Customer Transaction Prediction

Rohitkumar Keswani
200361138
MSc Big Data Science with
Industrial Experience

Abstract - The primary objective of this study is to find the transactional behavior of the customers irrespective of the money transacted in the past. The traditional machine learning algorithms like Logistic Regression, Support Vector Machine, and Random Forest Classifier were outperformed by Ensemble techniques like Boosting. This will help the banks and financial institutions to provide personalized customer service and achieve customer satisfaction which will not only help the business to grow but also an improved more enhanced service experience to end-users.

Keywords—transaction, machine learning, classification, ensemble methods, machine learning explainability

I. Introduction

Banks and financial institutions have huge amounts of data that can be potentially used to build the business and marketing strategies by gaining insights from the data and attract more customers and provide the customers with personalised service. However, this data cannot be used as is as it contains lots of personal data and so it is used by anonymising the features which contain personal data and so it makes it quite challenging to interpret the data.

The project is Customer Transaction Prediction model. In this project the aim is to build a model that which customers will make the transaction in the future irrespective of the transactions in the past. The data has 200 anonymized features and a binary target column consisting of values 0 and 1. Zero represents that the transaction will not happen and one represents that transaction will happen. Apart from the anonymized features the data or the values of these features are manipulated or some normalization or preprocessing of these values have already been done which hinders our understanding of the data which is the primary step while approaching any problem in data science. Moreover, we cannot understand the statistical properties of the data as it is not known to us which techniques have been applied to manipulate the data.

This project can help the business to grow in different ways. Determining customer lifetime value (LTV) of every customer and planning the business model and decisions on a long-term goal rather than short-term revenues. By getting this insight we can take necessary marketing actions on each customer, helping the business to achieve the goal. Moreover, segmenting the customers into mini groups and customers on the transactional behavior can help businesses significantly.

Contributions

In this study various data preprocessing techniques and hyper parameter tuning of traditional machine learning algorithms are applied to tackle the imbalance nature of the dataset however, the machine learning models were overfitting. To overcome this imbalance nature of the data, data augmentation (SMOTE) and boosting methods gave the best results. To use this model for predictions a flask app is deployed on Heroku which can be used by banks and financial institutions to predict the customers who will make the transaction in the future and can build their financial and marketing strategies according to it.

II. Related Work

After the use of Data Science in every field, Banks are now weighing more importance in utilizing and collating huge internal data for instance credit and debit transactions. Moreover with Digitalization the use of internet banking has increased significantly and as a result, increased the data flow in the financial institutions and this can be leveraged or used in providing the customers a more enhanced personalized product [6]. The data therefore, comes in various formats and it becomes a challenge to find meaningful insights from it and help the business, to overcome this problem many methods are used for example, hypothesis testing, time series analysis and visualization, machine learning, natural language processing. Personalised customer service, fraud detection, and profiling can be done with data science and machine learning with the amount of Data Banks and other Financial Institutions possess.

It has been researched and surveyed that Customers do not want different experiences across various banking products for example Auto Loans, Credit Cards. Customers tend to appreciate the Personalised-as-a-service experience compare to be treated like a number instead [3]. Also, the banks or the financial institutions that fail to satisfy these standards are in jeopardy as customers are now reevaluating their choices and considering their choice of financial providers given various non-traditional range of alternatives.

Customer Churn plays a major role in losing the customers. Customer churn is the loss of clients or customers for a certain period. Customer Relationship Management (CRM) plays a vital role here to minimize customer churn because it plays the major concern where a business makes a loss [2]. There has been a lot of studies where using Data Mining techniques and Machine learning techniques to predict the churn of the customers by modeling customer behaviors. Moreover with the amount of transactional data with the banks and financial institutions a new aspect of the CRM came which was customer relationship marketing where new services can be sell using data from the transactional databases and collecting new data from the customers which will give a probability of how new services will work with the customers and their competitors.

Profiling of the customers is the second important aspect in the business where customer transactional data is used [1]. Histogram techniques are used to generalize the customers from sparse transactional data but are not efficient enough to bin the complexity of the data to improve those probabilistic mixture models are used which automatically bins the customers and generate profiles from very large customer transactional data.

III. Methodology

The dataset for this problem can be downloaded online. The training dataset has 200,000 rows with 200 anonymized features and a binary target column 0 or 1. Zero representing that future transactions will not occur. The test dataset is also available with 200,000 features but it does not have a target column.

After loading the training set and starting the preprocessing of the dataset, we can see that the values of the 200 features are not the original values, there have already been done some preprocessing on it, for instance, Standard Scaler or Min-Max Scaler or Normalisation of the values of these features have been done. This has been done because this data is similar to the real dataset used in banks or financial institutions, due to the privacy issues the data has already been preprocessed and for the same reason, the features column names are also anonymized which makes the problem more challenging to solve as one cannot interpret which feature columns are more relevant or can be discarded for building machine learning models for prediction.

Machine Learning Explainability

The training data has 200 anonymized features which makes it very difficult to interpret the data. As the features were anonymized it was very important to get the idea of the features as to which features had the biggest impact on the predictions. This can be done by Permutation Importance [16]. It is only calculated after a model is fitted. After the model is fitted a single feature column of the dataset is shuffled leaving all other columns unchanged and then checking the effect in the accuracy of the predictions in the data. Reordering a simple column should not cause any significant difference in the predictions unless the feature column is more important compare to other columns. So, a basic Logistic Regression model is fitted to interpret the feature importance and the result achieved was var_139 was with the highest weight of 0.0015+-0.0002 and the second was var_81 with the weight of 0.0013+-0.0004. var_68 has 0 weight which means that it does not contribute towards the prediction, whereas features like var_101 and var_199 contribute inversely to the prediction with weights of -0.0001+- 0.0. It can be concluded that if we change the machine learning model the importance of features will change but it will not change significantly. This methodology is very useful in this scenario as we have 200 features which are anonymized and can be helpful to us to analyze the data and can help business to understand which features play key roles in the future transactions of the customers.

Feature Importance gives us insight into what features most affect the predictions, partial dependence plots on the other hand shows how predictions are affected by features. This is useful to answer questions like the predicted transaction difference between two customers is due to which feature or which any other factors contribute towards to it. This arises the question that can a single change in the feature value of two customers can change the prediction or how much probability increases or decreases for a customer transaction to occur in the future. Partial dependence plots can be interpreted similarly to the coefficients of linear or logistic models, however on sophisticated models, partial dependence plots can capture more complex information about the features and predictions [17]. After a model has been fitted Partial dependence plots are calculated, similar to permutation importance however, the data must not be manipulated artificially in any way for the partial dependence plots. Now to learn how a partial dependence plot separates the effect of each feature we start by taking a single row of data, in our data set we took var_81. After fitting the model we will repeatedly alter the value of one variable and will make series of predictions, for this dataset I chose var_81 and will change its value and will predict the outcomes for these changing feature values. The predicted outcomes are traced on the vertical axis as we move our values from small to large on the horizontal axis.

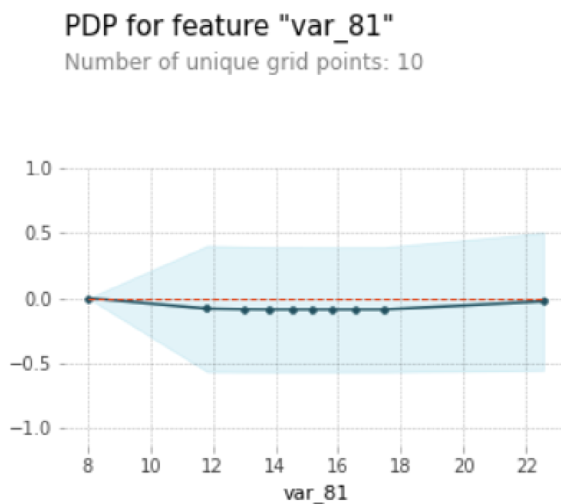


Figure. 1a. Partial Dependence Plot showing the change of prediction with the increase of the value of var_81

Few things that can be interpreted from the above graph is the y-axis is interpreted as the change in the prediction from what the prediction would be at the leftmost value. The blue shaded area indicates a level of confidence. We can see from the above graph that as the value of var_81 increases the probability of prediction changes from 0 to1. These techniques are used to extract general insights from the machine learning model. Now to learn how the prediction works for an individual prediction SHAP values (Shapely Additive exPlanations) are used which shows the impact of each feature[18]. This can be used in many business areas, especially in the finance and banking fields. For instance, a model predicts that a bank should not loan money to the customer and the bank is required to explain the basis for each loan rejection. Moreover, this can also be used in the Health sector where it can be identified as which factors are contributing to each patient's risk of disease, and can directly focus on those risk factors.

SHAP value interprets the impact of having a certain value of a given feature in comparison to the prediction we'd make if that feature took some baseline value. SHAP decomposes a prediction and gives a nice property. The decomposition of the prediction takes place with the following equation: sum (SHAP values for all features) = prediction – prediction for baseline values.

This means the SHAP values of all the features sum up to explain why the prediction was different from the baseline. The below SHAP summary plot gives us the overall view of important features and what is the reason behind it [19].

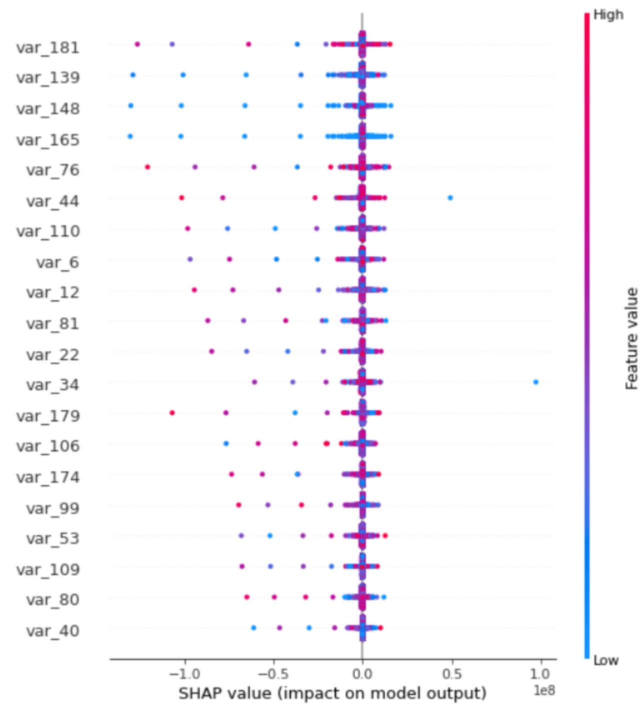


Figure. 1b. SHAP Summary Plot giving us the overall view of the feature importance.

After the analysis of these features, we can get an insight as to which feature has more weights or which feature is contributing more towards the prediction in a positive or in a negative way. This will help the business to grow more and can take necessary actions or plan in the right direction towards personalized marketing or personalized service to the customers which will help them to retain their customers and widen their scope of business with these insights.

Data Preprocessing

Now getting the insight of the data and preprocessing the data. There are no missing values in the dataset for 200,000 rows for 200,000 features. The next step will be analyzing the target column and checking the frequency of the occurrence of the binary values.

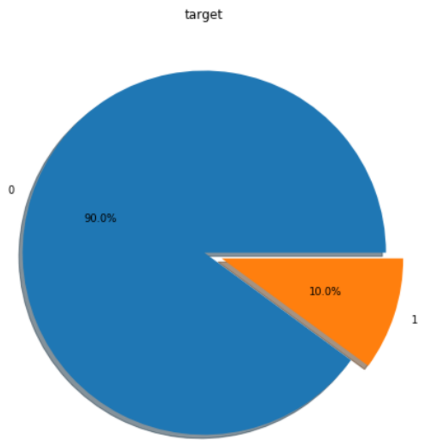


Figure. 2. Target Value Pie chart showing us the imbalance nature of the dataset

It can be concluded that 90% of the target values are 0 and only 10% of the binary values are 1. 179,902 values out of 200,000 are 0 and the remaining 20098 are 1 which makes this dataset highly imbalanced. Additionally, doing feature engineering and trying to gain more insight into the data. Checking the skewness and Kurtosis of the target column gives 2.657 and 5.063 respectively. Analysing the training data by checking Mean, Median, Standard Deviation, Min, Max, Skew, and Kurtosis.

Further, trying to find any correlations between the 200 features that can help to identify any insights from the anonymized features, no linear correlation was found neither in training data nor in test data. There is a possibility as the data is anonymised so perhaps they are also decorrelated by some transformations.

Now performing Exploratory Data Analysis and plotting some visualization to know the nature of data over 200 features and interpreting any insights from them. Plotting Mean Distribution of rows of training data. Mean is the average of the data, it is used to describe the sample of the data with a single value which is the centre of data, But one limitation of mean distributions is that they are not robust to outliers. It can be visualised from the below mean distribution plot that the mean of the training data is 6.75.

Figure. 3a. Mean distribution of rows in training data

Plotting mean Distribution of columns in training data gives us the following plot. The following plot is more scattered as compared to the mean distribution of the row values. The mean distribution of rows followed a normal distribution but the distribution of columns do not follow a normal distribution.

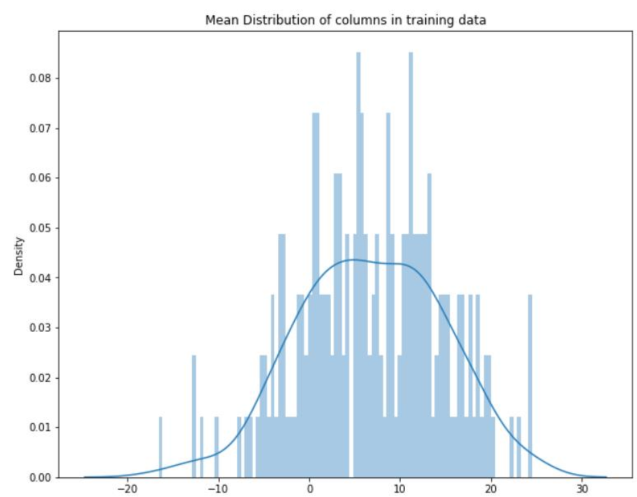


Figure. 3b. Mean distribution of columns in training data

The next distribution plotted is the Standard Deviation distribution per row. It can be visualised that the standard deviation per row follows a normal distribution.

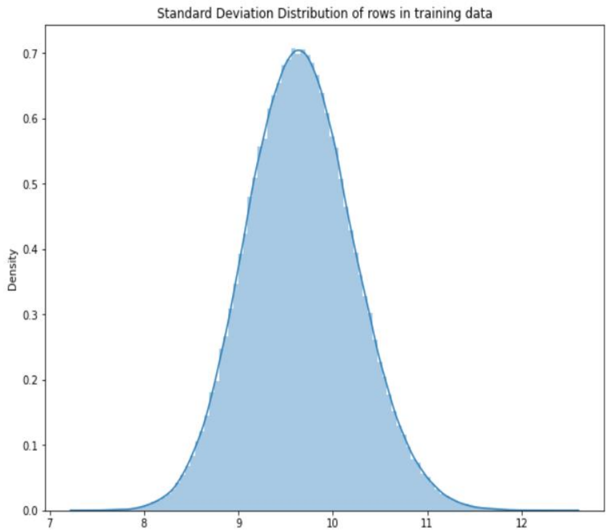


Figure. 4a. Standard Deviation of rows giving us insight into the amount of spread with respect to mean.

A pattern can be seen here that the row distributions follow a normal distribution and the values of the column distribution are more scattered and do not follow any distribution. The below distribution has a high standard deviation as it can be visualised from the graph.

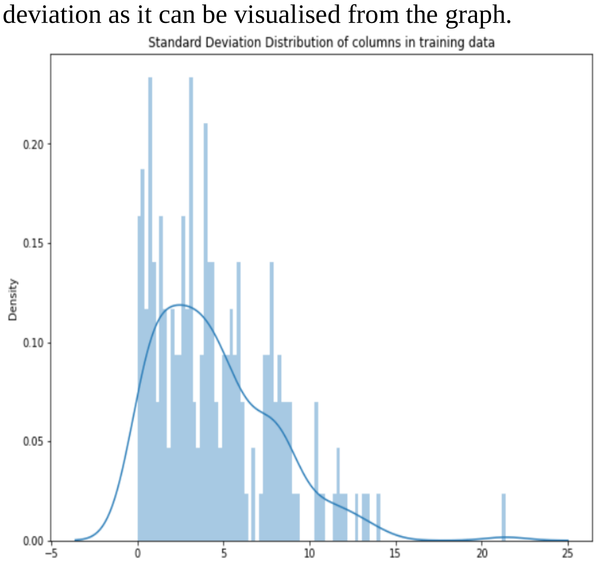


Figure. 4b. Standard Deviation of columns showing us high spread with respect to mean.

Standard Deviation is one of the most used measures for calculating the dispersion of the data from the mean. If the value of standard deviation is higher it means that the data is more spread and a lower standard deviation means that the data is closer to the mean value or closer to the center of the data.

IV. Results

After gaining insights through data visualization the next steps was the classification by building machine learning models and understanding the importance of preprocessing for different approaches using machine learning algorithms. Starting to build the machine learning model I dropped the target and ID_code from the training data and created a Y variable that has the target column of the dataset. Now, splitting the dataset in train test split where 80% of the data is used for training and the rest 20% for validation. Building a normal logistic regression model and 91% accuracy for the validation set and training dataset and 99% recall for zero targets for both training and validation set and only 19% recall for target 1 was achieved. But as the data is imbalanced accuracy score is not an appropriate evaluation metric. In this scenario, evaluation metric will be using the Area Under the Curve (AUC). The Area Under the Curve is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher the AUC, the better the performance of the model in distinguishing between the two binary targets, in this scenario Zero and One class. For logistic regression, we are getting a 0.58 roc auc score for both training and validation set which is very poor because 0.50 value is the worst a model can predict.

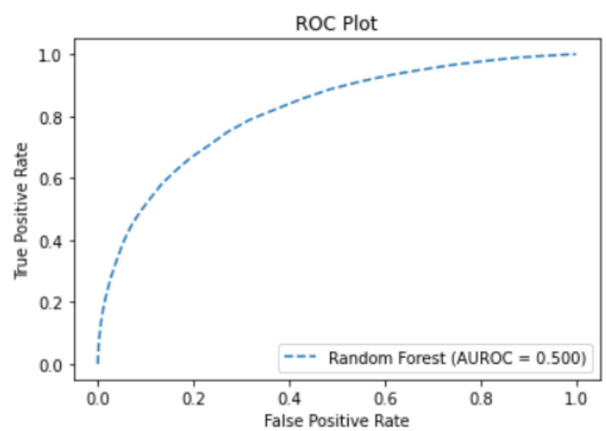


Figure. 5a. Roc curve of Random Forest with 0.50 roc score which is the worst-case scenario

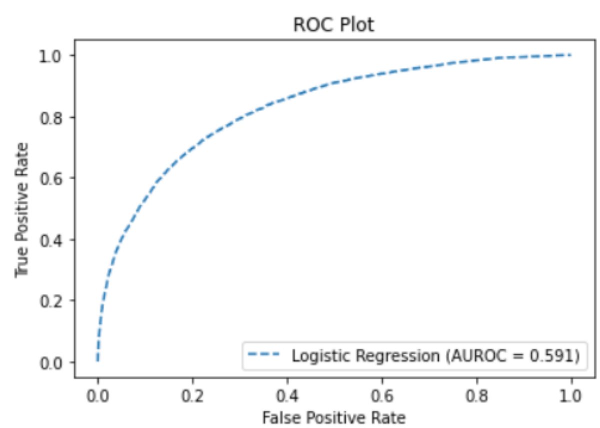


Figure. 5b. Roc curve of Logistic Regression with 0.58 score.

The next machine learning model was the Random Forest classifier which gave even less score for all the evaluation metric, achieving 1 recall for 0 target columns and 0 recall for 1 target column. Moreover, I achieved a 0.50 roc auc score which means that the classifier is not able to classify the features, it means that it just predicts 0 for all the validation set predictions.

It can be observed that directly applying the machine learning algorithms are not performing to the expectations of correctly classifying the two targets. To improve the results sampling technique can be used. Experimenting with under-sampling the data, using the imblearn library and random under sampler function. Fitting and under-sampling the data. After under-sampling getting 20,098 number 0 zero rows and the same number of 1 rows. Now, after under-sampling fitting the Random Forest Classifier after splitting the dataset in train test split. The result achieved was improved significantly compared to the previous experiment, achieving 0.74 recall for 0 and 0.75 for 1 which increased a lot for target 1 but decreased for feature 0. The roc auc score was 0.748 which improved significantly from the data without under-sampling.

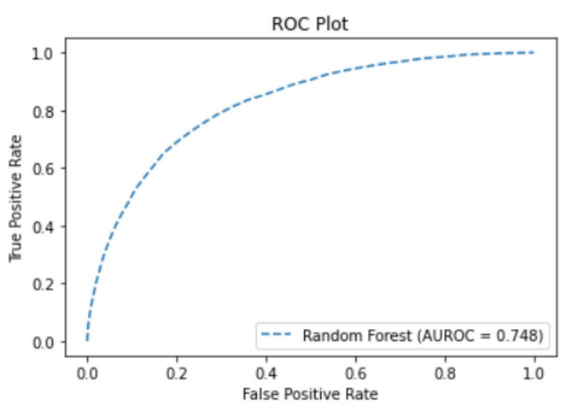


Figure. 5c. Roc curve of Random Forest modelled on the under sampled data.

Then experimenting with logistic regression and there was a slight improvement in the score with 0.77 and 0.78 for 0 and for 1 target respectively. Moreover, the accuracy achieved was 77% which decreased compared to the previous experiments on the whole imbalanced dataset, achieving roc auc score of 0.772.

Moreover, experimenting with the Support Vector machine it gave similar results compare to Logistic Regression but with Artificial Neural Network, the score dropped significantly. The recall dropped to 0.70 for both the target values and achieving the roc auc score of 0.697.

Models	Roc Auc Score
Random Forest	0.737
Logistic Regression	0.772
Support Vector Machine	0.764
ANN	0.707

Table. 5d. Roc Auc Score of Models in Undersampled data.

The limitation of Under Sampling is that the examples from the majority class are removed which may be useful and important to fit a robust decision boundary. In under-sampling given that the data is removed randomly, there is no way to detect which examples are rich in information that can be preserved from the majority class of an imbalanced dataset. To overcome this limitation I tried oversampling the data. In oversampling, contrary to under-sampling the data of the minority class is increased to the number of majority class, thus increasing the dataset. For this dataset, the shape of the dataset increased to (359804, 200) making the equal number of 0's and 1's in the training data equal to 179902 each of the target. Again splitting the dataset in train test split and fitting the logistic regression model achieving 0.77 recall for both the target values and achieving 0.772 roc auc score which is slightly better than under-sampling and also without the limitations of the under-sampling. I achieved a 100% score for all the evaluation metrics and 1 Auc Roc curve which clearly means that there is something wrong with the oversampling process because a model cannot be 100% accurate for whole validation data, so again have to adapt to the alternate process.

Following these methods, experimenting with Data Augmentation. Instead of duplicating the data from the minority class, synthesizing new data from the minority class gives better results. The Most widely adopted technique for synthesizing new samples is the Synthetic Minority Oversampling Technique (SMOTE). SMOTE randomly selects a minority class and finds it k nearest minority class. Then a random neighbor is chosen and new synthetic data is created at randomly selected points among these two points in the feature space.

After transforming the data set using SMOTE the new shape of the data was 359804, thus solving the problem of the imbalance dataset and splitting this dataset and taking 25% for the test set, achieving very high results to compare to the previous approaches. 1.00 Auc score for the training data set and 0.96 Auc score for the test set.

Next, trying to transform the data into normal distribution and then tried to apply machine learning algorithms. Applying Quantile Transformers, this is a robust preprocessing scheme as this will spread out the most frequent values, moreover will also reduce the effect of outliers significantly. Building a pipeline using the sklearn library and in that pipeline, combining the quantile transformer and Gaussian Naïve Bayes Classifier achieving very high results on the training as well as on the validation set. For the training set achieving an auc score and 0.888 for the validation which is almost the same.

It can be observed that after transforming our data into normal distributions there is a significant increase in the auc roc score compare to previous models which were trained on the original data set and also even though after under-sampling and oversampling.

Moreover achieving higher results with boosting techniques such as Cat Boost. Cat Boost is a gradient boosting technique on decision trees. Trying Cat Boost on the original dataset achieved higher results compared to previous algorithms, 0.72 auc score and 0.45 recall for 1 which is the minority class in training set and 0.679 auc score and 0.37 score for 1 in the test set. This results increased drastically when under sampling the data and applying the Cat Boost. On the training set achieving 0.842 auc score and 0.85 and 0.83 recall for 0 and 1 respectively and 0.80 on the test set achieving 0.82 and 0.80 recall for 0 and 1. The score further improved if Synthetic Minority Oversampling Technique (SMOTE), achieving training auc of 0.88 with 0.91 and 0.86 recall for 0 and 1 class, and 0.877 Test auc and 0.90 and 0.85 recall.

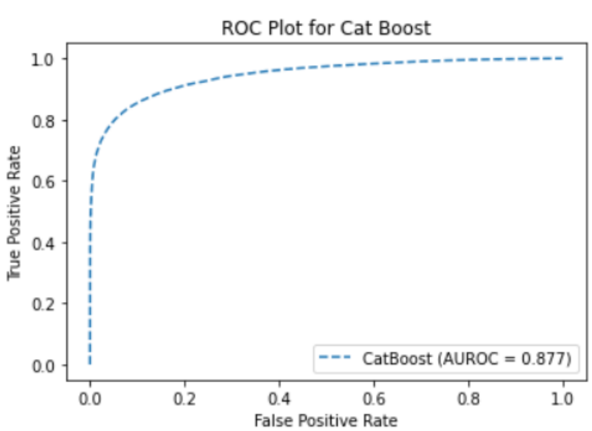


Figure. 5e. Roc curve of Cat Boost after Data Augmentation using SMOTE technique

Light GBM gave much higher results than Cat Boost algorithm after Data augmentation. It gave the score of 0.915. But Light GBM was also overfitting on the standard imbalance dataset, even after performing under sampling the model was over fitting, the problem solved after data augmentation using SMOTE technique it gave me 0.94 training roc score and 0.91 validation score.

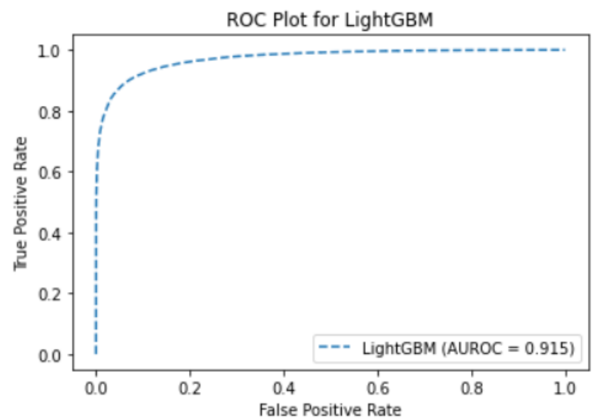


Figure. 5f. Roc curve of Light GBM after Data Augmentation using SMOTE technique

Ensemble Methods	**Auc Roc Score**
XGBoost	0.833
CatBoost	0.877
LightGBM	0.915

Table. 6. Auc Roc score of Ensemble methods on Augmented Data

Deployment

After building machine learning models, I build a flask app. The flask app is a simple application where a user can upload the data and can submit the file and will get the prediction in return whether the customer will make a specific transaction or not. For the prediction, using the pickle library and made a pickle file of the machine learning model. After making a pickle file, building a simple HTML page where a user can upload the data in CSV file format and building an API where after submitting the file the CSV file will be read and the pickle file will be used which will predict whether the customer will make a transaction or not in the future. After building the app It is deployed on Heroku cloud and It can be used by anyone around the globe and can get information on whether the customer will make a transaction or not.

V. Conclusion

The study has shown that the Imbalance dataset cannot be treated like a normal dataset and directly applying the machine learning algorithms will not give the required results. So to tackle this highly imbalanced dataset a different evaluation metric Roc Auc curve is required rather than Accuracy of the model because 90% of the data has 0 target value and the model will predict 0 most of the time and hence the accuracy will remain high.

As the dataset was anonymised it became more challenging to find insights from the data and to do that Machine Learning Explainability techniques like Permutation Importance and SHAP values are used. Permutation Importance gives us the features which contributed the most in the prediction or are the most important features and the features down in the list are the ones which did not contribute or inversely contributed means because of that feature there was some negative effect in the prediction which means these are least important features.

The Second Technique was the Partial Dependence plots that show how a feature affects the prediction, these plots are very useful as they can capture complex patterns and provide an insight into the data which will be very useful to tackle business problems. Moreover, the values of these 200 features were also manipulated that make it very difficult to pre pre-process the data because it was already preprocessed and preprocessing the pre-processed data will not make the model help to learn, so I performed a Standard Scaler to Scale the dataset by removing the mean with the value of variance 1.

After gaining insights from the dataset and performing feature engineering and standard scaling, building machine learning models was the next step. Adapting the standard Machine Learning approach was not enough as it gave very low results. These models gave high accuracy but the Roc curve score was 0.50 which is the worst-case scenario and thus different approach was required. In Logistic Regression using hyper parameter tuning using Grid Search CV but the result did not improve due to the highly imbalance nature of the dataset. To overcome this a slightly different approach can be used while tuning the parameters which is to use a dictionary of weights for the class_weights parameter and this improved the result significantly.

To further improve the score under-sampling technique is used to reduce the samples of the majority class, the results further improved, but it has one limitation that in under-sampling we cannot decide which samples will be deleted and because of that there is a high chance that important or sample with rich information will be deleted. To overcome this problem oversampling technique can be used where the minority class samples are increased equal to the majority sample class and the results were improved using this approach. But oversampling also has limitations in that it makes the same samples that are in the dataset thus it increases the chances of overfitting.

To overcome this problem Data Augmentation technique is used. In Data Augmentation I used Synthetic Minority Oversampling Technique (SMOTE). SMOTE overcomes the limitations of both under-sampling and oversampling and reduces the chance of overfitting. After these Feature Engineering processes building standard machine learning models did not give high results. whereas ensemble techniques, Boosting in particular gave high results. CatBoost and XGBoost were the two techniques that gave higher results than other standard algorithms moreover, lightgbm gave more improved results after tuning the hyper parameters.

VI. Future Scope

Building a Customer Transaction model will help businesses to grow by predicting which customers will transact in the future. Moreover with the amount of flow of data in the banks have increased the scope of using the data not only for prediction if the customer will make a future transaction but also in many different areas of businesses.

For instance, a Customer product recommendation system can be built which will consider past behaviour of the customers moreover, the data and the behavior of similar customers can be taken into account for modeling the machine learning algorithms. Moreover, a Customer Satisfaction model can be built with the data which will help us to classify which customers are satisfied with the banking experience. It will help the institutions to identify the customers very early if they are not having a good experience and thus can save from losing the customers and any more potential customers. A Pipeline can be created which will be customer-centric which can Predict the future transactions of the customers, which can recommend products to the potential customers looking at the past behaviors and behavior of similar customers, and lastly predicting the satisfaction of the customers which will help the institutions to improve their relationship with the customers and will help them grow.

References:

[1] Kamakura, W.A., Wedel, M., De Rosa, F. and Mazzon, J.A., 2003. Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. International Journal of Research in marketing, *20*(1), pp.45-65.

[2] Prasad, U.D. and Madhavi, S., 2012. Prediction of churn behavior of bank customers using data mining tools. Business Intelligence Journal, *5*(1), pp.96-101.

[3] Cadez, I.V., Smyth, P. and Mannila, H., 2001, August. Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 37-46).

[4] Jaiswal, D.P., Kumar, S. and Mukherjee, P., 2020. Customer transaction prediction system. Procedia Computer Science, 168, pp.49-56.

[5] Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C. and Haltmeier, M., 2020. A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, *281*(3), pp.588-596.

[6] 2021. [online] Available at: <https://financetrain.com/applications-data-science-banking/> [Accessed 5 August 2021].

[7] Kaggle.com. 2021. *Learn Python, Data Viz, Pandas* & More | Tutorials | Kaggle. [online] Available at: <https://www.kaggle.com/learn/machine-learning-explainability.> [Accessed 5 August 2021].

[8] Medium. 2021. Santander Customer Transaction Prediction *:* A Simple Machine Learning Solution. [online] Available at: <https://medium.com/@aganirbanghosh007/santander-customer-transaction-prediction-a-simple-machine-learning-solution-771613633843> [Accessed 5 August 2021].

[9] Mutanen, T., Nousiainen, S. and Ahola, J., 2010. Customer churn prediction–a case study in retail banking. In Data Mining for Business Applications (pp. 77-83). IOS Press.

[10] Benoit, D.F. and Van den Poel, D., 2009. Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. Expert Systems with Applications, *36*(7), pp.10475-10484.

[11] Ahmad, A.K., Jafar, A. and Aljoumaa, K., 2019. Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, *6*(1), pp.1-24.

[12] Li, J., Zhang, L., Meng, F. and Li, F., 2014. Recommendation algorithm based on link prediction and domain knowledge in retail transactions. Procedia Computer Science, *31*, pp.875-881.

[13] Tsai, C.F. and Chen, M.Y., 2010. Variable selection by association rules for customer churn prediction of multimedia on demand. Expert Systems with Applications, *37*(3), pp.2006-2015.

[14] Ilyas, S., Zia, S., Butt, U.M., Letchmunan, S. and un Nisa, Z., 2020. Predicting the future transaction from large and imbalanced banking dataset. *Int. J. Adv. Comput. Sci. Appl.*, *11*(1), pp.273-286.

[15] Kaggle.com. 2021. Use Cases for Model Insights. [online] Available at: <https://www.kaggle.com/dansbecker/use-cases-for-model-insights> [Accessed 18 August 2021].

[16] Kaggle.com. 2021. Permutation Importance. [online] Available at: <https://www.kaggle.com/dansbecker/permutation-importance> [Accessed 18 August 2021].

[17] Kaggle.com. 2021. Partial Plots. [online] Available at: <https://www.kaggle.com/dansbecker/partial-plots> [Accessed 18 August 2021].

[18] Kaggle.com. 2021. SHAP Values. [online] Available at: <https://www.kaggle.com/dansbecker/shap-values> [Accessed 18 August 2021].

[19] Kaggle.com. 2021. Advanced Uses of SHAP Values. [online] Available at: <https://www.kaggle.com/dansbecker/advanced-uses-of-shap-values> [Accessed 18 August 2021].

Appendix:

A. Plots

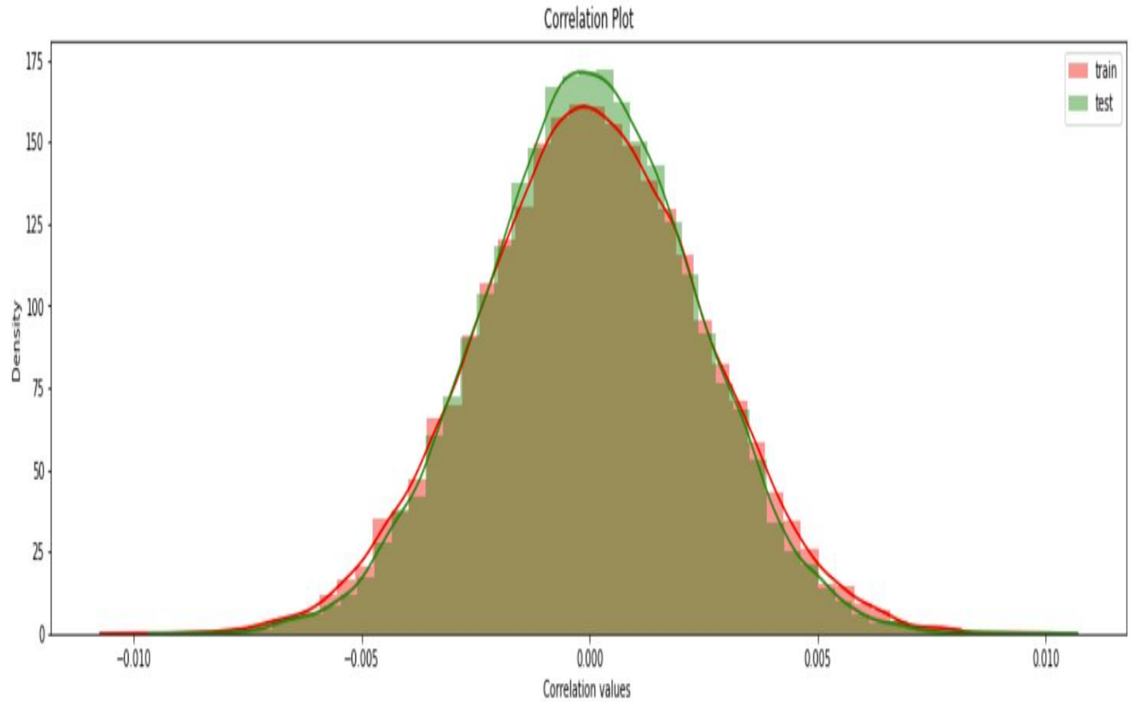


Fig. A 1. Correlation Plot of training and test data showing us there is no correlation in training or in the testing dataset.

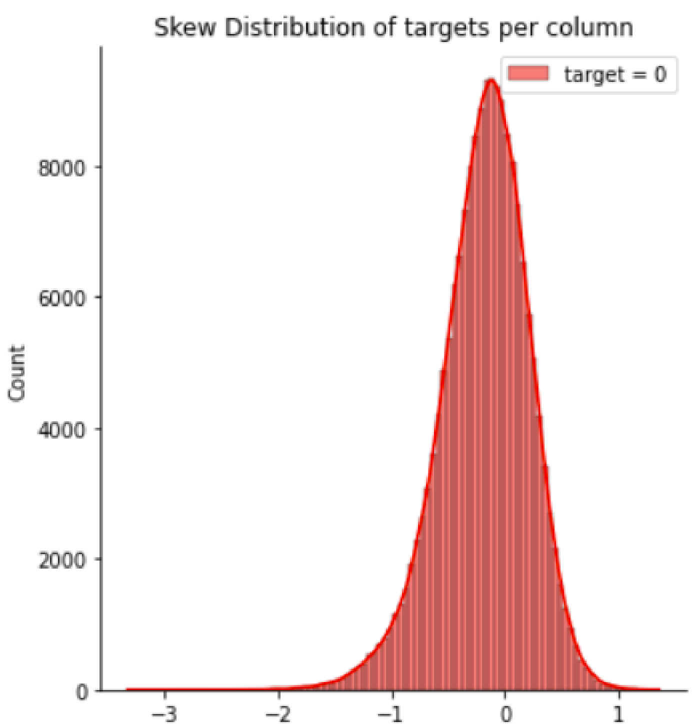


Fig. A 2. Skew Distribution of 0 target per column

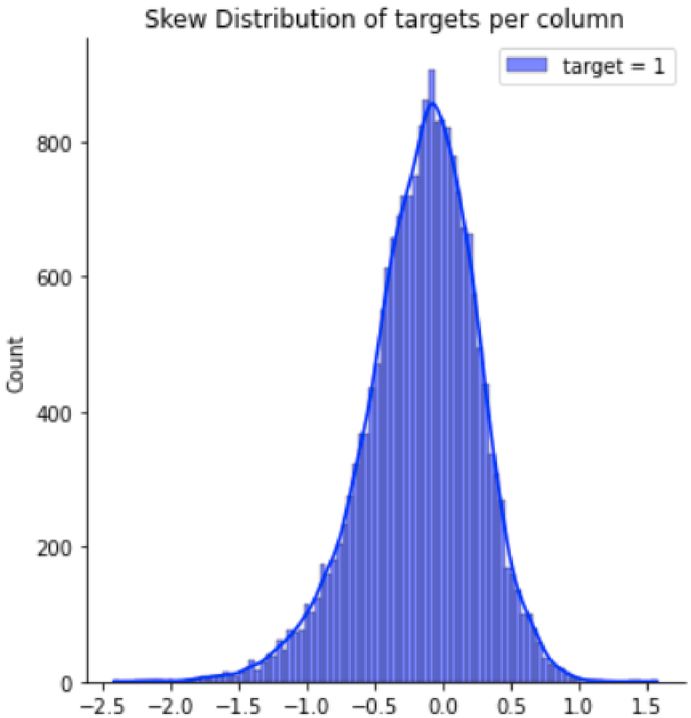
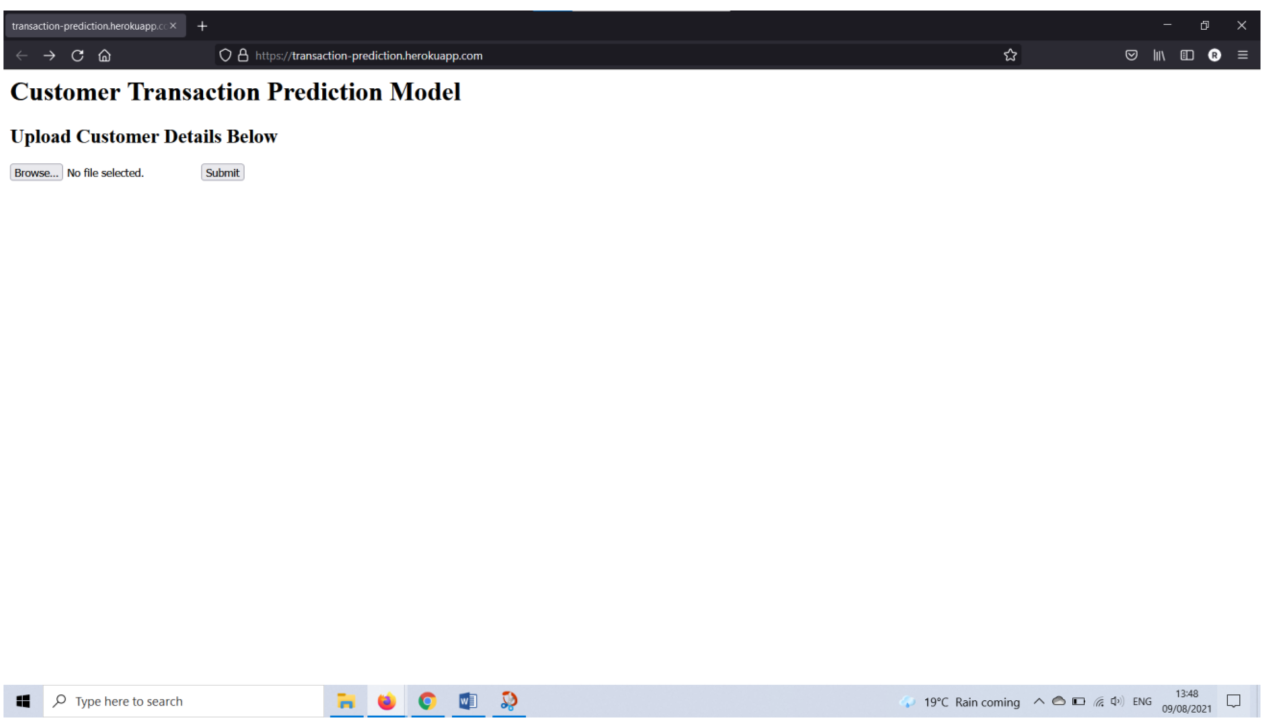
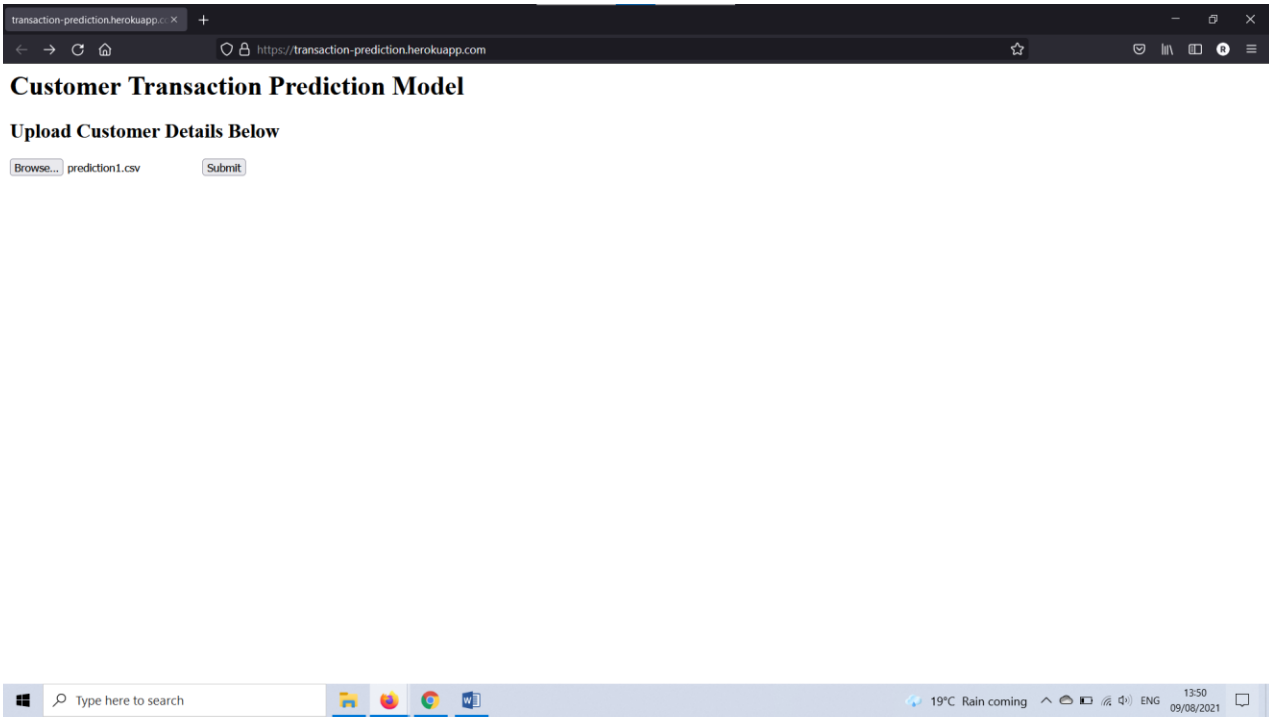
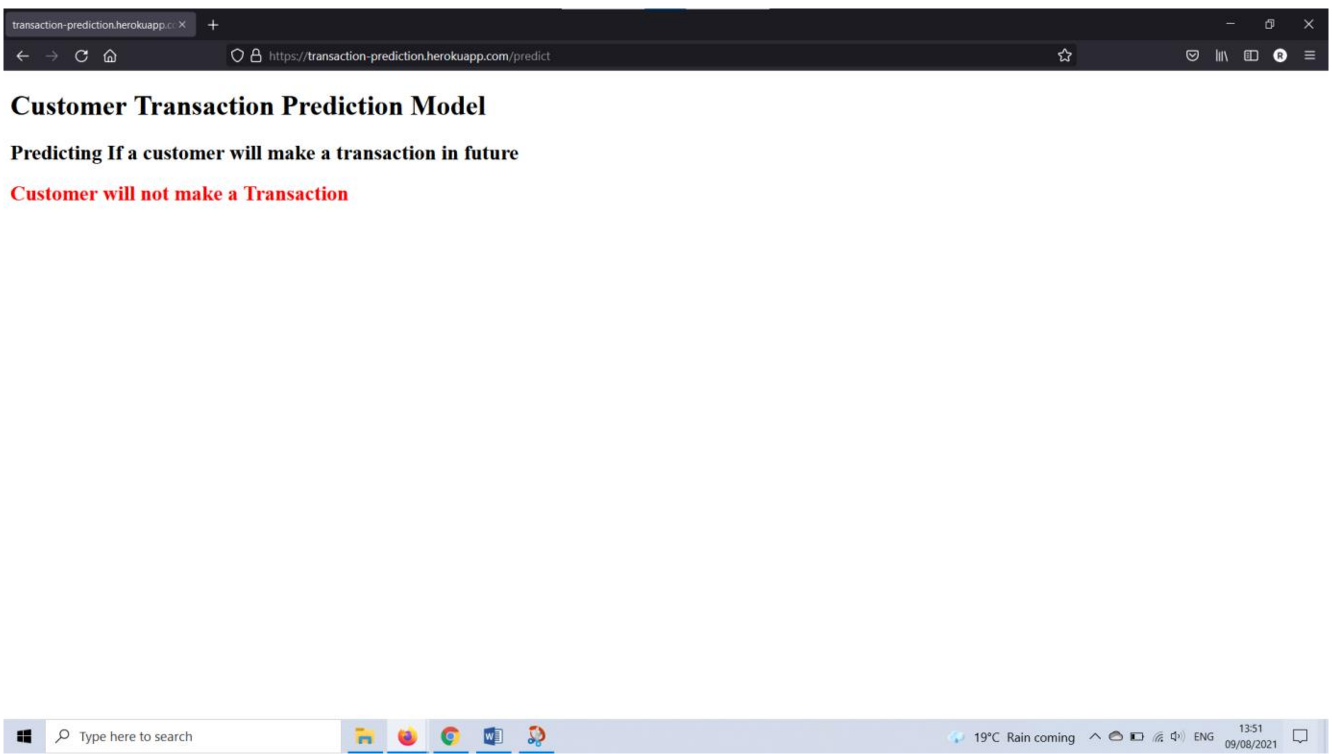


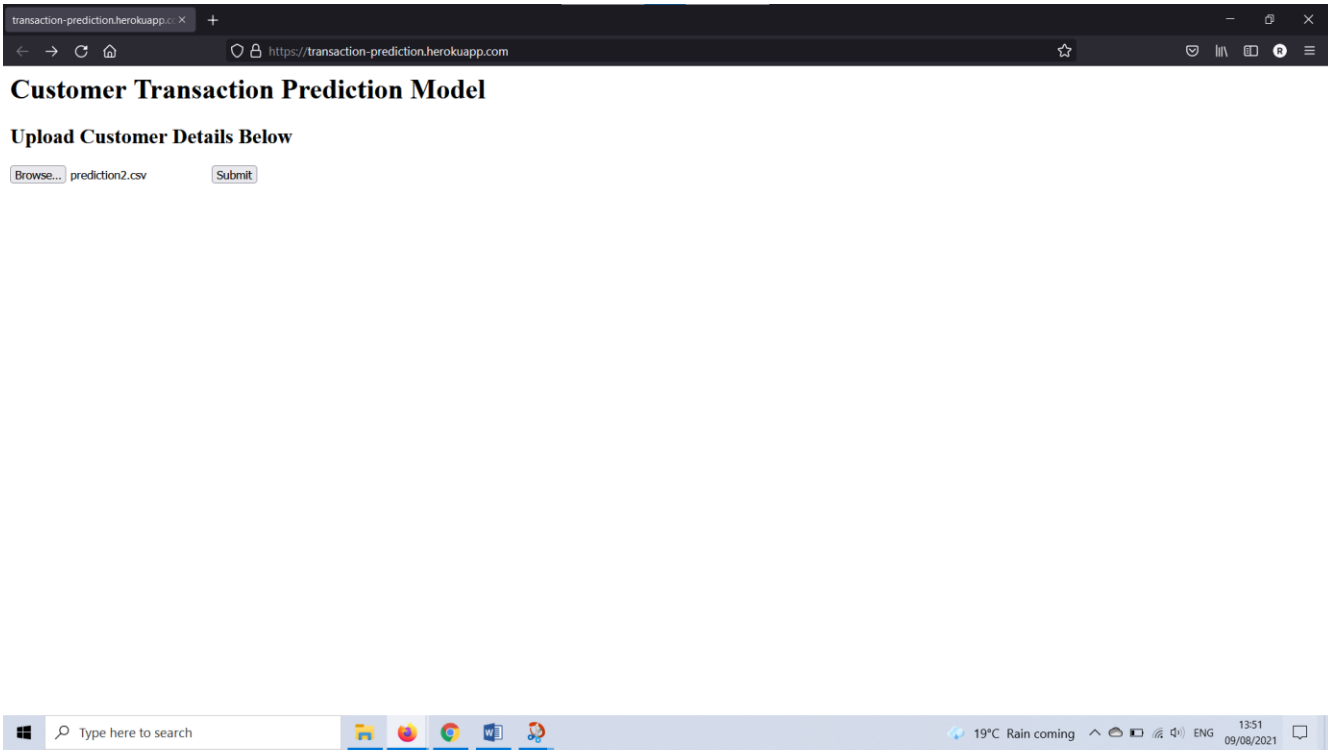
Fig. A 3. Skew distribution of 1 target per column

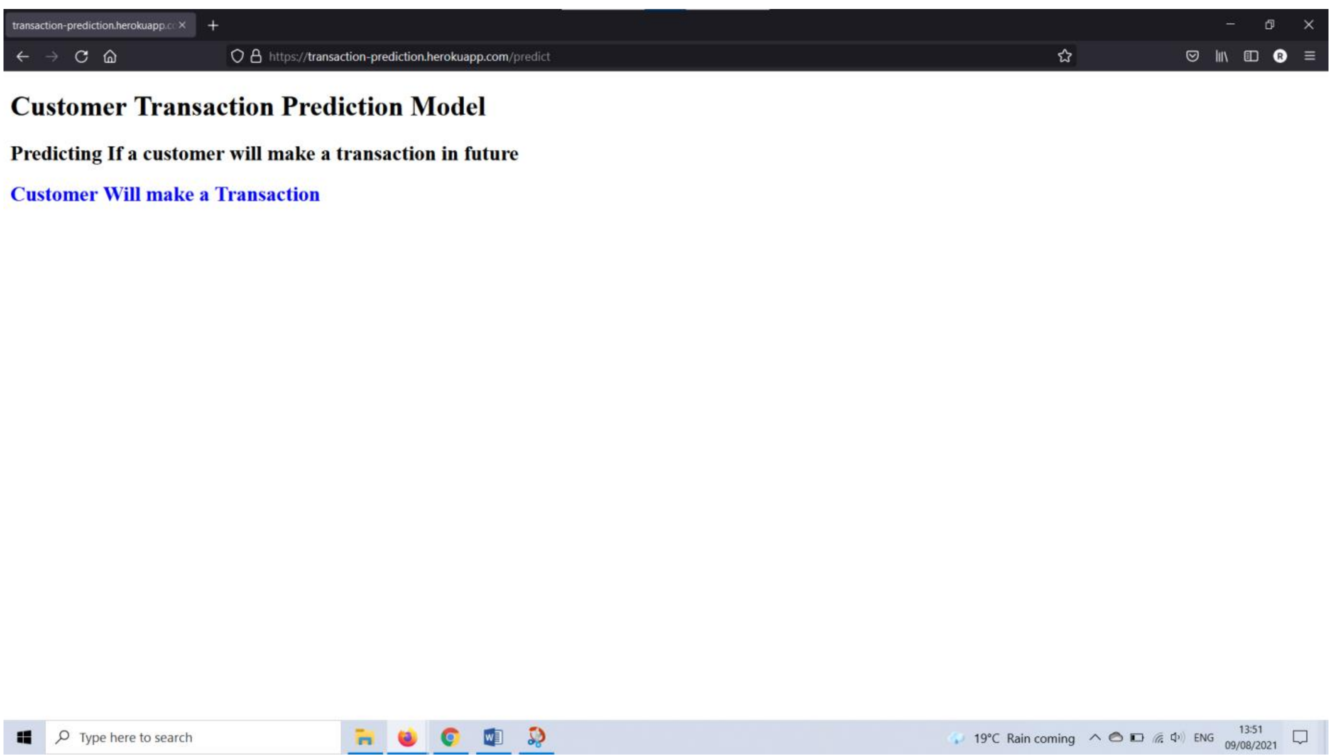
B. Screenshots of the flask app deployed on heroku.




transaction-prediction.herokuapp.c
https://transaction-prediction.herokuapp.com
Customer Transaction Prediction Model
Upload Customer Details Below
Browse...
prediction1.csv
Submit
Type here to search
19°C Rain coming
ENG
13:50
09/08/2021


transaction-prediction.herokuapp.c
https://transaction-prediction.herokuapp.com/predict
Customer Transaction Prediction Model
Predicting If a customer will make a transaction in future
Customer will not make a Transaction
Type here to search
19°C Rain coming
ENG
13:51
09/08/2021


transaction-prediction.herokuapp.co
https://transaction-prediction.herokuapp.com
Customer Transaction Prediction Model
Upload Customer Details Below
Browse... prediction2.csv
Submit
Type here to search
19°C Rain coming
ENG
13:51
09/08/2021


transaction-prediction.herokuapp.co
https://transaction-prediction.herokuapp.com/predict
Customer Transaction Prediction Model
Predicting If a customer will make a transaction in future
Customer Will make a Transaction
Type here to search
19°C Rain coming
ENG
13:51
09/08/2021