STOCK MARKET PREDICTION USING MACHINE LEARNING

Alisha Bhoir, Prof. Deepesh Jagdale

Abstract- In Stock Market Prediction, the aim is to predict the longer term value of the financial stocks of a corporation. The current running trend available stock market prediction technology is that the operation of ML (machine learning) which makes predictions supported the values of current stock market indices by training on their previous values. In the finance world stock trading is one among the leading important activities. Stock market prediction is an act of trying to work out the longer term value of a stock other financial appliance traded on a financial exchange. Stock market prediction is that the high level of correctness and accuracy is the key factor in predicting a stock market.

In, this paper we are going to present and review a more feasible method to predict the stock association with higher accuracy. The first thing we have taken into account is the dataset of the stock market prices from preceding year. For real analysis the dataset was pre-processed and tuned up. Hence, our paper will also focus on data pre-processing of the raw dataset. Secondly, after pre-processing the data, we will review the use of random forest, support vector machine on the dataset and the outcomes it generates. In this paper we survey of well-known efficient regression approach to predict the stock exchange price from stock exchange data based. In future the results of multiple correlation approach might be improved using more number of variables.

Index Terms— Stock Market, Prediction, Machine Learning, Support Vector Machine

INTRODUCTION

Stock market plays a really important role in fast economic process of the developing country like India. So us country and other developing nation's growth may depend on performance of stock market. A stock exchange may be a platform for trading of a company's stocks and derivatives at an agreed price. Supply and demand of shares drive the stock exchange. In any country stock exchange is one among the foremost emerging sectors. Nowadays, many of us are indirectly or directly associated with this sector.

The stock exchange is essentially an aggregation of varied buyers and sellers of stock. A stock (also known as shares more commonly) in general reprsents ownership claims on business by a particular indiviual or a group of people. The attempt to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient.

The stock market prediction is not a simple process, mainly as a consequence of the close to random- walk behaviour of a stock time series. Two methods are used for forecast stock prices like Fundamental and technical analyses. Artificial Neural networks (ANNs) is the most commonly used technique.

ANNs knowledge from over-fitting problem due to the great number of parameters to fix, and the little previous users knowledge about the relevance of the inputs in the analyzed difficulty in most cases. Also, Support vector machines (SVMs) had been developed as an alternate that avoids such limitations. Their practical successes are often attributed to solid theoretical foundations supported VC-theory. SVM compute globally optimal solutions, unlike those obtained with ANNs, which tend to fall under local minima.

In the recent years, increasing prominence of machine learning in various industries have enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results. This paper will build up a budgetary information indicator program in which there will be a

- *Alisha Bhoir is currently pursuing masters degree program in M.Sc IT in Mumbai University, India,7378584813 . E-mail: alishabhoir98@gmail.com.*
- *Prof. Deepesh Jagdale is currently proffesor in Mumbail University, India, 9028609874. E-mail: djadale@mes.ac.in.*

dataset putting away all authentic stock costs and information will be treated as preparing sets for the program. The main purpose of the prediction is to scale back uncertainty associated to investment deciding.

MACHINE LEARNING ALGORITHMS

1. Unsupervised learning

When the dataset isn't well defined or very hard for interpretation, it's called unsupervised learning. The labels for the data are not defined. There no right thanks to divide data set except performing iterations. Thus, in supervised learning the input is employed to get a structure by watching the relation of the input itself.

For example, Classification of animals.

According to this research, unsupervised learning isn't advisable for prediction.

2. Supervised learning

Supervised learning are often said as function approximation, training examples cause function generation. If the training is completed with right training set, a well behaved function are often expected. Supervised learning grows consistently with the data. It is a kind of induction learning, and it causes biased supervised learning sometimes.

E.g.: The function generated with supervised learning will be $X2$, if X is the input value and the output is self-multiplied.

Since, there is well defined data available from BSE itself and which is in well-defined numeric form it would be beneficial to use supervised learning algorithms. Supervised learning algorithms are of two variants:

1. Regression.
2. Classification

METHODOLOGY:

1. Classification

Classification is an instance of supervised learning where a group is analyzed and categorized supported a standard attribute. From the values or the info are given, classification draws some conclusion from the observed value. If quite one input is given then classification will attempt to predict one or more outcomes for an equivalent. A few classifiers that are used here for the stock exchange prediction includes the random forest classifier, SVM classifier.

i) Random Forest Classifier

Random forest classifier may be a sort of ensemble classifier and also a supervised algorithm. It basically creates a set of decision trees that yields some result. The basic approach of random class classifier is to require the decision aggregate of random subset decision tress and yield a final class or result supported the votes of the random subset of decision trees.

ii) SVM classifier

SVM classifier is a type of discriminative classifier. The SVM uses supervised learning i.e. a labelled training data. The output are hyper planes which categorizes the new dataset. They are supervised learning models that uses associated learning algorithm for classification and as well as regression.

2. Random Forest Algorithm

Random forest algorithm is getting used for the stock exchange prediction. Since it's been termed together of the simplest to use and versatile machine learning algorithm, it gives good accuracy within the prediction. This is usually used in the classification tasks. Because of the high volatility within the stock exchange, the task of predicting is sort of challenging. In stock market prediction we are using random forest classifier which has the same hyper parameters as of a decision tree. The decision tool has a model related to that of a tree. It takes the choice supported possible consequences, which incorporates variables like event outcome, resource cost, and utility. The random forest algorithm represents an algorithm where it randomly selects different observations and features to build several decision tree and then takes the aggregate of the several decision trees outcomes. The data is split into partitions based on the questions on a label or an attribute. The data set we used was from the previous year's 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.

3. Support Vector Machine Algorithm

The main task of the support machine algorithm is to spot an N-dimensional space that distinguishably categorizes the info points. Here, N stands for a number of features. Between two classes of knowledge points, there are often multiple possible hyper planes which will be chosen. The objective of this algorithm is to seek out a plane that has maximum margin. Maximizing margin refers to the space between data points of both classes. The benefit related to maximizing the margin is that it provides is that it provides some rein forcement in order that future data points are often more easily classified. Based on the position of the data points relative to the hyper plane they are attributed to different classes. The dimension of the hyper plane relies on the number of attributes, if the number of attributes is two then the hyper plane is just a line, if the number of attributes is three then the hyper plane is two dimensional.

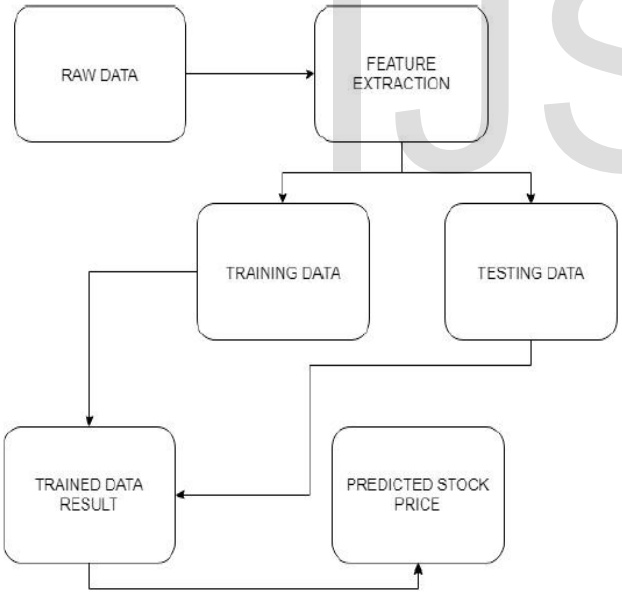

Figure: Stock Market Architecture

RESULT AND DISCUSSION

In this section, we show the results obtained for the models on the assorted stocks. The evaluation metrics are (1) directional accuracy, which analyzes the direction of the expected value with reference to yesterday close price, (2) Precision,which measures the relevancy of the result, (3)Recall,which measures what number true relevant results returned, and (4) F-measure, which measures the weighted average of precision and recall. Based on the directional accuracy metric, SVM outperforms RNN, SVR and DNN for various tested stocks. In table 1. We describe the input file.

Table 1. Stock Data Details

Stock Name	Total Data points	Total Articles	output direction
AAPL	19,243	78,036	1,478 positives 1,271 negatives
FB	11,515	30,198	886 positives 759 negatives
GOOGL	8,225	19,829	625 positives 550 negatives
AMZN	19,243	37,265	1,450 positives 1,299 negatives

Table 2. SVM Directional Accuracy Results

Sentiment-Window	Directional Accuracy			
	AAPL	GOOGL	AMZN	FB
S1-4	78.18%	70.94%	75.27 %	68.9%
S1-5	83.36%	80.34%	74.91 %	73.17%
S1-6	81.73%	79.62%	65.82 %	74.66%
S2-4	79.27%	70.94%	74.18 %	73.17%
S2-5	82.64%	77.78%	74.18 %	74.01%
S2-6	81.09%	79.76%	68.36 %	73.27%
S3-4	79.27%	70.09%	75.64 %	75%
S3-5	82.91%	76.92%	70.18 %	73.78%
S3-6	81.64%	76.62%	68.73 %	60.74%

According to Table 2, it's very clear that our SVM model is ready to attain accuracies way above the five hundred. When staring at Table 3, it also clear that SVM outperforms SVR, DNN, and RNN. All achieved accuracies are above 75% and within the caseof APPL, the achieved accuracy is around 83%. All our models achieved better results than those reported in literature as indicated in Table 4. Based on the reported results, we summarize our contributions as follows:

- We highlighted the effect of stories sentiments on the stock price movement
- We identified best measure for stock price prediction.
- We identified best news scenario which each stock is

affected differently by news.

• Our model analysis indicates that close price or trend with regard to yesterday close price are often predicted using various AI models.

• Our proposed model are often employed in alternative ways. Firstly, our model are often employed by traders without programing information. These traders can use our model either to only predict the variation in price and help traders in their analysis. Also they'll use our automated trading system with none moniting, where the system opens and closes trades supported the predictions. Finally, our code are often easily deployed to try and do short-term trading.

CONCLUSION

The aim of our research study is to help the stock brokers and investors for investing money within the stock market. The prediction plays a extremely important role available market business which is extremely complicated and challenging process due to dynamic nature of the stock market. The results of this study confirm that machine learning techniques are capable of predicting the stock market performance. This paper was an endeavour to figure out the long run prices of the stocks of a corporation with greater accuracy and reliability using machine learning techniques. Right now, built up a stock value pattern expectation framework. To form these models we've gathered data from two sources (i) Historical stock market data from Reuters and (ii) news sentiment released on a particular stock; this data was collected for 4 different stocks over 10 years. Technical features are calculated and used as input data for our model additionally to a few scenarios considered when adding sentiments to the calculated features.

After developing our model, and to point its performance we'd implement a risk strategy to check the profits we'd gain supported our predictions and a few enhancements is completed and studied for our prediction model. One direction is to feature extra technical indicators employed in stock market. Another direction would be trying different time-frames for grouping our data. Finally, we could attempt to enhance the prediction of the precise price.

Table 3. All Models Directional Accuracy

	SVM	SVR	DNN	RNN
APPL	82.91%	79.2%	81.32%	81.3%
AMZN	75.27%	72.26%	74.03%	74.56 %
GOOGL	80.34%	66.38%	80.1%	68.38 %
FB	75%	68.71%	72.68%	72.39 %

Table 4. Related Work Accuracies

Paper	Metric	Value
Arévalo, A. et al. (2016)	Directional Accuracy	66%
Schumaker, R. P. et al. (2009)	Directional Accuracy	71.18%
Ding, X. et al. (2014)	Accuracy	60%

REFERENCES

1. Schumacher, R.P., Chen, H.: Textual analysis of stock market prediction using breaking financial news: the AZFin text system. ACM Trans. Inf. Syst. 27, 1–19 (2009)

2. Giannini, R.C., Irvine, P.J., Shu, T.: Do local investors know more? A direct examination of individual investors' information set. Working paper (2014)

3. Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G.: Identifying and following expert investors in stock microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1310–1319 (2011)

4. Plakandaras, V., Papadimitriou, T., Gogas, P., Diamantaras, K.: Market sentiment and exchange rate directional forecasting. Algorithmic Financ. 4(1–2), 69–79 (2015)

5. Singh, Aishwarya. "Predicting the Stock Market Using Machine Learning and Deep Learning." Analytics Vidhya, 26 July 2019,

6. Author: W. HuangResearch paper: Forecasting stock market movement direction with support vector machine.Journal: Computers & Operations Research

7. Author: J. MoodyResearch paper: Learning to trade via direct reinforcement. Journal: IEEE Transactions on Neural Networks

8.https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/

9.https://azure.microsoft.com/enin/documentation/articles/machine-learning-algorithm-choice/