**Project Documentation for Credit Card Approval Project**

**Credit Card Approval Project**

**Introduction**

This project aims to address the challenge of predicting credit card approval using machine learning techniques. Accurately predicting credit card approval is crucial for financial institutions to manage risk and improve customer satisfaction. By leveraging a dataset that includes various applicant attributes, we aim to build a robust model that can automate the approval process while maintaining fairness and accuracy."

**Files in the Project**

credit\_risk\_dataset.csv: The raw dataset used for training and testing the models.

credit\_risk.csv: The cleaned and preprocessed dataset.

README.md or documentation.txt: The documentation file explaining the project.

**Steps**

**Import Libraries:**

* Import essential libraries for data manipulation, visualization, model building, and evaluation.

**Load the Dataset**:

* Load the dataset using pd.read\_csv and display the first few rows to understand its structure.

**Dataset Overview**:

* Check the size and basic information of the dataset to identify data types and missing values.

**Handle Missing Values**:

* Replace missing values in 'person\_emp\_length' with 0.
* Replace missing values in 'loan\_int\_rate' with the median value.

**Encoding Categorical Values**:

* Convert categorical variables into numeric values using dictionaries for 'person\_home\_ownership', 'loan\_intent', and 'cb\_person\_default\_on\_file'.

**Correlation Matrix**:

* Generate and visualize a correlation matrix to understand the relationships between features.

**Save and Reload Transformed Dataset**:

* Save the cleaned and transformed dataset to a CSV file and reload it for further processing.

**Feature and Target Variables**:

* Define the feature matrix X and target vector y.

**Train-Test Split**:

* Split the data into training and testing sets using train\_test\_split.

**Data Standardization**:

* Standardize the feature variables to have a mean of 0 and standard deviation of 1.

**Initialize Classifiers**:

* Initialize various classifiers including Logistic Regression, Decision Tree, K-Nearest Neighbors, XGBoost, and Random Forest.

**Voting Classifier**:

* Create a Voting Classifier that combines the predictions of the individual classifiers.

**Model Training and Evaluation**:

* Train the Voting Classifier and evaluate its performance.
* Identify and evaluate the best-performing individual classifier.

**Handle Class Imbalance**:

* Check for class imbalance in the target variable.
* Use SMOTE to generate synthetic samples for the minority class.

**Train-Test Split for Resampled Data**:

* Split the resampled data into training and testing sets.

**Standardization and Model Training on Resampled Data**:

* Standardize the resampled data and train the Voting Classifier.
* Evaluate the model's performance on the resampled data.

**Cross-Validation**:

* Perform cross-validation to assess the model's robustness and generalizability.

**Conclusion**

In conclusion, our project successfully built and evaluated a machine learning model for credit card approval prediction. The Voting Classifier emerged as the best performer, achieving an accuracy of 91%. Through cross-validation and handling class imbalance using SMOTE, we ensured the model's robustness and fairness in predicting both approval and denial cases. Future work could focus on exploring additional features or advanced modeling techniques to further enhance predictive accuracy.