**Machine Learning Techniques for Customer Churn Prediction**

1. **Problem Statement and Goal**

Customer churn is a significant issue and one of the most challenging problems for large companies. We know that, increasing the retention period of customers is more profitable strategy than acquiring new customers. For this reason, finding factors that increase customer churn is important to take necessary actions to reduce this churn. If we go into a little more detail; predicting customer churn behavior, analyzing the root causes of customer churn, finding the links that need to be improved in the process of operation and management, winning back churned customers, and establishing a stronger customer relationship have become the strategic focus. That’s why, we aim to determine the customer churn by analyzing their behavior and try to put effort into retaining the customers and seek to develop means to predict potential customer to churn. We have experimented a number of algorithms such as Gradient Boosting and XGBoost to build the predictive model of customer Churn after developing our data preparation. In addition, the logistic regression algorithm produces a better prediction effect, based on which the level of importance of customer churn factors can be seen. In this project, the logistic regression model is used to predict the trend in customer churn, assist enterprises in finding out the early warning signals of customer churn, and determine the tendency of customer churn. The dimensions of input features also are represented by using feature importance plots, which will be studied in this project.

1. **Literature Search**

Customer churn is a term used to describe the loss of important clients to competitors by service organizations in the telecommunications industry. There have been several developments in the telecommunications business in recent years, including market liberalization, increased competition, new services, and new technology. Customer churn results in a significant loss of communications services, making it a critical issue. Data mining approaches have recently evolved to address the difficult problem of customer turnover in the telecommunications industry.

Gavril et al. [1] reported an advanced data mining algorithm for predicting churn for prepaid customers using a dataset for 3333 customers' call information with 21 characteristics and a dependent churn parameter with two values. Some features include the quantity of incoming and outgoing messages, as well as voicemail for each client. To minimize data dimensionality, the author used the principal component analysis technique "PCA." To forecast churn factor, tree machine learning methods such as neural networks, support vector machines, and Bayes networks were applied. The author utilized AUC to evaluate the algorithms' performance. The AUC values for Bayes networks, neural networks, and support vector machines were 99.10 percent, 99.55 percent, and 99.70 percent, respectively.

Ahmad et al[2], because of their diversity and application in this form of prediction, four tree-based algorithms chose. The algorithms in question include Decision Tree, Random Forest, GBM Tree Algorithm, and XGBOOST Algorithm. Days since the last outgoing transaction, total balance, average radio access type, and local cluster coefficient are all crucial factors that the study highlights.

Praveen et al. [3] conducted a comparison of machine learning models for customer churn prediction, including support vector machine, decision tree, naive bayes, and logistic regression. They then investigated the impact of boosting methods on classification accuracy. SVM-POLY utilizing AdaBoost outperformed the rest in the findings. However, by combining feature selection procedures such as uni-variate selection and others, classification accuracy may be enhanced even more.

Horia Beleiu et al. [4] used three machine learning algorithms to predict customer churn: neural networks, support vector machines, and bayesian networks. To minimize the dimensionality of the data, principal component analysis (PCA) is used in the feature selection process. However, the feature selection process may be improved by using an optimization technique, which improves classification accuracy. Gain measure and ROC curve were employed in the performance evaluation.

The authors, J. Burez et al. [5], attempted to represent the class imbalance problem. They used logistic regression and random forest with resampling. Additionally, boosting methods were used. AUC and Lift are taken into account in the performance analysis. They also investigated the impact of improved sampling techniques such as CUBE, however the results did not increase performance. However, the class imbalance problem may be tackled more effectively by using optimization-based sampling strategies.

Huang et al. [6] investigated the issue of client attrition in a big data platform. The researchers' objective was to demonstrate that big data significantly improves the process of anticipating churn based on the amount, diversity, and velocity of the data. To manufacture the fissures in data from the Operation Support and Business Support departments at China's largest telecoms firm, a big data platform was required. AUC was used to evaluate the Random Forest method.

1. **Dataset Description**

The Telco Customer Churn data is acquired from Kaggle. (<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>). The raw data contains 7043 rows (customers) and 21 columns (features). Features have information about:

**Demographic Informations**

Gender (Female,Male),

SeniorCitizen: Whether they are a senior citizen or not (Yes,No),

Partner: Whether they have a partner or not (Yes,No),

Dependents: Whether they have a dependents or not (Yes,No)

**Services Provided to the Customer:**

PhoneService (Yes, No)

MultipleLines (Yes, No, No phone service)

InternetService (DSL, Fiber optic, No)

OnlineSecurity (Yes, No, No internet service)

 OnlineBackup (Yes, No, No internet service)

DeviceProtection (Yes, No, No internet service)

TechSupport (Yes, No, No internet service)

StreamingTV (Yes, No, No internetservice)

StreamingMovies (Yes, No, No internet service)

**Customer Account Details:**

Tenure: Number of months the customer has stayed with the company

Contract (Month-to-month, One year, Two year)

PaperlessBilling (Yes, No)

MonthlyCharges: The monthly charged amount to the customer

TotalCharges: The total amount charged to the customer

PaymentMethod (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)

**Our Target:** Churn: Whether the customer has churned or not

1. **Methods**

We used 5 machine learning models in our study, which are as follows.

1. Logistic Regression

2. Gradient Boosting

3. XGBoost

4. XGBoost with Hyperparameter Tuning

5. Gradient Boosting with Hyperparameter Tuning

The description of each method we used are given in the following section.

* 1. **Logistic Regression**

Logistic regression is a classification algorithm despite the "regression" in its name. That is, it is frequently used in two-class classification problems such as guessing whether the given information belongs to a man or a woman.

**Advantages:**

* Logistic regression is easy to implement and interpret.
* It performs quite well if the dataset is linearly separable.
* It is less prone to overfitting, but can be overfitted on large datasets.

**Disadvantages:**

* If the number of observations is less than the number of features, Logistic Regression should not be used, otherwise overfit may occur.
* For logistic regression to discriminate, the data set must be linearly separable.

Logistic regression uses the Sigmoid (Logistics) Function to classify. The sigmoid function is an “S” shaped curve. The sigmoid function is simply the function used to compress our data between 0 and 1. With this function we can classify. It is also frequently used under activation functions in Deep Learning [7].

* 1. **Gradient Boosting**

Gradient Boosting or GBM is another ensemble machine learning algorithm that works for both regression and classification problems. GBM uses the amplification technique, combining a number of weak learners to create a strong learner. Regression trees used as basic learners are based on errors calculated by a tree preceding each tree in series [8].

**Advantages:**

* It is extremely powerful machine learning classifier.
* Accepts various types of inputs that make it more flexible.
* It can be used for both regression and classification.
* It gives you features important for the output.

**Disadvantages:**

* It takes longer time to train as it can’t be parallelized.
* More likely to overfit as it obsessed with the wrong output as it learns from past mistakes.
* In some cases, Tuning is very hard as it has many parameters to tune [9].
  1. **XGBoost**

XGBoost (extreme Gradient Boost) is an advanced implementation of the gradient boosting algorithm. XGBoost has proven to be an extremely effective machine learning algorithm, widely used in machine learning competitions. XGBoost has high predictive power and is almost 10 times faster than other gradient boosting techniques. It also includes several adjustments that reduce overfitting and improve overall performance. For this reason, it is also known as the "regular boost" technique [8].

**Advantages:**

* Works well with massive data sets.
* Tree algorithms like XGBoost and Random Forest don't require normalized features and operate well with nonlinear, non-monotonic, or clustered data.

**Disadvantages:**

* Tree algorithms like XGBoost and Random Forest might over-fit the data, especially if the trees are very deep and the data is noisy.

In order to increase the performance of XGBoost and Gradient Boosting models, we conducted hyperparameter tuning. Grid Search method is used in hyperparameter tuning phase. The parameters, which are tuned and the tuned values of these parameters are given below.

* 1. **XGBoost with Hyperparameter Tuning**

While tuning the parameters of XGBoost model, the parameters to be tuned are given in Table 1.

**Table 1.** Tuned parameters and their values for XGBoost model

|  |  |
| --- | --- |
| **Parameter** | **Values** |
| n\_estimators | [100,250, 500] |
| subsample | [0.6, 0.8, 1.0] |
| gamma | [0,1,5] |
| max\_depth | [3, 4, 5] |
| learning\_rate | [0.1,0.3, 0.5] |

After implementing the Grid Search method and running the code, each parameter received a best value. These values are given below.

n\_estimators = 100

subsample = 0.6

gamma = 1

max\_depth = 3

learning\_rate = 0.1

The architecture of the Tuned XGBoost model is created by using the values above.

* 1. **Gradient Boosting with Hyperparameter Tuning**

The same methodology is used in the tuning of Gradient Boosting model. In Table 2, the parameters tuned and their values are given.

**Table 2.** Tuned parameters and their values for Gradient Boosting model

|  |  |
| --- | --- |
| **Parameter** | **Values** |
| n\_estimators | [5,50,250,500] |
| max\_depth | [1,3,5,7,9] |
| learning\_rate | [0.01,0.1,1,10,100] |

After implementing the Grid Search method and running the code, each parameter received a best value. These values are given below.

n\_estimators = 500

max\_depth = 1

learning\_rate = 0.1

The architecture of the Tuned Gradient Boosting model is created by using the values above.

1. **Experiment Results and Discussion**

For each model that we created, we acquired the results of four different performance measures, which are accuracy, precision, recall and F1 score. Results of each model can be found in Table 3.

**Table 3**. Performance measure results of machine learning models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| Logistic Regression | 0.808874 | 0.645669 | 0.550336 | 0.594203 |
| XGBoost | 0.806598 | 0.650704 | 0.516779 | 0.576060 |
| XGBoost (Tuned) | 0.809443 | 0.653846 | 0.532438 | 0.586930 |
| Gradient Boosting | 0.811149 | 0.662890 | 0.523490 | 0.585000 |
| Gradient Boosting (Tuned) | 0.814562 | 0.676385 | 0.519016 | 0.587342 |

Results presented in Table 3 show that Logistic Regression model has the highest recall and F1 score among all models created for this study, whereas Tuned Gradient Boosting has the highest accuracy and precision. Since Tuned Gradient Boosting model also has a F1 Score close to the highest value and has the highest accuracy and precision values, we selected Tuned Gradient Boosting model to be the best one for our dataset.

After selecting the best model, we constructed the confusion matrix of the results acquired from running the Tuned Gradient Boosting model. In Figure X, the confusion matrix can be seen.

Chart

Description automatically generated

**Figure 1**. Confusion Matrix of Tuned Gradient Boosting

Analyzing Figure 1, it is seen that the True Positive (TN) value of the model is 1200, False Positive (FP) value is 111, False Negative (FN) value is 215 and True Negative (TN) value is 232. Since the total of True Positive and True Negative values corresponds to around 80% of total predictions, we conclude that the model is successful. After the confusion matrix, we created the plots of feature importance for 3 different models of the study, namely Gradient Boosting, Tuned Gradient Boosting and XGBoost.

Chart, bar chart

Description automatically generated

**Figure 2**. Feature importance of tuned gradient boosting model

As it can be seen in Figure 2, four features show great importance compared with the rest of the features. Contract of the customer being month-to-month or not, tenure, internet service being fiber optic or not and payment method being electronic check or not shows great importance for Tuned Gradient Boosting model and variations in these variables have a high chance of affecting the prediction of the model.

Chart

Description automatically generated

**Figure 3.** Feature importance related to gradient boosting model

In Figure 3, the feature importance plot of Gradient Boosting model is given. Unlike Tuned Gradient Boosting model, first 6 features show high importance. This model includes the same important features and includes two additional important features, which are monthly charges and total charges.

Chart, bar chart

Description automatically generated

**Figure 4.** Feature importance related to XGBoost model

In Figure 4, the feature importance plot of XGBoost algorithm is given. It is seen that XGBoost model has a different feature importance distribution compared with other two models. In Gradient Boosting models, least important features show little to no importance at all. However, analyzing XGBoost algorithm’s feature importance plot, it is seen that even the least important feature shows importance in the model.

Analyzing the overall results of this study, we observed that hyperparameter tuning is in fact a very beneficial method to increase the overall performance of any machine learning model. As it can be seen in Table 3, tuning the parameters of Gradient Boosting and XGBoost models increased the values of performance measures for both models.

1. **Conclusion**

Acquiring new customers is the goal of every company, but since this is a difficult and costly task, it is also very important for companies to have loyal customers and retain the existing ones.Companies can take steps for retaining their customers if they can predict whether the customer has a potential to churn. Predicting whether their customers will churn is critical at this point.At this point, companies can use use machine learning algorithms to predict whether their customers will churn. In our study we used Logistic Regression,Gradient Boosting and XGboost algorithms.Also tried  XGboos and Gradient Boosting with Hyperparameter Tuning. We acquired more accurate results with Tuned Gradient Boosting model with using our dataset. We noticed that tenure and contract type of the customer have a great effect on customer churn but gender has not such an effect. Companies can determine the customers with a high probability of churn and take actions to retain them.After predicting whether customers will churn or not, promotions can be shaped according  to customer groups.

**References**

[1] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[2] Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data 6(1):28

[3] Asthana P (2018) A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics 119(10):1149–1169

[4] Brându¸soiu, I., Toderean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016 International conference on communications (COMM), pp. 97–100. IEEE (2016)

[5] Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Systems with Applications 36(3):4626–4636

[6] Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., Zeng, J.: Telco churn prediction with big data. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607–618 (2015)

[7] Akca, M. F. (2021). *Lojistik Regresyon Nedir? Nasıl Çalışır? - Mehmet Fatih AKCA*. Medium. <https://mfakca.medium.com/lojistik-regresyon-nedir-nas%C4%B1l-%C3%A7al%C4%B1%C5%9F%C4%B1r-4e1d2951c5c1>

[8] Singh, A. (2020). *Ensemble Learning | Ensemble Techniques*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>

[9] Kurama, V. (2021). *Gradient Boosting for Classification*. Paperspace Blog. https://blog.paperspace.com/gradient-boosting-for-classification/