**A CASE STUDY: BANK CHURNERS**

**SHIGHA\_VELLOTH**

**Abstract: what is churning?** Churn is the measure of how many customers stop usingaproduct. This can be measured based on actual usage or failure to renew (when the product is sold using a subscription model). Often evaluated for a specific period of time, there can be a monthly, quarterly, or annual churn rate. Churn rates are important because losing customersmeans losing revenue. So, the bottom line is that high churn could affect your bottom line. Another reason it's critical to improve customer retention and reduce churn is that it's generally more expensive to find new customers than it is to keep existing ones.

**Problem statement:**

This analysis aim to find why the customers are leaving credit card service in a Bank and to predict who is gonna get churned so Bank can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction.

Now age, salary, marital\_status, credit card limit, credit card category, etc. There are nearly 18 features of 10000 customers are collected for study.

**1.INTRODUCTION:**

Banks and financial services providers offer a range of services to both keep and draw in new consumers in the highly competitive banking sector. Most banks and financial services providers offer bonuses for first-time clients who sign up for services like credit cards, ATM services, debit cards, home banking, internet banking, mobile banking, etc. to entice users to use those services. They are frequently referred to as bonuses and sign-up deals. However, some clients abuse these benefits to increase their financial security. Which has emerged as one of the most difficult issues facing the banking sector. Therefore, banks and other financial institutions should take the appropriate steps to prevent the abuse of those bonuses and offers.

Heavy restrictions could result in the loss of current clients if they were applied to these matters. Due to the rigorous rules that a particular bank will have, customers are likely to switch allegiances from one bank to another as a result of the greatly larger customer base and improved awareness of the quality of the service. Customers may be extremely delighted with a banking service depending on a number of variables, including flexible rules and limits, accessibility to cutting-edge technology, customer-friendly bank employees, cheap interest rates, the location of the bank, the range of services it offers, etc.

Therefore, imposing severe regulations to prevent the abuse of the offers is not a workable approach. It is crucial to keep up good client relations if banks and other financial services providers are to survive in the digital age. Therefore, it is crucial to analyse customer behaviour and to pinpoint the causes of customer churn in various services offered by the bank or finance firm in order to implement the right regulations and constraints that assure quality relationships with the consumers**.**

The manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate if one could predict for them who is gonna get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction

Now, the dataset we used consists of 10,000 customers mentioning their age, salary, marital\_status, credit card limit, credit card category, etc. There are nearly 18 features.

Relation between big banks and small credit cards:

The use of credit cards is crucial in the banking industry. We might scratch credit cards as buyers with the most advantageous deals and financial stability. I use credit cards to gain airline miles, hotel discounts, food shopping savings, and sign-up incentives on different apps.For each transaction, users receive bonus prizes and their credit score. Users that use several credit cards and do not utilise just one are despised by banks. Churners are the name given to these users. Closing credit cards means doing so after the bonus has been credited to your account and before the subsequent annual fee is assessed Credit card reward programmes are offered by banks, corporate finance departments, and business finance managers as a way to earn cash advances, annual fees, interest charges, and in\_hand money

**2.OBJECTIVE**

To find why customers are leaving the credit card services in a bank and to predict who is gonna get churned.

**3.Data collectection and analysis**

**About dataset:**

Feature Description

1. Client number : Unique identifier for the customer holding the account
2. Attrition\_Flag : Account is closed or not.(Existing Customer / Attrited Customer.
3. Customer\_Age : Cusomer's age in years.
4. Gender : Male/Female.
5. Dependent\_Count : Number of Dependents.
6. Education\_Level : Education Qualification of the account holder.
7. Marital\_status : Married, Single, Divorced, Unkown.
8. Income\_Categpry : Annual Income Category of the account holder.
9. Card\_Category : Type of card(Blue, Silver, Gold, Platinum)
10. Credit\_limit : Credit limit on credit card

* These are the columns will be used during the analysis.

**Table 1:** Dataset of 10,000 customers mentioning their age, salary, marital\_status, credit card limit, credit card category, etc. There are nearly 18 features.

Python code:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.svm import SVC

from sklearn.tree import DecisionTreeClassifier

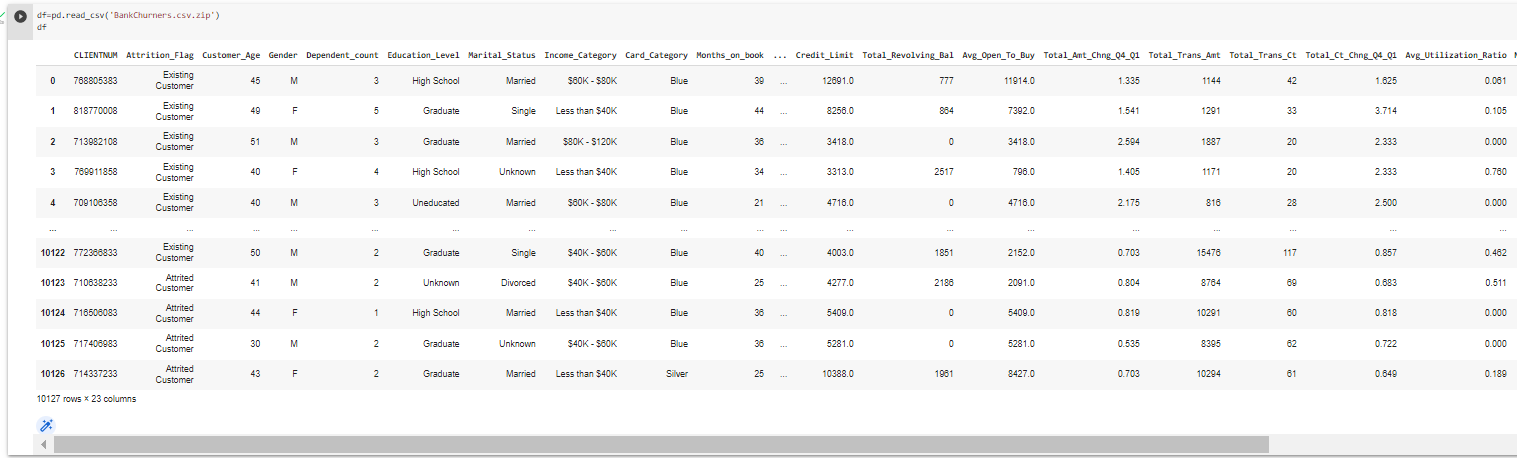
from sklearn.ensemble import RandomForestClassifier

from google.colab import files

uploaded=files.upload()

df=pd.read\_csv('BankChurners.csv.zip')

df

****

1. **Data understanding:**

**A picture containing text, indoor, computer, white

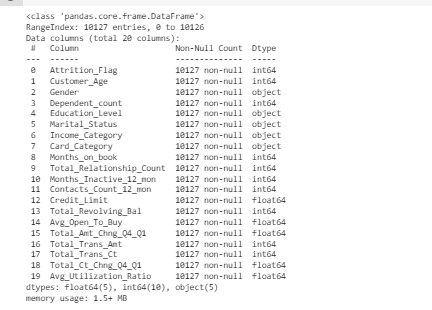
Description automatically generated**

**Background pattern

Description automatically generated**

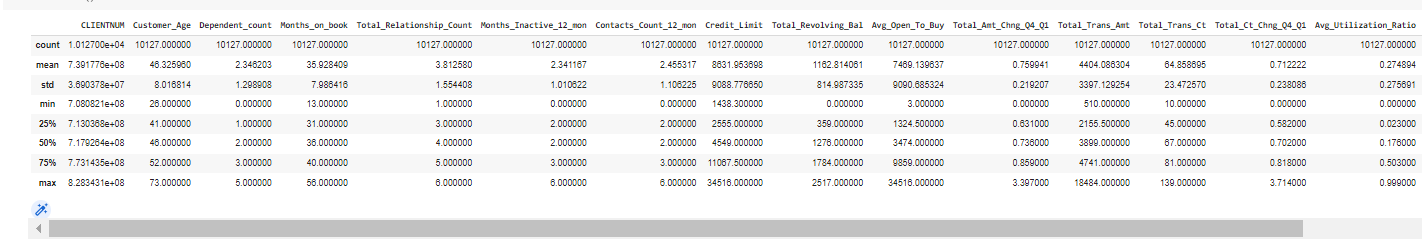
**Information about dataset columns**

df.info()

****

**Statistical Report:**

df.describe()

****

**Numbers of unique rows in each feature:**

#Checking number of unique rows in each feature

df.nunique().sort\_values()

****

1. **Exploratory Data Analysis**

#Finding duplicates value

counter = 0

r,c = df.shape

df1 = df.copy()

df1.drop\_duplicates(inplace=True)

df1.reset\_index(drop=True,inplace=True)

if df1.shape==(r,c):

    print('\n\033[1mInference:\033[0m The dataset doesn\'t have any duplicates')

else:

    print(f'\n\033[1mInference:\033[0m Number of duplicates dropped ---> {r-df1.shape[0]}')

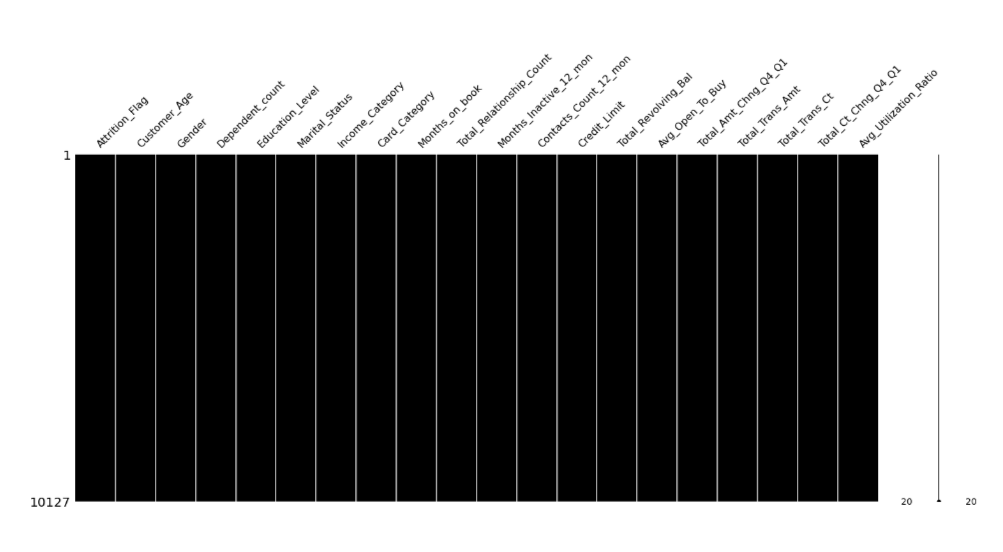
* The dataset not having any duplicate values

Checking for null values:

df.isnull().sum()

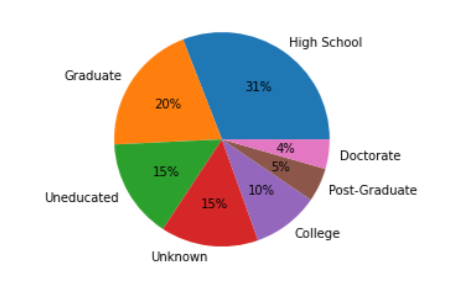


No null values present



colors = sns.color\_palette('dark')

plt.pie(df.Education\_Level.value\_counts(), labels=df.Education\_Level.unique(),autopct='%0.0f%%')



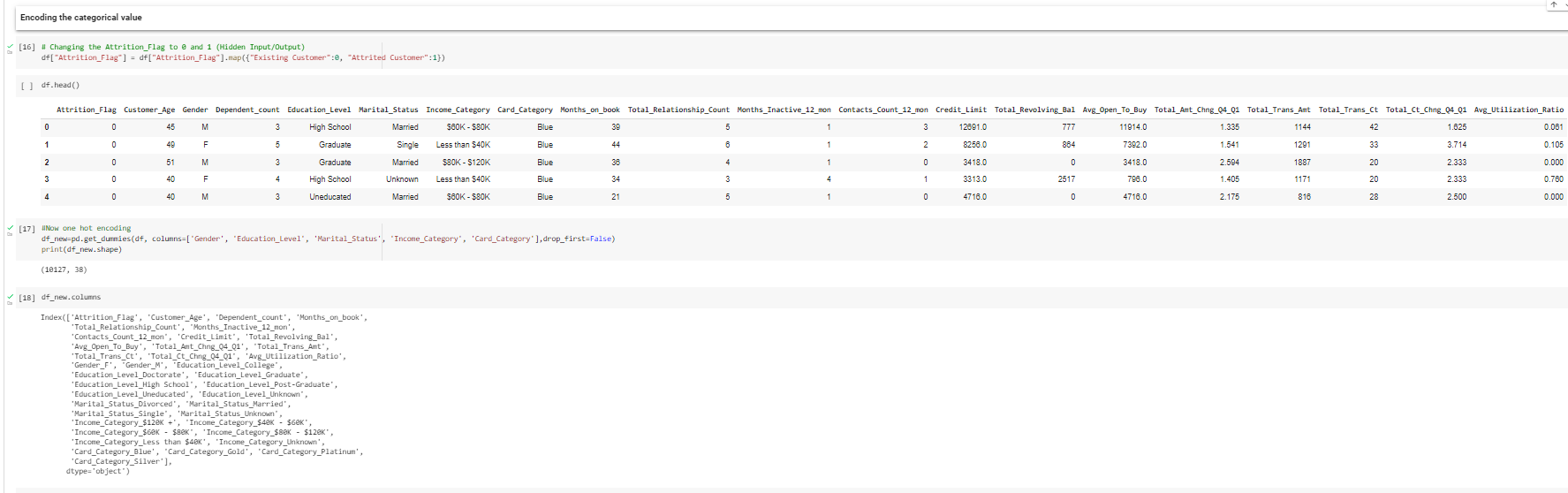
1. **Feature Engineering:**

Encoded the categorical value to numerical using one-hot encoding

One-hot encoding:

One hot encoding method involves transforming categorical information into a format that may be given to ML algorithms to help them perform better at prediction.

*FIGURE: PYTHON CODE AND RESULT*



**1V. FEATURE SELECTION**

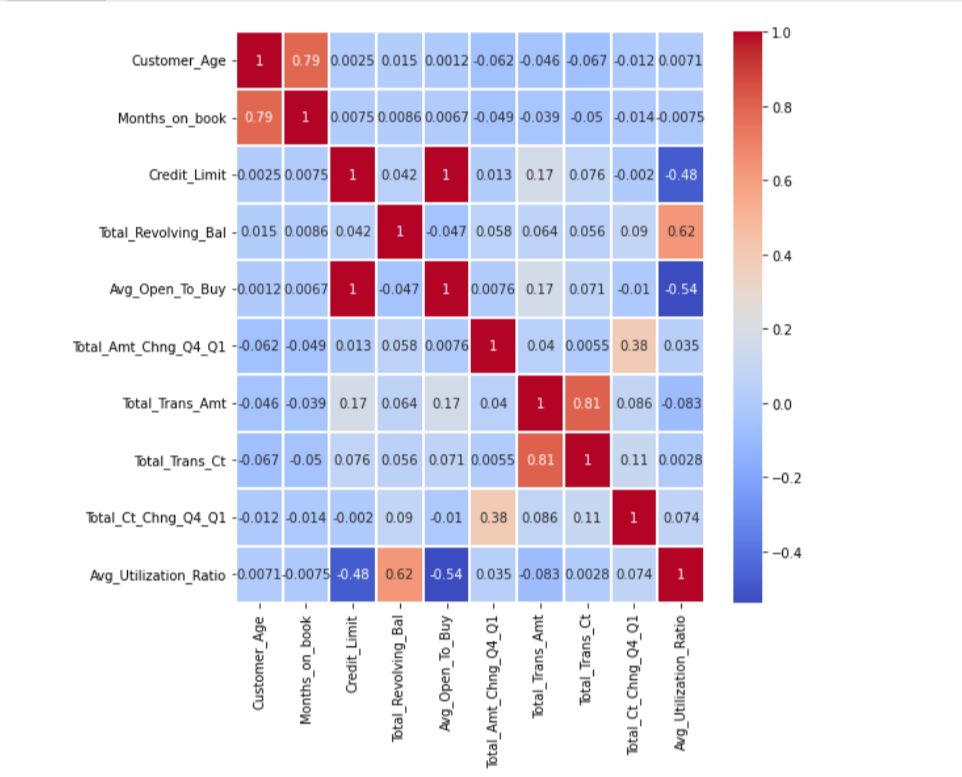
features = df\_new.columns

plt.figure(figsize=[30,25])

plt.title('Features Correlation-Plot')

sns.heatmap(df\_new[features].corr(), vmin=-1, vmax=1, center=0, annot=True) #,

plt.show()

****

Multicollinearity exists

*Figure: correlation of numerical data with target*

import matplotlib

background\_color = "#f6f6f6"

fig = plt.figure(figsize=(44,18), facecolor=background\_color)

gs = fig.add\_gridspec(1, 1)

ax0 = fig.add\_subplot(gs[0, 0])

colors = ["#0078d7"]

colormap = matplotlib.colors.LinearSegmentedColormap.from\_list("", colors)

ax0.set\_facecolor(background\_color)

ax0.text(-1.1, 1.25, 'Correlation of Numerical Features with Target', fontsize=20, fontweight='bold')

chart\_df = pd.DataFrame(df.corrwith(df['Attrition\_Flag']))

chart\_df.columns = ['corr']

sns.barplot(x=chart\_df.index, y=chart\_df['corr'], ax=ax0, color='#0078d7', zorder=3, edgecolor='black', linewidth=3)

ax0.grid(which='major', axis='x', zorder=0, color='#EEEEEE', linewidth=0.4)

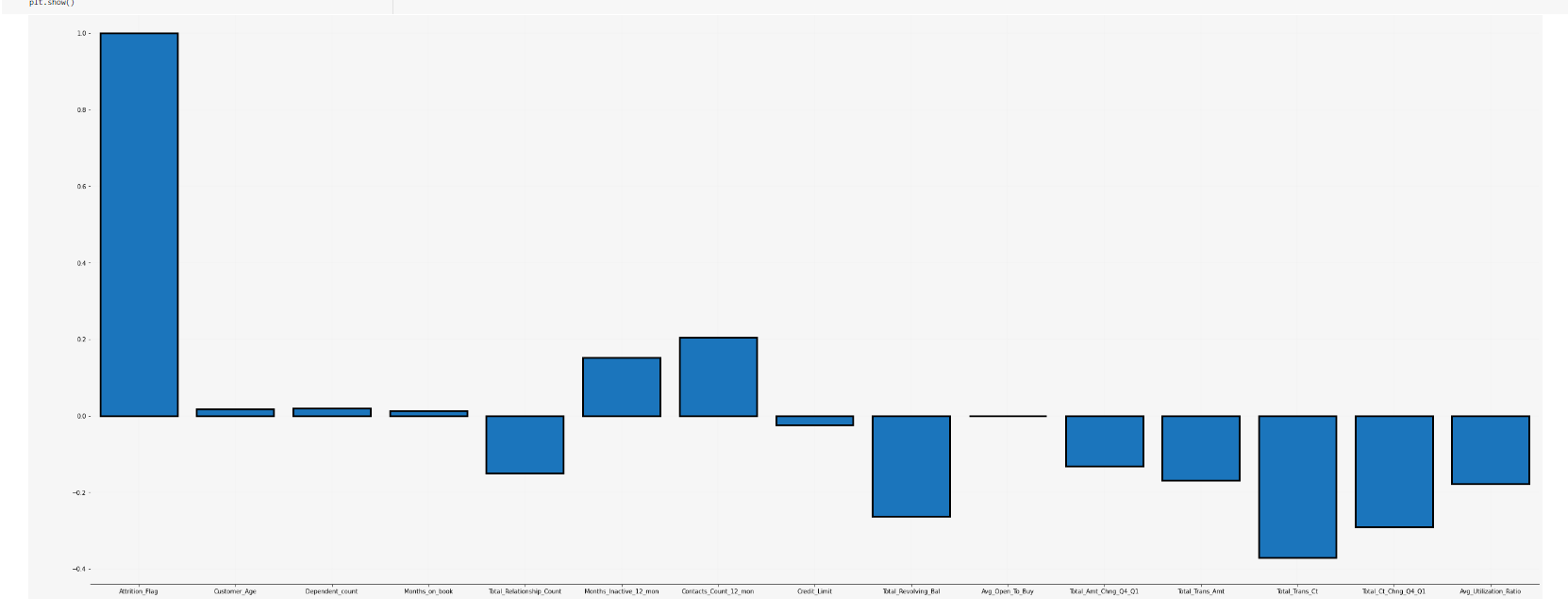
ax0.grid(which='major', axis='y', zorder=0, color='#EEEEEE', linewidth=0.4)

ax0.set\_ylabel('')

for s in ["top","right", 'left']:

    ax0.spines[s].set\_visible(False)

plt.show()



Based on these results we selected testing set and target set as x and y respectively

Graphical user interface, application

Description automatically generated

v. Modelling:

Decision tree:

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

The strengths of decision tree methods are:

1.Decision trees are able to generate understandable rules.

2.Decision trees perform classification without requiring much computation.

3.Decision trees are able to handle both continuous and categorical variables.

4.Decision trees provide a clear indication of which fields are most important for prediction or classification.

GINI INDEX

The internal working of Gini impurity is also somewhat similar to the working of entropy in the Decision Tree. In the Decision Tree algorithm, both are used for building the tree by splitting as per the appropriate features but there is quite a difference in the computation of both the methods.

RANDOM FOREST

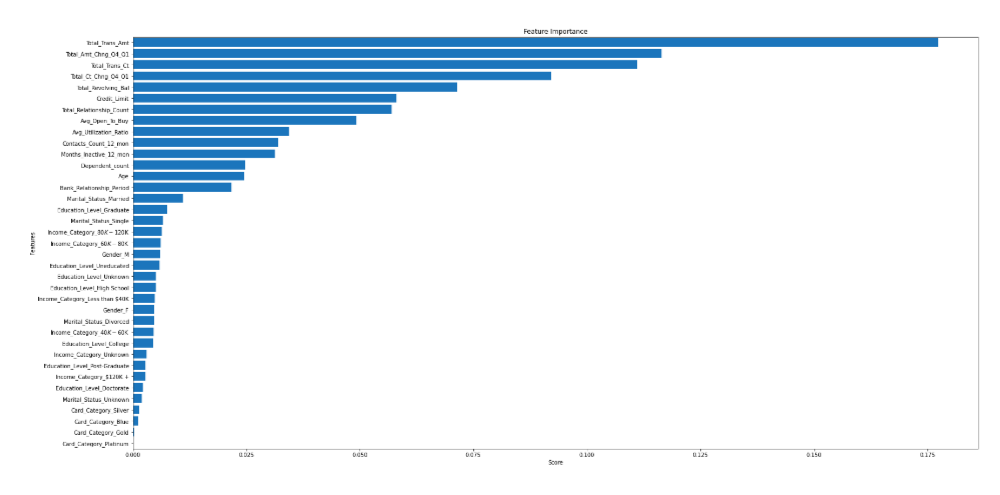
Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees

Python code and output:

Graphical user interface, application

Description automatically generated with medium confidence

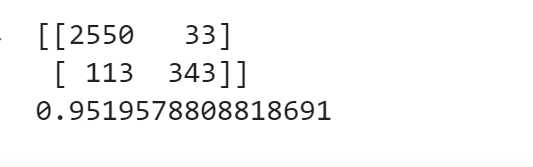
Feature importance:



**3.RESULTS**

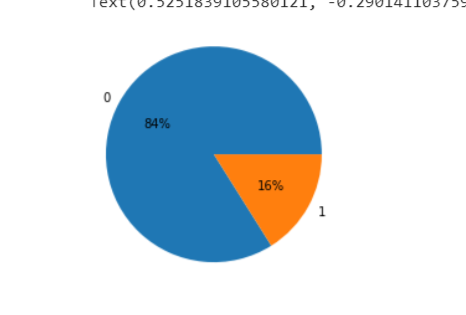
* By using decision tree, we get the accuracy of 93%
* When we give Gini index to decision-tree we got accuracy as 91%
* Random forest model given the accuracy 95%, which means our model got improved and its good.

Confusion matrix and accuracy of Random Forest:



**4.CONCLUSION:**

* There are 16.07% of customers who have churned.

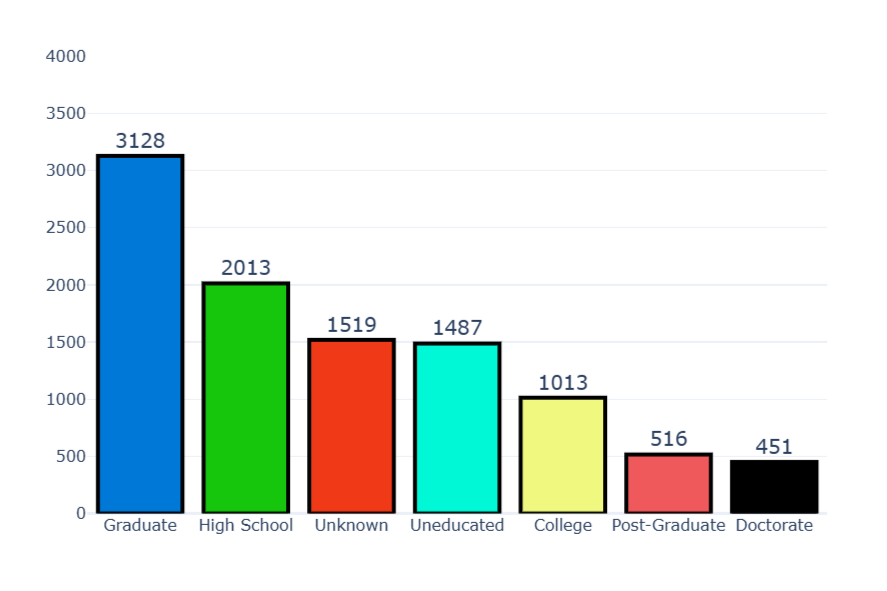


* The proportion of current and ascribed customers is considerably unbalanced (83.9% and 16.1%), but the proportion of gender count is fairly even distributed (52.9% male and 47.1%).
* There are 14.4% more male consumers who have left than female customers in terms of attrition.

**Chart, bar chart

Description automatically generated**

• Highly educated customers are more likely to churn - Graduate level education accounts for a large share of ‘attrited customers' education levels (29.9%), followed by post-graduate level education (18.8%).

****

• Married (43.6%) and Single (41.1%) are the marital statuses of consumers who have churned at a higher rate than Divorced (7.4%) and Unknown (7.9%) –

• As you can see from the proportion of attrition customer's income group, it is strongly concentrated around 60K80K income (37.6%), followed by Less than 40K income(16.7120K+) (11.5%). Marital status of the attributable customers is largely clustered in Married status and Single.

Let conclude that customers with higher income doesn't likely to leave their credit card services than meddle-income customer.

Python code:

<https://colab.research.google.com/drive/1oLf52K6w1ZEOAw7IemRI7pwQg5iqjh5K#scrollTo=oy40qg34qzzR>