**project proposal**

**Jahnavi jammula SP ID : 1022376**

**Project Title: Credit Scoring and Segmentation System**

**Introduction:**

The project aims to build a robust credit scoring and segmentation system using Python. By leveraging machine learning techniques, the system will analyze historical credit data to assess and segment customers based on their creditworthiness, enabling financial institutions to make informed lending decisions.

**Objectives:**

Gather and preprocess credit data for analysis.

Explore various machine learning algorithms suitable for credit scoring.

Develop a credit scoring model to predict customer creditworthiness.

Implement segmentation techniques to categorize customers based on risk profiles.

Create a user interface for visualizing credit scores and segmentation results.

**Methodology:**

**Data Collection and Preprocessing:**

Collect historical credit data from reliable sources.

Clean the dataset by handling missing values, outliers, and encoding categorical variables.

Conduct exploratory data analysis (EDA) to understand patterns and correlations.

**Feature Engineering:**

Identify and engineer relevant features such as credit history, income, loan amounts, etc.

Normalize or scale features to prepare them for modeling.

**Credit Scoring Model Development:**

Select appropriate machine learning algorithms (e.g., Logistic Regression, Random Forest, Gradient Boosting) for credit scoring.

Split the dataset into training and testing sets.

Train models using the training data and optimize hyperparameters.

**Model Evaluation and Selection:**

Evaluate model performance using metrics like accuracy, precision, recall, and ROC curves.

Choose the best-performing model based on evaluation results.

**Customer Segmentation:**

Utilize clustering techniques (e.g., K-means, DBSCAN) to segment customers based on credit risk profiles.

Visualize segmentation results to identify distinct customer groups.

**Interface Development:**

Develop a user-friendly interface to input customer information and obtain credit scores.

Display segmentation results through interactive visualizations.

**Deliverables:**

Detailed report documenting the methodology, model development, and segmentation results.

Trained machine learning model(s) for credit scoring.

Codebase and scripts used for data preprocessing, model training, and evaluation.

User interface for inputting customer data and accessing credit scores.

**Resources Required:**

Dataset containing historical credit data

Python programming language with libraries like Pandas, NumPy, Scikit-learn

Machine learning algorithms for classification and clustering

Web development tools/frameworks for building the user interface (e.g., Flask, Django)

**Conclusion:**

The project aims to deliver an accurate credit scoring model and segmentation system, empowering financial institutions to make data-driven decisions while assessing the creditworthiness of customers, thereby minimizing risks associated with lending.