**Credit Scoring for Lending Institutions**

**Group Members:**

Aashlesha Shirole

Bhakti Palkar

Dipali Avhad

**Abstract**

Credit scoring for lending institutions is a critical task to determine the credit worthiness of loan applicants, which directly impacts the institution's ability to mitigate risks and make informed lending decisions. Previous research in this field has predominantly focused on the development of predictive models that leverage applicant data, financial history, and loan performance information. Common methodologies employed include logistic regression, decision trees, and ensemble methods such as random forests or gradient boosting. Logistic regression is favored for its simplicity and interpretability, while decision trees and ensemble methods offer enhanced predictive accuracy. These models are trained on historical data, identifying patterns and relationships among various applicant attributes, and subsequently generating credit scores and risk assessments, helping lending institutions make more informed lending decisions while managing their portfolio risk effectively. Ongoing research in the field continues to explore innovative data sources and machine learning techniques to further improve the accuracy and robustness of credit scoring models.

**Problem Statement**

Lending institutions face the challenge of accurately assessing the creditworthiness of borrowers in an ever-evolving financial landscape. This problem statement seeks to address the need for the development and enhancement of credit scoring models.

**Introduction**

Credit scoring plays a pivotal role in the world of lending. It helps lending institutions make informed decisions about who to grant loans to and at what terms. Today, we 'll delve into the fascinating world of credit scoring and see how pattern recognition through machine learning is enhancing lending practices. In the past, most credit scoring managers assess applicants’ credit by visible indicators combined with their own experience. However, patterns of consumers with negative credit cannot always be summarized with explicit features, and judgments by experience result in defaulting of consumers and a loss for financial institutions. Determination of an applicant's credit by classification models results in a higher accuracy than human intuition.

Some traditional statistical approaches are firstly applied in credit scoring. The widely used model is logistic regression (LR). After that, machine learning algorithms are applied in credit scoring. Support vector machine (SVM) is one of the most successful machine learning algorithms applied in the field of credit scoring, along with decision tree (DT), artificial neural network (NN), etc.

**Objective**

Our primary objective in this project is to develop highly accurate credit scoring models for lending institutions.

We achieve this by harnessing the power of machine learning and pattern recognition techniques.

**Data Sources**

To build these models, we rely on a trio of critical data sources:

* First, we use applicant data to understand the personal and financial characteristics of loan applicants.
* Second, we examine financial history to gauge past financial behaviors and patterns.
* Lastly, we incorporate loan performance data, which informs us about how well loans have been repaid in the past.

**Methodology**

We employ a mix of modeling techniques to fulfill our objectives. These include Logistic Regression, Decision Trees, and Ensemble Methods.

**Logistic Regression** - This statistical method, which is also known as a logit model, is utilised rather frequently in the process of classification as well as predictive analytics. Logistic regression is a statistical method that determines the chance that an event will take place. Due to the fact that the result is a probability, the range of the dependent variable is between 0 and 1.

**Decision Trees** - Decision trees may be used for classification and regression issues, but in this case, we have just employed the classification method. A decision tree is a prediction model that links an observation of an item with branch representation to an assessment of a target value with leaf representation. It is among the most successful supervised learning strategies.

**Model Training**

To build effective credit scoring models, we train them on historical data. This data helps us identify patterns and relationships that lead to more accurate credit risk assessments.

**References:**

1. Varun S, Abishek Theagarajan, Shobana M, “Credit Score Analysis Using Machine Learning” IEEE 2023.
2. Baichuan Liu, Likun Lu, Qingtao Zeng, Yeli Li, “Implementation of credit scoring card model based on logistic regression and lightgbm”, IEEE 2021.
3. P. M. Addo, D. Guegan, and B. Hassani, “Credit risk analysis using machine and deep learning models,” Risks, vol. 6, no. 2, p. 38, 2018