Customer Churn Analysis

Aishwarya Paruchuri, Archita Chakraborty, Manjushree Rajanna, Rohit Chandra
San Jose State University
November 2021

***Abstract*—**

I. Introduction

The mobile services market is growing significantly and sustainably, not only due to the size of the market, but also due to the increasing variety of services offered and fierce competition in the telecommunication industry. Regardless of the earliest stages of this industry, the method of contest has moved from procuring new endorsers to holding existing customers. This has been accomplished by participating in showcasing endeavors and by luring customers from rival organizations.

As indicated by a survey in 2004, it cost around 300 dollars per record to secure new clients and 25 dollars to hold existing clients. This implies that it was significantly more costly to procure new clients than to hold existing ones. Hence, financially, it makes more sense for an organization to focus on retaining its existing customers. As a result, churn management is a major area of focus.

A. *Dataset*

The data set has been randomly collected from an Iranian telecommunication company's database over a period of 12 months. The data set contains 3150 customer data with the below mentioned columns-

Feature Name	Type	Description
Call Failures	Categorical	Number of call failures
Complains	Numerical	Binary (0: No complaint, 1: complaint)
Subscription Length	Numerical	Total months of subscription
Charge Amount	Categorical	0: lowest amount, 9: highest amount
Seconds of Use	Numerical	total seconds of calls
Frequency of use	Numerical	total number of calls
Frequency of SMS	Numerical	total number of text messages
Distinct Called Numbers	Numerical	total number of distinct phone calls
Tariff Plan	Categorical	binary (1: Pay as you go, 2: contractual)
AgeGroup	Categorical	1: younger age, 5: older age
Status	Categorical	binary (1: active, 2: non-active)
Customer Value	Numerical	The calculated value of customer
Churn	Categorical	binary (1: churn, 0: non-churn) - Class label

Fig. 1. Table : Dataset

Note: The output feature - "churn" has 495 records which belongs to the churned class and 2645 records belong to the non-churned class. This shows that the data is highly imbalanced.

B. *Association Analysis*

Multiple factors can affect customer churn(Fig. 1).

1. Customer dissatisfaction: Customer complaints and Service failure rates are positively associated and length of customer association is negatively associated with customer churn probability.

2. Level of service usage: Number of calls, Minutes of monthly use and Number of distinct calls are negatively associated with customer churn probability.

3. Switching costs: Type of service is positively associated with the probability of subscriber churn.

4. Customer demographic variable: Customer age is positively associated with customer churn probability.

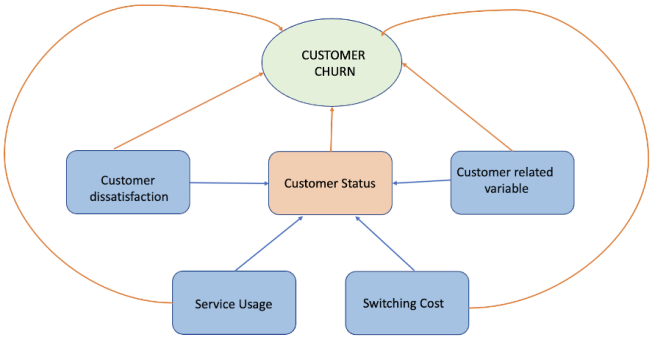

Fig. 2. Association Analysis

C. *Exploratory Data Analysis*

1. Data Preprocessing

It's a data mining approach for converting raw data into a usable and efficient format.

Steps Involved in Data Preprocessing:

(A) Data Cleaning: There may be various useless and missing elements in the raw data. Data cleaning is used to deal with this aspect. It entails dealing with missing data and noisy data.
Method used to find the null values: Missingno()

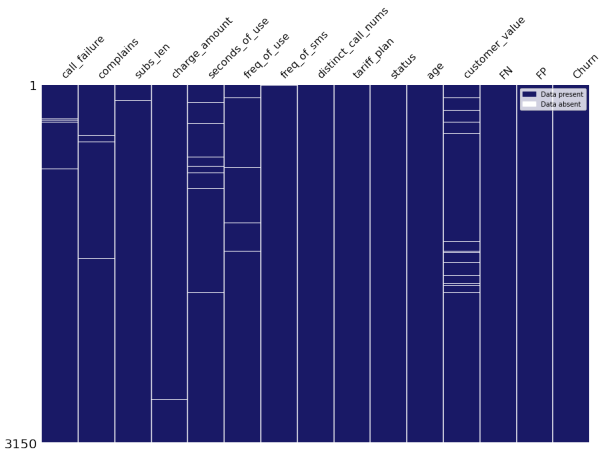

Fig. 3. Missingo Matrix Plot

Observation: Missing values are shown by white striped lines in each column. The columns such as, call failure, complaints, subscription length, charge amount, seconds of use, frequency of use, and customer value have missing values that must be cleaned.

Techniques used to handle null values in the data:

i) Median: Some attributes, such as secondsOfUse and customerValue, have outliers. We impute median value in place of null values for these columns because mean is prone to outliers.
ii) Mean : We impute mean values in place of null values for the remaining features that don't have outliers.

(B) Outlier Detection:

Outliers are extreme values that deviate from other observations on data , they may indicate a variability in a measurement, experimental errors or a novelty. In other words, an outlier is an observation that diverges from an overall pattern on a sample.

Observation- As we can see in Fig.4, there are outliers in our data set, especially seconds of use feature has the most outliers.

Methods used to treat Outliers:
i) Drop the outliers
ii) Replace with median or a constant value

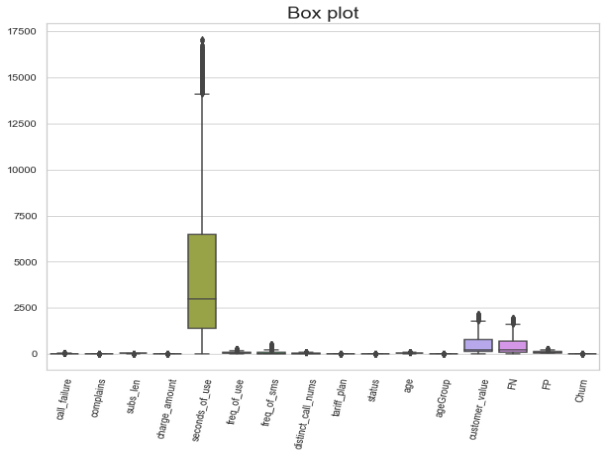

Fig. 4. Null Values in each column of the data set

2. Feature Scaling: This step is conducted to convert the data into a format that can be used in the mining process. Some of the classification models are based on probability, so we have scaled the data using MinMaxScalar() which transforms the values between 0 to 1 instead of StandardScalar() which transforms the data between -1 to 1.

3. Data Visualizations The graphical depiction of information and data is known as data visualisation. Data visualisation tools make it easy to examine and comprehend trends, outliers, and patterns in data by employing visual elements like charts, graphs, and maps.

Distribution of the output feature - "Churn":

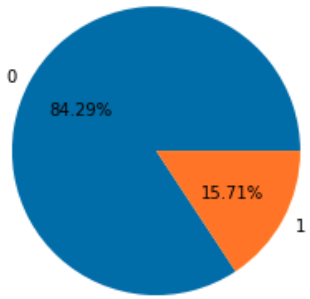

Fig. 5. Distribution of Churn

Observation- In the output column, we can see that there are 84.29 percent non-churn customers and 15.71 percent churn customers in total, indicating a data imbalance.

Distribution of all the categorical features:

Fig. 6. Frequency Distribution of Different Categorical Variable

Observation1- We notice that many of the customers are between the ages of 30 and 60, and they are less likely to churn
Observation2- We see a big difference in count between churn and non-churn active customers (where active customers are more likely to not churn) and churn and non-churn inactive customers (where inactive customers are more likely to churn).
Observation3- We notice that a higher percentage of consumers have no call failures and thus no complaints.
Observation4- We notice that a higher percentage of consumers in the 30-60 age group have a charge amount of o, which denotes the lowest salary

Distribution of all the numerical features

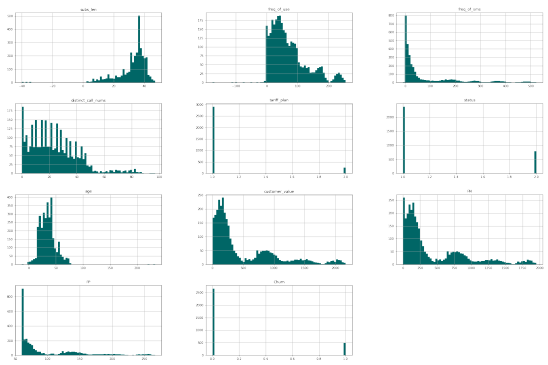

Fig. 7. Distribution of all columns

Observation- Above figure shows the frequency distribution of all the columns are shown above.

4. Feature Engineering

It involves deriving new features based on the existing features in the data set. In the data set, We have identified Age feature and performed feature engineering on it to create a new feature called AgeGroup to combine different age values in five different age intervals. The following table shows the age intervals:

Age group	Age interval
1	Less than 15
2	Between 15 and 30
3	Between 30 and 45
4	Between 45 and 60
5	Above 60

Fig. 8. Table: Age Group

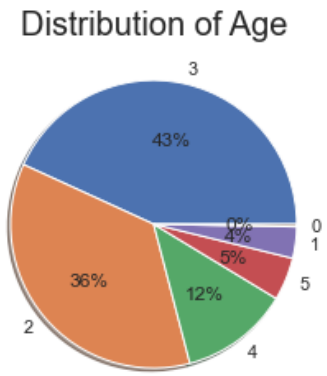

Fig. 9. Pie Chart displaying the age distribution of customers

Observation- As we can see from the pie chart, most of the customers belong to the age group 30-45.

5. Correlation analysis

From Fig.10 :
a: We observe a positive correlation between freq of use and distinct call numbers. The correlation coefficient value is 0.9389.
b: We observe a positive correlation between charge amount and call failure. The correlation coefficient value is 0.5817.
c: We observe a positive correlation between freq of use and customer value. The correlation coefficient value is 0.9249.
d: We observe very less correlation between call failure and churn. The correlation coefficient value is -0.0093.

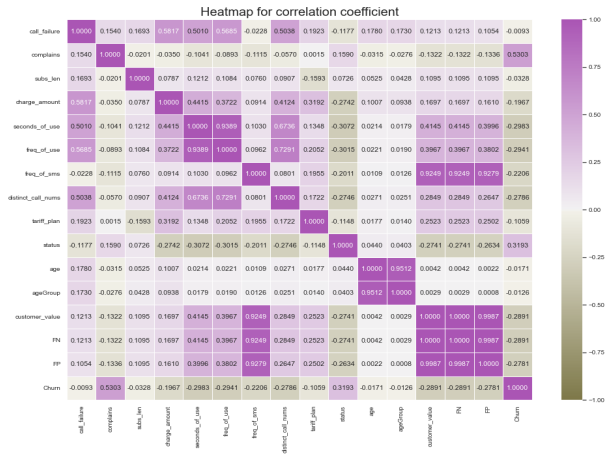

Fig. 10. Heatmap for correlation coefficient

6. Feature Selection

We used SelectKBest feature selection technique to select top 13 features to train different multi-classification model. We can visualize with the help of horizontal bar plot shown below.

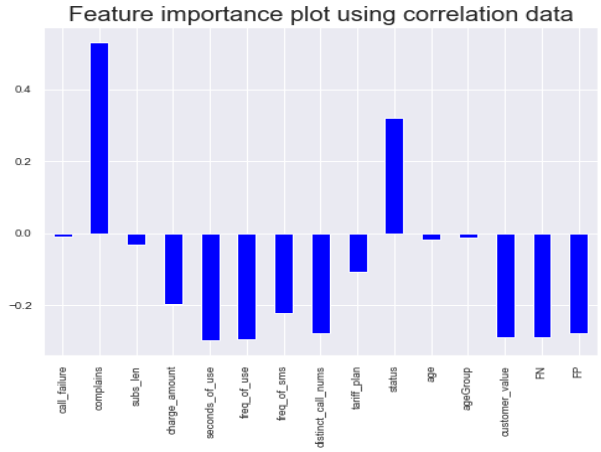

Fig. 11. Churn vs all features

Observation- From the graph we observe that status and complain are influential features for this dataset.

II. METHODS

As the data is highly imbalanced, we used Synthetic Minority Oversampling Technique(SMOTE), which involves duplicating samples in the minority class and Under Sampling Majority data set technique(UnderSampling), which randomly adds more minority observations by replication. All our analysis is done with imbalanced data, SMOTE generated data and under-sampling generated data. While training models on these datasets, we performed hyper-parameter tuning using GridSearchCV, which loops over predefined hyper-parameters and fits the model to the training data using best parameter values obtained. We considered following models to analyse the data:

1)XGBoost Classifier- XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. The goal is to improve weak learners by using a gradient descent approach to minimise errors. When compared to other algorithms, it is thought to be exceptionally efficient and quick. XGBoost classifier yields following results:

XGB_Classifier					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	94.07	97.64	84.32	76.35	80.14
Balance Data – SMOTE	96.86	99.53	96.5	97.23	96.87
Balance Data – Undersampling	89.9	96.14	88.3	91.8	90.00

Fig. 12. XGB Classifier Results

2) Naive Bayes Classifier- The Bayes Theorem-based probabilistic machine learning method Naive Bayes(NB) is employed in a wide range of categorization problems.

2a) Gaussian Naive Bayes Classifier- It is a naive bayes algorithm that is unique. When the features have continuous values, it's employed particularly. It's also expected that all of the characteristics have a Gaussian distribution, or a normal distribution. Gaussian Naive Bayes Classifier yields following results:

GaussianNB					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	68.36	89.06	32.15	91.89	47.63
Balance Data – SMOTE	76.52	89.84	69.75	93.59	79.93
Balance Data – Undersampling	77.44	91.73	70.55	93.91	80.57

Fig. 13. GaussianNB Classifier Results

2b) Multinomial Naive Bayes Classifier - It is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. Multinomial Naive Bayes Classifier yields following results:

MultinomialNB					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	89.42	89.63	91.37	35.81	51.45
Balance Data – SMOTE	81.48	91.03	82.49	79.89	81.17
Balance Data – Undersampling	83.16	92.56	85.5	79.72	82.51

Fig. 14. Multinomial Classifier Results

2c) Complement Naive Bayes Classifier - It is well-suited to dealing with unbalanced data sets. Instead of calculating

the probability of an item belonging to a certain class, we calculate the probability of the item belonging to all classes in complement Naive Bayes. Complement Naive Bayes Classifier yields following results:

ComplementNB					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	81.9	89.63	45.41	77.02	57.14
Balance Data – SMOTE	81.42	91.03	82.46	79.77	81.09
Balance Data – Undersampling	83.16	92.56	85.5	79.72	82.51

Fig. 15. ComplementNB Classifier Results

3) Support Vector Classifier(SVC) - It is a linear model that can be used to solve classification and regression problems. It can solve both linear and nonlinear problems.The algorithm generates a line or hyper-plane that divides the data into categories. Support Vector Classifier yields following results:

SVM					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	90.37	91.61	84.33	47.29	60.6
Balance Data – SMOTE	87.82	94.55	85.16	91.58	88.25
Balance Data – Undersampling	86.87	94.05	84.71	89.86	87.21

Fig. 16. SVM Classifier Results

4) Decision Tree - It is supervised machine learning that categorises or predicts outcomes based on the answers to a previous set of questions. Decision Tree yields following results:

Decision Tree					
DATASET -TYPE	ACCURACY	AUC	PRECISION	RECALL	F1-SCORE
Imbalance Data	100	100	100	100	100
Balance Data – SMOTE	92.72	93.19	93.36	91.95	92.65
Balance Data – Undersampling	84.18	88.16	83	85.81	84.38

Fig. 17. Decision Tree Classifier Results

III. COMPARISONS

A. *Performance Metrics*

1. Precision- The number of positive class predictions that actually belong to the positive class is measured by precision. 2. Recall- The number of positive class predictions made out of all positive examples in the dataset is measured by recall. 3. F1-Score- F1-Score generates a single score that accounts for both precision and recall concerns in a single number. 4. Accuracy- It's the proportion of correct predictions to total input samples. It only works if each class has an equal amount of samples. 5. AUC - The Area Under the Curve (AUC) is a curve that measures a classifier's ability to distinguish between classes. The greater the AUC, the better.
Note: Since our data is class imbalanced, we majorly rely on F1-Score, Recall and Precision.

B. *Comparison*

Among all the models built on imbalanced data, XGBoost is the winner as XGBoost can offer better performance on binary classification problems with a severe class imbalance.The model performed better with good precision(84.32 percent) as well as recall score(76.35 percent). The Fig.18 depicts the high AUC percentage(98 percentage) of XGBoost Classifier wrt other models.

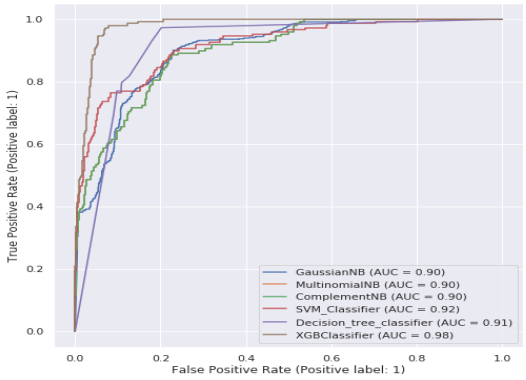

Fig. 18. Roc Curve of Models built on Imbalanced data

Among all the models built on Undersampling generated data, XGBClassifier is the best one. As we can see from the Fig.19, XGBClassifier has highest area under the curve(96 percent).

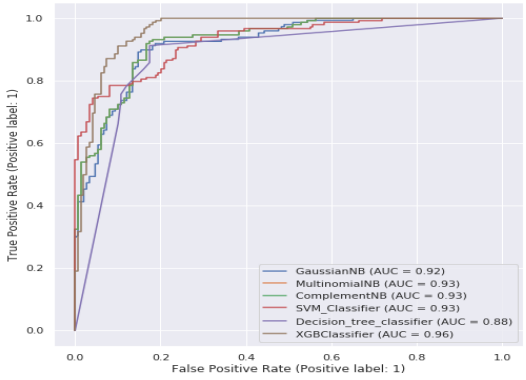

Fig. 19. Roc Curve of Models built on Undersampling data

As we can see from the Fig.20, XGBClassifier has highest area under the curve(98 percent) for balanced data generated using SMOTE technique.

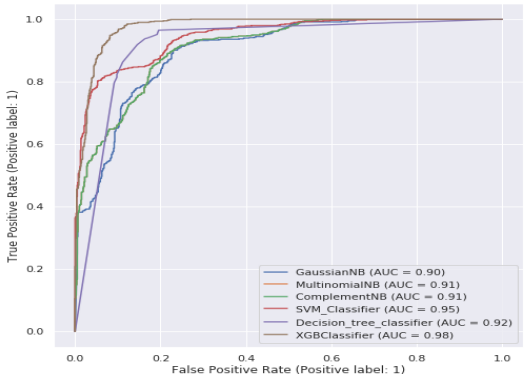

Fig. 20. ROC Curve of Models built on SMOTE data

The Fig.21 shows the ROC graph of all types of data discussed, built on various models specified above.

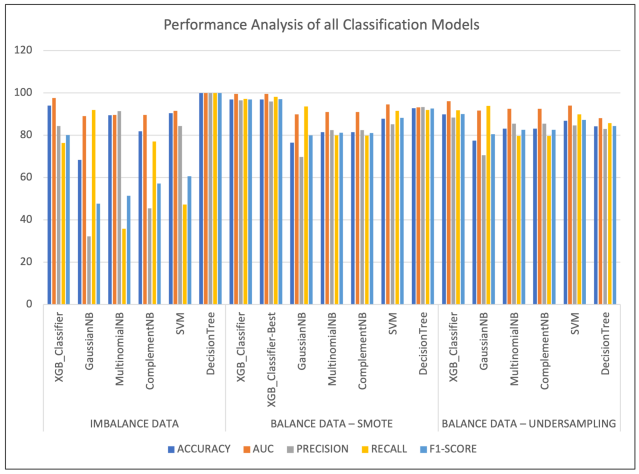

Fig. 21. Performance Analysis of all Classification Models

IV. CONCLUSIONS

V. REFERENCES

[1] *https://analyticsindiamag.com/tips-for-automating-eda-using-pandas-profiling-sweetviz-and-autoviz-in-python/*

[2] **Dataset Link:** *https://tinyurl.com/TelecomCustomerChurnDataset*

[3] *Ahmed U, Khan A, Khan SH, Basit A, Haq IU, Lee YS (2019) Transfer learning and meta classification based deep churn prediction system for telecom industry.*

[4] *Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A, Hussain A (2016) Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access 4:7940–7957*