Credit Risk Analysis Report

# 1. Project Overview

This project performs credit risk analysis using the 'Give Me Some Credit' dataset. The goal is to predict the likelihood that a customer will experience financial distress in the next two years and to flag high-risk individuals for financial institutions.

# 2. Data Preprocessing

Missing values in the dataset were handled using median imputation. The features 'MonthlyIncome' and 'NumberOfDependents' were particularly affected by missing data.  
Class imbalance in the target variable was addressed using the Synthetic Minority Oversampling Technique (SMOTE). Features were standardized using StandardScaler.

# 3. Feature Engineering

Several new features were engineered to enhance model performance:  
- DebtToIncomeRatio: Revolving utilization of unsecured lines divided by monthly income.  
- EstimatedMonthlyDebt: Product of DebtRatio and MonthlyIncome.  
- TotalLatePayments: Combined sum of all late payment counts.

# 4. Model Training

Three models were trained and evaluated:  
- Random Forest Classifier  
- Gradient Boosting Classifier  
- XGBoost Classifier  
Each model was trained on the resampled dataset after SMOTE. Performance was evaluated on a hold-out test set.

# 5. Model Evaluation

The models were evaluated using the following metrics:  
- Confusion Matrix  
- Classification Report (Precision, Recall, F1-Score)  
- AUC-ROC Score  
These metrics help assess the model’s ability to correctly identify high-risk customers.

# 6. Conclusion

The project successfully developed a credit risk analysis system capable of flagging potentially high-risk customers. This model can assist financial institutions in reducing default rates and improving credit decision-making.