ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 9, DOI:10.29027/IJIRASE.v4.i7.2021.891-904,March 2021

Exploring Data Visualization to Analyze and Predict Customer Loyalty in Banking Sector with Ensemble Learning

Jesmi Latheef[1], S.Vineetha[2]
Computer Science and Engineering, Rajiv Gandhi Institute of Technology
Kottayam, Kerala, India
Email: jesmilatheef@gmail.com[1], svineetha@rit.ac.in[2]

Abstract—Customer loyalty or customer churn, also known as customer attrition, refers to the phenomenon whereby a customer leaves a company. Some studies confirmed that acquiring new customers can cost five times more than satisfying and retaining existing customers. There are a lot of benefits that encourage the tracking of the customer churn rate. Customer value analysis and customer churn predictions will help marketing programs target more specific groups of customers. Churn's prediction could be a great asset in the business strategy for retention applying before customers' exit. In this paper, the banking sector's churn data is used and explores the data with Python data visualization packages such as Matplotlib, Seaborn, and the very new Plotly. This paper aims to identify and visualize which factors contribute to customer churn and to build a prediction model using Ensemble Learning Algorithm. Compare the system accuracy with each model and visualize the results. Preferably based on model performance, choose the model and that will make easier for organizations to target the customers with more chances to become churn. Thus it will allow avoiding loss of revenue of the corresponding organization.

Index Terms—Customer Relationship Management, Customer Churn, Visualization, Ensemble Algorithm

I. INTRODUCTION

Customer Relationship Management (CRM) requires different methods and approaches to maintain a positive relationship with both current and increasing customers. For higher customer retention, companies must ensure consumers are happy with their goods and services. CRM is a company-based strategic plan of action designed to accomplish a long-term or overall target of making or becoming greater sales and profitability, lowering costs and rising customer loyalty, which is more important than ever to companies because it will help you attract new customers and sustain existing ones. Fig.1 demonstrates that when clients or subscribers avoid doing or leave the business with a company or service, customer churn occurs. It is important because it costs more to attract new customers than it does to keep current customers in possession. In fact, a few customer retention increases will produce an increase in profit. CRM programmes use market intelligence and predictive models to determine the most profitable category of customers and target them to achieve higher customer retention rates. These models can predict customers with a high probability of churning based on customers' personal, behavioural data analysis.

Fig.1. Customer churn happens when customers stop doing or exist business with a company or service

Churn prediction is one of the most popular big data use cases in business. It includes finding clients who are likely to cancel a service subscription and offers an excellent opportunity to improve customer satisfaction and prevent revenue loss. In several market areas, including but not limited to subscriptions to telecommunications, insurance, retail, and cloud services, it is increasingly being investigated. By applying statistical models to historical data and trying to identify a trend in customers that can lead to churn, customer churn analysis can be created. Artificial intelligence (AI) and machine learning can predict future trends and patterns of customer churn and identify previously hidden indicators that help predict churn. Several attempts have been made to compare the current churn prediction techniques and benchmark them. A distinction was made between (Decision trees, Logistic regression and Neural Network) models in [1], [2]. There are also a number of similarities made between various models. This paper explores visualization to analyze the data behaviors and trends and implement an Ensemble-based algorithm using Voting Classifier and all other modeling techniques.

Visualization plays an important part of data analytics and helps interpret big data in a real-time structure by utilizing complex sets of numerical or factual figures. Due to the way the human brain processes information, presenting insights in charts or graphs to visualize significant amounts of complex data is more accessible than relying on spreadsheets or reports. The most compelling data visualization methods are used to succeed in presenting the data effectively. Together with the demand for data visualization and analysis, the tools and solutions in this area develop fast and extensively. In this study, we use some basic visualization libraries to explore the churn data and. Pythons' most effective exploring library named Plotly is used. Plotly is a Python graphing library that makes interactive, publication-quality graphs to analyze and understand the data.

This paper is structured as follows: Section 2 provides a peer review for churn prediction and visualization tools of different data mining techniques. The study's design and methodology are discussed in Section 3, implementation and assessment are given in Section 4, and this work is concluded in Section 5.

II. Literature Survey

Machine learning models have been commonly used to assess how often consumers can adjust or not. Most of the literature are discussed about various machine learning algorithms, Deep Learning algorithms, a brief of the literature of chosen algorithms are describing here,

A. *Logic Regression*

The regression analysis techniques are primarily aimed at conducting research and approximately estimating the relationships between a series of characteristics. Regression models were used to evaluate the relationship between one dependent variable and a set of independent variables. Logistic Regression (LR) is the optimal regression analysis model when the target variable is binary. LR is a predictive analysis used to describe the relationship between a target variable that is binary and a set of function variables. As a function of collecting variables or customer features for customer churn[3]-[5], LR has been widely used to describe the churn likelihood.

B. *Decision Tree*

To develop a structure-like tree that constitutes a set of decisions, a Decision Tree (DT) is used. Chances of becoming churn or scores of class membership is the output of the decision tree. The essential elements of the Decision tree are: a) Internal Nodes, a single variable/feature is represented by each node and to represents a point of the test at variable level; b) Branches; these are represented by lines that may be a connection to leaf nodes and which represent the outcome of test c) Leaf Nodes, which represent the class labels. DT assisted in finding both categorical and continuous data, so it is a flexible model. Their flexibility is attained as the most widely used models for finding whose is churn [6]-[7].

C. *Support Vector Machine*

One of the most common supervised learning techniques is SVM and developed for analysis of data to find patterns. The input is a collection of labeled training data; SVM represents data in multi-dimensional or a high dimensional space as points and tries to find the most suitable classifying hyperplanes between fragments of different categories of data. Based on their nearness in space to the separating gap, new occurrences are represented in the same space and are separated to a particular group. SVM techniques have been most frequently inspected and examined to be of high predictive performance for churn prediction [8]-[10].

D. *Random Forest*

The algorithm for Random Forest is a supervised algorithm for classification. It is to create a forest and make it random in some way. There is a direct connection between the number of trees in the forest and the results that can be obtained. The higher the number of trees, the more precise the outcome is. Overfitting is a crucial problem that can make the outcomes worse, but the classifier would not overfit the model if there were enough trees in the forest for the Random Forest algorithm. The benefit is that the Random Forest classifier can handle missing values, and the other advantage is that it is possible to model the Random Forest classifier based on categorical values [11]-[13].

E. Ensemble Algorithms

Ensemble learning helps by integrating multiple models to optimize machine learning outcomes. This approach offers better predictive results as opposed to a single model. In order to reduce the effect of variance (bagging), bias (boosting), and increase predictions (stacking), many machine learning techniques are merged into one predictive model, so these are referred to as the Ensemble of Meta-Algorithms [14]-[17]. Along with the ensemble technique, the data can be visualized using different methods. The ggplot2 package is represented as the term in graphics and a system for understanding graphics as composed of different layers used to make various plots for the same characteristics [18]-[19]. The most commonly-used library for creating simple and powerful visualizations in the Python community is Matplotlib and is the first Python data visualization tool. The library provides to make a variety of graphs from histograms to swarm plots to line plots. Finally, Seaborn can also be used with data visualization.

III. System Architecture

The proposed system architecture of churn prediction using the ensemble method with exploring data visualization techniques are discussed. Fig.2 shows the entire architecture of the system. The system begins with importing the churn data of the bank and the following steps are performing.

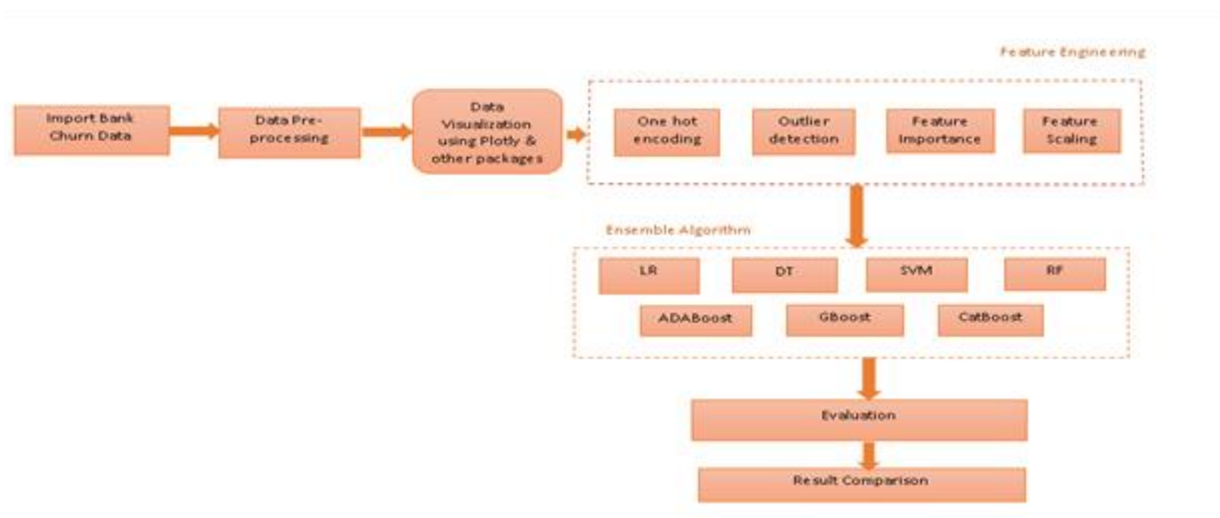

Fig.2. Overview of System Architecture

A. Data Pre-processing

From the row data null value removal, irrelevant feature removals are the first step in the data processing. If there are null values, they can be replaced either with the feature's mean value or remove from those data points. The next step is to identify and removes irrelevant features that affect the algorithm. Thus irrelevant features can be regarded as noise whose presence will affect the final result adversely.

B. Data Visualization

In this step, we can get more idea about the distribution of the data in different strategies. For that, Python data visualization packages such as matplotlib, Seaborn are used. In addition to this, we are using

Python's Plotly packages, which provide more interactive plots. Plotly is one of the important data visualization tools available and is built on top of visualization library D3.js, HTML and CSS. Plotly also allows you to share the plots with someone else in various formats.

C. Feature Engineering

Feature Engineering is for making the proper input dataset suitable with the machine learning algorithm requirements and improving the performance of the models that we will create. Following are some of the techniques used for feature engineering.

1) One Hot Encoding: In machine learning, the most commonly used feature encoding method is One Hot Encoding. In this method, flag values such as 0 or 1 assigns to a specific feature and convert to multiple flag columns. These binary values express the relationship between clustered and encoded columns. This method translates the categorical data that the algorithms are very difficult to understand into a numerical value and allows categorical data to be grouped without losing any details.

2) Handling Outliers: To handle outliers of data, first we have to detect the outliers. For that, visual techniques like box plots can be used. Several statistical methods are also available for identifying outliers but to take a decision with high precision; it is important to visualize the outliers. The outlier is an inspection that is numerically different from the rest of the data or it is an abnormality in the data. When examining a boxplot, an outlier is defined as a data point situated outside the box plot's railing.

3) Feature Scaling: The dataset's numerical characteristics do not have a certain range in most instances and vary from each other. In terms of the spectrum, after a scaling process, the continuous features become identical. Normalization is one type of function scaling (min-max normalization). The dataset's numerical attributes do not have a certain range in most instances and vary from each other. In terms of the spectrum, after a scaling process, the continuous characteristics become identical. Normalization is one type of function scaling (min-max normalization).

4) Feature Importance: In order to calculate the features with more importance, some machine learning models were used. In this paper, the Random Forest algorithm was used for finding which one of the features is most important. In machine learning, feature selection was used as the key concept and extremely influences the performance of the described model. Inappropriate or partially appropriate features can negatively impact model performance. Proper selection of features can lead to improving the accuracy of the system. Here also several techniques are used. In this paper, we are using the feature importance property of the model. We get the feature importance of each feature of our dataset by using the model's feature importance property. Finding feature importance gives a score for each part of the data; the higher value of the score indicates the more important part.

D. Model Creation

After completing the above steps, the data is ready for modeling. This paper proposes an ensemble-based learning algorithm for customer churn prediction in the banking sector. Ensemble learning helps improve machine learning results by combining several models. As compared to a single model, this method provides better predictive performance. Several machine learning techniques are combined into one predictive model in order to reduce the effect of variance (bagging), bias (boosting), and increase predictions (stacking); thus, these are referred to as an ensemble of meta-algorithms. Ensemble learning is used to lower the error, reduce overfitting. There are several techniques available for ensemble modeling; here, we are using the Voting method. This is one of the simplest ways of cumulating the predictions from different machine learning algorithms. The voting classifier is a wrapper for different models trained and examined concurrently to utilize each algorithm's various oddities and is not an actual classifier. Different algorithms can be combined as an ensemble and train the model, then predict the final output. The final output and thereby predictions can be taken as majority vote manner based on two different methods of either soft voting or hard voting. If it is hard voting select, the class gained the highest number of votes or predict labels with majority voting. If it is soft voting, the winning class is selected as the highest value of majority voting. In this work, soft voting is used. The

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 9, DOI:10.29027/IJIRASE.v4.i7.2021.891-904,March 2021

multiple algorithms that are going to ensemble are of the following:

1) Logistic Regression (LR)
2) Decision Tree (DT)
3) Support Vector Machine (SVM)
4) Random Forest (RF)
5) Ada Boost
6) Gradient Boost
7) Cat Boost

Along with LR, DT, SVM, RF, the following are the three boosting algorithms used in the model.

1) Ada Boost: Ada Boost algorithm is known as Adaptive Boosting. It is mainly used for classification problems. The objective is to convert a set of weak classifiers into a stronger one. For classification, the equation can be represented as

$$F(x) = sign(\sum_{m=1}^{M}(\theta_m f_m(x))) \quad (1)$$

Where *fm* stands for the *m*th weak classifier and *θm* is the corresponding weight. It is exactly the weighted combination of M weak classifiers. Ada Boost is developed for binary classification problems and is considered one of the best algorithms. Generally, short decision trees are the primary part of Ada Boost, after creating the first tree, the performance of the tree on each training instance is used. Also, understanding how much attention and care is needed for the next tree use the performance to weigh. Thus, it is important to pay attention to each training instance. Hence, more weight is given to training data which is difficult to predict, and less weight is given to easy to predict instances.

2) Gradient Boost: The gradient Boost algorithm is works similarly as of Ada Boost. Gradient Boosting trains models in a gradual, additive and sequential. The major difference between Ada Boost and Gradient Boosting Algorithm is how the two identify the shortcomings of weak learners (e.g. decision trees). While the Ada Boost model identifies the shortcomings by using high weight data points, gradient boosting performs the same by using gradients in the loss function. The loss function is a measure indicating how good model's coefficients are at fitting the underlying data.

3) Cat Boost: Cat Boost is an algorithm for gradient boosting on decision trees. Problems, such as regression, classification, multi-class classification and ranking, are solved using Cat Boost. Different modes of the algorithm differ by the objective function that we are trying to minimize during gradient descend. Moreover, Cat Boost has pre-built metrics to measure the accuracy of the model. It is the extension of the MatrixNet model, which is famous for rating jobs such as forecasting and making recommendations. It is a global version and can be used across a wide range of domains and to different problems.

Voting Classifier is used to combine or ensemble these algorithms. Voting is one of the finest ways to combine the observations from the above-discussed machine learning algorithms. Voting classifier is a classifier and a wrapper for set of different models that are trained and evaluated in at the same time to utilize the various mannerisms of each models. Thus, the ensemble absorbs the different algorithms' benefits and relents good performance than using a single one. In other words, it is better to include different classifiers.

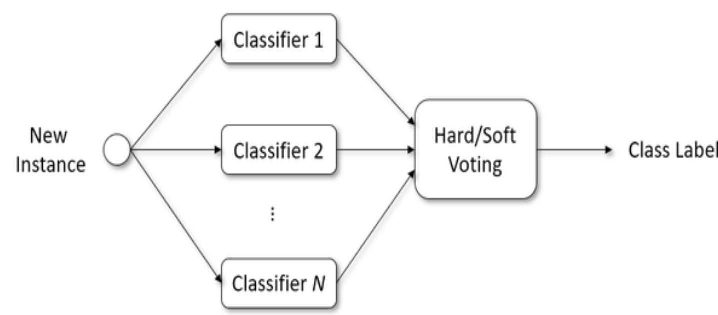

Fig. 3. Structure of voting classifier

Fig. 3 shows the structure or idea of an ensemble of different classifiers. There can use a number of classifiers and the voting classifier is able to wrap all of them. Then the data is given to the voting classifier and training is done, in a sense, each classifier is individually trained the data and output is produced. However, the voting type will decide who will win based on the type of voting was used. It may be hard voting or soft voting. Then the output is predicted as the winning class label is projected [20] – [29].

A. Evaluation of Models

In this step, the confusion matrix and ROC curve-based evaluations are used for analyzing the performance of the created model. Confusion Matrix is a metric to measure a classification model's performance and quality on validation data in machine learning. The table contains TP (True Positive) means the examined value is positive and is predicted as positive, TN (True Negative) means the examined value is negative and FP (False Positive) means the examined value is positive but is predicted as negative, and FN (False Negative) means the examined value is negative, but it is predicted as positive. ROC or Receiver Operating Characteristics Curve is another well-known metric for analyzing and evaluating a classification model's performance. There are two variables, True Positive Rate (TPR) and False Positive Rate (FPR), which are plotted by ROC curve. Area under ROC curve will indicate the quality of a prediction model in machine learning.

IV. IMPLEMENTATION AND EVALUATION

The system is implemented in Python3 available in Colab platform. The proposed model is implemented by selecting the available data set for churn prediction of the banking sector. So importing a bank churn modeling dataset from Kaggle with 14 features and 10,000 records. The data set id preprocessed by removing null values and removing irrelevant features. Then the next step is exploring data visualization using python's Plotly along with basic visualization packages.

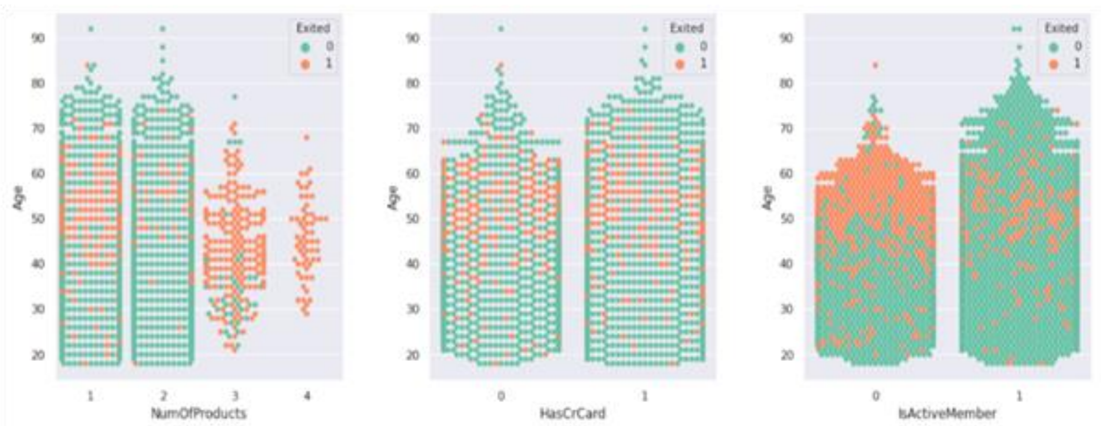

Fig. 4. Swarm Plot of churn and non-churn customers.

Fig. 4 shows the swarm plot using matplotlib. The swarm plot is a type of scatter plot, but helps in visualizing different categorical variables. This gives a better representation of the distribution of values, which implies a feature named a number of products with 3 and 4 have with age between 40-70 are more chance to churn and feature age between 50-70 are mostly not active members and chance to churn.

Fig.5 is the scatter plot of 40 to 70 years old customers, and are higher chances to churn and Customer with Credit Score less than 400 are higher chances to churn and also balance below 50000 are more chance to churn. Fig.6 shows the counts plot of age distribution. It will provide a great understanding about the age distribution over the data space.

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 7, DOI:10.29027/IJIRASE.v4.i7.2021.836-840,January 2021

Fig. 5. Scatter plot of balance and credit score against age

Plotly is available to put up on top of visualization library D3.js, HTML, and CSS and is an outstanding tool for data visualization. Plotly also provides plots in different formats based on user demand. In order to display the plots, we need require.js, it is a JavaScript file and in sometimes for different browser use it act as a module loader. RequireJS is a modular script loader that will improve the rate and standard of code. Plotly gives more user-friendly and interactive plots. Fig.5 shows Plotly scatter plot for an age-based credit score of males and females and when the mouse point hovers on the plot it will show all the details related to that also, it gives options for zoom in, zoom out, automatic scaling, etc.

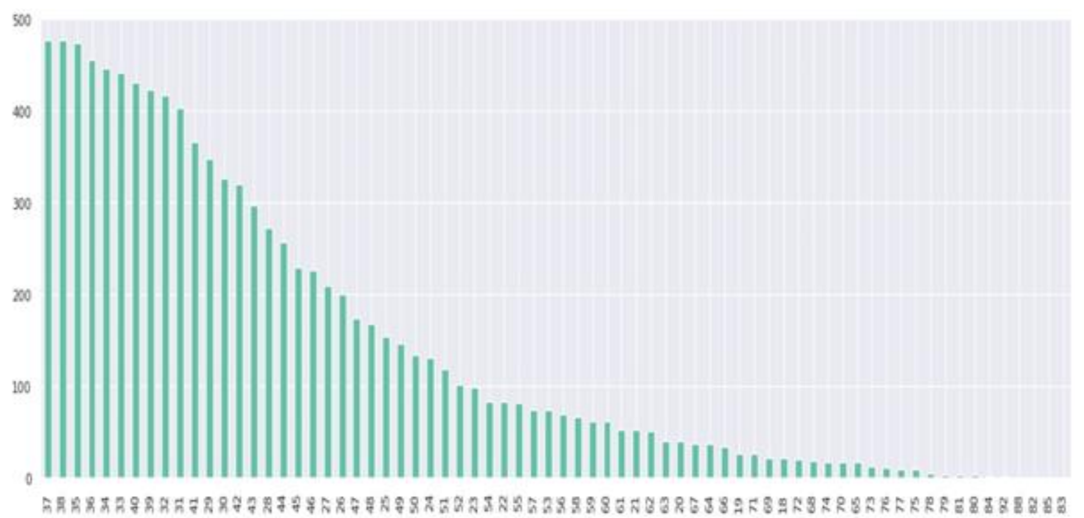

Fig. 6. Counts plot for customer's age.

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 7, DOI:10.29027/IJIRASE.v4.i7.2021.836-840,January 2021

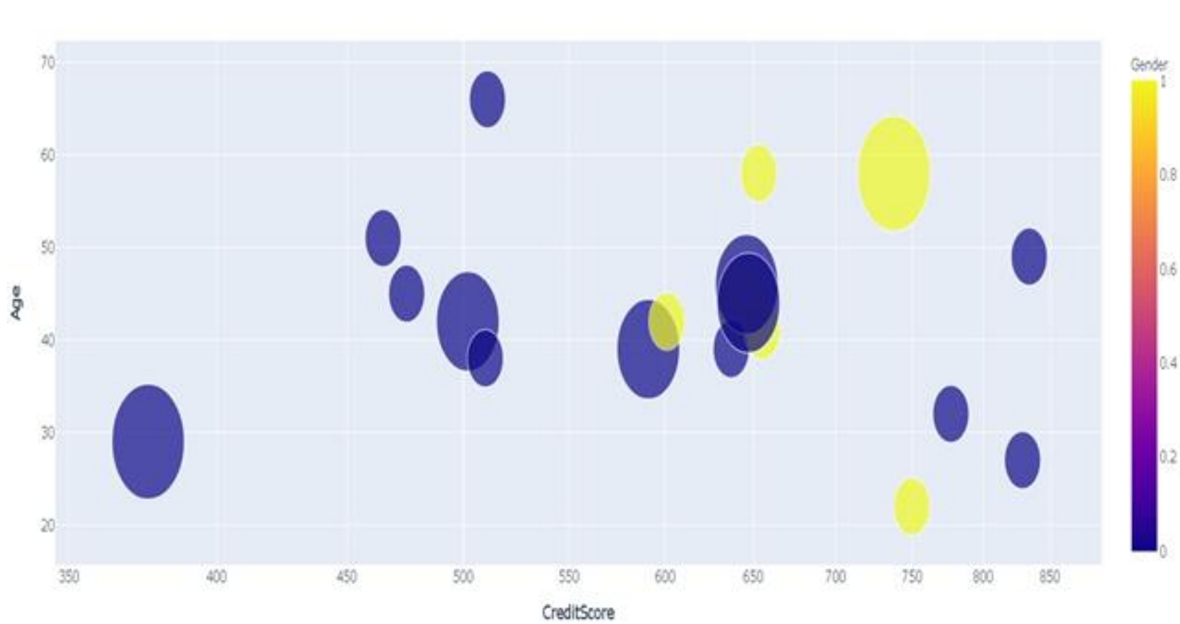

Fig. 7. Scatter plot for an age-based credit score of male and female using Plotly

Fig.7 shows the scatterplot for the age-based credit score of males and females. It is created using Plotly and which will provide interactive plots. There is an option for a special view of a selected portion or a detailed view specific area. It also provides a snapshot facility. The scale of the graph can be changed based on the scaling option.

Fig. 8. Box plot for Credit score distribution by geography.

Fig.8 and Fig.9 shows interactive box plots for credit score distribution by geography and number of product distribution using Plotly, respectively. Fig.10 shows the Correlation Matrix of the features.

After understanding the trends and distribution of data, it is time to prepare the data for modeling by encoding the categorical values using one-hot encoding, handling outliers by detecting it using box plot. To understand the most important feature, we use the Random Forest model's feature importance property and calculated the importance score for each feature.

To make the values of features in a fixed range between 0 and 1, normalization technique such as min max scaler is used. After completing the data preparation, the next step is to create an Ensemble model, so splitting the entire data set into a training set and testing set with a proportion of 80% and 20%, respectively. For ensemble modeling, the classification models such as Logistic Regression, Support Vector Machine, Decision Tree, Ada Boost, Gradient Boost, Cat Boost and Random Forest are used.

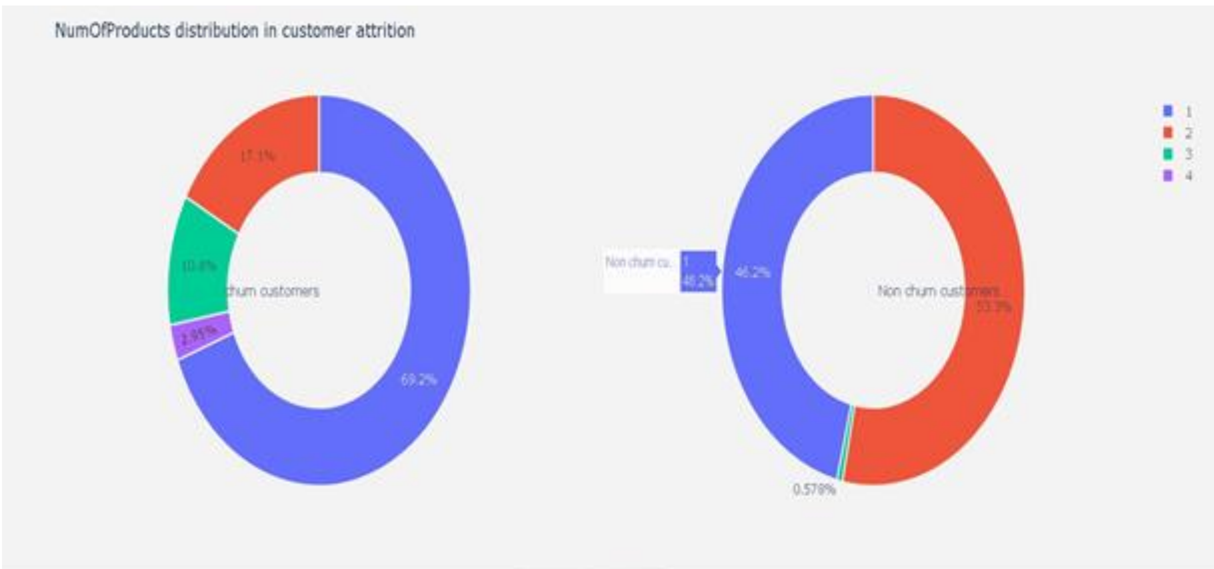

Fig. 9. No of product distribution of churn and non-churn customers

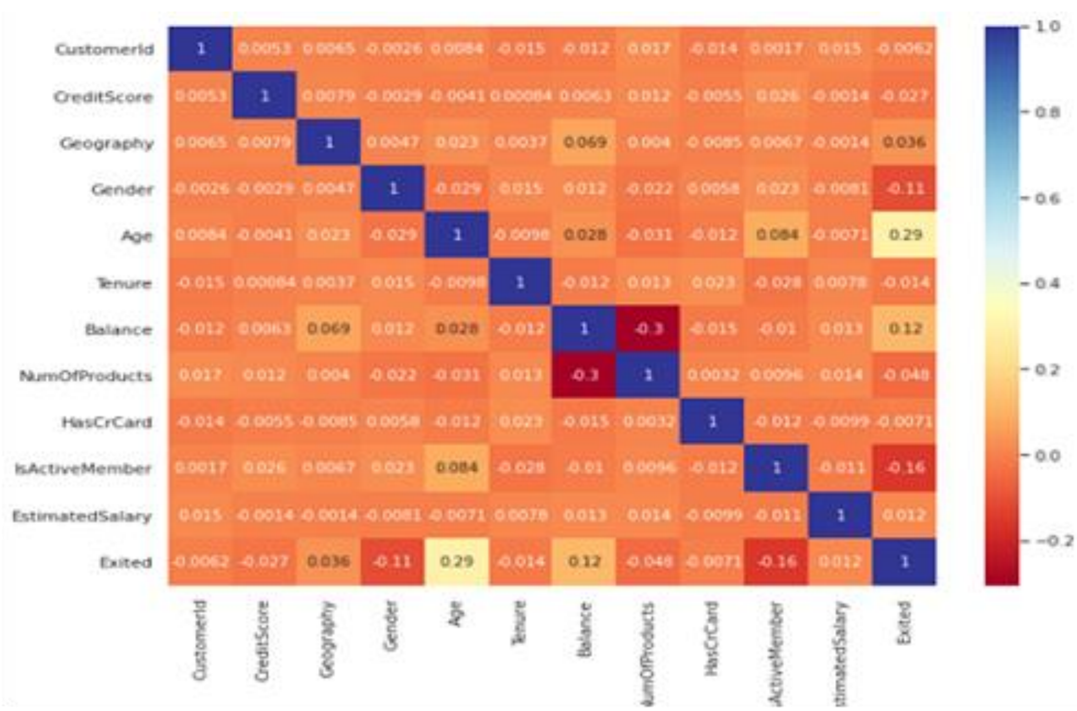

Fig. 10. Correlation Matrix of features

ISSN(Online) : *2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 7, DOI:10.29027/IJIRASE.v4.i7.2021.836-840,January 2021

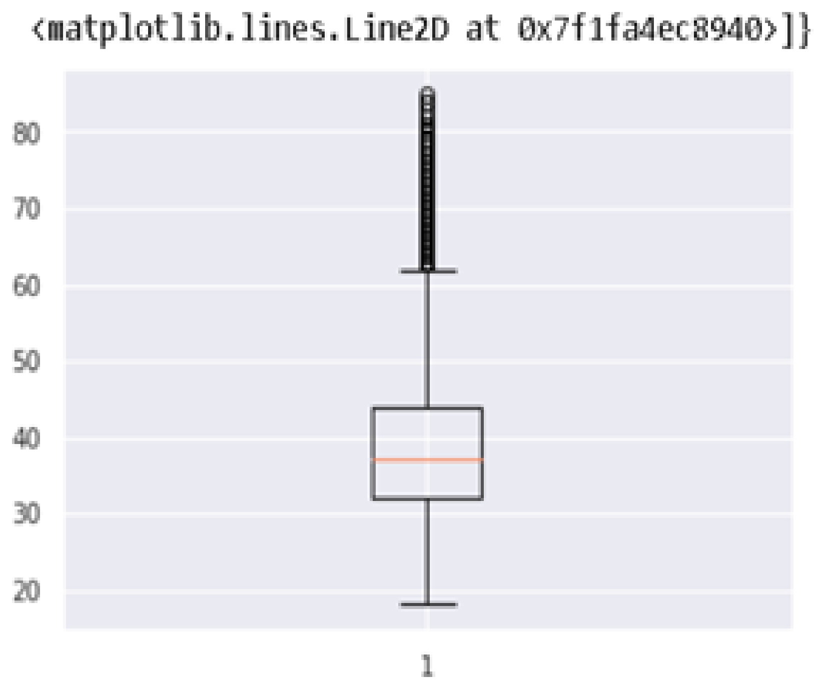

Fig. 11. Box plot for age feature

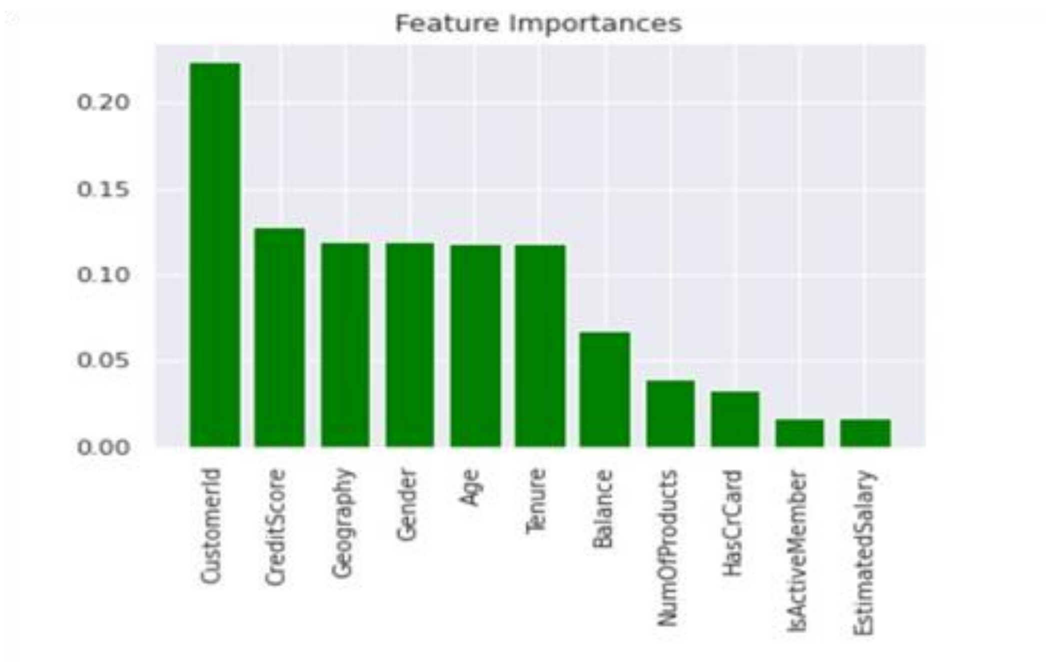

Fig. 12. Feature Importance of data set

Fig. 11 shows the Tukey boxplot of age feature, giving the detailed form of that feature by providing its minimum value, first quartile (Q1), median, third quartile (Q3) maximum. To handle outliers, there is a need of detect them. So boxplot will help to detect these outliers and after detecting the outlier, it can be removed by replacing with the average value. Fig. 12 shows the feature importance

property of the features in the data set. The feature importance can be calculated based on scoring based

ISSN(Online) : *2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 4, Issue 7, DOI:10.29027/IJIRASE.v4.i7.2021.836-840,January 2021

system. In this work, the feature importance property of a model is used, which will give a score for each and every data field. The highest scored feature will have the highest importance. Fig. 13 shows the feature importance score table

1)	CustomerId	0.223568
2)	CreditScore	0.127584
3)	Geography	0.119597
4)	Gender	0.119493
5)	Age	0.118488
6)	Tenure	0.117995
7)	Balance	0.067782
8)	NumOfProducts	0.039489
9)	HasCrCard	0.033452
10)	IsActiveMember	0.016279
11)	EstimatedSalary	0.016273

Fig. 13. Feature importance score

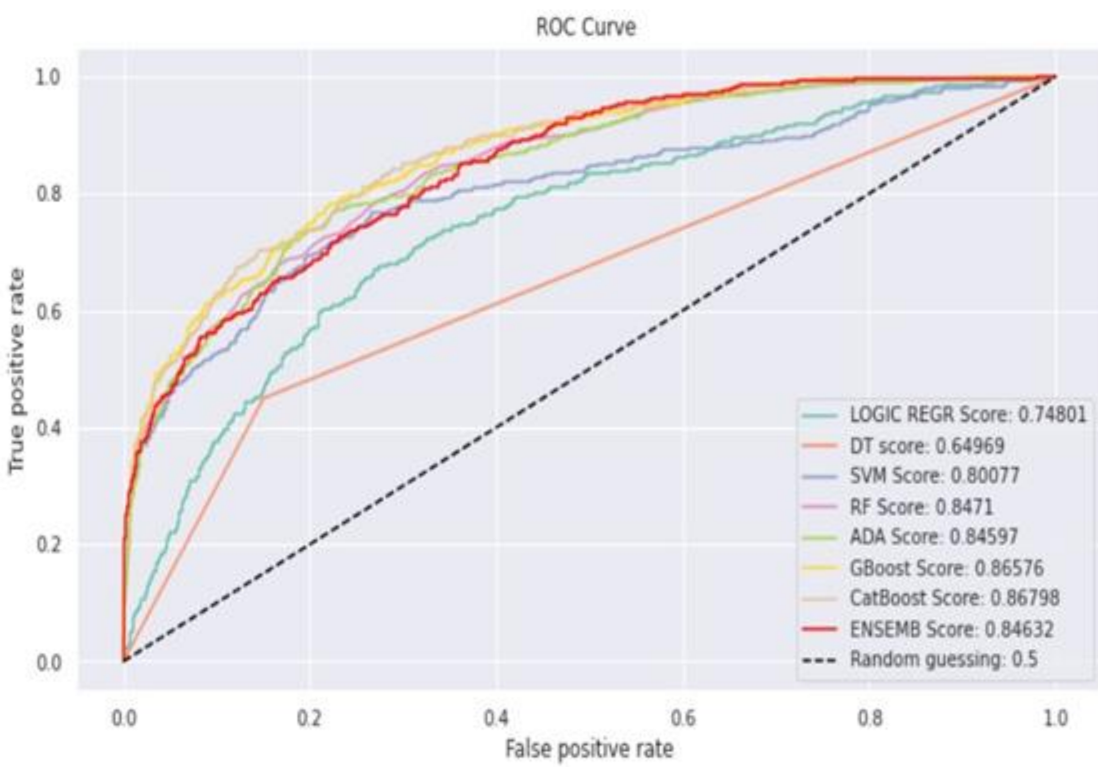

Fig. 14. ROC curves of individual models and Ensemble model

TABLE I
ACCURACY AND AUC-ROC OF DIFFERENT MODELS WITH ENSEMBLE MODEL

Model	Accuracy	AUC - ROC
Logistic Regression	80	0.72
Decision Tree	78	0.67
SVM	84	0.77
Random Forest	85	0.83
ADA Boost	84	0.83
Gradient Boost	85	0.84
Cat Boost	85	0.85
Ensemble (Voting Classifier)		**0.83**

Train each model individually using a training set, then combine each model's predictions using an ensemble classifier such as a voting Classifier. Our model uses the soft voting method and selects the majority voting class as output prediction. Each model is individually trained using training data and combines the models using the special classifier. After training of voting classifier, it should validate or test with the testing set.

```
Test Result:
============================================
Accuracy: 0.85

Classification Report:
        Precision: 0.7248908296943232
        Recall Score: 0.4058679706601467
        F1 score: 0.5203761755485893

Confusion Matrix:
 [[1528   63]
 [ 243  166]]
```

Fig. 15. Test result of ensemble model

Table.1 shows a comparison of individual model accuracies and ensemble model accuracy. From the table, it is clear to imply that the ensemble model selects the class with major voting. Hence the overall accuracy of the ensemble model 85%. Fig. 15 shows the test result of the ensemble model and Fig. 16 shows the comparison of the individual model with the ensemble model.

```
Accuracy: 0.80 [Logistic Regression]
ROC_AUC: 0.72
Accuracy: 0.78 [Decision Tree]
ROC_AUC: 0.67
Accuracy: 0.84 [SVM]
ROC_AUC: 0.77
Accuracy: 0.85 [RF]
ROC_AUC: 0.83
Accuracy: 0.84 [ADA]
ROC_AUC: 0.83
Accuracy: 0.85 [GBoost]
ROC_AUC: 0.84
Accuracy: 0.85 [CatBoost]
ROC_AUC: 0.85
Accuracy: 0.85 [Ensemble]
ROC_AUC: 0.83
```

Fig. 16. Test result comparison of individual models and ensemble model

To evaluate the system, we are using a confusion matrix and AUC - ROC curves. A confusion matrix is a summary of the prediction results of the classification problem. The classification Rate or Accuracy of the system can be evaluated from the confusion matrix. In AUC - ROC curves, the two variables, True Positive Rate (TPR) and False Positive Rate (FPR), are used to plot. The quality of performance of a classification model is measured based on the area under the curve. Fig.14 shows the ROC curves of individual models and the Ensemble model. The area under ROC of the ensemble is some average of the largest AUC-ROC accuracy values or AUC-ROC of majority voted models. The area under highlighted (red) color curve shows the performance efficiency of the ensemble algorithm in the figure.

I. CONCLUSION AND FUTURE WORK

In our system, the churn data of the banking sector is used to predict whether the customer is churn or not using the voting classifier of the ensemble method. One of the most powerful techniques for ensemble methods is Voting classifier that is in machine learning for a classification problem; it is really good to consider the results of some other models before taking the actual one. The ensemble's main advantage is that if the data is dynamic, then the accuracy of different learning algorithms may vary. Using the ensemble technique will consider the predictions of collection of algorithms and make decisions based on that. Also visualization of the dataset will give detailed information about the data distribution which can be used for feature engineering. For that, we are using Python’s Plotly library and it will generate more interactive plots. Thus customer loyalty or churn can be effectively find using ensemble modelling. However, this work can be enhanced by integrating deep learning models and combination of those models. Also in the case of evaluating, it can be extended by using other performance metrics.

ACKNOWLEDGMENT

I would like to acknowledge the Computer Science and Engineering Department of Rajiv Gandhi Institute of Technology's contribution and support.

REFERENCES

[1] Sabbeh, Sahar F, “Machine-learning techniques for customer retention: A comparative study”, International Journal of Advanced Computer Science and Applications 9.2 (2018).

[2] Afaq Alam Khan, Sanjay Jamwal, and M.M.Sepehri , “Applying data mining to customer churn prediction in an Internet Service Provider”, International Journal of Computer Applications Volume 9No.7,2010.

[3] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, “ Analysis of customer churn prediction in telecom industry using decision trees and logistic regression” , 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, 2016, pp. 1-4.

[4] L. Wu and M. Li, “Applying the CG-logistic Regression Method to Predict the Customer Churn Problem”, 2018 5th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS), Toronto, ON, 2018, pp. 1-5.

[5] S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra and S. Kumar, “Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron(MLP)”, 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Bangalore, India, 2018, pp. 436-438.

[6] Islam, Md, and Md Habib, “Data mining approach to predict prospective business sectors for lending in retail banking using decision tree”, arXiv preprint arXiv.1504.02018 (2015).

[7] Kumar, G. Ravi, K. Tirupathaiah, and B. Krishna Reddy, “Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques”, (2019).

[8] Kumar, A. Saran, and D. Chandrakala, “ An Optimal Churn Prediction Model using Support Vector Machine with Adaboost”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2.1 (2017): 225-230.

[9] Mahajan, Deepika, and Rakesh Gangwar, "Improved Customer Churn Behaviour by Using SVM", International Journal of Engineering and Technology (2017): 2395-0072.

[10] Kumar, Saran, S. Viswanandhne, and S. Balakrishnan, " Optimal Customer Churn Prediction System using Boosted Support Vector Machine", International Journal of Pure and Applied Mathematics 119.12 (2018) 1217-1231.

[11] ApurvaSree, G., Ashika, S., Karthi, S., Sathesh, V., Shankar, M.,Pamina, J. "Churn prediction in Telecom using classification algorithms" , International Journal of Scientific Research and Engineering Development 5 (2019) 19-28.

[12] Prajapati, Darshana, and Ratnesh Kumar Dubey . "Analysis Of Customer Churn Prediction In Telecom Sector Using Random Forest."

[13] Jaisakthi, S. M., Gayathri, N., Uma, K., Vijayarajan, V, "Customer Churn Prediction Using Stochastic Gradient Boosting Technique", Journal of Computational and Theoretical Nanoscience 15, no. 6-7 (2018) 2410-2414.

[14] A. Mishra and U. S. Reddy, " A comparative study of customer churn prediction in telecom industry using ensemble based classifiers", 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 721-725.

[15] Adnan Idris, Asifullah Khan, " Prediction System for Telecom using FilterWrapper and Ensemble Classification", The Computer Journal, Volume 60, Issue 3, March 2017, Pages 410430,

[16] Vijaya, J., and E. Sivasankar., "Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector.", Computing 100.8 (2018): 839-860.

[17] Wang, Xing, Khang Nguyen, and Binh P. Nguyen. "Churn Prediction using Ensemble Learning.", Proceedings of the 4th International Conference on Machine Learning and Soft Computing. 2020

[18] Garca, David L., Angela Nebot, and V. Alfredo, "Visualizing pay-per-view television customers churn using cartograms and flow maps" , ESANN 2016 Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2016.

[19] Mease, JonBringing ipywidgets Support to plotly. py, (2018).

[20] Dror simon et.al, " Auto annotation quality prediction for semi supervised learning with ensembles", 2009.

[21] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni. Hierarchical classification: Combining Bayes with SVM. In Proc. of the 23rd Int. Conf. on Machine Learning, pages 177–184. ACM Press, 2006

[22] AA.VV. Special Issue on Applications of Ensemble Methods, volume 9 (1) of Information Fusion Journal. Elsevier, 2008.

[23] Adnan O. M. Abuassba, Dezheng Zhang,, Xiong Luo, Ahmad Shaheryar, and Hazrat Ali, "Improving Classification Performance through an Advanced Ensemble Based Heterogeneous Extreme Learning Machines", 2017.

[24] Dietterich, Thomas G. "Ensemble methods in machine learning." International workshop on multiple classifier systems. Springer, Berlin, Heidelberg, 2000.

[25] Kazienko, Przemysław, Edwin Lughofer, and Bogdan Trawiński. "Hybrid and ensemble methods in machine learning J. UCS special issue." J Univers Comput Sci 19.4 (2013): 457-461.

[26] Sagi, Omer, and Lior Rokach. "Ensemble learning: A survey." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.4 (2018): e1249.

[27] Ardabili, Sina, Amir Mosavi, and Annamária R. Várkonyi-Kóczy. "Advances in machine learning modeling reviewing hybrid and ensemble methods." International Conference on Global Research and Education. Springer, Cham, 2019.

[28] Dong, Yan-Shi, and Ke-Song Han. "A comparison of several ensemble methods for text categorization." IEEE International Conference onServices Computing, 2004.(SCC 2004). Proceedings. 2004. IEEE, 2004.

[29] Valentini, Giorgio, and Francesco Masulli. "Ensembles of learning machines." Italian workshop on neural nets. Springer, Berlin, Heidelberg, 200.