**Predictive Analytics and Recommendation**

**Systems in Banking**

**Introduction**

This project focuses on predicting Loan Defaults using Supervised Learning, Segmenting Customers with Unsupervised Learning, and Recommending Bank Products through a Recommendation Engine.

#### **Technologies Used:**

* Python
* Pandas
* NumPy
* Scikit-learn
* Matplotlib
* Seaborn
* Streamlit
* Pickle
* Scipy
* Surprise

#### **Installation:**

* pip install pandas
* pip install numpy
* pip install scikit-learn
* pip install matplotlib
* pip install seaborn
* pip install streamlit
* pip install pickle
* pip install scikit-surprise

#### **Data Collection**

Data is collected using the Google.

* **Loan Default Prediction:** Customer demographics, loan amounts, interest rates, and repayment status.
* **Customer Segmentation:** Transaction details including amounts, types, and dates.
* **Loan Default Prediction:** Customer interactions with various banking products.

#### **Data Understanding:**

1. Identify Variable Types
2. Handle Invalid Values

#### **Data Pre-processing:**

1. Handle Missing Values using Mean, Median, Mode.
2. Detect Outliers using IQR or Isolation Forest.
3. Determine Skewnes Using Log, sqrt or Box-Cox transformations.
4. Encode Categorical Variables with One-Hot Encoding, Label Encoding, or Ordinal Encoding.

#### **Exploratory Data Analysis:**

1. Visualize Outliers and Skewness with Boxplot, Distplot or Violin plots.
2. Analyze and Treat Skewness.
3. Histogram,pairplot,crossplot,heatmap,scatterplot,countplots are visualized.

#### **Feature Engineering:**

1. Create New Features through Aggregation or Transformation.
2. Drop highly correlated columns using heatmaps.

#### **Model Building and Evaluation:**

* **Split Data**
* **Model Training and Evaluate**:
  + **Loan Default Prediction:** Use Classification models- Logistic Regression,Decision Tree Clasifier, Random Forest Classifier,Gradient Boosting. Metrics: Accuracy, Precision, Recall, F1 score.
  + **Customer Segmentation:** Use Clustering Algorithms- KMeans, DBscan, Hierarchical Clustering to segment customers based on transaction behavior. Metrics: Silhouette scores and Davies-Bouldin index to evaluate cluster quality
  + **Product Recommendations:** Use Collavorative filtering or Content-Based Filtering Algorithms. Metrics: Precision, Recall, Mean Average Precision(MAP), Normalized Discounted Cumulative Gain score.
* **Optimize with Hyperparameter Tuning**: Use Cross-Validation and Grid Search.

#### **Model GUI:**

1. Develop a Streamlit App for interactive predictions, customer segmentations and product recommendations.
2. Allow the users to input feature values and display predictions, customer segmentations and product recommendations.

#### **Usage**:

Steps to be followed for effectively using the application:

1. **Access the Streamlit App:** Open the application in your browser.
2. **Select target:** Choose from options such as Loan Default Prediction or Customer Segmentation or Product Recommendations from the Navigation menu.
3. **Input Data:**
   * **For Loan Default Prediction:** Input customer informations.
   * **For Customer Segmentation:** Provide Transaction details.
   * **For Product Recommendations:** Input customer interaction data
4. **Perform Prediction:** By clicking the button will able to get results based on the input data.
5. **Results:** The Prediction output will be displayed on the page, allowing you to analyze loan default risks, customer segments, or recommended products