

An Optimal Ensemble Classification for Predicting Churn in Telecommunication

J.Beschi Raja* and S. Chenthur Pandian

Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India

Received 21 December 2018; Accepted 14 December 2019

Abstract

Churn Prophecy is the most salient sphere in the telecom industry for planning, decision making and to know the future events. Due to the tremendous amount of data, prediction has been always a tangled work. The prominent objective of churn Prophecy model is to strongly classify the customers into clusters and predict the prospective churners. Efficient feature selection is also required for predicting the churners. In this paper, an overview of recent machine learning works for customer churn prediction has provided. This paper provides XG boost classifier which less focused in the previous works. XG boost classifier is applied on publicly available telecom dataset and experiential results are compared with KNN and Random Forest Classifiers. XG boost classifier performs superior out of three classifiers mentioned. The evaluated metrics such as Precision, Recall, F1-score, Support are calculated. Finally, it is observed that among all the features, Fiber Optic customers with greater monthly chargers has higher cognition for churn using XG boost classifiers.

Keyword: Customer churn, Telecommunication, XG boost classifier, Classification, Churn Prediction

1. Introduction

Rapid increase of telecom providers rises to tremendous competition and leads to churn [1][2][3]. Churn is the process of customers switching from one firm to another in given time [22] [19]. The pre-eminent reason for churn is the customer dissatisfaction [1] [24]. Retaining the existing customers is more profitable than fetching the new customers [9]. So, the telecom providers have to maximize concentration to the extant customers to avert churn [7] [8]. Further, long standing satisfied customers provoke more profit because they aren't going to churn [13] [17]. The customer churn leads to huge loss economically and affect company's status [28]. An appropriate churn prediction model is essentially needed to predict the churners [6] [23]. Many companies are concentrating on customers individual for giving best services to them. Churn is the inevitable action in all industry [1]. In recent past, various machine learning methods are used to tackle threat in customer churn in telecom [11] [12]. Lu et al.in (2014) suggested Gentle Ada boost algorithm [22] for churn prediction, because it reaches high accuracy and provided good separation of churn models and also a comparison is made with the results of logistic regression and single logistic regression model, which proves that the logistic regression model is the better one. Wenjie et al.in (2016) proposed Semantic-Driven Subtractive Clustering Method (SDSCM) which is proven to have more clustering strength than subtractive clustering method (SCM) and fuzzy c-means (FCM). Hadoop Map Reduce frame work is used to implement this algorithm. [28]. A.keramati et al.in (2014) compared the performance of DT, ANN, KNN, SVM and shown that hybrid of all four performs well, they also proposed a methodology for extracting the influential features from the dataset[18]. Pinar kisioglu et al.in (2010) applied Bayesian belief networks to explain the causal relationships

between the features that contributes to customer churn and the important features were found out and suggested promotions to reduce the churn rate [20]. Koen W.De Bock et al.in(2012) Proposed GAMensplus classification algorithm for strong classification and interpretability. The system also demonstrated the relationship between the predictors and probabilities of churn [16]. Muhammad Azeem et al.in (2018) compared the accuracy of several classifiers with fuzzy model and proved that fuzzy classifiers are more accurate in predicting customer churn dataset having noise [4]. Chih-Fong Tsai et al.in (2009) used neural networks and hybridised data mining technique for predicting customer churn with higher accuracy, this system combines ANN with ANN and ANN with SOM for churn higher prediction accuracy [25]. Wouter Verbeke et al.in (2010) proposed AntMiner+ and ALBA for improving the learnability providing comprehensible and accurate customer churn prediction model [27]. Zhen-Yu Chen et al.in (2012) proposed Novel approach known as hierarchical multiple kernel support vector machine (HMK-SVM) combines static customer and longitudinal behavioural data to improve churn prediction, a three-phase training algorithm was implemented. [5]. Long Zha et al.in (2017) proposed New K- local maximum margin feature extraction algorithm (KLMM), they followed the fact that extracting the features of the data will reduce the dimensionality of prediction [29]. T. Vafeiadis et al.in (2015) made a comparative study on a public domain dataset, next deals with improving the accuracy by using boosting algorithms. They used Monte Carlo simulation at different settings of each classification method and improved SVM-POLY with Ada boost classifier which gave higher accuracy [26]. Ying Huang et al.in (2013) proposed Novel hybrid model-based learning system, which integrates k-means clustering algorithm (unsupervised) and inductive technique (FOIL)-(supervised) for building an effective predictive model [10]. The rest of the paper contains the following session

*E-mail address: ljpamina@gmail.com

ISSN: 1791-2377 © 2020 School of Science, IITU. All rights reserved.

doi:10.25103/jestr.132.07

2. Machine Learning Classifiers

2.1. KNN Classifier

K nearest neighbour is widely used classification method. KNN classifier classifies new cases based on the similarity measures of the available cases. The similarity measure corresponds to any functions such as distance. A case is classified based on the majority of neighbourhood data points. The distance between the datapoints is calculated using the Euclidean distance function.

$$D = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Where D is the distance function and x and y correspond to the datapoints.

The above distance function is applied only for continuous variables, categorical variables will use Hamming distance.

$$D_H = \sum_{i=1}^K |x_i - y_i|$$

In the above equation if we the value of x and y are equal then the distance function will become zero, otherwise the value of D_H is one. These criteria can be tabulated as follows

X	Y	Distance
M	M	0
M	F	1

2.2. Random Forest Classifier

Random forest Classifier is one of the ensemble learning method and supervised learning algorithm which is used for both classification and regression [14] [15]. This Classifier generates a group of decision trees from the part of training set randomly selected. For each sample it creates a decision tree and predict the results from it. The higher number of trees gives the greater prediction accuracy. It then forms the combination and average of all the prediction votes from various decision trees to finalise the output.

The random forest algorithm is split into two stages namely

1. Creation of Random forest
2. Performing prediction from the created random forest.

Creation of Random forest

The pseudo code for the creation of random forest is as follows:

1. Select “k” features from total “m” features randomly.
2. using the best split point calculate the node “d” among the “k” features.
3. Split the node into child nodes by making use of the best split.
4. Steps 1 to 3 is repeated until “i” number of nodes has been reached.
5. The forest is built by repeating the steps 1 to 4 for “n” number times to create “n” number of trees.

Random forest prediction

The trained random forest algorithm uses the below pseudo code to perform prediction.

1. Take the test features and pass the features through the rules of each decision tree created randomly to predict the outcome. Then store the predicted outcome (target).
2. The votes for each predicted target are calculated.
3. The high voted predicted target is considered as the final prediction from the random forest algorithm.

3. XGBoost classifier

XG Boost stands for extreme Gradient Boosting. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. It is a supervised and ensemble learning method that combines trees to give a more generalizable Machine Learning model boost classifier generally built many numbers of trees and averaging the results for higher prediction. The trees are generally built one after another, in which the errors can be refined and corrected in the previous tree and focus on the improvement of prediction accuracy in next tree. Due its cache good optimization, the XG boost classifier makes good results in prediction model but it takes more training time for iteration process. XG boost classifiers handles the sparse dataset in which the missing values are handled properly. XG boost algorithm gives very good training model and performance for prediction of accuracy. The XG boost algorithm addresses one of the key problems in tree learning which is to find the best split. For this purpose, the XG boost algorithm uses the exact greedy algorithm. The exact greedy algorithm is shown in Algorithm 1.

Algorithm 1: Exact Greedy Algorithm for Split Finding

Input: I, instance set of current node

Input: d, feature dimension

gain ← 0

G ← ∑_{i∈I} *g*_i, *H* ← ∑_{i∈I} *h*_i

for *k* = 1 to *m* do

*G*_L ← 0, *H*_L ← 0

for *j* in sorted (*I*, by *x*_{*j**k*}) do

*G*_L ← *G*_L + *g*_{*i*}, *H*_L ← *H*_L + *h*_{*i*}

*G*_R ← *G* - *G*_L, *H*_R ← *H* - *H*_L

score ← max (*score*, $\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{G^2}{H + \lambda}$)

end

end

Output: Split with max score

4. Evaluation Measures

The classifiers Churn prediction performance for various appropriate attributes.

Precision is measure of correct positive cases predicted and evaluated by the following calculation.

$$precision = \frac{TP}{TP + FP}$$

Recall is the measure of identified correct positive cases and evaluated by the following calculation

$$recall = \frac{TP}{TP + FN}$$

Only precision or recall cannot elucidate the performance of the classifier, both measures must imply good

performance. As a consequence of this phenomena, the average of precision and recall is calculated as f- measure used as measure for predicting performance of classifier. F_1 score attain its best performance at 1 and worst at 0.

$$F - measure = \frac{2 * precision * recall}{Precision + recall}$$

The above three evaluation measures are used for KNN, RF and XGBoost classifier.

5. Simulation Setup

Our aim is to analyse the important classifiers of churn prediction. To that end, we implemented our simulation in two steps using the python language and for Analysis Libraries, pandas and keras have been used. In the first step, all tested classifiers are initially applied on the dataset. Their performance is evaluated using the accuracy score and F - measure criteria.



Fig. 1. Overall Process of KNN, RF and XGBoost classifier

Generally, the process involved in the classifiers have four stages. Data access is nothing but giving input to the model which is the first process for analytics. Data Wrangling is a process for selecting the distribution of data which we use certain data for analysis. Training is implemented by building the models with three classifiers and test it for better accuracy. The last process is acquiring Insights in which final output are predicted and understanding the results and parameters for good prediction.

Dataset

Telecom dataset is not publicly available more due to its customer’s personal privacy. Data set for this paper obtained from IBM Watson dataset released on 2015. The data set contains 7043 instances and 21 attributes. The last attributes denote churn or not in which 5174 are not churners and 1869 are churners. The percentage of churners is 26.53% and non-churners is 73.46%. This dataset helps to figure out customer Prophecy and build retention possibilities.

Data Pre-processing

Data Pre-Processing is the one of the basic and important work in the sphere of machine learning. In this study, two tasks were carried out for pre-processing the data. Firstly, we have performed the missing value analysis and it is found that there no missing value present in the dataset. Next, the conversion of data types ie. Object data type to numerical data type is implemented. Total charges attribute is converted to numerical data type from object data type for the classifiers we used for analysis. We have also checked for duplication of customer id and it shows that all the rows are unique.

a. Univariate analysis

The univariate analysis is very simplest task performed for analysing the data. In this work, we used the univariate analysis for understanding the unique values of each and

every features distribution to observe the objects and non-object data points.

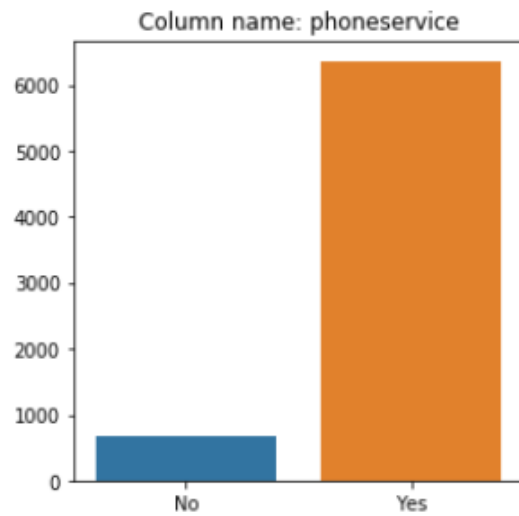


Fig. 2. Phone Services

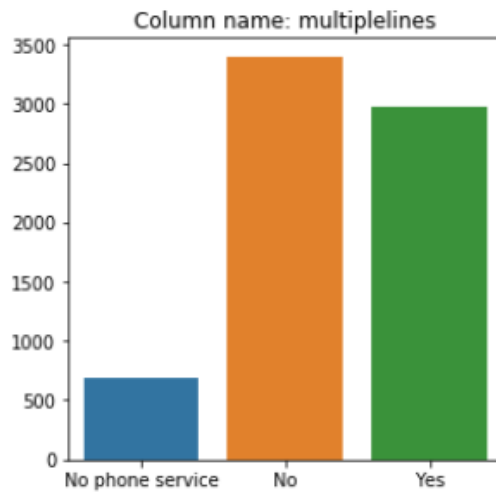


Fig. 3. Multiple Lines

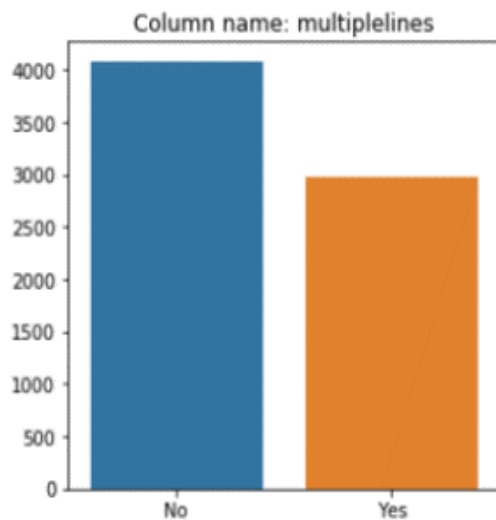


Fig. 4. Combination

Using univariate analysis, the features are normalized based on the value of subscriber’s services, there are no phone

service, no and yes. Here phone service and multiple lines are primary and secondary attributes respectively. Fig. 2 represents the customers having phone services or not in the dataset. Fig. 3 indicates the number of customers having multiple lines or not and no phone service. Fig. 4 represents the outcome of univariate analysis which aggregates no phone service and no multiple lines.

6. Results and Discussion

a. Data Correlation

Before building the churn, prediction model using classifiers of KNN, RF and XGBoost the correlation among the features are identified and sorted in the descending order. Above table indicates the correlation values of the attributes and it is plotted as a graph (Fig 5.1) Based on this assumption, it shows clearly that there is a strong correlation between the attributes churn and Month-to-month contracts. Fibre Optic ISP and monthly charges are the second and third attributes that shows strong correlation.

Table 1. Features and its correlation values

Attributes	Correlation values
churn	1.000000
contract_Month-to-month	0.405103
ISP_Fiber_optic	0.308020
payment_Electronic_check	0.301919
Monthlycharges	0.193356
paperlessbilling	0.191825
Seniorcitizen	0.150889
Streamingtv	0.063228
streamingmovies	0.061382
Multiplelines	0.040102
Phoneservice	0.011942

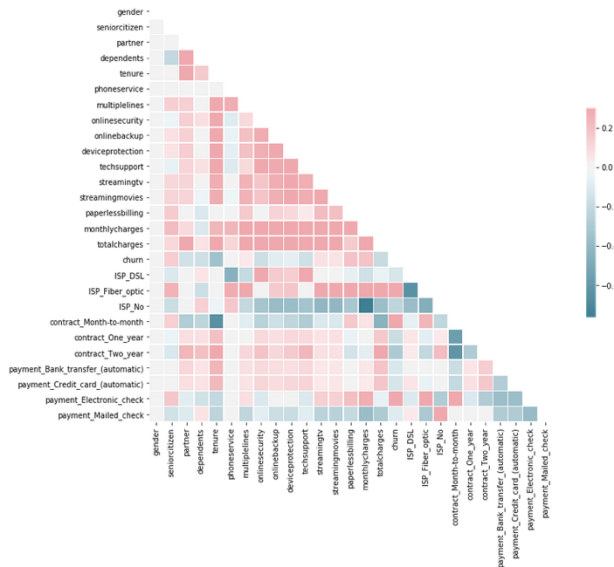


Fig. 5. Correlation among features

b. Comparative Analysis

The churn prediction model is created by three well known classifiers namely KNN, RF and XGBoost. The comparison of evaluation metrics such as accuracy score and F score are calculated. We have calculated the F score because accuracy alone is insufficient for higher churn prediction model. Experimental results show (Table2) that XGBoost classifier gives higher accuracy score and F score compared to KNN

and RF classifiers. Thus, XGBoost classifier is used for feature selection in the proposed system.

Table 2. Results Summary of KNN, RF and XGB

Classifier	Accuracy score	F score
KNN	0.754	0.495
RF	0.775	0.506
XGB	0.798	0.582

c. Implementation of Xgboost

Binary Logistic regression is a baseline algorithm used here. We have used XGBoost classifier in the rest of the work for feature selection of customer churn with same data set. Memory efficiency and the training speed is handled in XGBoost by an internal data structure called DMatrix. Based on F Score and taking 15 different features, the feature importance values are visualized as Fig.5.

Table 3. Confusion Matrix

	Predicted False	Predicted True
Actual False	1234	330
Actual True	150	399

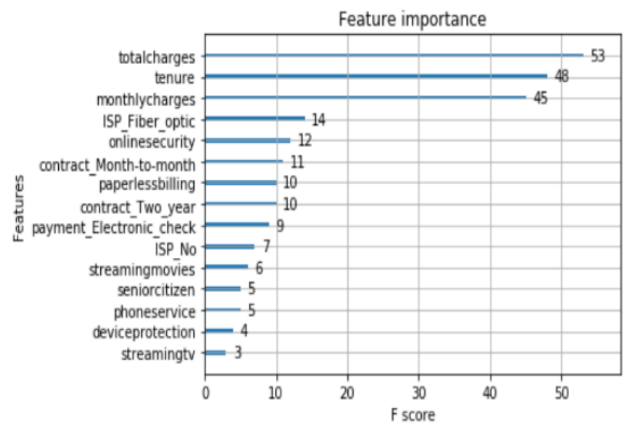


Fig. 6. Feature Importance

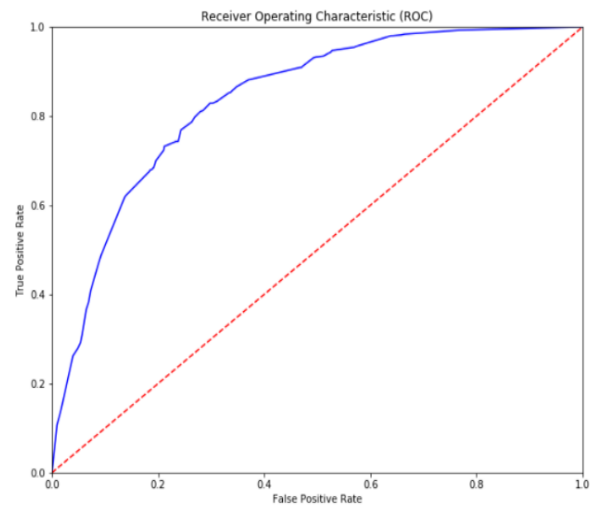


Fig. 7. ROC

Table 3 the confusion matrix calculated and it is identified that 1234 customers falls under non-churners and 399 customers falls under churners. From this, we observe that

number of non-churners are more than churners. So, our classification accuracy is correctly tested. Fig 7 displays the ROC curve of XG Boost classifier. ROC curve is mostly used to measure the test's ability as a criterion. Generally, final model accuracy is predicted using ROC curve. In the figure 6, the curve is stretched to true positive area which provides better prediction rates. The packages plot tree and graphviz are used to construct the decision tree. The plot_tree function takes best_xgb (Best XGB value), num_trees (Number of Decision trees) as parameters. The decision tree is plotted as figure 8. SHAP package is used for better visualization of XG boost classifier.



Fig. 9. Pruned Decision Tree

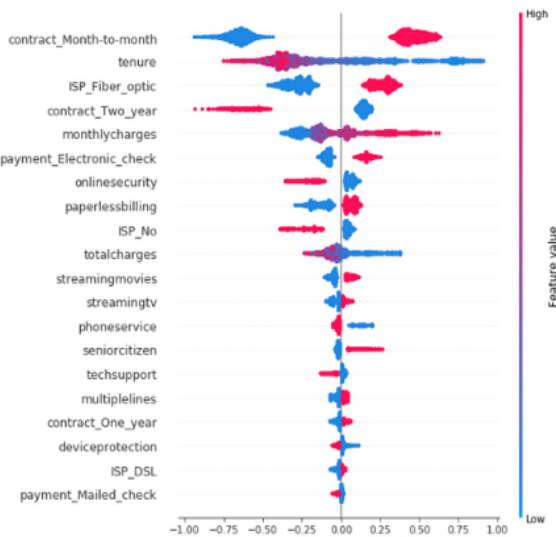


Fig. 9. Feature Importance

We have used TreeExplainer to the XGBoost model object. After training the model, we got SHAP values. For clear explanations of entire dataset and to visualize which feature is influencing the output, the data were rotated and stacked. We also displayed the features which are influencing the output. For the Tree SHAP implantation, we have focused on the margin output model not the actual or predicted output. This indicates values units are log odd ratios. Subscriber who has large positive values are likely to churn. Rest of the

attributes are used as feature importance plot in above figures. Finally, it is observed that Fiber Optic customers with greater monthly chargers are predicted as churners.

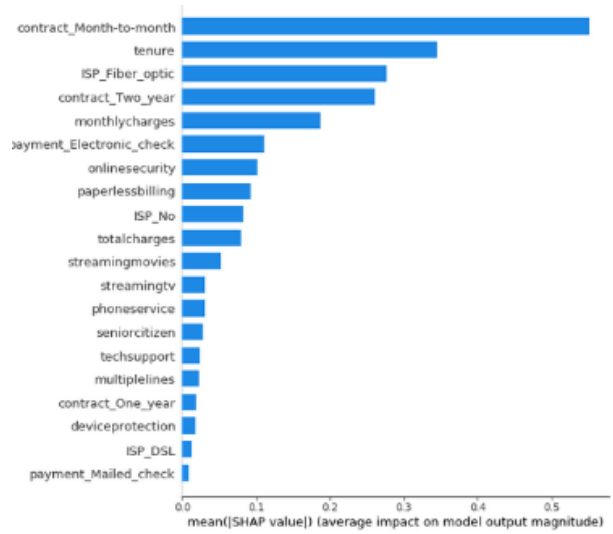


Fig. 10. Feature Importance

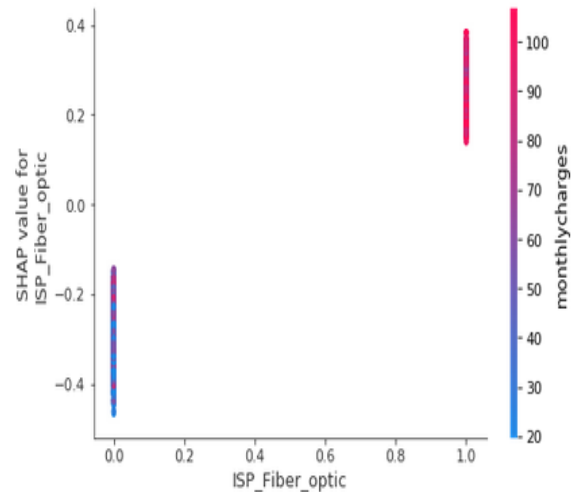


Fig. 11. ISP Fiber Optic

7. Conclusion

A Comparison study has made with three prominent classifiers namely K-NN, Random Forest and XG boost for improving accuracy of customer churn prediction. The XG boost classifier performs well out of three on the publicly available telecom dataset. Next, the work was focused to identify the attribute which has higher cognition for churn using XG boost classifier. The experiment result shows that Fiber Optic customers with greater monthly chargers has higher influence for churn. Anticipated directions can be to predict by hybrid of classifiers which gives high accuracy and desirable results.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License



References

1. Ahmed, Ammar & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal*. 18. 10.1016/j.eij.2017.02.002.
2. Amin, Adnan & Al-Obeidat, Feras & Shah, Babar & Adnan, Awais & Loo, Jonathan & Anwar, Sajid. (2018). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*. 10.1016/j.jbusres.2018.03.003.
3. Amin, Adnan & Shehzad, Saeed & Khan, Changez & Ali, Imtiaz & Anwar, Sajid. (2015). Churn Prediction in Telecommunication Industry Using Rough Set Approach. *Studies in Computational Intelligence*. 572. 83-95. 10.1007/978-3-319-10774-5_8.
4. Azeem, Muhammad & Usman, Muhammad & C. M. Fong, A. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*. 66. 10.1007/s11235-017-0310-7.
5. Chen, Zhen-Yu & Shu, Peng & Sun, Minghe. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*. 223. 461-472. 10.1016/j.ejor.2012.06.040.
6. Chu, Bong-Horng & Tsai, Ming-Shian & Ho, Cheng-Seen. (2007). Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems*. 20. 703-718. 10.1016/j.knsys.2006.10.003.
7. De Bock, Koen & Van den Poel, Dirk. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Syst. Appl.*. 38. 12293-12301. 10.1016/j.eswa.2011.04.007.
8. De Caigny, Arno & Coussement, Kristof & De Bock, Koen. (2018). A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees. *European Journal of Operational Research*. 269. 10.1016/j.ejor.2018.02.009.
9. Elango, Sivasankar & Vijaya, J. (2018). Hybrid PFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. *Neural Computing and Applications*. 10.1007/s00521-018-3548-4.
10. Huang, Ying & Kechadi, Tahar. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*. 40. 5635-5647. 10.1016/j.eswa.2013.04.020.
11. Huang, Bingquan & Buckley, B & Kechadi, Tahar. (2010). Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. *Expert Syst. Appl.*. 37. 3638-3646. 10.1016/j.eswa.2009.10.027.
12. Hung, Shin-Yuan and Hsiu-Yu Wang. "Applying Data Mining to Telecom Churn Management." *PACIS* (2004).
13. Idris, Adnan & Iftikhar, Aksam & Tanoli, Zia-Ur-Rehman. (2017). Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Cluster Computing*. 10.1007/s10586-017-1154-3.
14. Idris, Adnan & Rizwan, Muhammad & Khan, Asifullah. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*. 38. 1808-1819. 10.1016/j.compeleceng.2012.09.001.
15. Idris, Adnan & Khan, Asifullah & Lee, Yeon Soo. (2013). Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost based ensemble classification. *Applied Intelligence*. 39. 10.1007/s10489-013-0440-x.
16. K. W. De Bock & D. Van Den Poel, 2012. "Reconciling Performance and Interpretability in Customer Churn Prediction using Ensemble Learning based on Generalized Additive Models," Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 12/805, Ghent University, Faculty of Economics and Business Administration.
17. Karahoca, Adem & Karahoca, Dilek. (2011). GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system. *Expert Systems with Applications*. 38. 1814-1822. 10.1016/j.eswa.2010.07.110.
18. Keramati, Abbas et al. "Improved churn prediction in telecommunication industry using data mining techniques." *Appl. Soft Comput.* 24 (2014): 994-1012.
19. Keramati, Abbas & M.S. Ardabili, Seyed. (2011). Churn analysis for an Iranian mobile operator. *Telecommunications Policy*. 35. 344-356. 10.1016/j.telpol.2011.02.009.
20. Kisioglu, Pinar & Topcu, Ilker. (2011). Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Syst. Appl.*. 38. 7151-7157. 10.1016/j.eswa.2010.12.045.
21. Lee, Hyeeseon & Lee, Yeonhee & Cho, Hyunbo & Im, Kwanyoung & Kim, Yong Seog. (2011). Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decision Support Systems*. 52. 207-216. 10.1016/j.dss.2011.07.005.
22. N. Lu, H. Lin, J. Lu and G. Zhang, "A Customer Churn Prediction Model in Telecom Industry Using Boosting," in *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659-1665, May 2014.
23. Owczarczuk, Marcin. (2010). Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Syst. Appl.*. 37. 4710-4712. 10.1016/j.eswa.2009.11.083.
24. P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, Indore, 2016, pp. 1-4.
25. Tsai, Chih-Fong & Lu, Yu-Hsin. (2009). Customer churn prediction by hybrid neural networks. *Expert Syst. Appl.*. 36. 12547-12553. 10.1016/j.eswa.2009.05.032.
26. Vafeiadis, Thanasis & Diamantaras, Kostas & Sarigiannidis, G & Chatzivasvas, Konstantinos. (2015). A Comparison of Machine Learning Techniques for Customer Churn Prediction. *Simulation Modelling Practice and Theory*. 55. 10.1016/j.simpat.2015.03.003.
27. Verbeke, Wouter & Martens, David & Mues, Christophe & Baesens, Bart. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Syst. Appl.*. 38. 2354-2364. 10.1016/j.eswa.2010.08.023.
28. W. Bi, M. Cai, M. Liu and G. Li, "A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn," in *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1270-1281, June 2016.
29. Zhao, Long & Gao, Qian & Dong, XiangJun & Dong, Aimei & Dong, Xue. (2017). K- local maximum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing*. 20. 10.1007/s10586-017-0843-2.
30. Zhu, Bing & Baesens, Bart & K.L.M. vandenBroucke, Seppe. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*. 408. 10.1016/j.ins.2017.04.015.
31. Raja, J. Beschi, S. Chenthur Pandian, and J. Pamina. "Certificate revocation mechanism in mobile ADHOC grid architecture." *Int. J. Comput. Sci. Trends Technol* 5 (2017): 125-130.
32. Datasource-<https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/>