Model Comparison and Inferences

For this project, I applied three different models on the dataset, each targeting a specific problem formulation. The models evaluated were:

1. MLP Regressor:
     A neural network-based regression model was employed to predict a continuous target variable. The key performance metrics for this model were:

     - R^2 Score: 0.493
       This indicates that the model explains about 49% of the variance in the target variable.

     - Mean Squared Error (MSE): 1.243
       On average, the squared error between the predicted and actual values is 1.243. While the MLP Regressor captures a moderate amount of the data variability, there remains substantial unexplained variance.

2. Logistic Regression:
    Performance metrics of this linear classification model were as follows:

     - Accuracy: 39.8%
       The model correctly classified about 40% of the instances.

     - Confusion Matrix and Classification Report:
       The confusion matrix shows significant misclassification across classes, and the precision and recall values (ranging roughly from 0.35 to 0.47 and 0.21 to 0.53, respectively) indicate that the model is struggling to correctly identify instances for each class.

       Logistic regression, which relies on linear decision boundaries, seems insufficient to capture the complexity of the data's class structure. Its lower overall accuracy and imbalanced performance across classes suggest that a more flexible or non-linear model might be more appropriate.

3. Decision Tree Classifier:
    This model builds a tree-like structure to classify the data and is capable of modeling non-linear relationships. The performance metrics were:

- ○ Accuracy: 63.7%
  The decision tree correctly classified approximately 64% of the instances.

- ○ Confusion Matrix and Classification Report:
  With precision values ranging from 0.55 to 0.78 and recall values from 0.60 to 0.73 across classes, the decision tree provides a more balanced performance compared to logistic regression.
  The decision tree classifier outperformed logistic regression by capturing more complex patterns in the data, leading to higher overall accuracy and better class-wise performance.

- ● MLP Regressor (Regression):
  Achieved a moderate performance with an R² of 0.493 and an MSE of 1.243. This indicates that while the model explains nearly half of the variance.

- ● Logistic Regression (Classification):
  With an accuracy of about 40%, logistic regression struggled with the multi-class problem, likely due to its reliance on linear separability which does not sufficiently capture the data's complexity.

- ● Decision Tree Classifier (Classification):
  The best performer among the classification models with an accuracy of 63.7% and more balanced precision and recall scores across classes. Its ability to model non-linear relationships makes it a strong candidate for this dataset.