GREAT LEARNING

PGP in Data Science and Engineering

Final Report for:

FINANCIAL RISK ANALYSIS

CREDIT CARD DEFAULT

Group Members :
Enavamshi Gadikota
Mohit Sardar
Sruthi Basani
Radhakrishna Penugonda
Adarsh Raj

Mentor :
Naveen Koneti

Contents

1 Introduction to Domain

Credit cards are typically issued by banks, and it enables the cardholder to borrow funds from that bank and issuers customarily pre-set borrowing limits, based on an individual's credit rating and the banks also impose the condition that cardholders pay back the borrowed money, plus interest, as well as any additional agreed-upon charges.[7]

What is in it for banks?
Credit cards feature higher Annual Percentage rates (APRs) than other forms of consumer loans. Interest charges on the unpaid balance charged to the card are typically imposed one month after a purchase is made.

What makes this product attractive to customers?
Apart from borrowing money when in need, credit card users can avail discounts, travel points, cash-backs and many other perks unavailable to debit card holders. Consumers who pay off their cards in full and on time every month can profit substantially by running their monthly purchases and bills through reward points.

1.1 Taiwan Credit Card Crisis

The dataset we are analyzing comes from a bank in Taiwan and contains information from April 2005 to September 2005. There is some interesting background to this data because Taiwan underwent a credit card crisis around the same time.[8]

The Taiwanese government allowed the formation of new banks in 1990. Within a couple of years of this, the banks turned to credit and cash cards to expand their businesses. The banks spent lots of money on advertising their credit cards mainly targeting young people.

Requirement for credit card approvals were lowered so that even those without any income or any other means to pay back were able to acquire them. Moreover, the banks made it seem as if these credit cards could be consumed without any consequences thus attracting many customers. This careless and short-sighted act resulted in massive debts for many Taiwanese people.

By February 2006, debts from credit card reached $268 billion and more than 500,000 people were not able to repay their loans thus becoming what came to be known as "credit card slaves" (those who could only pay the minimum balance on their credit card debt every month).

This issue resulted in significant societal problems including crime, violence, homelessness and suicides. The suicide rate in Taiwan is the second highest in the world.

The suicide rate in 2008 increased 22.9% compared to the rate in 2005, and the main reason is unemployment and credit card debt.

The Taiwanese Finance Supervisory Commission took some strict counter measures to deal with these problems.

Banks were ordered to modify their requirements for issuing credit cards such as raising the income and job requirements, prohibiting improper credit card commercials, prohibiting inappropriate collection behaviors and prohibiting compound interest.

1.2 Problem Statement

Credit card business is one of the key areas in the banking industry where there exists a huge money flow. Credit card business has been a huge success in the banking market over the years. Most of the banks are also interested in this area expecting a good revenue. Many of the customers use their credit cards beyond their repayment capabilities leading to high debt accumulation and finally defaulting the credits.

Classifying the risky and non-risky customers has become a very big challenge for banks. So, the problem we are trying to analyse is to identify the risky and non-risky customers based on the available customer data, helping the bank to decide if a customer can repay his credits to the bank in the coming months.

It is important to identify the risky customers, those who will default soon but it is also important not to wrongly classify a good customer as a risky one since this lead to a bad customer experience.

2 Dataset Description

The dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card customers of a bank in Taiwan from April 2005 to September 2005.[2]

There are 25 variables:

1. **ID:** ID of each customer
2. **LIMIT_BAL:** Amount of given credit in NT dollars (includes individual and family/supplementary credit)
3. **SEX:** Gender (1=male, 2=female)

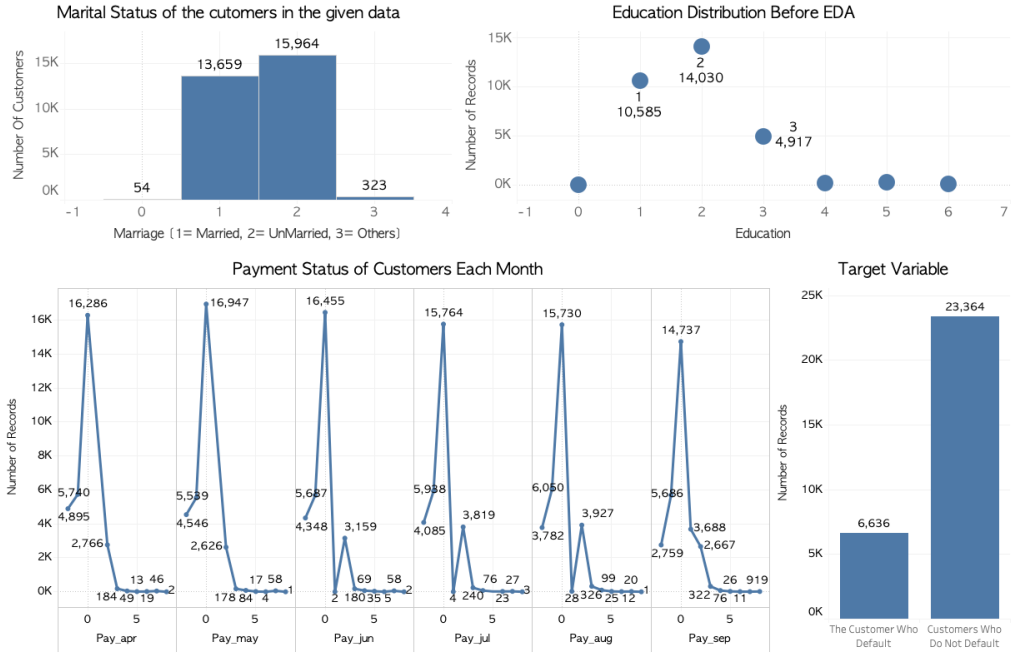
4. **EDUCATION:** (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
5. **MARRIAGE:** Marital status (1=married, 2=single, 3=others)
6. **AGE:** Age in years
7. **Repay_Sept:** Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, ... 8=payment delay for eight months, 9=payment delay for nine months and above)
8. **Repay_Aug:** Repayment status in August, 2005 (scale same as above)
9. **Repay_July:** Repayment status in July, 2005 (scale same as above)
10. **Repay_June:** Repayment status in June, 2005 (scale same as above)
11. **Repay_May:** Repayment status in May, 2005 (scale same as above)
12. **Repay_April:** Repayment status in April, 2005 (scale same as above)
13. **Bill_Amt_Sept:** Amount of bill statement in September, 2005 (NT dollar)
14. **Bill_Amt_Aug:** Amount of bill statement in August, 2005 (NT dollar)
15. **Bill_Amt_July:** Amount of bill statement in July, 2005 (NT dollar)
16. **Bill_Amt_June:** Amount of bill statement in June, 2005 (NT dollar)
17. **Bill_Amt_May:** Amount of bill statement in May, 2005 (NT dollar)
18. **Bill_Amt_April:** Amount of bill statement in April, 2005 (NT dollar)
19. **Pre_Pay_Sept:** Amount of previous payment in September, 2005 (NT dollar)
20. **Pre_Pay_Aug :** Amount of previous payment in August, 2005 (NT dollar)
21. **Pre_Pay_July:** Amount of previous payment in July, 2005 (NT dollar)
22. **Pre_Pay_June:** Amount of previous payment in June, 2005 (NT dollar)
23. **Pre_Pay_May:** Amount of previous payment in May, 2005 (NT dollar)
24. **Pre_Pay_April:** Amount of previous payment in April, 2005 (NT dollar)
25. **DEFAULT:** Default payment (1=yes, 0=no)

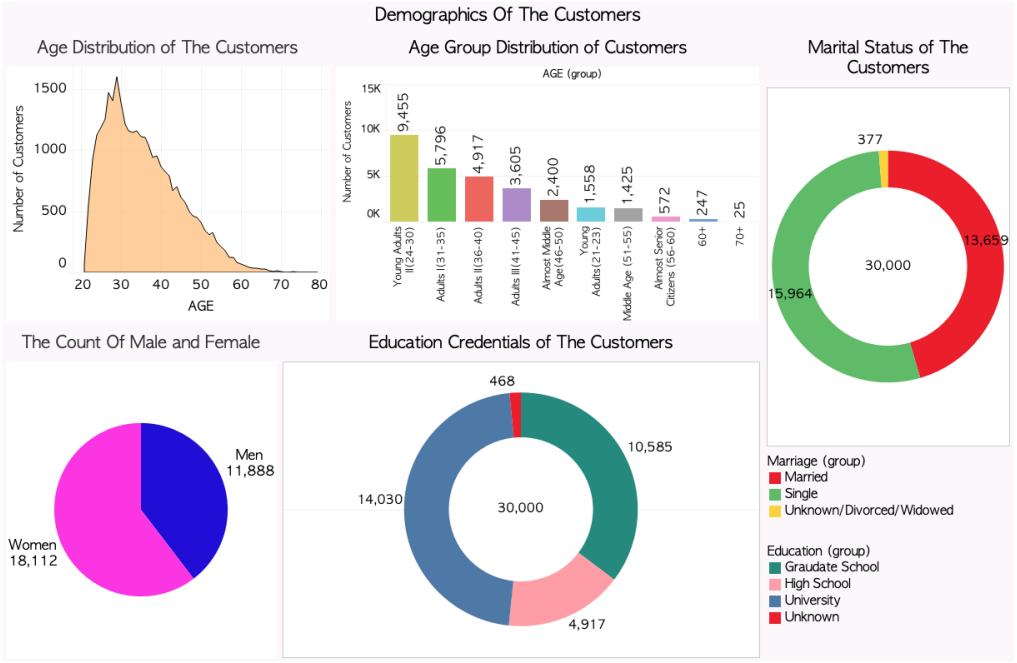
3 Exploratory Data Analysis

3.1 General Observations

- There are no missing values
- Every data point is important in the credit card business, so we did not consider any outliers in the bill amount, limit balance amount and the payment amount columns. Instead we just considered them as extreme values.
- There are some values in categorical columns that are not defined in the dataset description.
  - In the education column, as per the data description 1=graduate school, 2=university, 3=high school, 4=Others, 5=unknown, 6=unknown so we decided to replace 5 and 6 with 4 as Others.
  - In the marriage column, as per the data description 1=married, 2=single, 3=others, but we had some values in 0=unknown so we decided to replace the records with 0 as their marital status with 3 as others.
  - The payment status columns for all the months had values like -2,-1,0, 1 and as per the dataset description -1=pay duly, 1=payment delay for one month, 2=payment delay for two months, ... 8=payment delay for eight months, 9=payment delay for nine months and above. We did not have any explanation for what 0, -1 meant and a very important point to be noted is every months there are at least 50% of rows which had their payment status as 0. So we have taken values less than and equal to 0 as the group which made payments and the values greater as the group which had delay in their payment.
- The *Pre_Pay* amounts for each month correspond to the *Bill_Amt* of the previous month but the six months of *Pre_Pay* do not align with the six months of *Bill_Amt* given i.e *Pre_Pay_April* is given but Bill Amount for March is not given. Similarly Bill_Amt_Sept is given but the corresponding payment in October is not.

3.2 Univariate Analysis

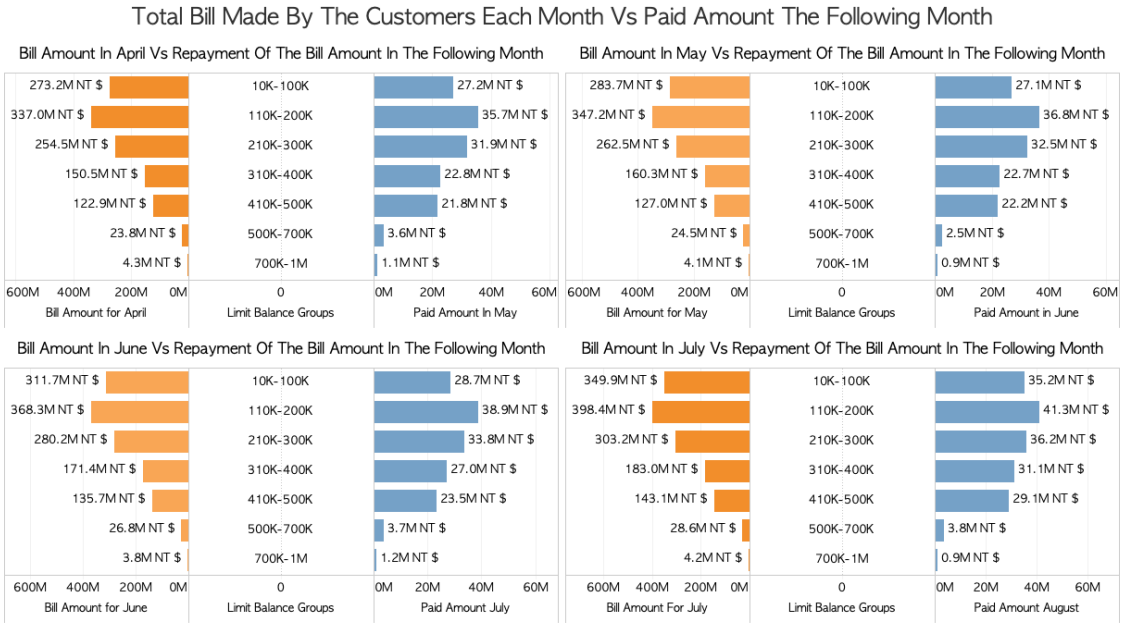




Inferences:

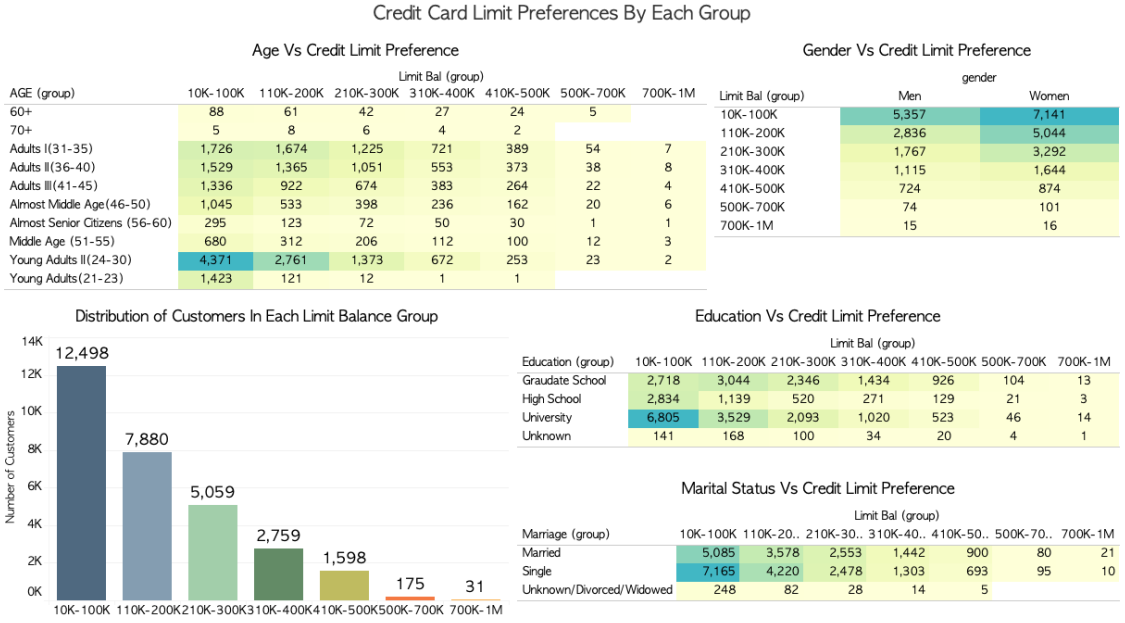
- Target Variable, *DEFAULT* is imbalanced.
- As mentioned in the facts of the case when new banks turned to other new business – credit cards and cash cards. In expanding this area of business, young people became target customers. Although young people tend not to have enough income, banks still issued credits cards to them And it is very evident from the age group distribution graph that the youngest customer in the data set is of 21 years and majority of the customers are between the age 24-35 which is 50% of the dataset.
- After the new regulations came into force in 2005, credit card applicants were required to have jobs and a specified level of income as per the acts of the case. It is very clear from the education credentials of the customers that the majority of the credit card holders have University level or Graduate level education. Meaning these customers have a steady source of income.

3.3 Bill Amount vs Amount Paid



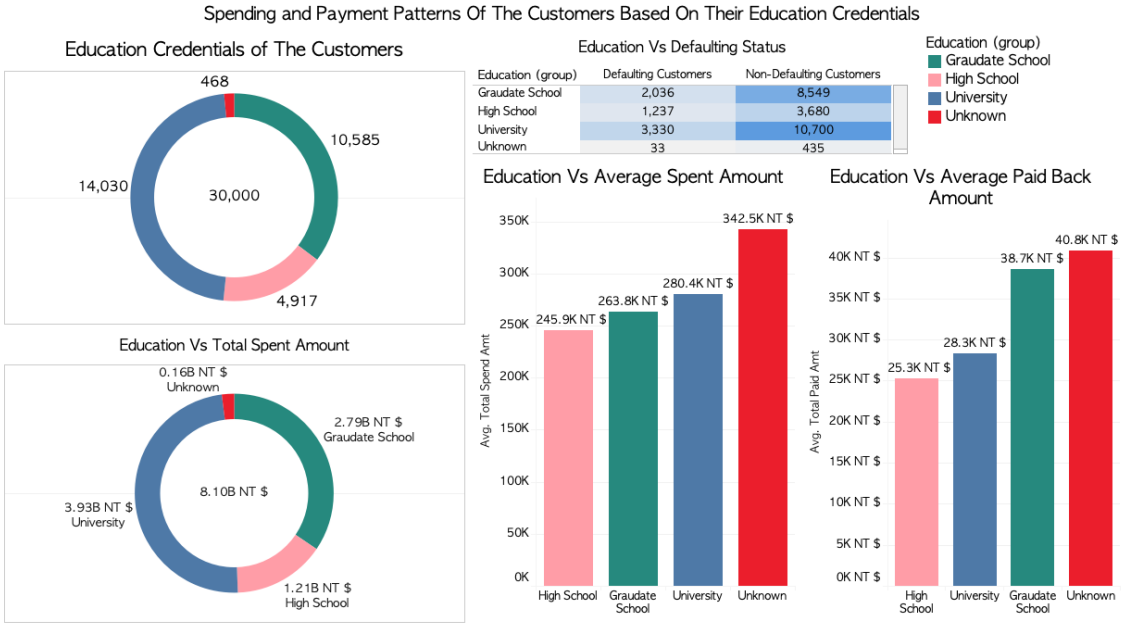
As Mentioned earlier in the the facts of the case, In Taiwan, in February 2006, debt from credit cards reached $268 billion USD. More than half a million people were not able to repay their loans and the above mentioned dashboard explains that majority customers were only able to repay just 10-20% of their total spend amount each month and hence the huge debt.

3.4 Credit Card Limit Preferences

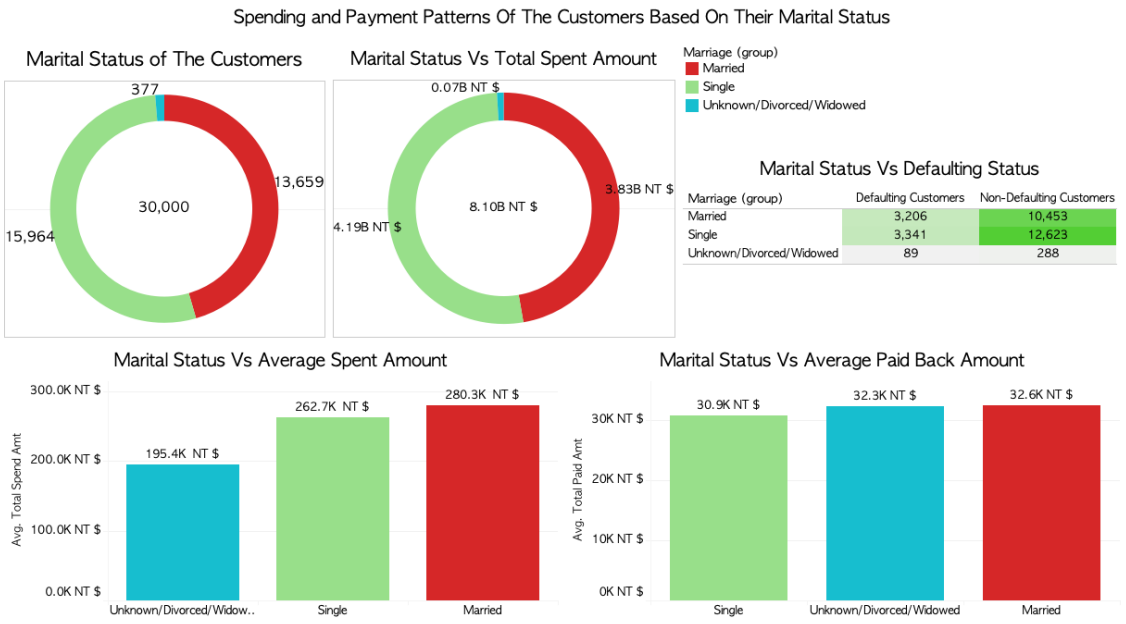


- 67% of the customer's credit limit was in the range 10K to 200K and 31% of the customers in the range 200K-500K and the last 2% customers who has credit of 500k+.
- None of factors like education, gender or marital status have any influence on the customer's credit limit.

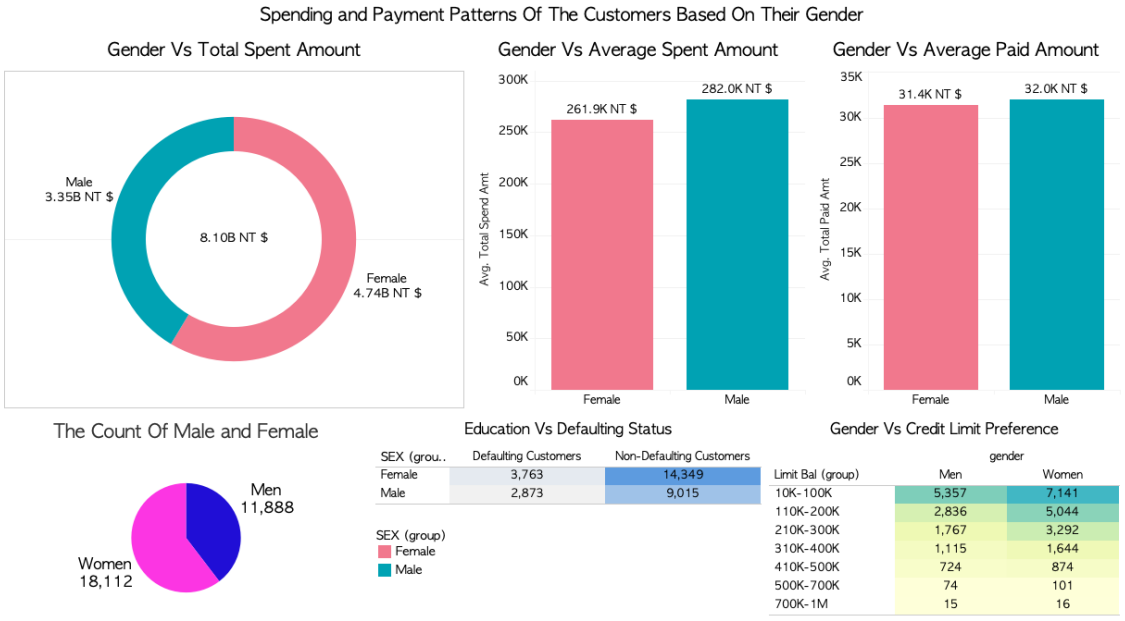
3.5 Spending Patterns



- There are a total of 468 customers who's education status is unknown. However, the average spent amount by these customers is the highest which is around 342.5K and these are also the customers who paid back the highest amount which is 40.8K
- When it came to the number of defaulters while considering education as a factor again those customers who education are very less likely to default because the are a total of 468 customers who's education status is unknown and out of which the no of defaulter in this group is only 33 which is less than 10% .
- When we consider the group of customers who's highest qualification is high school , University and Graduate School the number of defaulters for these are 25%, 23% and 19 % respectively.

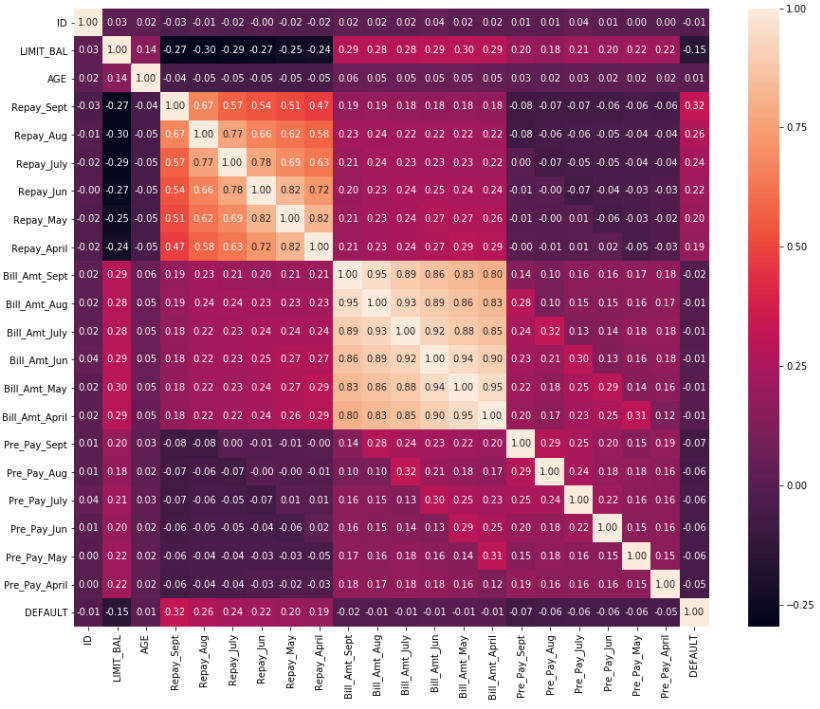


- Customers who are single together as a group spent the most (when sum is taken in to consideration). However, on average married customers spent the most around 280K.
- 30% of the married customer default out of the total 13659 married customers, while only 20% of the single people default out of the total 15964 customers who are single.
- All the three groups paid back almost the same amounts.



- Female customers when taken sum in to consideration they have spent around 4.7 billion which is the higher than males.However the average spent amount is higher for males which is around 282K while the average spent amount for females is around 261K
- 20% of the female customers are defaulters while 24% of the male customers are defaulters

3.6 Correlation



- There is high correlation between the *Bill Amt* variables but this correlation gradually decreases with difference in months.
- Similarly, there is moderate correlation between the *Repay* variables but this correlation also gradually decreases with difference in months.

4 Statistical Analysis

Statistical tests were performed to see the whether each of the 24 independent variables have a significant relationship with the dependent variable, *DEFAULT*

4.1 Chi-square Test

For the Categorical Columns, a Chi-square Test of independence was performed with the target variable, *DEFAULT* which is also a categorical column.

Here Null Hypothesis H_0: There is NO association between the two variables
And Alternate Hypothesis H_a: There is an association between the two variables

Variable	**p-value**	**Decision**
SEX	$4.94e^{-12}$	Reject H_0
EDUCATION	$1.23e^{-32}$	Reject H_0
MARRIAGE	$8.82e^{-8}$	Reject H_0

For all three variables above The null hypothesis have been rejected which means that they all have a significant relationship with the target variable.

For the *Repay* variables there are 12 categories so instead we will use the simpler version of the columns where ≤ 0 (Duly paid) or > 0 (Payment Delay)

Variable	**p-value**	**Decision**
Repay_Sept	0.0	Reject H_0
Repay_Aug	0.0	Reject H_0
Repay_July	0.0	Reject H_0
Repay_Jun	0.0	Reject H_0
Repay_May	0.0	Reject H_0
Repay_April	0.0	Reject H_0

For all six variables above The null hypothesis have been rejected with p-value = 0.0 (extremely low; almost 0) which means that they all have an extremely significant relationship with the target variable.

4.2 Two-sample t test

For all the numeric variables, A two-sample unpaired t tests was performed between values of the variable for two classes of target variables to compare their means.

Here Null Hypothesis H_0: The means of the two samples are EQUAL
And Alternate Hypothesis H_a: The means of the two samples are NOT EQUAL.

If the means of the two samples are significantly different form each other, then we can conclude that the variable does have a significant relationship with the target variable.

The preliminary Normality Tests (*Shapiro*) and Equality of Variance test (*Levene* and *Bartlett*) were done to determine whether to do a parametric test (*ttest_ind*) or a non-parametric test (*Mannwhitneyu*).

Variable	**p-value**	**Decision**
LIMIT_BAL	$6.12e^{-190}$	Reject H_0
AGE	0.186	Failed to Reject H_0
Bill_Amt_Sept	$5.75e^{-6}$	Reject H_0
Bill_Amt_Aug	0.003	Reject H_0
Bill_Amt_July	0.014	Reject H_0
Bill_Amt_Jun	0.074	Failed to Reject H_0
Bill_Amt_May	0.118	Failed to Reject H_0
Bill_Amt_April	0.494	Failed to Reject H_0
Pre_Pay_Sept	$2.31e^{-170}$	Reject H_0
Pre_Pay_Aug	$4.98e^{-151}$	Reject H_0
Pre_Pay_July	$4.49e^{-129}$	Reject H_0
Pre_Pay_June	$3.64e^{-109}$	Reject H_0
Pre_Pay_May	$5.62e^{-91}$	Reject H_0
Pre_Pay_April	$1.59e^{-98}$	Reject H_0

- Only *AGE* and the Bill amounts for the last three months failed to reject the null hypothesis which means that they do not have a significant relationship with the target variable.
- The rest of the numerical variables, *LIMIT_BAL*, Bill amounts for the first three months and all the six Pre Pay amounts rejected the null hypothesis and hence have a significant relationship with the target variable.
- Also, the p-value gradually increases for Bill Amounts from September to May. So, bill amounts become more and more significant the more recent it is.
- Similar gradual increase for the p-values corresponding to Pre_Pay can be observed. But here, the p-values are extremely low indicating that they have an even mroe singificnat effect on *DEFAULT* than the Bill Amounts.

5 Evaluation Metrics

The Evaluation Metrics that can be used for a Binary Classification problem are:

1. **Accuracy** - Proportion of correctly identified instances
2. **Precision** - proportion of positive predictions that are correct
3. **Recall** - Proportion of Actual positives predicted correctly
4. **F1 Score** - Harmonic mean of Precision and Recall
5. **ROC AUC** - Area Under Receiver's Operating Characteristics Curve (trade-off between sensitivity and specificity for different thresholds)
6. **Log Loss** - measures the uncertainty of the probabilities of the model by comparing with the true labels[6]
7. **Cohen's Kappa** - measure of inter rater reliability between two raters (actual and predicted labels)[1]

Due to an imbalance in the dataset, Accuracy is an unreliable measure to evaluate a model. For example, say the test dataset has a 70-30 split of the two classes of target variable. A model which simply predicts the majority class (in our case 0) will get 70% accuracy despite the model not predicting even one defaulter.

Since our main goal is to identify possible defaulters we need to focus on the True Positives. Recall is an excellent measure here since it will give us a measure of proportion of actual defaulters identified correctly.

Precision is still an important measure since it identifies how many of the ones predicted as defaulter are correct. From the credit card company's perspective, it is not desirable to predict too many non-defaulters as defaulters since this could result in losing good customers. In comparison Recall is still more important than precision but we should not compromise too much on precision.

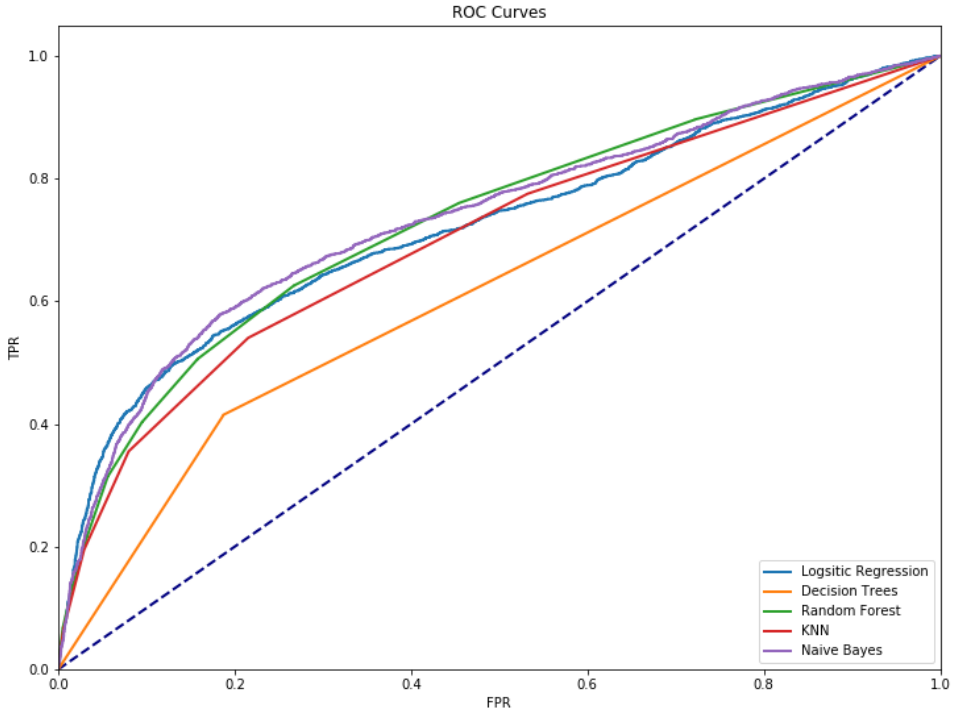
F1 Score is a harmonic mean of precision and recall. High F1 Score is achieved only if both precision and recall are high. This is also an excellent measure to evaluate our models.

ROC Curve plots True Positive Rate (sensitivity / recall) vs False Positive Rate (1 - specificity) for range of threshold values between 0 and 1. The Area Under the Curve can range from 0 to 1 and is a good measure to evaluate how well model is capable of distinguishing between the two classes. This is also a good evaluation metric for our problem statement.

6 Base Models

Before we launch into creating predictive models for the problem, we start with some baselines to give some sense of how the different classification models perform on our dataset. In the due process we have created Logistic Regression, Decision Trees, Random Forest, K Nearest Neighbors, Naïve Bayes.

	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.808	0.708	0.228	0.345	0.721
Decision Trees	0.723	0.349	0.409	0.397	0.611
Random Forest	0.808	0.637	0.320	0.426	0.732
KNN	0.795	0.560	0.356	0.435	0.702
Naive Bayes	0.745	0.446	0.602	0.512	0.736



Even though logistic regression has the highest accuracy and precision scores, its recall is too low comparatively. Since our main aim is to find defaulters logistic regression can't be the best suitable model for the data. Where as the Decision Trees has decent accuracy score but it failed in giving better precision and recall

scores and so on. Amongst all the models below, Naive Bayes has the best recall score with 60.2% and with the largest AUC. From the baseline classifiers that we have, Naive Bayes seems to be the best suitable for the data.

7 Feature Engineering

To help build better models, we created some new features from the existing ones.

7.1 Monthly Dues

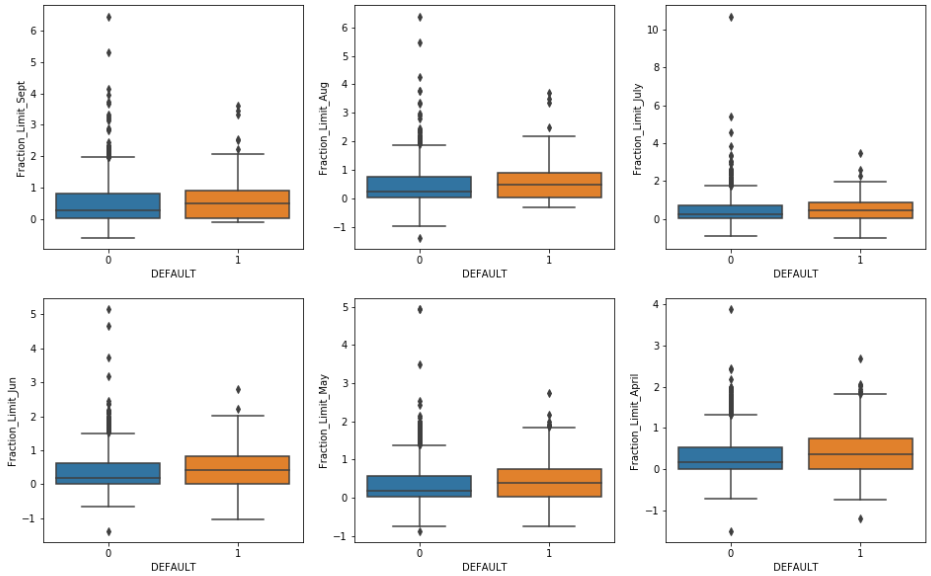
Instead of having Bill Amount and Pre Pay amount in separate columns, why not have a variable that measures how much payment is due from each month. Since we have five corresponding Bill amount and pre pay months, we will get five new variables:

1. **Due_Sept** = Bill_Amt_Aug - Pre_Pay_Sept
2. **Due_Aug** = Bill_Amt_July - Pre_Pay_Aug
3. **Due_July** = Bill_Amt_June - Pre_Pay_July
4. **Due_June** = Bill_Amt_May - Pre_Pay_June
5. **Due_May** = Bill_Amt_April - Pre_Pay_May

7.2 Fraction of Limit Spent

Comparing the Bill amounts of two different customers with different credit card limits will not tell us much about the possibility of defaulting. But looking at how much of the limit has been spent in a month could give us some new information:

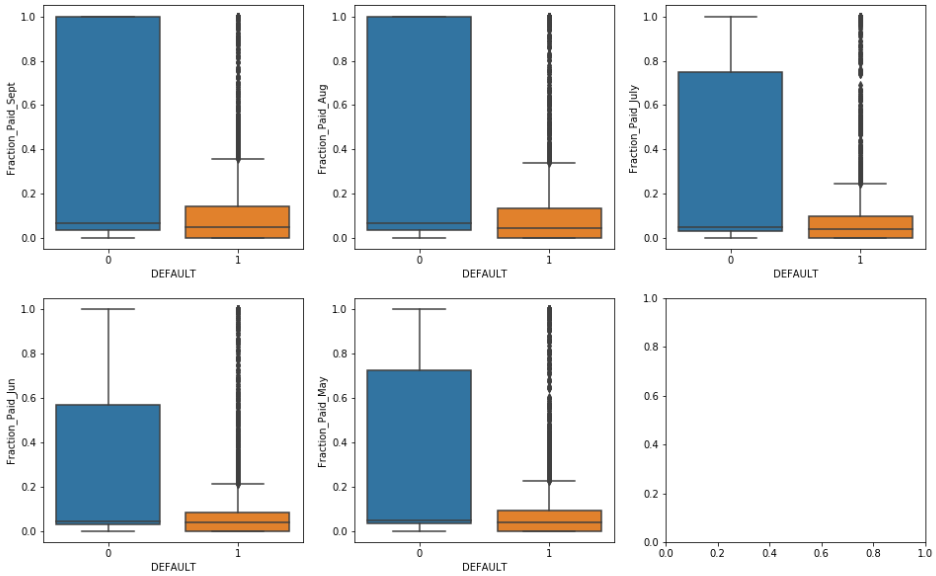
1. **Fraction_Limit_Sept** = Bill_Amt_Sept / LIMIT_BAL
2. **Fraction_Limit_Aug** = Bill_Amt_Aug / LIMIT_BAL
3. **Fraction_Limit_July** = Bill_Amt_July / LIMIT_BAL
4. **Fraction_Limit_June** = Bill_Amt_June / LIMIT_BAL
5. **Fraction_Limit_May** = Bill_Amt_May / LIMIT_BAL
6. **Fraction_Limit_April** = Bill_Amt_April / LIMIT_BAL



7.3 Fraction Paid per month

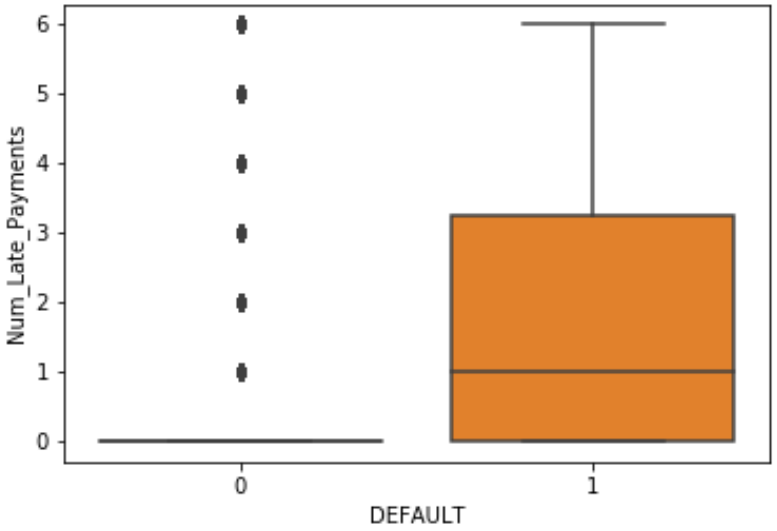
Again, comparing payment amount of two different customers with different bill amounts may not give much information but the fraction of the bill amount paid could be crucial in determining possibility of defaulting

1. **Fraction_Paid_Sept** = Pre_Pay_Sept / Bill_Amt_Aug
2. **Fraction_Paid_Aug** = Pre_Pay_Aug / Bill_Amt_July
3. **Fraction_Paid_July** = Pre_Pay_July / Bill_Amt_June
4. **Fraction_Paid_June** = Pre_Pay_June / Bill_Amt_May
5. **Fraction_Paid_May** = Pre_Pay_May / Bill_Amt_April



7.4 Number of Months of Late Payment

Combining all the six Repay status columns to give us how many months of the six there was a late payment would be very useful.



7.5 Statistical Analysis on new features

Performing the same hypothesis tests on the new features:

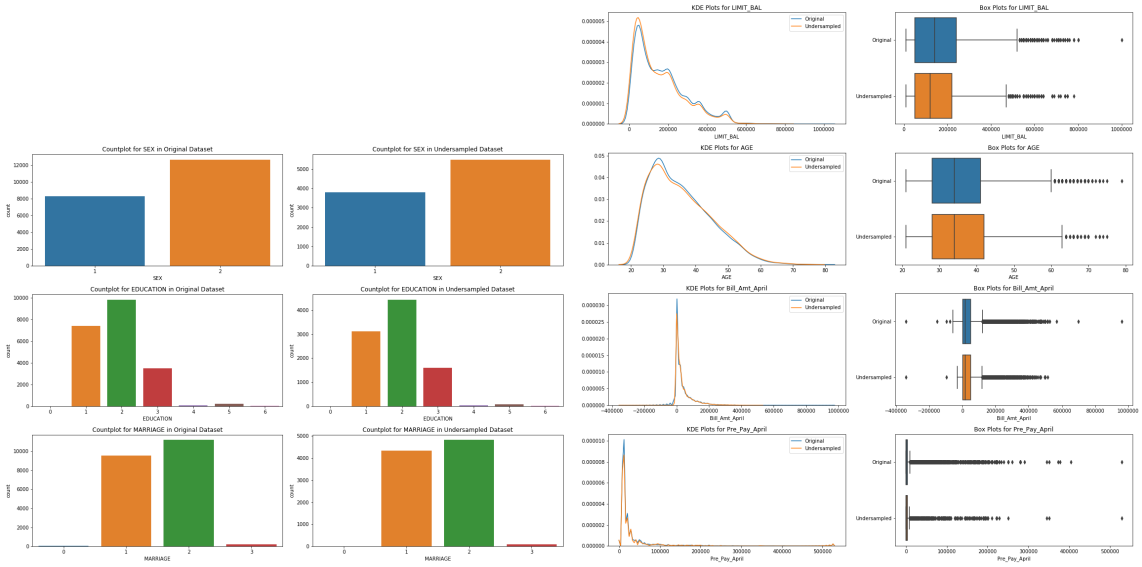
Variable	p-value	Decision
Due_Sept	$9.65e^{-}14$	Reject H_0
Due_Aug	$1.88e^{-}16$	Reject H_0
Due_July	$8.28e^{-}16$	Reject H_0
Due_June	$2.68e^{-}14$	Reject H_0
Due_May	$1.28e^{-}16$	Reject H_0
Fraction_Limit_Sept	$6.42e^{-}37$	Reject H_0
Fraction_Limit_Aug	$1.23e^{-}48$	Reject H_0
Fraction_Limit_July	$9.13e^{-}53$	Reject H_0
Fraction_Limit_June	$1.39e^{-}60$	Reject H_0
Fraction_Limit_May	$1.76e^{-}58$	Reject H_0
Fraction_Limit_April	$6.81e^{-}61$	Reject H_0
Fraction_Paid_Sept	$1.05e^{-}132$	Reject H_0
Fraction_Paid_Aug	$1.25e^{-}119$	Reject H_0
Fraction_Paid_July	$8.90e^{-}112$	Reject H_0
Fraction_Paid_June	$3.44e^{-}92$	Reject H_0
Fraction_Paid_May	$1.60e^{-}85$	Reject H_0
Num_Late_Payments	0.0	Reject H_0

8 Undersampling

One of the ways to deal with an imbalanced dataset is to undersample the majority class i.e choose a random sample of majority class that is of the same size as minority class so that we end up with a balanced dataset.

But, undersampling may lead to losing information. So it is important to look at the distribution of the independent variables before and after undersampling. If the distribution is the same then we can conclude that the undersampled data exhaustively covers all the data in the original dataset. If distirbution is significantly less then we have lost some information and undersampling is not a good idea.

After a stratified train-test split, undersampling was performed on the train data and distribution of each variables was compared. For categorical, we used countplots and for numeric variables, KDE plots.



Although, in the figure above only one Bill Amount and one Pre Pay column are shown, the distribution of all the numerical columns were compared.

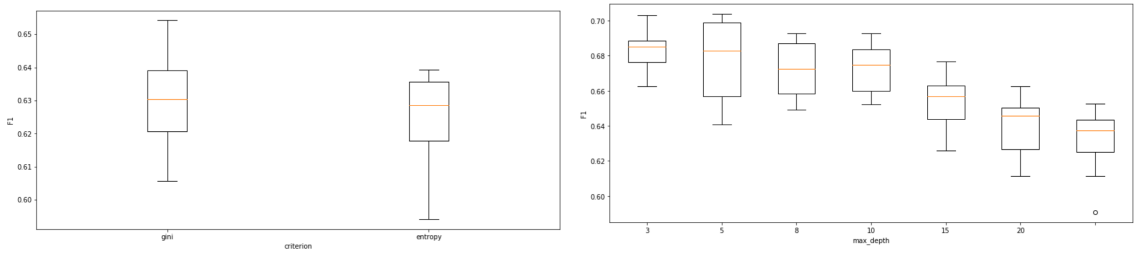
The distribution of all the independent features were similiar before and after undersampling. Although some extreme values are missing in the boxplots for the numerical variables. The major part of the distribution is the same for both.

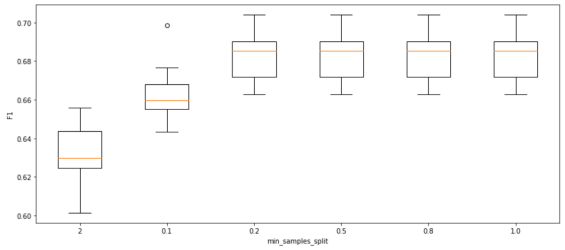
This indicates that undersampling could help build better models.

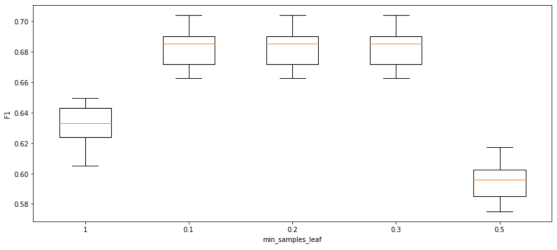
9 Parameter Tuning

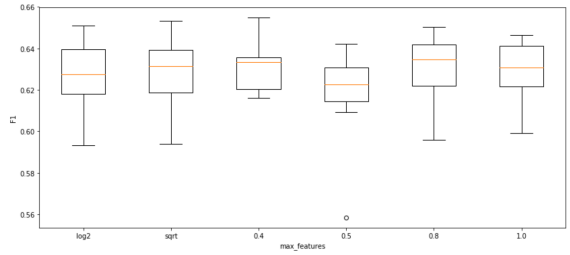
There are multiple hyper parameters for each classification model. We looked at the trend of the F1 score for each of this parameters individually. Then, applied *GridSearch* with the best values for each of them.

9.1 Decision Trees







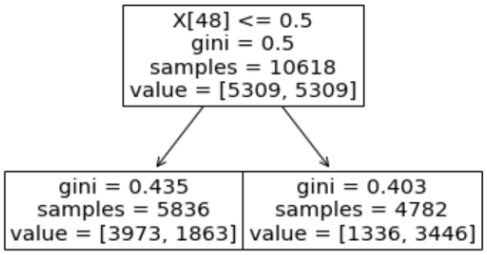


1. **Criterion** : Both Gini and Entropy give approximately the same median F1 score but the overall distribution is higher for Gini
2. **Max Depth** : F1 scores are decreasing with higher max depth. The best scores are for max depth = 3 to 10
3. **Min Samples Split** : F1 scores increase with higher min samples split but it plateaus at 0.2
4. **Min Samples Leaf** : F1 score is best in the range of 0.1 to 0.3
5. **Max Features** : There are no clear winners here but 0.4-0.5 is giving a smaller distributions t=so they are more reliable.

Using this, a Grid Search was done with the best values for each hyper parameter. The best parameters were found as :

1. criterion = 'gini'
2. max_depth = 3
3. min_samples_leaf = 0.3
4. min_samples_split = 0.8
5. max_features = 0.5

The following is a plot of the best Decision Tree found; the only feature that is considered in this tree is *Num_Late_Payments*:



9.2 Random Forests

The best parameters found after GridSearch were:

1. n_esitmators = 50
2. criterion = 'gini'
3. max_depth = 5
4. min_samples_leaf = 1
5. min_samples_split = 2
6. max_features = 'log2'

9.3 Extra Trees

The best parameters found after GridSearch were:

1. n_esitmators = 50
2. criterion = 'entropy'
3. max_depth = 15
4. min_samples_leaf = 1
5. min_samples_split = 2
6. max_features = 'sqrt'

10 Results

10.1 After Undersampling

	Accuracy	Precision	Recall	F1 Score	Cohen Kappa
Decision Trees	0.732	0.430	0.645	0.515	0.341
Random Forest	0.764	0.475	0.619	0.538	0.383
Extra Trees	0.765	0.477	0.614	0.537	0.383

10.2 After Oversampling with SMOTE Borderline

	Accuracy	Precision	Recall	F1 Score	Cohen Kappa
Logistic Regression	0.783	0.371	0.007	0.013	0.006
KNN	0.643	0.345	0.731	0.469	0.248
Naive Bayes	0.431	0.254	0.843	0.390	0.087
Gradient Boosting	0.783	0.497	0.557	0.525	0.385
XG Boost	0.783	0.254	0.843	0.525	0.385
Voting Classifier	0.633	0.341	0.754	0.470	0.246

The best F1 scores are for ensemble tree-based models like Random Forests and Extra Trees with around 0.54. Advanced Boosting techniques like Gradient Boosting and XG Boost are also giving close to that with 0.525.

For Random Forest, which has the best F1 Score, the following is the most important features that were found for this Random Forest in descending order:

1. Num_Late_Payments
2. Repay_Sept
3. Repay_Aug
4. Repay_July
5. Repay_Jun
6. Pre_Pay_Sum
7. Repay_May

8. Repay_April
9. Pre_Pay_Jun
10. LIMIT_BAL

11 Conclusions

Some valuable insights from EDA done earlier:

- Higher the educational qualifications, lower the default rate
- Overall Single people spend the most, but on an average married people spend more
- Married people more likely to default
- On an average, males spend more and have higher default rate

From the important features above for Random Forest, we can conclude that the important features that determine defaulting of a customer relate to the repayment status and payment amounts.

References

[1] Cohen's kappa statistic. https://www.statisticshowto.datasciencecentral.com/cohens-kappa-statistic/.

[2] Default of credit card clients data set. https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.

[3] Documentation for decisiontreeclassifier. https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html.

[4] Documentation for extratreesclassifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html.

[5] Documentation for randomforestclassifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[6] Log loss. http://wiki.fast.ai/index.php/Log_Loss.

[7] Andrew Bloomenthal. Credit card definition. https://www.investopedia.com/terms/c/creditcard.asp.

[8] Eric Wang. Taiwan's credit card crisis. https://sevenpillarsinstitute.org/case-studies/taiwans-credit-card-crisis/.