International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 2, February 2022, pg.75 – 84

Implementing Random Forest to Predict Churn

Linda Makurumidze¹; Wellington Simbarashe Manjoro²; Wellington Makondo³

¹Information Technology & HIT, Zimbabwe

²Information Technology & HIT, Zimbabwe

³Information Technology & HIT, Zimbabwe

¹leemak2011@gmail.com; ²wmanjoro@hit.ac.zw; ³wmakondo@hit.ac.zw

DOI: doi.org/10.47760/ijcsmc.2022.v11i02.009

Abstract— For one to remain afloat in business, the best marketing technique is to maintain the current customers rather getting new ones [1]. [2] Shows that it costs more to get a fresh client than maintaining the available ones. An organization that intends to keep its customers must speculate which of them is at risk of abandoning their service and put all their concentration on those customers in an effort to retain them.

This paper's contribution is to create a prototype, which aid banks to foretell clients that are prone to abandon their service. This paper makes use of four algorithms namely Gradient boost, Random forest, Adaboost and Decision tree to classify and segment bank clients based on a number of features. The paper then selects the best performing algorithm, that is Random forest, to build a prediction model that can be used by banks to identify the most likely clients to churn away.

Keywords— Customer churn, Customer churn management, Predictive data analytics, Machine learning, Evaluation metrics

I. INTRODUCTION

Over the past years, there has been stiff competition for customers amongst different service providers, as customers are the backbone of an organization. Customer turnover is a serious issue and it affects all profit centric organizations as it has a direct impact on the profits of an organization. Customer loyalty has gained more attention in the everyday business technique. Customer turnover is a potentially critical issue for every organization. Management of churn is the process of detecting consumers who intend to switch service to a competitor [3].

The turnover of clients can be termed attrition or defection, and is the loss of consumers. Consumer turnover has a huge effect on organizations and is the key point for the organizations to remain afloat in business. It is much more costly for a company to get fresh customers than maintaining the current ones [2]. Churn prediction enables organizations to keep as many consumers as possible, by implementing loyalty programs and retention efforts.

The expense of recruiting new consumers is more than five times that of maintaining current clients according to [2], [7] and [9]. According to [1] the cost of acquiring new clients, which requires a significant investment of resources, is confirmed to be costlier than the cost of retaining current clients. When an

organisation tries to reach out to new clients, the costs are rather high than when they maintain current consumers. As a result, a lot of effort has gone into finding innovative techniques to identify consumers who are at high risk of churning. This has led to a lot of innovative strategies that able to detect consumers who are at risk of abandoning the service in advance.

Churn is defined by [4] as the transfer of existing consumers from competitor to competitor. Consumer turnover is an inevitable action in all industries [5] and organizations are adopting machine-learning techniques to solve their problems and to understand customer behaviour in a better way. Customer turnover is the tendency of a consumer to abandon a service provider [6], [7]. According to [4] management of churn is a procedure that entails building a machine learning model based on historical data and identifying key attributes that cause customers to leave. As a result, the aim of predicting churn is to detect customers that are likely to exit ahead of time so that the service provider may use their marketing approach to keep them.

II. STATE OF ART

Author [7] conducted a comparative analysis for the performance of Decision tree, ANN, KNN, SVM and proved that hybrid of all the four performs well and the author proposed a methodology for extracting the important features from the dataset. According to [8] exploratory data analysis was implemented to explore the likely features within the data set that resulted in customer turnover. To anticipate customer turnover, a K-means clustering method was used to separate the customers, and then a decision tree model was developed in each cluster. Second, to anticipate churn, a Back Propagation Neural Network and a Decision Tree were used. [8] used data mining to create numerous churn prediction models, demonstrating that both decision tree and neural network techniques could produce accurate churn prediction models. After evaluating four algorithms (Gradient Boosted Machine Tree, Extreme Gradient Boosting, Decision Tree, and Random Forest), Ahmed et al [5] built a churn prediction model utilizing XGBOOST (XGBOOST). By mining Social Network Analysis features from the consumer social network, he was able to improve the model's performance up to 93.3 percent.

Author [10] proposed a geo-demographic segmentation and predictive analysis methodology that divides customers into groups based on their geographic or demographic data. He used backward elimination and forward selection procedures in stepwise logistic regression; a bidirectional model was built, and the best among them was chosen using the CAP analysis curve. Geodemographic segmentation according to [11], is a procedure that involves grouping individuals into clusters based on their geographical and analytical characteristics. For churn time prediction or tenure modelling, [12] employed ordinal regression. Ordinal regression was compared to multiclass classification and the Cox PH model for survival analysis and was found to perform better.

Author [13] suggested integrating structural equation modelling and Bayesian networks and used it for customer retention outcome assistance in a virtual community. The outcomes of the application gave practitioners ideas on how to keep their customers. Author [6] investigated the utility of popular data mining methods for predicting turnover and found that linear models, such as logistic regression, were the best choice for modelling prepaid client turnover. In cooperation with several feature reduction strategies, [9] evaluated the importance of a Particle Swarm Optimization based under sampling strategy to handle the imbalance data distribution. In a retention effort, [14] established a novel, profit-centric performance metric to choose consumers with the greatest anticipated probabilities of churn. Their study found that a modest number of characteristics were enough to accurately forecast attrition, and that oversampling did not increase performance considerably.

Bayesian belief networks was implemented by [15] to show the relationships between the features that contributes to customer churn, important features were found out some marketing was done so as to reduce the number of customers that turnover. Comparative analysis was done by [16] for accuracy of several classifiers namely Gradient Boosting , SVM ,Neural Network, Linear regression, Random Forest, AdaBoost, and C4.5 with fuzzy model and shows that fuzzy classifiers especially OWANN achieved a true positive rate of 0.98 and AUC score of 68% are more accurate in predicting customer churn using a noisy dataset.

A comparative study carried out by [17] on different classifiers, and cross validation was used to evaluate them, then parameter tuning was done and Monte Carlos simulations was used to detect the possible parameter combinations. The results reveal that boosted models outperform non-boosted models, with the best SVM-POLY classifier employing AdaBoost having 97 percent accuracy and an F score measure of over 80 percent. Tsai et al [18] combined two neural nets to forecast attrition and they are ANN and SOM. Compound representations consists of two ANNs; and SOM together with ANN. he initial compound algorithms reduces the dataset by screening out unclassified learning data, then utilizing the outcomes to make a forecasting model using the second method. From trials, the compound ANN performs better than the others do.

Table I
Related Literature about Customer Churn

Authors	Dataset	Techniques used	Evaluation metrics	Outcome
[J].Paminaa et al	Telecom Industry	Random Forest , K-NN, and XG boost	Accuracy score F-score	Out of the three, the XG boost classifier performs well
[20] Abinash Mishra et al	Telecom Industry	Naïve Bayes Bagging , Decision Tree Boosting, Random Forest, ,SVM	Accuracy Error rate Specificity Sensitivity	Random Forest classifier performs well
[21]Coussement K., et al	News paper subscription	random forests , support vector machine, Logistic regression,	PCC, AUC under sampling , cross validation,	support vector machines
[19] Manas et al	Bank	KNN, SVM, Decision Tree(DT) and RF	Accuracy	Random forest
[5] Ahmad et al.	Telecom	Extreme Gradient Boosting “XGBOOST”, Gradient Boosted Machine Tree “GBM”, Decision Tree, Random Forest.	Area Under Curve	XGBOOST
[22] Eiben A.E., et al	Telecom	CHAID , genetic programming, Logistic regression, rough data analysis,	OMEGA software ,Lift chart, hold-out , CoC , PCC,	Genetic technology

III.OBJECTIVES

- The primary objective is to develop a machine learning model that accurately foretells clients attrition.
- Secondly to discover the reasons why customers churn.
- The purpose is to segment clients so that a marketing strategy may be developed. As a result, the purpose of churn prediction is to spot clients who are going to abandon the service in advance, so that the service provider can utilize marketing to maintain them.

IV.METHODOLOGY

This study, machine learning is used to develop a model that is able to forecast the future before it actually happens. Machine learning is defined as an automated process that mines patterns from a dataset. The art of developing and using models that make predictions based on patterns extracted from data from the past is called predictive data analytics.

A. Dataset

In this study, we used a public bank data set from Kaggle, which comprises of 10127 customer records with 21 features. Attrition Flag feature denotes churn or not in which 8500 are not churners whilst 1627 are churners and the percentages are 16.1% and 83.9% of churners and non-churners respectively.

Table II
Dataset Feature Description

Feature	Description
Client number	Client ID
Attrition Flag	Account activity 0 or 1
Age	Age of the customer
Sex	M or F
Dependents	Number of dependents
Qualifications	Qualifications of the customer
Marital Status	Married, Divorced, Single ,Unknown
Income Class	Category of Income of the customer
Bank Card type	Card type (Blue, Silver, Gold, Platinum)

Feature	Description
Duration active	Period using bank
Number of products	Number of products held by the customer
Inactive Months	Number of inactive months
Count of Contacts	Number of Calls to the bank
Limit on Credit	Limit on Credit card
Total Revolving Bal	Balance revolving
Average to Buy	Amount available to buy
Total Amount Change	Variation in Transaction Amount
Total Transaction Amount	Total Transaction Amount
Total Transaction Count	Count in Transaction
Total Count Change	Variation in Transactions
Average Utilization Ratio	Utilization of the Card

B. Data Scrubbing / Feature Engineering

In this study, five tasks were carried out for data pre-processing. Firstly, we dropped some two features, which were irrelevant to the study. Then we performed the missing value analysis and it is found that no values were missing. Next, we checked for duplicates entries in our dataset and it shows that all the rows are unique. In addition, the conversion of data types like object data type to numerical data type is implemented for the classifiers we used for analysis.

Finally, a train and test data sets were created, and because the dataset was uneven, SMOTE was employed to balance it. Also cross validation is implemented to avoid over fitting. For cross validation K-fold is used.

C. The Predictive Model

The following classifiers were used, Gradient boost, Random forest, Adaboost and Decision tree were used to build and train the model.

- 1) *Decision tree*: In this supervised learning method, the data is continuously segregated according to a particular parameter, which is represented as a tree structure. The Attribute Selection Measure is a method for selecting the finest qualities (ASM). ASM employs two strategies: information gain and the Gini Index.
- 2) *Random forest*: The classifier uses the ensemble technique of several judgments and it is a supervised algorithm. It combines the predictions of numerous trees to provide an end result using the majority vote. The classifier is a supervised technique that comprises of several decisions. It combines the predictions of multiple decision trees to provide a majority vote-based result.
- 3) *Adaboost*: Ada-boost is an iterative ensemble boosting classifier that increases accuracy by combining numerous classifiers. In each iteration, the goal is to identify the weights of classifiers and train the data sample in order to provide reliable predictions of unexpected events.
- 4) *Gradient boost*: An ensemble algorithm, which unites the Boosting technique and gradient descent to predict the residuals of each of the base estimators. It creates a sequence of base estimators then; it indicates the residue for each of them so that the following estimator is more accurate and can reduce the residuals successively.

V. RESULTS AND MODEL EVALUATION

We evaluated our models using Precision, Recall, Accuracy, Confusion matrix and ROC/AUC evaluation metrics. Hyper parameter tuning involves tuning the model parameters for improved performance.

A. Evaluation metrics for the classification models

A machine-learning model's performance is measured using evaluation metrics. We conducted experiments on four classifiers, which are Gradient Boosting, Random forest, Adaboost and Decision tree classifier. Their performance were evaluated using recall, F1_score ,accuracy score and precision measures. ROC/ Area Under

Curve (AUC) was also implemented to evaluate and validate the models. The results were compared to find an appropriate model with higher precision, accuracy and predictability.

Below are the experimental results for the tested algorithms:-

1) *Classification reports:*

Decision tree

	precision	recall	f1-score	support
0	0.94	0.97	0.95	1647
1	0.83	0.72	0.77	379
accuracy			0.92	2026
macro avg	0.89	0.84	0.86	2026
weighted avg	0.92	0.92	0.92	2026

Random forest

	precision	recall	f1-score	support
0	0.97	0.98	0.98	1691
1	0.89	0.87	0.88	335
accuracy			0.96	2026
macro avg	0.93	0.92	0.93	2026
weighted avg	0.96	0.96	0.96	2026

Gradient boosting

	precision	recall	f1-score	support
0	0.95	0.98	0.97	1653
1	0.90	0.79	0.84	373
accuracy			0.95	2026
macro avg	0.93	0.89	0.91	2026
weighted avg	0.94	0.95	0.94	2026

Ada boost

	precision	recall	f1-score	support
0	0.94	0.98	0.96	1631
1	0.90	0.74	0.81	395
accuracy			0.93	2026
macro avg	0.92	0.86	0.89	2026
weighted avg	0.93	0.93	0.93	2026

2) F1-score Cross-validation for each model

Fig 1 : Compare F1 score for Cross-validation for each model

Fig 1 shows that Random forest has an F-score of 0.87 followed by Gradient boost with 0.84 then 0.81 and 0.77 respectively for Adaboost and Decision tree.

3) Confusion Matrix for all classifiers

Fig 2: Confusion matrix

From the above results, we see in both Fig 3 and classification report that both Random forest and Gradient boosting had accuracy above 95% and on Fig 2 Random forest has the highest F1-score. Adaboost and Decision tree had 93% and 92% accuracy respectively.

4) Receiver Operating Characteristic curve / Area Under Curve

AUC score for Adaboost 0.97
AUC score for Random Forest 0.99
AUC score for Gradient Boosting 0.98
AUC score for Decision Tree 0.89

Fig 3: ROC/AUC

In addition to the above metrics, ROC/AUC was applied to all classifiers to evaluate and to further compare the classifiers. From the experiment Fig 4 shows that Random forest has an AUC score of 99% followed by Gradient boost classifiers with 98%, which shows the model is learning the data well enough.

B. Hyper parameter tuning (Optimization)

From the above results, we will now consider Gradient Boosting and Random Forest classifiers. Hyper parameter tuning was done on Gradient boosting and Random Forest classifiers but they was no significant change in terms of accuracy. Below are the experimental results after hyper parameter tuning.

Random Forest Classifier (Optimized)

Fig 4: Confusion matrix for Random forest (optimized)

Gradient boosting classifier (optimized)

	precision	recall	f1-score	support
0	0.97	0.98	0.98	1699
1	0.90	0.84	0.87	327
accuracy			0.96	2026
macro avg	0.93	0.91	0.92	2026
weighted avg	0.96	0.96	0.96	2026

Fig 5: Confusion matrix for Gradient boost (optimized)

Table III:
Summary of Hyper Parameter Tuning Results

Classifier	Precision	Recall	F1-score	Accuracy
Gradient boost	0.90	0.84	0.87	0.96
Random forest	0.93	0.84	0.88	0.96

From the above hyper tuning results in Fig 5 and Fig 6 it shows that ,they was a significant improvement after parameter tuning for Gradient boost and Random forest in terms of f1-score ,recall and precision .

C. Comparative analysis of churn prediction classifiers

Experimental results show that Random Forest and Gradient Boost have the best model as compared to the other with other metrics score of precision, f1 and accuracy has shown at least 95% accuracy. In addition, both Gradient Boost and Random Forest have AUC score of at least 98% showing the model is learning the data well enough. Even though most of the models are improving in their precision and F1-score, Decision tree seem to have significant reduction in their precision, recall and F1-score.

Gradient Boost and Random forest classifiers gives higher accuracy score and F1-score compared to Decision tree and Adaboost classifiers. From ROC/AUC, F1-score, confusion matrix and hyper parameter tuning, Random forest seem to be performing well out of the three classifiers. Thus, Random forest classifier is used for feature selection in the proposed system.

D. Feature importance

Fig 6: Feature importance

Fig 6 shows that the most important feature that decides between the churning customer and the existing customer is the Total Transaction Amount. It makes logical sense since anyone who is planning on churning will leave the bank's services before closing down the current account.

VI. TECHNOLOGY OR BUSINESS CASE DESCRIPTION

The number of customers that are leaving their service concerns a bank and they have consulted a data analyst to develop a machine learning that can forecast clients' turnover using the available data. Lots of data has been accumulated within the bank environment and the manager wants to find a way of utilising the data to gain insights and to get value from it so as to make well informed decisions.

i) Business Benefits

- Models are used to make predictions that help businesses make informed decisions. Models that data analyst develop are able to forecast events before they actually happens using patterns extracted from data from the past. These predictions provide insights that help the organization make better decisions to solve their business problem. It helps in finding hidden patterns and extracting useful information from data.
- Churn prediction model helps businesses detect customer actions in advance using data from the past and also to adjust certain attributes and make them favourable for the customers to stay.
- Identifying customers who are going to churn in advance helps the business take appropriate action by applying some marketing strategy to retain these customers.
- High customer retention that is enabled by churn prediction models makes the business have high profits.

VII. CONCLUSIONS

A comparative study was carried out for four prominent classifiers namely Gradient boost, Random Forest, Adaboost and Decision tree for improving accuracy of customer churn prediction. Gradient boost and random forest classifiers performs well out of the four on the publicly available bank dataset, but for our proposed system Random forest is implemented.

Further, the importance of each customer attribute is analyzed and we found out that only seven attributes have more than 50% importance towards churn prediction. Next, the work focused on identifying the attribute, which has higher cognition for churn using Random forest classifier. The experiment result shows that customers with lower Total Transaction Amount have higher influence for churn.

Future works can be extended to hybrid models and the proposed model can be tested on large data sets.

ACKNOWLEDGEMENT

Firstly, I want to thank God taking me this far. Secondly, I just want to appreciate my family for their support during the process. Also, would like to convey my gratitude to Mr. Manjoro and Mr Makondo my supervisors for their guidance.

REFERENCES

- [1]. N. Singh et al, 2020 Machine learning based classification and segmentation techniques for CRM: a customer analytics. *International Journal of Business Forecasting and Marketing Intelligence* vol.6.
- [2]. James A. Rozum, Defining and Understanding Software Measurement Data. Software Engineering Institute of Carnegie Mellon University, paper 2002. 2-3.
- [3]. John Haddena,*, Ashutosh Tiwaria, Rajkumar Roy, Dymitr Rut 2005 , Computer assisted customer churn management: State-of-the-art and future trends in *Computers & Operations Research* 34 (2005) 2902 – 2917.
- [4]. Berson, A., Smith, S. and Thearling, K. (2000) *Building Data Mining Applications for CRM*, New York.
- [5]. Ahmed, A. A. and Maheswari, D. Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal* 18 (3) (2017) 215-220.
- [6]. Owczarczuk, M. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications* 37 (6): (2010) 4710-4712.
- [7]. Keramati, A., Jafari-Marandi, R., Aliannejadi, M., et al. (2014) Improved Churn Prediction in Telecommunication Industry Using Data Mining Techniques. *Applied Soft Computing*, vol 24 page994-1012.
- [8]. Hung, S., Yen, D.C. and Wang, H. (2006) Applying Data Mining to Telecom Churn Management. *Expert Systems with Applications*, volume 31, Issue 3, October 2006, Pages 515-524.
- [9]. Idris, A., Rizwan, M. and Khan, A. (2012) Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies. *Computers & Electrical Engineering*, 38(6): 1808-1819.
- [10].Hoang Viet Long et al A New Approach for Construction of Geodemographic Segmentation Model and Prediction Analysis. *Computational Intelligence and Neuroscience* 2019(1):1-10

- [11].H. S. Le, "Enhancing clustering quality of geo-demographic analysis using context fuzzy clustering type-2 and particle swarm optimization," *Applied Soft Computing*, vol. 22, September 2014 pp. 566–584, 2014.
- [12].R. K. Gopal and S. K. Meher, "Customer Churn Time Prediction in Mobile Telecommunication Industry Using Ordinal Regression," in *Proceedings of the 12th Pacific- Asia conference on Advances in knowledge discovery and data mining*, 2008, pp. 884-889
- [13].Gupta, S. and Kim, H.W. (2008) Linking Structural Equation Modeling to Bayesian Networks: Decision Support for Customer Retention in Virtual Communities. *European Journal of Operational Research*, 190(3):818-833.
- [14].Verbeke, W., Dejaeger, K., Martens, D., et al. (2012) New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach. *European Journal of Operational Research*, 218(2):211-229.
- [15].Kisioglu, P., Topcu, Y. I. (2011) Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Systems with Applications* 38 (6):7151-7157.
- [16].Azeem, M., Usman, M. and Fong, A. C. M (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems* vol 66 issue 4, page 603-614.
- [17].Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. and Chatzisavvas, K. C (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory* vol 55 pp 1-9.
- [18].Tsai, C. F. and Lu, Y. H (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications* vol 36 (10) pp 12547-12553.
- [19].Manas Rahman, V Kumar (2020) Machine learning based customer Churn prediction in banking. *Conference: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*
- [20].Abinash Mishra, U. Srinivasulu Reddy, (2017). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. *Conference: 2017 International Conference on Inventive Computing and Informatics (ICICI)*
- [21].Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, vol 34, issue 1, pp 313–327.
- [22].Eiben, A., Koudijs, A., & Slisser, F. (1998). *Genetic modeling of customer retention*. Lecture Notes in Computer Science, vol 1391, pp 178–186.
- [23].J. Pamina et al, 2019, An Effective Classifier for Predicting Churn Telecommunication. *Journal of Advanced Research in Dynamical and Control Systems* vol 11(1):221-229.