Project Title	Predictive Analytics and Recommendation Systems in Banking
Skills take away From This Project	Python, Datascience , SQL , ML
Domain	Banking

Problem Statement:

Banks deal with large volumes of customer data and transactions daily. There is a need to leverage this data to enhance customer experience, reduce risk, and optimize product offerings. This project aims to solve three critical problems:

1. Predicting loan defaults (Supervised Learning).
2. Segmenting customers based on transaction behavior (Unsupervised Learning).
3. Recommending suitable banking products to customers (Recommendation Engine).

Business Use Cases:

● **Loan Default Prediction:** Banks need to minimize financial risk by predicting which customers are likely to default on loans. Accurate predictions can lead to better loan approval processes and risk management.
● **Customer Segmentation:** By understanding customer behavior, banks can target their marketing strategies more effectively, enhancing customer satisfaction and increasing revenue.
● **Product Recommendations:** Recommending the right products to customers based on their past behavior can increase customer loyalty and boost cross-selling opportunities.

Approach:

A. Supervised Learning: Loan Default Prediction

- **Data Source:** Historical customer data including demographics, credit scores, income, loan details, and repayment history.
- **Algorithm:** Logistic Regression, Random Forest, or Gradient Boosting.
- **Steps:**
    1. **Data Preprocessing:** Cleaning and preparing the data, handling missing values, and encoding categorical variables.
    2. **Feature Engineering:** Identifying important features and creating new ones if necessary.
    3. **Model Training:** Splitting data into training and test sets, training models, and tuning hyperparameters.
    4. **Model Evaluation:** Evaluating model performance using metrics like accuracy, precision, recall, and F1-score.

B. Unsupervised Learning: Customer Segmentation

- **Data Source:** Transactional data including frequency, amount, and types of transactions.
- **Algorithm:** K-Means Clustering or Hierarchical Clustering.
- **Steps:**
    1. **Data Preprocessing:** Standardizing the data and selecting relevant features.
    2. **Clustering:** Applying clustering algorithms to group customers based on similar transaction patterns.
    3. **Evaluation:** Analyzing the clusters using visualization techniques and interpreting the results.

C. Recommendation Engine: Product Recommendations

- **Data Source:** Customer interaction data, product purchase history, and transaction data.
- **Algorithm:** Collaborative Filtering, Matrix Factorization, or Content-Based Filtering.
- **Steps:**
    1. **Data Preprocessing:** Preparing the data and handling sparsity in the customer-product matrix.
    2. **Model Building:** Implementing the recommendation algorithm and generating product recommendations.
    3. **Evaluation:** Measuring the effectiveness of recommendations using metrics like Precision and Recall.

Results:

- **Loan Default Prediction:**

- ○ Expected output is a model that can predict with high accuracy whether a customer will default on a loan.
- ○ The model will help the bank minimize risk and improve the loan approval process.
- **Customer Segmentation:**
  - ○ Expected output is a set of customer segments that the bank can use for targeted marketing and product development.
  - ○ The bank can tailor its services and products to different customer groups based on their behavior.
- **Product Recommendations:**
  - ○ Expected output is a personalized list of products for each customer.
  - ○ The bank can use this to enhance customer satisfaction and increase sales through cross-selling.

Project Evaluation metrics:

Supervised Learning (Loan Default Prediction):

- Accuracy, Precision, Recall, F1-Score, ROC-AUC Score.

Unsupervised Learning (Customer Segmentation):

- Silhouette Score, Davies-Bouldin Index, Cluster Visualization (e.g., PCA plot).

Recommendation Engine:

- Precision, Recall, Mean Average Precision (MAP), Normalized Discounted Cumulative Gain (NDCG).

Technical Tags:

NA

Data Set:

Please create the synthetic dataset using python libraries like faker , etc and also do the data collection process based on business requirement and feature engineering by adding more derived attribute

Below are the sample features

A. Loan Default Prediction

- **Data Source:** Loan data from a fictional or anonymized dataset (e.g., LendingClub data).
- **Columns:**
    - `customer_id`: Unique identifier for each customer.
    - `age`: Customer age.
    - `income`: Annual income of the customer.
    - `credit_score`: Credit score of the customer.
    - `loan_amount`: Amount of the loan.
    - `interest_rate`: Interest rate on the loan.
    - `loan_term`: Term of the loan in months.
    - `repayment_status`: Binary indicator (1 = default, 0 = no default).

B. Customer Segmentation

- **Data Source:** Transactional data from a fictional bank.
- **Columns:**
    - `customer_id`: Unique identifier for each customer.
    - `transaction_id`: Unique identifier for each transaction.
    - `transaction_amount`: Amount of the transaction.
    - `transaction_type`: Type of transaction (e.g., deposit, withdrawal).
    - `transaction_date`: Date of the transaction.

C. Recommendation Engine

- **Data Source:** Customer-product interaction data.
- **Columns:**
    - `customer_id`: Unique identifier for each customer.
    - `product_id`: Unique identifier for each product.
    - `interaction_type`: Type of interaction (e.g., purchased, viewed, clicked).
    - `interaction_date`: Date of interaction.

Data Set Explanation:

MEntioned above

Project Deliverables:

Format: A zip file containing all project components (EDA slides, code, documentation, and deployment artifacts).

Platform: Option to submit via the course's LMS or GitHub.

EDA Quality: Depth of analysis, insights, and data visualization.

Model Performance: Accuracy, robustness, and relevance of the models used.

Business Relevance: How well the models address the banking problem.

Deployment Feasibility: Successful deployment and documentation of the process.

Presentation and Communication: Clarity, conciseness, and effectiveness of the slides and final report.

Project Guidelines:

Data Handling: Ensure data is preprocessed correctly, with missing values handled and features properly engineered.

Model Selection: Justify the choice of algorithms and provide insights into model performance.

Visualization: Use clear and informative visualizations for EDA and result interpretation.

Documentation: Maintain detailed documentation of the entire project, including challenges faced and solutions implemented.

Deployment: Clearly document the deployment process, whether on-premises or in the cloud, with step-by-step instructions.

Presentation: Summarize the project in a concise, clear manner, focusing on business relevance and key takeaways.

Format: Documentation and screenshots of the deployment.

Content: Deployment strategy, steps taken to deploy the models on-premises or in the cloud, and a demonstration of the models in action.Key insights, business implications, and recommendations for stakeholders.

Timeline:

5 days

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Tuesday, Thursday, Saturday (5:00PM to 7:00PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open on Saturday and Sunday Only on Every Week

Timing: Monday-Saturday (11:30PM to 12:30PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA