Machine Learning Task Report
Credit Scoring Model

1. Objective

The objective of this task was to predict an individual's creditworthiness using supervised machine learning algorithms based on financial and demographic data.

2. Dataset Overview

The dataset used in this task was sourced from Kaggle and includes the following features:

- **Numerical**: Age, Income, Debt, Loan Amount, Loan Term, Number of Credit Cards, Payment History
- **Categorical**: Gender, Education, Employment Status, Residence Type, Marital Status.
- **Target**: Creditworthiness (binary classification: creditworthy or not)

3. Data Preprocessing

- Missing values were **dropped**.
- Numerical columns were **standardized** using StandardScaler.
- The dataset had an **imbalance** in the target variable (Creditworthiness), which was handled by **random oversampling** of the minority class to balance the training data.

4. Model Training

Three supervised learning models were trained:

- **Logistic Regression**
- **Decision Tree Classifier**
- **Random Forest Classifier**

5. Evaluation Metrics

The models were evaluated using:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**
- **ROC-AUC**

During evaluation, warnings were encountered due to the model failing to predict one of the classes (e.g., predicting only "not creditworthy"). This was resolved by:

- Applying zero_division=0 in classification_report()
- Adjusting prediction thresholds based on probability outputs.

6. Conclusion

This project successfully demonstrated how machine learning techniques can be applied to predict an individual's creditworthiness using financial and demographic data. The initial dataset suffered from class imbalance, which led to misleading evaluation metrics. By applying proper data preprocessing, feature scaling, and oversampling techniques, the model performance was significantly improved.

Among the models tested, Random Forest and Logistic Regression with class weighting showed the best balance between precision and recall. This highlights the importance of handling imbalanced data thoughtfully in credit risk prediction tasks.

Overall, the system developed in this task provides a reliable framework that could assist financial institutions in making informed lending decisions, minimizing risk, and improving loan approval processes.