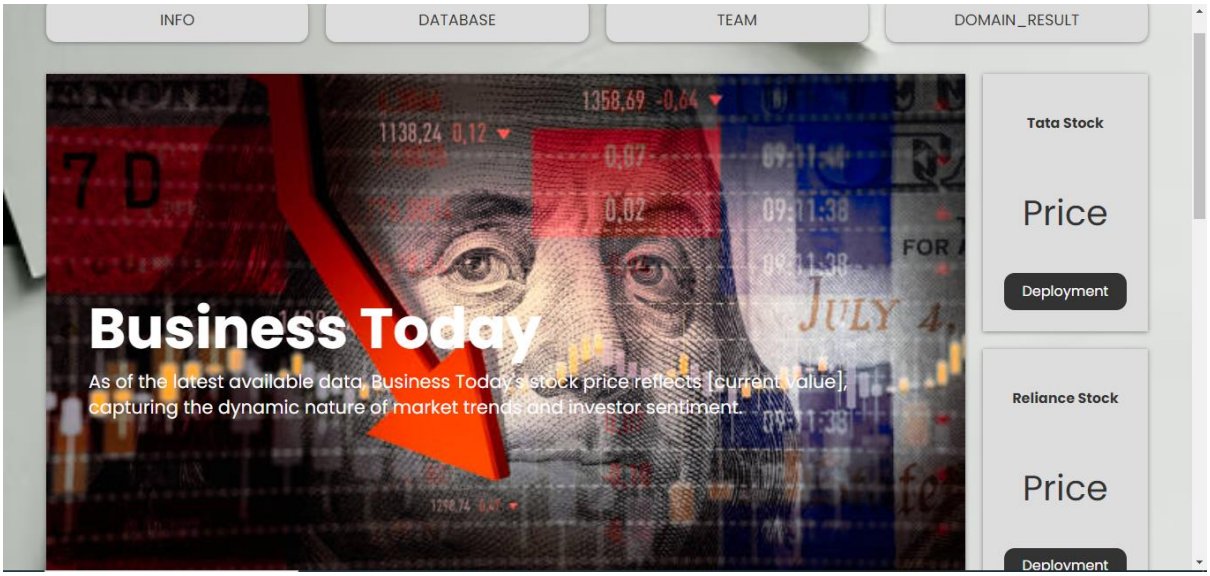


Figure 4.1 Home Page

4.2.2 REGISTRATION PAGE

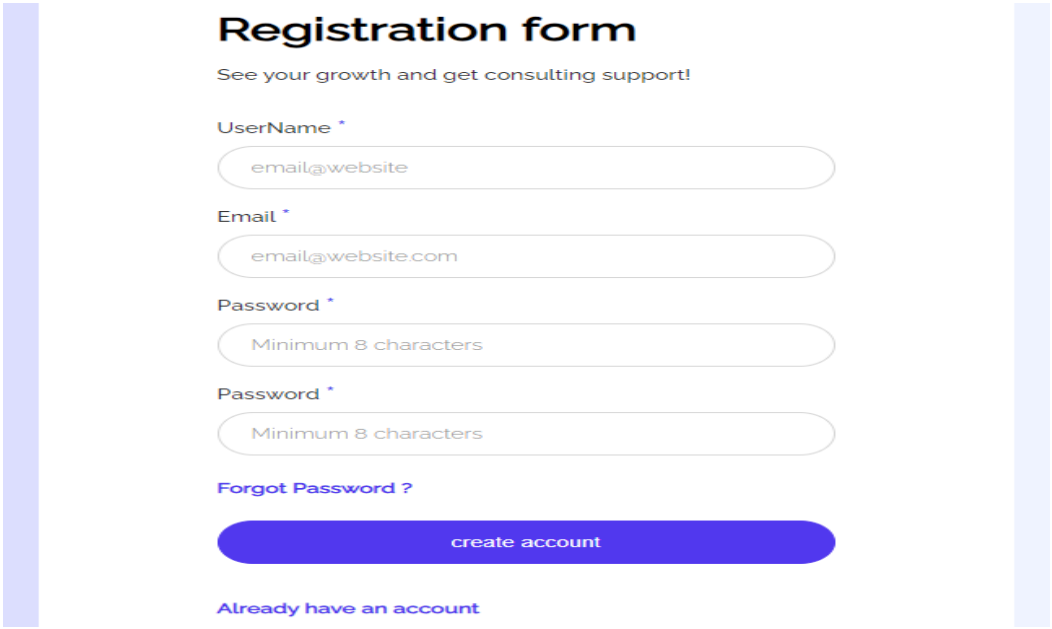


Figure 4.2 Registration Page

4.2.3 LOGIN PAGE

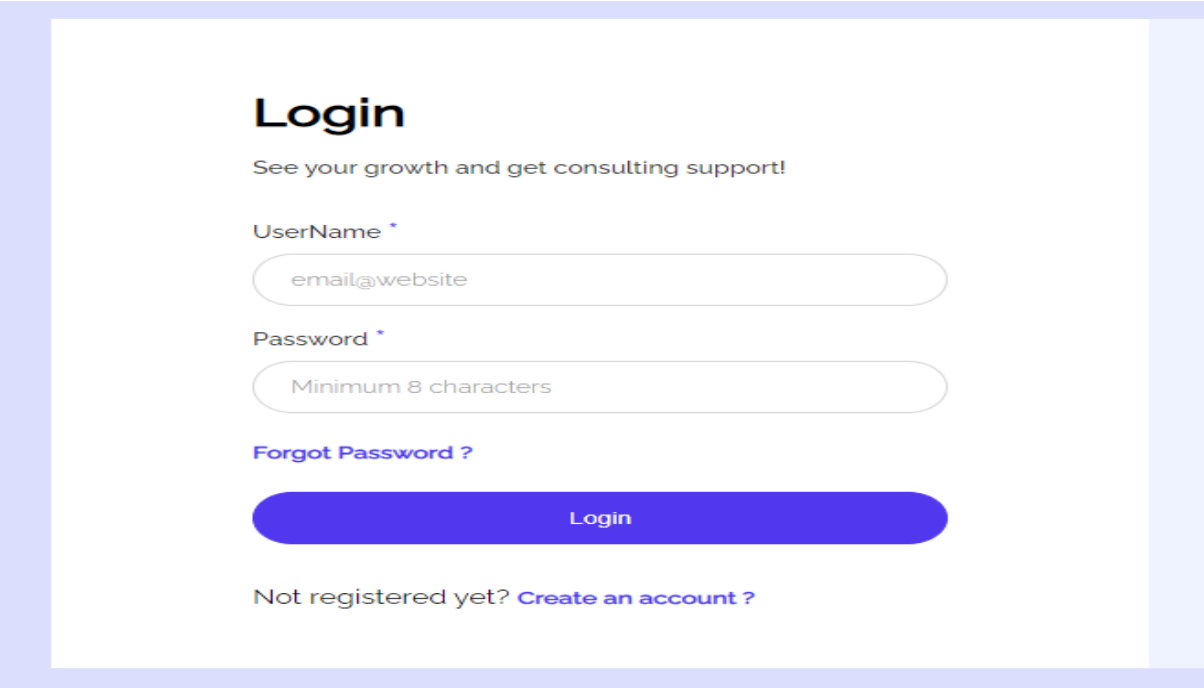


Figure 4.3 Login Page

4.2.4 PERSONAL INFORMATION PAGE

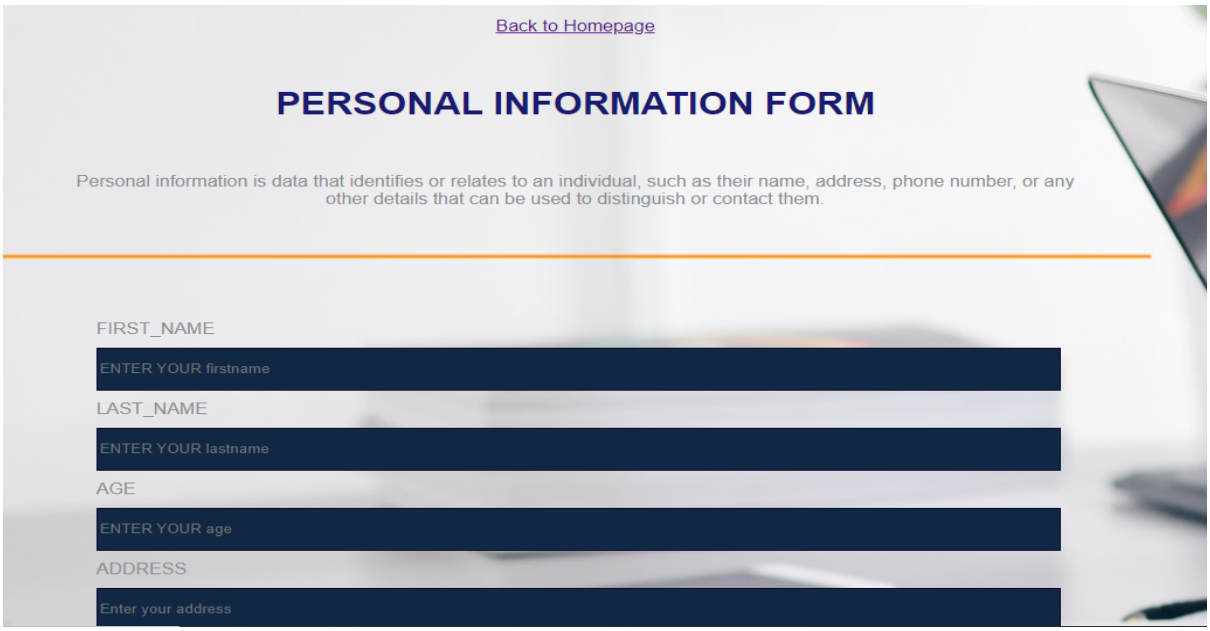


Figure 4.4 Personal Information Page

4.2.5 PERSONAL INFORMATION DATABASE PAGE

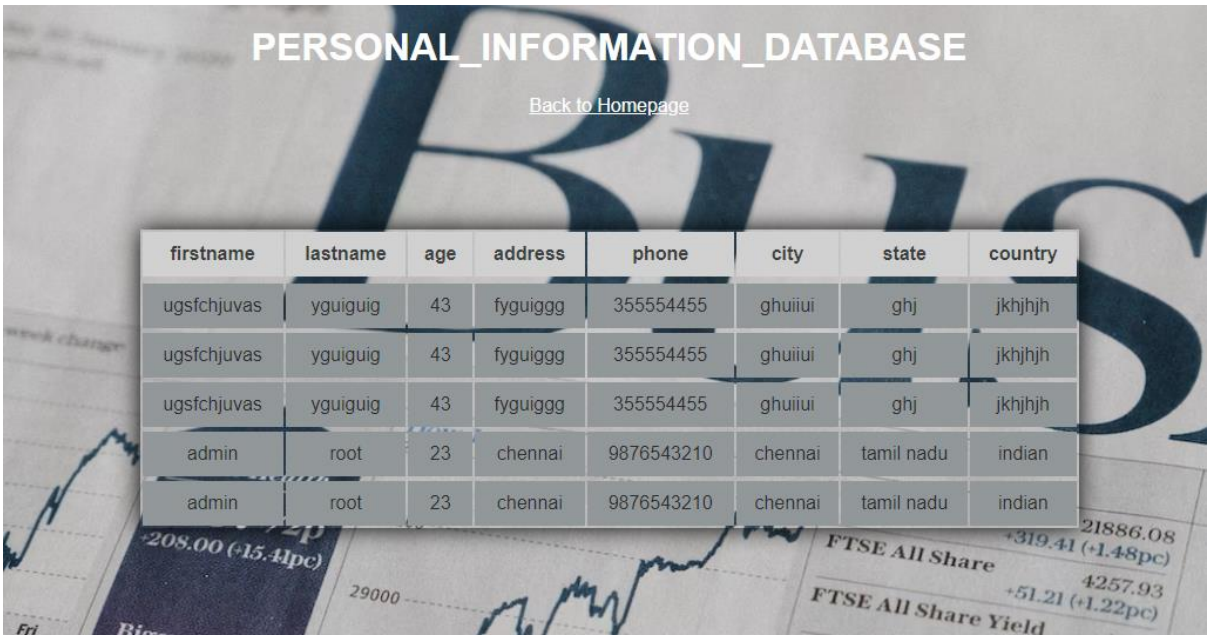


Figure 4.5 Personal Information Page

4.2.6 ABOUT US PAGE

Figure 4.6 About Us Page

4.2.7 ALGORITHM LAYOUT PAGE

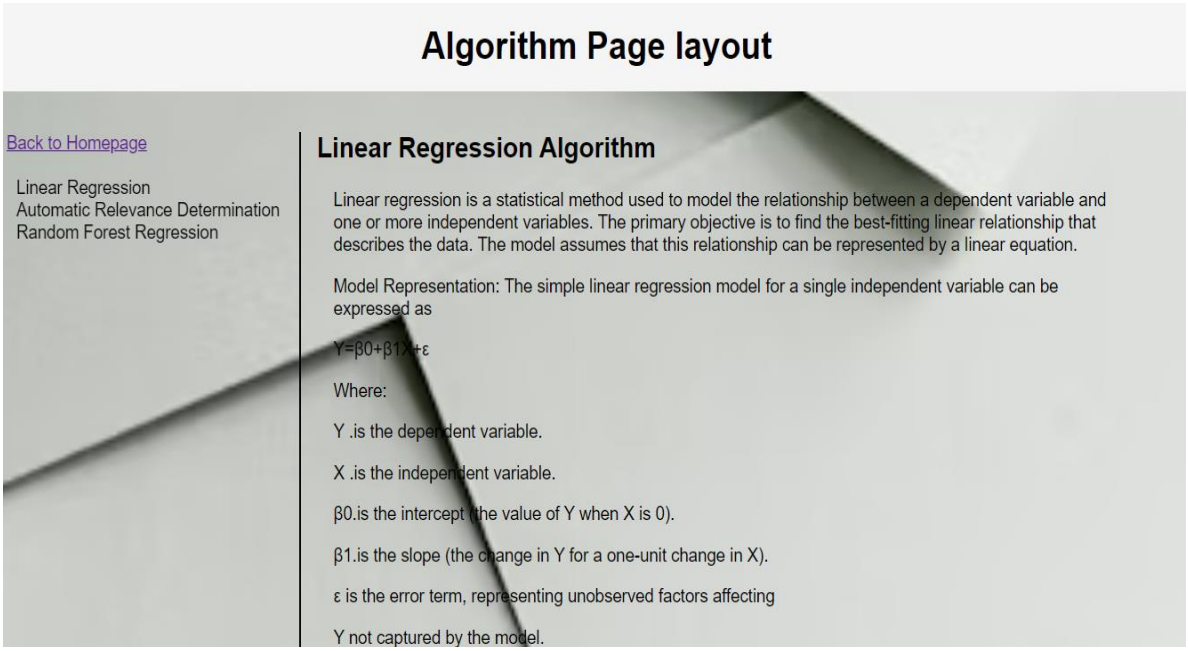


Figure 4.7 Algorithm Layout Page

4.2.8 TATA DEPLOYMENT PAGE

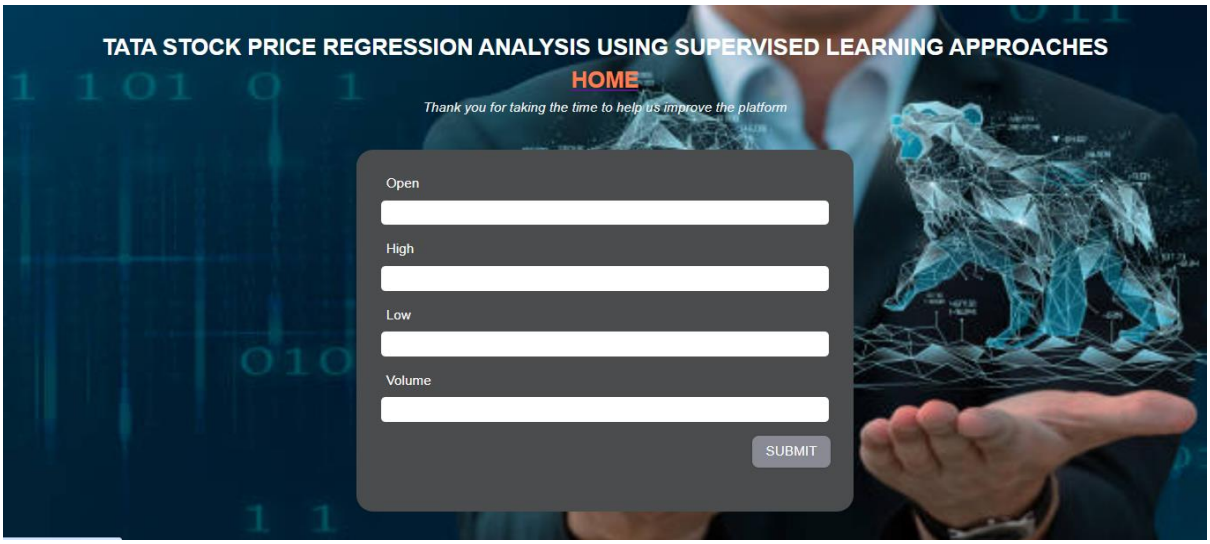


Figure 4.8 Tata Deployment Page

4.2.9 RELIANCE DEPLOYMENT PAGE

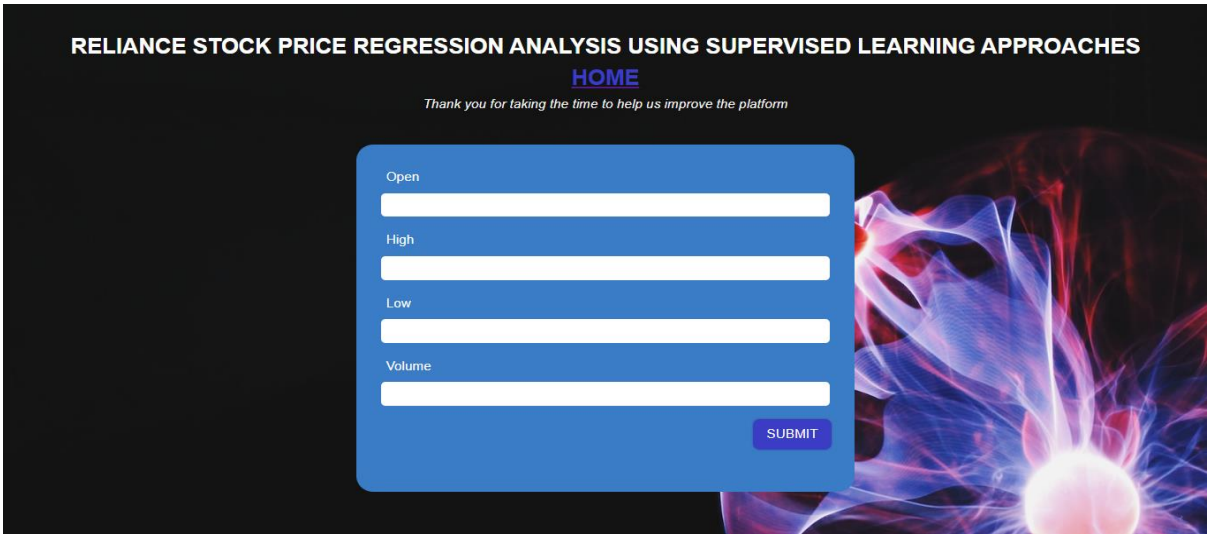


Figure 4.9 Reliance Deployment Page

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The Best accuracy on public test set is higher accuracy score is will be found out. This application can help out to find the Tata and Reliance Stock Price.

5.2 FUTURE WORK:

- Tata and Reliance Stock Price prediction to connect the AI Model
- To automate this process by show the prediction result in web application or desktop application.
- To optimize the work to implement in Artificial Intelligence Environment.

REFERENCE

[1] Thirunavukkarasu Anbalagan a, S. Uma Maheswari b, Classification and Prediction of Stock Market Index Based on Fuzzy Meta graph, Elsevier B.V, 1877-0509 © 2015, DOI: 10.1016/j.procs.2015.03.200

[2] Devanshu Upadhyay1, Soumya Sharma2, Stock Price Prediction Using Concept of Long-Short Term Memory, International Journal of Scientific Engineering and Research (IJSER), Volume 9 Issue 12, December 2021

[3] Dr. P. K. Sahoo, Mr. Krishna charlapally, Stock Price Prediction Using Regression Analysis, International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015 ISSN 2229-5518

[4] Kathika Sai Krishna, Mettupalli Hari Naveen Reddy, Appidi Koushik Reddy, Mukkamalla Naveen Reddy, Stock price prediction using machine learning models in python, *International Journal of Advance Research, Ideas and Innovations in Technology,* Volume 7, Issue 2 - V7I2-1525 © *2021*