FTI Proceedings: 4th international conference on computer applications and information security (iccais'2021) / March 19 / 2021/ Tunisia: https://fti-tn.net/iccais-2021-list-of-papers

Automated Prediction of Customer Churn in Telecom Sector

Eman Saeed Alamoudi
College of Computers and Information Technology
Taif University
Taif, Saudi Arabia
eman.alamoudi@tu.edu.sa

Albtool Alaidah
College of Computer and Information Sciences
Princess Nourah Bint Abdulrahman University
Riyadh, Saudi Arabia
alaidah.albtool@gmail.com

Malak AlQahtani[1], Hajar AlNamshan[2]
School of Computing
Dublin City University
Dublin, Ireland
[1]malak.m.q@hotmail.com, [2]hajar.alnamshan2@mail.dcu.ie

Musfira Jilani
Independent Researcher
India
musfira.jilani@ucdconnect.ie

Abstract—**For major businesses, customer churn is a primary concern and one of the most critical issues. Due to the direct impact on a company's revenues, notably in the telecommunications sector, companies are attempting to develop techniques to predict potential customer churn. Most businesses focus on retaining clients rather than adding new ones, as the expense of gaining a new client is significantly greater than retaining a current one. It is, therefore, crucial to find factors that increase customer churn to take appropriate measures to reduce this churn. The primary contribution of this study is to develop a model of churn prediction that allows telecommunications companies to determine which clients are most likely to be subject to churn. Toward this end, eight different machine learning algorithms were studied for customer churn prediction. In addition, advanced feature selection methods and parameter optimization were leveraged for improving the model performance. Model performance was studied in terms of metrics such as accuracy, sensitivity, and AUC as well as by comparing with previous studies. XGBoost algorithm-based model with an accuracy of 80.3% gave the best results. Moreover, this study discusses the results in a business-oriented way that makes them more applicable. Thus, this will help decision-makers subjectively and objectively choose the right model based on their business situation.**

Keywords-customer-churn; logistic regression; telecommunications; machine learning; grid search; XGBoost

I. INTRODUCTION

Today, there is fierce competition in the field of telecommunications as companies compete for products and customers, which has led to significant drops in revenue for some companies [1]. One of the risks is that customers are switching from one service provider to a competitor, which is known as customer churn (customer loss). Thus, addressing the issue of customer churn in the telecommunications industry has become unavoidable. The primary and most common cause of churn is either due to a customer's dissatisfaction with a provider's services or because another service provider can offer improved services at a lower price. Many telecommunications providers undertake enormous investments to retain their current customers. Most of the companies focus on retaining current customers rather than attracting new customers as the cost of acquiring a new customer is usually much higher than the cost of retaining an existing one [2]. Churn analysis, therefore, is crucial to retaining valuable customers for many companies. Retention strategies are used by several telecom companies to keep customers for a longer period. In order for telecom companies to minimize their customer turnover, it is crucial that they be able to predict which customers are at risk of moving to another service provider so that their needs can be met by preventing churn at an early stage. Machine learning has been used for churn prediction to help detect the problem early and also identify the reasons behind the decisions of customer churns [3]. The telecommunications industry's huge amount of data, such as customers' data and billing data, provide a great opportunity for analysis and processing throughout machine learning and data analytics.

This study is organized as follows. Section II presents a review of the literature in the domain of customer churn prediction in the telecom sector. The methodology including the machine learning based model development is presented in Section III. The experiments and results are discussed in Section IV, and the discussion and analysis are provided in Section V. Finally, the conclusion, future directions, and limitations are presented in Section VI.

II. RELATED WORK

Customer churn is a critical problem in telecommunications and it is an issue that has been addressed by many studies. A wide variety of techniques, in both supervised and unsupervised learning algorithms, have been applied to predict churn. In [4], the researcher developed a model on a dataset acquired from a PAKDD – 2006 data mining competition. The major focus was on data preprocessing followed by building a decision tree with an accuracy of 98.88%. The researchers in [5] tried three different algorithms namely: Naïve Bayes, decision tree, and random forest. The results demonstrated the effectiveness of ensemble learning, with the highest accuracy of 97.5% gained by random forest. With more complicated models, authors in [6] obtained the highest accuracy of 98% using fuzzy-based K-nearest neighbor classifiers compared to benchmark models.

Much of the previous research has relied upon the domain expert's knowledge to select the features for the prediction models. In contrast, the researchers in [7] performed Boruta and random forest algorithms for feature selection. Hence, it improved the model's performance, where the highest results were obtained by random forest and adaptive boost with 96% accuracy. Moreover, researchers in [8] used the same dataset used in this present study. They explored different feature selection techniques such as correlation, gain ratio, OneR and information gain. The algorithms used include classification and regression trees (CART), partial decision trees (PART), and bagged (bootstrap aggregation or bagging) CART classifiers; the highest accuracy obtained was 79%, which was obtained by a combination of Correlation feature selection and CART model.

Some researchers focused mainly on proposing new feature selection techniques. Researchers in [9] proposed a hybrid feature selection framework using stacked auto-encoder and Fisher's ratio analysis. Authors in [10] performed a hybrid two-phase feature selection using a Markov blanket discovery method. Furthermore, authors [11] developed a two-phase feature selection method using an AUC index. Authors in [12] demonstrated the effectiveness of three feature selection tools: permutation importance, partial dependence plot, and SHapley Additive exPlanations (SHAP). In contrast, researchers in [13] claimed that there was no need for the feature selection phase when using only deep learning-based models.

We note that little relevant research on the issue related to optimization methods for customer churn prediction has been conducted, creating a gap in the literature. Most related experiments focus only on models for machine learning and on feature selection techniques. Therefore, this research gap will be filled by this present study. This study proposes the development of a special system that enables businesses to efficiently stop customer churn by designing accurate predictive models. An optimization approach for enhancing model results will be introduced.

III. METHODOLOGY

In this section, we present our framework of effective customer churn prediction. Different approaches were used to achieve various combinations of techniques and optimal results at each point. In the following, parts are presented as well as the goals of every step, and explanations of the techniques.

A. Data Description

In this study, the dataset was obtained from the Kaggle website[1] and was originally downloaded from IBM Sample Data Sets for customer retention programs. There are 7,043 client records in the dataset, with 21 features: 20 descriptive and one target feature (churn). The data types include numerical and categorical data as defined in Table I.

TABLE I. FEATURES DESCRIPTION

Feature	Description and Data Type
Customer ID	The customer's ID - Categorical
Gender	The customer's gender (male/female) - Categorical
Senior Citizen	Customer is a senior citizen or not (1, 0) - Numerical
Partner	Customer has a partner or not (Yes, No) - Categorical
Dependents	Customer has dependents or not (Yes, No) - Categorical
Tenure	Number of months the customer has stayed with the company - Numerical
Phone Service	Customer has a phone service or not (Yes, No) - Categorical
Multiple Lines	Customer has multiple phone lines or not (Yes, No, No phone service) - Categorical
Internet Service	Customer's internet service provider (DSL, Fiber optic, No) - Categorical
Online Security	Customer has online security or not (Yes, No, No internet service) - Categorical
Online Backup	Customer has online backup or not (Yes, No, No internet service) - Categorical
Device Protection	Customer has device protection or not (Yes, No, No internet service) - Categorical
Tech Support	Customer has tech support or not (Yes, No, No internet service) - Categorical
Streaming TV	Customer has streaming TV or not (Yes, No, No internet service) - Categorical
Streaming Movies	Customer has streaming movies or not (Yes, No, No internet service) - Categorical
Contract	The contract term of the customer (Month-to-Month, 1 Year, 2 Years) - Categorical
Paperless Billing	The customer has paperless billing or not (Yes\No) - Categorical
Payment Method	The customer payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)) - Categorical
Monthly Charges	The amount charged to the customer monthly - Numerical
Total Charges	The total amount charged to the customer - Categorical
Churn	Does the customer churn? (Yes\No) - Categorical

B. Data Cleaning

In this section, we provide the techniques to turn the raw data into an appropriate and adequate format for machine learning models. The missing value issue exists when such instances have lacking values for one or more features. In our case, the TotalCharges attribute contains 11 missed values, so the optimal approach is to delete these instances with missing values. The outlier is the value that is far away from a feature's central tendency. The minimum and maximum values of the numerical characteristics have been observed to find outliers within the used dataset and according to the domain information; the result showed that data do not suffer from outliers. Generally, having continuous features with hugely different ranges in a dataset can cause difficulties for some machine learning algorithms, so, standardization of a dataset is

[1] https://www.kaggle.com/blastchar/telco-customer-churn.

a common requirement for many machine learning estimators [14]. it is preferred that the individual features look like standard, normally distributed data. The standard score, a python method, was applied on the two continuous features MonthlyCharges and TotalCharges. Moreover, binning is used to convert continuous features to categorical features to facilitate the work of machine learning algorithms. An equal-width binning was applied on tenure to distribute instances equally among five different ranges. To uniform all features values, the 0, 1 value in SeniorCitizen was replaced to "No" and "Yes," respectively. Furthermore, in the following features: OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV and StreamingMovies, the "No internet service" values were replaced with "No" as they both refer to the same class. Moreover, TotalCharges is assumed to be a numeric data type, but it was stored as a string, for easy handling; this feature type is converted to a float data type. Finally, a separate variable was generated by one-hot encoding for each categorical feature with more than two classes in order to ensure the dataset was applicable for machine learning models.

C. *Data Analysis*

In order to explain the data and sum up its key characteristics, the exploratory data analysis (EDA) approach was applied. Some associations between descriptive and objective characteristics were visualized for insight into the relevant characteristics. Customers with a month-to-month contract, tenure category, who have been with the company for a year or longer, internet subscribers with fiber optics category, and paperless customers all appeared to be churner candidates. Whereas, two-year contract clients, customers who do not use an internet service, customers who have been with the company for four years or more, and customers with low average annual costs tend to have a negative relation with turnover, and are thus not viewed as churner candidates. Finally, the distribution of the used dataset has 26.6% churn customers and 73.4% non-churn customers; as stated in [15], many studies did not use sampling techniques but worked on the current distribution of the data.

D. *Feature Selection*

A feature selection technique is a process to select the appropriate set of features among all the other features, and to aim to construct a precise and fast predictive model. Information gain (IG) and SubsEval are the classifiers used in this experiment, and the features selection phase was applied using Weka software[2]. The IG formula is a measurement of the entropy of a system [16]. For each attribute, information gain can be measured, and the values vary between 0 (not informative) and 1.0 (maximum information). Attributes that give more information (high score) can be selected. SubsEval classifiers[3] aim to identify the significance of a subset of attributes that are strongly correlated with the target feature (class) and with a low intercorrelation. We applied two different experiments in the first experiment and the IG classifier was used and resulted in selecting 16 features with IG rank equal to 0.015 for the last feature. Thereafter the IG rank started to drop significantly. The second experiment was done using the SubsEval classifier; and 13 features were extracted from it. Finally, the cross-cutting features between the two experiments were selected, which equal eight features: SeniorCitizen, TotalCharges, tenure_group_Tenure_0-12, Contract_Month-to-Month, InternetService_Fiber optic, Contract_OneYear, PaymentMethod_Electronic check and PaperlessBilling.

E. *Machine Learning Models*

In this study, the following algorithms were selected with the consideration as to their suitability with our domain problem (churn prediction). The chosen algorithms are: decision tree (DT), which is a type of supervised machine learning where the data constantly divides as per specific parameters. The decision is known as the leaves, while the place of dividing the data is called the decision nodes. DT is used for classification purposes, and to be built, calculation entropy and information gain should be done, and then the shortest decision tree is set [17]. Random forests (RF) is an ensemble learning method that operates based on constructing multiple decision trees using an aggregation and bootstrap approach [18]. XGBoost (extreme gradient boosting) is a boosting algorithm that uses the gradient boosting machine/regression tree framework at its core, and the tree ensembles as the default base learners [19]. XGBoost provides a parallel tree boosting. Logistic regression predicts the maximum likelihood for transforming the dependent variables into a logistic variable by passing a linear regression into a sigmoid function. The prediction in random forest is taken by the majority vote. Support vector machines (SVMs) are used for classification, regression, and outliers' detection [17]. In classification, it is based on the idea of margin maximizing hyperplanes which try to find support vectors to separate different classes; it also uses constraints optimization. Kernel trick adds more value to SVM, by calculating the dot product of vectors in a very high dimensional space. K-nearest neighbor algorithm (KNN) is a non-parametric method that stores all available cases and classifies new instances by a majority vote of its neighbors based on a similarity measure (e.g., distance functions) [17]. Naive Bayesian classifier is a probabilistic classifier, based on Bayes' theorem with the independence assumptions between predictors [17]. Finally, the voting classifier takes more than one algorithm model for predictions, creates voting, and sets the prediction based on the majority [20]. In this study, a combination of KNN, SVM, decision tree, and random forest algorithms have been used.

F. *Optimization Methods*

The grid search is primarily an optimization algorithm that allows the best parameters of the machine learning methods to be selected from a list of parameter options [21]. It is most widely known for its use in machine learning to fine-tune a model by identifying the model's parameters to provide the

[2] https://www.cs.waikato.ac.nz/ml/weka/.

[3] http://www.cs.tufts.edu/~ablumer/weka/doc/weka.attributeSelection.CfsSubsetEval.html/.

highest accuracy [21]. Grid search optimization was used to boost the efficiency of the model. On the other hand, cross-validation is a strong preventive approach to over fitting, which results in more accurate and reliable results [22]. Cross-validation was used to achieve a less predicament prediction. In our experiments, the 10-fold cross-validation from Sklearn library[4] was used.

G. Evaluation Criteria

There are prominent metrics for evaluating the classification models in our problem domain, such as accuracy, sensitivity, and AUC [23]. On the basis of the positive class identification problem (i.e., churn), the best model was selected based on the accuracy, whereas other metrics were used for comparison with other similar studies. Accuracy is a ratio of correctly predicted observations to the total observations [24]. Sensitivity is the ratio of correctly predicted positive observations to the total observations in an actual class (i.e., the churn). Of all of the customers that truly churned, how many did the model label? [24]. AUC measures the area under the curve of sensitivity versus specificity for a binary classifier system [24].

IV. Experiment and Results

Eight algorithms were applied to make sure the prediction model was capable of predicting customer churn. The algorithms applied were: LR, NB, RF, DT, KNN, SVM, Voting, and XGBoost. This phase assessed different ways to increase the effectiveness of the classification models through fine tuning. At the beginning of our experiment, two different ways of feature selection were applied, and ended with selecting the most informative features regarding the prediction class (churn). The majority of the models showed accuracy result improvement after applying the grid search technique [25], thus, the parameters resulted from this fine-tunning phase were selected when building the models. The best parameters values resulted from the fine-tunning phase are as follows: first, in SVM, the best parameters values were {'Kernal': 'poly', 'C': 10}. Second, in logistic regression, the best parameters values were {'max_iter': 50, 'multi_class': 'auto', 'solver': 'newton-cg'}. Third, in KNN, the best parameters values were {'algorithm': 'auto', 'n_neighbors': 50}. Fourth, in random forest, the best parameters values were {'bootstrap': True, 'criterion': 'gini', 'max_depth': 5, 'n_estimators': 1000}. Fifth, in the decision tree, the best parameters values were {'criterion': 'gini', 'max_depth': 8}. Finally, the best parameters values were integrated into the voting algorithm, which included KNN, SVM, DT, and RF. NB is not applicable with grid search as it does not have hyperparameters. The accuracy of the XGBoost model has not changed much with the grid search technique due to the fact that the default hyperparameters of the XGBoost model are a combination of the best parameters.

The main experiment was done to obtain the best version of each algorithm for the given problem. By trying different feature selections and the grid technique, all algorithms received an accuracy higher than 70%. Table II shows the final result of all algorithms.

TABLE II. THE ACCURACY OF APPLIED MODELS

Model	**Accuracy**
XGBoost	0.802
SVM	0.799
LR	0.797
KNN	0.797
RF	0.794
Voting	0.794
DT	0.776
NB	0.727

V. Discussion and Analysis

As shown in Table II, boosting (XGBoost) resulted in the highest accuracy of 80%. Different studies demonstrated that XGBoost provided state-of-the-art results on a wide range of problems [26]. For example, in a claim prediction experiment, XGBoost outperformed the other ensemble learning models, i.e., AdaBoost (adaptive boosting), Stochastic gradient boosting, random forest, and online learning-based methods, i.e., neural network, in terms of accuracy [27]. One factor behind the success of XGBoost is the algorithmic optimizations, which uses a novel tree learning algorithm [26]. Additionally, a support vector machine (SVM) provides similar accuracy to XGBoost due to its power in dealing with a high-dimensional feature space and kernel trick. On the other hand, Naive Bayes (NB) algorithm results in the lowest accuracy of 72%, where it shows a significant decrease compared to XGBoost, possibly because Naive Bayes is comes from a weak learner compared to a boosting algorithm. Additionally, some of the descriptive features have a high correlation with one another and the Naive Bayes assumes that all descriptive features are independent of one another [28]. When comparing the results of these models with the researchers who used the same dataset of IBM, which is Mishra's study [8], where the highest accuracy obtained was by using decision tree. However, as it is shown in Table III our accuracy is higher by 1.0%. Another important metric for our problem domain that could prove the enhancement over other studies is sensitivity, as our result shows a 10% improvement. Additionally, in Mishra's study, they focused on increasing AUC, however, our result shows an improvement of nearly 20%. Mishra's study focused their experiment on balancing the data. Conversely, our results demonstrate that the data were not heavily imbalanced and could provide high accuracy and sensitivity. In addition, in [8], correlation, as the feature selection technique, was used, but in our experiment, information gain was selected. Lastly, XGBoost uses parallel tree boosting and votes on the best model, which provides an advantage over regular decision trees (CART).

TABLE III. COMPARISON OF THE RESULTS OF THE STUDY WITH RESULTS OF PREVIOUS STUDIES

[4] https://scikit-learn.org/.

	Model	Sensitivity	AUC	Accuracy
Our Study	XGBoost	0.53	0.843	0.802
Mishra's Study	DT	0.401	0.6581	0.792

In model analysis, in the terms of business goal and strategy in this problem, sensitivity is as important as accuracy, as identifying churn customers is critical to for businesses to take actions to prevent customer churn [29]. Thus, obtaining a higher sensitivity provides value to our model compared to others studies. Another viewpoint is that a company may build an expensive retention program, such hiring sales reps to contact each possible churner, and they may utilize a logistic regression model, which provides a relatively high accuracy of 79%. The company could then deal only with customers who have a high probability to churn since the logistic regression provides a probabilistic interpretation prediction for each customer. Lastly, as it is mentioned in the related work, some of the algorithms obtained a high accuracy of 90% or above, although the result of this study shows the higher result in terms of accuracy, and this could be interpreted by the difference in the features existing in the datasets.

VI. Conclusion

Churn prediction along with client retention strategies are becoming necessary in any telecommunications business where there is significant competition from peer businesses and relative loss related to customer churn. An optimum churn prediction model alone is not adequate, so the right combinational model is a prediction model with an effective retention strategy that can address today's competitive environment.

In this paper we addressed the various data quality issues present in the dataset. We also tried several machine learning algorithms belonging to different family of classifiers. In addition, to make the methodology more suitable and applicable for solving business problems, we also addressed the problem of feature selection. One of the techniques used to enhance model accuracy is the grid search algorithm. Both accuracy and sensitivity have been enhanced by the XGBoost algorithm, compared to the result obtained by the researcher who worked with the same dataset. Some of the insights from the dataset could be helpful in building a retention program. For example, if customers are not pleased with the optical fiber compared to other internet services, customers with month-to-month contracts are more willing to leave than customers with contracts, and paperless and electronic billing customers are more willing to leave compared to paper-billed clients. As a result, it can be concluded that the customers' unhappiness with the services could be due to other related services. Some of the retention strategies that could be used include continuously asking customers for feedback, making special offers, improving customer service, and hiring a convincing sales representative to contact probable churners.

Several limitations have been found, which include churners in the dataset not being clearly specified as to whether their leaving the company's service was done voluntarily or involuntarily, such as customers that were terminated by the company due to unpaid bills. This would affect the predictive result, thus a clear separation between those two types of customers is necessary. Other limitations include the lack of some features, such as the number of minutes a customer used that has changed from one month to another, and the amount of a customer's monthly payment that has changed from one month to another.

For future work, the same experiment could be performed using telecommunications big data with additional features, which can be obtained from different departments, such as the Customer Relationship Management Department. Additionally, some of the recent algorithms could be applied to this domain problem. Moreover, sampling techniques can be also applied to see how certain sampling techniques affect improving results.

References

[1] T. J. Gerpott, W. Rams, and A. Schindler, "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market," *Telecomm. Policy*, vol. 25, no. 4, pp. 249–269, 2001.

[2] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, "Telecommunication subscribers' churn prediction model using machine learning," *8th Int. Conf. Digit. Inf. Manag. ICDIM 2013*, no. 2014, pp. 131–136, Nov. 2013.

[3] V. Umayaparvathi and K. Iyakutti, "A survey on customer churn prediction in telecom industry: datasets, methods and metrics," *Int. Res. J. Eng. Technol*., pp. 2395–56, 2016.

[4] M. Balasubramanian, "Churn prediction in mobile telecom system using data mining techniques," Int. *J. Sci. Res. Publ*., vol. 4, no. 4, pp. 1–5, 2014.

[5] A. Alamsyah and N. Salma, "A comparative study of employee churn prediction model," *Int. Conf. Sci. Technol*., no. 2, pp. 3–6, 2018.

[6] M. Azeem M, Usman A. C., Fong, "A churn prediction model for prepaid customers in telecom using fuzzy classifiers." *Telecommunication Systems* vol. 66, no. 4), pp. 603–14, Dec. 1, 2017.

[7] S. F. Sabbeh, "Machine-learning techniques for customer retention: A comparative study," *International Journal of Advanced Computer Science and Applications,* vol. 9, no. 2, 2018.

[8] K. Mishra and R. Rani, "Churn prediction in telecommunication using machine learning," *Int. Conf. Energy, Commun. Data Anal. Soft Comput.*, no. 2012, pp. 2252–2257, 2017.

[9] Li R, Wang P, Chen Z. A feature extraction method based on stacked auto-encoder for telecom churn prediction. In *Theory, methodology, tools and applications for modeling and simulation of complex systems*, 2016 Oct 8 (pp. 568-576). Springer, Singapore.

[10] H. Xu, Z. Zhang and Y. Zhang, "Churn prediction in telecom using a hybrid two-phase feature selection method," *in 2009 Third International Symposium on Intelligent Information Technology Application,* Shanghai, 2009, pp. 576–579, doi: 10.1109/IITA.2009.392.

[11] J. Qi and Y. Li, "A novel and convenient variable selection method for choosing effective input variables for telecommunication customer churn prediction model," In *2009 IEEE International Conference on Systems, Man and Cybernetics,* San Antonio, TX, 2009, pp. 3217–3222, doi: 10.1109/ICSMC.2009.5346166.

[12] D. Andreea, A. M. Alexandra, S. Stancu et al. (2020). "Churn prediction in telecommunication industry: Model interpretability," *Journal of*

Eastern Europe Research in Business and Economics, 2020, 241442, doi.org/10.5171/2020.241442.

[13] V. Umayaparvathi, K. Iyakutti, "Automated feature selection and churn prediction using deep learning models," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4 no. 3, 1846–54, Mar. 2017.

[14] B. Kumar Singh, K. Verma and A. S. Thoke, "Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification," *Int. J. Comput. Appl.*, vol. 116, no. 19, pp. 11–15, 2015.

[15] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1414–1425, 2012.

[16] H. Fairweather and S. J. Hutt, "On the rate of gain of information in children," *J. Exp. Child Psychol.*, vol. 26, no. 2, pp. 216–229, 1978.

[17] J. Kelleher, B. Mac Namee, and A. D'Arcy, Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press; 2020 Oct 20.

[18] L. Breiman, 2001. Random forests. *Machine learning*, vol. 45, no. 1, 2001 Oct. Available: <https://link.springer.com/article/10.1023/A:1010933404324> .

[19] T. Chen, and C. Guestrin, XGBoost: A scalable tree boosting system. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016 [Online], Available: <https://arxiv.org/abs/1603.02754> .

[20] Scikit-learn.org. 2020. 1.11. Ensemble methods-Scikit-learn 0.23.2 Documentation [Online], Available: <https://scikit learn.org/stable/modules/ensemble.html#voting-classifier>

[21] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *30th Int. Conf. Mach. Learn. ICML* 2013, no. PART 1, pp. 115–123, 2013.

[22] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, pp. 40–79, 2010.

[23] S. Khodabandehlou and M. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 1-2, pp. 65–93, 2017.

[24] S. Khodabandehlou and M. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," *Journal of Systems and Information Technology*, vol. 19, no. 1-2, pp. 65–93, 2017.

[25] G. Behera and N. Nain, "Grid search optimization (GSO) based future sales prediction for big mart," in *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2019.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD*, 16, 2016.

[27] M. A. Fauzan and H. Murfi, "The accuracy of XGBoost for insurance claim prediction," *International Journal of Advances in Soft Computing and its Applications*, vol. 10, no. 2, pp. 159–171, 2018.

[28] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, 3(22), pp. 41–46.

[29] M. Yildiz and S. Albayrak, "Customer churn prediction in telecommunication," in *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, 2015, pp. 256–259.