Churn Prediction in Telecommunication using Machine Learning

Kriti Mishra*, Dr. Rinkle Rani
Computer Science & Engineering Department, Thapar University, Patiala, Punjab, India
(e-mail: kritimishra94@gmail.com* and raggarwal@thapar.edu)

***Abstract*—The telecommunication industry always has a tough competition with its competitors to retain customers, and therefore has become one of the research sectors in machine learning and data mining. Since the customers' churn behavior is to be monitored closely and efficiently it requires for a methodical churn prediction model to monitor the customers' churn. The main setbacks in achieving the desired performances in a classifier are the enormous datasets, large feature space and imbalanced class distribution. In this work, we explore the implication of Synthetic Minority Over-sampling TEchnique (SMOTE) to reduce the imbalance in data in collaboration with different feature reduction techniques such as Co-relation feature extraction, Gain ratio, Information gain and OneR feature evaluation method. Classification and Regression Trees(CART), Bagged CART and Partial Decision Trees(PART) classifiers are trained to analyze the performance on balanced and reduced feature space dataset. Prediction performance of the classifiers is evaluated through measures such as Area Under the Curve(AUC), sensitivity and specificity. Finally, it is concluded through simulations that our proposed method based on SMOTE, co-relation, and ensembling performs well for predicting churners as against simply applying learners on the unrefined dataset. Therefore, this methodology can be helpful for the telecommunication industry to predict churn.**

***Index Terms*—Data Mining, Machine Learning, Churn Prediction, Data balancing, Decision tree, Ensembling**

I. Introduction

In past few years, there has been an explosion of data. But in this large amount of data, it is important to extract information which is hidden in raw data. Meaningful information can be extracted from this raw data to make appropriate decisions and come to conclusions that help an industry to grow. Data mining and machine learning has been used to extract this hidden information from large repositories of data. Well-known examples where data mining and machine learning have already been explored are, market basket analysis in the retail sector, breast-cancer detection in the biomedical sector and credit scoring in the financial sector . This paper however focuses on the use of data mining to predict customer churn. The telecommunication industry faces a challenge to retain customers. The customers may switch over to other networks for varying reasons such as, better services, better feasibility of other companies' pricing plan, better call quality, cutomers' billing problems etc. In telecommunication industry, there is this problem of customers churning out and switching to other competitors which may be due to a number of reasons. Therefore, we need a good classifier to correctly predict customers' churn tendencies. Hence, in a telecommunication industry, customers' churn prediction is an important requirement of the customer relationship management . The effectiveness of a churn prediction model depends on learning achieved from the dataset provided. An appropriately preprocessed dataset gives high performance to the classifiers as it removes all unnecessary and redundant data in the dataset. Telecommunication companies collect information about their customers not all of which is relevant. Usually, this data has large feature space and imbalanced class distribution as the number of churners is far less compared to non-churners. The imbalanced dataset causes faulty and weak learning by a classifier. Therefore, proper preprocessing is required to remove any redundant or useless features which do not have any relation to the target class. A popular feature selection technique is PCA(Principal Component Analysis) [1] which linearly operates on data to select the most relevant features. Others are Gain Ratio, Information Gain, OneR and Co-relation based techniques. Similarly, some common sampling methods are Random Oversampling (ROS) and Random Undersampling (RUS) [2]. In Random Oversampling, random instances from the minority class are simply replicated. In Random Undersampling, random instances from the majority class are discarded. These methods are not effective and show inconsistency and varying performances due to the random act of duplicating and discarding instances. Due to this random selection, ROS is prone to overfitting and RUS may discard some useful instances. There are various models in machine learning that use mathematics and statistics to find patterns in data and classify them. Broadly there are 4 categories wherein the models can be classified. They are Trees and Rules(CART, J48, PART) [3], Ensemble of Trees(C5.0, Bagged CART, Random Forest, Stochastic Gradient Boosting) [4], Linear(Linear Discriminant Analysis, Logistic Regression) [5], Non-Linear(Neural Network, Support Vector Machine, k-Nearest Neighbor, Naive Bayes) [6]. To improve the performances of classifiers, ensemble approach can be applied which is a combination of 2 or more classifiers to obtain the best of all the classifiers.

II. Related Work

Adnan Idris et al (2012) [7] discusses about churn prediction of telecommunication using Random forest and KNN method. In order to handle the imbalanced data distribution, Particle Swarm Optimization is used for under-sampling the

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

dataset, along with feature reduction techniques like PCA, F-score, Fisher's ratio and Minimum Redundancy Maximum Relevance. Performance of the modeling methods is evaluated using area under the curve, sensitivity and specificity methods. These simulations were found to be successful in positive prediction of churn.

Asifullah Khan et al [8] worked on churn prediction for the telecom using ensemble model. Minimum redundancy and maximum relevance techniques were used for discriminative feature selection which led to enhanced feature label association and reduced feature set. Different base classifiers ensemble is applied as a predictor technique. Random forest, Rotation Forest and KNN classifiers were used for final predictions using majority voting. Results evaluated using area under the curve, sensitivity, specificity and Q-statistic based measures predicted that the churn model prediction is very efficient.

Yeon Soo Lee et al (2012) [9] have also worked on the churn prediction for customer retention using Genetic algorithm approach. For each class, various programs were generated using Adaboost method. These programs were used for predictions using the higher output, from weighted sum of the outputs of programs per class. A 10-fold cross validation technique was used to check the prediction accuracy and the area under the curve score of 0.89 was found.

Javed Basiri et al (2010) [10] used OWA (Ordered Weighted Average) to fuse the output of each learned classifier to introduce a hybrid approach to improve the accuracy of the results obtained. In this study, bagging and boosting is implemented to train the classifiers and also LOLIMOT algorithm is learned using different number of important features. The results generated showed that the approach was good enough than some well-known classifiers.

III. Materials and Methods

In telecommunication datasets, there is a problem of skewed data distribution. Due to this, classification algorithms perform unsatisfactorily in predicting customer churn. Therefore, in this methodology, we handle these problems and also implement suitable ensemble approach [11] to build a classifier that shows better performance than all the three individual models . The block diagram shown in Fig.1 highlights various steps involved in this approach. The dataset is initially preprocessed to tackle the missing values in the dataset. SMOTE based sampling method [12] is used to determine its effectiveness in enhancing the prediction results. Some feature extraction algorithms such as Information Gain, Gain Ratio, Co-relation and OneR attribute evaluation are employed individually and their respective results on classification is assessed. In this work, CART, Bagged CART, PART [13] are employed to study their individual efficiency and Adaboost [14] is applied to the results given by these algorithms to train the newly blended dataset that has been achieved by incorporating the results of base algorithms to act as new features. Sampling is employed in combination with a few feature selection techniques and classifiers are used to grade the results of data balancing and feature extraction techniques using specificity, sensitivity and AUC. The approach based on SMOTE, Co-relation feature extraction and Adaboost classifier gives the best results.

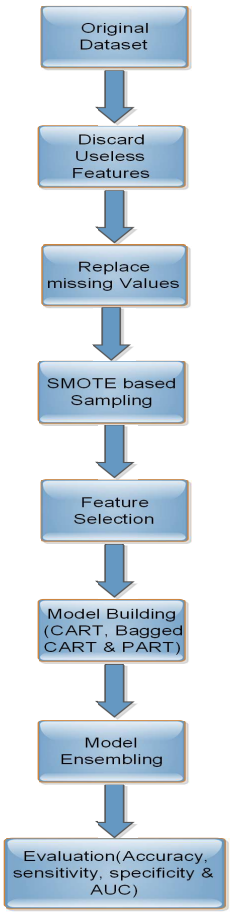

Fig. 1. Basic Block Diagram of proposed approach

A. Dataset

The dataset has been taken from IBM Watson Analytics. The dataset in this work has 7,043 instances and 21 attributes. The dataset is composed of 4 numerical and 17 nominal attributes. It has 1,869 records of minority and 5,174 records of majority class that amounts to 26.5% contribution of minority class in the entire dataset.

B. Dataset preprocessing

We have preprocessed the dataset so that useless features are discarded and also minority class of churners is not overwhelmed by the majority class of non-churners. The preprocessing phase involves removing useless features. We further process the dataset for sampling using SMOTE technique. Thereafter, Co-relation, Gain Ratio, Information Gain and OneR feature selection methodologies are used for selecting the most relevant features before applying classification models.

C. SMOTE sampling technique

Classification upon class-imbalanced dataset is prone to be in favor of the majority class as the minority class samples are far less to represent a class adequately. So, undersampling and oversampling techniques have been developed to balance the classes. Undersampling randomly discards some of the samples of the majority class while random oversampling replicates samples from the minority class. Random OverSampling(ROS) causes overfitting due to replication. Therefore, it is not a very helpful technique. SMOTE technique [15] was developed to overcome the drawback of Random OverSampling(ROS) technique. In our dataset the dimensionality of the dataset is not too high instead number of instances greatly outnumber the feature space but SMOTE analysis is done to analyze its effect on our dataset.

D. Co-relation Feature Selection

The co-relation based feature selection method [16] is a heuristic for measuring the worth of individual features for predicting the class label. There are mainly 2 techniques to find co-relation of an attribute with the class label. They are Pearson's Co-relation Coefficient and Spearman's Co-relation Coefficient. The attributes which show higher co-relation with the class label are selected and those with lower co-relation are discarded. In co-relation feature selection, features are rated according to the importance of features.

E. Information Gain Attribute Selection

IG(Information Gain) [17] measures the entropy of a system which means the degree of disorder in a system. It is a term-goodness criterion in the field of machine learning. Therefore, the entropy of a subset is a fundamental calculation to compute IG.

F. Gain Ratio Attribute Selection

A decision tree is a simple structure of nodes and edges wherein nodes represent test cases on attributes, the edges represent the various answers possible on the test nodes while the terminal nodes represent decision outcomes to these test cases. The IG measure is used to select attributes for the terminal nodes of the decision tree. The IG measure has a limitation that it selects the attributes with large number of values. This limitation of the IG method is overcome by the Gain Ratio method [18]. ID3 is a decision tree based on IG. C4.5 is an enhancement of ID3 that uses gain ratio which is an extension of IG measure aimed at improving ID3.

G. OneR Attribute Selection

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, it constructs a frequency table for each predictor against the target. It has been seen that "one rule" produces rules that are easy to interpret and also these rules do not exhibit very low classification accuracy as compared to other complex classification algorithms

H. Decision tree based classification

The CART programs [19] build classification or regression models of a very general structure using a two stage procedure; the resulting models can be represented as binary trees. The tree is built by the following process: first the single variable is found which best splits the data into two groups. The data is separated, and then this process is applied separately to each sub-group, and so on recursively until the subgroups either reach a minimum size or until no improvement can be made. The resultant model is, with a certainty, too complex, and the question arises as it does with all stepwise procedures of when to stop. The second stage of the procedure consists of using cross-validation to trim back the full tree.

I. Partial Tree based classification

Partial Decision Trees(PART) [20] are decision trees that prune the decision tree on their own. They are an improvisation of C4.5 algorithm. Unlike C4.5, they do not have to perform global optimization to produce appropriate rules.

J. Bagged tree based classification

Bagged tree classification [21] is a type of ensemble machine learning algorithm called Bootstrap Aggregation or Bagging. Bagging is a simple yet powerful ensemble method. An ensemble method is a technique that combines the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model. The bagging technique has been invented to reduce the variance of those classifiers which have a high variance. Algorithms such as Classification and Regression Trees(CART) have been known to have high variance and low bias. Decision trees are specific to the data on which it is trained which means that if they are trained on some other data then they may result in a completely different decision tree and in turn different predictions. Bagging is the application of the Bootstrap procedure to a high-variance machine learning algorithm, typically decision trees. Assuming we have a sample dataset of 1000 instances (x) and we are using the CART algorithm. Bagging of the CART algorithm would work as follows.

- Create many (e.g. 100) random sub-samples of our dataset with replacement.
- Train a CART model on each sample.
- Given a new dataset, calculate the average prediction from each model.

Just like the decision trees themselves, bagging can be used for classification and regression problems.

K. Boosted Classification trees

Classification trees are increasingly being used in a wide range of applications, especially binary classification systems to predict a binary outcome. A limitation of classification trees is their limited predictive accuracy. In data mining, boosting was developed in order to improve upon the accuracy of classification trees. Boosting [22] grows classification trees iteratively in a sequence of re-weighted datasets. In a given iteration, the instances that were mis-classified in the previous

iteration are assigned more weight as compared to those that were correctly classified. Classifications from each of the classification trees in the sequence are combined through a weighted majority vote to produce a final classification.

L. Performance measures

The performance criteria used in this study to evaluate performances of classifiers on processed dataset are Specificity, sensitivity and Area Under the Curve(AUC) [23]. AUC is a helpful measure for evaluating comprehending, visualizing and understanding the performances of the techniques applied. The AUC measures the area under the curve in the graphical plot of sensitivity vs specificity [24] for a binary classifier system. AUC is used to assess the prediction power of CART, Bagged CART and PART on processed dataset.

IV. Proposed Approach

A number of combinations of sampling, feature extraction and classifiers have been used and it is observed that SMOTE based sampling along with co-relation based feature extraction and ensemble approach outputs best prediction performance. Our proposed method effectively uses a SMOTE based sampling method, which appropriately oversamples the dataset. This improves the CART, Bagged CART, PART as well as ensemble performances upon testing the dataset on 10-fold cross validation. The SMOTE technique works by synthetically creating new minority class samples using an algorithm. This is better than conventional oversampling as it tends to overfit the train set. The newly produced instances are combined with the original dataset to produce a balanced dataset. Finally an optimal dataset is produced that shows better classification score for CART, Bagged CART and PART in terms of AUC. After the process of model building, ensembling is applied on the dataset by combining the results of all the three algorithms.

V. Results and discussion

Initially, the unprocessed dataset is used to deploy the various combinations of feature extraction and classification methods. Then, similar experimentation is conducted upon the processed dataset. The performance of the various techniques applied into the study are compared as to performance they display on processed as well as unprocessed dataset. The testing is done on 10-fold cross validation and performance measures used are AUC, sensitivity and specificity.

A. Performance analysis after basic preprocessing

Initially, CART, Bagged CART and PART are implemented on original dataset without applying any sampling or feature reduction strategy. Table 1 shows, all three classifiers show poor performance in terms of specificity, sensitivity and AUC. The original unprocessed dataset contains 7,074 records and 21 attributes. In basic preprocessing the only optimization done is to discard useless features and instances with missing values have been treated.

TABLE I
The performance of CART, Bagged CART and PART on original dataset

	CART	Bagged CART	PART
Accuracy	0.7643	0.7619	0.7686
Sensitivity	0.275	0.441	0.507
Specificity	0.937	0.88	0.863
AUC	0.6063	0.6603	0.6851

B. Performance based on SMOTE sampling technique

SMOTE (Synthetic Minority Over-Sampling Technique) is one of the most adopted approaches due to its simplicity and effectiveness. It is a combination of oversampling and undersampling, but the oversampling approach is not by replicating minority class but constructing new minority class data instance via an algorithm. In traditional oversampling, minority class is replicated exactly. In SMOTE, new minority instances are constructed in this way: For each minority class instance c

- neighbours=Get KNN(k)
- Randomly pick one from neighbours
- Create a new minority class instance r using c's feature vector and the feature vector's difference of n and c multiplied by a random number i.e., r.feats = c.feats + (c.feats - n.feats) * rand(0,1)

The intuition behind the construction of this algorithm is that oversampling causes overfitting [25] because repeated instances cause the decision boundary to tighten. SMOTE instead, creates similar examples instead. To the machine learning algorithm, these new constructed instances are not exact copies and thus softens the decision boundary. As a result, the classifier is more general and does not overfit. The dataset obtained through the process of SMOTE based balancing gives a better performance, as now the data is evenly distributed among the classes i.e., the churners and the non-churners which is bound to improve the classification score. This therefore extends enhanced learning to the base classifiers. All three classifiers show improvements in AUC as shown in Table II. CART, Bagged CART and PART achieve 0.6445, 0.6998 and 0.7139 AUC values respectively, which show that there definitely has been an improvement in performance due to the sampling technique used. The dataset obtained after the SMOTE based sampling evenly justifies the presence of both the classes. Now the minority class instances have not been suppressed by the majority class instances. Due to this an improved training level was possible by the used classifiers. This was possible because of KNN classifier [26] involved in the instance generation of SMOTE based sampling. Therefore, the SMOTE based sampling serves the idea of achieving balanced data for better performance.

C. Performance analysis after feature selection

It was observed that after feature selection, the performance of the individual models has improved substantially. The AUC of CART on sampled dataset is 0.6445 and that after applying

TABLE II
THE PERFORMANCE OF CART, BAGGED CART AND PART ON SAMPLED DATA

	CART	Bagged CART	PART
Accuracy	0.7661	0.7690	0.7701
Sensitivity	0.259	0.410	0.436
Specificity	0.940	0.891	0.923
AUC	0.6445	0.6998	0.7139

TABLE III
PERFORMANCE OF CART AFTER CO-RELATION FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7922	0.401	0.933	0.6581

co-relation is 0.6581. The AUC shows an improvement of 0.0136 on sampled data after applying Co-relation feature selection. The AUC increased by 0.0204 from 0.6998 to 0.7202 after applying co-relation on Bagged CART classifier. AUC improved from 0.7139 to 0.7243 after applying co-relation on sampled data for PART classifier. Overall, we observe that co-relation feature selection performs much better than Gain Ratio, Information Gain and OneR feature selection methods.

VI. CONCLUSIONS

This work validates the argument that adequate pre-processing and data balancing in case of imbalanced datasets are bound to improve the classification performances of the used classifiers. The SMOTE based classifier extends desired performance level to the classifiers measured in terms of AUC. Further, appropriate feature extraction strategies are employed to explore the power of discriminating features in the training of the classifiers and hinder their performance and lower the learning of the models. Finally, to further enhance the productivity of the learners, ensemble approach has been used, which works by combining the results of the base classifiers. The proposed approach is a promising contribution of SMOTE analysis, Co-relation based feature extraction and ensemble of classifiers with CART, Bagged CART and PART as the base classifiers. The AdaBoost classifier gives the best results in terms of prediction performance measures(AUC, sensitivity, specificity). Thus, the proposed approach can be understood as a viable solution for accurately predicting customer churn in telecommunication industry. The comparative analysis graph

TABLE IV
PERFORMANCE OF BAGGED CART AFTER CO-RELATION FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7709	0.465	0.882	0.7202

TABLE V
PERFORMANCE OF PART AFTER CO-RELATION FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7638	0.516	0.921	0.7243

TABLE VI
PERFORMANCE OF CART AFTER GAIN RATIO BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7492	0.205	0.950	0.6317

TABLE VII
PERFORMANCE OF BAGGED CART AFTER GAIN RATIO BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7473	0.243	0.929	0.7024

TABLE VIII
PERFORMANCE OF PART AFTER GAIN RATIO BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7738	0.519	0.866	0.6973

TABLE IX
PERFORMANCE OF CART AFTER ONER BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7553	0.266	0.932	0.6373

TABLE X
PERFORMANCE OF BAGGED CART AFTER ONER BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7515	0.185	0.954	0.6993

TABLE XI
PERFORMANCE OF PART AFTER ONER BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7686	0.503	0.868	0.7172

TABLE XII
PERFORMANCE OF CART AFTER IG BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7440	0.574	0.803	0.645

TABLE XIII
PERFORMANCE OF BAGGED CART AFTER IG BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7530	0.519	0.850	0.7181

TABLE XIV
PERFORMANCE OF PART AFTER IG BASED FEATURE SELECTION

Accuracy	Sensitivity	Specificity	AUC
0.7406	0.527	0.838	0.7035

of each technique with each of the classifiers and ensemble classifier is shown in Fig 2.

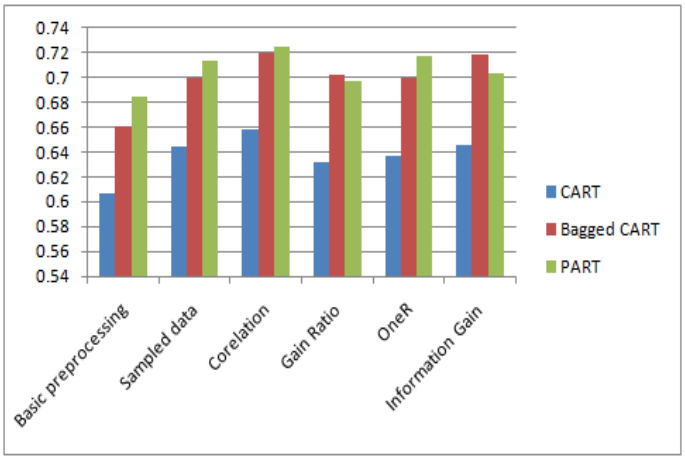

Fig. 2. Performance comparison of base classifiers based on AUC scores

REFERENCES

[1] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[2] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 853–867.

[3] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," 2007.

[4] M. Gashler, C. Giraud-Carrier, and T. Martinez, "Decision tree ensemble: Small heterogeneous is better than large homogeneous," in *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on*. IEEE, 2008, pp. 900–905.

[5] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 197–200, 1992.

[6] I. Rish, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22. IBM New York, 2001, pp. 41–46.

[7] A. Idris, M. Rizwan, and A. Khan, "Churn prediction in telecom using random forest and pso based data balancing in combination with various feature selection strategies," *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1808–1819, 2012.

[8] A. Idris, A. Khan, and Y. S. Lee, "Intelligent churn prediction in telecom: employing mrmr feature selection and rotboost based ensemble classification," *Applied intelligence*, vol. 39, no. 3, pp. 659–672, 2013.

[9] ——, "Genetic programming and adaboosting based churn prediction for telecom," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1328–1332.

[10] J. Basiri, F. Taghiyareh, and B. Moshiri, "A hybrid approach to predict churn," in *Services Computing Conference (APSCC), 2010 IEEE Asia-Pacific*. IEEE, 2010, pp. 485–491.

[11] T. G. Dietterich, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, pp. 110–125, 2002.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] R. L. Lawrence and A. Wright, "Rule-based classification systems using classification and regression tree (cart) analysis," *Photogrammetric engineering and remote sensing*, vol. 67, no. 10, pp. 1137–1142, 2001.

[14] T. G. Dietterich, "Machine-learning research," *AI magazine*, vol. 18, no. 4, p. 97, 1997.

[15] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: Improving prediction of the minority class in boosting," in *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, 2003, pp. 107–119.

[16] M. A. Hall, "Correlation-based feature selection of discrete and numeric class machine learning," 2000.

[17] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, vol. 97, 1997, pp. 412–420.

[18] A. G. Karegowda, A. Manjunath, and M. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[19] R. Kohavi and J. R. Quinlan, "Data mining tasks and methods: Classification: decision-tree discovery," in *Handbook of data mining and knowledge discovery*. Oxford University Press, Inc., 2002, pp. 267–276.

[20] H. Berger, D. Merkl, and M. Dittenbach, "Exploiting partial decision trees for feature subset selection in e-mail categorization," in *Proceedings of the 2006 ACM symposium on Applied computing*. ACM, 2006, pp. 1105–1109.

[21] A. M. Prasad, L. R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: bagging and random forests for ecological prediction," *Ecosystems*, vol. 9, no. 2, pp. 181–199, 2006.

[22] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, "Boosted decision trees as an alternative to artificial neural networks for particle identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2, pp. 577–584, 2005.

[23] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[24] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity." *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.

[25] D. M. Hawkins, "The problem of overfitting," *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1–12, 2004.

[26] T. Denoeux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE transactions on systems, man, and cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.