**Name: -** Vatsal Rameshbhai Gohel

**Mail id: -** [**gohel.vatsal11@gmail.com**](mailto:gohel.vatsal11@gmail.com)

**Contact Detail: -** +1-774-253-3960

**LinkedIn Id: -** <https://www.linkedin.com/in/gohelvatsal/>

**GitHub Id: -** <https://github.com/vatsalgohel>

**Project Portfolio**

**Loan Eligibility Prediction and Alternative Loan Amount Estimation**

**Project Title:**

Comprehensive Loan Decision Framework for Financial Institutions

**Problem Statement:**

The challenge is to develop a predictive model that can accurately determine loan eligibility and provide alternative loan amounts for non-eligible applicants, thereby streamlining the loan application process for banks and financial institutions.

**Project Description:**

**Objective:**

* Develop a predictive model to determine loan eligibility based on the applicant's personal and financial background.
* Identify the most important factors that influence loan approval.
* Estimate the maximum loan amount for applicants who do not meet the initial loan criteria.

**Scope:**

* Use historical loan application data to train and test the models.
* Implement logistic regression and random forest classifiers for prediction.
* Optimize the models for better accuracy using grid search cross-validation.
* Provide a user-friendly framework for bank employees to assess loan applications and offer tailored solutions.

**Deliverables:**

* A predictive model with high accuracy for loan eligibility.
* A framework for estimating alternative loan amounts for non-eligible applicants.
* Documentation detailing the methodology, results, and impact of the models.

**Methodology:**

**Tools and Technologies Used:**

* Python (for data preprocessing, model building, and evaluation)
* Scikit-learn (for machine learning algorithms and grid search cross-validation)
* Pandas and NumPy (for data manipulation and analysis)
* Matplotlib and Seaborn (for data visualization)

**Process:**

* **Data Preprocessing:**
  + Handle missing values, perform one-hot encoding for categorical variables, and standardize the data.
  + Create a new feature "TotalIncome" by adding ApplicantIncome and CoapplicantIncome.
  + Drop irrelevant columns such as Loan\_ID.
* **Model Building and Evaluation:**
  + Split the dataset into training and testing sets.
  + Train logistic regression and random forest classifier models.
  + Evaluate models using accuracy, precision, recall, and F1-score metrics.
  + Optimize the random forest model using grid search cross-validation.
* **Model Tuning:**
  + Perform grid search cross-validation to find the best parameters for the random forest model.
  + Implement the final model with optimized parameters and evaluate its performance.
* **Prediction and Estimation:**
  + Use the final model to predict loan eligibility on the test dataset.
  + Estimate the maximum loan amount for non-eligible applicants using a logistic regression model.

**Roles and Responsibilities:**

* **Coordinator:** Oversaw the entire project, from task assignment to submission. Managed data preprocessing, model building, evaluation, and documentation.
* **Team Members:** Conducted exploratory data analysis, feature engineering, model training, and result validation.

**Results and Impact:**

**Outcome:**

* The logistic regression model achieved an accuracy of 81.17% for predicting loan eligibility.
* The random forest classifier, optimized with grid search cross-validation, achieved an accuracy of 81.6%.
* Identified credit history, total income, and loan amount term as the most important factors affecting loan approval.

**Impact:**

* The framework allows bank employees to efficiently assess loan applications and offer tailored loan amounts for non-eligible applicants.
* Provides customers with a transparent way to check their loan eligibility and improve their chances of approval.

**Challenges and Solutions:**

**Challenges Faced:**

* Handling missing values and standardizing the data.
* Ensuring the model's interpretability and transparency.
* Optimizing model parameters for better accuracy.

**Solutions Implemented:**

* Imputed missing values and standardized the data using appropriate techniques.
* Selected logistic regression and random forest models for their balance between accuracy and interpretability.
* Used grid search cross-validation to optimize model parameters.

**Conclusion and Learnings:**

**Summary:**

This project developed a comprehensive loan decision framework that accurately predicts loan eligibility and estimates alternative loan amounts for non-eligible applicants. The framework streamlines the loan application process for both bank employees and customers, enhancing decision-making and fostering better customer relationships.

**Learning Experience:**

* Gained experience in data preprocessing, feature engineering, and model optimization.
* Improved understanding of machine learning algorithms and their application in financial decision-making.
* Enhanced project management skills through coordinating a team project and ensuring timely delivery of results.

**Link to the File:**

<https://drive.google.com/drive/folders/1hmxvvausgAvsbINb6drwDd0kMssaIEJd?usp=sharing>