Credit Risk Modeling Using Machine Learning

Rituj Upadhyay (2020570), Hardik Sachdeva (2022193), Mananya Kohli (2022275), Akshat (2022056)

I. MOTIVATION

In today's rapidly evolving financial landscape, credit risk assessment is more crucial than ever. With increasing levels of consumer debt and economic uncertainty due to global events like the COVID-19 pandemic, traditional credit risk models are struggling to adapt to new dynamics.ML offers a solution by providing more flexible and accurate models that can quickly respond to changes in borrower behavior.

A. Current Situation

The financial sector is facing unprecedented challenges due to economic fluctuations, increasing non-performing loans (NPLs), and a rise in defaults. Financial institutions are under pressure to improve their credit risk management practices to mitigate potential losses and maintain profitability.

B. Impact of Machine Learning

ML models can analyze vast amounts of data in real-time, uncovering patterns and correlations that traditional models might miss. By integrating ML, financial institutions can not only predict defaults more accurately but also offer more personalized lending products, reduce the risk of financial crises, and enhance overall financial stability.

II. RELATED WORK

A. Predicting Credit Risk for Unsecured Lending

The paper can be found at this link.

B. Machine Learning-Driven Credit Risk: A Systemic Review

The review is available here.

III. TIMELINE

A. Phase 1 (Weeks 1-2): Data Collection and Preparation

- Collect historical credit data from publicly available sources or financial institutions.
- Clean and preprocess the data, including handling missing values and outliers.

B. Phase 2 (Weeks 3-4): Exploratory Data Analysis (EDA)

- Perform EDA to identify patterns and correlations.
- Visualize key features and their relationship with the target variable.

C. Phase 3 (Weeks 5-6): Model Development

- Develop and train various machine learning models (e.g., Logistic Regression, Random Forest, Gradient Boosting).
- Evaluate models using cross-validation and select the best-performing one.

D. Phase 4 (Weeks 7-8): Model Interpretation and Validation

- **Model Validation on Test Data**: Use the test dataset to evaluate the model's performance. This step ensures that the model generalizes well to unseen data.
- **Performance Metrics**: Calculate metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to assess the model's predictive power.
- **Feature Importance**: Identify the most influential features in the model using algorithms like Random Forest or Gradient Boosting.
- **Confusion Matrix**: Analyze the confusion matrix to assess the model's ability to distinguish between different classes.

E. Phase 5 (Weeks 9-10): Deployment and Reporting

- Deploy the model using Flask for real-time credit risk assessment.
- Prepare a detailed report documenting the methodology, results, and potential applications.

IV. INDIVIDUAL TASKS

- **Hardik & Akshat: Data Collection and Preprocessing** - Responsible for sourcing the data, cleaning it, and preparing it for analysis.
- **Akshat & Rituj: Exploratory Data Analysis and Feature Engineering** - Focus on performing EDA, identifying key features, and engineering new features to improve model accuracy.
- **Rituj & Mananya: Model Development and Evaluation** - Develop, train, and evaluate different machine learning models. Select the best model based on performance metrics.
- **Mananya & Hardik: Model Interpretation and Deployment** - Work on interpreting the model using explainability tools and deploy the final model to a web interface.

V. FINAL OUTCOME

The primary expectation from this project is to develop a highly accurate and reliable machine learning model that can predict credit risk effectively. This model should help financial institutions make informed decisions about loan approvals, reducing the likelihood of defaults and improving overall financial stability. Additionally, the project aims to provide actionable insights into the factors that most significantly affect credit risk, helping institutions refine their lending strategies.