International **EU**ropean **J**ournal of **M**anagerial **R**esearch
Dergisi (EUJMR)
ISSN: 2602 – 4179 / Güz / Cilt: 5 / Sayı: Özel Sayı 1
Araştırma Makalesi

GÜÇLENDIRME ALGORİTMALARI İLE TELEKOMÜNİKASYONDA MÜŞTERİ KAYBI TAHMİNİ

Ezgi NAZMAN[1]

ÖZ

Müşteri kaybı, özellikle bankacılık ve telekomünikasyon alanındaki büyük şirketler için en büyük sorunlardan biridir. Son zamanlarda, telekomünikasyon şirketleri, yeni müşteriler kazanmanın maliyeti, mevcut müşterileri elde tutmaktan daha fazla olduğu için müşteri kaybını önleme eğilimindedir. Bu nedenle şirketler, makine öğrenimi algoritmaları gibi farklı tahmin yöntemlerini kullanarak potansiyel kayıpları belirlemek istemektedir. Karar ağacı öğrenme yöntemlerine ait olan XGBoost, Adaptive ve Gradyant güçlendirme algoritmaları, denetimli makine öğrenimi yöntemleri olarak yaygın şekilde kullanılmaktadır. Bu algoritmaların performansları, veri seti oldukça dengesiz olduğu durum için halen karşılaştırmaya açıktır. Çalışmada, güçlendirme algoritmalarını değerlendirmek için % 26,4'ü kaybedilen müşterileri içeren veri seti dikkate alınmıştır. Özellikler, müşterilerin cinsiyet, çevrimiçi güvenlik, internet hizmeti, çevrimiçi yedekleme gibi müşteri kaybetme durumuyla ilişkili olabilecek değişkenlerden oluşmaktadır. İlk olarak, müşterilerin ilgili özellikler açısından dağılımını anlamak için keşifsel veri analizi uygulandı. Daha sonra, dengesiz veri problemini ortadan kaldırmak için Adaptive-oversampling yöntemi kullanılmıştır. Son olarak, karşılaştırılan algoritmaların tahmin sonuçlarını değerlendirmek amacıyla doğruluk, kesinlik ve F1 ölçümleri hesaplandı. Tahmin sonuçlarını doğrulamak için 10 kat çapraz geçerlilik de uygulandı.

Anahtar Kelimeler: Müşteri Kaybı, Telekomünikasyon, Makine Öğrenmesi, Dengesiz Veri

Jel Kodları: C45, L96, M1

[1]Arş. Gör. Dr., Sivas Cumhuriyet University, ezgicabuk@cumhuriyet.edu.tr, (ORCID: 0000-0003-0189-3923)

Geliş Tarihi/ Received: 28.12.2020
Kabul Tarihi/ Accepted: 11.11.2021

CUSTOMER CHURN PREDICTION IN TELECOMMUNICATION WITH BOOSTING ALGORITHMS

ABSTRACT

Customer churn is one of the major problems for large companies, especially in banking and telecommunication. Recently, telecommunication companies tend to prevent customer churn since the cost of gaining new customers is more than retaining existing customers. Therefore, the companies would like to have to determine potential churns using different prediction methods such as machine learning algorithms. XGBoost, Adaptive, and Gradient Boosting algorithms, which belong to the decision tree learning methods, are widely used as supervised machine learning methods. The performances of these algorithms are still open to comparison for the case where the data set is highly imbalanced. In the study, the data set including 26.4% of customer churn was considered for the study to evaluate the performances of the boosting algorithms. Features are consist of the variables which can be related to the churn decision of the customers such as gender, online security, internet service, online backup, etc. Firstly, Exploratory data analysis was applied to understand the distribution of customers in terms of the related features. Then, the Adaptive-oversampling method was used to eliminate the imbalanced data problem. Lastly, accuracy, precision, and F1 metrics were calculated to evaluate the prediction results of the algorithms. 10 fold cross-validation was also applied in order to validate accuracy results.

Keywords: Customer Churn, Telecommunication, Machine Learning, Imbalanced Data

JEL Codes: C45, L96, M1

INTRODUCTION

A telecommunication company wishes to predict possible customer churn to take required measures for this customer movement (Wei and Chiu, 2002). It is known that gaining new customers is 5 to 10 times costlier than keeping existing ones (Athanassopoulos, 2000; Kotler 2009). In today's world, customers have numerous choices of telecommunication companies; therefore, they are often willing to switch the company when they are not satisfied with the services. Customer churn can be defined as a movement of the customer who is moving from the existing company to another company (Mishra and Reddy, 2017; Óskarsdóttir et al., 2017). The customer churn can be due to various reasons including dissatisfaction, higher cost, low quality of the services and products (Sharma & Rajan, 2017).

The customer churn prediction can not only improve a company's revenue and but also a reputation in the market (Maria et al., 2016). Nowadays, a telecommunication company is suffering from customer churn due to the competition of the market and dynamic conditions. The average churn rate for a telecommunication company is estimated at 2.2% per month (Berson et al. 2000). Therefore, marketing departments are always willing to launch new attractive offers (Óskarsdóttir et al., 2017). Since the telecommunication companies can easily keep track of their customers' demographic information and activities on the services, churn prediction can be easily applied using several machine learning algorithms (Ćamilović, 2008). Although boosting algorithms are known as superior algorithms since they increase the accuracy of single classifiers by combining several of them, the performances of these models can be greatly affected when the data set is highly imbalanced (Galar et al. 2011). However, highly imbalanced data stands as a challenge for the prediction.

In this study, the Adaptive-oversampling method was used to eliminate the imbalanced data problem. Then, XGBoost, Adaptive, and Gradient Boosting algorithms were compared in terms of the customer churn prediction performances in terms of a telecommunication company data set. The remaining of the paper is organized as follows. Section 2 gives the literature review on the customer churn prediction where machine learning algorithms were used. Section 3 introduces the boosting algorithms and oversampling method for the imbalanced data case. Section 4 gives the detail about the data set and the comparison results of the boosting algorithms. Finally, conclusions on the current study and useful suggestions for future studies are provided in Section 5.

1. THEORETICAL FRAMEWORK

A wide range of approaches or methods hase been applied to predict customer churn based on machine learning algorithms. Idris & Asifullah, 2014 proposed an ensemble approach which considers initially Max-Relevance, and Min-Redundancy feature selection (Peng et al. 2005), then majority votings of K-Nearest Neighbor(KNN), Random Forest (RF), and Rotation Forest methods. They indicated that the proposed approach was a promising method for customer churn prediction in telecommunication in terms of classification metrics. Farquad et al., 2014 proposed a hybrid approach including 3 main steps. While the first step contains the feature selection using Support Vector Machine and Recursive Feature Elimination (SVM-RFE), the SVM method was used for the churn prediction as to the second step. They mentioned that the proposed hybrid method using SMOTE oversampling method revealed the best sensitivity. Ahmed & Maheswari, 2017 proposed a hybrid Firefly algorithm based on a metaheuristic approach. Even though, Firefly algorithm superior on churn prediction, it was seen that the hybrid Firefly algorithm was effective and faster. Amin et al., 2019 grouped customers using the distance factor, and they indicated that this method could be used to select cases for efficient and more accurately training. Coussement et al. (2016) studied enhancing prediction performance of Logistic Regression with appropriate data preparation. LR was compared with the bagged CART, Bayesian Network, J4.8 decision tree, Multilayer perceptron neural network, Naive Bayes, Random Forest, Radial basis kernel support vector machine, and Stochastic gradient boosting. The study indicated that LR was competitive with advanced single and ensemble methods after the appropriate data preparation. Umayaparvathi&Iyakutti (2016) compared the performances of LR, KNN, RF, SVM, Ridge classifier, Decision tree, and Gradient Boosting algorithms, and they mentioned that the Gradient Boosting method had the best performance. Esteves&Mendes (2016) compared the performances of KNN, Naive Bayed, C4.5, RF, AdaBoost, and Artificial Neural Network algorithms for the churn prediction in telecommunication. According to the area under the Receiver Operating Characteristic (ROC) curve, sensitivity and specificity, RF had the best performance.

2. METHOD

The objective of supervised machine learning algorithms is to increase the overall accuracy rate of the prediction. Boosting algorithms consider training data set resulting in having misclassifications (Rahman et al. 2020). Adaptive Boosting (AdaBoost) initially assigns equal weights to each observation in the training data set. AdaBoost uses multiple weak models and places higher weights on the observations. Therefore, the accuracy rate of the misclassified observations is improved during the iteration (Freund& Schapire, 1997). Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function (Friedman, 2002). XGboost, a scalable tree boosting system, consists of a histogram-based algorithm to find the split. As a result, the calculation time of this algorithm is more than the gradient-based one-side sampling (Chen and Guestrin, 2016). The imbalanced data occurs when group label distribution significantly dominates the instance space compared to other data distributions in group labels. Customer churn data sets have generally imbalanced group labels means that the rate of the churned customers is mostly lower than that of existing customers. This case can cause bias in the prediction process using machine learning classification methods. oversampling methods can provide broad and well-defined decision regions for minority group labels (Li and Mao, 2014). There are several oversampling methods to overcome this situation. One of the most widely used oversampling methods is Adaptive-oversampling (Adasyn) which helps reducing bias and learning adaptively (He at al. 2008).

Table 1. Confusion Matrix

		Predicted class	
		+	-
True class	+	**T**rue **P**ositives (TP)	**F**alse **N**egatives (FN)
	-	**F**alse **P**ositives (FP)	**T**rue **N**egatives (TN)

Reference: (Fawcett, 2006)

In supervised machine learning, prediction results are evaluated by using evaluation metrics which can be calculated from a confusion matrix as in Table 1. These metrics of the classification results are calculated as below:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+FN+TN},\quad \text{Precision}=\frac{TP}{TP+FP},\quad \text{Recall}=\frac{TP}{TP+FN},\quad \text{F1}=\frac{2\cdot Precision\cdot Recall}{Precision+Recall}$$

In the study, Python programming language was used for all calculations. matplotlib (Hunter 2007) and seaborn (Waskom et al. 2017) libraries were used for data visualizations shown in the figures. Scikit-learn (Pedregosa et al. 2011) library was used for AdaBoost and Gradient Boosting classifications and XGBoost (Chen and Guestrin 2016) library was used for XGBoost classification.

3. RESULTS

The Telecommunication customer churn data is a fictional telecommunication company with 7043 customers in California. There are fourteen important demographic and telecommunication-related variables for each customer (IBM Cognos Analytics 11.1.3). The churn distribution based on customers who have left or stayed in the company is imbalanced as in Figure 1 (26.5%).

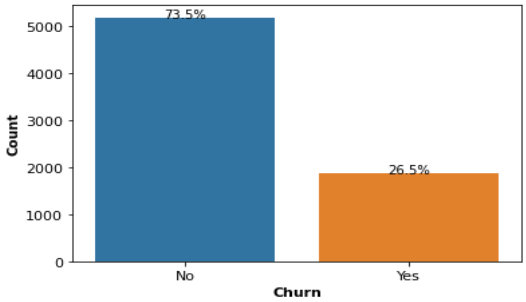

Figure 1: Churn Distribution of the Customers

Two continuous independent variables are tenure (32.37±24.56) and monthly charges (64.76±30.09).

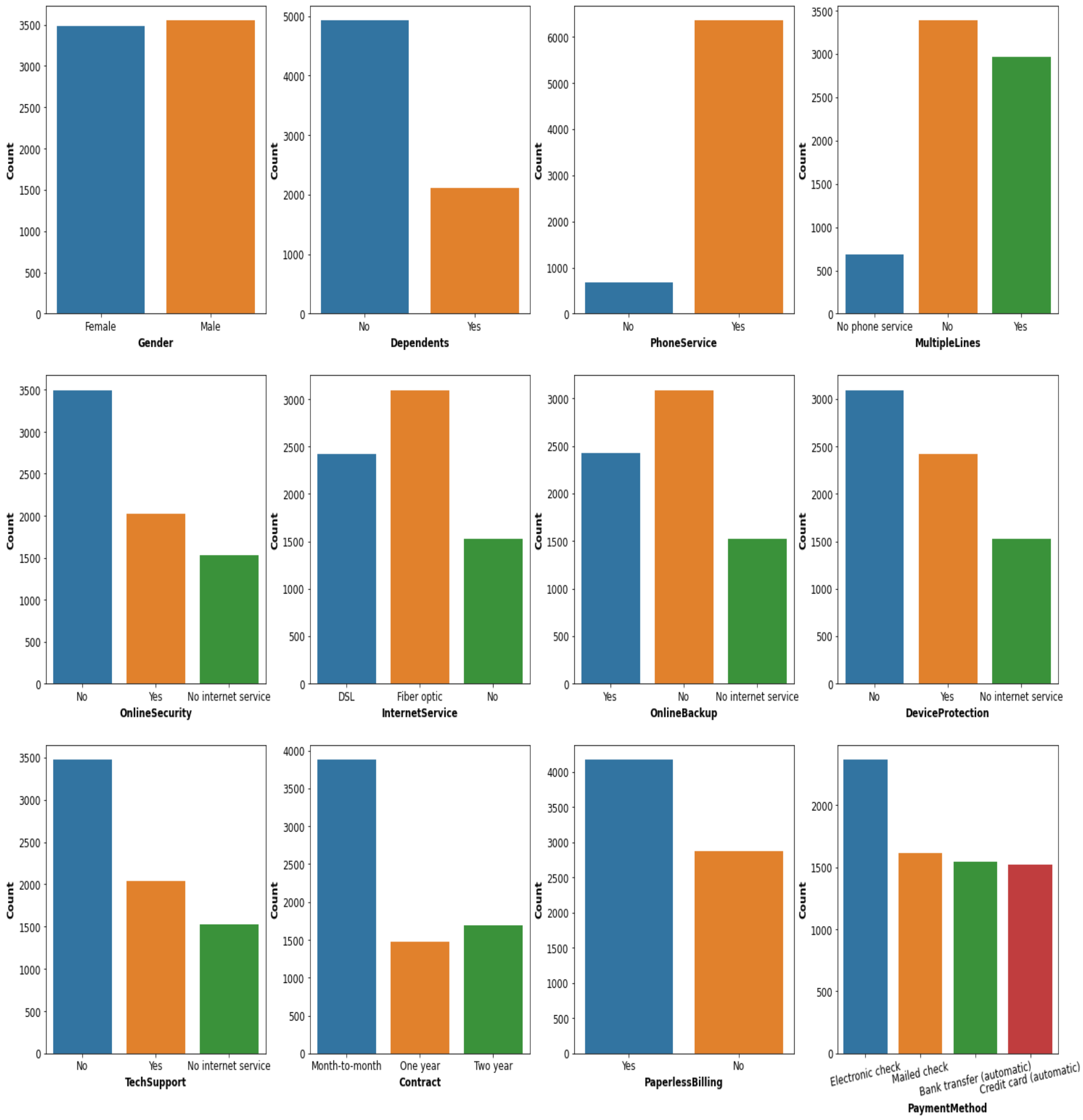

Figure 2: Distributions of the Categorical Variables

The distributions of the categorical variables are shown in Figure 2. The number of female and male customers are approximately equal. The number of fiber optic users is higher than the DSL users and non-users. The phone service users are much higher than non-users. Most users do not use multiple services and online security. The number of customers who have a month-to-month contract is higher. The number of customers who do not have technical support is higher.

First, the Adasyn-oversampling method was applied to the data set since the distribution of customer churn is imbalanced. The data set was split into the train (80%) and test (20%) data sets after the oversampling. Then, Gradient Boosting, AdaBoost, and XGBoost algorithms were compared in terms of classification performances. Receiver Operating Characteristic (ROC) curve, cross-validation accuracy, and F1 results were obtained as shown in Figure3. Here AUC indicates the area under the ROC curve.

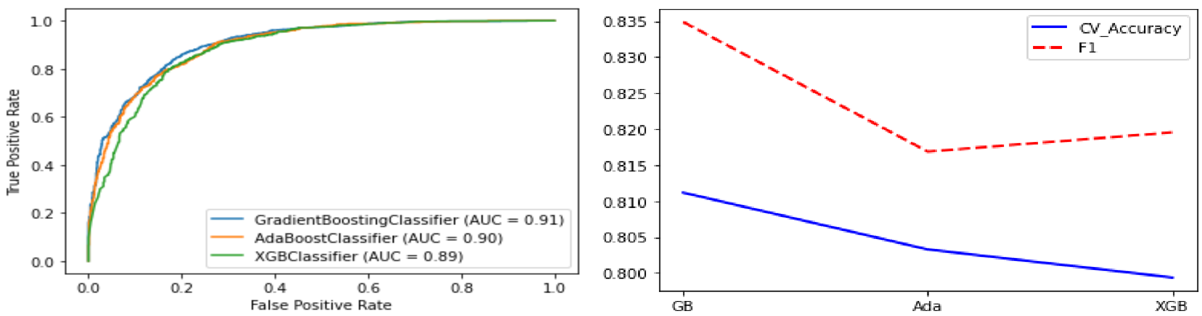

Figure 3: ROC Curve, Cross-Validation (CV) Accuracy, and F1 Score Rates.

According to the results, Gradient Boosting has the highest AUC, cross-validation accuracy, and F1 results. Adaboost has the least F1 score whereas the Xgboost algorithm has the least cross-validation accuracy. The confusion matrices support the results shown in Figure 3. Gradient Boosting tends to predict more precisely customer churn as in Figure 4.

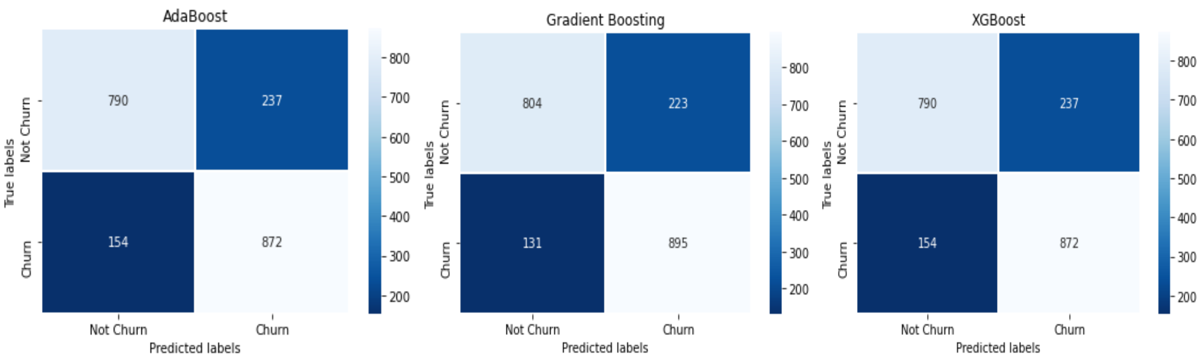

Figure 4: Confusion Matrices of the AdaBoost, Gradient Boosting, and XGBoost Prediction Results

RESULT

In the telecommunication industry, the customers have many choices in terms of the services and products; therefore, they are willing to switch the company when other companies are more convenient for them. As it is known that churn decision of the customers can be due to various reasons including displeasure with the services and cost. A telecommunication company can predict customers who tend to exit with the help of machine learning algorithms and take measures early. In this study, XGBoost, Ada, and Gradient Boosting machine learning algorithms were compared in terms of the customer churn prediction performances on real telecommunication company data set. As a general result, the boosting algorithms have high prediction accuracy rates. When it comes to the comparison, it is seen that the Gradient Boosting algorithm is superior in terms of classification metrics such as cross-validation accuracy rate, F1 score, and confusion matrix results. Therefore, the Gradient Boosting algorithm is suggested to the researchers who would like to predict customers who tend to exit from a telecommunication company.

Declaration: This study was presented online at the International Communication, Economics & Organization Social Sciences Congress, 18-20 December 2020, Gorazde University, Bosnia Herzegovina.

REFERENCES

Ahmed, A. A., Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal*, 18(3), 215–220.

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., Anwar, S. (2019). “Customer churn prediction in telecommunication industry using data certainty”. *Journal of Business Research*, 94, 290-301.

Athanassopoulos, A. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research*, 47(3), 191-207.

Berson, A., Smith, S., Thearling, K. (2000). *Customer retention*. Building data mining applications for CRM, New York:McGrawHill, Chapter 12.

Ćamilović, D. (2008). Data mining and CRM in telecommunications. *Serbian Journal of Management*, 3(1):61-72.

Chen, T., Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD, International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 785–794. http://doi.acm.org/10.1145/2939672.2939785.

Coussement, K., Lessmann, S., Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction:a case study in the telecommunication industry. *Decision Support Systems*. 95:27-36.

Esteves, G., Mendes, M.J. (2016). *Churn prediction in the telecom business*, The Eleventh International Conference on Digital Information Management (ICDIM 2016), 254-259.

Farquad, M., Ravi, V., Raju, S. B. (2014). Churn prediction using comprehensible support vector machine:an analytical CRM application. *Applied Soft Computing*, 19:31–40.

Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters,* 27:861-874.

Freund, Y., Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis,* 38: 367–378.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F. (2012). Review on ensembles for the class imbalance problem:bagging-, boosting-, and hybrid-based approaches, *IEEE Transactions on Systems, Man, and Cybernetics-Part C* (Applications and Reviews), 42(4), 463-484, doi: 10.1109/TSMCC.2011.2161285.

He, H., Bai, Y., Garcia, E.A., Li, S. (2008). *ADASYN: Adaptive synthetic sampling approach for imbalanced learning*. International Joint Conference on Neural Networks (IJCNN 2008), IEEE.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment, *Computing in Science & Engineering,* 9(3):90-95.

IBM Cognos Analytics 11.1.3. https://community.ibm.com/community/user/businessanalytics/

blogs/steven-macko/2019/07/11/telco-customer-churn-1113

Idris, A., Asifullah, K. (2014). *Ensemble based efficent churn prediction model for telecom*. 12th International Conference on Frontiers of Information Technology (FIT), 1–7.

Kotler, P., Keller, K.L. (2009). *Marketing management*. Pearson Prentice Hall.

Li, Q. And Mao, Y. (2014). A review of boosting methods for imbalanced data classification, *Pattern Analysis and Applications,* 17: 679-693.

Maria, O., Verbeke, W., Sarraute, C., Baesens, B., Vanthienen, J. (2016). *A comparative study of social network classifiers for predicting churn in the telecommunication industry*. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),1151–1158.

Mishra, A., Reddy, U.S. (2017). *A comparative study of customer churn prediction in telecommunication industry using ensemble based classifiers*. Proceedings of the International Conference on Inventive Computing and Informatics. IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9.

Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., Vanthienen, J. (2017). Social network analytics for churn prediction in Telco:model building, evaluation and network architecture. *Expert Systems with Applications*, 85:204–220.

Peng, H., Long, F., Ding, C. (2005). *Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and MinRedundancy*. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): 1226-1238.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al. (2011). Scikit-learn: Machine Learning in Python, JMLR 12:2825-2830.

Rahman, S., Irfan, M., Raza, M., Ghori, K.M., Yaqoob, S., Awais, M. (2020). Performance analysis of boosting classifiers in recognizing activities of daily living. *International Journal of Environmental Research and Public Health*. 17(3):1082, 1-15.

Sharma, R. R., Rajan, S. (2017). Evaluating prediction of customer churn behavior based on artificial bee colony algorithm. *International Journal of Engineering and Computer Science*, 6(1): 20017–20021.

Umayaparvathi, V., Iyakutti, K. (2016). *Customer churn prediction using big data analytics*. PhD Thesis from Blekinge Institute of Technology.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D.C., et al., (2017). *mwaskom/seaborn: v0.8.1*. Zenodo. https://doi.org/10.5281/zenodo.883859.

Wei, C.P., Chiu, I.T. (2002). Turning telecommunications call details to churn prediction:a data mining approach. *Expert Systems with Applications*, 23:103-112.