Index of the project

1. **Problem statement**

   As a top focus for acquiring new credit card customers, the credit card department uses data driven credit assessment to evaluate the credit score of a customer.

   Based on the credit score the credit card approval for a customer is done. This credit score is determined by a model known as an application scorecard.

   Thus, customers are asked to fill an application (physically or online) to apply for credit card and based on this application data in addition to the Credit Bureau Score (e.g. CIBIL SCORE) and some other internal information of the applicant, their credit worthiness is evaluated, thereby approving or disapproving their credit card issue for the same.

2. **Data sourcing**

   The source of data for this project is Odin school

3. **Data description**

- Features name: (Credit_Card.csv)
- Ind_ID: Client ID
- Gender: Gender information
- Car_owner: Having car or not
- Propert_owner: Having property or not
- Children: Count of children
- Annual_income: Annual income
- Type_Income: Income type
- Education: Education level
- Marital_status: Marital_status
- Housing_type: Living style
- Birthday_count: Use backward count from current day (0), -1 means yesterday.
- Employed_days: Start date of employment. Use backward count from current day (0). Positive value means, individual is currently unemployed.
- Mobile_phone: Any mobile phone
- Work_phone: Any work phone
- phone: Any phone number
- EMAIL_ID: Any email ID
- Type_Occupation: Occupation
- Family_Members: Family size

Another data set (Credit_card_label.csv) contains two key pieces of information

- ID: The joining key between application data and credit status data, same is Ind_ID
- Label: 0 is application approved and 1 is application rejected.

4. **Initial Hypothesis**
   - A person who has a larger family will inevitably need more resources, which will result in higher expenses. Thus, these individuals are more likely than those with one or two family members to apply for credit cards on a regular basis.

- Those with significant annual incomes have a high possibility of receiving credit cards. Considering that they are less likely to miss credit payments than people with low income.
- In general, work experience grows with age, and persons with more years of experience are more likely to earn more and have good credit.
- Those with a decent educational background and a steady source of income have a better chance of getting credit cards.

5. **Exploratory data analysis**
    a. Basic data exploration
    b. Missing value treatment
    c. Outlier treatment
    d. Univariate
    e. Bivariate
6. **Insights from EDA**
7. **Data Pre- processing**
    a. Selecting independent and dependant variables
    b. Encoding categorical data
    c. Splitting data into train and test
    d. Feature scaling

8. **Model training**
    a. Smote resampling
    b. Logistic regression
    c. Decision tree
    d. K neighbour's classifier
    e. Support vector machine
    f. Random forest classifier

Highest accuracy achieved by Random Forest Classifier - 91%

9. **K Fold Cross Validation**

From the cross validation and individual testing of different models, we can see, Random forest and Decision tree default are good models for this type of imbalanced data set with smote technique. Best model is random forest with 87% pass rate.

As data is highly imbalanced, we have to use the SMOTE sampling technique, without the technique, we are getting a higher model pass rate, but that is completely biased towards the size of prediction 0 or rejection of a credit card with greater number of type 2 errors.