Homework Assignment: Vehicle MultiClass Classification Using Machine Learning Algorithms

Classifiers:

- Logistic Regression
- Decision Tree
- Bayes

BACKGROUND

You have been hired by a car manufacturer to build a machine learning model that can classify different types of vehicles based on their features. The manufacturer wants to use this model to automatically classify vehicles during production to ensure that they are correctly labeled and packaged.

The dataset you will be using contains information on different types of vehicles such as sedans, SUVs, and trucks. Each record in the dataset contains various features such as weight, horsepower, and fuel efficiency.

DATASET

The dataset you will be using can be downloaded from:

https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data

The dataset contains the following columns:

- price: The price of the vehicle
- year: The year that the vehicle was made
- manufacturer: The manufacturer of the vehicle
- condition: The condition of the vehicle (e.g., new, like new, excellent, etc.)
- cylinders: The number of cylinders in the vehicle's engine

- fuel: The type of fuel used by the vehicle (e.g., gasoline, diesel, hybrid, etc.)
- odometer: The number of miles the vehicle has been driven
- title_status: The status of the vehicle's title (e.g., clean, salvage, etc.)
- transmission: The type of transmission used by the vehicle (e.g., automatic, manual, etc.)
- drive: The type of drive used by the vehicle (e.g., 4wd, fwd, rwd, etc.)
- type: The type of vehicle (e.g., sedan, SUV, truck, etc.)
- paint_color: The color of the vehicle's paint
- description: A description of the vehicle

TASKS

1. Preprocess the dataset to prepare it for machine learning algorithms. This should include handling missing values, converting categorical variables to numerical ones, and splitting the data into training and testing sets.

2. Build a logistic regression model to classify the vehicles. Train the model using the training set and evaluate its performance using the testing set. Report the accuracy, precision, recall, and F1 score of the model.

3. Build a decision tree model to classify the vehicles. Train the model using the training set and evaluate its performance using the testing set. Report the accuracy, precision, recall, and F1 score of the model.

4. Build a Naive Bayes model to classify the vehicles. Train the model using the training set and evaluate its performance using the testing set. Report the accuracy, precision, recall, and F1 score of the model.

5. Compare the results obtained from the three models and discuss which model performed the best

and why.

DELIVERABLES

1. A Jupyter Notebook containing your code for preprocessing the dataset and building the three machine learning models.

2. A report summarizing your findings, including the accuracy, precision, recall, and F1 score of each model and a discussion of which model performed the best and why.

GRADING

You will be graded on the following:

- The quality of your code and how well it is documented.

- The correctness of your models and their performance metrics.

- The quality of your report and how well you explain your findings.