Project Report

On

**STOCK MARKET PREDICTION & ANALYSIS**

Submitted to

Sharda University, Greater Noida



in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

Department Of

Computer Science & Engineering

By

**SWATI DHOUNDIYAL**

**SIMRAN MOHAKUD**

**AKSHAT ARORA**

**KARTIK PATADIA**

under the guidance of

**Mr. Abhishek Singh Verma**

**Assistant Professor**

Department of Computer Science & Engineering

SCHOOL OF ENGINEERING AND TECHNOLOGY

SHARDA UNIVERSITY , GREATER NOIDA

May-June2020

**BONAFIDE CERTIFICATE**

This is to certify that this project report entitled “**STOCK MARKET PREDICTION & ANALYSIS”** submitted to **SHARDA UNIVERSITY ,GREATER NOIDA**, is a bonafide record of work done by **“SWATI, SIMRAN , AKSHAT & KARTIK”** under my supervision from **“07TH JAN’20” to “25TH APR’20”.**

**Mr. ABHISHEK SINGH VERMA**

ABHISHEK SINGH VERMA

Department of Science & Technology

SCHOOL OF ENGINEERING AND TECHNOLOGY

SHARDA UNIVERSITY,GREATER NOIDA

**PLACE: SHARDA UNIVERSITY**

**DATE: May 27, 2020**

**DECLARATION BY AUTHORS**

This is to declare that this report has been written by us. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. We aver that if any part of the report is found to be plagiarized, we are shall take full responsibility for it.

**Swati Dhoundiyal**

Roll number: 180101346

**Simran Mohakud**

Roll number: 180101324

**Akshat Arora**

Roll number: 180101037

**Kartik Patadia**

Roll number: 180101143

**Place: Sharda University**

**Date: May 27, 2020**

**ABSTRACT**

Our aim is to create software that analyzes previous stock data of RELIANCE Company, with help of certain parameters that affect stock value. We are going to implement these values in data mining algorithms and we will be able to decide which algorithm gives the best result. This will also help us to determine the values that particular stock will have in near future. We will determine the patterns in data with help of data mining algorithms.

The software will be developed using Asp .Net and will be cross checked and verified using “DB MINER” which is standard data mining tool.

An extension to the project that sends an SMS to the user that is triggered on certain events can be added as an optional feature.

**TABLE OF CONTENTS**

**TITLE PAGE NO.**

**ABSTRACT iii**

**CONTENTS iv**

**INTRODUCTION………………………………………………….. 1**

1. **PROPOSED SYSTEM……………………..………………….. 2**
2. **SYSTEM ARCHITECTTURE……………………… 3**
3. **MODULE IDENTIFICATION……………………… 4-5**
4. **TECHONOLOGIES USED……………………………………. 6-7**
5. **FINAL PROJECT OUTPUT…………………………………. 8**
6. **CONCLUSION………………………………………………… 9**
7. **FUTURE SCOPE……………………………………………… 10**
8. **REFERENCES………………………………………………… 11**

**INTRODUCTION**

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate. Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends.

The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the

effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of

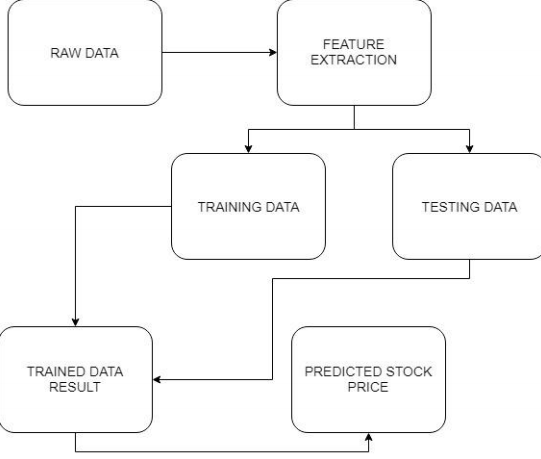
almost all investors to correctly and consistently predict these hyper parameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result.

**PROPOSED SYSTEM ARCHITECTURE**

In this proposed system, we focus on predicting the stock values using machine learning technique **Data base mining.** In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the previous year stocks to train the model. The two majorly used machine-learning libraries to solve the problem are **NumPy & Matplotlib.** First NumPy which is used to clean and manipulate the data, and getting it into a form ready for analysis. The other is Matplotlib,which, is a plotting library for the Python programming language and its numerical mathematics extension NumPy.. The data set we used was from the previous years stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python pandas library for **data processing** which combined different datasets into a data frame. The tuned up dataframe allowed us to prepare the data for feature extraction. The dataframe features were date and the closing price for a particular day. We used all these features to train the machine on training data set and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and the actual values. The proposed system touches different areas of research including data base mining ,data pre-processing, and so on.

**SYSTEM ARCHITECTURE**

**Block Diagram**

****

**MODULES IDENTIFICATION**

The various modules of the project would be divided into the segments as described.

1. **DATA COLLECTION**

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external. Our data

mainly consists of the previous year stock prices. Initially, we will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

1. **PRE-PROCESSING**

Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and

finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

1. **TRAINING THE MACHINE**

Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the hyper parameters like the Dollar value, Corporate value, Inflation . We perform the entire cross-validation loop on each set of hyper parameter values.Finally, we will calculate a cross-validated score, for individual sets of hyper parameters. Then, we select the best hyper parameters. The idea behind the training of the model is that we some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

**TECHNOLOGIES USED**

* **Python (programming language)**

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

**Some of the important Python libraries used are:**

* **KERAS**

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

* **NUMPY**

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

* **PANDAS**

In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

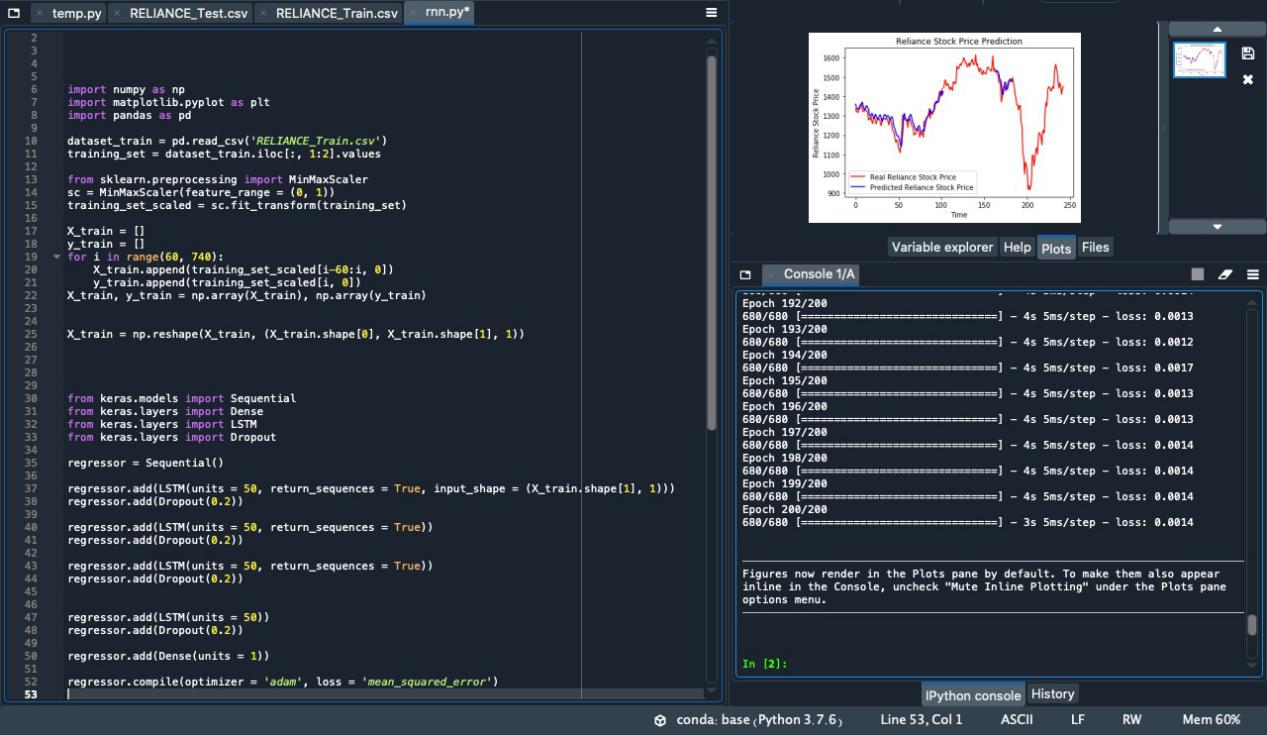
* **Anaconda Navigator:-**

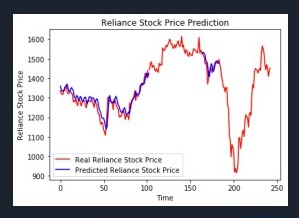
Anaconda is a free and open-source Python distribution and collection of hundreds of packages related to data science, scientific programming, development and more. Python is included in the Anaconda distribution. It is not an IDE though it can be configured with most IDEs. It also comes with a platform-agnostic package manager called conda.

* **Spyder:-**

Spyder is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages in the scientific Python stack, including NumPy, SciPy, Matplotlib, pandas, IPython, SymPy and Cython, as well as other open source software. It is released under the MIT license.

**PROJECT OUTPUT**

****

****

**CONCLUSION**

By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the data mining technique. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interesting metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

**FUTURE SCOPE**

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation‟s performance structure as a whole.

**Application of Analysis of stocks in our selected domain:**

* Analyzing the factors.

We have to obtain the data in the same period for the following

factors.

1. **Dollar value**: We will obtain the variation of dollar value as compared to the rupee.
2. **Corporate results:** Companies declare their performance results and profit at the end of each quarter.
3. **Inflation:** From financial experts we can obtain inflation rate over a period of time.

* Helps the users in detecting the market trend patterns and other

conditions.

* The project report contains a number of filter elements ranging as per the market trends that helps the users to analyze the registered stocks
* The project report contains a number of shares, their prices and Volume Breakouts
* It contains stocks that have an unanticipated rise in the business volume and a rapid rate escalation in terms of cost.

**REFERENCES**

* <https://en.wikipedia.org/wiki/Data_mining>
* <https://en.wikipedia.org/wiki/Stock_market_prediction>
* <https://www.securities-services.societegenerale.com/en/insights/expert-views/technology/tech-magazine/twenty-years-research-stock-market-prediction-from-text-mining/>
* <https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/>