

## **Project Report:** Credit Card Approval Prediction using Machine Learning

### **Introduction:**

Credit card approval is a crucial process in the banking and financial industry that requires a careful evaluation of various factors, such as an individual's financial history, creditworthiness, and other relevant features. To make informed decisions regarding credit card approval, banks and other financial institutions can use machine learning models to predict an applicant's credit card approval status accurately. In this project, we aim to build a machine learning model that can predict credit card approval status based on various features of the applicants.

### **Dataset Description:**

The dataset used in this project is provided by Kaggle and contains information on credit card applications from a bank. The dataset comprises 400,000 observations with 20 explanatory variables, including numerical, categorical, and binary variables. The target variable is a binary variable that indicates whether the credit card application was approved or not. The dataset is split into two files – application\_record.csv and credit\_record.csv.

### **Project Objective:**

The primary objective of this project is to build a machine learning model that can accurately predict the credit card approval status of an applicant based on various features such as income, education, age, and credit history.

To achieve this objective, we followed the following steps:

1. **Data Preprocessing:** We first examined the dataset for missing values, outliers, and other issues affecting the model's accuracy. We then imputed missing values using appropriate methods such as mode imputation. We also encoded categorical variables and performed feature scaling to prepare the data for modeling.
2. **Exploratory Data Analysis:** We performed various statistical analyses and visualizations to understand the relationships between the variables and identify any patterns and trends in the data. We examined the distribution of the target variable (credit card approval status) and looked for any correlations between the target variable and the explanatory variables. We also created scatter plots, box plots, and histograms to visualize the distribution and relationships between the variables.
3. **Feature Selection:** We used feature selection techniques such as correlation analysis, mutual information, and recursive feature elimination to identify the most critical variables that contribute to the prediction of credit card approval. We then selected a subset of the most important variables for use in the classification models.
4. **Model Selection:** We experimented with various classification models such as logistic regression, decision tree, random forest, support vector machine, Gaussian Naïve Bayes, Voting, Neural Network and XG boost to select the best model for predicting the credit card approval status. We trained the models on the training dataset and evaluated their performance using cross-validation. We also tuned the hyperparameters of each model to optimize its performance.

5. **Model Evaluation:** We evaluated the performance of the model using various metrics such as accuracy, precision, recall, and area under the receiver operating characteristic (ROC) curve. We also created visualizations such as confusion matrices to identify any areas for improvement in the model.

### **Results:**

After completing the project, we found that the Logistic Classifier, Random Forest, Support Vector Machine, Voting, XG Boost and Neural Network outperformed other models in predicting the credit card approval status, with an accuracy of 98%. We identified that some of the most critical variables for predicting credit card approval include the applicant's age, experience, and family status.

### **Conclusion:**

In conclusion, this project aimed to build a machine learning model that could accurately predict credit card approval status based on various features of the applicants. We achieved this objective by following the steps mentioned above, and we found that the Logistic Classifier, Random Forest, Support Vector Machine, Voting, XG Boost and Neural Network outperformed other models in predicting credit card approval status. The results of this project can be used by banks and other financial institutions to make informed decisions regarding credit card approval and minimize the risks associated with credit card fraud and default. The project can also be extended