**SMART INVESTMENT PREDICTION**

**PROJECT REPORT**

**Submitted by**

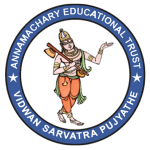
**MIND BENDERS**

* S.V.V Anki Reddy
* B. Divya
* K. DILLI
* L. Pooja
* K. Prathyusha

***In partial fulfilment for the award of the Certificate***

**of**

**SUMMER INTERNSHIP PROGRAM**

**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village , Renigunta Mandal , Tirupati , Andhra Pradesh**

**517520**

**July 2019.**

### *BONAFIDE CERTIFICATE*

This is to certify that the project entitled ***”MIND BENDERS”*** submitted by **s.V.V Anki Reddy, B.Divya, K.Dilli, L.Pooja, K.Prathyusha** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### Signature of Supervisor                                            Signature of Head of the Department

   Akshay Kumar Kothuri, M.Tech                                 Mrs.B.Rupa Devi M.Tech,Ph.D.,

  AI & IoT Developer Assistant Professor & HOD,

  SmartBridge Educational Services PVT LTD               Department of CSE,

  Hyderabad – 500096                                                        AITS,Tirupati.

INDEX

1. Title of the project

1.1. Introduction

1.2. Objective of Research

1.3. Problem Statement

1.4. Industry Profile

2. Review of Literature

3. Data Collection

4. Methodology

4.1. Exploratory Data Analysis

4.1.1. Figures and table

4.2 Statistical techniques and visualization

4.3 Data Modeling and visualization

5. Findings and Suggestions

6. Conclusion

# 1.1 INTODUCTION:

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange, as well as stock that is only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms.

Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.

# 1.2 Objective Of Research:

## The main objectives of stock market are:

### Raising Money For Business:

Stock exchanges around the world enable companies around the world to raise money.

### Capital Formation:

The primary function of a stock exchange is to help companies raise money.

### Security And Transparency:

The legitimate sale of stock on any exchange requires reliable and accurate information.

Companies plan to use the newly-raised funds to invest in productive business assets and grow revenues and profits.

## 1.3 Problem Statement:

We’ll dive into the implementation part of this article soon, but first it’s important to establish what we’re aiming to solve. Broadly, stock market analysis is divided into two parts – Fundamental Analysis and Technical Analysis. Fundamental Analysis involves analyzing the company’s future profitability on the basis of its current business environment and financial performance. Technical Analysis, on the other hand, includes reading the charts and using statistical figures to identify the trends in the stock market.

Machine Learning methods that can actually analyse the stock prices over time and gain intelligence, then use this intelligence in prediction, can be used to model such a tool

## 1.4 Industry Profile :

The main significant approach, used in this paper for the predicting result is a concept of machine learning and result tested on the Bombay Stock Exchange (BSE) index data set. To seize the best accurate output, the approach decided to be implemented is machine learning along with supervised classifier.

# 2. REVIEW OF LITERATURE:

In fact, different input variables are used to predict the same set of stock return data. Some researchers used input data from a single time series where others considered the inclusion of heterogeneous market information and macro economic variables. Some researchers even pre-processed these input data sets before feeding it to the ANN for forecasting.

1. A recent study (et al.Risul Islam Rasel ,Nasrin Sultana ,NasimulHasan, IEEE 2016) has shown that ANN model can be more advantageous compared to other SVM or LR models and the Advantages are Increase in accuracy with multiple attributes[4]. Works well even if attributes and output do not have a clear relation. Also, some disadvantages also must be considered which are Time required for prediction is more than other methods can face over fitting problem.

2. Chan., Wong. And Lam., implemented a neural network model using the technical analysis variables for listed companies in Shanghai Stock Market. In this paper performance of two learning algorithm and two weight initialization methods are compared. The results reported that prediction of stock market is quite possible with both the algorithm and initialization methods but the performance of the efficiency of the back propagation can be increased by conjugate gradient learning and with multiple linear regression weight

## 3. DATA COLLECTION:

We need a training data set. It is the actual data set used to train the model for performing various actions. In our experiment, we created our own dataset for investment prediction that contains the following set of columns as:

• Grade of the company - contains the grade value from 1 to 5 which represents from low to high grade.

• No. of offline projects and No. of online projects - refers to the projects that are taken by the related company.

• Net turnover - refers to the overall profits that gained by the company in that specific year.

• Share price - refers to the values holed by the shareholders in the company.

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter.

The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

An example of a decision tree can be explained using above binary tree. Let’s say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc.

The decision nodes here are questions like ‘What’s the age?’, ‘Does he exercise?’, and ‘Does he eat a lot of pizzas’? And the leaves, which are outcomes like either ‘fit’, or ‘unfit’. In this case this was a binary classification problem (a yes no type problem).

## There are two main types of Decision Trees:

1. Classification trees (Yes/No types) What we’ve seen above is an example of classification tree, where the outcome was a variable like ‘fit’ or ‘unfit’. Here the decision variable is Categorical.

2. Regression trees (Continuous data types) Here the decision or the outcome variable is Continuous, e.g. a number like 123.

# 4. METHDOLOGY:

## 4.1 Exploratory Data Analysis:

Exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA),[1] which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.

The objectives of EDA are to:

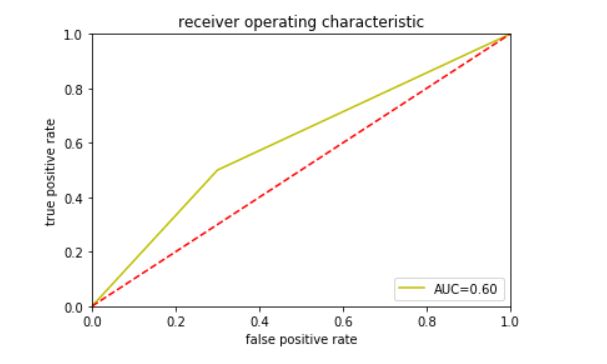
• Suggest hypotheses about the causes of observed phenomena

• Assess assumptions on which statistical inference will be based

• Support the selection of appropriate statistical tools and techniques

• Provide a basis for further data collection through surveys or experiments

## 4.1.1 Figures And Tables:



## 4.2 Statistical Techniques And Visualization:

Statistics is a collection of tools that you can use to get answers to important questions about data. You can use descriptive statistical methods to transform raw observations into information that you can understand and share. You can use inferential statistical methods to reason from small samples of data to whole domains. Statistics is a pillar of machine learning. You cannot develop a deep understanding and application of machine learning without it.

• Problem Framing: Requires the use of exploratory data analysis and data mining.

• Data Understanding: Requires the use of summary statistics and data visualization.

• Data Cleaning. Requires the use of outlier detection, imputation and more.

• Data Selection. Requires the use of data sampling and feature selection methods. • Data Preparation. Requires the use of data transforms, scaling, encoding and much more. Model Evaluation. Requires experimental design and resampling methods.

• Model Configuration. Requires the use of statistical hypothesis tests and estimation statistics. Model Selection. Requires the use of statistical hypothesis tests and estimation statistics. Model Presentation. Requires the use of estimation statistics such as confidence intervals. Model Predictions. Requires the use of estimation statistics such as prediction intervals.

## Numpy:

The data contains information on various attributes of wines, such as pH and fixed acidity, along with a quality score between 0 and 10 for each wine. The quality score is the average of at least 3 human taste testers. As we learn how to work with NumPy, we’ll try to figure out more about the perceived quality of wine. Numpy 2-Dimensional Arrays with NumPy, we work with multidimensional arrays.

We’ll dive into all of the possible types of multidimensional arrays later on, but for now, 8 we’ll focus on 2-dimensional arrays. A 2-dimensional array is also known as a matrix, and is something you should be familiar with.

## Creating A NumPy Array :

We can create a NumPy array using the numpy.array function. If we pass in a list of lists, it will automatically create a NumPy array with the same number of rows and columns. Because we want all of the elements in the array to be float elements for easy computation, we’ll leave off the header row, which contains strings. One of the limitations of NumPy is that all the elements in an array have to be of the same type, so if we include the header row, all the elements in the array will be read in as strings. Because we want to be able to do computations like find the average quality of the wines, we need the elements to all be floats.

In the below code, we:

• Import the Numpy package.

• Pass the list of lists wines into the array function, which converts it into a NumPy array.

• Exclude the header row with list slicing.

• Specify the keyword argument dtype to make sure each element is converted to a float.

***Pandas:***

**Pandas DataFrame** is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the **data**, **rows**, and **columns**.

We will get a brief insight on all these basic operation which can be performed on Pandas DataFrame :

* Lists
* dict
* Series
* Numpy ndarrays
* Another DataFrame

[**Column Selection**](https://www.geeksforgeeks.org/how-to-select-multiple-columns-in-a-pandas-dataframe/)**:**  In Order to select a column in Pandas DataFrame,can either access the columns by calling them by their columns name.

[**Row Selection**](https://www.geeksforgeeks.org/python-pandas-extracting-rows-using-loc/)**:**  Pandas provide a unique method to retrieve rows from a data frame. [DataFrame.loc[]](https://www.geeksforgeeks.org/python-pandas-extracting-rows-using-loc/) method is used to retrieve rows from Pandas data frame. Rows can also be selected by passing integer location to an [iloc[]](https://www.geeksforgeeks.org/python-extracting-rows-using-pandas-iloc/" \t "_blank) function.

# Matplotlib:

Matplotlib is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc. It supports a very wide variety of graphs and plots namely - histogram, bar charts, power spectra, error charts etc. It is used along with NumPy to provide an environment that is an effective open source alternative for MatLab. It can also be used with graphics toolkits like PyQt and wxPython.

Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source. Several toolkits are available which extend Matplotlib functionality. Some are separate downloads; others ship with the Matplotlib source code but have external dependencies.

• **Basemap**: map plotting with various map projections, coastlines, and political boundaries. 10

• **Cartopy**: a mapping library featuring object-oriented map projection definitions, and arbitrary point, line, polygon and image transformation capabilities. (Matplotlib v1.2 and above)

• Excel tools: utilities for exchanging data with Microsoft Excel

• GTK tools: interface to the GTK+ library.

## 4.3 Data Modelling And Visualization:

Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices,

APIs and online services as part of the Internet of Things. Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions. Elements of applications can be saved or shared for reuse. The runtime is built on Node.js. The flows created in Node-RED are stored using JSON. Since version 0.14 MQTT nodes can make properly configured TLS connections. In 2016, IBM contributed Node-RED as an open source JS Foundation project.

## 5. FINDINGS AND SUGGESTIONS:

The Securities market emerges out to mobilize primary savings from the public to serve as a resource of funds by issuing shares and providing liquidity to these instruments through regular quotations in the financial markets and thus traded and pave the path for wealth creation. Millions of investors are backbone of Indian Securities market. The Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) are the leading stock exchange in India.BSE has the distinction of being largest in the country. Securities market is known for tax-free returns, an effortless, easy entry into the stock market, higher returns, any time liquidity and to deliver higher real returns than any other investments. Based on the findings, following suggestions are given to investors, brokers and SEBI to overcome the problems faced by them.

## 6. CONCLUSION:

Stock market prediction is important factor in finance. It is considered to be dynamic in nature. The paper presented how to predict stock values based on the data using Machine Learning algorithms: Decision Tree. We also concluded this algorithm is better than the other algorithms because, within a certain range, the difference between actual price and predicted price is quite small as compared to those in Logistic Regression and Random Forest. Also, Random Forest is better than Logistic Regression, but inferior to MLP and Decision Tree, in predicting stock values

Thank you