Educational Administration: Theory and Practice
2024, 30(1), 3241-3251
ISSN: 2148-2403
https://kuey.net/
Research Article

Customer Churn Prediction Using Ensemble Techniques And Algorithms

Pallabi Baruah[1*], Bhairab Sarma[2]

[1*]Research Scholar, Department of Computer Science, University of Science and Technology,Ri-Bhoi, Meghalaya, India
[2]Associate Professor, Department of Computer Science, University of Science and Technology,Ri-Bhoi, Meghalaya, India

***Corresponding Author:** Pallabi Baruah
*Email: pallavibaruahguha@gmail.com

Citation: Pallabi Baruah , (2024) Customer Churn Prediction Using Ensemble Techniques And Algorithms, *Educational Administration: Theory And Practice, 30(1),* 3241-3251
Doi: 10.53555/kuey.v30i1.6126

ARTICLE INFO

Submitted- 12/March/2024
Reviewed- 15/April/2024
Accepted- 05/May/2024
Published-30/May/2024

ABSTRACT

Customer churn or customer attrition is the phenomenon where customers of a business discontinue buying or communication with the organization. A high churn means that a higher number of customers no longer want to buy goods and services from the business. Customer churn rate or customer attrition rate is the mathematical calculation of the percentage of customers who are not likely to make another purchase from a business. Customer churn happens when customers decide to not continue purchasing products/services from an organization and end their association. It is an integral parameter forthe organization since acquiring a new customer could cost almost 7 times more than retaining an existing customer. Customer churn can prove to be a roadblock for an exponentially growing organization and a retention strategy should be decided in order to avoid an increase in churn rates. Customer churn is a critical aspect of business operations, and understanding its causes and implementing effective strategies to mitigate it is essential for long-term success.In this research work, different models are used to predict the rate of customer churn rate and it is found that Random forest gives a preferable accuracy.

Keywords: Customer, churn, analysis, prediction, accuracy, models.

Introduction

Customer churn is defined as when customers or subscribers discontinue doing business with a firm or service. Customers in the telecom industry can choose from a variety of service providers and actively switch from one to the next. The telecommunications business has an annual churn rate of 15-25 percent in this highly competitive market. Individualized customer retention is tough because most firms have a large number of customers and can't afford to devote much time to each of them. The costs would be too great, outweighing the additional revenue. However, if a corporation could forecast which customers are likely to leave ahead of time, it could focus customer retention efforts only on these "high risk" clients. The ultimate goal is to expand its coverage area and retrieve more customers loyalty. The core to succeed in this market lies in the customer itself. Customer churn is a critical metric because it is much less expensive to retain existing customers than it is to acquire new customers. To reduce customer churn, telecom companies need to predict which customers are at high risk of churn. To detect early signs of potential churn, one must first develop a holistic view of the customers and their interactions across numerous channels, including store/branch visits, product purchase histories, customer service calls, Web-based transactions, and social media interactions, to mention a few. As a result, by addressing churn, these businesses may not only preserve their market position, but also grow and thrive. More customers they have in their network, the lower the cost of initiation and the larger the profit. As a result, the company's key focus for success is reducing client attrition and implementing effective retention strategy.

Importance of Predicting Customer Churn

The ability to be able to predict that a certain customer is at a very high risk of churning, while there is still some time to do something significant about it, itself represents a great additional potential revenue source for any business.

- It's a fact acquiring new customers is a costly affair but losing the existing customers will cost even more for the business or the organization. As existing paying customers are usually returning customers who if happy will purchase repeatedly from your brand.
- The competition in any market is on a rise and this encourages organizations to focus not only on new business but also on retaining existing customers.
- The most essential step toward predicting customer churn is to start awarding existing customers for constant purchases and support.
- An entire customer journey leads to customer churn and not just a few incidents. Due to the priority of avoiding customer churn, organizations should start offering incentives for purchases of these soon-to-churn customers.
- As mentioned earlier, a customer's intention to stop using a particular product/service may always be a decision formed over time. There are various factors that lead to this decision and it's important for organizations to understand each and every factor so that customers can be convinced to stay and keep making purchases. This can be done by constantly conducting customer satisfaction surveys and analyzing the received feedback.

Review of Literature

According to Ahmad A.K(2019)Four tree based algorithms were chosen because of their diversity and applicability in this type of prediction and they were Decision Tree, Random Forest, GBM tree algorithm, and XGBOOST algorithm. Qureshii SA(2013) reported that predicting the customers who are likely to leave the company will represent potentially large additional revenue source if it is done in the early phase.

Suh Y(2023) in his research identifed and calculated the inf;uence of key variables on individual customer churn to enable a business person (rental care customer management stfff) to carry out customer-tailored marketing to address the cause of the churn.

Lalwani P.,Mishra M.K.(2022) in their research work provided a compara-tive study of Customer Churn prediction in Telecommunication Industry using famous machine learning techniques such as Logistic Regression, Naïve Bayes, Support Vector Machines, Decision Trees, Random Forest, XGBoost Classifier, CatBoost Classifier,AdaBoost Classifier and Extra tree Classifier. In the research further the experimental results show that two ensemble learning techniques that is Adaboost classifier and XGBoost classifier gives maximum accuracy with respect to others with an AUC score of 84% for the churn prediction problem with respect to other models.

Kazemi M.(2023) in his research work carefully examines and compares different optimized XGB models to predict the penetration rate of rotary drilling. In this work further, these models were created by combining XGB with four hyperparameter tuning methods including random search, grid search, and intelligent optimization algorithms like HHO and DA.

Sweta P.(2021) in her study presents a prescient examination way to deal with improve client churn in the telecom business just as the utilization of a method normally utilized in retail settings known as "strategically pitching" or "market container investigation".

Customer Churn

Customer Churn is the phenomenon in which a client stops doing business with an entity. Users can stop using a company's product or service for a variety of reasons, such as affordability, dissatisfaction with the offering, and bad customer service. More often than not, customers who churn from one company will start doing business with their competitor. For instance, if one isn't happy with hir/her current mobile service provider due to slow Internet speed, they are likely to switch to an alternative. The act of churning isn't one that happens suddenly. If one experience low network bandwidth, he is likely to tolerate it for a month or two. During this period, he/she would probably contact customer support, check his network speed, and leave a review on social media expressing his dissatisfaction.

If the data scientists at a current provider can collect this data and ascertain that the behavior is similar to that of other customers who have churned in the past, they will immediately alert the marketing team, who will then reach out and attempt to cater to the needs in the best way possible. They may provide one with special promotions, upgrade plans, and work on creating a satisfactory user experience for that customer to prevent him from leaving.

Customer churn prediction is one of the most popular use cases of data science in marketing. Companies incur a lot of costs when users churn since it is expensive to replace an existing customer. Due to this, most mid to large-sized organizations will have some sort of churn prediction mechanism in place.
Offering streamlined experiences, competitive pricing, good service, and strong CRM contact center solutions are crucial for increasing customer retention. But being able to predict churn and address the customers' concerns timely is the key to nurturing loyal clients.

Telco Customer Churn dataset from Kaggle is used for this analysis. And there is a need of a Python IDE to run the codes provided here. Finally, it should have the following libraries installed - pandas, Matplotlib, Seaborn, Scikit-Learn, and Imblearn.
First, the dataframe should be loaded into Python with the pandas library and take a look at its head.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information - how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

#	Column	Non-Null Count	Dtype
0	customerID	7043 non-null	object
1	gender	7043 non-null	object
2	SeniorCitizen	7043 non-null	int64
3	Partner	7043 non-null	object
4	Dependents	7043 non-null	object
5	tenure	7043 non-null	int64
6	PhoneService	7043 non-null	object
7	MultipleLines	7043 non-null	object
8	InternetService	7043 non-null	object
9	OnlineSecurity	7043 non-null	object
10	OnlineBackup	7043 non-null	object
11	DeviceProtection	7043 non-null	object
12	TechSupport	7043 non-null	object
13	StreamingTV	7043 non-null	object
14	StreamingMovies	7043 non-null	object
15	Contract	7043 non-null	object
16	PaperlessBilling	7043 non-null	object
17	PaymentMethod	7043 non-null	object
18	MonthlyCharges	7043 non-null	float64
19	TotalCharges	7043 non-null	object
20	Churn	7043 non-null	object

The dataframe has 21 columns related to telecom user subscription behavior. In total there are 7043 rows.The target is used to guide the exploration is Churn. Entire dataset is checked for missing values and also is observed that it has no peculiar pattern that stands out.
Each user is identified through a unique customer ID. There are 19 independent variables used to predict the target feature – customer churn. In this dataset, customer churn is defined as users who have left within the last month. 26.6 % of customers switched to another firm.
Customers are 49.5 % female and 50.5 % male.

Fig 2: Gender and Churn Distributions

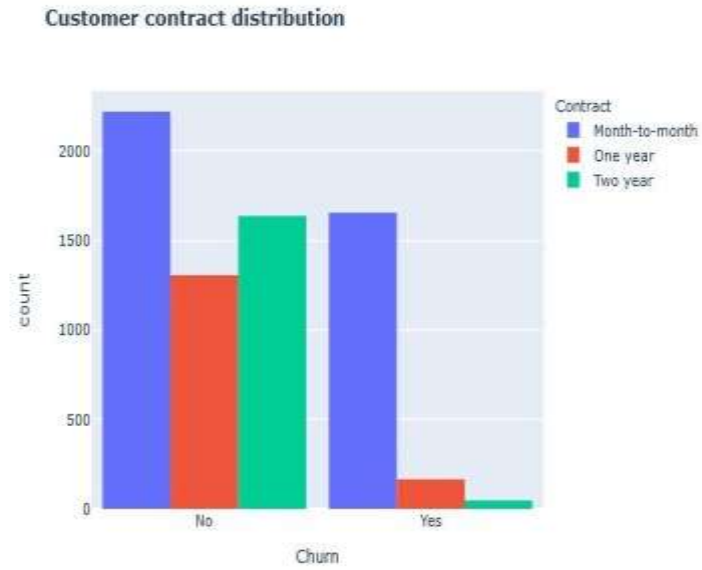


Fig 2: Customer Contract Distribution

About 75% of customer with Month-to-Month Contract opted to move out as compared to 13% of customers with One Year Contract and 3% with Two Year Contract

Fig 3: Customer Payment Distribution W.R.T Churn

Further it has been observerd that major customers who moved out were having Electronic Check as Payment Method. Customers who opted for Credit-Card automatic transfer or Bank Automatic Transfer and Mailed Check as Payment Method were less likely to move out.

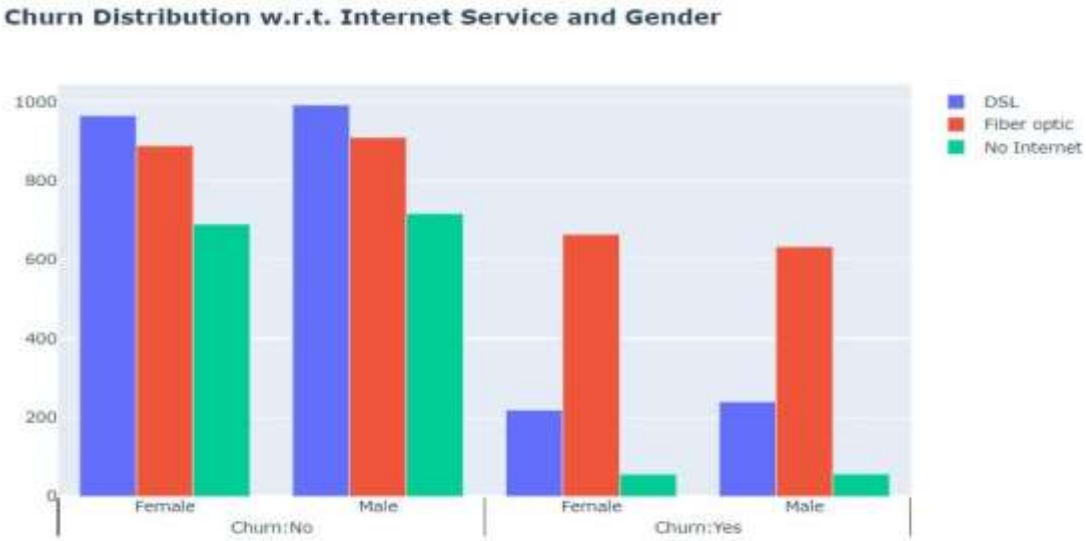


Fig 4: Churn Distribution W.R.T Internet Service and Gender

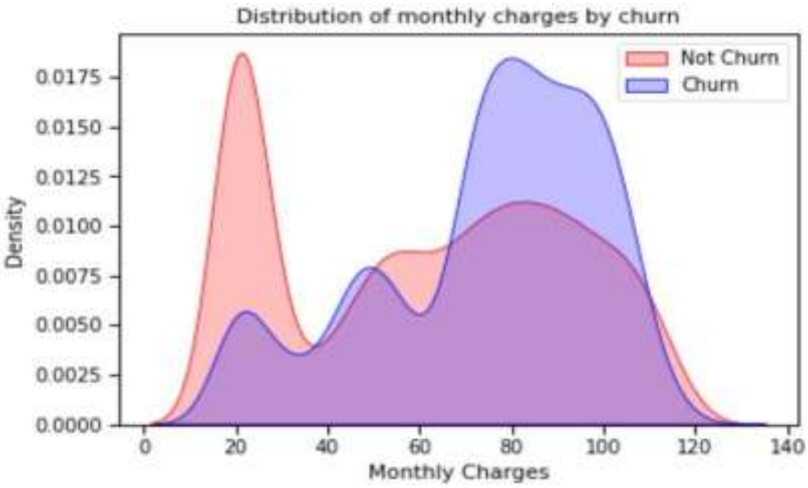


Fig 5: Distribution of monthly charges by churn

A lot of customers choose the Fiber optic service and it's also evident that the customers who use Fiber optic have high churn rate, this might suggest a dissatisfaction with this type of internet service.Customers having DSL service are majority in number and have less churn rate compared to Fibre optic service.

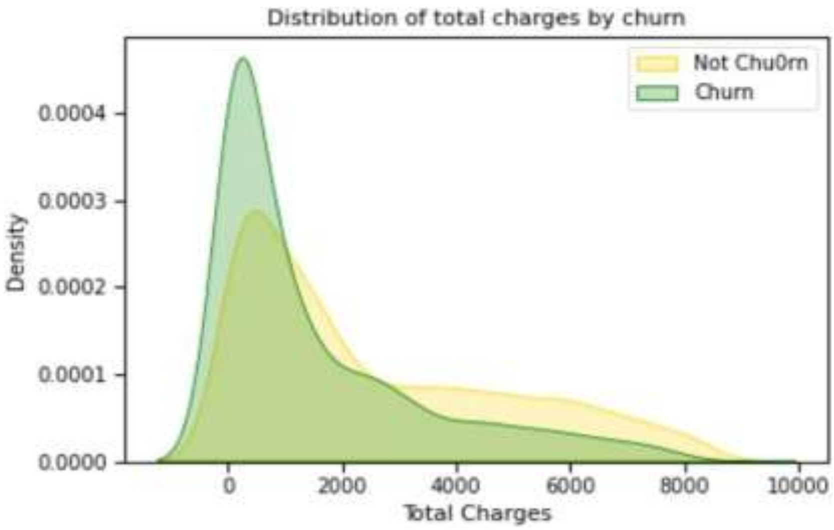
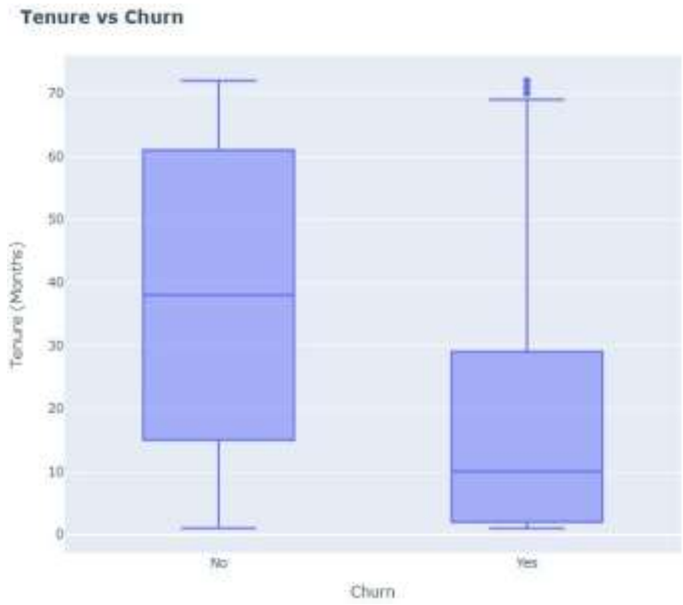


Fig 6: Distribution of total charges by churn

Customers without dependents are more likely to churn.It can be observed that the fraction of senior citizen is very less. Most of the senior citizens churn. Customers who have joined newly are very much prone to churn.

Fig 7: Tenure vs Churn



Data Preprocessing

The splitting of dataset is based on the random manner, system automatically divided the two different dataset in terms of ratio 70:30 manner which is one of the standards mapping parameter to train and test the dataset in machine learning model. Dataset form business industry is used to study on churn customer and its related features which are significant towards the non-churning features of customers in business industry.Since the numerical features are distributed over different value ranges, standard scalar is used to scale them down to the same range.

Model Evaluation and predictions

The three algorithms are used mainly for the comparison and evaluation of prediction-Logistic Regression, AdaBoost Classifier and Gradient Boosting Classifier. The prediction accuracy of Logistic Regression is found to be 80.9%, for AdaBoost Classifier is 80.7 % and Gradient Boosting Classifier is 80.8%.

Logistic Regression

It is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

The logistic regression model transforms the linear regression function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. Kanade V.(2023) in hos research work explained this function as the logistic function.

The sigmoid function is referred to as an activation function for logistic regression and is defined in Equation of Logistice regression(1) as:

$$f(x) = \frac{1}{1+e^{-x}} \qquad \text{------------(1)}$$

where,

- e = base of natural logarithms
- value = numerical value one wishes to transform

The following equation(2) represents logistic regression, Sigmoid Function:

$$y = \frac{e^{(b_0 + b_1X)}}{1 + e^{(b_0 + b_1X)}} \qquad \textbf{-----------(2)}$$

here,

- x = input value
- y = predicted output
- b0 = bias or intercept term
- b1 = coefficient for input (x)

This equation is similar to linear regression, where the input values are combined linearly to predict an output value using weights or coefficient values. However, unlike linear regression, the output value modeled here is a binary value (0 or 1) rather than a numeric value.

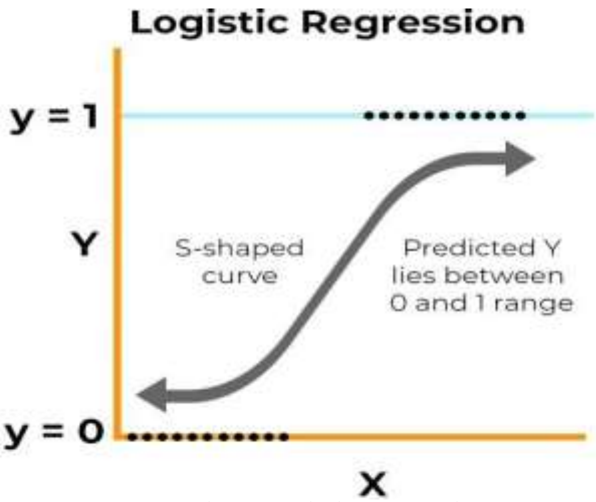


Fig 8: Logistic Regression

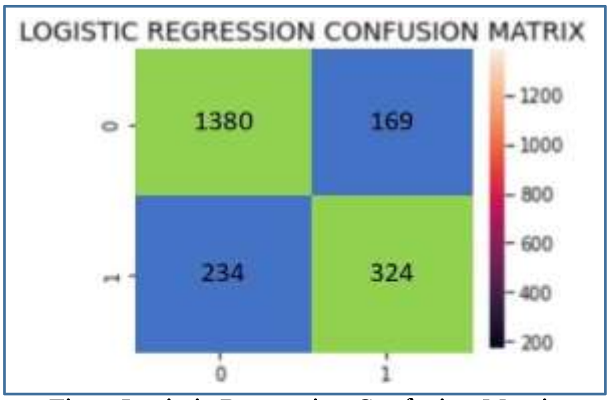


Fig 9: Logistic Regression Confusion Matrix

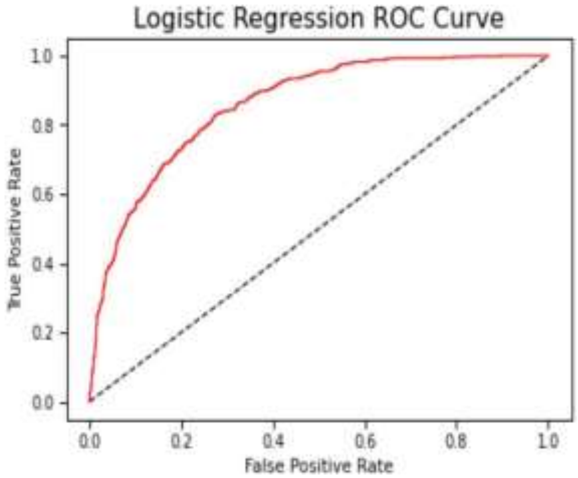


Fig 10: Logistic Regression ROC Curve

	precision	recall	f1-score	support
0	0.85	0.90	0.87	1549
1	0.67	0.55	0.60	561
accuracy			0.81	2110
macro avg	0.76	0.72	0.74	2110
weighted avg	0.80	0.81	0.80	2110

AdaBoost Classifier

Ada-boost or Adaptive Boosting is one of ensemble boosting classifier proposed by Yoav Freund and Robert Schapire in 1996. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier. The basic concept behind Adaboost is to set the weights of classifiers and training the data sample in each iteration such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as base classifier if it accepts weights on the training set. Adaboost should meet two conditions:

1. The classifier should be trained interactively on various weighed training examples.
2. In each iteration, it tries to provide an excellent fit for these examples by minimizing training error.

AdaBoost refers to a particular method of training a boosted classifier. A boosted classifier is a classifier of the form -

$$F_T(x) = \sum_{t=1}^{T} f_t(x) \qquad \text{.................................-(4)}$$

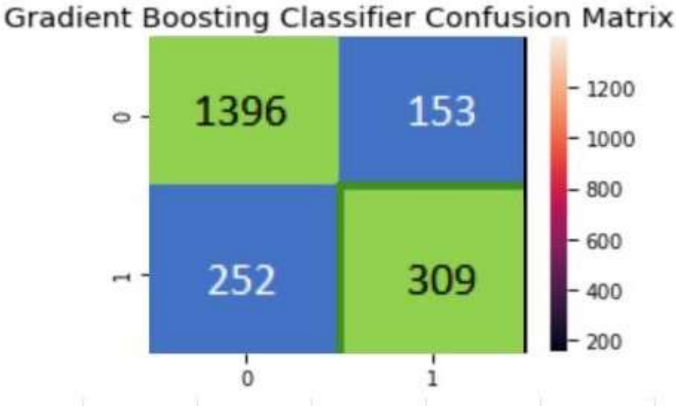
where each f_t is weak learner that takes an object x as input and returns a value indicating the class of the object.

Fig 11: AdaBoost Classifier Confusion Matrix

	precision	recall	f1-score	support
0	0.85	0.90	0.87	1549
1	0.67	0.55	0.60	561
accuracy			0.81	2110
macro avg	0.76	0.72	0.74	2110
weighted avg	0.80	0.81	0.80	2110

Gradient boosting classifiers

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting. Gradient boosting models are becoming popular because of their effectiveness at classifying complex datasets, and have recently been used to win many Kaggle data science competitions. Gradient boosting classifiers are specific types of algorithms that are used for classification tasks, as the name suggests. Masui T(2022), in his research work explained the algorithm with the following steps-



Gradient Boosting Algorithm

1. Initialize model with a constant value:

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

2. for $m = 1$ to M:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ $for\ i = 1,\dots,n$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\dots,J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ $for\ j = 1,\dots,J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

Fig 12: Gradient Boosting Classifier Confusion Matrix

Voting Classifier Voting Classifier

Prediction is done based of the final model based on the highest majority of voting and it's score.

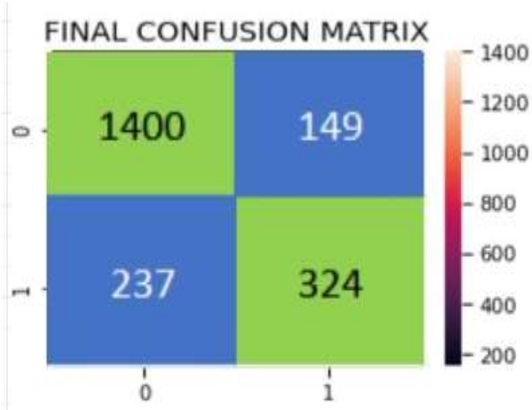


Fig 13: Final Confusion Matrix

Let's now predict the final model based on the highest majority of voting and check it's score.

	precision	recall	f1-score	support
0	0.86	0.90	0.88	1549
1	0.68	0.58	0.63	561
accuracy			0.82	2110
macro avg	0.77	0.74	0.75	2110
weighted avg	0.81	0.82	0.81	2110

Result and Discussion

It is observed from the confusion matrix that: there are total 1400+149=1549 actual non-churn values and the algorithm predicts 1400 of them as non churn and 149 of them as churn. While there are 237+324=561 actual churn values and the algorithm predicts 237 of them as non churn values and 324 of them as churn values.

Customer churn accuracy achieves 82% f1-score which is a fairly good score. Various strategies can be implemented to eliminate customer churn. The best way to avoid customer churn is for a company to truly know its customers. This includes identifying customers who are at risk of churning and working to improve their satisfaction. Improving customer service is, of course, at the top of the priority for tackling this issue. Building customer loyalty through relevant experiences and specialized service is another strategy to reduce customer churn. Some firms survey customers who have already churned to understand their reasons for leaving in order to adopt a proactive approach to avoiding future customer churn.

Conclusion

This Research paper proposes an three algorithm - Logistic Regression, AdaBoost algorithm and Gradient Boost Classifier and proposes an algorithm that predict the final model based on the highest majority of voting and its score is checked based on the performance Satisfactory results are obtained. It is more accurate and the calculation time is medium. Therefore, the subsequent research can be based on the algorithm to improve the calculation speed of the algorithm.

References

1. Ahmad, A.K(2019), Customer churn prediction in telecom using machine learning in big data platform, Springer Open Journal of Big Data, https://doi.org/10.1186/s40537-019-0191-6.
2. Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. (2013), Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p.131–6
3. Suh Y.(2023), Machine learning based customer churn prediction in home appliance rental business, Springer Open Journal of Big Data, https://doi.org/10.1186/s40537-023-00721-8.
4. Lalwani P.,Mishra M.K.,Chadha J.S., Sethi P.(2021), Customer churn prediction system: a machine learning Approach, https://doi.org/10.1007/s00607-021-00908-y
5. 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)

6. . Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Systems with Applications 36(3):4626–4636
7. Sweta P.(2021). Customer Churn Prediction In The Telecom Domain Using Machine Learning Algorithms, https://shodhganga.inflibnet.ac.in/handle/10603/471774.
8. . Amali A. I.(2020). An Extensive Performance Analysis on Framework of Customer Churn Prediction Through Stochastic Gradient Descent based Deep Learning Model, https://shodhganga.inflibnet.ac.in/handle/10603/324818
9. .Nabavi, Z., et al.(2023), A hybrid model for Backbreak prediction using XGBoost machine learning and metaheuristic algorithms in Chadormalu iron minswetae. J. Min. Environ, https://doi.org/10.22044/jme.2023.12796.2323
10. Kanade V.(2023), What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices, April 18, 2022,https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/#lg=1&slide=0.
11. Masui T,(2022), All You Need to Know about Gradient Boosting Algorithm – Part 1. Regression, https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502.
12. Zhang1 Y.,Ni M., Zhang1 C., Liang1 S., Fang S., Li K,Tan Z.(2019),Research and Application of AdaBoost Algorithm Based on SVM, 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC 2019)