CENG4515: Data Science and Analytics

Final Project for 2025-2026 Fall Semester

Instructor: Assist. Prof. Dr. Zeynep Filiz EREN

Project Overview

- Project focuses on constructing an integrated analytical dataset derived from various customer-related data sources, with the objective of predicting **customer status** using classification techniques. The dataset provides a multidimensional view of each customer, including product interaction patterns, financial indicators, support activities, satisfaction metrics, demographic information, and engagement behavior.

The project requires the integration, examination, and modeling of these heterogeneous data components to gain insights into customer behavior and to develop a robust predictive system for customer status categorization.

Dataset Overview

+The dataset consists of multiple CSV files, each representing a distinct dimension of customer information. The following section outlines the content and column structure of each file.

customer_satisfaction_scores.csv

Contains customer satisfaction survey responses collected periodically.

Column	Description
Customer ID	Customer identifier.
Year	Survey year.
Quarter	Fiscal quarter in which the survey was conducted.
Survey Date	Date the survey was sent.
Response Date	Date the survey was completed by the customer.
How likely are you to recommend insider to a friend or colleague	Likelihood to recommend the company or product.
How would you rate the value you gain from our company	Perceived value assessment.
How frequently are you using our platform	Frequency of platform usage.
Please rate the overall quality of our products	Evaluation of product quality.
Please rate the usability of the panel	Assessment of platform usability.
Please rate your understanding of our reporting capabilities in the panel	Self-reported understanding of reporting features.

customer_demographics.csv

Represents the customer's age within the system, reflecting duration of engagement.

Column	Description
CUS ID	Customer identifier.
Customer Age (Months)	Number of months the customer has been using the platform.

customer_monthly_recurring_revenue.csv

Contains the specifies the monetary amount that each customer pays to system on a monthly basis for the subscribed products or services.

Column	Description
Customer ID	Customer identifier.
MRR	The amount of money that the customer is paying to the system on a monthly basis.

customer_revenue_history.csv

Indicates the total amount of money that the customer has earned by using the products of system.

Column	Description
Customer ID	Customer identifier.
Total Revenue	Total amount of money the customer has earned by utilizing system's products.

support_ticket_activity.csv

Contains information about customer interactions with the support team.

Column	Description
Customer ID	Customer identifier.
Help Ticket Count	Total number of support tickets submitted.
Help Ticket Lead Time (hours)	Average resolution time for support requests (in hours).

newsletter_engagement.csv

Represents customer engagement with corporate newsletters.

Column	Description
Customer ID	Customer identifier.
Company Newsletter Interaction Count	Number of interactions with company newsletters.

product_bug_reports.csv

Contains records related to product issues reported by customers.

Column	Description
Customer ID	Customer identifier.
Product Bug Task Count	Total number of bug or issue reports submitted by the customer.

customer_region_and_industry.csv

Provides regional and industry classification for each customer.

Column	Description
Customer ID	Customer identifier.
Region	Geographic region where the customer operates.
Vertical	Primary industry classification of the customer.
Subvertical	Detailed subcategory within the primary industry.

customer_status_level.csv (Target File)

Includes the target variables used for status and level prediction.

Column	Description
Customer ID	Customer identifier.
Status (Target Variable)	Customer condition (Onboarding, Churn, Retained).
Customer Level	Assigned customer tier level.

Assessment Criteria Overview

1) Data Exploration and Visualization

In this stage, the project is expected to include a thorough examination of the provided datasets to gain an initial understanding of their structure, distribution, and key characteristics. The objective is to identify general patterns, trends, and notable observations through visualizations and descriptive insights. The emphasis is on demonstrating the ability to interpret data and extract meaningful impressions at first glance.

2) Data Cleaning and Preprocessing

This part of the project involves preparing the data for analysis and modeling by ensuring that it is organized, consistent, and suitable for further processing. The expectation is to detect and address evident issues within the data, establish a coherent structure, and construct a clean dataset that can reliably support the modeling workflow. The focus is on showing careful handling of the data and a deliberate approach to preparing it for subsequent steps.

3) Feature Engineering

The project is expected to incorporate the creation of additional meaningful information derived from the existing data. The purpose is to enrich the dataset by generating features that highlight relevant relationships, enhance interpretability, or strengthen the predictive capacity of the model. This component reflects analytical thinking and the ability to contribute original insights to the dataset.

4) Modeling

This part of the project involves designing a predictive approach aimed at classifying customers into the status categories provided in the dataset. The task requires selecting and implementing a suitable classification strategy based on the characteristics of the data, while clearly explaining the reasoning behind the chosen approach. The emphasis is on constructing a coherent classification model that aligns with the problem structure and demonstrates a thoughtful interpretation of the predictive objective.

5) Performance Evaluation

The project should include an analysis of the trained models, focusing on their strengths, limitations, and comparative performance. The objective is to interpret evaluation results meaningfully and assess which model performs most effectively for the task at hand. The emphasis is not only on producing performance metrics but also on understanding what these results reveal about the model’s behavior.

6) Reporting and Interpretation

The final component involves presenting the entire workflow in a clear, coherent, and well-structured manner. The project report is expected to summarize the overall process, describe the reasoning behind methodological choices, and interpret the outcomes in a meaningful way. This section highlights the ability to articulate insights, reflect on the analytical process, and convey conclusions in a thoughtful and comprehensive manner.

THANK YOU