

Assignment 2

Possible Points: 100

Due Date: 28th Feb 2025

Assignments should be done on an individual basis.

Download the dataset **Food_delivery_times.csv** provided to perform the following tasks:

Q1: Build a predictive linear regression model to estimate delivery time based on distance, weather, traffic level and time of the day to predict delivery time. Divide the data into 70% training and 30% testing. **(Total 25 points)**.

- Perform one-hot encoding on the categorical variables. Check for null values and remove if present. **(5 points)**
- Report the coefficient values using standard Least Squares Estimates. **(5 points)**
- Evaluate RSS, Adjusted r square and MSE. Explain the results. **(7 points)**
- Analyze whether there is a relationship between distance and delivery time. **(5 points)**
- What is the difference between linear and logistic regression? **(3 Points)**

Q2: Using the dataset provided, perform regression using Ridge and Lasso regression models to predict the delivery time **(Total 50 Points)**

- Use all the attributes given in the dataset, check for null values and drop them. Plot bar chart and visualize deliveries during different times of the day **(5 Points)**
- Apply label encoding to categorical values **(5 Points)**
- Check for outliers in the column courier experience. Plot a histogram for the column distance. **(5 Points)**
- Divide the dataset into 60% training and 40% testing data **(3 Points)**
- Train the data on Ridge and Lasso regression models. **(10 Points)**
- Report MSE and R-squared metrics for both the models. Which model performs better? Explain. **(15 Points)**
- Apply cross validation on the underperforming model. **(7 Points)**

Q3: Develop a logistic regression model to predict the traffic level based on distance, weather, time of the day and vehicle type. Split the dataset into 70% training and 30% testing **(Total 25 points)**.

- Apply label encoding to the categorical variables. Drop null values. Apply standard scaler on distance **(7 points)**
- Generate a pie chart for time of the day. **(3 Points)**
- Create a heatmap to visualize the correlation between preparation time and delivery time. **(5 points)**

- Assess the model's performance by calculating accuracy, precision, recall, and the area under curve (AUC). (**5 points**)
- Discuss the significance of the predictors in determining vehicle type based on p-values. (**5 points**)