Credit Card approval prediction by Hesham Ali

Content :

- Data set
- Data Handling & Feature engineering
- Relations between features and target
- Handling imbalanced data
- Scaling and normalization
- ML training and Evaluation
- Results

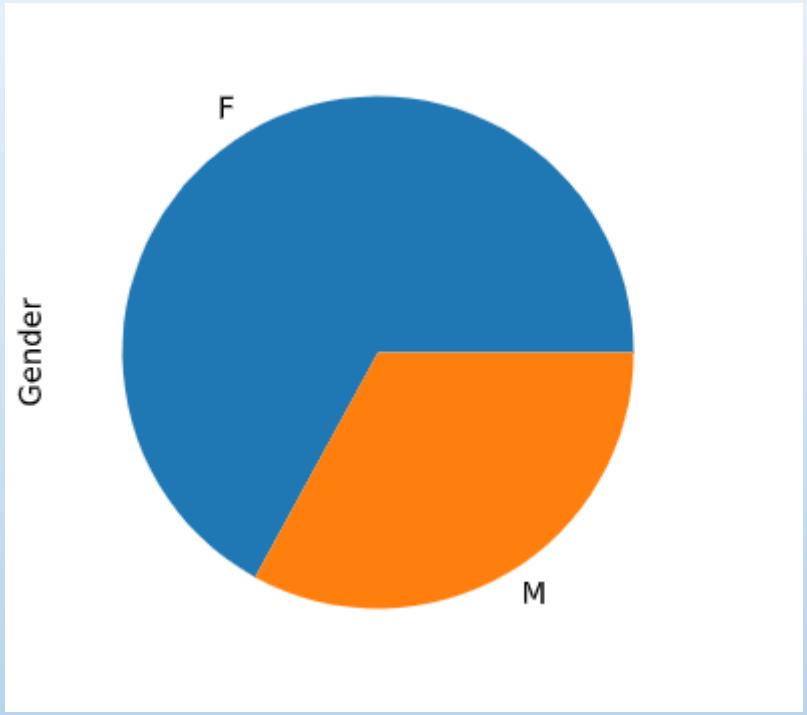
Credit score cards are a common risk control method in the financial industry. It uses personal information and data submitted by credit card applicants to predict the probability of future defaults and credit card borrowings. The bank is able to decide whether to issue a credit card to the applicant. Credit scores can objectively quantify the magnitude of risk.
We have two data sets the first one is for users data and second for credits operations data and we will merge both based on ID

Data set link :
https://www.kaggle.com/rikdifos/credit-card-approval-prediction

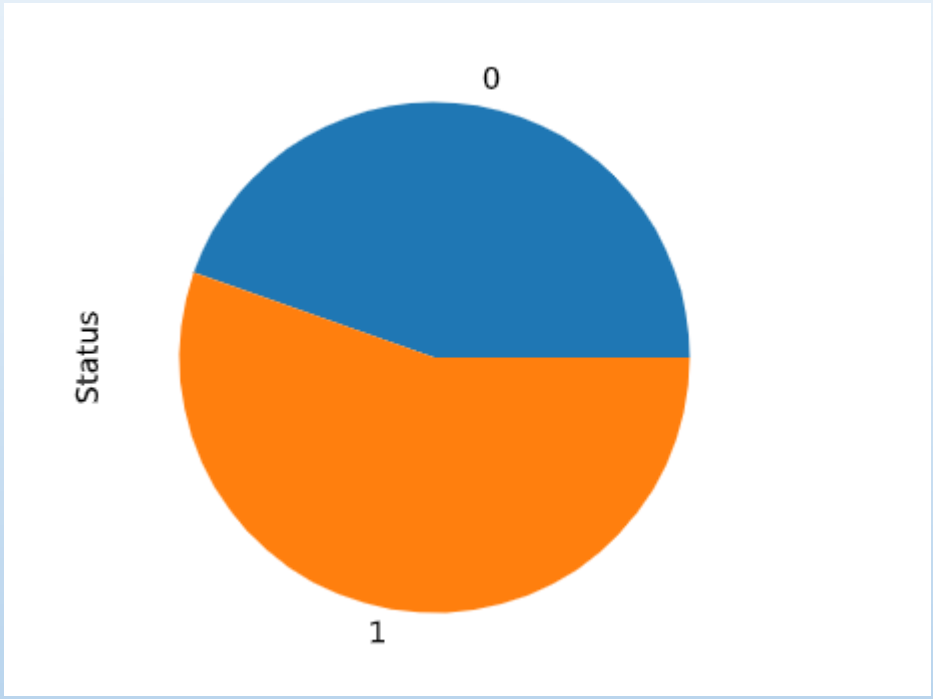
Data Handling & Feature engineering :

- Merge both datasets
- The user will classified good or bad based on the majority of his records in the 2nd dataset
- Handling missing data (remove records with missing data)
- encode categorical features with label encoder
- Modify the format of age and years_of_experience to positive

EDA



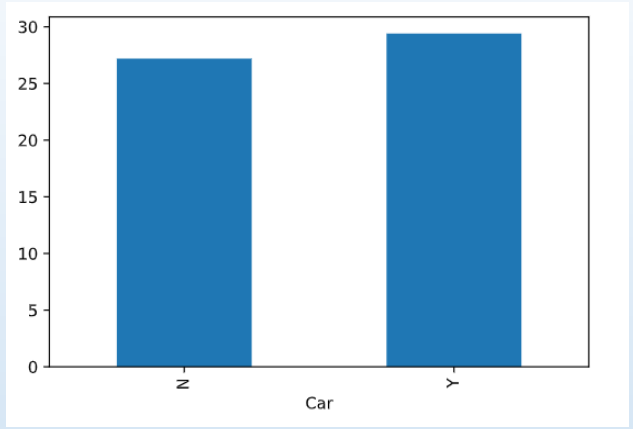
MALE VS FEMALE

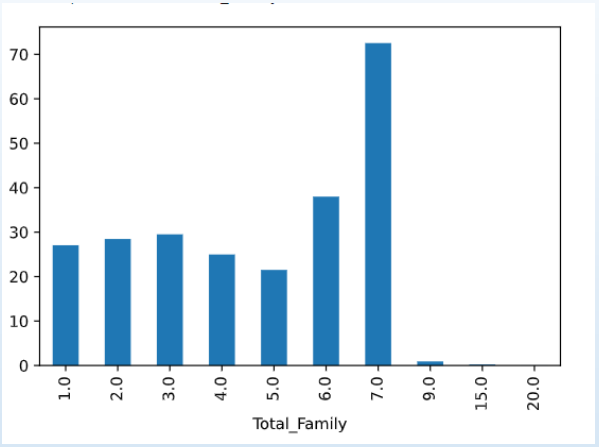


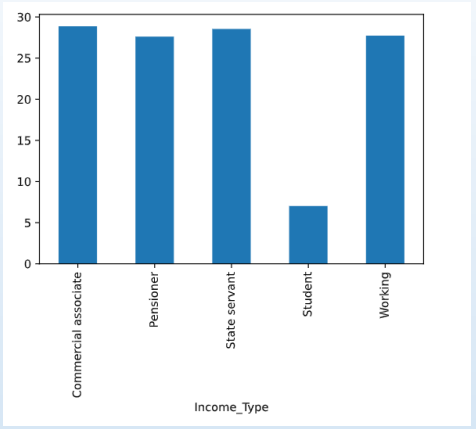
GOOD (1) VS BAD(0)

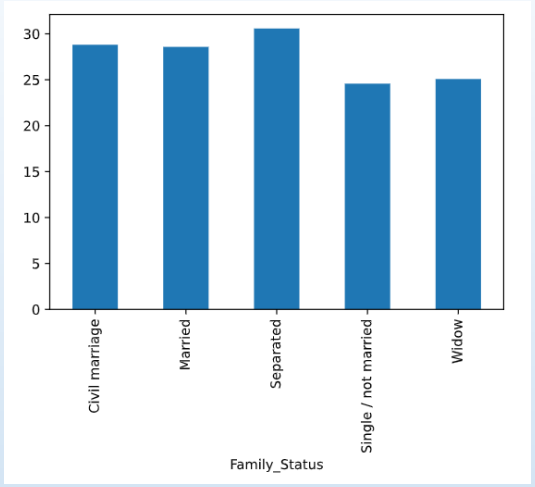
Insights :

- Female is the majority gender in our data
- Data is slightly imbalanced between two classes





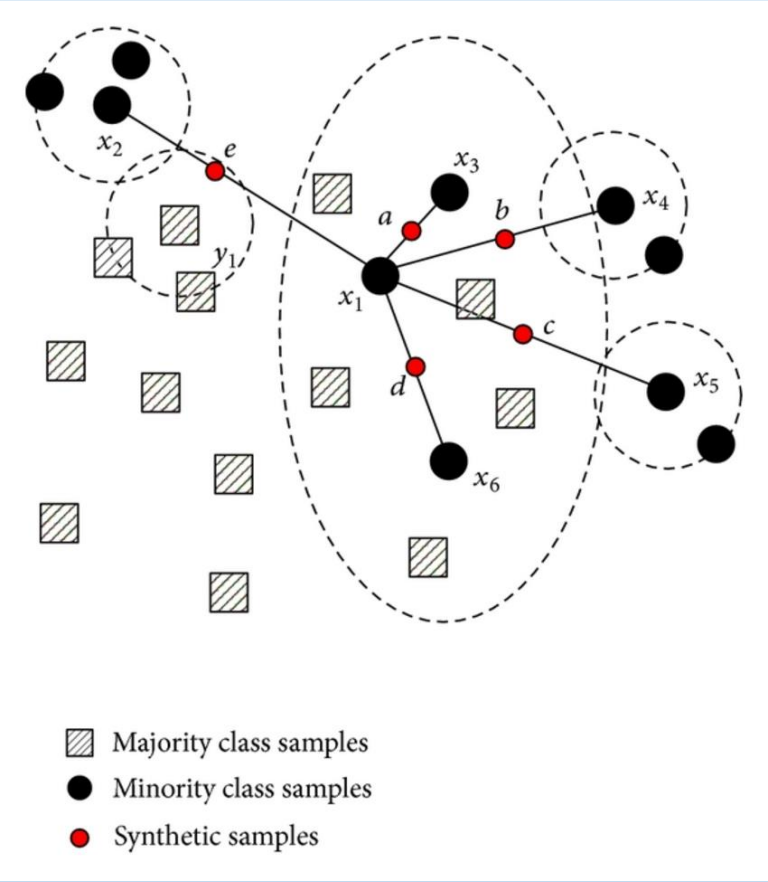


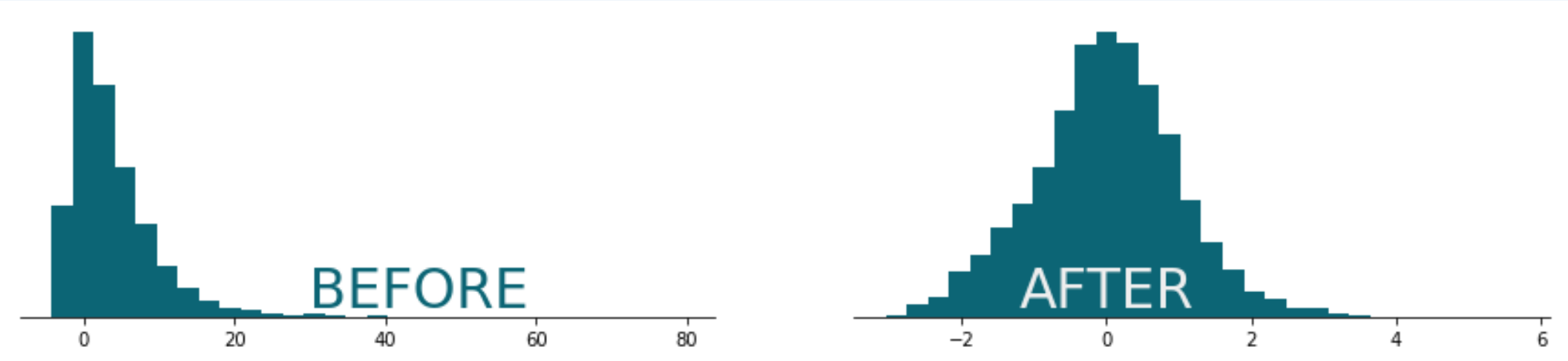


By comparing 4 features with (GOOD/BAD) ratio we can find :

- User who own cars have better history
- Users with 6 & 7 family members have good rates !!
- Students have worst rates
- Married person rate is better than single

Handling imbalanced data using ADASYN Method

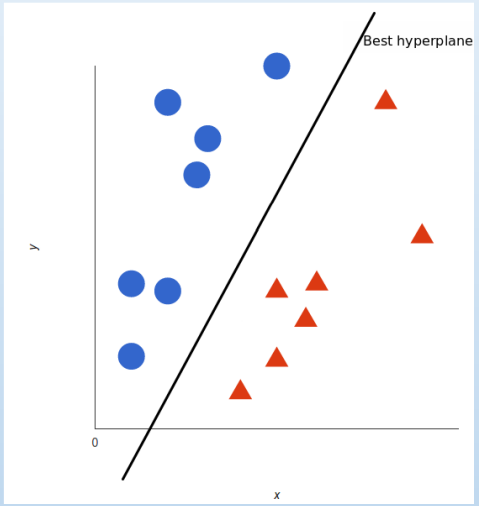


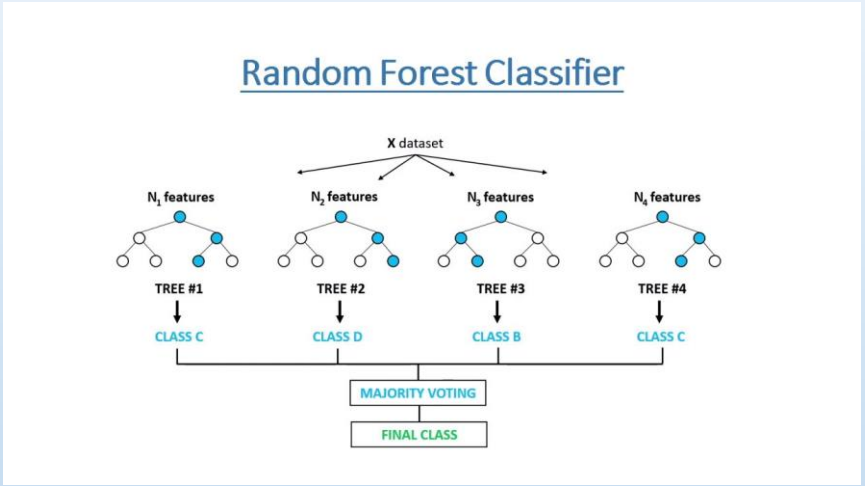


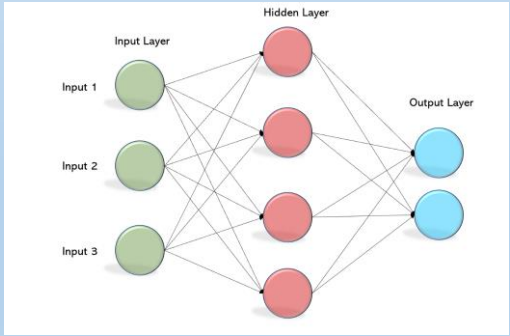
Data normalization :

- Scaling using standard scaler
- Transforming using power transformer

ML Models


Best hyperplane
y
0
x


Random Forest Classifier
X dataset
N1 features
N2 features
N3 features
N4 features
TREE #1
TREE #2
TREE #3
TREE #4
CLASS C
CLASS D
CLASS B
CLASS C
MAJORITY VOTING
FINAL CLASS


Hidden Layer
Input Layer
Input 1
Input 2
Input 3
Output Layer

Results

	ACCURACY	F1-score
Random forest	96.2%	96.3%
SVM	98.3%	98.4%
MLP	99.7%	99.7%
XGB	96.2%	96.3%