

Graduate Certificate in Artificial Intelligence with Machine Learning
AIGC 5002 - Machine Learning and Deep Learning
Fall 2023

Assignment 1

~~Due on: October 3, 2023 at 11:59 PM~~

Due on: October 4, 2023 at 11:59 PM

Plagiarism and the use of any form of generative AI will result in a zero grade for this assignment, at the least. Please always cite your sources.

Submission guidelines:

- For this assignment, you will need to submit **1 Jupyter notebook as a PDF file. No other formats will be accepted.**
- Name the PDF as follows: `firstname_lastname_Assignment1.pdf`
- Go to the course Blackboard → Assignments folder → Assignment 1 → and submit the pdf.

(3 marks) Q1: (Max 600 Words)

Answer this question using a markdown cell in the Jupyter notebook.

Ethical Considerations in AI:

- Discuss potential ethical concerns that can arise with the use of AI and ML in decision-making processes in sensitive areas such as finance, healthcare, or criminal justice.
- Investigate a real-world incident where the use of AI in decision-making led to unintended consequences. Detail the incident, identify where things went wrong, and discuss what could have been done to avoid the incident.

(12 marks) Q2:

The dataset is available on the BlackBoard in the Assignment folder.

Scenario: You are the Chief Machine Learning engineer at an e-commerce start-up. You have collected data about 3000 of your customers that describes their purchasing behaviors. Your goal is to develop a Multiple Linear Regression model to predict the Customer Lifetime Value (CLV) score.

Dataset description

The dataset captures the behaviors and metrics of customers in an e-commerce environment over a period of a year. It has the following variables:

- **Total_Purchases:** This is a representation of the total monetary amount a customer has spent within the e-commerce store over the last year. In this dataset, it is simulated to range from \$10 to \$1000. This helps understand a customer's spending power or affinity towards the store's products.
- **Number_of_Purchases:** This represents how many times a customer has made a purchase in the past year. The frequency of transactions can be an indicator of loyalty or repeated need.
- **Average_Rating:** This denotes the average rating (out of 5) given by the customer for the products they've purchased. A higher average rating might indicate satisfaction with the product quality or alignment with customer needs, while a lower rating may signify dissatisfaction.
- **Days_Since_Last_Purchase:** This metric can be crucial in identifying how engaged a customer is. A smaller value may signify that the customer is active, while a larger value might indicate the customer hasn't shopped in a while and could be at risk of churning.
- **Number_of>Returns:** This signifies the number of times a customer has returned a product in the past year. Frequent returns might indicate dissatisfaction or issues with product fit, quality, or alignment with expectations. It's also a potential marker for costing the company more in return processing fees.
- **CLV (Customer Lifetime Value):** the total revenue you as an e-commerce business earns from a customer over time.

For e-commerce, using such a dataset, can:

- Identify high-value customers and potentially offer them loyalty programs or personalized discounts.
- Gauge product and service quality based on ratings and returns.
- Predict future sales or revenues.
- Implement targeted marketing strategies based on customer engagement metrics.

Your tasks for this assignment:

1. Data Exploration and Visualization:
 - a. Load the dataset using pandas.
 - b. Visualize the relationships between features and the target variable. State your observations.
2. Data Pre-processing:
 - a. Split the data into training and test sets (80% train, 20% test).
3. Model Development:
 - a. Using the libraries and functions we used in the lab session of LR and MLR, create an MLR model.
 - b. Train the model on the training set.
 - c. Predict CLV values on the test set.

4. Model Evaluation:
 - a. Evaluate the model using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score.
 - b. Interpret the coefficients of the model.
 - c. Interpret the significance and impact of each feature on the CLV based on the coefficients.
 - d. Use WEKA on the same dataset and compare your results. State your analysis. Transfer the results from WEKA to your Jupyter notebook.
-