

The document titled "Feature Forge" outlines the process of implementing feature engineering techniques and evaluating machine learning models using the "Loan Prediction" dataset from Analytics Vidhya. Here's a concise summary:

## **Project Overview**

Participants are tasked with enhancing machine learning models through feature engineering and evaluation. The project involves refining the dataset, applying machine learning techniques, and documenting the results in a Jupyter notebook.

## **Key Steps**

### **1. Data Preparation:**

- Load the "Loan Prediction" dataset.
- Perform exploratory data analysis (EDA) to understand data structure and address missing values using imputation techniques.

### **2. Feature Engineering:**

- Encode categorical variables using methods like one-hot or label encoding.
- Create or transform features (e.g., income-to-loan ratio, family size).
- Select the most relevant features using statistical tests and model-based importance.

### **3. Model Development:**

- Split the dataset into training and testing subsets.
- Build a baseline model (e.g., logistic regression) for initial performance comparison.
- Evaluate models using metrics such as accuracy, precision, recall, and F1 score.

### **4. Model Optimization:**

- Refine the model through hyperparameter tuning or by exploring advanced algorithms like Random Forest and XGBoost.
- Compare model performance and analyze the impact of feature engineering steps.

### **5. Documentation:**

- Compile the findings, including feature engineering methods, model results, and insights, into a final report.

The document encourages thoughtful exploration and systematic improvement of models, emphasizing the importance of documenting both the process and the results. Let me know if you'd like assistance with any specific aspect of the project!