**Credit Risk Analysis Report**

# 1. Objective

The objective of this task was to build a machine learning model to assess the creditworthiness of customers and flag high-risk individuals for financial institutions. The model should help reduce default rates by predicting customers who are more likely to default on loans.

# 2. Dataset

We used the 'Give Me Some Credit' dataset, which includes customer demographics, income, debt history, loan attributes, and a target column indicating loan default status. Key features include:  
- Age, Income, Home Ownership, Employment Length  
- Loan Intent, Grade, Amount, Interest Rate  
- Credit History, Default on File, Percent of Income

# 3. Preprocessing and Feature Engineering

We handled missing values and converted categorical features using One-Hot Encoding. The target variable was converted to integer format for compatibility with classification algorithms. To address class imbalance in the dataset, we used SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic samples for the minority class.

# 4. Model Training

We trained a Gradient Boosting Classifier as it performs well on imbalanced data and is robust to overfitting. The dataset was split into training and testing sets using an 80-20 split. The model was evaluated using classification metrics such as precision, recall, F1-score, and ROC-AUC score.

# 5. Evaluation Metrics

The confusion matrix, classification report, and ROC curve were used for evaluation. These showed good model performance in distinguishing between defaulters and non-defaulters, with a satisfactory ROC-AUC score.

# 6. Challenges and Resolutions

- Problem: The SMOTE method threw an error due to incompatible target format.  
 Solution: We ensured the target variable was cast to integer type.  
- Problem: Class imbalance led to biased predictions.  
 Solution: We applied SMOTE to balance the dataset.  
- Problem: Need for interpretability and visualization.  
 Solution: We plotted a confusion matrix and ROC curve to assess model performance.

# 7. Outcome

We successfully built a credit risk assessment system that flags high-risk customers. The model is saved using `joblib` and can be reused or deployed. Visualizations including confusion matrix and ROC curve assist in understanding model decisions.