Journal of Advanced Computer Science & Technology, 12 (2) (2024) 60-66

SPC

Journal of Advanced Computer Science & Technology

Website: www.sciencepubco.com/index.php/JACST

Research paper

Customer churn prediction in telecom using machine learning

Ahmed Faraz [1] *, Sibghat Ullah Bazai [1], Shumaila Hussain [2], Hamayoun Yousaf Shahwani [1], Mehmood Baryalai [3], Shariqa Fakhar [3]

[1] Department of Telecommunication Engineering BUITEMS
[2] Department of Computer Engineering BUITEMS1, Department of Computer Science SBKWU
[3] Department of Computer Science BUITEMS4,Department of Computer Science SBKWU
**Corresponding author E-mail: faraz.awan130@gmail.com*

Abstract

The customer churn prediction has become crucial for retaining customers in rapidly growing telecommunication industry. It enables telecommunication industry to avoid sizeable revenue losses. There are many techniques available to analyze the customer churn. It is still a challenging task due to large data size, imbalanced class distribution, and high dimensionality of telecom datasets.
This study contributes to formalize customer churn prediction. The study investigates the most promising machine learning algorithms XGboost and Logistic Regression. Furthermore; the feature extraction will be used before implementing the classification algorithms in order to analyze the impact.

Keywords: *Customer churn prediction, Telecommunication industry, Machine learning algorithms, Feature extraction, Imbalanced data*

1. Introduction

Customer churn is a major problem and one of the most important concerns for many industries. The customer churn directly effects the revenue of the company [1]. The telecommunication sector is the most adopted and influenced sector in the society. The telecommunication industry is growing day by day. The rapid growth is creating a competitive market. It is very important for the telecommunication companies to survive themselves in the competitive market and adopt numerous strategies to sustain. The sustainability of telecommunication companies totally relies on their customer. It is therefore that the companies are adopting various techniques to maintain a good ratio of the customers.
The most important factors of generating a good revenue in the telecommunication industry mainly focuses on customers

1) Attaining new customers
2) Upsell the existing customers and
3) Increasing the retention period of the customers.

It has been observed in any business that retaining customers is most profitable as compare to attaining new customers.
The availability of better choices for customer is creating customer churn. The customer churn is found to be most frequent in telecommunication industry as compare to others. It is therefore that the telecommunication industry is desperately seeking to adopt different techniques to predict customer churn. The customer churn management is found to be the most important task to maintain a good revenue.
Customer churn prediction is a complicated task. It has been observed that only gathering the data but data processing is a challenging task as the customers are to be segmented on the basis of the processed data [2]. Moreover; identifying the factors that increase customer churn is very important in order to take necessary actions to reduce this churn [3].
The prediction of the customer with the highest probability of leaving any telecommunication company is the most important factor in the customer churn management.
Literature Review
The customers are the priority of every telecommunication industry. The telecommunication industry is growing on daily basis [4]. The rapid growth of telecommunication industry is building a competitive market. This huge growth is providing customers with many opportunities in terms of better services, packages, new technical features, latest technological adoption and other facilities and in turn is increasing customer churn for the telecommunication industry [5] [6]. We therefore can say that the competitive market is promoting customer churn. The prevention of the customer churn can insure the revenue generation [7].
The customer churn prediction plays a very important part in the customer churn management. The customer churn prediction is a very complex and challenging task therefore telecommunication companies are adopting different model for this purpose [8] [9] [10]. The machine learning techniques are proven to be among the most adopted and promising as compare to others [11] [12] [13] [14] [15] [16].

Copyright © Ahmed Faraz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are many machine learning algorithms proven to be most effective in predicting the customer churn. The most implemented algorithms for customer churn predictions are support vector machine (SVM), linear regression, Naïve Bayes, decision tree and random forest etc. [17] [18].

The researchers presented Multilayer Perceptron Neural Networks for customer churn prediction [19]. Another study proposes an intelligent rule-based decision-making technique, based on rough set theory (RST), to extract important decision rules related to customer churn and non-churn. The proposed approach effectively performs classification of churn from non-churn customers, and proved that RST based on Genetic Algorithms (GA) showed better results [20]. The research focused on the k-mean clustering for the customer churn prediction. In another research two machine learning techniques the logistic regression and logit boost are applied to predict the customer churn and showed improved results [21].

The study focused on using the Ordered Weighting average (OWA) technique for customer churn prediction and showed improved results [22]. There are many hybrid techniques used for customer churn prediction like Ordered Weighting average (OWA) is combined with K-mean clustering technique showed improved results [23]. Similarly; another hybrid model Logistic Regression and Decision Tree was implemented for customer churn. This study proposed LLM (Logit Leaf Model) that includes the decision tree to divide the dataset into smaller subset based on some decision rules and then logistic regression was applied on each subset and showed improved results [24].

Similarly another research conducted [25], they adopted three machine learning approaches, namely, neural network, support vector machine and Bayesian networks for customer churn prediction. They have considered the feature selection and used principal component analysis (PCA) in order to reduce the dimension of the used data. They concluded that the performance of the feature selection can be improved by using optimization algorithms.

Y. Huang et al. [26], author's applied various classifiers on churn prediction dataset. They concluded that the random forest performs better than others in terms of area under the curve AUC and PR-AUC analysis. The have added that the optimization algorithms can improve the accuracy while considering the feature selection. Another research focused on analyzing the feature selection during customer churn prediction. The research concluded that the Fisher score and random forest were the techniques are the most important feature selection techniques [27]. The researcher. Another researcher used decision tree algorithm to identify the customer churn in e-commerce and concluded that they have achieved 90 % accuracy based on F-measure [28]. Another research for the customer data pattern recognition focused on block chain mechanism. They have used two kinds of cryptographic algorithms in block chain that are asymmetric-key algorithms and hash functions and achieve good performance for customer churn prediction [29].

The researcher Praveen et al. [30], focusing on implementing the machine learning techniques for customer churn prediction has evaluated logistic regression, naïve bays, support vector machine, random forest, decision tree, Adaboost and XGboost. The research has preprocessed the data and extracted the features using gravitational search algorithm and contributed that feature selection can improve the performance. They have analyzed that the Adaboost nad XGboost classifiers outperformed as compare to all other models showing accuracy of 80% and 84% respectively.

Another researcher focused on applying XGboost algorithm to improve the prediction business financial distress and achieved good accuracy [31]. Another researcher for customer churn prediction has used logistic regression and logit boost algorithms. They concluded that the logistic regression showed an accuracy 85.2385% where Logit Boost had an accuracy 85.1785% [32]. Another research conducted regarding the customer retention focused on XGboost. The experiments conducted for predicting the customer churn reveal that the oversampling methods improve the performance of Gradient Boosted Trees evaluated using F-measure value which is around 84% can be reached with SMOTE method at oversampling ratio of 20%. They overall concluded that the XGboost showed promising results as compare to other techniques [33].

The research focusing on building the customer value model that integrates the value of social network in order to help the companies subdivides the customer accurately. They have used machine learning algorithm XGboost for customer churn prediction before and after subdivision of the customer. They have concluded that the accuracy was found higher after sub division and added that the XGboost algorithm performed better as compare to other prediction algorithms [34].

Another research conducted to predict customer churn in internet service provider evaluated that the imbalance between the number of churners and non-churners can be reduced using SMOTE oversampling technique before implementing churn prediction models. They furthermore compared the customer churn prediction technique that are AdaBoost, Extra Trees, KNN, Neural Network and XGBoost and concluded that XGboost performed better. The XGboost showed 98% accuracy and the precission and recall of the model was found 45.71% to 42.06 % respectively [35]. The research focusing customer churn prediction has presented a hybrid model based on XGboost and multilayer perception MLP and showed better predictive performance [36]. Similarly another hybrid model based on K-mean and XGboost showed better customer churn prediction in telecommunication industry [37]. Another research to analyze the churn prediction algorithm compared XGboost, K nearest-neighbor KNN and Random Forest algorithms and identified that XGboost outperformed than the others and added that the fiber optics customers with greater monthly charges have higher influence for churn in the telecommunication industry [38]. Another research concluded that the algorithm called logistic regression used for customer churn prediction showed 93.4 % accuracy [39]. Another research concluded that the gradient boosted tree algorithm showed good accuracy for customer churn prediction and suggested to apply oversampling method to improve the accuracy of the churn prediction techniques [40].

Table 1: Summary of Literature Review

Author	Year	Techniques	Evaluation
Ming Zhao et al.	2021	Logistic Regression	Customer churn prediction is vital for churn management moreover; introduced management strategies for churn control.
Lawchak Fadhil Khalid et al.	2021	Data Mining strategies	Data mining for proper data segmentation is important for accurate customer churn prediction in telecommunication.
Ayesha Siddika et al.	2021	Convolutional Neural Network (CNN), Multilayer Perception (MLP) and Random Forest	Random Forest performed better
Kiran Dahiya et al.	2015	Decision Tree and Logistic Regression	Decision Tree performed better.
Mirjana Pejic-Bach et al.	2021	The hybrid approach of clustering(k-mean) and classification (CHAID decision tree) is proposed.	The proposed hybrid approach showed good results.
Vishal Mahajan et al.	2017	Reviewed different techniques	Reviewed churn prediction techniques
Abdelrahim Kasem Ahmed et al.	2019	Decision Tree, Random Forest, Gradient Boosted Machine Tree "GBM" and Extreme Gradient Boosting (XGboost).	XGboost performed better than all other algorithms.

T.vafeiadis et al.	2015	Multilayer perception, support vector machine, decision tree, naïve bayes and logistic regression were compared.	SVM-POLY using AdaBoost with accuracy of almost 97% and F-measure over 84%.
Ionuţ Brânduşoiu et al.	2016	Proposed implementation of datamining before implementing Neural networks, support vector machines, and Bayesian networks	Applying Datamining before implementing predictive model improved the accuracy till 99%.
J. Burez et al.	2009	Weighted random forests and random forest.	Weighted random forests performed better than random forest.
Irfan Ullah et al.	2019	Proposed hybrid technique based on classification and clustering.	Proposed technique showed 88.63 % accuracy.
Parveen Lalwani	2021	logistic regression, naive bayes, support vector machine, random forest, decision trees, K-fold cross validation , boosting	Adaboost and XGboost Classifier gives the highest accuracy of 81.71% and 80.8%
Omar Adwan	2014	Multilayer perception with back-propagation learning.	Proposed model performed well.
Amin Adnan et al.	2017	Rough set theory using four rule generation mechanism Exhaustive Algorithm (EA), Genetic Algorithm (GA), Covering Algorithm (CA) and the LEM2 algorithm (LA).	The proposed mechanism showed improved results.
Hemlata jain et al.	2020	Logistic regression and logit boost	Logit boost performed better.
Javad Bisiri et al.	2010	Ordered weighted average	OWA proved good accuracy in customer churn prediction in telecommunication.
Jie coa et al.	2015	Data mining and Ordered weighted average	The hybrid approach is promising in customer churn prediction.
Yiqing Huang et al.	2015	Big Data using Volume, Variety, Velocity	Improved accuracy while using big data analytics for customer churn prediction.
Yuwadi Thein Naing et al.	2022	Proposed Feature selection before implementing prediction algorithms.	Fischer Score and random forest were found to be the most prominent feature selection techniques.
Sulim kim and Heeseok lee	2022	Decision tree	Decision tree showed 90 % accuracy in customer churn prediction.
Mohammad Tabrez Quasim et al.	2022	Block chain	Block chain showed improved results while used for customer churn prediction.
Pedro Carmona et al.	2022	XGboost	XGboost showed improved results.
Attala M. et al.	2020	XGboost	XGboost improved the accuracy of customer churn prediction.
Yayun Zhuang	2018	XGboost	XGboost has improved accuracy of churn prediction.
Duyen Do et al.	2017	SMOTE oversampling before Adaboost, extra tree, KNN, neural networks and XGboost	XGboost showed more accuracy as compare to others.
Qi Tang et al.	2020	XGboost and multilayer perception (MLP)	XGboost performed better.
Pan Tang	2020	Hybrid based in k-mean and XGboost	Proposed hybrid model showed good accuracy.
j. pamina et al.	2019	K nearest neighbor (KNN), Random forest, XGboost	XGboost showed improved accuracy as compare to other techniques.
Zhang tianyuan et al.	2022	Fisher discriminant with logistic regression	Proposed technique showed improved accuracy
Alrence Santiago Halibas et al.	2019	Feature extraction and naïve bayes, generalized linear model, logistic regression, deep learning model, decision tree, random forest and gradient boosted tree.	Gradient boosted tree performed best.

2. Methodology

The study considers the most promising machine learning solutions for customer churn prediction. The dataset is retrieved from kaggle in order to avoid biased data. The study incorporates the feature selection technique in order to maximize the accuracy of the algorithms. The Logistic regression algorithm and XGboost algorithms will be implemented and evaluated using precision and recall evaluation indicators. The diagram below describes the research model

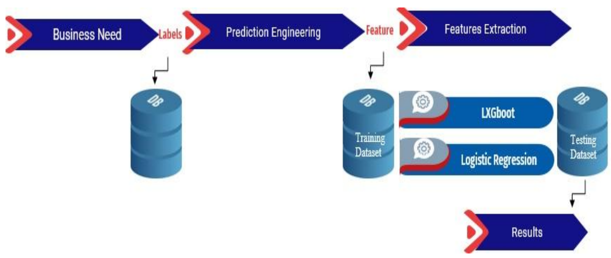


Fig. 1: Research Model.

3. Logistic regression

Logistic regression algorithm is a machine learning technique. It is a linear algorithm with non-linear output which uses logistic function. It can either be binomial, multinomial or ordinal. We have used binomial logistic regression model as we have two groups churner and non-churner customers. It considers the prediction in class 0 if it is >0.5 otherwise it takes output as class 1 the class 0 is referred as non-churner and 1 as the customers with possibility to churn.

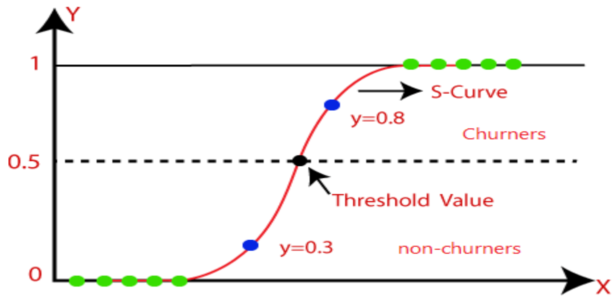


Fig. 2: Logistic Regression.

We can mathematically describe the hypothesis of the logistic regression as

$$0 \le h_\theta(x) \le 1$$

The equation θ = p / 1 – p where θ denotes the probability of churners and non- churners Now taking the log
log (p(x) / 1 – p(x)) = $\beta_0 + \beta_1$ x
Now taking the exponent on both sides

$$e^{ln}\left(\frac{p(x)}{1-p(x)}\right) = e^{\beta_0+\beta_1 x}$$

$$\left(\frac{p(x)}{1-p(x)}\right) = e^{\beta_0+\beta_1 x}$$

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

The logistic regression has proven to be one of the most frequently used machine learning algorithms for prediction.

4. XGboost

The Extreme Gradient Boosting (XGBoost) is a machine learning algorithm. It is highly saleable tree boosting algorithm. This algorithm is a modified and improved version of gradient boosting algorithm. The XGBoost adopted Regularized, Gradient and Stochastic boosting in the gradient boosting algorithm to enhance its performance. Moreover, it has decrease the computational time by parallel execution, optimal memory resource utilization, and handling the missing values while generating the tree construction [41] [42].

The XGboost algorithm in case of minimizing the loss in our case the mean square error which is $F_0(x)$

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$$

$$argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) = argmin_\gamma \sum_{i=1}^{n} (y_i - \gamma)^2$$

Taking the first differential with respect to γ, it is seen that the function minimizes at the mean i=1nyin. So, the boosting model could be initiated with:

$$F_0(x) = \frac{\sum_{i=1}^{n} y_i}{n}$$

5. Experiments

5.1. Data Acquisition

The data acquisition from a telecommunication industry is very important yet difficult task to predict the customer churn. Moreover; the biased data is another point in analyzing the performance of any algorithm. It is therefore that the study utilized dataset of telecommunication industry from Kaggle. The dataset contains 4000 customer's information with multiple attributes. The customer's attributes or behavior will be used to predict either they are going to churn or not.

5.2. Feature extraction

The random forest technique is used for feature selection before implementing the prediction algorithms in order to analyze the impact. It has been observed that the feature selection boosted the accuracy of the churn prediction algorithms.
The confusion matrix of XGboost algorithm is given below

Table 2: Confusion Matrix of Xgboost Similarly; the F Score of the Xgboost Algorithm for the Customer Churn Prediction Is Depicted Below

[1544, 65]
[193, 198]

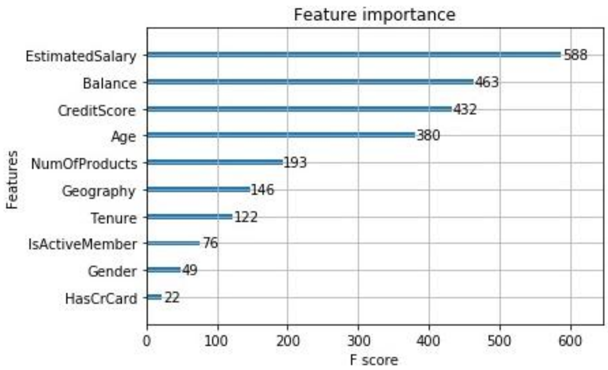


Fig. 3: F-Measure of XGboost.

The precision, recall and f1 score of customer churn prediction is given below

	precision	recall	f1-score	support
0	0.89	0.96	0.92	1609
1	0.75	0.51	0.61	391
micro avg	0.87	0.87	0.87	2000
macro avg	0.82	0.73	0.76	2000
weighted avg	0.86	0.87	0.86	2000

Fig. 4: Precision and Recall of XGboost.

The classification algorithm XGboost showed 87% accuracy while incorporating the feature engineering the performance was boosted to 90%.

Comparing the Logistic Regression classifier, the precision, recall and f1 score of logistic regression for customer churn prediction is given below

	precision	recall	f1-score	support
0	0.84	0.92	0.88	1021
1	0.71	0.54	0.61	388
avg / total	0.80	0.81	0.80	1409

Fig. 5: Precision And Recall of Logistic Regression.

The accuracy of customer churn prediction with Logistic Regression was found 82 % and using the classifier with feature engineering was 87%.

The categorical plots given below shows the churner and non-churners with respect to the features or categories of the customers so that the companies can easily identify the factors and can manage the customer churn accordingly.

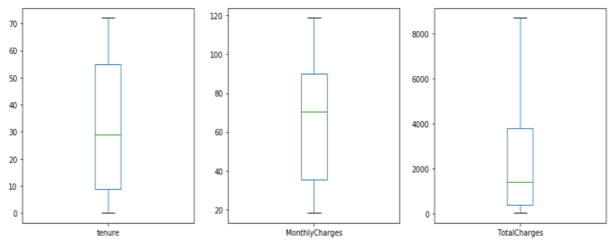


Fig. 6: Boxplot of Churners and Non-Churners.

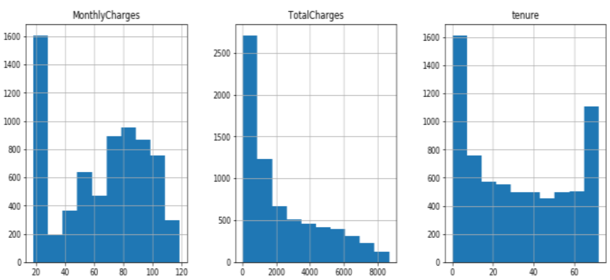


Fig. 7: Histogram of Churners and Non-Churners.

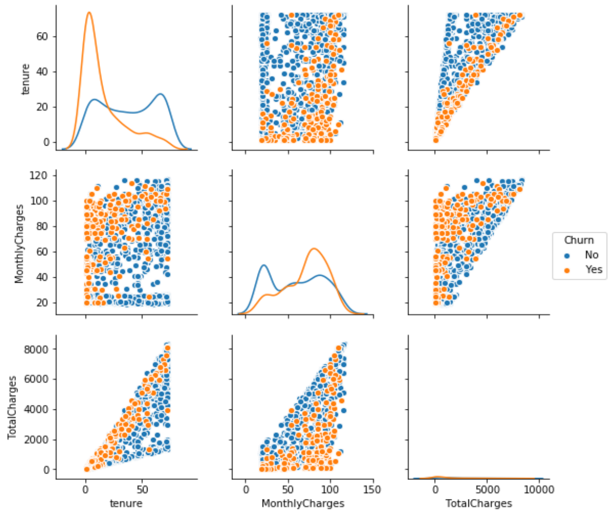
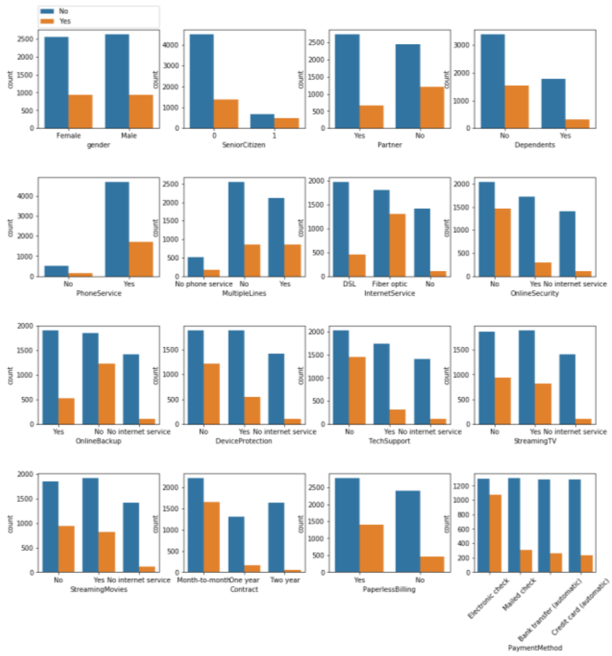


Fig. 8: Kernel Density Estimation of Churners and Non-Churners.



The comparison of the most promising machine learning classifiers was interesting as both the classifiers were using different techniques for customer churn prediction. Moreover; the impact of the feature engineering in boosting the performance of the classifiers has been observed. The study concludes that the feature engineering before implementing the machine learning techniques for churn prediction has boosted the performance of the classifiers. The results of the experiment demonstrate that the XGboost classifier after using with feature extraction has boosted from 87% to 90% accuracy similarly the Logistic Regression when used with feature engineering improved from 82% to 87% accuracy. The XGboost outperformed the customer churn prediction with 90% accuracy as compare to Logistic Regression.

6. Conclusion

This paper presented a classification solution with the feature extraction before implementing the churn prediction algorithms to identify the users at the risk of leaving the service in telecommunication industry. The study mainly focused on benchmarking the classifiers for customer churn prediction in telecommunication industry and enhances the role of feature engineering in boosting the performance of the algorithms. This is a vital contribution for the telecommunication industry seeking for the solutions to retain the customers for better revenue generation.

References

[1] M. Zhao, Q. Zeng, M. Chang, Q. Tong and J. Su, "A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China," Discrete Dynamics in Nature and Society, 2021. https://doi.org/10.1155/2021/7160527.

[2] L. F. Khalid, A. M. Abdulazeez, D. Q. Zeebaree, F. . Y. H. Ahmed and D. A. Zebari, "Customer Churn Prediction in Telecommunications Industry Based on Data Mining," in IEEE Symposium on Industrial Electronics & Applications (ISIEA), 2021. https://doi.org/10.1109/ISIEA51897.2021.9509988.

[3] A. Siddika, A. Faruque and A. K. Muhammad, "Comparative Analysis of Churn Predictive Models and Factor Identification in Telecom Industry," in International Conference on Computer and Information Technology (ICCIT), 2021. https://doi.org/10.1109/ICCIT54785.2021.9689881.

[4] M. Schaarschmidt and T. Kilian, "Impediments to customer integration into the innovation process: A case study in the telecommunications industry," European Management Journal, vol. 32, no. 2, pp. 350-361, 2014. https://doi.org/10.1016/j.emj.2013.04.004.

[5] C. Czarnecki and C. Dietze, "Understanding Today's Telecommunications Industry," in Reference Architecture for the Telecommunications Industry, Springer Link, 2017. https://doi.org/10.1007/978-3-319-46757-3_2.

[6] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," in International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015. https://doi.org/10.1109/ICRITO.2015.7359318.

[7] K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015. https://doi.org/10.1109/ICRITO.2015.7359318.

[8] M. Zhao, Q. Zeng, M. Chang, Q. Tong and J. Su, "A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China," Discrete Dynamics in Nature and Society, vol. 2021, pp. 1-12, 2021. https://doi.org/10.1155/2021/7160527.

[9] M. P. Bach, J. Pivar and B. JaKovic, "Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees," Journal of Risk and Finanacial Management, pp. 1-25, 2021.

[10] V. Mahajan, R. Misra and R. Mahajan, "Review on factors affecting customer churn in telecom sector," International journal of Data Analysis Techniques and Strategies , vol. 9, no. 2, pp. 122-144, 2017. https://doi.org/10.1504/IJDATS.2017.085898.

[11] A. K. Ahmad, A. Jafar and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, 2019. https://doi.org/10.1186/s40537-019-0191-6.

[12] T. V. K. D. G. S. and K. C. , "A comparison of machine learning techniques for customer churn prediction," Simulation Modelling Practice and Theory, vol. 55, pp. 1-9, 2015. https://doi.org/10.1016/j.simpat.2015.03.003.

[13] I. Brândușoiu, G. Toderean and H. G. Beleiu, "Methods for Churn Prediction in the Prepaid Mobile Telecommunications Industry," in International Conference on Communications, 2016. https://doi.org/10.1109/ICComm.2016.7528311.

[14] J. B. and D. V. d. Poel, "Handling class imbalance in customer churn prediction," Expert Systems with Applications, vol. 36, no. 3, pp. 4626-4636, 2009. https://doi.org/10.1016/j.eswa.2008.05.027.

[15] I. Ullah, B. Raza, A. K. Malik, M. A. Imran, S. U. I. and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE Access, pp. 60134-60149, 2019. https://doi.org/10.1109/ACCESS.2019.2914999.

[16] P. Lalwani, M. K. Mishra, J. S. Chadha and P. Sethi, "Customer churn prediction system: a machine learning approach," Computing, pp. 1-24, 2021. https://doi.org/10.1007/s00607-021-00908-y.

[17] A. K. Ahmad, A. Jafar and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, 2019. https://doi.org/10.1186/s40537-019-0191-6.

[18] V. U. and . K. I. , "A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics," International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056, no. 43, pp. 1065-1070, 2016.

[19] O. Adwan, H. Faris, K. Jaradad and O. Harfoushi, "Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks: Modeling and Analysis," Life Science Journal, vol. 11, no. 3, pp. 75-81, 2014.

[20] 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, vol. 237, pp. 242-254, 2017. https://doi.org/10.1016/j.neucom.2016.12.009.

[21] H. Jain, A. Khunteta and S. Srivastava, "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost," Procedia Computer Science, vol. 167, pp. 101-1012, 2020. https://doi.org/10.1016/j.procs.2020.03.187.

[22] J. Basiri, F. Taghiyareh and B. Moshiri, "A Hybrid Approach to Predict Churn," in IEEE Asia-Pacific Conference on Services Computing (APSCC), 2010. https://doi.org/10.1109/APSCC.2010.87.

[23] J. Cao, X. Yu and Z. Zhang, "Integrating OWA and data mining for analyzing customers churn in E-commerce," Journal of Systems Science and Complexity, pp. 381-392, 2015. https://doi.org/10.1007/s11424-015-3268-0.

[24] A. De Caigny, K. Coussement and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on Logistic Regression and Decision Tree," European Journal of operational research, vol. 267, pp. 760-772, 2018. https://doi.org/10.1016/j.ejor.2018.02.009.

[25] I. Brandusoiu, G. Toderean and H. Beleiu, "Methods for churn prediction in the pre-paid mobile telecommunications industry," in International conference on communications (COMM), 2016. https://doi.org/10.1109/ICComm.2016.7528311.

[26] Y. Huang, F. Zhu, M. Yuan, K. Deng, Y. Li, B. Ni, W. Dai, Q. Yang and J. Zeng, "Telco Churn Prediction with Big Data," in SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Malbourne, 2015. https://doi.org/10.1145/2723372.2742794.

[27] Y. T. Naing, M. Raheem and . N. K. Batcha, "Feature Selection for Customer Churn Prediction: A Review on the Methods & Techniques applied in the Telecom Industry," in 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2022. https://doi.org/10.1109/ICDCECE53908.2022.9793315.

[28] S. K. and H. Lee, "Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees," Procedia Computer Science, vol. 199, pp. 1332-1339, 2022. https://doi.org/10.1016/j.procs.2022.01.169.

[29] M. T. Quasim, A. Sulaiman, A. Shaikh and M. Younus, "Blockchain in churn prediction based telecommunication system on climatic weather application," Sustainable Computing: Informatics and Systems, vol. 35, 2022. https://doi.org/10.1016/j.suscom.2022.100705.

[30] P. Lalwani, M. K. Mishra, J. S. Chadha and P. Sethi, "Customer churn prediction system a machine learning approach," Computing, pp. 271-294, 2021. https://doi.org/10.1007/s00607-021-00908-y.

[31] P. Carmona, A. Dwekat and Z. Mardawi, "No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure," Research in International Business and Finance, 2022. https://doi.org/10.1016/j.ribaf.2022.101649.

[32] H. Jain, A. Khunteta and S. Srivastava, "Churn Prediction in Telecommunication using Logistic Regression and Logit Boost," Procedia Computer Science, pp. 101-112, 2020. https://doi.org/10.1016/j.procs.2020.03.187.

[33] A. M. AL-Shatnwai and M. Faris, "Predicting Customer Retention using XGBoost and Balancing Methods," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, 2020. https://doi.org/10.14569/IJACSA.2020.0110785.

[34] Z. Yayun, "Research on E-commerce Customer Churn Prediction Based on Improved Value Model and XG-Boost Algorithm," Management Science and Engineering, pp. 51-56, 2018.

[35] D. Do, P. Huynh, P. Vo and T. Vu, "Customer Churn Prediction in an Internet Service Provider," in IEEE International Conference on Big Data (BIGDATA), 2017. https://doi.org/10.1109/BigData.2017.8258400.

[36] Q. Tang, G. Xia, . X. Zhang and F. Long, "A Customer Churn Prediction Model Based on XGBoost and MLP," in International Conference on Computer Engineering and Application (ICCEA), 2020. https://doi.org/10.1109/ICCEA50009.2020.00133.

[37] P. Tang, "Telecom customer churn prediction model combining K-Means and XGboost algorithms," in IEEE International conference on mechanical, control nad computer engineering (ICMCCE), 2020. https://doi.org/10.1109/ICMCCE51767.2020.00248.

[38] J. P. . J. B. Raja, S. S. Bama and S. S. , "An Effective Classifier for Predicting Churn in Telecommunication," Journal of Advance Research in Dynamical & Control Systems, vol. 11, no. 01, pp. 221-229, 2019.

[39] Z. Tianyuan, S. Moro and R. F. Ramos, "A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation," Future internet, 2022.

[40] A. S. Halibas, A. C. Matthew, I. G. Pillai, J. H. Reazol and E. G. Delvo, "Determining the Intervening Effects of Exploratory Data Analysis and Feature Engineering in Telecoms Customer Churn Modelling," in MEC International Conference on Big Data and Smart City (ICBDSC), 2019. https://doi.org/10.1109/ICBDSC.2019.8645578.

[41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, Claifornia, 2016. https://doi.org/10.1145/2939672.2939785.

[42] T. Chen, T. He and M. Benesty, "XGBoost: R-package," R package version 0.4-2, p. 104, 2015.