International Journal of Computational and Experimental Science and Engineering
(IJCESEN)
Vol. 8-No.3 (2022) pp. 77-84
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Research Article

The Effectiveness of Homogeneous Classifier Ensembles on Customer Churn Prediction in Banking, Insurance and Telecommunication Sectors

Zeynep Hilal KİLİMCİ*

Kocaeli University, Faculty of Technology, Department of Information Systems Engineering, 41001, Kocaeli-Turkey
***Corresponding Author Email**: zeynep.kilimci@kocaeli.edu.tr - **ORCID:** 0000-0003-1497-305X

Article Info:

DOI: 10.22399/ijcesen.1163929
Received : 18 August 2022
Accepted : 08 November 2022

Keywords

Churn Prediction
Ensemble Learning
Machine Learning
Random Forest
Customer Loyalty

Abstract:

The prediction of customer churn is a big challenging problem for companies in different sectors such as banking, telecommunication, and insurance. It is a crucial estimation for many businesses since obtaining new customers frequently costs more than holding present ones. For this reason, analysts and researchers are focus on to investigate reasons behind of customer churn analyzing behaviors of them. In this paper, an ensemble-based framework is proposed to predict the customer churn in various sectors, namely banking, insurance, and telecommunication. To demonstrate the effectiveness of proposed ensemble framework, k-NN, logistic regression, naïve Bayes, support vector machine, decision tree, random forest, multilayer perceptron algorithms are employed. Moreover, the effects of the inclusion of feature extraction process are investigated. Experiment results indicate that random forest algorithm is capable to predict churn customers with 89.93% of accuracy in banking, 95.90% of accuracy in telecommunication, and 77.53% of accuracy in insurance sectors when feature extraction procedure is carried out.

1. Introduction

In the present-day world, there is a sequence of telecom, bank, and insurance companies contending in the market to raise their customer portion. Moreover, customers have various alternatives in the shape of better quality and less costly getting service. The eventual objective of companies is to maximize their gain and subsist in a competitive market [1-2]. The churn of customers occurs when customers are not pleased with the service of any company. This concludes in service transmigration of customers who initiate replacing to other companies that operate in the same area. Because of this reason, customer satisfaction is staminal for the companies by preventing customer churn and retaining the valuable customers. [3-4].

From this point of view, companies make very serious investments to retain their customers. The companies have improved methods to determine and keep their clients since it is less costly than fascinating the new ones [3]. This is because of the expenses comprised in advertisements, workforce, and grants which can boost to nearly five to six times than holding present customers [5]. The necessity of retaining clients requires to advance a precise and high-performance approach for determining churn clients. Among the investments, there resides many data mining tools and machine learning approaches which can be employed to analyze this kind of data. The methods are evaluated to investigate the data and explore reasons behind customer churning. Moreover, the approaches can be utilized for the purpose of detecting customers that tend to churn.

Ensemble techniques utilize multiple classifiers to acquire a better success than a single learner. An ensemble classifier is a set of consolidated weak classifiers. During training stage, each classifier is trained respectively on a given training data set. An ensemble methodology is generally constructed an ensemble generation and fusion steps. In the first step, a diversified group of learners is generated from the training data set. In the second phase, the outputs of the classifiers are consolidated to acquire a final decision. The main approach in ensemble methodology is therefore to produce many classifiers and consubstantiate outputs of classifier such that the consolidation of learners boosts the

success of a base classifier [6-9]. If all the classifiers are constructed by utilizing the same technique, the ensemble system is named as homogeneous, otherwise it is entitled heterogeneous. Heterogeneous ensemble approach mainly employs more than one classifier to perform diversity [9-10]. In this work, we concentrate on the efficiency of homogenous classifier ensembles for forecasting customer churn in different sectors namely, banking, telecommunication, and insurance. For this purpose, proposed homogeneous ensemble technique, random forest, is compared with the traditional machine learning methods (k-NN, logistic regression, naïve Bayes, support vector machine, decision tree) and multilayer perceptron. Experiment results represent that the utilization of random forest method as a homogenous ensemble model outperforms conventional machine learning techniques in all sectors.
The rest of the research is designated as follows: Section 2 presents related works on the churn prediction. Section 3 presents framework and the techniques used in the research. Experiment results are given in Section 4. Section 5 concludes the paper with a conclusion and discussion.

2. Literature Background

This section presents a summary of the literature studies on churn detection. In [11], authors propose to predict customer churn in telecom sector with the help of machine learning techniques and big data platform. For this purpose, social network analysis (SNA)-based features are introduced before modeling SyriaTel telecommunication data. After that, decision tree, random forest, gradient boosted machine tree, and extreme gradient boosting techniques are assessed. Experiment results prove that the inclusion of SNA-based features boosts the prediction performance of churn customers with the increment nearly 9% in AUC. Moreover, authors emphasize extreme gradient boosting model exhibit superior performance compared to the others.
In [12], authors concentrate on the churn prediction problem and factors behind of it in telecom sector. As a first step, feature selection procedure is performed employing information gain, and correlation attribute ranking filter. Then, classification task is carried out with the aid of random forest algorithm to determine whether the customers tend to churn or not. After specifying the prediction performance, the reasons behind of churn is investigated by segmenting customers into different groups through k-means clustering. Authors conclude the study that the utilization of random forest and k-means techniques are capable to determine customers that likely to churn and factors behind of it, respectively. In [13], authors present a machine learning approach to construct a system that detect the churn customers in telecom sector. In this approach, data-preprocessing and feature analysis is performed. Then, gravitational search algorithm is carried out for the purpose of picking features up. Finally, data set is modeled with Adaboost and XGboost techniques. Authors conclude the paper that Adaboost exhibits superior performance with 81.71% of accuracy score. In [14], authors introduce a model that infers parameter estimation for the prediction of telecom clients. Principal component analysis is utilized as feature selection technique. To model the data, five different machine learning models are employed with 3 different parameter settings. These are k-nn, decision tree, random forest, stochastic gradient descent (SGD), multilayer perceptron (MLP). Based on experiment results, it is observed that third decision tree model outperforms all other models with 71% of F-measure by performing minimum sample split.
In [15], authors propose a system that predicts the customer churn using different machine learning techniques in banking sector. These methods, namely, k-nn, support vector machine, decision Tree, and random forest are blended with various feature selection techniques. Authors report that the usage of random forest method exhibits better classification accuracy compared to the others. In [16], authors focus on the churn prediction and retention of IT, banking, and telecom sectors. For this aim, classification performance of logistic regression, random forest, support vector machine, and XGBoost models are compared, comprehensively. In [17], authors propose to predict customer churn in retail banking with different machine learning models. Random forest, support vector machine, stochastic boosting, logistic regression, classification and regression trees, and multivariate adaptive regression splines are carried out on real data. Experiment results indicate that random forest, and stochastic boosting generally present better classification accuracies in different periods.
In [18], authors propose a novel technique using multi-objective rain optimization algorithm with the help of the combination of synthetic minority over-sampling (SMOTE) model with optimal weighted extreme machine learning (OWELM) method for the purpose of predicting churn of customers. The proposed technique comprises of three stages, namely pre-processing, dataset balancing, and categorization. Authors report that experiment results of proposed technique exhibit remarkable accuracy scores on three datasets. In [19], authors concentrate on metaheuristic optimization technique namely, chaotic pigeon to handle the customer churn

by predicting it with long short-term memory network model. To show the effectiveness of the proposed model, authors perform experiments on there datasets. Approximately, 95.56% accuracy score is provided on the first dataset as the best result. In [20], authors utilize the spoken contents in phone communication for the purpose of discovering possible churn of customers. A large-scale call center dataset is composed of two million calls from more than two hundred thousand clients. The experimental results indicate that the proposed model can forecast the risks of customer churn with the aid of machine learning techniques.
Our study differs from the literature studies in that both evaluating three different sectors and eight different machine learning classifiers extensively in terms of predicting customer churn.

3. Methodology

3.1. Linear Discriminant Analysis (LDA)

Linear discriminant analysis [21] is proposed by R. Fischer. It comprises discovering the projection hyperplane that diminishes the intercategory variance and maximizes the space between the reflected means of the categories. The hyperplane is utilized for categorization, reduction of dimension.

3.2. Logistic Regression (LR)

Logistic model is a statistical technique that models the probability of an event among out of two options performing by accomplishing the logarithm of the odds for the event be a linear consolidation of one or more independent predictors. In regression analysis, logistic regression [22] forecasts the variables of a logistic model or the coefficients in the linear consolidation. As in this work, there is a unique binary dependent parameter in binary logistic regression where the two values are tagged as "0" and "1".

3.3. K-Nearest Neighbor (K-NN)

K-nn is a genre of lazy learning method that is utilized for classification, and regression tasks. K-nn is a basic algorithm that is readily carried out through the nearest neighbor, distance functions. The method allocates all cases and classifies new conditions depending on similarity metrics namely, Euclidean, Manhattan, and Minkowski [23-24]. In classification and regression tasks, the input is related to the k closest training samples in the attribute space. In this study, K-nn algorithm is employed for the categorization problem enforcing Euclidean distance and setting k as 3.

3.4. Naïve Bayes (NB)

Naïve Bayes is a classifier based on Bayes' theorem that proposes independency between features [25]. The naïve Bayes technique is particularly evaluated when the size of input features is high. In the literature studies [26-32], the naïve Bayes classifier is assessed due to its superior performance, training time, and simplicity. In the experiments, Gaussian naïve Bayes model is employed.

3.5. Support Vector Machine (SVM)

The support vector machine (SVM) is a supervised machine learning technique that is one of the most popular, precise, and vigorous models for classification and regression tasks. SVM is proposed by Cortes and Vapnik [33] for the purpose of categorizing data that are linearly distinguishable. In addition to linear categorization, SVMs effectively ensure a non-linear categorization with the help of kernel trick, indirectly mapping inputs into high-dimensional attribute space. In this work, linear kernel is employed in the experiments.

3.6. Decision Tree (DT)

Decision tree [34] model constructs a classification technique that employs a graph like tree for decision making. Decision tree utilizes a tree demonstration to learn function of categorization that estimates the value of a dependent parameter, given the values of the independent features. It is a supervised learning algorithm that splits a tagged data set into smaller subsets while constructing decision tree. The procedure is sustained until the last subset has only similar objects.

3.7. Random Forest (RF)

Random forest is an ensemble learning model for regression problems and classification. The machine learning method of random decision forests is presented by Ho [35]. RF constructs multiple trees and consolidates them for the purpose of obtaining more robust and precise system outputs. To compose the forest, randomness is provided with consolidating bagging and random subspace methods. Thus, RF assesses random thresholds for each individual feature to construct different trees instead of utilizing best possible thresholds. The number of learners is set to 100 in the experiments.

3.8. Multilayer Perceptron (MLP)

The multilayer perceptron is a popular supervised learning model among other methods of artificial neural networks that is carried out for forecasting problems in the literature. MLP composes of fundamentally three layers, namely, input layer, hidden layer, and output layer. Hidden layer may compose of one or more than of more activation specified [36]. MLP is a feed-forward neural network that maps data into a set of related outputs. The MLP has robust computing sufficiency because of facilitating the remedy of non-linear problems. In the experiments, the MLP model consists of 8 neurons and 1 dense layer. ReLU is chosen as the activation function used in the layer. In addition, the heap size is 8 and the iteration value is 500. The remaining parameters are set as default.

3.9. Model

In this work, the extensive analysis of churn prediction is proposed by utilizing machine learning algorithms. The datasets are employed to analyze churn tendency in three different sectors, namely telecom, banking, and insurance. The telecom, banking, and insurance data sets include 3333, 22067, and 6321 instances, respectively. The number of features is 14 for Insurance data set, 21 for banking and telecom data sets as seen in the Figure 1. After gathering data set, principal component analysis (PCA) is applied for the purpose of feature reduction to improve the prediction accuracy of the system. After applying PCA, 11 features for Insurance data set, 15 features for banking data set, and 17 features for telecom data set are remained. In insurance data set, the features namely, row number, surname, and HasCrCard are eliminated. In banking data set, customer_id, gender, city, dependents, branch_code, and vintage features are removed. In telecom data set, state, area code, international plan, and voice mail plan attributes are extracted. After that, data sets are modeled with six different machine learning algorithms, an ensemble algorithm, and an artificial neural network (ANN). The models are linear discriminant analysis (LDA), logistic regression (LR), k-nearest neighbor (k-NN), naïve Bayes (NB), support vector machine (SVM), decision tree (DT), random forest (RF), multilayer perceptron (MLP).

4. Experiment Results

In this section, extensive experiments are performed to predict the customer churn employing in three different sectors employing machine learning algorithms. Accuracy (AC), f-measure (FM), precision (PR), and recall (RC) are employed as evaluation metrics to demonstrate the performance of the models. The accuracy score to categorize as false when false and true when true to the total number of data presented in Equation 1. In the experiments, dataset is splitted into 80% as training set, and remaining is test set.

$$Accuracy = \frac{TP+TN}{TP+NP+TN+FN} \quad (1)$$

Precision presents the proportion of accurately estimated positive observation outcomes to sum of positive forecasts. Negatives offers the proportion of precisely forecasted negative observation scores to sum of negative predictions. In Equation 2, precision is given.

$$Precision = \frac{TP}{TP+NP} \quad (2)$$

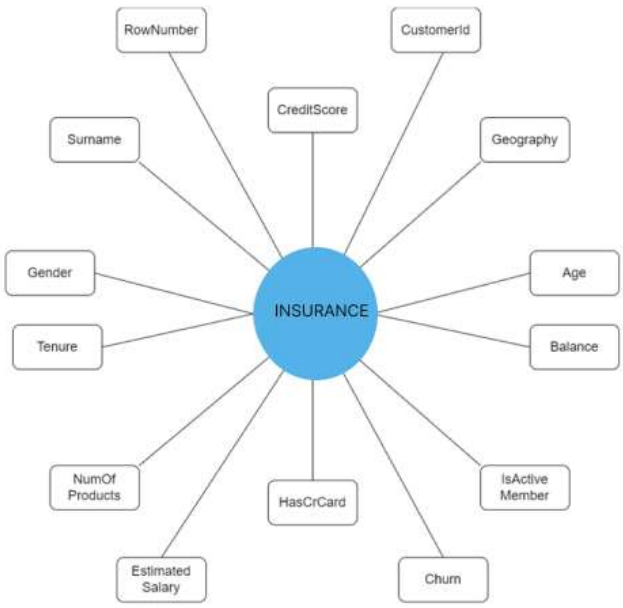

Figure 1. *The feature details of Insurance data set.*

Recall is employed for the purpose of evaluating the precision of the method. RC is calculated as the ratio of correctly estimated positive forecasts to the number of all essentially correct samples given in Equation 3.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Both false-positive and false-negative scores are included in the F-measure calculation as in Equation 4. When the class distribution is skewed, F-measure as an evaluation metric present more realistic scores compared to the sum of accuracy criteria when assessing the performance of the method.

$$F-measure = \frac{2*Recall*Precision}{Recall+Precision} \quad (4)$$

The following abbreviations are employed in the tables: RF: Random Forest, k-NN: k-nearest neighbor, LR: Logistic Regression, LDA: Linear Discriminant Analysis, NB: Naïve Bayes, SVM: Support Vector Machine, DT: Decision Tree, MLP: Multilayer Perceptron. The best scores are shown as bold in all tables. We also present the efficiency of using feature extraction by exhibiting experiment results in the tables.

In Table 1, the accuracy results of all classification techniques are demonstrated in insurance dataset without applying feature extraction. It is obviously observed that random forest exhibits superior classification performance as a homogeneous ensemble algorithm with 72.05% of accuracy. It is followed by MLP, SVM, NB, and LR nearly 67% of accuracy. The usage of k-NN techniques does not seem meaningful because of the worst classification success with 61.60% of accuracy score. Even though MLP is the second-best technique that fits to classify churn customers, nearly 5% difference in accuracy is significant when the superior performance of RF is considered. As a result of Table 1, the utilization of ensemble model exhibits remarkable result even if feature extraction is not carried out.

Table 1. *The results of all classification techniques are demonstrated in insurance dataset without feature extraction.*

Insurance	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	66.98	68.34	72.21	77.45
LR	67.38	70.13	71.36	75.36
k-NN	61.60	64.57	65.13	70.30
NB	67.39	65.42	65.86	74.67
SVM	67.30	66.36	68.40	75.89
DT	65.80	67.80	67.33	72.50
RF	**72.05**	73.05	75.27	80.33
MLP	67.83	71.57	70.30	76.08

In Table 2, the performance results of all classification techniques in terms of evaluation metrics in insurance dataset are presented when feature extraction procedure is applied.

Table 2. *The results of all classification techniques are demonstrated in insurance dataset with feature extraction.*

Insurance	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	70.52	70.40	71.00	80.42
LR	72.80	73.35	72.81	82.60
k-NN	68.45	70.22	71.96	81.02
NB	71.94	73.90	74.57	83.55
SVM	72.07	75.48	77.85	84.46
DT	70.48	72.06	73.09	75.57
RF	**77.53**	79.10	82.55	86.92
MLP	72.14	71.90	73.10	82.13

As it is clearly seen that in Table 2, RF outperforms other techniques with 77.53% of accuracy. It is followed by LR with 72.80%, MLP with 72.14%, SVM with 72.07%, NB with 71.94%, LDA with 70.52%, DT with 70.48%, k-NN with 68.45% of accuracy scores. Furthermore, the inclusion of feature extraction process boosts the classification performance approximately 5% in RF model. The order of model performances can be summarized as: RF> LR> MLP> SVM> NB> LDA> DT> k-NN. The performance order is not like Table 1. The reason behind of this, the classification success of LR, MLP, SVM, and NB is very close to each other. The inclusion of feature extraction procedure contributes significantly to predict customer churn with minimum %4 advancement, and maximum %8 enhancement. In Table 3, the classification results of each learning model are presented in terms of different evaluation metrics for banking dataset. The superior performance of RF is clearly seen in the banking dataset with 86.38% of accuracy when feature extraction is not performed. In banking dataset, RF exhibits better classification success compared to the insurance data set with roughly 14% advancement in classification accuracy. The poorest classification performance is carried out by DT in banking dataset with 77.68% of accuracy score. Even if the success of DT is considered in insurance data set, 12% improvement is provided in banking data set. In Table 2, k-NN algorithm is surprisingly competitive with 84.23% of accuracy when the complexity of MLP algorithm with 84.60% of accuracy is considered. The success of classification models is ordered as: RF> MLP> k-NN> LR> SVM> LDA> NB> DT. In Table 4, classification performance of each technique is demonstrated in terms of different evaluation measures in banking dataset. RF outperforms others with 89.93% of accuracy. RF is capable to classify customers' churn especially in banking dataset with the help of feature extraction method. MLP is also competitive with 87.40% of accuracy score.

Table 3. *The results of all classification techniques are demonstrated in banking dataset without feature extraction.*

Banking	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	81.83	83.14	81.25	81.60
LR	83.51	85.10	76.23	85.02
k-NN	84.23	87.35	81.37	88.55
NB	80.79	88.23	83.93	84.68
SVM	82.46	82.84	85.00	87.02
DT	77.68	83.10	80.59	85.94
RF	**86.38**	87.45	87.65	86.06
MLP	84.60	83.48	85.23	85.90

Table 4. *The results of all classification techniques are demonstrated in banking dataset with feature extraction.*

Banking	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	84.42	85.70	84.36	89.45
LR	85.57	87.43	86.90	90.13
k-NN	86.30	85.30	84.21	90.51
NB	81.22	83.01	82.57	87.24
SVM	85.55	86.40	85.68	88.00
DT	80.38	80.27	79.26	85.02
RF	**89.93**	90.51	89.75	93.57
MLP	87.40	88.13	87.59	90.45

Similarly, k-NN exhibits superior performance when Table 1 and Table 2 is considered in insurance data set. The inclusion of feature extraction approach does not influence the success of NB, significantly nearly with 1% improvement in accuracy result. The performance order of the models is the same as the results of Table 3. Although DT has the poorest classification success, it provides approximately 3% improvement with the inclusion of feature extraction procedure.

In Table 5, classification scores of the models are presented in telecom dataset in terms of different evaluation metrics when feature extraction process is not applied. Apart from the other data sets, NB exhibits the poorest classification performance with 84.41% of accuracy. It is followed by LDA with 84.86%, MLP with 85.22%, SVM and LR with 85.31%, k-NN with 87.41%, DT with 90.40%, and RF with 95.90% of accuracies. Differently from the Table 1, Table 2, Table 3, and Table 4, DT demonstrates the second-best classification success while in other datasets it exhibits the poor classification success. The performance sequence of the models varies as: RF> DT> k-NN> LR~SVM> MLP> LDA> NB. It can be arisen since the number samples and features of telecom dataset are proportionally less than the other datasets.

Table 5. *The results of all classification techniques are demonstrated in telecom dataset without feature extraction.*

Telecom	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	84.86	89.37	86.67	88.43
LR	85.31	84.04	82.00	88.16
k-NN	87.41	90.40	85.91	89.59
NB	84.41	86.39	86.88	85.92
SVM	85.31	85.80	84.57	85.10
DT	90.40	89.81	88.91	89.97
RF	**92.05**	91.36	90.14	89.72
MLP	85.22	86.13	85.48	85.81

In Table 6, the effect of feature extraction process is clearly observed when the improvements on each classification model are considered. RF indicates the best classification performance (95.90% of accuracy) with inclusion of feature extraction process when compared to the other datasets namely, insurance with 77.53%, and banking 89.93% of accuracy results. This means the inclusion of RF as an ensemble algorithm boosts classification success in each data set with different improvements. On the other hand, LDA exhibits the worst classification success with 85.10% of accuracy. As a result of Table 2, Table 4, and Table 6, process of feature extraction is beneficial in terms of advancing classification performance of the system to predict the customer churn for all three sectors. The best performance for each dataset is achieved by

Table 6. *The results of all classification techniques are demonstrated in telecom dataset with feature extraction.*

Telecom	Evaluation Metrics			
Models	**AC**	**FM**	**PR**	**RC**
LDA	85.10	87.22	87.95	88.36
LR	87.43	89.23	88.12	89.40
k-NN	90.15	91.55	92.00	91.23
NB	88.59	89.07	89.18	90.52
SVM	91.77	91.50	89.27	89.13
DT	93.46	90.59	91.27	90.00
RF	**95.90**	93.17	94.38	95.54
MLP	89.05	90.03	89.45	88.75

proposed RF model. The performance of the proposed model according to the training set percentages is presented in Figure 2. To show the effectiveness of the proposed model, experiment results are compared with the state-of-the-art studies. In [37], authors present transfer learning and ensemble learning based framework to categorize churn of customers. After transfer learning stage, GP-AdaBoost is employed to classification purpose. 75.40% of accuracy is obtained on orange dataset. In our experiments, after feature selection procedure, 95.90% of accuracy is achieved. In [38], authors focus on the hybrid firefly based to predict the churn of customers in telecommunication sector. Authors report that firefly-based algorithm can classify the churn customers with 86.38 of accuracy. There is a difference in terms of the number of instances when Orange telecom dataset is considered. Although dataset includes 50,000 records in [38], proposed model in our study outperforms with approximately 9% enhancement in accuracy score.

5. Conclusion and Discussion

In this work, the efficiency of homogeneous classifier ensembles on customer churn prediction in banking, insurance, and telecommunication sectors is investigated. For this purpose, different machine learning techniques are modeled to indicate the

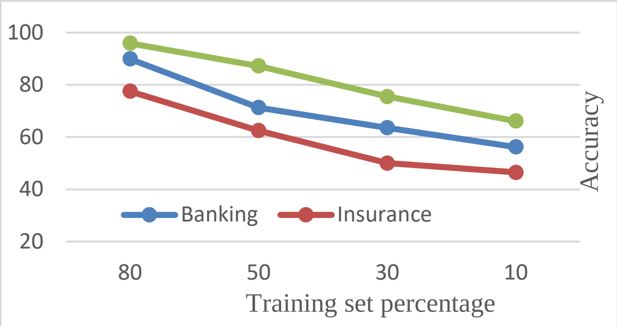

Figure 2. *The feature details of Banking data set.*

effectiveness of proposed ensemble framework. The models used in the study are k-NN, logistic regression, naïve Bayes, support vector machine, decision tree, random forest, multilayer perceptron algorithms are employed. Moreover, the impact of the inclusion of feature extraction stage are analyzed. Experiment results show random forest model can estimate churn customers with 89.93% of accuracy in banking, 95.90% of accuracy in telecommunication, and 77.53% of accuracy in insurance sectors when feature extraction procedure is performed. In future, we plan to develop deep learning methodology for predicting customers' churn by extending datasets and using different feature extraction techniques.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Babu, S., Ananthanarayanan, D. N., Ramesh, V. (2014). A Survey on Factors Impacting Churn in Telecommunication using Data Mining Techniques. Int. J. Eng. Res. Technol., 3: 1745-1748.

[2] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., Kim, S. W. (2019). A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. IEEE Access, 7: 60134-60149. DOI: 10.1109/ACCESS.2019.2914999.

[3] Idris, A., Khan, A. (2012, December). Customer churn prediction for telecommunication: Employing various features selection techniques and tree based ensemble classifiers. In International Multitopic Conference (pp. 23-27). DOI: 10.1109/INMIC.2012.6511498

[4] Kaur, M., Singh, K., Sharma, N. (2013). Data mining as a tool to predict the churn behaviour among Indian bank customers. Int. J. Recent Innov. Trends Comput. Commun., 1(9): 720-725.

[5] Verbeke, W., Martens, D., Mues, C., Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Syst. Appl., 38(3): 2354-2364. DOI: 10.1016/j.eswa.2010.08.023

[6] Rokach, L. (2010). Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2): 1–39. DOI: 10.1007/s10462-009-9124-7

[7] Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3): 21-45. DOI: 10.1109/MCAS.2006.1688199

[8] Gopika, D., Azhagusundari, B. (2014). An analysis on ensemble methods in classification tasks. International Journal of Advanced Research in Computer and Communication Engineering, 3(7): 7423–7427.

[9] Ren, Y., Zhang, L., Suganthan, P., N. (2016. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. IEEE Computational Intelligence Magazine, 11(1): 41-53. DOI: 10.1109/MCI.2015.2471235

[10] Kilimci, Z. H., Akyokus, S., Omurca, S. I. (2016, August). The effectiveness of homogenous ensemble classifiers for Turkish and English texts. In IEEE International Symposium on INnovations in Intelligent SysTems and Applications (pp. 1-7). DOI: 10.1109/INISTA.2016.7571854

[11] Ahmad, A., K., Jafar, A., Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. J Big Data, 6(28): 1-24. DOI: 10.1186/s40537-019-0191-6

[12] Ullah, I., Raza, B., Malik, A., K., Imran, M, Islam, S., U., Kim, S., W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access, 7: 60134-60149. DOI: 10.1109/ACCESS.2019.2914999

[13] Lalwani, P., Mishra, M., K., Chadha, J., S., Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 104: 271–294. DOI: 10.1007/s00607-021-00908-y

[14] Pamina, J., Beschi Raja, J., Sam Peter, S., Soundarya, S., Sathya Bama, S., Sruthi, M.S. (2020). Inferring Machine Learning Based Parameter Estimation for Telecom Churn Prediction. In: Smys, S., Tavares, J., Balas, V., Iliyasu, A. (eds) Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in Intelligent Systems and Computing, vol 1108. Springer. DOI: 10.1007/978-3-030-37218-7_30

[15] Rahman, M., Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In IEEE International Conference on Electronics, Communication and Aerospace Technology (pp. 1196-1201). DOI: 10.1109/ICECA49313.2020.9297529

[16] Jain, H., Yadav, G., Manoov, R. (2021). Churn Prediction and Retention in Banking, Telecom, and IT Sectors Using Machine Learning Techniques. In: Patnaik, S., Yang, XS., Sethi, I. (eds) Advances in Machine Learning and Computational Intelligence. Algorithms for Intelligent Systems. Springer, Singapore. DOI: 10.1007/978-981-15-5243-4_12

[17] Dias, J., Godinho, P., Torres, P. (2020). Machine Learning for Customer Churn Prediction in Retail Banking. In:, et al. Computational Science and Its Applications – ICCSA 2020. ICCSA 2020. Lecture Notes in Computer Science, vol 12251. Springer, Cham. DOI: 10.1007/978-3-030-58808-3_42

[18] Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M., & Shankar, K. (2021). Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. Complex & Intelligent Systems, 1-13. DOI: 10.1007/s40747-021-00353-6

[19] Pustokhina, I. V., Pustokhin, D. A., Aswathy, R. H., Jayasankar, T., Jeyalakshmi, C., Díaz, V. G., & Shankar, K. (2021). Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. Information Processing & Management, 58(6): 102706. DOI: 10.1016/j.ipm.2021.102706

[20] Vo, N. N., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. Knowledge-Based Systems, 212: 106586. DOI: 10.1016/j.knosys.2020.106586

[21] Xanthopoulos, P., Pardalos, P.M., Trafalis, T.B. (2013). Linear discriminant analysis. In: Robust Data Mining. SpringerBriefs in Optimization. Springer, New York, NY. DOI: 10.1007/978-1-4419-9878-1_4

[22] Tolles, J., Meurer, W., J. (2016). Logistic regression relating patient characteristics to outcomes. JAMA. 316 (5): 533–534. DOI:10.1001/jama.2016.7653

[23] Tsigkritis, T., Groumas, G., Schneider, M. (2018). On the use of k-NN in anomaly detection. Journal of Information Security, 9: 70-84. DOI: 10.4236/jis.2018.91006

[24] Martín-Valdivia, M., T., Rushdi, Saleh, M, Ureña-López, L., A., MontejoRáez, A. (2011). Experiments with SVM to classify opinions in different domains. Expert Systems with Applications, 38(12): 14799-14804. DOI: 10.1016/j.eswa.2011.05.070

[25] Ren, J., Lee, S., D., Chen, X., Kao, B., Cheng, R., Cheung, D. (2009, December). Naive Bayes classification of uncertain data. In: IEEE International Conference on Data Mining (pp. 944 – 949). DOI: 10.1109/ICDM.2009.90

[26] Horn, C. (2010). Analysis and classification of Twitter messages (Master's thesis). Graz University of Technology, Graz, Austria.

[27] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence (pp. 41-46).

[28] Jiang, L., Zhang, H., Cai, Z. (2008). A novel Bayes model: Hidden naive Bayes. IEEE Transactions on knowledge and data engineering, 21(10): 1361-1371. DOI: 10.1109/TKDE.2008.234

[29] Frank, E., Trigg, L., Holmes, G., Witten, I. H. (2000). Naive Bayes for regression. Machine Learning, 41(1): 5-25. DOI: 10.1023/A:1007670802811

[30] Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15). DOI: 10.1007/BFb0026666

[31] Kilimci, Z. H., Ganiz, M. C. (2015, September). Evaluation of classification models for language processing. In: 2015 IEEE International Symposium on Innovations in Intelligent SysTems and Applications (pp. 1-8). DOI: 10.1109/INISTA.2015.7276787

[32] Kilimci, Z. H., Akyokuş, S. (2016, May). N-gram pattern recognition using multivariate-Bernoulli model with smoothing methods for text classification. In: 2016 24th IEEE Signal Processing and Communication Application Conference (pp. 597-600). DOI: 10.1109/SIU.2016.7495811

[33] Cortes, C., Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3): 273-297. DOI: 10.1007/BF00994018

[34] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1): 81-106. DOI: 10.1007/BF00116251

[35] Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). DOI: 10.1109/ICDAR.1995.598994

[36] Murtagh, F. (1991). Multilayer perceptrons for classification and regression. Neurocomputing, 2(5-6), 183-197. DOI: 10.1016/0925-2312(91)90023-5

[37] Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer learning and meta classification based deep churn prediction system for telecom industry. arXiv preprint arXiv:1901.06091.

[38] Ahmed, A. A., & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly-based classification. Egyptian Informatics Journal, 18(3): 215-220. DOI: 10.1016/j.eij.2017.02.002