**Loan Default Prediction Report**

# Objective

The objective of this project was to build a classification model using financial data to predict whether a loan applicant is likely to default. The dataset used was the Lending Club Loan Dataset which contains extensive financial, demographic, and loan-related attributes.

# Date Preprocessing

Several preprocessing steps were taken to prepare the data:  
- Missing values were handled by imputing medians for numeric columns and modes for categorical columns.  
- The 'loan\_status' column was used as the target, and mapped to binary labels (1 for Default, 0 for Not Default).  
- Categorical columns were encoded using Label Encoding.  
- SMOTE (Synthetic Minority Over-sampling Technique) was applied to handle class imbalance.

# Model Training

We used the LightGBM classifier, which is well-suited for large datasets with many features and handles categorical features efficiently. After splitting the data into training and testing sets, we trained the model using the oversampled data.

# Model Evaluation

The model was evaluated using Precision, Recall, F1 Score, and Confusion Matrix. The results showed balanced performance, especially in identifying high-risk (default) applicants.

# Challenges Faced

- Several features had missing values, which required careful imputation.  
- The dataset was highly imbalanced with fewer default cases.  
- Feature selection and encoding required attention to maintain model interpretability and performance.

# Solutions Implemented

- Handled missing data using column-wise imputation.  
- Used SMOTE to generate synthetic examples for the minority class.  
- Employed LightGBM for efficient model training.  
- Visualized the Confusion Matrix to interpret model predictions.  
- Suppressed future warnings to improve notebook readability.

# Recommendations for Lenders

- Use the model to flag high-risk loan applicants before disbursal.  
- Periodically retrain the model with recent data to maintain accuracy.  
- Include additional behavioral or external credit scoring metrics to improve performance.