**Group : Saba Alaeddini, Ruprekha Baruah, Vidushi Bundhooa**

**Project 4 Proposal: Loan Eligibility and Interest Rate Prediction System**

### **Introduction**

Financial institutions have to make decisions on loan eligibility daily. This requires a vigorous assessment of the client’s creditworthiness based on several factors such as income, age, debts and other factors. This project aims to develop a machine learning-based algorithm to predict loan eligibility and predicting interest rates.

### **2. Objectives**

* Develop a predictive model based on applicants' financial data to determine loan eligibility.
* Predict appropriate interest rates by assessing risk levels.
* Enhance fairness and transparency in loan approval decisions.

**3. Research Questions**

**Loan Eligibility Prediction**

* **H**ow do person\_age and person\_income influence loan approval (loan\_status)?
* Does home ownership status (person\_home\_ownership) impact the likelihood of loan approval?
* What is the relationship between employment length (person\_emp\_length) and loan approval?
* How does loan\_intent (purpose of the loan) affect approval rates?
* Does a previous default record (cb\_person\_default\_on\_file) significantly impact loan eligibility?
* How does credit history length (cb\_person\_cred\_hist\_length) affect the probability of loan approval?
* Can a model predict loan status more accurately using a combination of income, employment length, and credit history?

**Interest Rate Prediction**

* What factors most influence loan interest rates (loan\_int\_rate)?
* How does the loan amount (loan\_amnt) correlate with the assigned interest rate?
* Does loan grade (loan\_grade) effectively represent risk levels and predict interest rates?
* Can borrowers with higher income (person\_income) secure lower interest rates?
* How does the loan percentage of income (loan\_percent\_income) affect the assigned interest rate?
* Are there significant differences in interest rates for different loan\_intent categories (e.g., education, business, medical, etc.)?
* Can previous defaults (cb\_person\_default\_on\_file) predict higher interest rates?

### **5. Methodology**

#### **5.1 Data Collection**

* **Datasets**: Publicly available loan datasets-Kaggle

**Features**: Age, Income, Employment Length, Loan amount, Loan interest, Loan Status

#### **5.2 Data Preprocessing**

* Handling missing values and outliers.
* Feature selection using statistical analysis and correlation matrices.
* Data normalization and encoding categorical variables.

#### **5.3 Model Development**

* **Loan Eligibility Prediction**: Logistic Regression/ Random Forest/ XGBoost.
* **Interest Rate Prediction**: Regression models (Linear Regression/ Neural Networks).
* **Evaluation Metrics**: Accuracy, Precision-Recall, Mean Squared Error (MSE) for interest rate prediction.

#### **5.4 Deployment**

* Development of a web-based application for users to input details and receive instant predictions.

### **6. Expected Outcomes**

* A functional machine learning model capable of predicting loan eligibility and interest rates.
* Improved loan approval process reducing bias and manual errors.
* Increased financial inclusion by providing fair evaluations for individuals with limited credit history.

### **9. Tools & Technologies**

* **Programming Languages**: Python (Scikit-learn, TensorFlow, XGBoost)
* **Data Processing**: Pandas, NumPy
* **Data Visualisation:** Tableau/Pandas
* **Deployment**: Flask, JavaScript, HTML
* **Database**: PostgreSQL

**10. Conclusion**

This project aims to facilitate the loan approval process by analysing different features that affect loan eligibility.Providing accurate predictions for loan eligibility and interest rates will enhance financial decision-making for lenders while increasing accessibility to loans for borrowers. The system will contribute to a fairer, more efficient, and data-driven lending ecosystem.