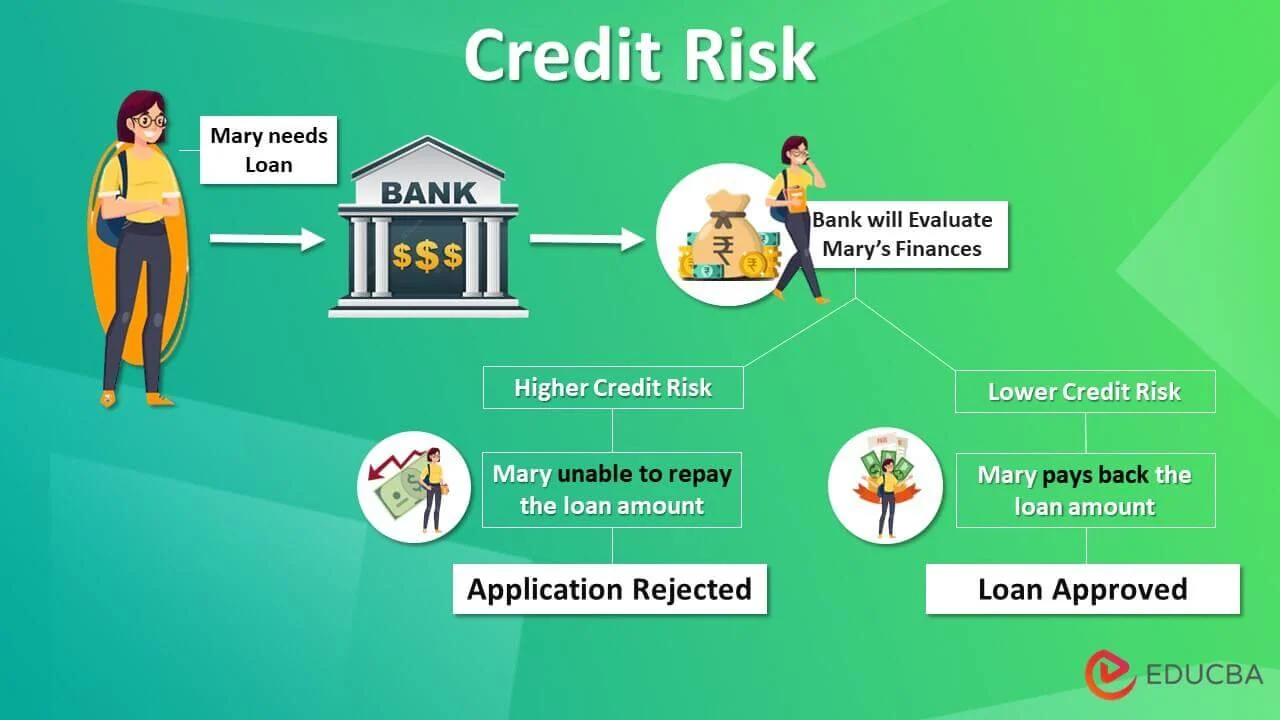
**Credit Risk Analysis**

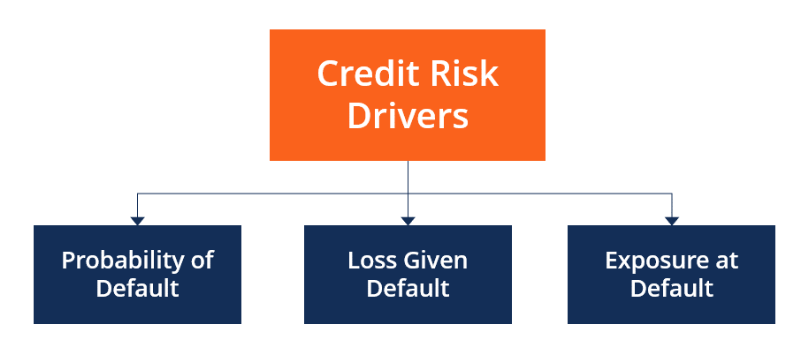
**Overview**

Credit risk analysis is crucial for financial institutions to minimize loan defaults and ensure responsible lending. This project aims to develop a predictive model that assesses the creditworthiness of customers using historical financial data. By leveraging machine learning techniques, the system will identify high-risk individuals, helping institutions make informed lending decisions and mitigate financial risks.



**Objectives**

* Preprocess the **"Give Me Some Credit"** dataset, handling missing values and class imbalances using techniques like **SMOTE**.
* Engineer meaningful features related to **income, debt levels, and repayment history** to improve model accuracy.
* Train machine learning models such as **Random Forest, Gradient Boosting, and XGBoost** to classify high-risk customers.
* Evaluate model performance using metrics like **AUC-ROC, precision-recall, and F1-score** to ensure reliability.
* Develop a system that accurately flags potential defaulters, assisting financial institutions in making data-driven credit decisions.



* **Dataset:** use from Kaggle (Give Me Some Credit Dataset)

****

**Dataset Link :** [**https://www.kaggle.com/c/GiveMeSomeCredit**](https://www.kaggle.com/c/GiveMeSomeCredit)

**Models:**

**Supervised Learning Models:**

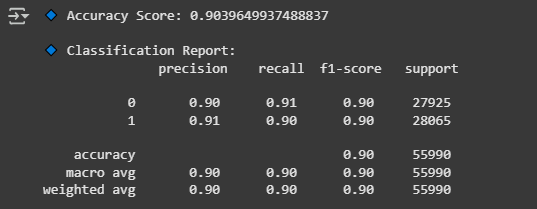
1. Random Forest – A robust ensemble learning method that reduces overfitting and improves accuracy in credit risk classification.
2. Gradient Boosting Machines (GBM) – A boosting algorithm that enhances weak learners iteratively to improve predictions.
3. XGBoost – An optimized gradient boosting model known for high accuracy and efficiency in handling imbalanced financial data.

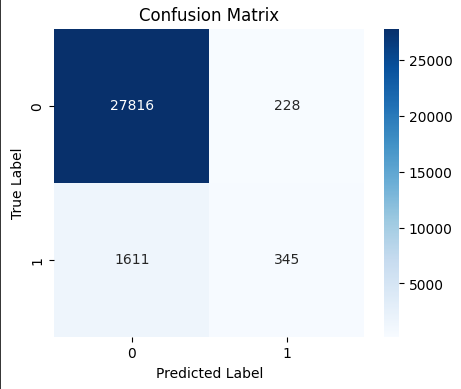
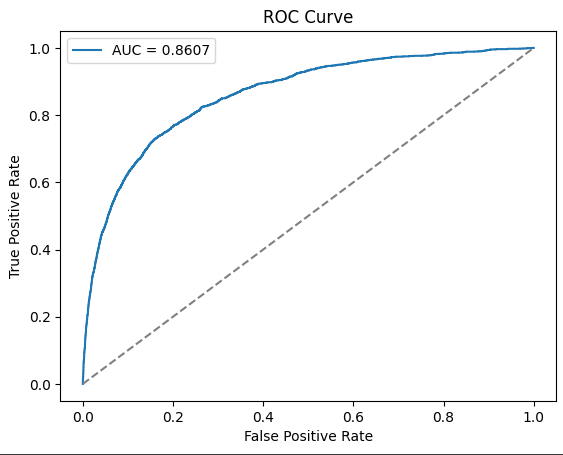
Data Preprocessing Techniques:

* SMOTE (Synthetic Minority Over-sampling Technique) – Used to balance the dataset by generating synthetic examples of the minority class.

These models are well-suited for binary classification (creditworthy vs. high-risk customers) and provide strong predictive capabilities for financial risk assessment.

**Accuracy:**

****

****

**Environment:**

Google colab

Framwork: Torch

**Google colab link :**

[**https://colab.research.google.com/drive/18HR2IOwczxs5m144F8BGRYjpWwE77O5O?usp=sharing**](https://colab.research.google.com/drive/18HR2IOwczxs5m144F8BGRYjpWwE77O5O?usp=sharing)