Grenze International Journal of Engineering and Technology, June Issue

Boosting Comparison with Customer Churn

Kshitij Singh Chouhan[1] and Dr Nilamadhab Mishra[2]
[1-2]VIT Bhopal University Madhya Pradesh, India- 466114
Email: kshitijsinghkvj@gmail.com, nmmishra77@gmail.com

***Abstract*— Boosting algorithms is one of the greatest tool in machine learning and over the years it has various versions and they all have their own identities. The main objective of the paper is to provide comprehensive study on various algorithms and try to solve a challenging classification problem of customer churn and observe the result of these algorithms too see which of them work the best in the raw state possible. In this paper of we used a dataset "Telecommunication Customer Churn Dataset" and we will few preprocessing to level the ground for all our models/algorithms. Finally we will run some test like accuracy, cross validation to find the best performing model among the boosting algorithm which we will use**

***Index Terms*— Machine Learning, Boosting Algorithm, Customer Churn.**

I. Introduction

In this research paper we are comparing and seeing the effects of various boosting algorithm in their raw state (which mean we are not going to tune our algorithms we only do the things which are needed by our algorithms to work) on a predicting problem and for that we are using customer churn prediction as our target problem. Customer Churn which is better known as customer turnover which is a very important term in the aspect of business as it show the incites that why a customer has stop engaging with their company, product or services. Customer churn has its significance in various industries such as e-commerce, telecom and internet service providers, insurance, subscription based services etc. In this paper we are implementing some of the top boosting algorithms (AdaBoosting, XG Boosting, Gradient Boosting, and LGBM) in their raw sate to compare and see out these advance machine learning algorithms which is the most suitable algorithm for the problems which have similar kind of environment. To ensure that one model performed better/worse than other models are going through accuracy test to project a clear picture of their performance. This comprehensive study of these boosting algorithms will fill us the information about boosting eco system and the implementation will help us to understand the raw state working of all these boosting algorithms.

II. Literature Review

A. Existing Work

Problem of customer churn has existed for a very long time in the Artificial Intelligence and Machine Learning community as it directly affect many online business and help them to plan and strategies their next steps moving forward with their business. Multiple studies and implementation have been done to address the problem and they used various regression and classification algorithms like logistic regression, linear classification, naïve Bayes, decision tree, support vector machine and etc [5] and some implementation where algorithm like random forest is

Grenze ID: 01.GIJET.10.2.244

combined with SMOTE-ENN to get above 90% on F1 score [1].
Few advance level studies and implementation have also used multiple layered perceptron neural networks to tackle the problem of customer churn. There are some studies which leverage the power of boosting algorithms as well. The dataset which is used for this research. Some studies are also cvovered effects of transforming and not transforming data to predict customer churn [3].

B. Boosting Algorithums over Traditional alorithums

Boosting algorithms are ensemble learning techniques which have ability to outperform the traditional classification and regression algorithms in various real world problems. Boosting algorithm also have the ability to combine various weak learns and create a highly accurate predictive model. The basic idea which boosting algorithm carry is that it learns from the failures/mistakes from the previous model and emphasizes more on misclassified instances. This gives boosting algorithm more adaptability which helps them to capture more complex patterns in a given dataset. The ability of capturing more complex pattern in a data gives boosting algorithms a strong edge over the traditional classification and regression algorithms [2].

C. Boosting Algorithums over Neural Networks

Neural Networks and boosting algorithms are both powerful machine learning techniques. Neural Networks also work great to predict customer churn [6] .One of the advantages of boosting algorithms is its interpretability which help us to understand the decision making process with more clarity. Boosting algorithm are volatile to over fitting when compare to the deep neural networks. Boosting algorithms can work in the situation where we need to deal with high dimensional data and due to its ensemble nature it handle diverse data effectively [2]. Due to its interpretability and less complexity boosting becomes a little bit better solution to the some problem like customer churn regardless of the power of the neural networks as it help us to keep our complexity less and more readable .

III. DATA AND PREPROCESSIN

The dataset which we are using for our comprehensive study of comparing various boosting algorithm is one the most popular dataset for the customer churn problem it is “Telecommunication Customer Churn Dataset” which is obtained directly from Kaggle. In this dataset each row represents a customer and in that each column contains various customer attributes.

A. Inside look of the dataset

There are various categories of individual and independent features in the dataset which each of them having their own significance and purpose.

- 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 such as for how long they have been a customer, contracts, payment method, monthly charge and total charge.
- Demographic info about customers – gender, age range, and if they have partners and dependents
- Customers who left within the last month – the column is called Churn

B. Preprocessing

Preprocessing or we can say data preprocessing is one of the most important step to make the data ready to before feeding it to the model or the algorithm and there are various techniques to perform the same. In the area of artificial intelligence and machine learning it involves cleaning, transforming and reorganizing the raw data into a suitable format which is best for our algorithms [3]. Number one step comes in the form handling the missing data in the entire data set column by column and for that there are various techniques to deal with it. We are considering using the mean, mode or median imputation method to deal with the missing data in every column as it will fill the missing value with the mean, mode or median of the observed data in a particular column.
Next important step in the preprocessing is to scale the variables to a range of 0 to 1 as it will equalize the variable scales and it will improve convergence for the gradient based methods. It will also promote the regularization in the dataset and it will also enhance the interpretability of the data. We will leave the preprocessing after applying these techniques as we do not need to augment the data so much that it will affect the model/algorithm performance. These preprocessing steps are allowing us to provide the data in the most suitable form to our algorithms for target objective.

IV. DATA VISUALIZATION

Before feeding our data to the algorithms we need to study the data and analysis it as it will help us to better under the data as well the decision making process. Visualization will also communicate the data information effectively and it will help us to spot any anomalies or outliners which may be still present in our data [4]. As this dataset have multiple features in it visualization of it will allow us to better understand the relationship between these various features.

A. Correlation of churn with other variables

As the final goal is to predict the customer churn, so having visual presentation of the customer churn with other variables in the data help us to understand all the necessary relationship which directly affect the our dependent variable (in our case that is customer churn) and we can observe that in figure 1.

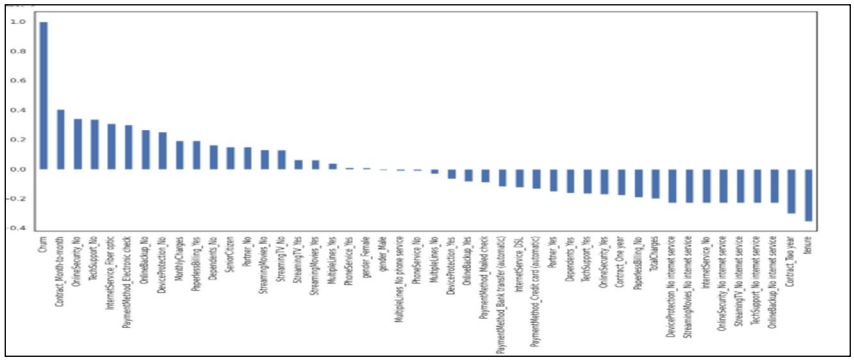

Figure 1: Customer Churn relation with parameters

B. Gender and Age Distribution in the dataset

Gender distribution in the dataset will provide us with the information about distribution of the customer in the two major categories (male and female for this particular dataset) and this will also provide us the clarity about the percentage of customer belonging to particular gender and will tell us the sex ratio towards consumption of the product.

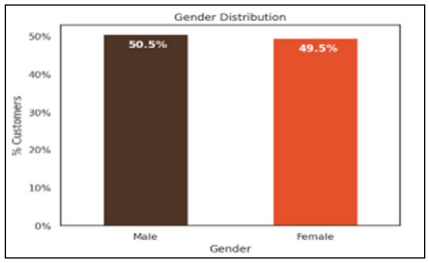

Figure 2: Gender distribution Dataset

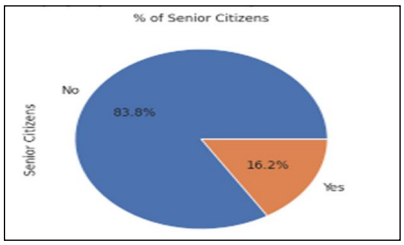

Figure 3: Senior Citizens and their churn

With the help of this visualization we can see that our data is pretty evenly distributed in the terms of the gender distribution figure 2. Similarly we can also visualize the distribution of age and one of the best ways to do that is to get the percentage of senior citizens in the total population. As figure 3 we can the senior citizen distribution with churn percentage.

C. Churn rate and Churn by Contract type

One of the most important element in our data is the actual churn and it is very important for us to know the ratios and proportion of how the actual churn is looking .

Looking into the visualization we have drawn a baseline towards the customer and churn percentage which will help us to draw some post processing conclusions and to further understand the vanilla churn we can dig deep into the churn by contract type (in our data we have three major types month-to-month, one year contract, two year contract). Having the visualization of the also help to draw conclusion regarding the best effective and more profitable contract type and that will help us to navigate the final conclusion in our research.

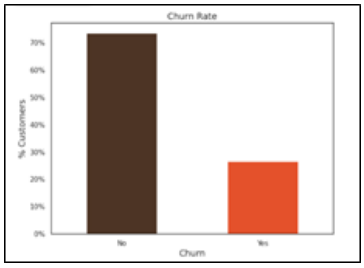

Figure 4: Gender distribution Dataset

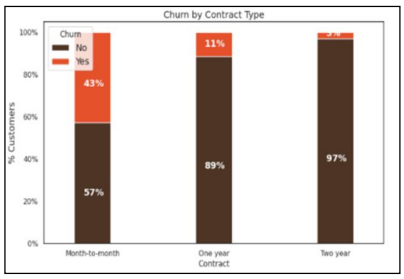

Figure 5: Churn by contract type

V. BOOSTING ALGORITHMS

Boosting algorithms are one of the powerful ensemble learning techniques which target to combine the output s of multiple weak learns and convert them into a strong learn. In boosting sequential training of multiple weak models happen and every subsequent model corrects the errors made by the previous model. At the end of the training final model is the combination of the earlier weak learns and every previous learn contributes to the final prediction using their individual strengths [2].Finally we can have a deeper look into the algorithms which we are going to use in our research and can get hold a better understanding of the tools which we are using.

A. Ada Boost

AdaBoost which is better known as Adaptive Boosting is one of the ensemble learning technique which have a purpose to improve weak classifier to make a strong classifier. Simplify the working of the AdaBoost Algorithm will show us that it basically train the weak learns and with that it assign weights to the each instance and it have a huge emphasis on the misclassified ones which improves the overall performance.

AdaBoost in the final model weak learns which have performed well are given higher weights and their prediction contributes more to overall prediction. The actual final prediction is made with the weighted sum of the weak learns predictions. One of the most interesting thing about this algorithm is that it is robust to the overfitting and that is because it have significant focus on the instances that are challenging to classify.To adjust and influence the performance of the

AdaBoost algorithm we have various parameter to play with such as the number of iteration, the choice of weak learns etc. Although it is a great algorithm it also have its limitation such it sometime struggles with noisy data and outliners and also sometime to become sensitive to the choice of the weak learners.There so many ways we can apply and implement the AdaBoost algorithm to perform the classification and regression problems and in various domains like computer vision, speech processing.

B. XG Boosting

XG Boosting which better known as Extreme Gradient Boosting is a scalable and efficient machine learning boosting algorithm. It was developed by Tianqi Chen and it is an extension of the gradient boosting framework and it has a special focus on overall model speed.

XG Boosting have some incredible key features like it have inbuilt feature which can handle missing values at the time of its training and the best part about it is that it can learn the best possibilities to handle the missing data. It is also equipped with an efficient cross validation system helping in minimizing the risk of overfitting.

XG Boost also have high scalability strength as it have ability to leverage parallel and distributed computing resources. It is one of the most flexible algorithm as it can be applied to the various machine learning problem for example prediction in various industries (health care, finance) etc. XG Boost is an incredible combination robustness, speed and scalability which make it one of the most reliable choice for high performing predictive models

C. Gradient Boosting Classifier

Gradient Boosting Classifier or simple gradient boosting can be described as the source material for the XG Boost, Light GBM, etc because they some most popular implementation of gradient boosting which improved performance and some particular adjustments .

Gradient Boosting is basically the root for various algorithm in this boosting categories which first introduced by Jerome Firedman in the year 1999 and after some time from its introduction to the world it become one of the most highly adopted techniques which results in various version of the boosting algorithms in the current time.

At its core Gradient Boost goes by the general definition of the boosting algorithm which it is use ensemble learning, use weak leans to make a strong learn, it a very robust algorithm and its ability to deal with the complex data of various domains. And to unlock it maximum capability we need adjust and fine tune its hype parameters just like any other machine learning algorithm.

D. Light GBM

Light GBM which also known as light gradient boosting machine is a open source gradient boosting framework. It was developed by Microsoft and it was created to address the limitation of the vanilla/traditional gradient boosting algorithm, It is one of the most preferred choice in various AI/ML compactions. Like any other boosting algorithm have its own sets of features.

As the name says LightGBM it is a lightweight framework but with a great efficiency. LightGBM also use Histogram-Based Learning and Gradient Based One Side Sampling (GOSS). Histogram based learning is a novel technique in which continuous features are discretized into bins and after that histograms are constructed which help us to find best possible split points. This method finally helps in to significantly speeds up the entire training process and also result in lowering the required memory.

A significant difference between the traditional gradient boosting and LightGBM is that it use leaf wise tree growth where as traditional boosting algorithm use depth wise growth and this leads to more balanced trees in LightBGM.

E. CAT Boost

CAT which is better known as categorical boosting, is a boosting algorithm which was developed by Yendex. In this version of boosting it try to address the challenges which comes with categorical features in machine learning.CAT Boost also allow user to define custom function (objective functions) which provide a flexible environment which help in solving diverse machine learning algorithm.

The interesting thing about the CAT Boost is that it use a new boosting technique which also known as the ordered boosting technique which consider the ordering of categorical variables during the training process.

VI. Implementing Algotithm

For the best possible results for the research we have implemented this boosting algorithm using python programming language along with library and frame works for: Scikit-Learn, Numpy, Pandas which help in initial data processing, plotting the data and adjustment and implementing the various algorithms. For our development environment we will be using Google Colab.

```
import numpy as np
import pandas as pd
import seaborn as sns # For creating plots
import matplotlib.ticker as mtick # For specifying the axes tick format
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

sns.set(style = 'white')
```

Figure 6: Importing Libraries

The major heavy lifting in implementing the algorithm (AdaBoost, XGBoost, Gradient Boosting Classifier) and processing is done by using Scikit-Learn and for the other algorithm (LightGBM , CAT Boost) we will be importing the necessary library as shown in figure 7.

```
[ ] import sklearn.metrics as metrics
    from sklearn.metrics import classification_report

    from sklearn.ensemble import AdaBoostClassifier
    from xgboost import XGBClassifier
    from sklearn.ensemble import GradientBoostingClassifier
    from lightgbm import LGBMClassifier

    import catboost
    from catboost import CatBoostClassifier
```

Figure 7: Importing Models

```
[ ]  accuracy_model1 = 0
     accuracy_model2 = 0
     accuracy_model3 = 0
     accuracy_model4 = 0
     accuracy_model5 = 0

[ ]  report_model1 = 0
     report_model2 = 0
     report_model3 = 0
     report_model4 = 0
     report_model5 = 0
```

Figure 8: Default Variables

To keep a record and to evaluate our models later we will initiate multiple variables as in Figure 8 with a value of zero which will contribute in storing the accuracy's and the evaluation reports for our model.

VII. Results And Analysis

The evaluation process is divided into two stages. In first stage we compare our five models on the test accuracy matrices and in the second stage we will apply K fold cross validation with 3 folds on the models and get cross validation scores. From those conclusions we will draw ROC curve on the top two performing models to just get little bit more incites on it.

```
Accuracy for Model 1 (AdaBoost)      : 0.8104265402843602
Accuracy for Model 2 (XG Boosting)   : 0.7848341232227488
Accuracy for Model 3 (Gradient boost): 0.7644549763033175
Accuracy for Model 4 (Light GBM)     : 0.7507109004739336
Accuracy for Model 5 (CAT Boost)     : 0.8075829383886256
```

Figure 9: Accuracy of Models

```
CrossValidation Score for Model 1 (AdaBoost)      : 0.8019055745164961
CrossValidation Score for Model 2 (XG Boosting)   : 0.7849829351535836
CrossValidation Score for Model 3 (Gradient boost): 0.7555460750853241
CrossValidation Score for Model 4 (Light GBM)     : 0.7558304891922639
CrossValidation Score for Model 5 (CAT Boost)     : 0.8033276450511946
```

Figure 10: Cross Validation Scores

From figure 9 we can observe the accuracy of the model. LightGBM gave an accuracy of 75% after that comes Gradient Boosting / Gradient Boosting Classifier 76.4% after that comes the XGBoost with an accuracy of 78.4%. After that comes CAT Boost with 80.7% and finally comes AdaBoost with and accuracy of 81%. From the observation of cross validation in figure 10 we can see that they are almost similar to the accuracy's by which we can assume that our models are consistently across different subset of our dataset. With these results we can tell our top performers for this research are Ada Boost and Cat Boost.

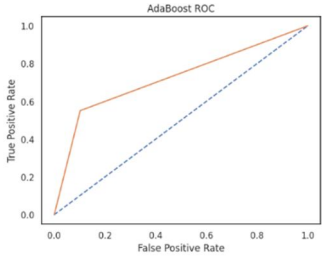

Figure 11: ROC of Ada Boost

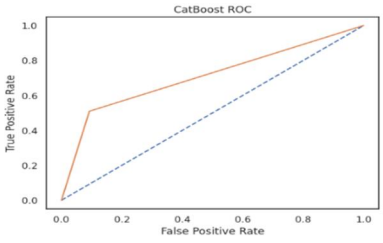

Figure 12: Roc of Cat Boost

We have also calculate the f1 score for both Ada and Cat Boost which came around ~ 0.60 which fairly decent in the context of our research. Looking at Figure 11 and 12 we can see that the performance of both Ada and Cat Boost and we can although Ada and Cat Boost work almost same but still by looks of the roc scores Ada Boost slightly outperformed the Cat Boost and Came at the top.

VIII. Conclusion

In this comparative study we study and worked with some top boosting algorithms. We also worked with one of the top problem which revolves around businesses in the name of customer churn. We created a common base for our models by which we can see the raw state working of these algorithms and we successful implemented these algorithm. The results are positive as all of our models delivered over 75% on scale of accuracy and our top performers gave results over 80% which is very decent. We backed our results with the help of cross validation and draw roc curves for top performers Ada Boost with 81% accuracy and 0.80 on cross validation and our second best Cat Boost have 80% accuracy with 0.80 on cross validation

References

[1] S. De, P. P and J. Paulose, "Effective ML Techniques to Predict Customer Churn," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 895-902, doi: 10.1109/ICIRCA51532.2021.9544785.

[2] N. Lu, H. Lin, J. Lu and G. Zhang, "A Customer Churn Prediction Model in Telecom Industry Using Boosting," in IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1659-1665, May 2014, doi: 10.1109/TII.2012.2224355.

[3] A. Amin, B. Shah, A. M. Khattak, T. Baker, H. u. Rahman Durani and S. Anwar, "Just-in-time Customer Churn Prediction: With and Without Data Transformation," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 1-6, doi: 10.1109/CEC.2018.8477954.
[4] Senthilnayaki, B. & M, Swetha & D, Nivedha. (2021). CUSTOMER CHURN PREDICTION. IARJSET. 8. 527-531. 10.17148/IARJSET.2021.8692.
[5] Rajendran, Srinivasan & Devarajan, Rajeswari & Elangovan, G.. (2023). Customer Churn Prediction Using Machine Learning Approaches. 1-6. 10.1109/ICECONF57129.2023.10083813.
[6] Khan, Yasser, et al. "Customers churn prediction using artificial neural networks (ANN) in telecom industry." International journal of advanced computer science and applications 10.9 (2019).