

EVALUATION OF ENSEMBLE LEARNING APPROACH FOR OPTIMISED STOCK PREDICTION

¹ Uma S, ² Susithra S, ³ Kapilavani R K,

^{1,2} Student, ³ Assistant Professor, Department of Computer Science and Engineering,
Prince Shri Venkateshwara Padmavathy Engineering College, Chennai, India.

ABSTRACT

Stock prediction is a process of determining the longer term value of the corporate stock. The successful prediction of a stock's future could yield significant profit. This hypothesis provides stock prices that reflect all presently available information and any fluctuation in price that's not supported newly revealed information thus are inherently unpredictable. Stock prediction is taken into account to be one among the foremost complicated areas of applied finance, where machine learning techniques can play an important role in predicting future stock. Therefore, for maximizing financial gain and minimizing loss to urge the optimum output, there's a requirement for accurately predicting the trends available. The aim is to predict machine learning techniques for predicting results providing accuracy. The analysis of the given sample of data by machine learning algorithms for capturing the information like, variable identification, univariate analysis, bivariate and statistical method, missing value treatment and analyzing the info validation, data cleaning, preparing, data visualization are

going to be administered on the whole given dataset. It aims to propose a machine learning-based method to accurately predict the stock price index value by prediction leads to the shape of stock increase or stable state, thus best accuracy from ensemble methods comparing various classifiers in machine learning algorithms is obtained.

Index Terms: dataset, machine learning algorithm-ensemble methods, python, predicting the accuracy of result.

1. INTRODUCTION

Prediction of the varying trends in stock prediction is meant to be a really challenging task due to various uncertainties involved in it. The upcoming success of application of AI within the financial area caused more firms counting on stochastic models for prediction of market behavior. Everyday quantitative analysis strives to achieve better accuracies from their machine learning models for forecasting returns from stock. With the increasing success of Machine Learning algorithm in almost every domain, the financial industry was faster to support AI driven

mathematical models, to extend profits of the varied organizations.

A stock basically represents ownership claims on business by a specific individual or a gaggle of individuals. The attempt [3] to work out the longer term value of the stock exchange is understood as a stock exchange prediction. The prediction is predicted to be robust, accurate and efficient. The system must work consistent with the real-life scenarios and will be suited to real-world settings. The system is additionally expected to require under consideration all the variables which may affect the stock's value and performance. There are varied methods and ways of accomplishing the prediction system like Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and statistic aspect structuring. With the furtherance of the digital era, the prediction has moved up into the technological realm. the foremost prominent and [3] promising technique involves the utilization of Artificial Neural Networks, Recurrent Neural Networks, that's basically the implementation of machine learning. Machine learning involves AI which entitles the system to find out and brush up from past experiences without being programmed time and again. Conventional methods of prediction in machine learning use algorithms like Backward Propagation, also referred to as back-propagation errors. Recently, many researchers are using more of

ensemble learning approaches. it credible to utilize low price and time [3] lags to predict future highs whereas another network would use lagged highs to predict future highs. These predictions were used to form stock prices. [1]

Stock prediction for brief time windows appears to be a random process. The stock price movement over an extended period of your time usually develops a linear curve. People tend to shop for those stocks whose prices are expected to rise within the near future. The uncertainty within the stock exchange refrain people from investing in stocks. Thus, there's a requirement to accurately predict the stock exchange which may be utilized in a real-life scenario. The methods wont to predict the stock exchange includes a statistic forecasting along side technical analysis, machine learning modeling and predicting the variable stock exchange. The datasets of the stock exchange prediction model include details just like the price opening price, the info and various other variables that are needed to predict the thing variable which is that the price during a given day. The previous model used traditional methods of prediction like statistical method with a prediction statistic model. stock exchange prediction outperforms when it's treated as a regression problem but performs well when treated as a classification. The aim is to style a model that gains from the market information utilizing machine learning

strategies and gauge the longer term patterns available value development. The Support Vector Machine (SVM) are often used for both classification and regression. it's been observed that SVMs are more utilized in classification based problem like ours. The SVM technique, we plot every single data component as some extent in n-dimensional space (where n is that the number of features of the dataset available) with the value of feature being the value of a specific coordinate and, hence classification is performed by finding the hyperplane that differentiates the 2 classes explicitly.

Predictive methods like Random forest technique are used for an equivalent . The random forest algorithm abide to an ensemble learning strategy for classification and regression. The random forest takes the typical of the varied subsamples of the dataset, this increases the predictive accuracy and reduces the over-fitting of the dataset.

2. MOTIVATION BEHIND THIS STUDY

- ❖ Monitoring and preserving stock price has become one of the most essential activities in many industrial. Thus we aim to predict the daily adjusted closing prices of Dow Jones Industrial Average from Reddit World News Channel. The model will be trained using the train set, model hyper parameters will be tuned using the validation set, and finally the

performance of the model will be reported using the test set.

- ❖ Hence, the main focus of this work is to investigate a dataset of stock price prediction using ensemble machine learning technique. We try to reduce the risk factor behind predicting a company stock so as to save a lot of industrial economy efforts and assests, thus predict stock price decreases or stayed the same
- ❖ This paper features existing Machine Learning Algorithms to predict the stock increase or decrease for an organization.

3. MODEL OF PROPOSED SYSTEM

In our project, an ensemble model is build using various classifiers, which train the machine with the given dataset and predicts the future stock of an organization.

The following are the steps for creating the model: Once the ensemble model is created, the machine gets trained based on the given dataset by using eight different ML algorithms(For example, Random forest, Decision tree, SVM, Logistic regression, KNN, Naive Bayes, Adaboosting, Voting Classifier)and then aggregate the prediction of each base model. When new data is provided, based on the past trained data the model provides accurate prediction. By using this

model, the generalization error of the prediction is reduced and further can be used by an organization in predicting future stock fluctuations and stabilize its economy.

We have taken the data from the past years stock to train the model. We majorly used two machine-learning libraries to unravel the problem. The first one was numpy, which cleans and manipulates the data, and getting it into a form ready for analysis. The second library was scikit, which was used for real analysis and prediction. We have used dataset from the previous years stock collected from the public database available online, out of which 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 find out the patterns and relationships within the data from the training set then reproduce them for the test data. We used the python pandas library for processing which combined different datasets into a knowledge frame. The tuned up dataframe allowed us to organize the info for feature extraction. The dataframe features were date and therefore the price for a specific day. We used of these features to coach the machine on random forest model and predicted the thing variable, which is that the price for a given day. We also quantified the accuracy by using the predictions for the test set and therefore the actual values. Thus, stock price of a company

reflects the perception of the investors ability to earn and grow its profits in the future. If the company is doing good, the shareholders are happy as it seems by its share price, the management would likely remain and receive increases in compensation and economy. Another reason that a corporation might be concerned with its stock price is for the prevention of a takeover. If the stock price of a company is performing good along with the company, company is probable to receive more favourable press from analysts and the media

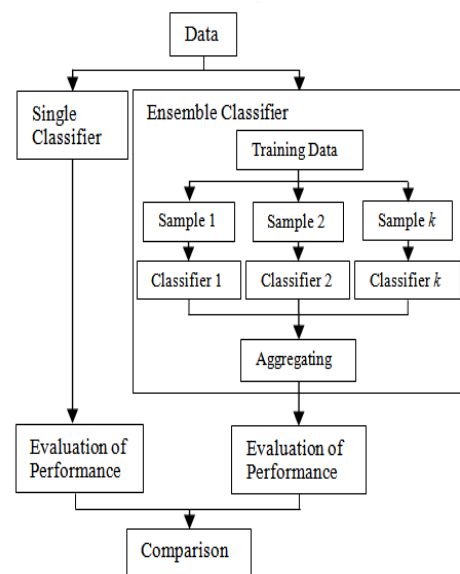


Fig 1: Block diagram describing how various modules interface with each other

4. METHODOLOGY

Here we are applying machine learning algorithms to a dataset that is considered. A machine learning project may not be linear, but it has a number of well-known steps:

- Define Problem.
- Prepare Data.
- Evaluate Algorithms.
- Improve Results.
- Present Results.

The best way to really come to terms with a new platform or tool is to work through a machine learning project end-to-end and cover the key steps. Namely, from loading data, summarizing data, evaluating algorithms and making some predictions. Classification is a specification of supervised learning where a group is analyzed and categorized based on a standard attribute. From the given data set or values, classification recognizes 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 . Some classifiers that are used here for the stock prediction of an organization includes the logistic regression classifier, decision tree classifier, random forest classifier, SVM classifier, Naive Bayes Classifier, KNN classifier, Adaboosting, Voting Classifier.

LOGISTIC REGRESSION

It is a statistical procedure for analysing a data set in which there are one or more independent variables that determine an outcome. The result is calculated with a binary variable (i.e. only two possible outcomes). The

objective of logistic regression is to reach out the best fitting model to elucidate the connection between the dichotomous characteristic of interest (dependent variable = response/result variable) and a set of independent (predictor/explanatory) variables.

Logistic regression is a ML classification algorithm that predict the probability of a implied dependent variable. In logistic regression classifier, the binary variable contains data coded either as 1 (yes, success, etc.) or 0 (no, failure, etc.).

Logistic regression Assumptions: In Binary logistic regression needs the dependent variable to be binary. For a binary regression, the factor level 1 of the variable should denote the specified result. Only the significant variables should be included. The independent variables used must be independent of each other. The independent variables are sequentially related with the log odds. Logistic regression needs quite large sample sizes.

DECISION TREE

It is one of the leading and popular algorithm. Decision-tree algorithm comes under the kind of supervised learning algorithms. It performs for both continuous as well as categorical output variables. Assumptions of Decision tree: At the initial stage, we consider the whole training set as the root. Attributes are

deduced to be categorical for information gain, attributes are assumed to be continuous. Based on attribute values records are distributed iteratively. We use analytical methods for ordering attributes as root or internal node.

Decision tree constructs tree structure like models for performing classification or regression. It splits the data set into smaller subsets while at the same instance an associated decision tree is incrementally constructed. A decision node consists of two or more branches and a leaf node denotes a classification or decision. The highest decision node in a tree which harmonize the best predictor called root node. Decision trees can operate on both categorical and numerical data. It makes use of an if-then rule set which is mutually exclusive and comprehensive for classification. The rules are learned linearly using the training data one at a time. Once the rule is learned at a given instance, the tuples covered by the rules are removed. This process is resumed on the training set until meeting a termination condition. It is build in a top-down recursive divide-and-conquer manner. The attributes used must be categorical. Otherwise, they should be detached in advance. Attributes at topmost position of the tree have more influence towards the classification and they are recognized using the information gain concept. A decision tree can be easily over-fitted

producing too many branches and may consider anomalies due to noise or outliers.

SUPPORT VECTOR MACHINE

SVM classifier is a sort of discriminative classifier. The SVM makes use of supervised learning i.e. a labeled training data. The output are hyperplanes that is used to categorize the new dataset. This model uses related learning algorithm for classification and as well as regression.

The SVM classifier consists of certain tuning parameters like kernel parameter, gamma parameter and regularization parameter. Kernels can be designated as linear and polynomial kernels that calculates the prediction line. In linear kernels prediction for a new input is estimated by the dot product between the input and the support vector. C parameter is also called as the regularization parameter; it estimates whether the accuracy of model is increases or decreases. The C parameter has a default value of 10. Lower regularization value results in misclassification. Gamma parameter computes the regulation of a single training on the model. Low values denotes far from the plausible margin and high values denotes closeness from the plausible margin.

The main purpose of the support machine algorithm is to identify an N-dimensional space that discernibly categorizes the data points, here N stands for a number of features. There can be multiple possible hyperplanes that can be chosen between two classes of data points. The intension of this algorithm is to seek out a plane that has maximum margin. Maximizing margin stipulates to the distance between data points of both classes. The advantage of maximizing the margin is that it provides some reinforcement in order that in long term the data points are often more easily classified. The data points are classified using the decision boundaries, these boundaries are called hyperplanes. Based on the data point position, relative to the hyperplane they're attributed to different classes. The dimension of the hyperplane relies on the number of attributes, the hyperplane is just a line provided the number of attributes is two and the hyperplane is two dimensional provided the number of attributes is three.

RANDOM FOREST

Random forest classifier is an instance of ensemble classifier and also a supervised algorithm. It basically generates a set of decision trees, that provides some result. The central approach of random class classifier is to take the decision lump of random subset decision tress and yield a final class or result

based on the votes of the random subset of decision trees.

The random forest classifier uses parameter like `n_estimators` which is total number of decision trees, and other hyper parameters like `oobscore` to establish the generalization accuracy of the random forest, `max_features` which includes sum of features for best-split. `min_weight_fraction_leaf` is the least weighted fraction of the sum of weights of all the input samples required to be at a leaf node. Sample weight is not provided, they have equal weight.

Random forest algorithm is here used for the stock prediction. Since it's been termed as simplest to use and a versatile machine learning algorithm, it provides good accuracy within the prediction. This is basically used in the classification tasks. Because the stock returns involve high volatility , prediction becomes a sort of challenging. In stock prediction we are using random forest classifier which has the same hyper parameters similar to a decision tree. The decision tool has a structure similar to that of a tree. It makes the decision based on possible consequences, which involves variables like event outcome, resource cost, and utility. The random forest algorithm constitute an algorithm where it randomly choose different observations and features to

generate several decision tree then takes the mixture of the several decision trees outcomes

K-NEAREST NEIGHBOR

K-Nearest Neighbor is categorized under supervised machine learning algorithm which reserves all instances corresponding to training data points in n-dimensional space. On receiving an unknown discrete data, it analyzes the closest k number of instances saved (nearest neighbors) and returns the foremost common class as the prediction and it yields the mean of k nearest neighbors for real-valued data. Confined of distance-weighted nearest neighbor algorithm, it weights the contribution of every k neighbors consistent with their distance using the subsequent query giving greater weight to the closest neighbors.

NAIVE BAYES

Naive Bayes classifier is a supervised learning algorithm, It is utilized for the text classification of high-dimensional training dataset. Naïve Bayes Classifier is considered to be one of the simple and most effective Classification algorithms which helps in constructing the fast machine learning models that can make quick predictions. It is a probabilistic classifier, that is it predicts based on the probability of an object. Some known examples of Naïve Bayes Algorithm are spam

filtration, Sentimental analysis, and classifying articles. The Naïve Bayes algorithm is composed of two words Naïve and Bayes, Which can be illustrated as:

- Naive means that the occurrence of a certain feature is independent of the occurrence of other features. Such as if an object is associated on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature independently contributes to associate that it is an apple without depending on each other..
- Bayes' theorem is also called as Bayes' Rule/Bayes' law, which is used to denote the probability of a hypothesis with prior knowledge. It is based on the conditional probability. Bayes' theorem formula is given as:

$$P(A/B) = \{P(B/A) P(A)\}/P(B)$$

Where, P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B. P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

ADABOOSTING

Adaboost is shortform for Adaptive Boosting. Basically, Adaboosting was the basic

and really successful boosting algorithm established for dichotomous classification. Also, it's the simplest start line for understanding boosting. Besides, modern boosting methods build on AdaBoost, is most specifically stochastic gradient boosting methods. Basically, Adaboost is used with short decision trees, followed by tree creation. The performance of the tree on each training instance is utilized. Thus, it is tree should pay attention to each training instance. Hence, training data that is difficult to predict is given more weight, whereas easily predicted instances are given less weight.

Learn Adaboost Model from Data: Adaboosting is specifically used to boost the performance of decision trees and for resolving binary classification problems. Adaboost was formerly called AdaBoost.M1 by the author. More recently it's going to be mentioned as discrete Adaboost. As because it's used for classification instead of regression. Adaboost can be best used with weak learners. In the training dataset, each occurrence is weighted. The initial weight is given by: $weight(x_i) = 1/n$, Where x_i is the i 'th training instance and n is the number of training instances.

VOTING CLASSIFIER

A voting classifier is a classification technique that uses multiple classifiers to carry out the predictions. It is much applicable in situations

when a data scientist or machine learning engineer is not sure about which classification method to use. Therefore, the voting classifier makes predictions based on the most frequent prediction from multiple classifiers

5. SYSTEM ARCHITECTURE

Kaggle is a web community for data analysis and predictive modelling. It also consists dataset of various fields, which is bestowed by data miners. Various data scientist competes to make the simplest models for predicting and depicting the knowledge . It authorize the users to use their datasets so that they can set up models and work with varied data science engineers to resolve various real-life data science challenges. Our proposed project makes use of the dataset downloaded from Kaggle. However, this data set is initially present in raw format. The data set may be a collection of stock exchange information an organization. The initial step is the data pre-processing, that is conversion of the raw data into processed data. This process is carried out through feature extraction, since the raw data might consist of multiple attributes but only a couple of those attributes are useful for the impetus of prediction. So the first step is feature extraction, where in the key attributes are extracted from the entire list of attributes present in the raw dataset. Feature extraction begins from an initial stage of measured data and assembles derived values or features. These

features are deliberately considered to be informative and non-redundant, hastening the next learning and generalization steps. Feature extraction is a process under dimensionality reduction, where the initial set of raw variables is shrunk to progressively reasonable features for ease of management, while still precisely and totally illustrating the first informational collection. The feature extraction process is accompanied by a classification process wherein the data that was obtained after feature extraction is split into two different and perceptible segments. Classification is the process of identifying to which set of categories a fresh observation belongs. The model is trained using the trained data set whereas the test data is used for prediction on the accuracy of the model. The splitting process is completed as a way to ensure that the training data maintain a better proportion than the test data.

6. MODULE IDENTIFICATION

6.1 DATA COLLECTION

Data collection is a very fundamental module and the initial step towards the project. It generally compact with the gathering of the correct dataset. The dataset that's to be utilized in the market prediction has got to be wont to be filtered based on various aspects. Data collection also complements to reinforce the dataset by adding more data that are external. Our model uses the dataset that mainly consists of the previous year stock returns. Initially, we'll be analyzing the Kaggle dataset and consistent with the accuracy, we'll be using the model with the data to investigate the predictions accurately.

6.2 DATA PRE-PROCESSING

Data pre-processing is a a part of data mining, which involves transforming data into a more reasonable format. Since the raw data is typically , inconsistent or incomplete and usually contains many errors. the data pre-processing involves finding out for the missing values, trying to find categorical values, splitting the data-set into training set and test set and eventually do a feature scaling to limit the range of variables so that they can be compared on common environs.

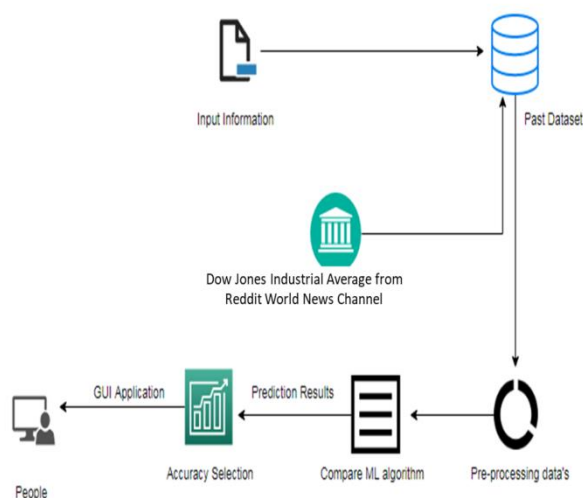


Fig 2: System Architecture

6.3 DATA VISUALIZATION

Data visualization is a crucial skill in applied statistics and machine learning. Statistics does indeed specialise in quantitative descriptions and estimations of data. Data visualization contributes a crucial suite of tools for securing a qualitative understanding. this will be helpful when exploring and going to know a dataset and may help with identifying patterns, corrupt data, outliers, and far more. With appropriate domain knowledge, data visualizations are often used to express and illustrate the key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it'll recommend a deeper dive into some the books mentioned at the end. Sometimes data doesn't add up until it can check out in a visual form, like with charts and plots.

6.4 TRAINING THE MACHINE

Training the machine is similar to providing the data to the algorithm to amend the test data. The model is thus trained using the trained data set. The test set are unconsumed, as a model shouldn't be judged based on unseen data. The training of the model includes data cross-validation so that by making use of the training data we get a well-grounded approximate performance of the

model. Tuning models are meant to particularly tune the hyper parameters like the number of trees in a random forest. The motive behind the training the model is that we use some initial values with the dataset and then optimize the parameters which we want to be in the model. Until we get the optimal values, this process occurs in repetition Thus, from the trained model, we take the predictions on the inputs from the test dataset. Hence, it is further divided in the ratio of 80:20 where 80% is referred to the training set and the rest 20% for a testing set of the data.

6.5 DATA SCORING

The method of applying a predictive model to a set of data is known as scoring the data. The approach is used to process the dataset using eight ML algorithms, i.e. logistic regression, decision tree, random forest, SVM, Naive Bayes, KNN., adaboosting, voting classifier. Random forest necessitates an ensemble method, which is usually used, for classification and as well as regression. Based on the learning models, we are provides with interesting results. The final module thus express how the result of the model can help to make prediction on the probability of a stock to rise and sink based on certain parameters. It also shows the susceptibility of a particular stock or entity.

7. RESULT ANALYSIS

7.1 Comparing Algorithm with prediction within the type of best accuracy result

It's important to match the performance of multiple different machine learning algorithms consistently and it will discover to form a test harness to match multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to match . Each model will have different performance characteristics. Using re-sampling methods like cross validation, you will get an estimate for how accurate each model could even be on unseen data. It must be able to use these estimates to choose one or two best models from the suite of models that you simply simply have created. When have a replacement dataset, it's an honest idea to ascertain the data using different techniques so on seem at the data from different perspectives. the same idea applies to model selection. you need to use sort of various ways of watching the estimated accuracy of your machine learning algorithms so on choose the one or two to finalize. the way to attempt to the present is usually to use different visualization methods to means the standard accuracy, variance and other properties of the distribution of model accuracies. The key to an honest comparison of machine learning algorithms is ensuring that each algorithm is evaluated within

the same way on an equivalent data and it are ready to do that by forcing each algorithm to be evaluated on a consistent test harness.

Now we'll create the confusion matrix here to ascertain the accuracy of the classification. to form it, we'd wish to import the confusion_matrix function of the sklearn library. After importing the function, we'll call it employing a replacement variable cm. The function consists of mainly two parameters, mainly y_true(the actual values) and y_pred (the targeted value return by the classifier).

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Fig 3: Confusion Matrix

7.2 PREDICTION RESULT

The application reads the input and applies the input and applies the prediction algorithm to it to generate the output. The output consists of the confusion matrix and accuracy results. The output is generated for all the stock. The close analysis of the output of the eight algorithms i.e logistic regression, decision tree, SVM, random forest, KNN, naïve bayes, adaboosting, voting classifier algorithms reveal that the logistic regression gives better results than other algorithms. The following table gives the

comparison of output of all the algorithms used in the prediction system:

S.NO	ALGORITHM	ACCURACY%
1	Logistic Regression	97.48
2	Decision Tree	63.92
3	SVM	53.52
4	Random Forest	62.75
5	KNN	70.13
6	Naïve Bayes	49.66
7	Adaboosting	61.91
8	Voting Classifier	61.91

Table.1 Prediction Results

Logistic regression algorithm has the highest accuracy rate of 97% and uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. It need the output of the algorithm to be classified variable data. Higher accuracy predicting result is logistic regression model by comparing the best accuracy.

8. CONCLUSION

The project concentrates mainly on the static data obtained from the Dow Jones industrial average of the Reddit World News Channel for a given period of time . this will further be optimized for real time stock exchange prediction by implementing other sources of stock data,

news journal policy etc. The goal is to get maximum accuracy by comparing various ensemble learning algorithms to make prediction on the simplest accuracy within stock fluctuation for an organisation. By measuring the accuracy of the algorithms, we found that the foremost suitable algorithm for predicting the longer term price of a stock of an organisation supported various data points from the historical data. Thus, the project demonstrates the ensemble machine learning model to predict the stock value with more accuracy as here various ML classifiers are compared or higher precision.

We collected the stock returns of a news channel to coach and test the system. Five years of data is downloaded from the Kaggle website, of which 75% is employed to coach the system and therefore the remaining 25% is employed for testing. During this paper, we study the utilization of eight ML algorithms to predict stock movement direction. of those algorithms, we saw the Logistic Regression gave us the simplest results(97% Accuracy). Logistic Regression may be a promising sort of tool for financial forecasting. Logistic Regression is superior to the opposite individual classification methods in forecasting daily movement which may cause a capital gain. However,

each ML methods has its own strengths and weaknesses. during this model, the principal components identified by the Logistic Regression are used along with internal and external factors in it for forecasting. We also observed that the selection of the indicator function can dramatically improve/reduce the accuracy of the prediction system. Also a specific ML Algorithm could be better suited to a specific sort of stock, say Technology Stocks, whereas an equivalent algorithm might give lower accuracies while predicting another sorts of stocks, say Energy Stocks.

9. FUTURE ENHANCEMENT

The future scope of this project will involve adding more parameters and factors just like the financial ratios, multiple instances, etc. The more the parameters are taken under consideration the more are going to be the accuracy. To automate this process by show the prediction end in web application or desktop application. Eight algorithms, Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbour, Naïve Bayes, Adaboosting and Voting Classifier utilized in this study and only one dataset from Dow Jones industrial Average from Reddit World News Channel was applied to coach and test the models. The system can only predict the direction (up/down) for the

future trading days of a specific organization. In the future, these algorithmic models are going to be utilized in order to predict the worth moment for the longer term and therefore the results are going to be compared with other data mining techniques by applying different dataset from different stock index.

10. REFERENCES

JOURNAL REFERENCES:

- [1]Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [2]Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.
- [3]Xi Zhang¹, Siyu Qu¹, Jieyun Huang¹, Binxing Fang¹, Philip Yu², "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2018.
- [4]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [5]F. Lin, S. Chen, and C. Hsu, "Intelligent backstepping control using re current feature selection fuzzy neural network for synchronous reluctance motor position servo drive system," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 3, pp. 413–427, Mar. 2019

- [6]Mahdi Pakdaman Naeini, Homa Baradaran Hashemiand Hamidreza Taremiian, “.Stock Market Value Prediction Using Neural Networks,” International Conference on Computer Information Systems and Industrial Management Applications (CISIM), Pp132– 136 2010. *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1881–1890, Sep. 2019.
- [7]NeelimabBudhani,C. K. Jha, and K Sandeep.Budhani,. “Application of Neural Network In Analysis of Stock Market Prediction,”International Journal of Computer Science And Engineering Technology (IJCSET), Vol 3No 4,, Pp61-68. Vol. 3 No. 4 , April 2012.
- [8]Pei-Chann Chang and Chen-Hao Liu,. “A Neural Network With A Case Based Dynamic Window for Stock Trading Prediction;Expert Systems With Applications “ Vol 36, No3,Pp 6889-6898, 2008
- [9]Sheng-Hsun Hsu and JJ Po-An Hsieh, “A Two-Stage Architecture For Stock Price For Recasting By Integrating Self-Organizing Map and Support Vector Regression, Expert Systems with Applications,” Vol36, No 4, Pp 7947-7951. 2008
- [10]H. Zhou, H. Ying, and C. Zhang, “Effects of increasing the footprints of uncertainty on analytical structure of the classes of interval type-2 Mamdani and TS fuzzy controllers,”