CREDIT CARD DEFAULT

PREDICTION

LOW LEVEL DESIGN

Lokendar Tirunahari

Objective

Credit cards are big source of income when deftly managed. But errors in risk assessment can leads to stains on the balance sheet. The climbing delinquencies will result in significant amount of money loss from the lending institutions, such as those commercial banks. Therefore, it is very crucial for banks to have a risk prediction model and be able to classify the most relative characteristics that are indicative of people who have higher probability to default on credit card loans. A robust model is not only a useful tool for the lending institutions to make decision on credit card applications, but it can also help the clients to be aware of the behaviors that may damage their credit scores.

This model will help financial institutions, commercial banks and other loan institutes to predict a defaulter earlier.

Problem statement

Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen. In this way, one of the biggest threats faces by commercial banks is the risk prediction of credit clients.

The goal is to predict the probability of credit default based on credit card owner's payment history,credit limit,age etc

#### Data set Information

The data set for training and testing contains the information of customers default payment. There are 30,000 different instances and 25 attributes total; each instance represents one customer, and attributes consist demographic information about the customers and their past payment history from April to September.

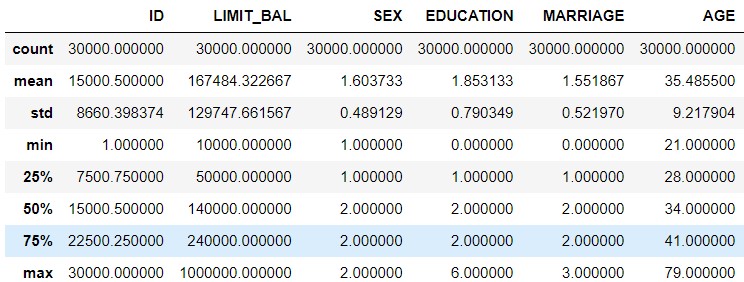
#### Attribute information

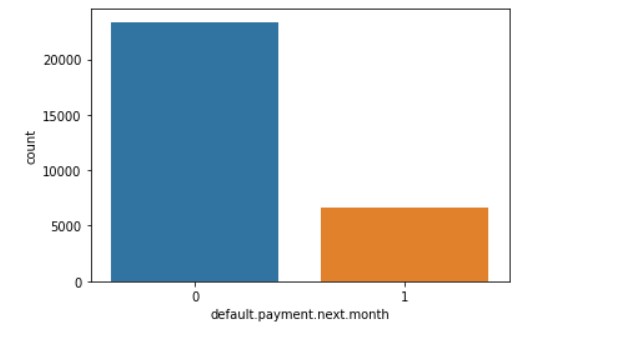
* ID : unique identification number assigned to each customer
* LIMIT BAL : amount of given credit access line
* SEX : gender (1 = male; 2 = female)
* EDUCATION : highest degree obtained (1 = graduate school; 2 = university; 3 = a high school; 4 = others; 5 = unknown; 6 = unknown)
* MARRIAGE : marital status (1 = married; 2 = single; 3 = others)
* AGE : age in year
* PAY 0 : monthly payment record in September
* PAY 2 : monthly payment record in August

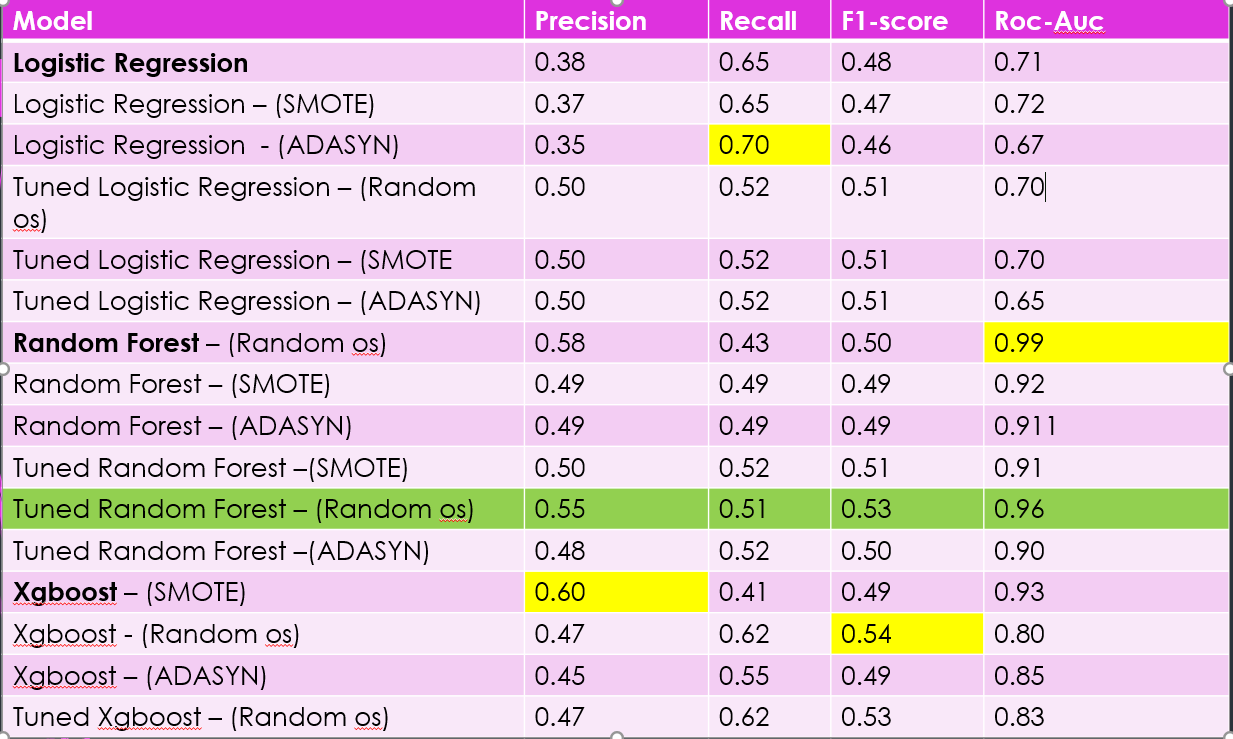
#### PAY 3 : monthly payment record in July

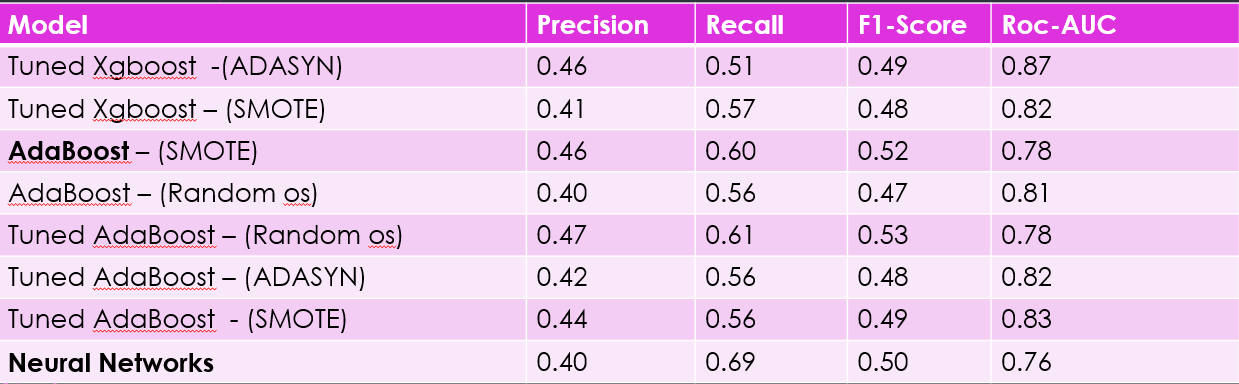
* PAY 4 : monthly payment record in June
* PAY 5 : monthly payment record in May
* PAY 6 : monthly payment record in April
* BILL AMT1 : total amount owed in September
* BILL AMT2 : total amount owed in August
* BILL AMT3 : total amount owed in July
* BILL AMT4 : total amount owed in June
* BILL AMT5 : total amount owed in May
* BILL AMT6 : total amount owed in April
* PAY AMT1 : amount of previous payment in September
* PAY AMT2 : amount of previous payment in August
* PAY AMT3 : amount of previous payment in July
* PAY AMT4 : amount of previous payment in June
* PAY AMT5 : amount of previous payment in May
* PAY AMT6 : amount of previous payment in April

Exploratory Analysis :



* From Table we can see that the average amount of given credit access line is 167,484 with an extremely high standard derivation close to 129,747, which can be explained by the large number of maximum credit card limit of 1 million.
* The average education level is 1.853, which indicates that most customers obtained bachelor or master degrees.
* More customers are either married or singled, compared to the number of unknown. The median age is 35.4 with 9.2 standard deviation. The youngest customer is 21 years old, while the oldest is 79 years old.
* 
* 6,636 customers (approximate 22% of the entire data set) will default in the next months payment, while 23,364 will not default.





* Though main objective is to find out the defaulter which basically means we need high recall for our model it is desirable to have a high precision since the bank wouldn’t want to lose customers who were denied a loan based on the model’s prediction that they would be defaulter.
* Keeping all the above things in mind random forest is suggested.
* Random Forest has the best precision and recall **balance**.
* Higher recall can be achieved if low precision is acceptable.
* This model serves as an aid to human decision.