ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Predictive Analysis of Market Trends for Stock Price Predictions Using Machine Learning and Deep Learning

Nirmala Choudhary

Asst. Professor, RIET, Jaipur , Rajasthan , India

ABSTRACT: The project aims at building a predictive model of market trends for stock price predictions using Deep learning and Machine learning. There are a lot of factors such as past prices, sentiment analysis, brand value among others which can be used in harmony to predict the results with a high accuracy rate. We have implemented different algorithms to achieve highest accuracy. The factors taken into consideration are opening price and Google trends value. Implemented machine learning algorithms are Linear Regressor, SGD Regressor, Decision Tree Regressor, Random Forest Regressor and KNN Regressor. Besides, we have also implemented Deep Neural Networks Algorithm.

KEYWORDS: Stock prediction; machine learning; deep learning; predictive analysis; regression; random forest; neural networks

I. OUTLINE

People have been investing in stock markets for a long time. People have become very rich, when stock price of a particular company they have invested in rises and vice-versa. However, people often faced difficulties in handling stock market because they were unable to respond quickly to changes. This is where computers came in. Different algorithms were used to learn the behavior of stock market which was later utilized to predict the outcomes.

II. STOCK PRICE FLUCTUATION

Stock prices are directly proportional to the demand. If the demand of any product increases, the stock price or market value of that product increases. Similarly, if the demand of any product falls, we will face a sudden fall in the stock price of the product. To make a profit by investing in the stock market, we need to analyze and understand the time of the fall and rise of the stock market. To be able to do this, we need to have a detailed study of the behaviour of the stock market history. However, we have to keep in mind that if the stock market behaved in a certain way to the occurrence of a particular event, it is not necessary that the stock market will behave in the exact similar pattern this time. To say the least, stock market is unpredictable. However, we can study and provide the best available results.

III. STOCK PRICE DETERMINATION

At any given instance, stock is dependent on demand, as we have stated earlier. Moreover, it is also dependent on the supply. Suppose, there is a demand for a particular product. For the stock to soar, it is necessary that there must be enough supply to fulfill the requirement. So, to determine the stock price of a product in the near future, we need to analyze and understand what will be the demand for that particular product and also the corresponding supply. If the supply and demand of the product is healthy, the stock price will also be very high.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

IV. PROBLEM STATEMENT

Can we predict stock prices with the aid of machine learning? Investors make educated guesses by analyzing data. They read the news, company history, industry trends among others to make an educated decision. There are lots of features that go into making a decision. Top firms like Morgan Stanley, Citigroup, D.E. Shaw hire quantitative analysts to build predictive models. We are now living in the age of algorithms. Records of prices for traded commodities go back thousands of years. In Finance, the field of quantitative analysts is about 25 years old and even now it is still not fully accepted, understood or widely used.

V. MACHINE LEARNING

Machine learning has become the state of the art computer science field in recent years. It provides computers the ability to learn from the data and predict the outcome based on the user input without being explicitly programmed. Machine learning does an excellent work on finding the pattern that exist between the data. These patterns can vary from data set to data set and the machine learning algorithms does an eloquent job in finding them provided apt algorithm is chosen to build the model.

VI. SYSTEM DESIGN

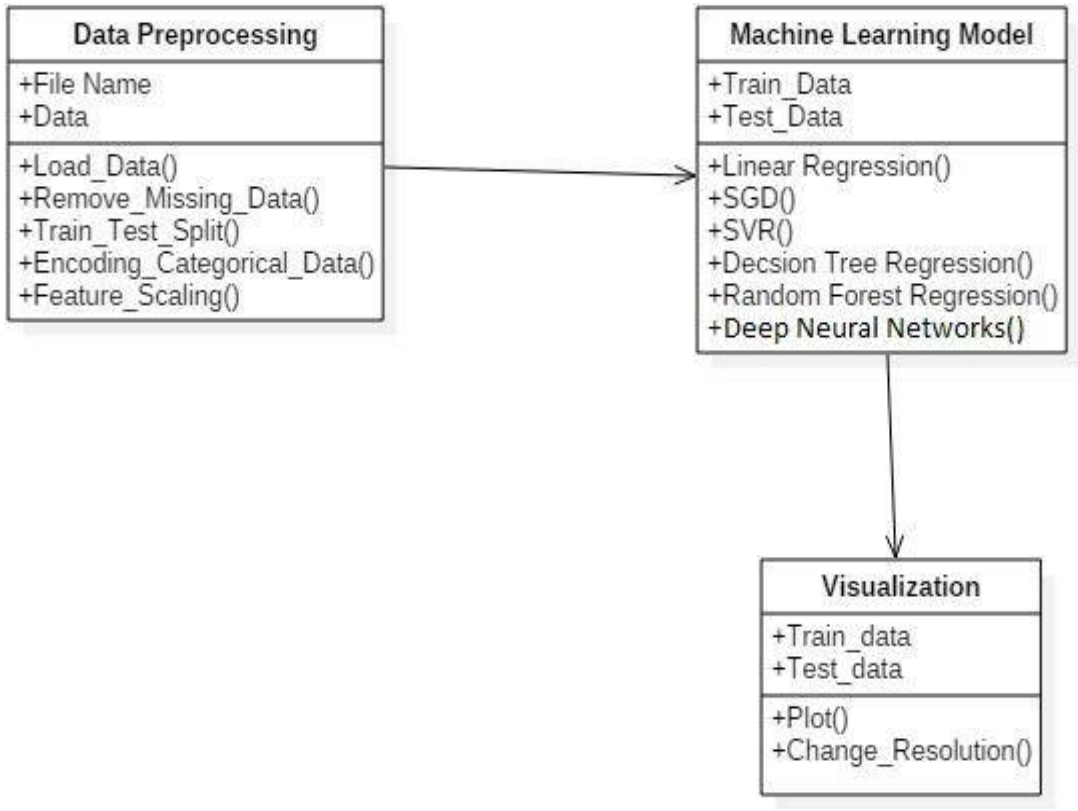

Figure 1: Class Diagram

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

VII. SYSTEM TESTING

The dataset is divided into train and test data and then we match the actual values with the predicted values. For simplicity, we are attaching only ten actual and their corresponding predicted values for each algorithm we have performed.

VIII. LINEAR REGRESSION

Linear Regression is a machine learning algorithm in which one variable is depen-dent while other variables are independent. In order to apply this algorithm, there should be a linear dependency between the dependent and independent variable.
The Actual Data in the table is from the test dataset corresponding to a partic-ular date. The Predicted data is the value predicted by Linear Regression algorithm on test data. Percentage Change represents the percentage of absolute difference between predicted data and actual data.

Date	Actual Values	Predicted Values	Percentage Change
2016-12-15	115.379997	114.9825627	-0.344456847
2016-12-16	116.470001	115.0076082	-1.255596108
2016-12-19	115.800003	115.4340115	-0.316054828
2016-12-20	116.739998	115.4590571	-1.097259656
2016-12-21	116.800003	115.4841026	-1.126627026
2016-12-22	116.349998	115.5091481	-0.722690085
2016-12-23	115.589996	115.5341936	-0.04827615
2016-12-27	116.519997	115.927098	-0.508838839
2016-12-28	117.519997	115.9521435	-1.334116355
2016-12-29	116.449997	115.977189	-0.406018044
			-0.715993394

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

STOCHASTIC GRADIENT DESCENT

It is a unsupervised machine learning algorithm in which the gradient of mini-batch of samples is calculated rather than taking the entire dataset at a time. It is computationally very faster in comparison to Batch Gradient Descent.

The Actual Data in the table is from the test dataset corresponding to a partic-ular date. The Predicted data is the value predicted by SGD Regressor algorithm on test data. Percentage Change represents the percentage of absolute difference be-tween predicted data and actual data.

Date	Actual Values	Predicted Values	Percentage Change
2016-12-15	115.379997	108.972979	-5.552970937
2016-12-16	116.470001	109.1948491	-6.246374034
2016-12-19	115.800003	113.0789682	-2.349770924
2016-12-20	116.739998	113.300838	-2.945999708
2016-12-21	116.800003	113.5227078	-2.805903353
2016-12-22	116.349998	113.7445776	-2.239295612
2016-12-23	115.589996	113.9664474	-1.404575358
2016-12-27	116.519997	117.536018	0.871971358
2016-12-28	117.757887	117.7578877	0.20242572
2016-12-29	116.449997	117.9797575	1.313662979
			-2.115682987

DECISION TREE TEST

It is a supervised learning method that can be used for both classification as well as regression problems. In this, each node represents a test on an attribute, each branch represents the output of the test and each leaf node represents class label. The model created by it predict the value of target variable by learning some rules inferred from data model. The Actual Data in the table is from the test data set corresponding to a particular date. The Predicted data is the value predicted by Decision Tree Regressor algorithm on test data. Percentage Change represents the percentage of absolute difference between predicted data and actual data.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Date	Actual Values	Predicted Values	Percentage Change
2016-12-15	115.379997	113.6999985	-1.45605698
2016-12-16	116.470001	113.6999985	-2.378296966
2016-12-19	115.800003	113.6999985	-1.813475342
2016-12-20	116.739998	116.5613327	-0.153045488
2016-12-21	116.800003	116.5613327	-0.204341005
2016-12-22	116.349998	116.5613327	0.181637047
2016-12-23	115.589996	116.5613327	0.840329383
2016-12-27	116.519997	116.5613327	0.035475198
2016-12-28	117.519997	116.5613327	-0.815745681
2016-12-29	116.449997	116.5613327	0.09560816
			-0.566791167

VIII. IMPLEMENTATION

DATA PREPROCESSING

The dataset contains following columns:

Date - Date
Open - Opening price of particular date
High - Highest price at particular date
Low - Lowest price at particular date
Close - Closing price at particular date
Adj Close - Adj. Close price at particular date
Volume - Volume of stock traded

From all the columns, we need Date and Open for our research work. We extracted Date and open from the table to make our new dataset.

Date is in the format YYYY-MM-DD so we need to break it in numerical form i.e, YY,MM,DD. We decompose the Date column into three columns.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

New Dataset Structure before Decomposition of Date:

Date(YYYY-MM-DD) (Date Object)	Open (Float)

After Decomposition of Date:

Year(YYYY) Integer	Month(MM) Integer	Day(DD) Integer	Open (Float)

- Dataset contains data from Year 2007 to 2016. We split the data into train and test data. Train dataset contains data from range 2007-2014 and Test Dataset contains from 2015-2016.
- The Date here is the features and the Open price is our target values which needs to be predicted. X(features) = [Year,Month,Day], y(target) = [Open]
- Feature scaling has been done on data for faster convergence rate of algorithms and also to maintain standardisation in data. This concludes our data prepro-cessing.
- Function to plot graph is written which takes dataset as parameters and plots the graph.
- To train the model different Regression algorithms have been used. The Train data is fitted into the algorithm and the model is trained.
- The models accuracy is then tested by giving Test data as input and evaluating the predicted result against the known value from Test data.

The graph is plotted between Actual value vs Predicted value.
Below is the flow chart for above procedure:

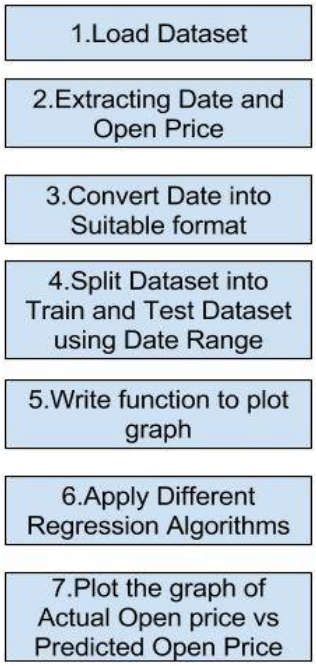

Figure 2 : Flow Chart

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

IX. ALGORITHMS

LINEAR REGRESSION

Linear Regression is a machine-learning algorithm in which one variable is dependent while other variables are independent. In order to apply this algorithm, there should be a linear dependency between the dependent and independent variable.

```
from sklearn.linear model import LinearRegression,Ridge
from sklearn.metrics import explained variance score
model linear reg = LinearRegression()
model linear reg.fit(X train,y train)
df lr = pd.DataFrame(index=df test.index)
df lr['Actual'] = y test
df lr['Predicted'] = model linear reg.predict(X test)
print("% Change : "+ str(calculateError(df lr['Predicted'],df lr['Actual'])
))
plotDataframe(df lr,"Linear Regression")
```

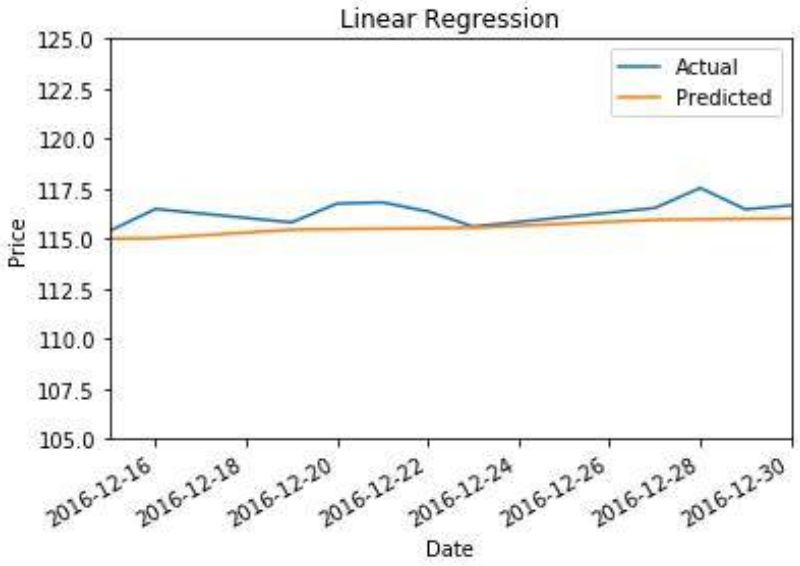

Figure 3 : Linear Regression Graph

Error: Margin of error ranges from -0.04827615 to -1.334116355

X. COMPARATIVE ANALYSIS

We have implemented 5 Machine Learning Regression algorithms namely, Linear Regression, SGD, Decision Tree, Random Forest, KNN along with Deep Neural Networks. We have pre-processed the data and used it as input to our algorithms. The results or outputs have been obtained and we have conducted a detailed study of the results to identify the best algorithm for stock market prediction.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

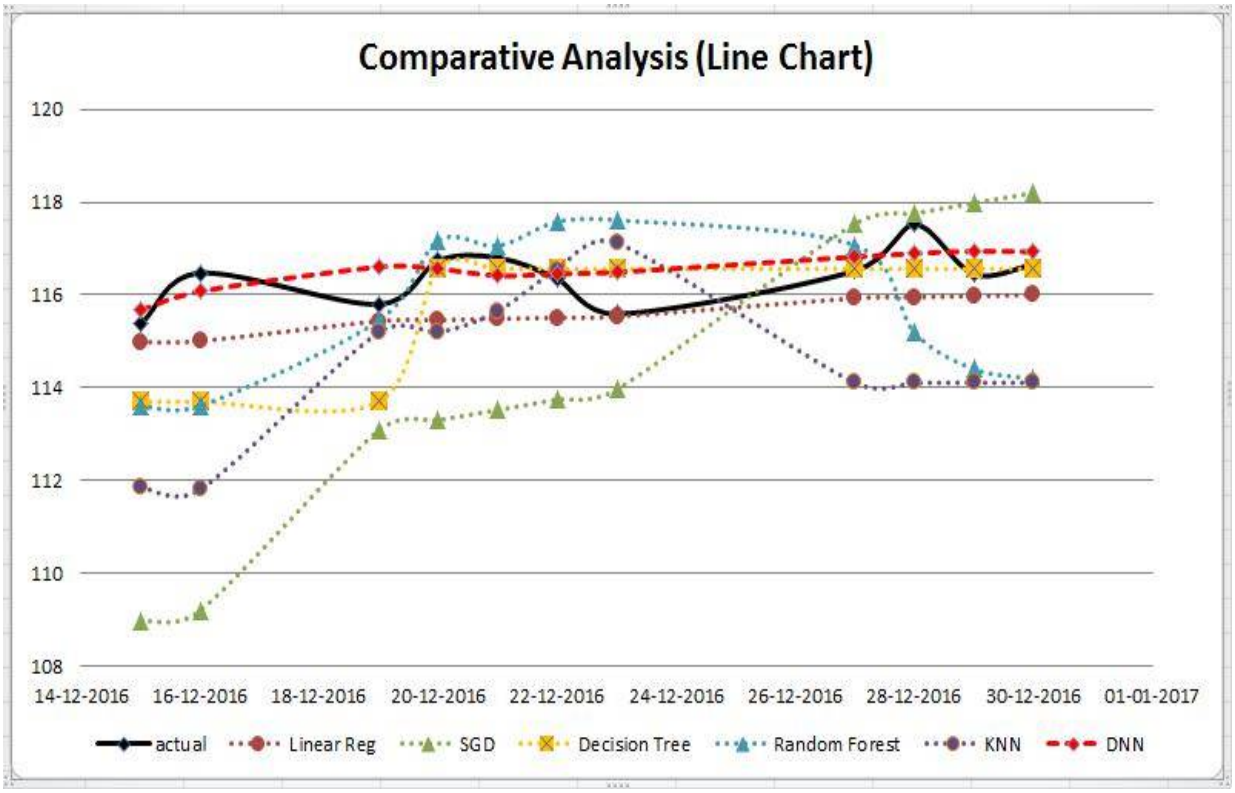

Figure 4 :Comparative Analysis (Line Graph)

As we can see in the line chart, the black line represents the actual values while the predicted values are indicated by other colors and designs. We have highlighted the predicted values obtained from Deep Neural Networks as it is the best result obtained among the implemented algorithms.

XI. CONCLUSION

Working with Machine Learning on stock market analysis and prediction, we have come to the conclusion that there is a huge scope of work in this eld .As this eld is comparatively new, there are endless possibilities. The algorithms we have so far implemented on our dataset are:

- Linear Regressor
- SGD Regressor
- Decision Tree Regressor
- Random Forest Regressor
- KNN Regressor
- Deep Neural Networks

The main aim of this project is to understand and analyses the future state of the market. Having knowledge of the market beforehand will help people avoid losses. Our machine-learning model can pave the way for a better future. Among the machine learning algorithms, Linear Regressor provided the best result. This result was sur-passed while implementing the Deep Neural Networks. The error is calculated in each of the algorithm using the formula:

ErrorPercentage : ((PredictedValue ActualValue)=ActualValue) 100

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

The average error among all the implemented machine-learning models is -1.07083 while the error in the deep neural networks is 0.114519. We can conclude from the obtained results that deep neural networks algorithm is most suitable for stock price prediction and can be widely used in the near future.

REFERENCES

1 A Hybrid Machine Learning System for Stock Market Forecasting ; Rohit Choudhry, Kumkum Garg.
2 Forecasting Stock Market Short-term trends using neuro-fuzzy based method-ology ; George S. Atsalakis, Kimon P. Valavanis.
3 Predicting direction of stock price index movement using artificial neural net-works and support vector machines ; Yaup Kara et al.
4 Surveying stock market forecasting techniques ; George S. Atsalakis et al.
5 Using artificial neural network models in stock market index prediction ; Erkam Guresen et al.
6 Credit rating analysis with support vector machines and neural networks ; Zan Huang et al.
7 Forecasting stock market movement direction with support vector machine ; Wei Huang et al.
8 Forecasting returns: New European evidence ; Steven J. Jordon et al.
9 Stock Market Prediction Using Articial Neural Networks; Birgul Egeli et al.
10 Deep Learning for Event-Driven Stock Prediction; Xiao Ding et al.
11 A fusion model of HMM, ANN and GA for stock market forecasting; MD. Raul Hassaneta.