Final Team Project

BUSI 651 Summer 2023

Objective:

The objective of this final project is to guide you through the process of predicting credit credibility using machine learning techniques. You will work with a provided dataset containing customer information and credit-related features. Through this assignment, you will be able to apply quantitative reasoning, critical and creative thinking, and communication, and problem-solving skills.

Problem: Customer Credit Prediction:

In this project, you will work with a dataset containing customer information and credit-related features. Your task is to explore, analyze, clean, preprocess, and engineer features from the data provided to you and finally develop a machine learning / deep learning model to predict customer credit. You have the flexibility to define the problem and choose the approach that best suits your interests and skills.

Data Dictionary	
Column	**Description**
CLIENTNUM	Client number. Unique identifier for the customer holding the account
Customer_Age	Demographic variable - Customer's Age in Years
Gender	Demographic variable - M=Male, F=Female
Dependent_count	Demographic variable - Number of dependents
Education_Level	Demographic variable - Educational Qualification of the account holder
Marital_Status	Demographic variable - Married, Single, Divorced, Unknown
Income_Category	Demographic variable - Annual Income Category of the account holder
Card_Category	Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
Months_on_book	Period of relationship with bank
Total_Relationship_count	Total no. of products held by the customer
Months_Inactive_12_mon	No. of months inactive in the last 12 months
Contacts_Count_12_mon	No. of Contacts in the last 12 months
Credit_Limit	Credit Limit on the Credit Card
Total_Revolving_Bal	Total Revolving Balance on the Credit Card
Total_Amt_Chng_Q4_Q1	Change in Transaction Amount (Q4 over Q1)
Total_Trans_Amt	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Change in Transaction Count (Q4 over Q1)
Avg_Utilization_Ratio	Average Card Utilization Ratio

Project Phases: (whichever applies to your problem)

1. **Data Exploration and Analysis (EDA):**
   - Understand the dataset's structure, dimensions, and types of data.
   - Explore the distribution of each feature and target variable.
   - Visualize relationships between features and the target variable using appropriate plots and charts.
   - Identify potential patterns, trends, and outliers.
2. **Data Cleaning:**
   - Identify and handle missing values appropriately (imputation, removal, etc.).
   - Address any inconsistencies or errors in the data.
   - Ensure data integrity and consistency.
3. **Feature Engineering:**
   - Select relevant features based on domain knowledge and exploration.
   - Create new features that might improve predictive performance.
   - Convert categorical variables into numerical representations (one-hot encoding, label encoding).
   - Scale or normalize numerical features if needed.
4. **Problem Definition:**
   - Define the specific problem you want to solve. This part is where you can show creativity in defining a problem and apply your machine knowledge to solve it. Examples include predicting:
     - credit limit value
     - credit limit level (for example high, average, low)
     - credit segmentation
     - credit risk
     - **…**

     Note: No need to limit yourself to this list!
   - Clearly state your goals and objectives.
5. **Model Selection and Implementation:**
   - Choose suitable machine learning algorithms for your defined problem.
   - Split the dataset into training and testing sets.
   - Train, tune, and validate your chosen models.
6. **Evaluation and Interpretation:**
   - Evaluate the performance of your models using appropriate evaluation metrics (accuracy, precision, recall, MSE, R-squared, etc.).
   - Interpret the results and provide insights into the features that are most influential for predictions.
7. **Conclusion and Recommendations:**
   - Summarize your findings and conclusions.
   - Recommend actions or strategies based on your model's predictions.
   - Discuss the potential future directions to improve the performance of your model.

Report Submission Instruction:

- Maximum 25 pages including relevant graphs, charts, illustrations, tables, reference, and Appendix.
- No need to include your python code. (I may ask to see your python code during the presentation day)
- Report due date: Sep 10, 23:59
- Presentation: Sep 10, during class
- All the team members must submit the report.

Report Expectations:

Your report should cover the following content or may include relevant sections where appropriate. Use your own creativity to provide additional sections if needed.

- Executive summary
- Introduction
- Project justification, problem definition, project goal, etc.
- Team roles
- Data exploration and analysis
- Quantitative analysis, and data visualizations
- Data Preprocessing
- Predictive model selection and development
- Model performance assessment
- Predictions, discussion, and recommendations
- References/Appendix

1.0 Grading Rubric

Key Points	Grade Allocation (%)
General Report Format (fonts, tables, etc.)	10
Technical content completeness	65
Comply with APA style	5
Creativity in the problem, approach and results	20

N.B. Failure to comply with the above would result in low grades.

2.0 Deadline

Submission: A single word/pdf document for group report.

Deadline: 23:59, Sep 10, 2023