

## Machine Learning

### Project Proposal: Loan approval prediction using machine learning

#### Group-15

Roll No:20

Name: Poojith Sonti

Student Id: 02080247

Email: [psonti@umassd.edu](mailto:psonti@umassd.edu)

Roll No:23

Name: Bharath kumar Tanniru

Student Id: 02078210

Email: [btanniru@umassd.edu](mailto:btanniru@umassd.edu)

**Objective:** The objective of the project is to develop a machine learning model using Python that can predict whether a loan applicant is eligible for a loan based on key features such as Marital Status, Education, Applicant Income, Credit History, and others. This model aims to simplify and streamline the decision-making process for banks by automating the evaluation of an applicant's profile for loan approval, thereby increasing efficiency and potentially reducing the time and resources spent on manual assessments.

#### Methodology:

- 1. Data Collection:** Determine the sources from which the loan application data will be collected. This could include internal bank records, publicly available datasets, or synthetic data that mimics real-world loan applications. Collect a comprehensive dataset that includes a variety of loan applications, capturing details such as Marital Status, Education, Applicant Income, Credit History, and other relevant features that could influence loan approval decisions.
- 2. Data Preprocessing:** Address missing values, remove duplicates, and correct inconsistencies in the dataset to ensure the quality of the data. Divide the dataset into training and testing sets to facilitate the evaluation of the model's performance on unseen data.
- 3. Model selection:** Evaluate different machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting Machines to identify the most suitable one for predicting loan approval. Use cross-validation techniques to ensure that the model selection process is robust and avoids overfitting to the training data.
- 4. Training and Testing:** Train the selected model on the training dataset, allowing it to learn the patterns and relationships between the input features and the target variable (loan

approval). Assess the model's performance on the testing set to gauge how well it generalizes to new, unseen data.

5. **Evaluation:** Use appropriate evaluation metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve (AUC-ROC) to quantify the model's effectiveness in predicting loan approvals. Implement additional validation techniques, such as confusion matrix analysis or external validation sets, to further ensure the model's reliability and robustness.

**Dataset information:** The dataset provides a comprehensive overview of loan applications, incorporating a diverse spectrum of applicant profiles, including various marital statuses, education levels, income brackets, and employment types. It covers multiple property areas, ranging from urban to rural settings, to ensure a broad representation of loan scenarios. This variety in data attributes, including credit histories and loan amounts, ensures the development of a robust machine learning model capable of accurately predicting loan approval outcomes across a wide array of applicant conditions and requirements.

**Software and tools:** We use python programming language with libraries like NumPy, Pandas, Scikit-learn, Seaborn and Matplotlib for functions like data manipulation, data handling, algorithms for classification, regression and for data visualization. We use jupyter notebook for coding and visualizations.

**Expected outcome:** A predictive model capable of accurately assessing loan applications and determining their approval status. This model aims to streamline the decision-making process for financial institutions by automating the evaluation of critical applicant information such as income, credit history, and property details.