### **Project Outline: Predicting Consumer Personal Loan Eligibility Based on Credit Scores**

#### **1. Project Ideation**

* **Objective**: Create a machine learning algorithm that predicts consumer personal loan eligibility based on credit scores.
* **Industry Focus**: Finance
* **Key Questions**:
  + How accurately can credit scores predict personal loan eligibility?
  + What other factors significantly influence the prediction?
* **Possible Data Source**: Kaggle, UC Irvine Machine Learning Repository, or Data.gov for credit score and loan data.

#### **2. Data Acquisition**

* **Identify and Source Data**:
  + Locate datasets with at least 1,000 records that include credit scores, loan application outcomes, and possibly additional financial indicators (income, debt-to-income ratio, etc.).
* **Document Data Sources**: Ensure datasets comply with fair use and are properly licensed for the intended use.

#### **3. Data Exploration and Cleanup**

* **Data Exploration**:
  + Conduct an initial analysis to understand the distribution and characteristics of the data.
  + Identify and handle missing or inconsistent data.
* **Data Cleanup**:
  + Remove duplicates, normalize variables, and handle missing values.
  + Perform feature engineering, including creating new features or modifying existing ones to improve model accuracy.

#### **4. Model Implementation**

* **Model Selection**:
  + Choose a supervised learning model (e.g., Decision Tree, Random Forest, Logistic Regression) suitable for classification.
* **Training the Model**:
  + Split data into training and testing sets.
  + Train the model using Scikit-learn.
* **Evaluation**:
  + Evaluate the model using metrics like accuracy, precision, recall, F1 score, and ROC-AUC.
  + Use Matplotlib to visualize performance metrics and feature importance.
  + Document model performance and optimization steps.

#### **5. Model Optimization**

* **Iterative Improvements**:
  + Experiment with different hyperparameters to improve model performance.
  + Test different feature combinations and scaling techniques.
  + Record all iterations and results in a CSV or within the Python script.
* **Final Evaluation**:
  + Print and visualize the final model performance.
  + Ensure the model achieves at least 75% classification accuracy.

#### **6. API Integration and Data Visualization**

* **API Requests**:
  + Use APIs to fetch additional data if necessary (e.g., real-time credit score updates).
* **Visualization**:
  + Create plots to visualize loan eligibility distribution, feature importance, and model accuracy.
  + Consider using Pandas plotting and Matplotlib for clear visualizations.

#### **7. Documentation and GitHub Setup**

* **Jupyter Notebook**:
  + Document the data extraction, cleanup, transformation, and analysis processes in a well-organized Jupyter notebook.
* **Python Script**:
  + Include the script that initializes, trains, and evaluates the model.
* **GitHub Repository**:
  + Clean up the repository by removing unnecessary files.
  + Use a .gitignore file appropriately.
  + Create a polished README file that presents the project overview, objectives, and findings.

#### **8. Presentation Preparation**

* **Executive Summary**:
  + Prepare an overview of the project goals and their relevance to the finance industry.
* **Data and Model Overview**:
  + Describe the data collection, cleanup, and exploration process.
  + Present the model training and evaluation approach.
* **Results and Conclusions**:
  + Summarize the model’s performance and its implications.
  + Discuss potential future developments and next steps.
* **Slide Preparation**:
  + Design visually appealing slides that clearly communicate key points.
  + Rehearse the presentation as a group to ensure smooth delivery.

#### **9. Internal Milestones**

* **Week 1**:
  + Complete project ideation, data fetching, exploration, and cleanup.
  + Begin initial model training and evaluation.
* **Week 2**:
  + Focus on model optimization, documentation, and presentation preparation.
  + Finalize and rehearse the presentation.
  + Submit GitHub repository links by the end of the second week.

### **Project Proposal: Predicting Consumer Personal Loan Eligibility Based on Credit Scores**

#### **Project Title: Predicting Consumer Personal Loan Eligibility Using Machine Learning**

#### **Team Members: Brandon, Ravi, Rasesh**

#### **Project Description / Outline of the Problem:**

The goal of our project is to develop a machine learning (ML) model that predicts a consumer's eligibility for a personal loan based on their credit score and other financial indicators. In the finance industry, accurately assessing loan eligibility is crucial for both financial institutions and consumers. Traditional credit scoring models often fail to capture the full picture of a borrower’s financial health, leading to suboptimal lending decisions.

Our project aims to address this issue by creating a more nuanced prediction model that considers multiple factors beyond just the credit score, thereby improving the accuracy of loan eligibility predictions.

#### **Why is ML Uniquely Qualified to Solve This Problem? What Approach Will You Use?:**

Machine learning is uniquely qualified to solve this problem due to its ability to analyze large datasets and detect complex, non-linear patterns that traditional statistical models might miss. By training our model on historical loan application data, we can identify key factors that influence loan eligibility and use this information to make more accurate predictions.

Our approach will involve using supervised learning techniques to develop a classification model that predicts loan eligibility. We will explore several algorithms, including Decision Trees, Random Forest, and Logistic Regression, to determine the most effective model for our dataset.

#### **Dataset(s) and Algorithm(s) to Be Used:**

* **Dataset**: We plan to use publicly available datasets from sources such as Kaggle or the UC Irvine Machine Learning Repository, which include credit scores, loan application outcomes, income, debt-to-income ratios, and other relevant financial indicators.
* **Algorithms**:
  + **Decision Tree**: For initial model creation and feature importance analysis.
  + **Random Forest**: For improving accuracy and handling overfitting.
  + **Logistic Regression**: As a baseline model to compare against more complex models.

#### **Rough Breakdown of Tasks with Owners:**

1. **Project Ideation and Planning**:
   * **Tasks**: Define project goals, create the proposal, and outline project structure.
2. **Data Acquisition and Exploration**:
   * **Tasks**: Identify and source the dataset(s), perform initial data exploration, and handle any missing or inconsistent data.
3. **Data Cleanup and Feature Engineering**:
   * **Tasks**: Clean and preprocess data, engineer new features, and prepare the dataset for model training.
4. **Model Selection and Initial Training**:
   * **Tasks**: Select appropriate algorithms, train the initial models, and evaluate their performance.
5. **Model Optimization and Finalization**:
   * **Tasks**: Optimize model parameters, test different feature combinations, and finalize the model.
6. **API Integration and Data Visualization**:
   * **Tasks**: Implement API requests for additional data (if necessary), visualize key metrics, and create graphs to support the findings.
7. **Documentation and GitHub Setup**:
   * **Tasks**: Document the project process, clean up the repository, and create a detailed README file.
8. **Presentation Preparation**:
   * **Tasks**: Prepare slides, rehearse the presentation, and finalize the executive summary.