Determining the Intervening Effects of Exploratory Data Analysis and Feature Engineering in Telecoms Customer Churn Modelling

Alrence Santiago Halibas
Faculty of Computing Sciences
Gulf College
Muscat, Oman
alrence@gulfcollege.edu.om

Anju Cherian Matthew
Faculty of Computing Sciences
Gulf College
Muscat, Oman
anjuit@gulfcollege.edu.om

Indu Govinda Pillai
Faculty of Computing Sciences
Gulf College
Muscat, Oman
indu@gulfcollege.edu.om

Jay Harold Reazol
College of Engineering and Architecture
La Salle University
Ozamiz City, Philippines
jayharold.reazol@lsu.edu.ph

Erbeth Gerald Delvo
College of Engineering and Architecture
La Salle University
Ozamiz City, Philippines
erbethgerald.delvo@lsu.edu.ph

Leslyn Bonachita Reazol
College of Engineering and Architecture
La Salle University
Ozamiz City, Philippines
leslyn.reazol@lsu.edu.ph

***Abstract*—The telecoms industry is a highly competitive sector which is constantly challenged by customer churn or attrition. In order to remain steadfast in the consumer business, companies need to have sophisticated churn management strategies that will harness valuable data for business intelligence. Data mining and machine learning are tools which can be used by telecoms companies to monitor the churn behaviour of customers. This study implemented exploratory data analysis and feature engineering in a public domain Telecoms dataset and applied seven (7) classification techniques namely, Naïve Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees. The results are analyzed using different metrics such as Accuracy, Classification error, Precision, Recall, F1-score, and AUC. This study discussed how these results are essential in reducing customer churn and improving customer service. The results obtained in the experiment demonstrate that the best classifier is Gradient Boosted Trees. It outperforms the other classifiers in almost all evaluation metrics. Further, all classifiers showed remarkable improved performance after the oversampling method is applied.**

Keywords—business intelligence, customer churn, exploratory data analysis, feature engineering, machine learning

I. INTRODUCTION

Customer churn is a good indicator of service quality and customer service satisfaction [1][2]. The telecommunications industry is a dynamic business sector that is primarily composed of companies operating in a subscription-based model. These companies are constantly pressured with higher rates of customers who churned and shifted to rival companies that offer competitive products and services. Thus, some of them employ measures in determining the reasons why their customers churn and seek innovative strategies to improve customer satisfaction and increase the customer base.

Customer Relationship Management (CRM) is a strategic process of managing customer relations and customer retention [3][4]. Some companies mine customers' data to better understand the behaviour of their customers and gain actionable insights that help improve customer service [5]. When machine learning is embedded in a CRM software, it can track churn rates, identify churn determinants, and pinpoint customers who are at risk of churning. It can also help a company decide and employ proactive retention strategies. Several studies have agreed that it is cheaper to retain existing customers than to find new ones [6][7][8][9][10]. Hence, having a churn prediction solution allows a company to cut down its customer-related cost and further drive its sales growth.

Earlier studies have paved the work on customer churn prediction in the telecoms industry [11][12][13], However, the study of Amin et.al. [14] identified that there is still a lack of efficient customer churn prediction approaches in the telecommunications sector. Accordingly, Tsai and Chen [15] identified that there are only a few studies that perform pre-processing and feature selection in data mining. These identified gaps have led the authors to conduct an experimental study on churn prediction in the telecoms industry. The primary focus of this study is to examine the intervening effects of exploratory data analysis and feature engineering and selection in finding trends and patterns in a telecoms dataset. Moreover, this study compares the churn prediction performance of several machine learning algorithms namely, Naïve Bayes, Generalized Linear Model, Logistic Regression, Deep Learning, Decision Tree, Random Forest, and Gradient Boosted Trees on a publicly available telecoms dataset. Metrics such as Accuracy, Classification error, Precision, Recall, F1-score, and Areas Under Curve (AUC) are used to evaluate and compare the prediction performance of the classifiers. Japkowitz [16] and Weiss [17]stated that studies that relate imbalances and the performance of standard classifiers need to be conducted because it can provide beneficial results in data mining. Hence, this study attempted to use an oversampling method in the unbalanced dataset to determine whether there is an improvement in the prediction performance of each of the classifier.

This paper is organized as follows: Section 2 presents a brief related work on several relevant concepts such as exploratory data analysis, feature engineering, the machine learning classifiers used in this study, and the metrics used to evaluate the performance of the classifiers. Section 3 demonstrates the experimental setup and discussion of the results. Section 4 concludes the study.

978-1-5386-8046-9/19/$31.00 ©2019 IEEE

II. BACKGROUND AND RELATED WORK

A. Classification Methods

Churn prediction models have gained popularity and increasingly utilized in aggressive retention campaigns. According to [18], a model needs to have good accuracy in predicting future churners, comprehensible and intuitive in identifying determining factors that will lead the customer to churn, and justifiable in developing retention strategies.

A classification problem uses supervised learning which is performed in two phases: training and testing. A training dataset is utilized to construct the classification model in the training phase whose target class label is known. The resulting model is then applied in the testing phase to classify instances with unknown target class label. The succeeding statements will briefly describe the different classification algorithms. Likewise, Table 1 lists other machine learning classifiers that have been used in several kinds of literature related to churn prediction using a telecoms dataset.

1. Naïve Bayes

Naïve Bayes (NB) is a high-bias classifier which works well for datasets of any size. The basic assumption of the classifier is that the value of an attribute is independent of the value of any other attributes. This assumption simplifies the computations needed to build the model. The NB classifier is reported to achieved good results in solving a prediction problem [19].

2. Generalized Linear Model

The Generalized Linear Model (GLM) is an improved linear model that maximizes the log-likelihood [20]. This model is extremely fast and works well for models with a limited number of features having non-zero coefficients. GLM is used in classification problems with binary responses[21].

3. Logistic Regression

The Logistic Regression (LR) is a probabilistic statistical model that produces a binary prediction of a categorical variable. The predictor variable depends on one or more variables which can be nominal or numerical [3]. The study of [22] revealed that LR performed well after data transformation is applied to the data.

4. Deep Learning

Deep Learning (DL) is based on a multi-layer feed forward neural network which comprises a large number of hidden layers of neurons. The network uses back-propagation that is trained with stochastic gradient descent. DL comprises a complex artificial network which is a popular algorithm used in solving complex problems including churn prediction [3].

5. Decision Tree

A Decision Tree (DT) is a classifier having a collection of nodes whose goal is to create a tree-like decision structure of attributes relevant to the target class attribute. The DT is able to accept nominal and numerical predictor variables but the label variable needs to be nominal. Decision Trees are usually utilized in classification problems because of its interpretability and ability to produce good accuracy levels [23][24].

Table1. Studies in Churn Prediction using a telecoms dataset

Authors	Classifiers/Techniques used in Churn Prediction	Result(s)
Keramati et.al.[25]	Decision Tree, Artificial Neural Networks, K-Nearest Neighbors, and Support Vector Machine	Introduced a hybrid methodology that improved the performance of evaluation metrics
Huang et.al. [26]	Logistic Regression, Linear Classification, Naive Bayes, Decision Tree, Multilayer Perceptron Neural Networks, Support Vector Machines, and Evolutionary data mining algorithms	Used seven prediction techniques and produced a new set of features
Junxiang [27]	Survival analysis techniques	Predicted the membership duration of customers
Vafeiadis et.al. [3]	Artificial Neural Networks, Decision Trees, Support Vector Machines, Naive Bayes classifiers, and Logistic Regression classifiers,	Demonstrated a clear advantage of the boosted versions of the models vs non-boosted versions
Lemmens and Croux [28]	Bagging and Boosting Classification Techniques	Recommended the use of balanced sampling scheme in predicting a large dataset
Tsai & Chen [15]	Neural Network, Decision Tree	Applied neural networks and Decision Tree to predict customer churns as well as four evaluation measures including prediction accuracy, precision, recall, and F-measure Selected variables using association rule
Sharma & Kumar [10]	Neural Network	Neural Networks (NNs) can predict customer churn with accuracy>92% and medium-sized NNs perform best in a customer churn prediction
Tsai & Lu [29]	Artificial neural networks (ANN), Self-Organizing Maps (SOM)	The hybrid models (ANN+ANN and ANN+SOM) outperform the single neural network baseline model in terms of accuracy
Hung & Wang [30]	Decision Tree, Neural Network	Achieved good prediction accuracy using customer profile, billing and call information, and service change records

6. Random Forest

A Random Forest (RF) is an ensemble of random trees on different subsets of the dataset. Each node of the tree represents a splitting rule for an attribute in the subset. The result is based on the voting of these trees. This process minimizes overtraining. The study of [31] revealed the effectiveness of an improved balanced RF in predicting customer churn, coupled with sampling techniques and cost-sensitive learning.

7. Gradient Boosted Tree

A Gradient Boosted Tree (GBT) is a forward-learning ensemble method that achieves its prediction results through continuing improved estimations [32]. It is a machine learning algorithm that is used in classification problems. Boosting is applied to help improve the accuracy of the trees.

GBT is used by [33] because it is adaptable, interpretable, and produces highly accurate models.

B. *Evaluation Metrics*

This study will evaluate the performance of the classifiers in churn prediction using the different evaluation metrics which are derived from the result of the confusion matrix. There are four (4) possible outcomes in the matrix namely,

- True Negative (*TN*) – meaning that the class outcome was successfully predicted 0.
- False Negative (*FN*)- meaning that the class outcome was predicted 0 instead of 1.
- False Positive (*FP*) - meaning that the class outcome was predicted 1 instead of 0.
- True Positive (*TP*)- meaning that the class outcome was successfully predicted 1.

Accuracy is the percentage of predicted cases and computed by dividing the total number of correct predictions to the total number of data in the dataset. See eq. (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision is computed by dividing the number of true positives to the number of true positives and false positives. See eq. (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is computed by dividing the number of true positives to the number of true positives and the number of false negatives. See eq. (3).

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F-measure (or F1 Score) carries the balance between precision and recall. See eq. (4).

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

AUC is a good metric for binary classification because it is not sensitive to imbalanced classes. This is the advantage of AUC over accuracy. An AUC score of 1 is considered a perfect classifier. The curve is plotted using the true positive against the false positive.

III. EXPERIMENTAL STUDY

A. *Data Source and Tool Used*

Due to the exclusive nature of most telecoms data, this study will utilize a synthetic churn dataset as shown in Table 2 which is provided by Kaggle [34] and used by IBM Analytics [35]. The dataset comprises customer information and transactions including account information, service preferences, payment scheme, and others. There are 7043 records in the dataset.

Table 2 reveals that it is a mixed dataset having 21 attributes which include the nominal and numerical variables and their descriptions. The predictor variables (or features) are the attributes 1 to 20 in the table. Likewise, the target variable is the “Churn” attribute, which can hold a value of “yes”, meaning, the customer churned, or a value of “no”, meaning, the customer did not churn.

Rapidminer Studio software provides an integrated environment for data preparation, machine learning, predictive analytics, and many others [36]. The software is used in this study to produce model validation and evaluation and result visualization using tables, pie charts, and histogram.

B. *Modelling Approach*

The modelling approach is illustrated in Fig. 1. It starts with exploratory data analysis to visually summarize the data and find patterns and data anomalies, followed by data pre-processing to format the incomplete, inconsistent, and/or missing data, followed by feature engineering to select features that most likely influence the target class attribute. Once all of these steps are undertaken, model building follows using Rapidminer’s Automodel feature which builds the seven (7) classification models. Finally, these models will be compared using the aforementioned evaluation metrics.

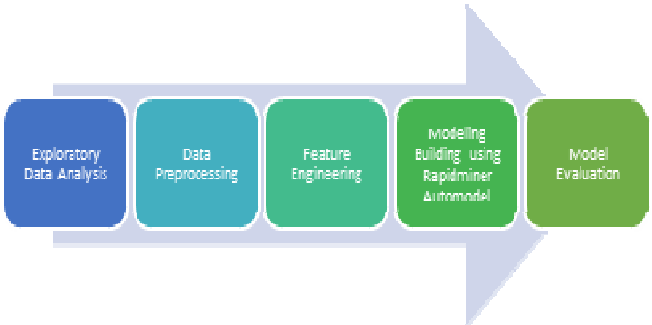


Fig. 1 Modelling Approach

1) *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) using univariate frequency analysis was conducted to describe key characteristics of each attribute including, minimum and maximum value, average, standard deviation and others. It was also used to produce a value distribution and identify missing values, and outliers.

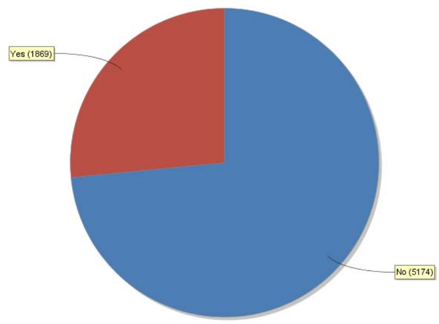


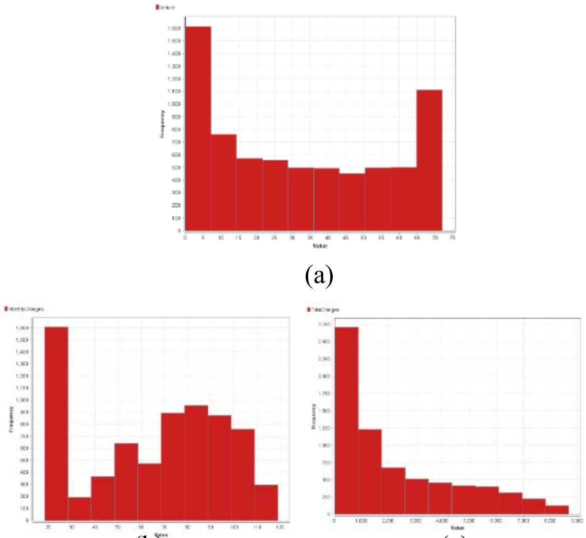
Fig. 2 Churners vs non-churners

As shown in Fig. 2, there are 5174 instances of customers *(in blue color)* who did not churn and only 1869 instances of customers *(in red color)* who churned. The dataset is highly biased towards the non-churned class. This study will attempt to resolve the issue using an oversampling method which is discussed in the succeeding section.

Table 2. Metadata for the Telecoms dataset

	Attribute Name	Description	DataType	Values
1	customerID	Customer ID	Nominal	7590-VHVEG, 5575-GNVDE, Etc.
2	gender	Customer gender	Nominal	Female, male
3	SeniorCitizen	Whether the customer is a senior citizen or not	Nominal	1, 0
4	Partner	Whether the customer has a partner or not	Nominal	Yes, No
5	Dependents	Whether the customer has dependents or not	Nominal	Yes, No
6	tenure	Number of months the customer has stayed with the company	Numeric	10,16, etc.
7	PhoneService	Whether the customer has a phone service or not	Nominal	Yes, No
8	MultipleLines	Whether the customer has multiple lines or not	Nominal	Yes, No, No phone service
9	InternetService	Customer's internet service provider	Nominal	DSL, Fiber optic, No
10	OnlineSecurity	Whether the customer has online security or not	Nominal	Yes, No, Internet service
11	OnlineBackup	Whether the customer has an online backup or not	Nominal	Yes, No, Internet service
12	DeviceProtection	Whether the customer has device protection or not	Nominal	Yes, No, Internet service
13	TechSupport	Whether the customer has tech support or not	Nominal	Yes, No, Internet service
14	StreamingTV	Whether the customer has streaming TV or not	Nominal	Yes, No, Internet service
15	StreamingMovies	Whether the customer has streaming movies or not	Nominal	Yes, No, Internet service
16	Contract	The contract term of the customer	Nominal	Month-to-month, One year, Two year
17	PaperlessBilling	Whether the customer has paperless billing or not	Nominal	Yes, No
18	PaymentMethod	The customer's payment method (Electronic check, Mailed check, Bank transfer	Nominal	Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)
19	MonthlyCharges	The amount charged to the customer monthly	Numeric	29.85, 16.23, etc.
20	TotalCharges	The total amount charged to the customer	Numeric	29.85, 1899.95, etc.
21	Churn	Whether the customer churned or not	Nominal	Yes or No

The distribution of values in continuous numerical attributes for Tenure, MonthlyCharges, and Totalcharges attributes are shown in Fig. 3. As noticed, the distribution for the three (3) attributes is skewed because of the varying range of values.



(a) (b) (c)

Fig. 3 Skewness of (a) Tenure, (b) MonthlyCharges, and (c) TotalCharges attributes

2) *Data Preprocessing*

A dataset may contain noise, missing values, and inconsistent data, thus, pre-processing of data is essential to improve the quality of data and time required in the data mining. Utilizing raw data in building predictive models is not desirable because it would produce weak results. Further, some machine learning algorithms are also not sophisticated enough in extracting meaningful information. The data pre-processing techniques in this study includes data cleaning, integration, discretization, and reduction.

a. Data Cleaning

This experiment used the Filter Examples Operator which removed the examples with missing values. The dataset contains 11 missing values in the TotalCharges column which were excluded, hence, they were deleted from the dataset. The new dataset contains 7032 records with 5163 non-churners and 1869 churners.

b. Data Transformation

Data transformation is the process of normalizing and aggregating the data to further improve the efficiency and accuracy of data mining. The nominal-to-binomial operator is used to transform and map the selected nominal value to its equivalent binomial value. The attributes that were transformed from nominal to binomial are Gender, SeniorCitizen, Partner, Dependents, PhoneService, PaperlessBilling, and Churn.

c. Data Reduction

Data reduction is the process of reducing data representation while still producing similar results. Discretization (or binning) is applied to numeric attributes by transforming and grouping the continuous values into discrete categories (or interval levels) because some data mining algorithms only accept nominal attributes and are not able to handle numeric attributes. Likewise, using raw numeric data into the model can adversely affect the performance of the model.

Table 3 Tenure

Index	Nominal value	Absolute count	Fraction
1	range2 [14.500 - 47.500]	2369	0.337
2	range1 [-∞ - 14.500]	2360	0.336
3	range3 [47.500 - ∞]	2303	0.328

Table 4 Monthly Charges

Index	Nominal value	Absolute count	Fraction
1	range1 [-∞ - 50.425]	2345	0.333
2	range2 [50.425 - 84.025]	2345	0.333
3	range3 [84.025 - ∞]	2342	0.333

Table 5 Total Charges

Index	Nominal value	Absolute count	Fraction
1	range1 [-∞ - 678.325]	2344	0.333
2	range2 [678.325 - 2745....	2344	0.333
3	range3 [2745.450 - ∞]	2344	0.333

Using the Discretize by Binning operator, the Tenure, MonthlyCharges, and TotalCharges feature values are assigned into specific bins. The operator automatically generated bins of equal ranges and the absolute count in the different bins are shown in Tables 3-5.

3) Feature Engineering

Feature engineering allows raw data to be transformed into features to help improve the overall performance of predictive models. In most cases, the number of features needs to be decreased in order to increase the accuracy of the algorithm. The features that may be influential to the target attribute are retained while the features that likely overfit the models are discarded. Feature selection, which reduces the number of predictors, is applied. Doing so decreases the error and increases the accuracy.

Feature Selection is performed using the Correlation Matrix Operator which produces a pairwise table of correlation coefficient. It is used to determine the correlation between all attributes. Generally, a correlation coefficient above 0.75 is strong, between 0.45 and 0.75 is moderate, and below 0.45 is weak. This experiment computed the correlation of the different features to discard features with high correlation. All features having a correlation coefficient of greater than or equal to 0.45 are listed in Table 6.

Table 6. Correlated Attributes

First Attribute	Second Attribute	Correlation
Tenure	TotalCharges	0.826
MonthlyCharges	TotalCharges	0.651

The features that will not contribute to the predictive power will be discarded. The following features are:

- Customer ID attribute will not be included because it will not contribute to the prediction.
- TotalCharges attribute will be discarded as it highly correlates with both Tenure and MonthlyCharges attributes and this attribute can be derived from other attributes.

Rapidminer software provides the relevant features which are important to the target attribute. The weights are computed using the Pearson correlation coefficient and displayed in Table 7. From the table, the top five (5) key drivers that will likely influence a customer churn are Contract, OnlineSecurity, TechSupport, Tenure, and DeviceProtection.

Table 7. Weight

Attribute	Weight
Contract	1
OnlineSecurity	0.836
TechSupport	0.828
Tenure	0.803
DeviceProtection	0.703
PaymentMethod	0.656
StreamingMovies	0.512
StreamingTV	0.508

4) Model Building

This study used the Automodel feature in Rapidmner Studio which provides a fast process of building and validating the models. The model, as shown in Fig. 4, comprises several data operators including preprocessing, replace missing values, reorder attributes, filter examples, stratified sampling, split data, cross-validation, model simulator and many others. First, the Automodel performs a guided data pre-processing, and feature engineering and selection. These processes have been undertaken in the early part of this experiment; it is redundancy but nonetheless useful. Second, it splits the data using a stratified sampling approach and performs cross-validation. Finally, it uses a model simulator that will display the results of the prediction

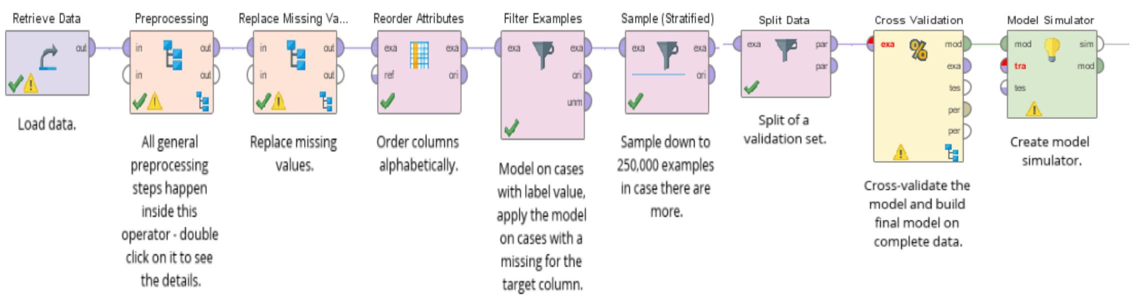


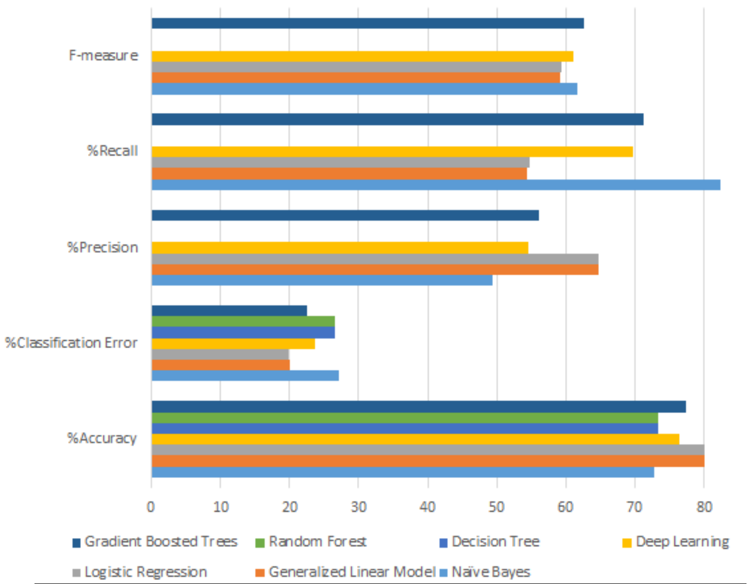
Fig. 4 The process model in Rapidminer

5) *Model Evaluation*

This section will compare the performance of the different classifiers to select the best method in predicting customer churn. The output of the model prediction is analyzed using accuracy, precision, recall, F-score, and AUC values which are described in the succeeding paragraphs. These values will be compared against the seven models. As revealed in Fig. 5, GLM and LR achieved the highest accuracy prediction of 80.0% and 80.1%, respectively given a churn rate of 26%, whereas, NB performed the least accuracy prediction of 72.8% of the same churn rate.

Accuracy is a good metric to use for a non-biased dataset, however, this study used a biased churn dataset. Hence, we employed a sampling technique to improve the prediction performance of the classifiers. The dataset, which has 26% churn rate, is acceptable in the industry standard of 80/20 churn distribution [38]. However, we used an oversampling method to further improve the prediction performance. We resampled the minority class in an attempt to balance the class distribution. In this study, the churn rate was doubled from 26.6% churn rate (with 1869 churners and 5163 non-churners) to 41% churn rate (with 3738 churners and 5163 non-churners).

The results indicate that GBT achieved the highest accuracy prediction of 79.1% (an increase of almost 2%) after oversampling was applied. This supports the study of [37] as regards the high performance of gradient boosted trees for binary classification. Notice that when oversampling was applied to achieve a near-to balanced class distribution in the dataset (churn rate=41%), the accuracy of RF and GBT increased whereas the accuracy of NB, GLM, LR, DL, and DT decreased. The precision, recall, and F1-score of all classifiers remarkably increased after oversampling was applied. Their values are more than 60% which is considered satisfactory.



Model	%Accuracy		%Precision		%Recall		F1 score	
ChurnRate	26%	41%	26%	41%	26%	41%	26%	41%
NB	72.8	73.0	49.3	63.6	82.3	83.9	61.6	72.3
GLM	80.0	75.7	64.7	70.4	54.4	72.8	59.1	71.6
LR	80.1	75.7	64.8	70.4	54.8	72.9	59.3	71.6
DL	76.4	74.3	54.5	66.0	69.8	80.9	61.1	72.6
DT	73.4	70.7	-	61.1	0	84.6	-	70.7
RF	73.4	75.2	-	70.6	0	70.5	-	70.5
GBT	77.4	79.1	56.0	73.1	71.2	79.6	62.6	76.2

Fig.5. Comparison of Model Performance

However, good accuracy alone is not sufficient to conclude the performance of a classifier. This study also looked at the other evaluation metric such as AUC. It is a better evaluation metric as compared to accuracy [39] when used in an imbalanced class problem. Fig. 6 and Fig. 7 illustrate the AUC of the different classifiers before and after sampling.

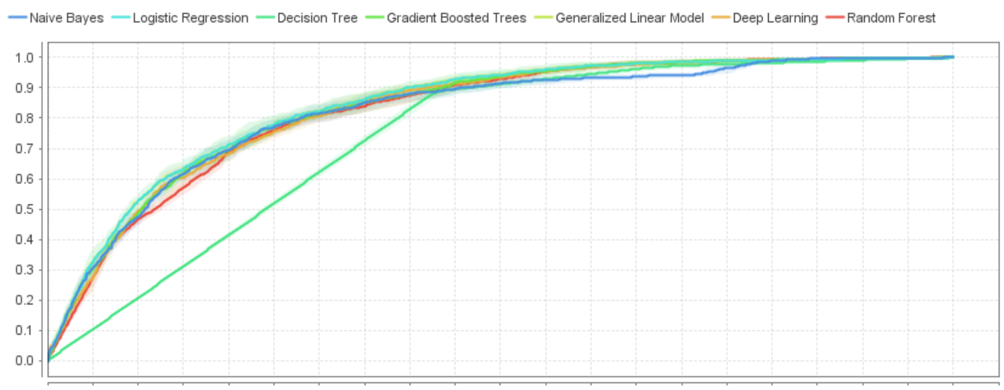


Fig. 6. ROC Comparison - Before Oversampling

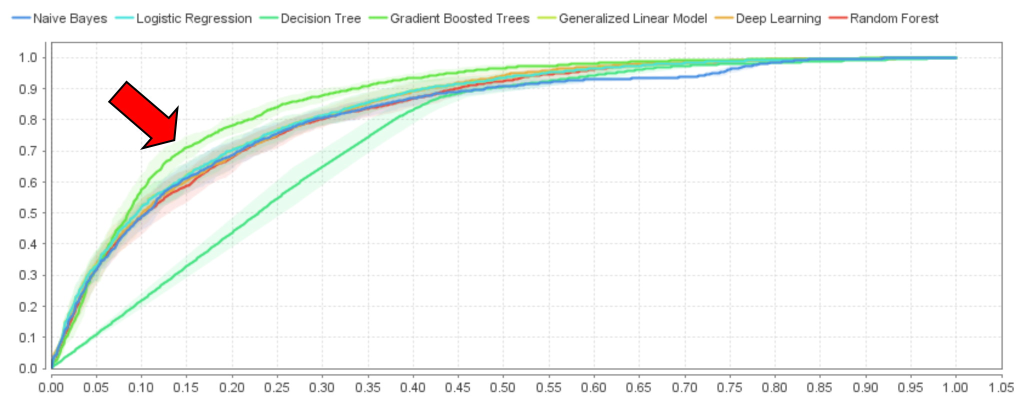


Fig. 7 ROC Comparison- After Oversampling

Table 8 AUC before and after oversampling

Model	*AUC*		
	Churn Rate = 26%	Churn Rate = 41%	%Difference
NB	0.822	0.818	~0%
GLM	0.841	0.837	~0%
LR	0.841	0.837	~0%
DL	0.830	0.834	~0%
DT	0.739	0.745	+1%
RF	0.825	0.827	~0%
GBT	0.834	***0.865***	+3%

A closer inspection of Fig. 7 *(see red arrow)* and Table 8 reveal that there is almost no significant improvement in the AUCs of these classifiers except for the DT and GBT classifiers where there are slight improvements in their AUC of +1% and +3%, respectively.

IV. CONCLUSION

This study applied exploratory data analysis and feature engineering on a telecoms dataset and use these techniques to improve the performance of seven classifiers for churn prediction. This study also discussed the prediction results of the different classifiers and the results revealed that all classifiers have achieved more than 70% accuracy. It also examined the use of oversampling the minority class. The experiment results demonstrated that GBT outperformed the other classifiers in almost all evaluation metrics. Moreover, all of the classifiers achieved a remarkably improved performance when the oversampling method was applied. Finally, this study recommends the GBT as an algorithm of choice for churn prediction modelling.

REFERENCES

[1] J. Sweeney and J. Swait, "The effects of brand credibility on customer loyalty," J. Retail. Consum. Serv., 2008.

[2] A. Gustafsson, M. D. Johnson, and I. Roos, "The Effects of Customer Satisfaction , Dimensions , and Triggers on," J. Mark., 2005.

[3] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," Simul. Model. Pract. Theory, 2015.

[4] K. Coussement, D. F. Benoit, and D. Van den Poel, "Improved marketing decision making in a customer churn prediction context using generalized additive models," Expert Syst. Appl., 2010.

[5] C. H. Cheng and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," Expert Syst. Appl., 2009.

[6] M. A. Khan, M. A. I. Khan, M. Aref, and S. F. Khan, "Cluster & rough set theory based approach to find the reason for customer churn," Int. J. Appl. Bus. Econ. Res., 2016.

[7] A. T. Jahromi, S. Stakhovych, and M. Ewing, "Managing B2B customer churn, retention and profitability," Ind. Mark. Manag., 2014.

[8] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," Expert Syst. Appl., 2011.

[9] P. Kisioglu and Y. I. Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey," Expert Syst. Appl., 2011.

[10] A. Sharma and P. Kumar Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," Int. J. Comput. Appl., 2011.

[11] J. H. Ahn, S. P. Han, and Y. S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," Telecomm. Policy, 2006.

[12] C. P. Wei and I. T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," Expert Syst. Appl., 2002.

[13] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," Eur. J. Oper. Res., 2012.

[14] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, and K. Huang, "Customer churn prediction in the telecommunication sector using a rough set approach," Neurocomputing, 2017.

[15] C. F. Tsai and M. Y. Chen, "Variable selection by association rules for customer churn prediction of multimedia on demand," Expert Syst. Appl., 2010.

[16] N. Japkowicz, "Learning from Imbalanced Data Sets: A Comparison of Various Strategies *," in In Proceedings of the 2000 International Conference on Artificial Intelligence, 2000.

[17] G. M. Weiss, "Mining with Rarity: A Unifying Framework," SIGKDD Explor. Newsl., 2004.

[18] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Syst. Appl., 2011.

[19] S. V Nath and R. S. Behara, "Customer Churn Analysis in the Wireless Industry: A Data Mining Approach," Proceedings-Annual Meet. Decis. Sci. Inst., 2003.

[20] J. A. Nelder, R. W. M. Wedderburn, A. J. a Nelder, and R. W. M. Wedderburn, "Generalized Linear Models," J. R. Stat. Soc. A., vol. 135, no. 3, pp. 370–384, 1972.

[21] T. J. Hastie and D. Pregibon, "Generalized linear models," in Statistical Models in S, 2017, pp. 195–247.

[22] D. Radosavljevik, P. Van Der Putten, and K. K. Larsen, "The Impact of Experimental Setup in Prepaid Churn Prediction for Mobile Telecommunications: What to Predict, for Whom and Does the Customer Experience Matter?," Trans. Mach. Learn. Data MiningJournal), 2010.

[23] R. C. Barros, M. P. Basgalupp, A. C. P. L. F. de Carvalho, and A. A. Freitas, "Towards the Automatic Design of Decision Tree Induction Algorithms," in Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation, 2011.

[24] K. H. Liao and H. E. Chueh, "Applying Fuzzy Data Mining to Telecom Churn Management," in Communications in Computer and Information Science, 2011.

[25] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," Appl. Soft Comput. J., 2014.

[26] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," Expert Syst. Appl., 2012.

[27] L. Junxiang, "Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS," Techniques, 2002.

[28] A. Lemmens and C. Croux, "Bagging and Boosting Classification Trees to Predict Churn," J. Mark. Res., 2006.

[29] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks," Expert Syst. Appl., 2009.

[30] S.-Y. Hung, D. C. Yen, and H. Wang, "Applying data mining to telecom churn," Expert Syst. Appl., 2006.

[31] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Syst. Appl., 2009.

[32] "Rapidminer Documentation," 2018. .

[33] J. Ye, J.-H. Chow, J. Chen, and Z. Zheng, "Stochastic gradient boosted distributed decision trees," in Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09, 2009, p. 2061.

[34] BlastChar, "Telco Customer Churn Focused customer retention programs," Kaggle, 2017. .

[35] McKinley Stacker IV, "Using Customer Behavior Data to Improve Customer Retention," 2015. .

[36] M. Hofmann;, R. Klinkenberg;, M. Hofmann, and R. Klinkenberg, RapidMiner: Data Mining Use Cases and Business Analytics Applications. 2013.

[37] L. Li, Y. Wu, and M. Ye, "Experimental comparisons of multi-class classifiers," Inform., vol. 39, no. 1, pp. 71–85, 2015.

[38] S. Dziaugyte and M. Mzyk, "Churn Analysis," Bloomington, 2016.

[39] C. Xing and C. Li, "Data Mining for Direct Marketing: Problems and Solutions," in Proceedings of International Conference on Knowledge Discovery from Data, 1998, pp. 73–79.