STOCK PRICE PREDICTION USING MACHINE LEARNING

[1]B.Vijay Kumar, [2] B.Sai Kiran, [3] Komal Singh [4] B.Srikar

[1]Assistant Professor, [2]Student, [3]Student, , [4]Student
[1]Department of Information Technology ,
[1]JBIET, Hyderabad, India

Abstract : Investing, in stock markets, is highly risky for investors, particularly those who are new to the stock market. As people's interest in forecasting stock prices has been recently increased, research on stock price analysis through big data and machine learning has been actively conducted. In this paper, we performed decision tree analysis for stock price prediction. A Decision tree is a tree-like model of decisions and their possible consequences. Here we are using a Decision support tool which builds decision trees from a set of training data, using the concept of information entropy. Entropy is one of the most commonly used discretization measures. It was first introduced by Claude Shannon in pioneering work on information theory and the concept of information gain. Entropy-based discretization is a supervised, top-down splitting technique. This makes this unpredictable format of business a little more predictable. We shall see how this simple implementation will bring acceptable results. Here we implemented through three modules: Data entry module, is used to provide the needed data to the project. The Analysis module is used to analyze the stock prices using multivariate analysis. The Front end module is used to create the needed GUI screens for the project.

IndexTerms - **Decision tree, Supervised Learning, Discretization, Entropy, Top-Down Splitting**

I. INTRODUCTION

The stock markets in the recent past have become an integral part of the global economy. Any fluctuation in this market influences our personal and corporate financial lives, and the economic health of a country. The stock market has always been one of the most popular investments due to its high returns. However, there is always some risk to investment in the stock market due to its unpredictable behavior. Stock market prediction is a classic problem which has been analyzed extensively using tools and techniques of Machine Learning. The successful prediction of a stock's future price could yield significant profit. So, an intelligent prediction model for stock market forecasting would be highly desirable and would be of wider interest. Machine learning techniques are very popular in building real- time application model which provide better result than other methods.

Many methods have been developed to predict the stock market related data using various techniques and models where they have used sentimental, past prices, sales growth and dividends we know that to predict the stock efficiently we need data along with one of the above factor where we can build the efficient market hypothesis where major companies are investing their money into we selected machine learning with other techniques just because they give us the better results than the other random prediction model. Supporting vectors in machine learning can provide the classification and regression for the model and few other trending models such as deep neural network, linear regression etc

Data collection is very important in this aspect as we rely on the data very much as we know predicting the stock market data. So data must capture based on the model if we are predicting for intraday trading then we have to capture the data every minute or so that we can train the model in real time. If we are trying to predict swing, positional and options strategies then we have to collect the data based on that.

This Project uses PyQt tool to create the needed Graphical User Interfaces, PyUIC module to automatically generate the code. PyUic tool is used automatically generate the code for the Front end user interfaces created by PyQt. All the front end python code is automatically generated by this tool, by converting the user interface (.ui) files into .py files.

A decision tree is a tree whose internal nodes can be taken as tests (on input data patterns) and whose leaf nodes can be taken as categories (of these patterns). These tests are filtered down through the tree to get the right output to the input pattern. Decision Tree algorithms can be applied and used in various different fields. It can be used as a replacement for statistical procedures to find data, to extract text, to find missing data in a class, to improve search engines and it also finds various applications for prediction. Many Decision tree algorithms have been formulated. They have different accuracy and cost effectiveness. It is also very important for us to know which algorithm is best to use. The decision support tool analyzes the variations in stock prices, using decision tree analysis. A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences. The tool builds decision trees from a set of training data, using the concept of information entropy. Traditional approaches to this problem is to collect relevant factors pertaining to the head injury, values and then search for the most optimum combinations of these factors that influence the actual outcome after a severe head injury. These approaches are based on gathering in large numbers the data of patients with severe head injury and on the application of statistical methods. Sometimes we are given a set of records, which we have to analyze and form a conclusion accordingly. In order to do so we often use statistics. But instead, now we propose the use of decision trees instead of statistics as they are more accurate and also they highlight some important attributes which we may often overlook.

LITERATURE SURVEY

Fundamental analysis is built on the belief that human society needs capital to make progress and if a company operates well, it should be rewarded with additional capital and result in a surge in stock price. Fundamental Analysts are concerned with the company that underlies the stock itself. They evaluate a company's past performance as well as the credibility of its accounts. Many performance ratios are created that aid the fundamental analyst with assessing the validity of a stock, such as the P/E ratio. Technical analysis is rather used for short- term strategies, than the long-term ones. And therefore, it is far more prevalent in commodities and forex markets where traders focus on short-term price movements. Technical analysts or chartists are not concerned with any of the company's fundamentals. They seek to determine the future price of a stock based solely on the trends of the past price (a form of time series analysis).

In the past, various approaches have been applied to analyze stock trend. Radu Jacomin analyzed stock with the help of SVM .The shortcoming of the approach was discussed and moving forward, two new techniques were then evolved namely, GASVM and PCASVM. Priti Saxena and Bhaskar Pant along with other fellow authors , generated the linguistic forms for raw data comprising of opening and closing price. Apriori model was then applied on the data to find out the most frequent patterns Prof. Dr. K. Raza analyzed stock with the help of four different versions of Artificial Neural Network. It went on to establish an association of stock behavior with change in petrol prices and finally concluded that it was possible to analyze stock with the help of machine learning techniques. In our model, we first applied different existing algorithms on stock market and compared their results. Then in our proposed model, we provided a model which would check itself at each stage for the correctness in analyzing share trends.

The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy . Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. To achieve those objectives, some research used the techniques of fundamental analysis, where trading rules are developed based on the information associated with macroeconomics, industry, and company. The fundamental analysis assumes that the price of a stock depends on its intrinsic value and expected return on investment. Analyzing the company's operations and the market in which the company is operating can do this. Consequently, the stock price can be predicted reasonably well. Most people believe that fundamental analysis is a good method only on a long-term basis. However, for short- and mediumterm speculations, fundamental analysis is generally not suitable.

Some other research used the techniques of technical analysis, in which trading rules were developed based on the historical data of stock trading price and volume. Various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit.

PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Decision Tree Multivariate analysis and by using the Entropy. We proposed the system "Stock price prediction using Machine Learning" we have predicted the stock market price using the Multi variate Analysis. In this proposed system, we were able to train the machine from the various attributes of data points from the past to make a future prediction. We took data from the previous year stocks to train the model. We majorly used two machine-learning libraries to solve the problem. The first one was pandas, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other was sklearn, which was used for real analysis and prediction and it has containing various in-built functions like Linear regression and Decision Tree Classifier(). The data set we used was from the previous years stock markets collected from the public database available online, Some of data was used to train the machine and the rest some data is used 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 raw data makes us to prepare the data for feature identification. The attributes were date, open, low, high and the closing price for a particular day. We used all these features to train the machine on Decision Tree model 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 gives the Predicted price.

SYSTEM ARCHITECTURE

The first step in this is collection of raw data from the various sources and dataset can be of any historical data of the organization. From the raw data we can extract the attributes which are used for the prediction. After extraction, we can train the data model using these previous datasets. Here we should give Testing data(input) to data analytical tool it will compare data with the trained data and gives the Predicted price as output to us.
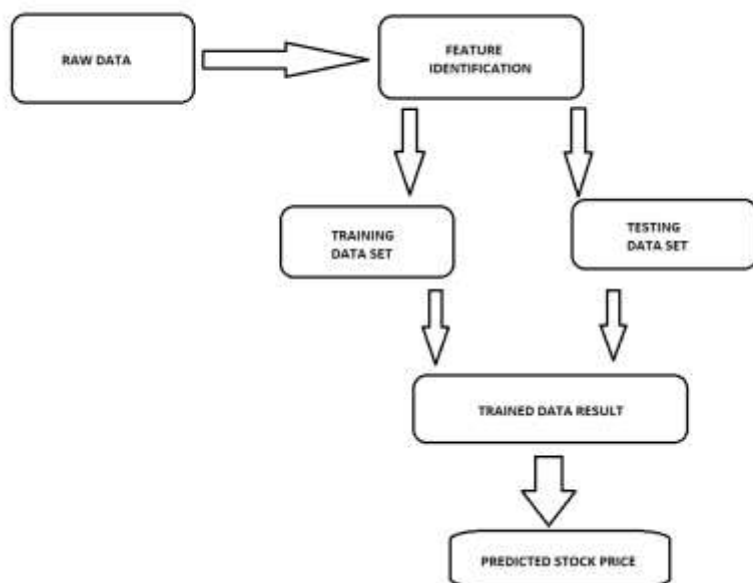
ALGORITHM :

ENTROPY CALCULATION

We used the python pandas library for data processing which combined different datasets into a data frame. The raw data makes us to prepare the data for feature identification. The attributes were date, open, low, high and the closing price for a particular day. We used all these features to train the machine on Decision Tree model 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.

1.Importing Data manipulation library files. //Pandas
2.Initialize attributes. #col_names= ['low','high','low','high','riskcategory']
3.Load dataset and read using read_csv.
4.Split dataset in features and target variable.
5.Calculate total risk category count using entropy.
6. Taking total risk category count as t=t1+t2+t3. 7. Calculate using Entropy Formula
$E=(-1*((t1)/(t)*\log((t1/t),2))-((t2)/(t)*\log((t2/t),2))-((t3)/(t)*\log((t3/t),2)))$

8..dot file generation:
9.from sklearn import tree
10.Used for Representation. //from sklearn.tree import export_graphviz
11.Put dataset feature columns in X. //X = eqts[feature_cols] #Features
12.Set Target variable. //y=dataset riskvalue
13.classify the model//clf = tree.DecisionTreeClassifier(criterion="entropy")
14.Fit values to the model //clf = clf.fit(X, y)
15.Represent Decision Tree.
  ///export_graphviz(clf, out_file="descntree.dot")
16.Built linear model using linear regression
17.Predict stock price

MODULES

I. *Data Collection*

Firstly, Dataset can be collected from various sources of any organization. The right dataset helps for the prediction and it can be manipulated as per our requirement. Our data mainly consists of the previous year stock prices. The data can be collected from the organization based on the market trading, noof days traded, shares ratio, no of shares traded. By collecting these it makes accurate in prediction.

II. *Data Processing*

At the beginning, when the data was collected, all the values of the attributes selected were continuous numeric values. Data transformation was applied by generalizing data to a higher-level concept so as all the values became discrete. The criterion that was made to transform the numeric values of each attribute to discrete values depended on the closing price of the stock. If the values of the attributes open, min, max, last were greater than the value of attribute previous for the same trading day, the numeric values of the attributes were replaced by the value Positive. If the values of the attributes mentioned above were less than the value of the attribute previous, the numeric values of the attributes were replaced by Negative. If the values of those attributes were equal to the value of the attribute previous, the values were replaced by the value Equal.

III. *Training the Data*

After the data has been prepared and transformed, the next step was to build the classification model using the decision tree technique. The decision tree technique was selected because the construction of decision tree classifiers does not require any domain knowledge, we can done by using the Decision Tree Classifier() which is in the sklearn library. By using the attribute we have considered in the dataset we train the model by using the algorithm. The training sets are used to tune and fit the models.

IV. *Deploying the Model*

The classification rules are generated from the decision tree algorithm. The trained data can be used for the Testing the data. It helps to give the output or accurate Predicted price of the stock using this model.

CONCLUSION

This project entitled "Stock Price Prediction Using Machine Learning Algorithm." is useful to the stock exchanges, in predicting stock prices, and thereby to guide their customers accordingly. The proposed system is also useful to the newbie's to the stock market, in properly investing into the stocks. This helps finally leads to the improvement of stock markets customers' satisfaction.

REFERENCES

1. Dr.Devpriya Soni, Sparsh Agarwal, Tushar Agarwal, Pooshan Arora, Kopal Gupta- "Optimised Prediction Model For Stock Market Trend Analysis" In: IC3 2018, IEEE.

2. K.Hiba sadia, Aditya Sharma , Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal "Stock Price Prediction Using Machine Learning Algorithms". IJEAT, 2019.

3. Han Lock Siew and Md Jan Nordin.-"Regression Techniques for the Prediction of Stock Price Trend" 2012,IEEE.

4. Tanapon Tantisripreecha and Nuanwan Soonthornphisaj –"Stock Market Movement Prediction using LDA-Online Learning". 2018 IEEE.