International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Prediction of Option Price using Ensemble of Machine Learning Algorithms for Indian Stock Market

Payal Shrivastava, Chandan Kumar Verma

Abstract: The non-deterministic behavior of stock market creates ambiguities for buyers. The situation of ambiguities always finds the loss of user financial assets. The variations of price make a very difficult task to predict the option price. For the prediction of option used various non-parametric models such as artificial neural network, machine learning, and deep neural network. The accuracy of prediction is always a challenging task of for individual model and hybrid model. The variation gap of hypothesis value and predicted value reflects the nature of stock market. In this paper use the bagging method of machine learning for the prediction of option price. The bagging process merge different machine learning algorithm and reduce the variation gap of stock price.

***Index Terms*: Stock Market, NSE, Ensemble, SVM, Machine Learning, Glowworm.**

I. INTRODUCTION

The future of country depends on the health of stock market. The health of stock market depends on the behaviors of minimum price variation. The variation of price born the volatile nature of market and induced the loss of customer assets. For the study of the nature of stock market used various parametric and non-parametric models [1-2]. The Prediction accuracy of non-parametric models is varying according to the variation of data. The dynamic nature of stock market influences the factor of non-parametric models for the process of estimation. The dynamicity of stock market data behaves like non-linear then used non-linear non-parametric models for the balancing of uncertainty and Prediction of stock market data [3-5]. The advancement of artificial neural network and feature optimization proposed the new Prediction algorithms which is more accurate instead of conventional artificial neural networks [3]. Machine learning offers various non-linear regression models for the Prediction of stock market data. The nature of non-linearity of machine learning algorithms improved the Prediction accuracy of option price. The machine learning algorithms such as deep neural network, support vector machine and decision tree. Ensemble of machine learning algorithms also enhanced the performance of stock market Prediction. The techniques of ensemble basically depend on the process of bagging, bossing and random forest [5-8]. In this paper used the process of bagging and bossing. The process of bagging used two different classifiers and ensemble the process of classification [9]. The accuracy of Prediction of ensemble classifier is high instead of normal machine learning classification algorithms. Perhaps the most common algorithm for ensemble learning is known as Boosting. Boosting incrementally builds the ensemble model by training each new model instance to emphasize samples were the previous models miss-classified. Bagging or Bootstrap aggregating, often abbreviated bagging, uses multiple learners that have equal weight in the ensemble committee. Variance is achieved by training each model on a randomly drawn subset of the data. Decision Trees are not Ensemble Learners by their self but serve as the basis for several ensemble learning algorithms. The Decision Tree is a structure like flowchart, where each node contains a test on a feature, each branch represents the outcome of the test, and every end node contains a class label [10-13]. In this paper proposed the ensemble-based classifier using machine learning algorithms. For the ensemble used three machine learning algorithms are used support vector machine, decision tree and KNN. For the betterment of ensemble used feature optimization algorithm is called glowworm optimization algorithm. The glowworm algorithm reduces the variation of noise in stock market data.

The contribution of this paper is summarized as follows.

1. Enhance the Prediction accuracy of option price.
2. Ensemble based classifier reduces the Prediction variation of predicted vale and hypothesis.
3. Design optimized ensemble classifier for the stock market Prediction.
4. Conduct experimental task used NSE data and measure four parameters for the validation of results.

II. MACHINE LEARNING AND ENSEMBLE CLASSIFIER

Machine learning is new area of data science for the better categorization of data.

Revised Manuscript Received on 30 July 2019.
* Correspondence Author
Payal Shrivastava*, Department of Applied Mathematics, MANIT, Bhopal, MP, India.
Chandan Kumar Verma , Department of Applied Mathematics, MANIT, Bhopal, MP, India.

The categorization of data shows the process of Prediction and accuracy of classification process. In the machine learning two types of data are involved one is test data and other is trained data. The techniques of data sampling and training decide the future of machine learning algorithms [14]. The machine learning offers various linear and non-linear classification algorithms.

The support vector machine, nearest neighbor, decision tree and statically based classification algorithms is major methods of machine learning. Improvement of classification and Prediction ratio is always a challenge task. The Prediction and classification ratio depend on the nature of data. The nature of stock market data is uncertain and behaves like a non-linear. The process of ensemble classifier enhances the Prediction ratio of individual classifier of machine learning algorithms. The methods of ensemble merge two or more classification algorithms according to their requirements. The methods of ensemble used three basic techniques bagging, boasting and random forest. These techniques improve the process of Prediction. Processing block diagram of ensemble of machine learning algorithm Support vector machine shown in Figure 1. In this paper used three machine learning algorithms for ensemble classifier [15-17].

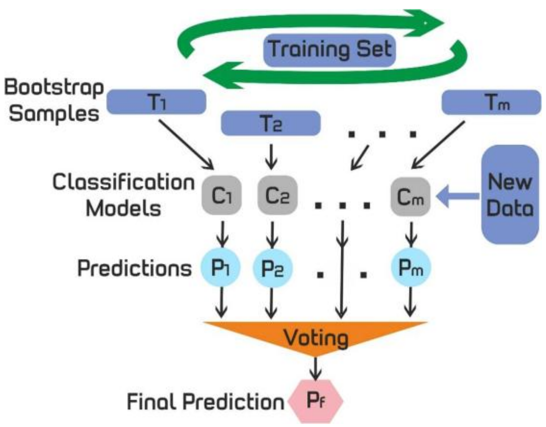

Figure 1 process block diagram of ensemble of machine learning algorithm

A. *Support Vector Machine*

SVM learning algorithm is binary classification algorithm. This SVM classifier has a slack variable and penalty function for solving non-separable problems. First, given a set of points $x_i \in R^d$, i=1,……,l and each point x_i belongs to either of two classes with the label $y_i \in \{+1,-1\}$.These two classes can be applied to anomaly attack detection with the positive class representing normal and negative class representing abnormal. Suppose there exists a hyper-plane $W^T x_i$ + b = 0 that separates the positive examples from the negative examples. That is, all the training examples satisfy [18]:

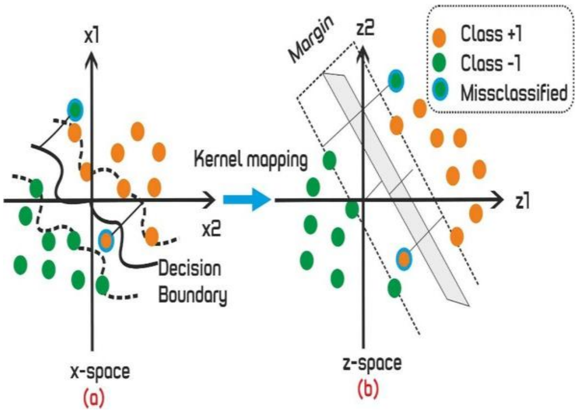

Figure 2 process block diagram of non-linear support vector machine

$$W^T x_i + b \geq +1 \; for \; all \; x_i \in P$$

$$W^T x_i + b \geq -1 \; for \; all \; x_i \in P \qquad (1)$$

W^T s an adjustable weight vector, x_i is the input vector and is the bias term.

Equivalently:

$$y_i.(W^T.x_i - b) \geq 1 \forall i, = 1 \ldots N \qquad (2)$$

In this case, we say the set is linearly separable.

B. *Decision Tree*

The DT (Decision Tree) classifier consists of decision and leaf nodes. Each decision node corresponds to a test over a single attribute of the given instances. It has different branches on other decision or leaf nodes that represent the possible values of the actual feature. Leaf nodes represent the possible attack and normal class labels that can serves as an output when classifying a new example. Process block diagram of non-linear support vector machine decision tree shown in Figure 2. Generally, the DT classifier is generated relying on two phase's process. The dimensionality reduction is the first phase in DT building process [19-21].

C. *KNN*

Nearest neighbor classifiers are based on learning by analogy. The training samples are described by n dimensional numeric attributes. Each sample represents a point in an n-dimensional space. In this way, all of the training samples are stored in an n-dimensional pattern space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is defined in terms of Euclidean distance, where the Euclidean distance, where the Euclidean distance between two points, X=(x_1, x_2,……, x_n) and Y=(y_1, y_2,…., y_n). The unknown sample is assigned the most common class among its k nearest neighbors. When k=1, the unknown sample is assigned the class of the training sample.

D. *Bagging*

Techniques of bagging in selection of different data samples shown in Figure 3. Bagging (bootstrap aggregating) is a method for generating multiple versions of a predictor and using these to get an aggregated predictor.

The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when expecting a class. Making bootstrap replicates of the learning sets forms the multiple versions. Bagging can give substantial gains in accuracy when used on classification methods. Moreover, bagging can make unstable models stable and thus improve accuracy [22].

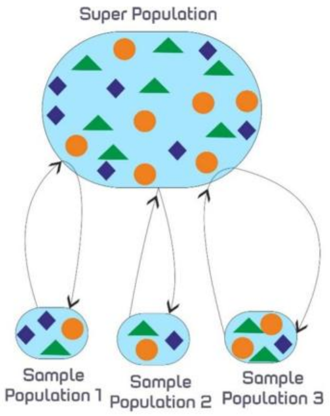

Figure 3 techniques of bagging in selection of different data samples

E. *Boosting*

Methods of boosting and resample of data and hypothesis shown in Figure 4. Boosting is a general method for improving (or boosting) the accuracy of any given learning algorithm. More precisely, boosting refers to a general and effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules. By sequentially fitting models, later models see more of the samples misclassified by earlier ones. Combination uses weighted average where later models get more weight. Boosting works well on weak models and it reduces both bias and variance [23-24].

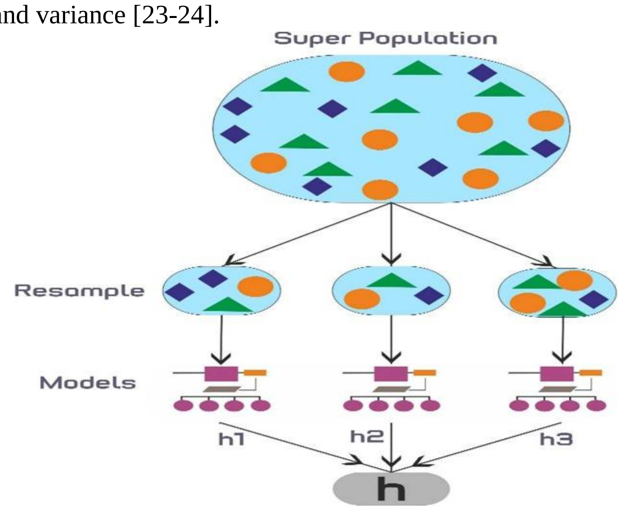

Figure 4 methods of boosting and resample of data and hypothesis

F. *Random forest*

The random forests are an ensemble of unpruned classification or regression trees. Random forest generates many classification trees. Each tree is constructed by a different bootstrap sample from the original data using a tree classification algorithm. After the forest is formed, a new object that needs to be classified is put down each of the tree in the forest for classification. Each tree gives a vote that indicates the tree's decision about the class of the object. The forest chooses the class with the most votes for the object. Flow process of random forest ensemble of machine learning algorithm shown in Figure 5. The main features of the random forests algorithm are listed as follows [25-26]:

- It is unsurpassable in accuracy among the current data mining algorithms.
- It runs efficiently on large data sets with many features.
- It can give the estimates of what features are important.
- It has no nominal data problem and does not over-fit.
- It can handle unbalanced data sets.

In random forests, there is no need for cross validation or a test set to get an unbiased estimate of the test error [27].

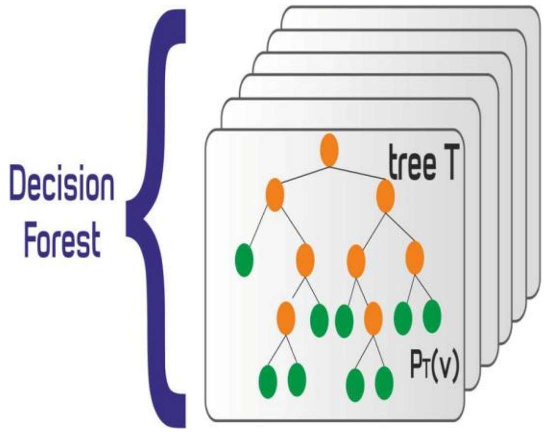

Figure 5 flow process of random forest ensemble of machine learning algorithm

Glowworm Algorithm

Initially all data of NSE is distributed in from of glowworm and process for the local decision. The mapped data designed the objective function J(xi(t)) at its current location xi(t) into α luciferin value li and broadcasts the same within its neighborhood. The set of neighbors (Ni(t)) of glowworm i consists of those glowworms that have relatively higher luciferin value that are located within a dynamic decision domain and updating by formula 1 at each iteration.

Local decision range update is given by equation 1

$$r_d^i(t+1) = min\left\{rs, max\left\{0, r_d^i(t) + \beta\left(nt - |Ni(t)|\right)\right\}\right\} \ldots\ldots (1)$$

And r_d^i (t+1) is glowworm is local decision range at the t+1 iteration, rs is the sensor range, nt is the neighborhood range. The number of glow in local decision range is given by equation (2)

$$N_{i(t)} = \left\{j : \|xi(t) - xi(t)\| < r_d^i ; li(t) < li(t)\right\} \ldots\ldots (2)$$

And xi(t) is the glowworm I position at the t iteration(t) is the glowworm i luciferin at the t iteration the set of neighbor of glowworm i consist of those glowworm that have relatively higher luciferin value and that are located within dynamic decision domain whose range r_d^i is bounded above by a circular sensor range.

Each glowworm is given in equation (3)

$$p_{ij}(t)=\frac{l_j(t)-l_i(t)}{\sum_{k\in N_i(t)} l_k(t)-l_i(t)} \ldots\ldots(3)$$

Movement update is given in equation (4)

$$x_i(t+1)=x_i(t)+s\left(\frac{x_j(t)-x_i(t)}{\|x_j(t)-x_i(t)\|}\right) \ldots\ldots(4)$$

Luciferin update is given in equation (5)

$$l_i(t)=(1-\rho)l_i(t-1)+\gamma J(x_i(t)) \ldots\ldots(5)$$

And li(t) is a luciferin value of glowworm i at the t iteration, P belong (0,1) lead to the reflection of the cumulative kindness of the path followed by the glowworm in their current luciferin values the parameter Y only scale the function values, J(xi(t)) is the value of test function [28].

III. METHODOLOGY

In this section describe the proposed algorithm for design of ensemble classifier. In the ensemble classifier base class is support vector machine. We used glowworm optimization algorithm to optimize the feature set of NSE stock market data. The glowworm optimization algorithm optimized the feature of NSE data and set to all classifier for the process of ensemble. Block diagram of feature optimization with glowworm and training phase of ensemble of classifier shown in Figure 6. The process of ensemble used F features and M different classifier. The size of ensemble classifier is F*M. Figure 7 represents of process block diagram of ensemble classifier with ensemble algorithm SVM, KNN and DT

Process of ensemble

1. Input: set of features data $Fd = \{f_1, \ldots\ldots, f_n\}$

2. Output: set of base classifier (SVM)={KNN(fn),……..,DT(f_n)}.
3. Check feature space F*M
4. $[F * M] \leftarrow 0;$ {null feature space}
5. for all $FD \in base\ classifier$ do
6. $knn(fn) \leftarrow$Fd(M)
7. if the set of features data process in DT
8. $(DT) \leftarrow SVM - KNN(fn)$
9. for all $base \in\ fd$do
10. Estimate $Fd - knn(fn), M(f), DT(fn)$
11. end for
12. return $base\ classifer\ SVM$

Condition of ensemble

Consider the base class classifier data is D. the D data set is optimized feature subset of glow-worm. The ensemble classifier as $C_1, C_2, \ldots\ldots\ldots C_n$. each ensemble classifier categorized by non-separable plan of SVM. These categorizations are non-overlapping, that is $C_d \cap C_e = \emptyset\forall_{d\neq e}$.

1. Consider base class feature data F $D = \{f_1, f_2, \ldots\ldots\ldots f_n\}$, where eac $f_j = \{o_{j1}, o_{j2}, \ldots\ldots . o_{js}\}$ is a set of features and the ensemble classes $\{C_1, C_2, \ldots\ldots\ldots C_k\}$, determine the non-overlapping categorization of FD, a ensemble set ES of class C_k, ES(C_k), is the mapping of features of ensemble class. That is, $ES(C_K)$ should predict instances of C_k.

$$ES(C_k) = \{\{o_j, o_k, \ldots\ldots o_m\} \subseteq \{o_1, ao_2, \ldots\ldots o_s\}\} \ldots\ldots\ldots (1)$$

2. Let L_i be the line of separation of base class and ensemble class the condition of hyper plan of ensemble class is

$$L_{i+1}[ES(C_g) = ES(C_g) \cup o_r] > P_i[ES(C_K)] \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$L_{i+1}[ES(C_k) \cup ao_r] > L_{i+1}[ES(C_K) \cup o_t], \forall t \neq r \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

3. Voting class the voting class process of mapped feature of ensemble class of features is$ES(C_k)$.

$$VOT_{ES(C_K)} = Vot(ES(C_K)) \leq M \ldots\ldots\ldots\ldots\ldots (4)$$

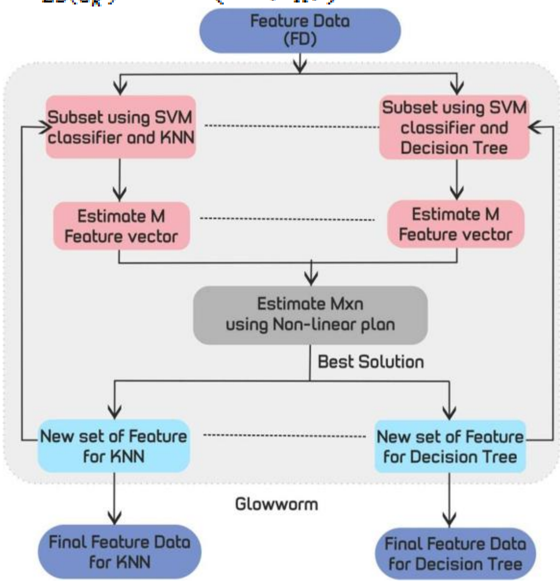

Figure 6 block diagram of feature optimization with glowworm and training phase of ensemble of classifier

Algorithms

Input: training data$FD = \{f_1, f_2, \ldots\ldots\ldots fn\}$,, where each $f_j = \{o_{j1}, o_{j2}, \ldots\ldots . o_{js}\}$ is a set of features of ensemble classes $\{C_1, C_2, \ldots\ldots\ldots . C_k\}$

Output: final Prediction ES(Ck)

for each $C_i,\ i{:}\ 1\ to\ k$

$ES\_voting[] = 0$

$ES(C_i) = \emptyset$

$vot = 0$

$\boldsymbol{repeat}$

$space = 0$

$feature = 0$

$f = o_j$

$for\ each\ feature\ o_{jh} \in f_j, where\ h{:}1\ to\ f$

$ES(C_i)' = ES(C_i)' \cup o_{jh}$

$f = f - o_{jh}$

$train\ classifier\ with\ SVM$

$test\ classifier\ with\ ES)$

$Compute\ voting\ \ f$

Retrieval Number: B2683078219/19©BEIESP
DOI: 10.35940/ijrte.B2683.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$if voting > F * M$$
$$voting = pridication$$
$$vot = vot + 1$$
$$ES_{voting[vot]} = predication$$
$$ES(C_i) = ES(C_i)' \cup FD$$
$$Until\{ES_{voting[vot]} \le CS_{voting[vot1]}\}$$
$$ES[i] = ES(C_i)$$

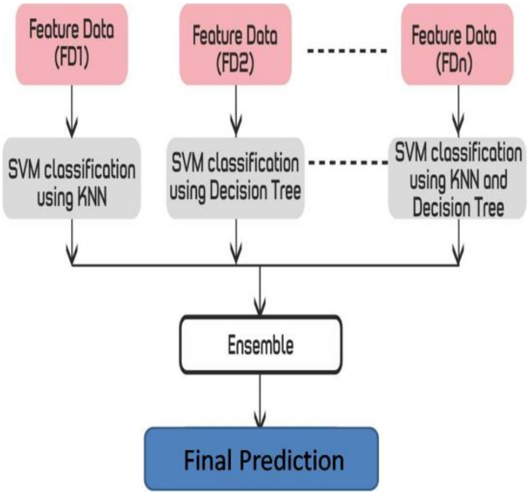

Figure 7 process block diagram of ensemble classifier with ensemble algorithm SVM, KNN and DT

IV. DATASET DESCRIPTION

In this section, discusses the process of result analysis of proposed algorithm and optimization of swarm intelligence. The proposed algorithm implemented in MATLAB 14.0. for the process of analysis used NSE dataset. For the evaluation used standard parameters NMSE, RMSE, MAE and MI. The proposed (bag machine learning) result compares with deep neural network, heterogeneous auto regression, and cascaded regression method [41-44]. Table 1, 2 and 3 are shown Input Data taken from National Stock Exchange of India (NSE) Stock option of Andhra Bank, ICICI and RBL.

Table 1 Input Data taken from National Stock Exchange of India (NSE) Stock option of Andhra Bank

Symbol	Strike Price	Settle Price	Underlying Value
ANDHRA BANK	75	0.05	49.85
ANDHRA BANK	75	0.05	50.65
ANDHRA BANK	75	0.1	51.05
ANDHRA BANK	75	0.05	51.5
ANDHRA BANK	75	0.05	51.85
ANDHRA BANK	70	0.05	49.5
ANDHRA BANK	67.5	0.1	47.85
ANDHRA BANK	75	0.05	53.2
ANDHRA BANK	70	0.05	49.85
ANDHRA BANK	75	0.05	53.95

Table 2 Input Data taken from National Stock Exchange of India (NSE) Stock option of ICICI Bank

Symbol	Strike Price	Settle Price	Underlying Value
ICICIBANK	310	0.2	190.75
ICICIBANK	300	0.05	192
ICICIBANK	310	0.05	198.45
ICICIBANK	300	0.05	193.55
ICICIBANK	310	0.35	204.05
ICICIBANK	300	0.15	199.25
ICICIBANK	280	0.05	187
ICICIBANK	270	0.05	183
ICICIBANK	300	0.25	203.5
ICICIBANK	270	0.2	184.8

Table 3 Input Data taken from National Stock Exchange of India (NSE) Stock option of RBL Bank

Symbol	Strike Price	Settle Price	Underlying Value
RBLBANK	620	0.05	505.85
RBLBANK	600	0.75	490.3
RBLBANK	600	0.8	490.35
RBLBANK	620	2.3	508.15
RBLBANK	620	0.05	508.95
RBLBANK	620	1.75	509.5
RBLBANK	620	0.15	511.7
RBLBANK	600	0.9	497.2
RBLBANK	620	0.65	516.35
RBLBANK	620	3	516.55

V. RESULT ANALYSIS

We evaluate prediction performance using four measures: normalized mean squared error (NMSE), root mean squared error (RMSE), mean absolute error (MAE), and mutual information (MI).

A. Normalized Mean Squared Error, NMSE

Given a set of target returns and their predicted values, $\{r^n_{t+1}, r^{\wedge n}_{t+1}\}^N_{n=1}$, NMSE is defined as

$$NSME = \frac{1}{N}\frac{\sum_{n=1}^{N}(r^n_{t+1} - r^{\wedge n}_{t+1})^2}{var(r^n_{t+1})}$$

Where $var(\cdot)$ denotes variance. recall that $var(r^n_{t+1}) = min_c \frac{1}{n=1}(r^n_{t+1} - C)^2$; NMSE is a mean squared error (MSE) normalized by the least MSE obtained from a constant prediction.

B. Root Mean Squared Error, RMSE

RMSE is the square root of MSE, defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(r_{t+1}^{n} - r^{\wedge n}_{t+1})^2}$$

C. Mean Absolute Error, MAE

MAE is defined as follows:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|r_{t+1}^{n} - r^{\wedge n}_{t+1}|$$

Note that inequality holds for the last two measures, $MAE \le RMSE \le \sqrt{N}\,MAE$; both error measures are known to informative, e.g., while MAE gives the same weight to all error amounts, RMSE is more sensitive to outliers, and is more suit for normal distributed error. For more interpretations, see Chai and Draxler, 2014.

D. Mutual Information, MI

MI measures dependency between r_{t+1} and u_t, and is denned as follows:

$$MI(r_{t+1}; u_t) = \sum_{r_{t+1}, u_t} p(r_{t+1}, u_t)\, log\frac{p(r_{t+1}, u_t)}{p(r_{t+1})P(u_t)} \approx \frac{1}{N}\sum_{n=1}^{N} \log\frac{p(r_{t+1}^{n} | u_t^{n})}{p(r_{t+1}^{n})}$$

$MI(r_{t+1}; u_t)$ is zero when the two variables are independent, and bounded to the information entropy, $H(r_{t+1}) = -\sum_{r_{t+1}} p(r_{t+1})\; log\; p(r_{t+1})$, when the two variables are fully dependent. From the assumption made earlier, we have $r_{t+1} | u_t \sim N(r^{\wedge}_{t+1}, \beta)$. With an additional assumption, $r_{t+1} \sim N(\mu, \sigma)$ time at the parameters β, μ and σ from the sample and evaluate MI from (20). In Fig. 8 indicates the variation of root mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the Andhra Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of RMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method RMSE values are 1.5986, 1.1524, 0.9024, 0.8698 and 0.79321 heterogeneous auto regression method RMSE values are 1.3545, 0.986, 0.8077, 0.7612 and 0.6674 cascaded regression method RMSE values are 0.7388, 0.6762, 0.5625, 0.62023 and 0.51147 and similarly proposed bag machine learning method RMSE values are 0.62481, 0.42417, 0.51325, 0.5531 and 0.48481.

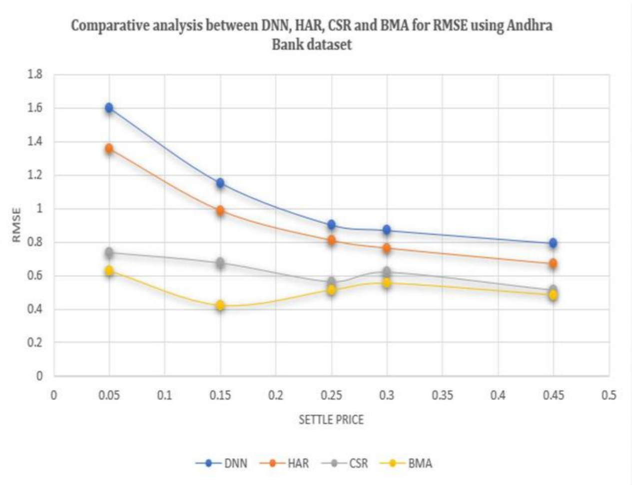

Figure 8 Comparison DNN, HAR, CSR and BMA for RMSE using Andhra Bank dataset

In Fig. 9 indicates the variation of normalized mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the Andhra Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of NMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method NMSE values are 0.88127, 0.67392, 0.9829, 0.84758 and 0.87001 heterogeneous auto regression method NMSE values are 0.85217, 0.66498, 0.9024, 0.7254 and 0.81471 cascaded regression method NMSE values are 0.6752, 0.62712, 0.72996, 0.67854 and 0.71254 and similarly proposed bag machine learning method NMSE values are 0.57951, 0.60741, 0.72004, 0.65991 and 0.66244.

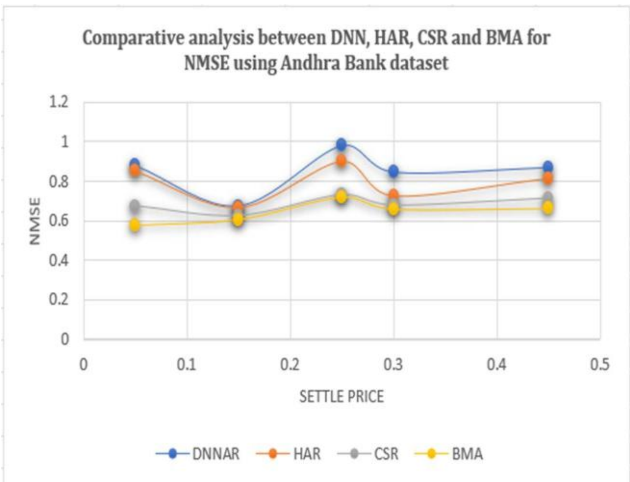

Figure 9 Comparison DNN, HAR, CSR and BMA for NMSE using Andhra Bank dataset

In Fig. 10 indicates the variation of mean absolute error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the Andhra Bank dataset.

Retrieval Number: B2683078219/19©BEIESP
DOI: 10.35940/ijrte.B2683.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MAE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MAE values are 1.964, 1.5252, 1.828, 0.9027 and 1.2247 heterogeneous auto regression method MAE values are 1.6784, 1.6621, 1.5215, 0.98761 and 0.78949 cascaded regression method MAE values are 0.8975, 1.2311, 0.96887, 0.82478 and 0.77697 and similarly proposed bag machine learning method MAE values are 0.77965, 0.84621, 0.76842, 0.72144 and 0.65651.

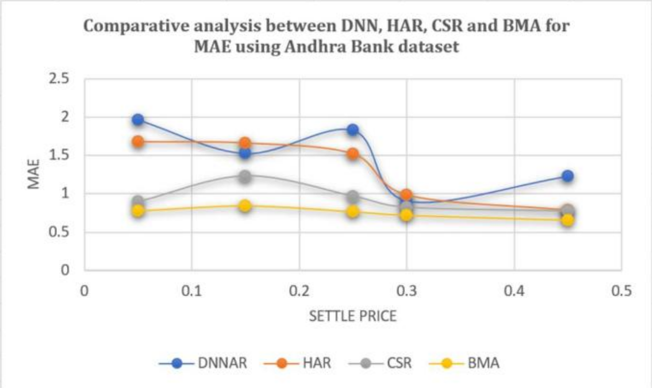

Figure 10 Comparison DNN, HAR, CSR and BMA for MAE using Andhra Bank dataset

In Fig. 11 indicates the variation of mutual information between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the Andhra Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MI is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MI values are 0.95689, 0.865881, 0.68985, 0.77589 and 0.826528 heterogeneous auto regression method MI values are 1.0388, 0.94627, 0.92776, 0.92313 and 0.91542, cascaded regression method MI values are 1.28745, 0.99684, 0.98652, 1.2567 and 0.99241 and similarly proposed bag machine learning method MI values are 1.53594, 1.67952, 1.24151, 1.88169 and 1.4427.

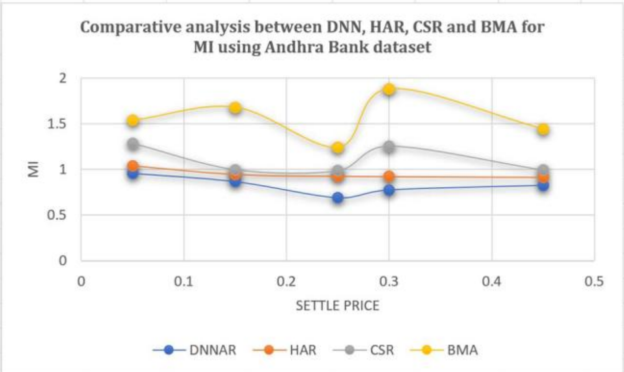

Figure 11 Comparison DNN, HAR, CSR and BMA for MI using Andhra Bank dataset

In Fig. 12 indicates the variation of root mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the ICICI dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of RMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method RMSE values are 1.1174, 0.87008, 0.92452, 0.91274 and 0.786952, heterogeneous auto regression method RMSE values are 0.986, 0.6985, 0.82114, 0.76261 and 0.77294, cascaded regression method RMSE values are 0.80042, 0.62122, 0.73271, 0.75569 and 0.59622 and similarly proposed bag machine learning method RMSE values are 0.59856, 0.52325, 0.46852, 0.5531and 0.57414.

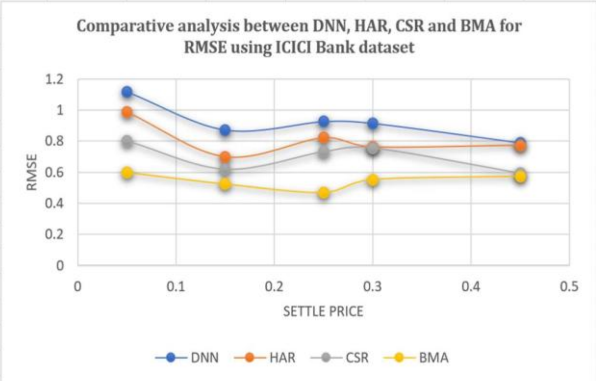

Figure 12 Comparison DNN, HAR, CSR and BMA for RMSE using ICICI Bank dataset

In Fig. 13 indicates the variation of normalized mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the ICICI Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of NMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method NMSE values are 0.945895, 0.85476, 0.786259, 0.95652 and 0.88287, heterogeneous auto regression method NMSE values are 0.97521, 0.84624, 0.74581, 0.8952 and 0.74258, cascaded regression method NMSE values are 0.78587, 0.68956, 0.66845, 0.85641 and 0.70014 and similarly proposed bag machine learning method NMSE values are 0.6895, 0.65147, 0.55211, 0.79588 and 0.68781.

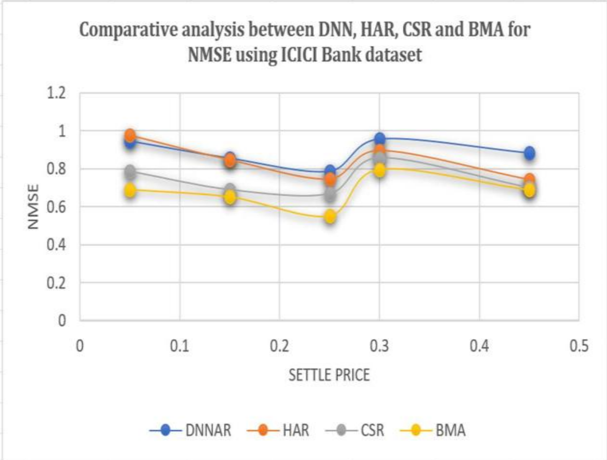

Figure 13 Comparison DNN, HAR, CSR and BMA for NMSE using ICICI Bank dataset

In Fig. 14 indicates the variation of mean absolute error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the ICICI Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MAE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MAE values are 1.56878, 1.2501, 0.95621, 1.02489 and 0.97691, heterogeneous auto regression method MAE values are 1.2741, 1.11741, 0.98952, 0.99001 and 0.90174, cascaded regression method MAE values are 0.87567, 0.89478, 0.90174, 0.68585 and 0.79569 and similarly proposed bag machine learning method MAE values are 0.80274, 0.85475, 0.7732, 0.66846 and 0.54681.

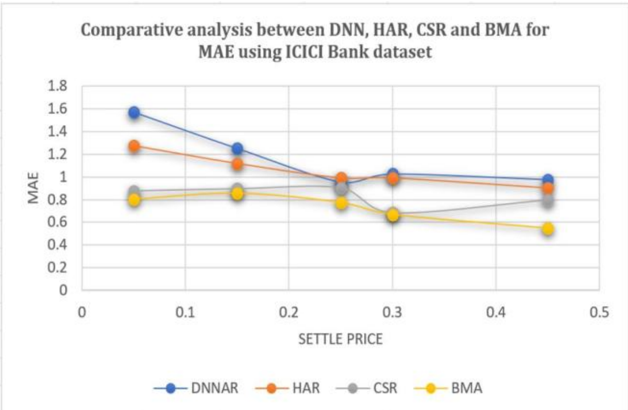

Figure 14 Comparison DNN, HAR, CSR and BMA for MAE using ICICI Bank dataset

In Fig. 15 indicates the variation of mutual information between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the ICICI Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MI is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MI values are 0.69758, 0.759864, 0.89887, 0.74698 and 0.46985, heterogeneous auto regression method MI values are 0.76895, 0.84576, 0.92471, 0.85254 and 0.667416, cascaded regression method MI values are 0.88249, 0.94785, 0.98851, 1.0024 and 0.76985 and similarly proposed bag machine learning method MI values are 1.0785, 0.98652, 1.19891, 1.2741 and 1.0741.

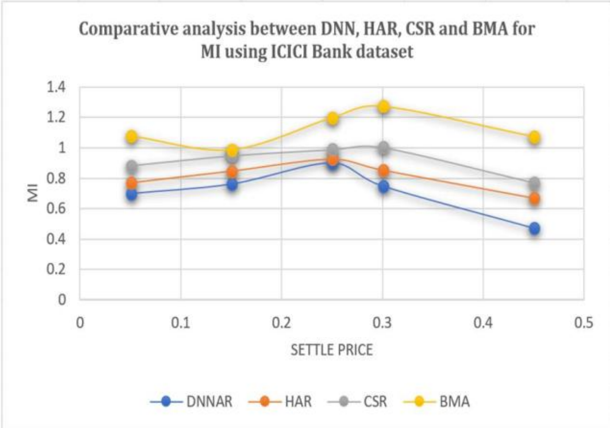

Figure 15 Comparison DNN, HAR, CSR and BMA for MI using ICICI Bank dataset

In Fig. 16 indicates the variation of root mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the RBL dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of RMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method RMSE values are 1.5864, 1.687952, 1.69852, 1.7757 and 1.72485, heterogeneous auto regression method RMSE values are 1.59875, 1.25895, 1.114784, 1.02522 and 1.00254, cascaded regression method RMSE values are 0.98697, 1.06385, 0.87687, 0.94287 and 0.82585 and similarly proposed bag machine learning method RMSE values are 0.77498, 0.94685, 0.68745, 0.7898 and 0.80248.

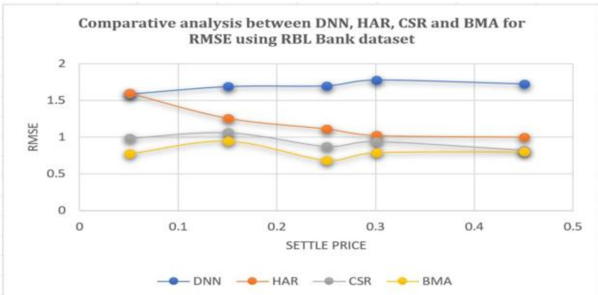

Figure 16 Comparison DNN, HAR, CSR and BMA for RMSE using RBL Bank dataset

In Fig. 17 indicates the variation of normalized mean square error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of NMSE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method NMSE values are 1.3241, 1.76421, 1.3421, 1.2541 and 1.12415, heterogeneous auto regression method NMSE values are 1.2414, 1.59342, 1.11478, 1.102 and 0.98524, cascaded regression method NMSE values are 0.97584, 0.96854, 0.98652, 0.92487 and 0.902487 and similarly proposed bag machine learning method NMSE values are 0.88476, 0.90462, 0.69881, 0.78942 and 0.88794.

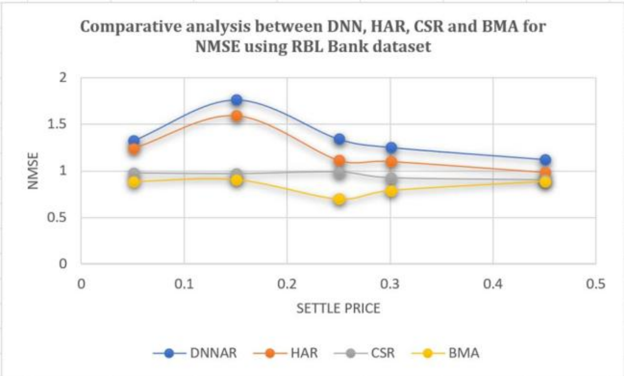

Figure 17 Comparison DNN, HAR, CSR and BMA for NMSE using RBL Bank dataset

In Fig. 18 indicates the variation of mean absolute error between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MAE is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MAE values are 1.54784, 1.22415, 1.3101, 1.22468 and 1.1589, heterogeneous auto regression method MAE values are 0.98544, 0.95685, 1.04934, 0.97856 and 1.27468, cascaded regression method MAE values are 0.97846, 0.86589, 0.990017, 0.92548 and 1.00457 and similarly proposed bag machine learning method MAE values are 0.89872, 0.78951, 0.78614, 0.89475 and 0.98546.

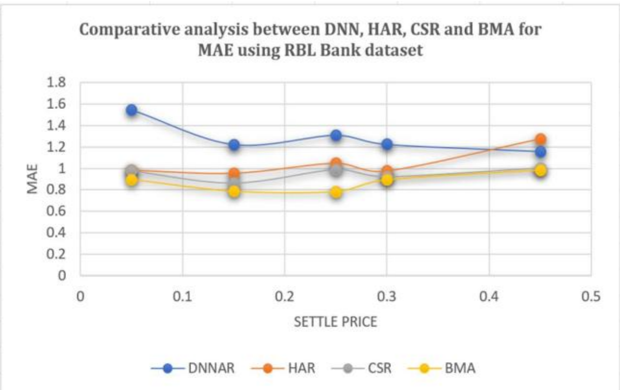

Figure 18 Comparison DNN, HAR, CSR and BMA for MAE using RBL Bank dataset

In Fig. 19 indicates the variation of mutual information between deep neural network, heterogeneous auto regression, cascaded regression method and bag machine learning method as a proposed method for the RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MI is optimized due to the process of optimization and better prediction of bag machine-learning. According to the settle price (0.05, 0.15, 0.25, 0.30 and 0.45) sequentially deep neural network method MI values are 0.88475, 0.98768, 1.2475, 1.0041 and 1.44274, heterogeneous auto regression method MI values are 0.85474, 0.95254, 1.0047, 0.98654 and 1.2041, cascaded regression method MI values are 0.775421, 0.79547, 0.89685, 0.95955 and 0.93685 and similarly proposed bag machine learning method MI values are 0.68694, 0.726895, 0.800247, 0.78842 and 0.79813.

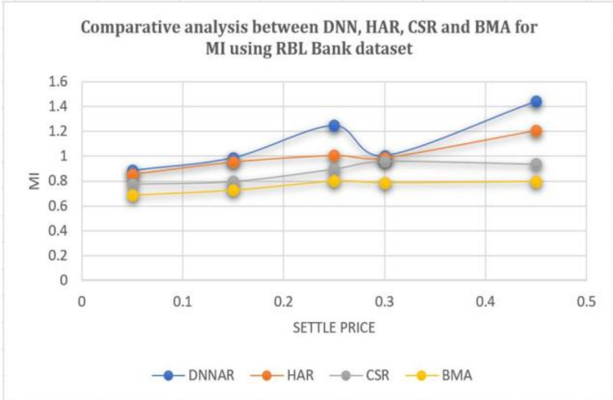

Figure 19 Comparison DNN, HAR, CSR and BMA for MI using RBL Bank dataset

In Fig. 20 indicates the variation of root mean square error between deep neural network methods heterogeneous auto regression, cascaded regression method and bag machine learning method as proposed method for Andhra Bank, ICICI Bank and RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of RMSE is optimized due to the process of optimization and better prediction of bag machine learning.

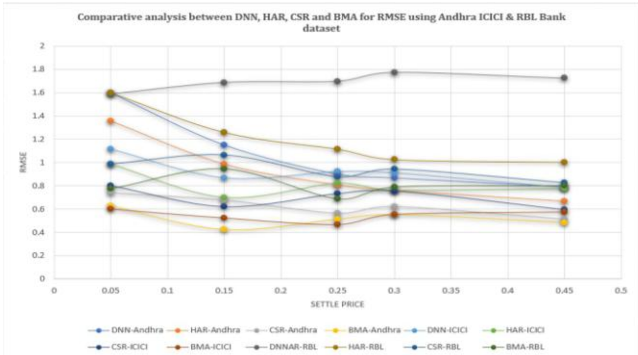

Figure 20 Comparison DNN, HAR, CSR and BMA for RMSE using Andhra, ICICI and RBL Bank dataset

In Fig. 21 indicates the variation of normalized mean square error between deep neural network methods heterogeneous auto regression, cascaded regression method and bag machine learning method as proposed method for Andhra Bank, ICICI Bank and RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of NMSE is optimized due to the process of optimization and better prediction of bag machine learning.

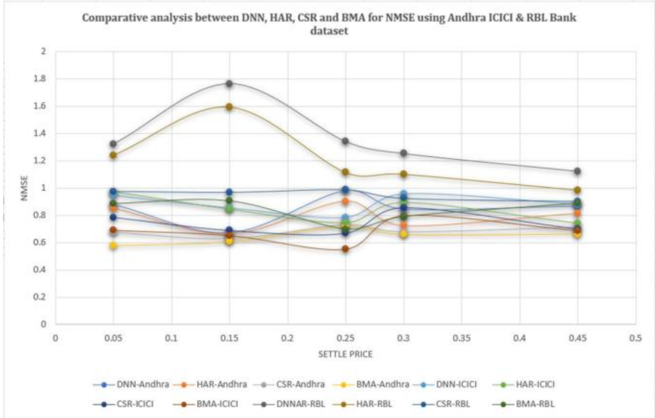

Figure 21 Comparison DNN, HAR, CSR and BMA for NMSE using Andhra, ICICI and RBL Bank dataset

In Fig. 22 indicates the variation of mean absolute error between deep neural network methods heterogeneous auto regression, cascaded regression method and bag machine learning method as proposed method for Andhra Bank, ICICI Bank and RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MAE is optimized due to the process of optimization and better prediction of bag machine learning.

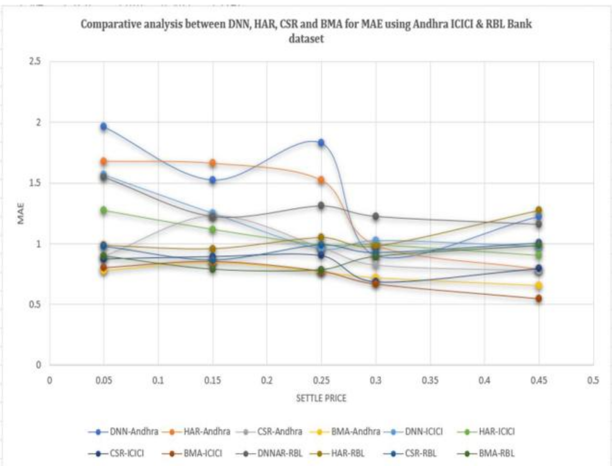

Figure 22 Comparison DNN, HAR, CSR and BMA for MAE using Andhra, ICICI and RBL Bank dataset

In Fig. 23 indicates the variation of mutual information between deep neural network methods heterogeneous auto regression, cascaded regression method and bag machine learning method as proposed method for Andhra Bank, ICICI Bank, RBL Bank dataset. The result of variation distributed in different settle price such as 0.05, 0.15, 0.25, 0.30 and 0.45 the variation indicates the value of MI is optimized due to the process of optimization and better prediction of bag machine learning.

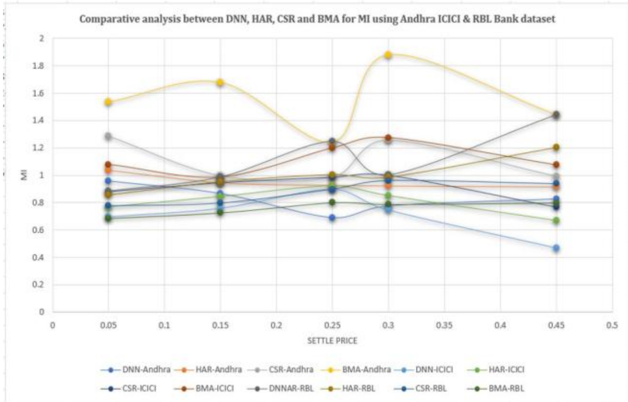

Figure 23 Comparison DNN, HAR, CSR and BMA for MI using Andhra, ICICI and RBL Bank dataset

VI. CONCLUSION & FUTURE WORK

This study proposed machine learning based ensemble classifier for the Prediction of option price in Indian stock market. In process of ensemble classifier used three classifier, support vector machine, nearest neighbor and decision tree. The support vector machine acts as base classifier and KNN and DT work as variable classifier for the optimization selection of data in base class. For the evaluation of the performance of proposed model used Indian stock market data approx. 6000 instances. The machine learning based ensemble classifier increase the value of mutual information and reduces the values of normalized mean square error. The increase value of mutual information indicates the reduces the value of variation of data and increases the possibility of Prediction. The base Prediction model of Prediction is used DNN (deep neural network) and HAR (heterogeneous autoregressive model). Validation of ensemble algorithms by four standard parameters NMSE, RMSE, MAE and MI. The values indicate better performance of ensemble classifier instead of CSR and HAR classifier. The glowworm optimization algorithm reduces the noise of stock data. The process of ensemble classifier enhanced the performance of option price in Indian stock market.

REFERENCES

1. Elham Ahmadi, Milad Jasemi, Leslie Monplaisir, Mohammad Amin Nabavi, Armin Mahmoodi and PegahAmini Jam "New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic", Elsevier, Expert Systems with Applications, 2018, Pp 21-31.
2. Madhu Iyengar, Nirmal Iyengar and HarmishSampat "Impact of US election results on Indian stock market: An event study approach", International Journal of Applied Research, 2017, Pp09-13.
3. Geetanjali Patel and Prof. Avinash Sharma "Data Classification using Multiple Support Vector Machine with Ant Colony Optimization", IJESC, 2017, Pp 13000-13003.
4. Kashif Bin Zaheer, Mohd Ismail Bin Abd Aziz, Amber Nehan Kashif and Syed Muhammad Murshid Raza "Two Stage Portfolio Selection and Optimization Model with the Hybrid Particle Swarm Optimization", Matematika, 2018, Pp 125-141.

5. Tamal Datta Chaudhuri and Indranil Ghosh “Forecasting Volatility in Indian Stock Market using Artificial Neural Network with Multiple Inputs and Outputs”, International Journal of Computer Applications, 2015, Pp 7-15.
6. LuckysonKhaidem, SnehanshuSaha and Sudeepa Roy Dey “Predicting the direction of stock market prices using random forest”, Applied Mathematical Finance, 2016, Pp 1-20.
7. MohamadrezaTahmasebi and Hossein Soleimani “Designing the Best Stock Portfolio of Tehran Stock Exchange Using the Bee Colony Algorithm”, RevistaPublicando, 2018, Pp 471-482.
8. David M. Q. Nelson, Adriano C. M. Pereira and Renato A. de Oliveira “Stock Marketâ€™s Price Movement Prediction with LSTM Neural Networks”, IEEE, 2017, Pp 1419-146.
9. Chuong Luong and Nikolai Dokuchaev "Forecasting of Realised Volatility with the Random Forests Algorithm", J. Risk Financial Manag., 2018, Pp 1-15.
10. J. Shahrabi and S. MottaghiKhameneh “development of a hybrid system of artificial neural networks and artificial bee colony algorithm for prediction and modelling of customer choice in the market”, Journal of Fundamental and Applied Sciences, 2016, Pp 154-183.
11. Thomas Trier Bjerring, Omri Ross and Alex Weissensteiner “Feature Selection for Portfolio Optimization”, Annals of Operations Research, 2016, Pp 1-25.
12. Armin Jabbarzadeh, Saeed Shavvalpour, Hossein Khanjarpanah and DavoodDourvash “A Multiple-Criteria Approach for Forecasting Stock Price Direction: Nonlinear Probability Models with Application in S&P 500 Index”, International Journal of Applied Engineering Research, 2016, Pp 3870-3878.
13. S. C. Nayak, B. B. Misra and H. S. Behera “ACFLN: artificial chemical functional link network forÂ prediction ofÂ stock market index”, Springer, 2018, Pp 1-27.
14. Sarat Chandra Nayak, Bijan Bihari Misra and Himansu Sekhar Behera “An Adaptive Second Order Neural Network with Genetic-Algorithm-based Training (ASONN-GA) to Forecast the Closing Prices of the Stock Market”, International Journal of Applied Metaheuristic Computing, 2016, Pp 39-57.
15. Mahmut Taha Yazici, ShadiBasurra and Mohamed Medhat Gaber “Edge Machine Learning: Enabling Smart Internet of Things Applications”, big data an d cognitive computing, 2018, Pp 1-17.
16. Taiyong Li, Min Zhou, Chaoqi Guo, Min Luo, Jiang Wu, Fan Pan, Quanyi Tao and Ting He “Forecasting Crude Oil Price Using EEMD and RVM with Adaptive PSO-Based Kernels”,Energies, 2016, Pp 1-21.
17. Asheesh Pandey and Sanjay Sehgal “Explaining Size Effect for Indian Stock Market”, Asia-Pacific Finan Markets, 2016, Pp 46-68.
18. Jason M. Hunter, Holger R. Maier, Matthew S. Gibbs, Eloise R. Foale, Naomi A. Grosvenor, Nathan P. Harders and Tahali C. Kikuchi-Miller “Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems”, Hydrol. Earth Syst. Sci., 2018, Pp 2987-30006.
19. Garima Malik, Aakansha Rathore and Sonakshi Vij “Utilizing Various Machine Learning Techniques to Classify Data in the Business Domain”, IJIRST, 2017, Pp 118-122.
20. S.O. Edeki, O.O. Ugbebor and E.A. Owoloko “On A Dividend-Paying Stock Options Pricing Model (Sopm) Using Constant Elasticity of Variance Stochastic Dynamics”, International Journal of Pure and Applied Mathematics, 2016, Pp 1029-1036.
21. Thuy-AnDinh and Yung-Keun Kwon “An Empirical Study on Importance of Modeling Parameters and Trading Volume-Based Features in Daily Stock Trading Using Neural Networks”, informatics, 2018, Pp 1-12.
22. Mitesh Patel, Dr. Munjal Dave and Dr. Mayur Shah “Stock Price and Liquidity Effect of Stock Split: Evidence from Indian Stock Market”, IJMRR, 2016, Pp 1030-1039.
23. Kiran Kumar Kotha and Bhawna Sahu “Macroeconomic Factors and the Indian Stock Market: Exploring Long and Short Run Relationships”, International Journal of Economics and Financial Issues, 2016, Pp 1081-1091.
24. Pramod Kumar Naik and Puja Padhi “Investor sentiment, stock market returns and volatility: evidence from National Stock Exchange of India”, Int. J. Management Practice, 2016, Pp 213-237.
25. Josip Arneric and Tea Poklepovic “nonlinear extension of asymmetric garch model within neural network framework”, CCSEIT, 2016, Pp 101-111.
26. Ayman E. Khedr, S.E.Salama and Nagwa Yaseen “Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis”, I.J. Intelligent Systems and Applications, 2017, Pp 22-30.
27. Amirhossein Ghaznavi, Mohammad Aliyari and Mohammad Reza Mohammadi “Predicting Stock Price Changes of Tehran Artmis Company Using Radial Basis Function Neural Networks”, International Research Journal of Applied and Basic Sciences, 2016, Pp 972-978.
28. OzgurEvren and Stefania Minardi "Warm-Glow Giving and Freedom to be Selfish*", Econ Papers, 2017, Pp 1-49.

AUTHORS PROFILE

Ms. Payal Shrivastava, (17th June, 1987) has done her Bachelor of Science in 2007 (Barkatullah University, Bhopal), M.Sc. (Mathematics) in 2011 (Barkatullah University, Bhopal). Her research interest includes financial mathematics and Machine learning. She is having 7 years of teaching experience and is working as Research Scholar in the Department of Mathematics and Computer Application at Maulana Azad National Institute of Technology, Bhopal-462051 (M.P.) India.

Dr. C. K. Verma (8th June 1975) has done Master of Science in Mathematics from Govt. Science College Jabalpur India in 1998 and has Qualified National Eligibility Test (NET) in 2000.He has done his Ph.D from National Institute of Technology, Bhopal in 2012. His area of research includes Computational Biology and Financial Mathematics. He is having 18 years of teaching experience and is working as Associate Professor in the Department of Mathematics at Maulana Azad National Institute Bhopal, India. He has published 35 papers in international journal, 05papers in national journal and 05 papers in proceedings of international and national conference. He guided more than 4 Ph.D.