JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

STOCK MARKET CLOSING PRICE ANALYSIS AND PREDICTION USING THE STUDY OF MACHINE LEARNING TECHNIQUE.

Prof.Mr.Subash Chandar
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Koushik R
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Kishore Kumar S
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Hariharan E
Department of Computer Science and Engineering
Jeppiaar Engineering College
Chennai,India

Mr.Subash Chandar
Professor
Jeppiaar Engineering College
Chennai,India

Abstract:

Stock price forecasting is a popular and important topic in financial and academic studies. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Accurate prediction of stock market returns is a very challenging task due to volatile and non-linear nature of the financial stock markets. To create software that analyses previous stock data of certain companies, 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 machine learning algorithms.

I.INTRODUCTION :

In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers .The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price .Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market

will perform is a hard task to do. There are other factors involved in the prediction, such as physical and psychological factors, rational and irrational behaviour, and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy.Nowadays, many people are indirectly or directly related to this sector. Therefore, it becomes essential to know about market trends. Thus, with the development of the stock market, people are interested in forecasting stock price. But, due to dynamic nature and liable to quick changes in stock price, prediction of the stock price becomes a challenging task. Stock markets are mostly a non-parametric, non-linear, noisy and deterministic chaotic system.

II.LITERATURE SURVEY

[1]TITLE: Research on Stock Price PredictionMethod Based on Convolutional NeuralNetwork.

AUTHOR:LounnaphaSayavong, Zhongdong Wu , Sookasame Chalita.

YEAR: 2019

This paper tends to make a prediction model for stock price in which the convolutional neural networks is given the most importance, that has exceptional capability of learning on its own. The data set that is provided is taught and tested relating to the behaviours of both Convolutional Neural Networks(CNN) and Thai stock market . The result shows that the model centered at Convolutional Neural Networks can effectively recognize the changing trends in stock market closing price and envisage it which provides significant illusion for stock price forecast. The accuracy of the prediction is found to be increased, and it could also be promoted in the field of finance and so on.

2]TITLE: Enhancing Profit by Predicting Stock Prices using Deep Neural Networks.

AUTHOR: Soheila Abrishami, et al.,

YEAR: 2019

The prediction of economic time series is quite a tough task, which has fascinated the attentiveness of many experienced financial people and is extremely vital for the investors who invest in stocks. This paper tends to focus on presenting a deep learning system, which makes use of a wide range of facts for the stocks on the NASDAQ exchange to

predict the price value of the stock. This model has been given a training on the smallest of data for a particular stock data and it precisely estimates the final value of that stock for multi-step-ahead. It consists of an auto-encoder which is used to remove the noise and makes use of time series data engineering to syndicate the advanced features with the original features. These new

features are then given to a Stacked LSTM Autoencoder for multi-step-ahead estimation of the stock concluding value. Further, the estimation that is found precisely is used by a profit maximization approach to give the assistance on the right time for buying and selling a particular right stock. The results indicate that the suggested method outclasses the state of the time series forecasting methodologies with respect to analytical accuracies ,efficiency and effectiveness.

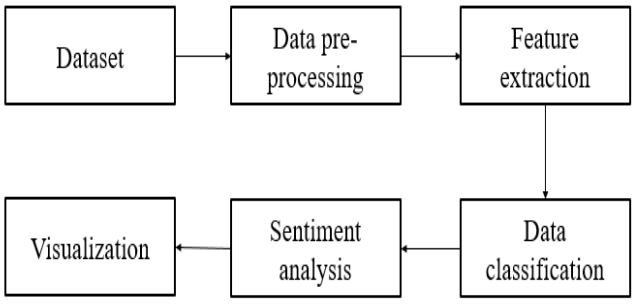
[3] TITLE: STOCKS MARKET PREDICTION USING SUPPORT VECTOR MACHINE.

AUTHOR: ZHEN HU , JIE ZHU , KEN TSE

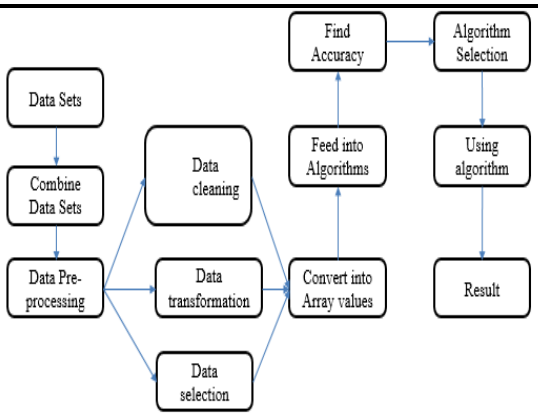
YEAR:2013

Recent studies suggest particular strategies that overcome these problems. Support Vector Machine (SVM) is a relatively advanced machine learning algorithm that has the desirable properties of the control of the decision function, the use of the kernel method etc... Through this paper, we present a theoretical and experimental explanation to apply the Support Vector Machines(SVM) method to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may affect the stock trend are grouped for further stock multivariate analysis. Secondly, the Support Vector Machine(SVM) is used to evaluate the factors one on one relationships and to predict the performance. Our results proves that SVM is a powerful predictive tool for stock predictions in the financial market and so on.

III.BLOCK DIAGRAM



SYSTEM ARCHITECTURE DIAGRAM



IV.IMPLEMENTATION

- **Data Collection**
  - Data cleaning
  - Data transformation
  - Data selection
- **Pre-processing**
- **Training Dataset**
- **Testing Dataset**
- **Algorithm**
- **Evaluation**
- **Result**

Collecting Dataset

- Data Collection is one of the most important tasks in building a machine learning model.We collect the specific dataset based on requirements from internet.
- The dataset contains some unwanted data also. So first we need to pre-process the data and obtain perfect data set for algorithm.

Pre-processing

- However, some of the data may be noisy, i.e. may contain inaccurate values, incomplete values or incorrect

values.Hence, it is must to process the data before analysing it and coming to the results.

Pre-processing techniques

- Data cleaning includes Fill in missing values, smooth noisy data, identify or remove outliers, and resolve in consistencies.
- Data transformation may include smoothing,aggregation, generalization, transformation which improves the quality of the data.

Training Dataset

Training dataset is a subset of the dataset used to build and fit predictive models. A training set is generated by building a training dataset script, which generates the training set features from the input options and the raw stock price data. The data is fed into the model for training. The model learns from this data and runs on the train set.

Testing Dataset

Testing Dataset is a subset of the dataset to assess the likely future performance of a model. It is a good standard for evaluating the model. Testing set is used against the predicted dataset and testing the model that is trained. It is used for evaluation purposes.

V.ALGORITHMS

[1]Random forest is an ensemble of decision trees. This is to say that many trees, constructed in a certain "random" way form a Random Forest. Each tree is created from a different sample of rows and at each node, a different sample of features is selected for splitting. Each of the trees makes its own individual prediction.

These predictions are then averaged to produce a single result. The averaging makes a Random Forest better than a single Decision Tree hence improves its accuracy and reduces overfitting**.**

[2]Decision trees tend to be the method of choice for predictive modeling because they are relatively easy to understand and are also very effective. The basic goal of a decision tree is to split a population of data into smaller segments. There are two stages to prediction. The first stage is training the model—this is where the tree is built, tested, and optimized by using an existing collection of data. In the second stage, you actually use the model to predict an unknown outcome. Simple to understand and to interpret. Requires little data preparation. The cost of using the tree (i.e., predicting data) is logarithmic in the number of data points used to train the tree. Able to handle both numerical and categorical data.Able to handle multi-output problems.

[3]LTSMs are a type of Recurrent Neural Network for learning long-term dependencies. It is commonly used for processing and predicting time-series data. Long-Short-Term Memory Recurrent Neural Network belongs to the family of deep learning algorithms. It is a recurrent network because of the feedback connections in its architecture. It has an advantage over traditional neural networks due to its capability to process the entire sequence of data. Its architecture comprises the cell, input gate, output gate and forget gate.

[4]Linear regression is commonly used for predictive analysis and modeling. For example, it can be used to quantify the relative impacts of age,

gender, and diet (the predictor variables) on height (the outcome variable). ***Linear regression*** quantifies the relationship between one or more *predictor variable(s)* and one *outcome variable*. Linear regression is commonly used for predictive analysis and modeling. For example, it can be used to quantify the relative impacts of age, gender, and diet (the predictor variables) on height (the outcome variable). Linear regression is also known as *multiple regression*, *multivariate regression*,*ordinary least squares (OLS)*, and *regression*. This post will show you examples of linear regression, including an example of *simple linear regression* and an example of *multiple linear regression*.

[5]Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

VI.EXISTING SYSTEM

Existing approaches to stock market analysis and stock price prediction include fundamental analysis, which looks at a past performance of stocks and the general credibility of the company itself, and statistical analysis, which aims in identifying patterns in stock price variation. The latter is commonly achieved with the help of Artificial Neural Networks, but these fail to capture correlation between stock prices in the form of long-term temporal dependencies,also Artificial neural networks prediction get worse with increased noise variation.Many used algorithms like Arima model which is suitable only for short term prediction.An alternative approach to stock market analysis is to reduce the dimensionality of the data that is fed and to apply feature selection algorithms to filter the core set of features that have the greatest impact on stock prices or currency exchange rates across markets .By using different algorithms, we tend to calculate and understand the algorithms accuracy and efficiency.

VII.PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest,Long Short Term Memory,Linear Regression, Decision Tree ,Support Vector Machines. We proposed the system “Stock market Closing price Analysis and Prediction Using Machine Learning Algorithms” we have predicted the stock market price using the above mentioned algorithm. 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. We majorly used machine-learning libraries like numpy and scikit to solve the problem. The first one was numpy,
which was used to clean and manipulate the data, and getting it ready for the analysis. The other was scikit, which was used for real analysis and

prediction. The data set we used was from the previous years stock markets collected from the public database available online from Data Reader using the corresponding companies Keys, 70 % of data was used to train the machine and the rest 30 % 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 using the things that it learned from training set ,it applies them for the test data. We used the python pandas library for data processing which combined different datasets into a data frame. The Features includes Open,Close,Date,Volume, Adjacent Close etc... The dataframe features were date and the closing price for a particular day. We used all these features to train the machine on the above mentioned models 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 data values. The proposed system touches different areas of research including data pre-processing, data visualization,data validation and so on.

VIII CONCLUSION AND FUTURE GOALS:

By measuring the accuracy of the different algorithms, we can conclude that Long short term memory and random forest gives us a good accuracy and hence forecasts the stock price for future with a good accuracy percentage. These algorithm will be a great asset for people who invest 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. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.we are predicting the closing stock price of any given organization, we have developed an application for predicting close stock price using LSTM,RandomForest,Support Vector Machines, Linear Regression, Decision Tree algorithm. We have used datasets belonging to Apple and achieved above 93% accuracy for these datasets.

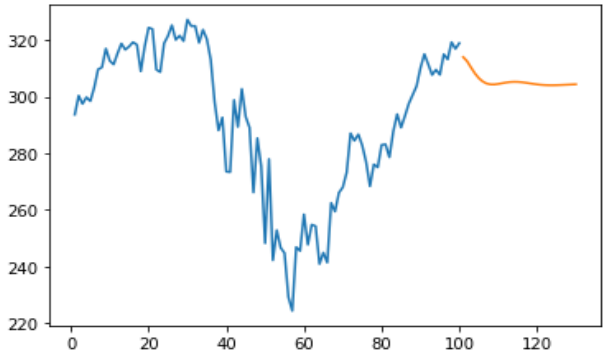
Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. In the future, we can extend this application for predicting cryptocurrency trading and also, we can add sentiment analysis for better predictions.

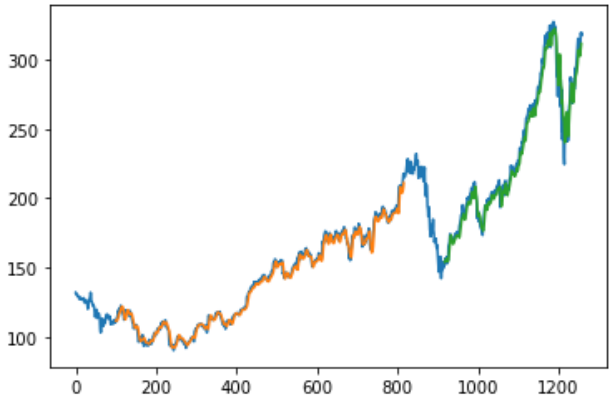
IX OUTPUTS:

LSTM:

```
1 day input [0.8866419  0.87431394 0.88431985 0.87836697 0.8986321  0.92582116
 0.92877649 0.95676771 0.93869797 0.93304061 0.94950604 0.96424048
 0.95512117 0.95989192 0.96635143 0.96246728 0.92295027 0.9598497
 0.98792536 0.98594106 0.92531453 0.92172591 0.96474711 0.97572406
 0.99159841 0.96972895 0.97614625 0.96795575 1.         0.99016297
 0.99050072 0.96538039 0.98488559 0.97086887 0.94026007 0.87748037
 0.83483915 0.85413324 0.77336823 0.77269273 0.88014017 0.84007431
 0.89673225 0.85527316 0.83884995 0.74233725 0.82327113 0.78143207
 0.6665963  0.7921557  0.64118044 0.68614371 0.66001013 0.65203074
 0.58642236 0.56586169 0.66089673 0.65515494 0.70970193 0.66452757
 0.69437642 0.69218104 0.63569197 0.65266402 0.63780292 0.7267162
 0.71388162 0.74191506 0.75002111 0.77222832 0.83049059 0.8194292
 0.8289707  0.8125475  0.78776492 0.75162543 0.78426074 0.77974331
 0.81326522 0.8141096  0.79473106 0.83336148 0.85898843 0.83901883
 0.85628641 0.87486279 0.88782403 0.90095415 0.92793211 0.948535
 0.93333615 0.91746179 0.92544119 0.91771511 0.9483239  0.94064004
 0.96635143 0.9563033  0.96491598 0.94413203]
1 day output [[0.9379593]]
```

```
[<matplotlib.lines.Line2D at 0x2d1b0f352b0>]
```



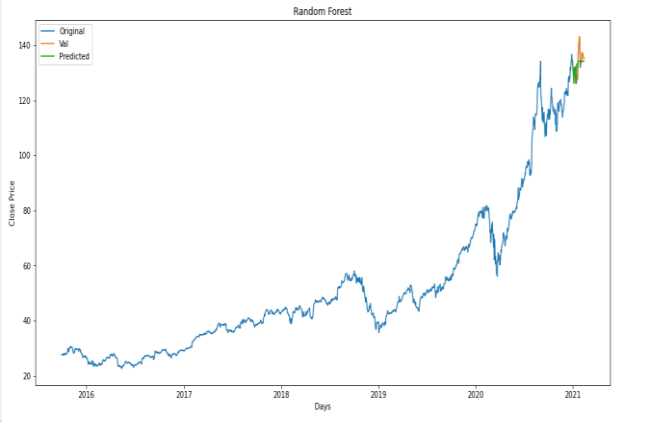


RANDOM FOREST:

```
# Split the data into 80% training and 20% testing
X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size=0.3, random_state=44)
```

```
from sklearn.ensemble import RandomForestRegressor
regr = RandomForestRegressor(random_state=0)
regr.fit(X_train, Y_train)
rf_confidence = regr.score(X_test, Y_test)
print("rf accuracy: ", rf_confidence)
```

```
rf accuracy:  0.9555319654164233
```



X REFERENCE:

[1] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3] H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

[4] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[5] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8] T. Gao, Y. Chai and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.