Project Proposal

Project Title:

Advanced Predictive Analytics for Loan Default Risk Assessment

Team Members:

1. Venkateshwar Reddy Kasturi – A20518749
2. Anshuman Raturi – A20531913

Description of the Problem:

Financial institutions face significant challenges in assessing loan default risks, which can lead to substantial financial losses and affect the stability of the financial system. Traditional models for predicting loan defaults rely heavily on historical data and basic borrower information, which may not fully capture the complexities of modern financial behaviors and economic conditions. There's a pressing need for more advanced analytical models that can accurately predict loan defaulters, incorporating a wider array of data points, including non-traditional variables such as behavioral patterns, and economic indicators.

A Brief Survey of What Has Been Done and How the Proposed Work Is Different:

Several studies and models have been developed to predict loan defaults, using a variety of statistical and machine learning techniques. These models have traditionally focused on borrower-specific information, such as credit scores, income levels, and employment history. Recent advancements have introduced machine learning models, including Decision Trees and Support Vector Machines, to improve prediction accuracy. However, these models often do not incorporate real-time data or non-traditional data sources, which can offer significant insights into borrower behavior and economic conditions.

The proposed work seeks to differentiate itself by integrating machine learning models with a broader dataset, including non-traditional data sources such as economic indicators and behavioral analytics. Additionally, the project proposes to apply deep learning techniques to capture complex patterns and relationships within the data, potentially offering superior predictive performance over traditional models. Overall, it outlines a comprehensive plan to advance the field of loan default prediction using modern data science techniques and a wide array of data sources which aims to significantly improve the accuracy and reliability of loan default predictions, providing valuable insights for financial institutions.

Preliminary Plan (Milestones):

TASK NAME	DUE DATE	DATE_COMPLETE	ASSIGNED TO
Data Gathering			
Kaggle Dataset:	03/01/2024		Venkateshwar Reddy Kasturi
Data Preparation			
Description of data	03/09/2024		Anshuman Raturi
Data Exploration / Visualization	**03/19/2024**		Venkateshwar Reddy Kasturi
Building Model			
Splitting data as training and testing set	03/21/2024		Venkateshwar Reddy Kasturi
Feature selection	03/28/2024		Anshuman Raturi
Residual Plot to Identify Non-Linearity	04/01/2024		Anshuman Raturi
Generate Models	04/05/2024		Venkateshwar Reddy Kasturi / Anshuman Raturi
Testing Hypothesis (F-Test / T-Test)	04/07/2024		Anshuman Raturi
Evaluating the Model			
Run model on test set	04/10/2024		Venkateshwar Reddy Kasturi
K-fold Cross Validation for model fit	04/12/2024		Anshuman Raturi
Model Accuracy Test	04/15/2024		Venkateshwar Reddy Kasturi
Deploy the model and documentation			
Model Conclusions and insights	04/18/2024		Anshuman Raturi / Venkateshwar Reddy Kasturi
Final Report Generation	04/20/2024		Venkateshwar Reddy Kasturi / Anshuman Raturi

Reference (A List of Papers):

- Alonso, A., & Carbo, J. M. (2021). Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3774075#

-Jumaa, M., Saqib, M., & Attar, A. (2023). Improving Credit Risk Assessment through Deep Learning-based Consumer Loan Default Prediction Model. *International Journal of Finance & Banking Studies (2147-4486)*, *12*(1), 85-92. https://www.ssbfnet.com/ojs/index.php/ijfbs/article/view/2579/1822

- Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012042). IOP Publishing. https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012042/meta

- Arora, S., Bindra, S., Singh, S., & Nassa, V. K. (2022). Prediction of credit card defaults through data analysis and machine learning techniques. *Materials Today: Proceedings*, *51*, 110-117. https://www.sciencedirect.com/science/article/abs/pii/S2214785321035148