Name: Md. Raihan Sobhan
Student ID: 1905095

Implementation of a Decision Tree algorithm.

Objective:

The primary objective of this assignment was to implement a decision tree learning algorithm and apply it to the car evaluation dataset for classification. The specific tasks included implementing the decision tree learning algorithm using the information gain criteria, dividing the dataset into training and testing sets, running the experiment multiple times, and calculating the average test accuracy along with the standard deviation.

Implementation Steps:

1. Implementation of a Decision Tree Learning Algorithm, ID-3:

I implemented the decision tree in Java using the ID-3 algorithm. The algorithm uses the information gain criteria to select attributes at each node. The decision tree is constructed recursively by choosing the best attribute to split the data.

2. Data Splitting:

We randomly divided the dataset into a training set (80%) and a testing set (20%) for each iteration of the experiment. This ensures that the decision tree is trained on a portion of the data and tested on a separate portion.

3. Multiple Experiments:

We conducted the experiment 20 times to assess the performance of the decision tree classifier on different random data splits.

4. Accuracy Calculation:

For each experiment, we calculated the accuracy of the decision tree classifier on the test set by comparing the predicted labels to the actual labels. The accuracy is defined as the ratio of correct predictions to the total number of test instances.

5. Mean Accuracy and Standard Deviation:

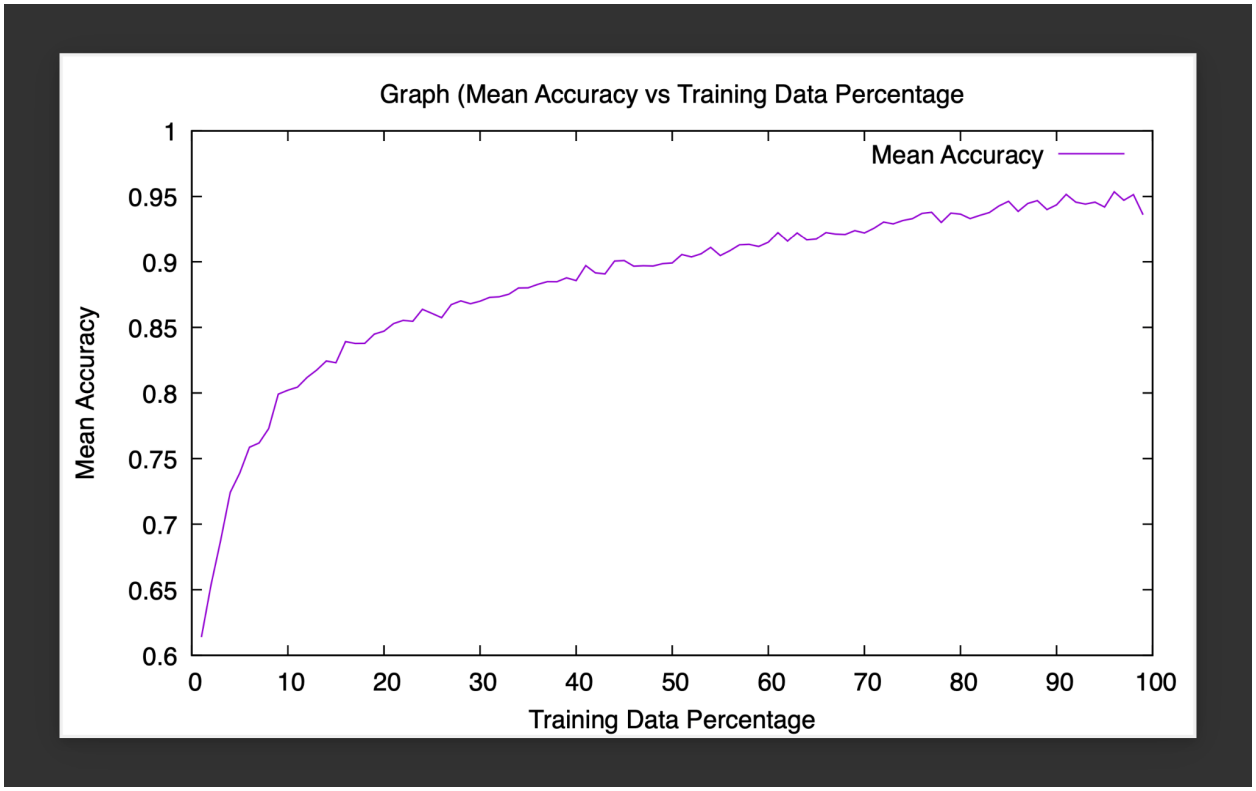
After running the experiment 20 times, we calculated the mean accuracy and standard deviation of the accuracies across all experiments. The mean accuracy provides an estimate of the classifier's overall performance, while the standard deviation indicates the variability in accuracy across different runs.

Results:

- Mean Accuracy: The mean accuracy of the decision tree classifier across 20 experiments was found to be approximately 93.77%.
- Standard Deviation: The standard deviation of the accuracies was 0.016, indicating the variability in accuracy across experiments.

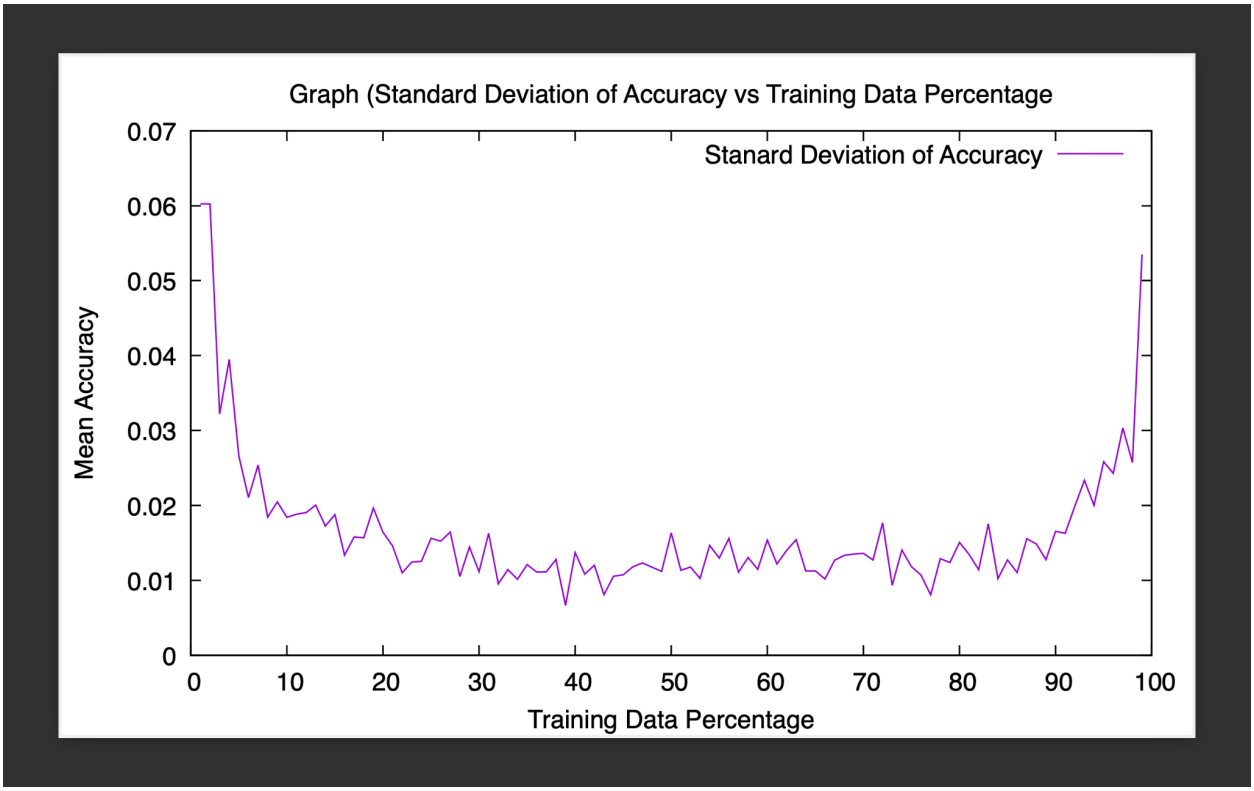
1. ### Change of mean accuracy with respect to change of the percentage of training data :

   The accuracy increases as the percentage of training data is increased.



2. Change of standard deviation with respect to change of the percentage of training data :

   Standard deviation is higher for lower percentage of training data and for lower percentage of test data too. It's lower , when there is a balance between training and testing data.



Discussion:

The decision tree classifier achieved an average accuracy of 93.77%, indicating its effectiveness in classifying car evaluations based on the provided dataset. The standard deviation of 0.016 suggests that there is some variability in performance across different data splits, but the classifier's performance remains consistent on average.
It's important to note that the accuracy of the decision tree classifier may be influenced by factors such as the choice of attribute selection criteria, the depth of the tree, and the specific random splits used for training and testing. Fine-tuning these parameters could potentially lead to improved accuracy.

Conclusion:

In this assignment, we successfully implemented a decision tree learning algorithm and applied it to the car evaluation dataset. By running multiple experiments and calculating the mean accuracy along with the standard deviation, we gained insights into the classifier's performance.

The decision tree classifier showed promise in accurately classifying car evaluations, and further optimization can potentially enhance its performance.
This assignment provided valuable hands-on experience with decision tree algorithms, dataset splitting, and performance evaluation, which are essential skills in machine learning and data analysis.
Overall, the assignment was a valuable exercise in understanding and implementing machine learning algorithms for classification tasks.