

## Classification and Regression Trees (CART)

- [Siva.Jasthi@metrosate.edu](mailto:Siva.Jasthi@metrosate.edu)
- ICS 412: Data Mining (Spring 2023)
- Decision Tree Regression Using sklearn
- <https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/>
- Decision Tree Classification Using sklearn
- <https://www.geeksforgeeks.org/decision-tree-implementation-python/>
- Decision Tree (Evaluating the performance with different depths)
- <https://medium.com/geekculture/applying-7-classification-algorithms-on-the-titanic-dataset-278ef222b53c>

### ▼ Lab 8: Perform Steps 1 to 5 towards Lab 8

Homework 4: Perform Decision Tree Analysis to answer 5 Questions in the end.

#### ▼ 1. Do the necessary imports

```
##@title 1. Do the necessary imports
import pandas as pd
```

#### ▼ 2. Load the Cars\_Co2.csv File

```
##@title 2. Load the Cars_Co2.csv File
# and display the dataframe
# The link to the raw file is given here https://raw.githubusercontent.com/sjasthi/python_input_files/main/MachineLearning/cars_co2.csv
url="https://raw.githubusercontent.com/sjasthi/python_input_files/main/MachineLearning/cars_co2.csv"

df = pd.read_csv(url)
df.head()
```

|   | Car        | Model      | Volume | Weight | CO2 |  |
|---|------------|------------|--------|--------|-----|--|
| 0 | Toyoty     | Aygo       | 1000   | 790    | 99  |  |
| 1 | Mitsubishi | Space Star | 1200   | 1160   | 95  |  |
| 2 | Skoda      | Citigo     | 1000   | 929    | 95  |  |
| 3 | Fiat       | 500        | 900    | 865    | 90  |  |
| 4 | Mini       | Cooper     | 1500   | 1140   | 105 |  |

#### ▼ 3. Create a new colum called "CO2\_Category"

```
##@title 3. Create a new colum called "CO2_Category"

# Divide the CO2 column into 3 equal bins (Call those as 'High', 'Medium', 'Low')
# Highest Value in CO2 == High
# Lowest Value in CO2 = Low
# You need to divide the range into three buckets (three equally spaced ranges) and categorize each numerical value
# into High, Medium, Low
co2_range = df['CO2'].max() - df['CO2'].min()

bin_size=[df['CO2'].min(), df['CO2'].min() + co2_range/3, df['CO2'].min() + 2*co2_range/3, df['CO2'].max()]

categorize_labels = ['Low', 'Medium', 'High']
```

```
df['CO2_Category'] = pd.cut(df['CO2'], bins=bin_size, labels=categorize_labels, include_lowest=True)
```

4. Display the new dataframe showign the column "CO2\_Category"

```
##title 4. Display the new dataframe showign the column "CO2_Category"
display(df.head(10))
```

|   | Car        | Model      | Volume | Weight | CO2 | CO2_Category |
|---|------------|------------|--------|--------|-----|--------------|
| 0 | Toyoty     | Aygo       | 1000   | 790    | 99  | Low          |
| 1 | Mitsubishi | Space Star | 1200   | 1160   | 95  | Low          |
| 2 | Skoda      | Citigo     | 1000   | 929    | 95  | Low          |
| 3 | Fiat       | 500        | 900    | 865    | 90  | Low          |
| 4 | Mini       | Cooper     | 1500   | 1140   | 105 | Medium       |
| 5 | VW         | Up!        | 1000   | 929    | 105 | Medium       |
| 6 | Skoda      | Fabia      | 1400   | 1109   | 90  | Low          |
| 7 | Mercedes   | A-Class    | 1500   | 1365   | 92  | Low          |
| 8 | Ford       | Fiesta     | 1500   | 1112   | 98  | Low          |
| 9 | Audi       | A1         | 1600   | 1150   | 99  | Low          |

Encode the column "CO2\_Category" and create a new column called "CO2\_Category\_Code"

```
##title Encode the column "CO2_Category" and create a new column called "CO2_Category_Code"
# Use the sklearn Label Encoder to encode the "CO2_Category" column
from sklearn.preprocessing import LabelEncoder
le_CO2_Category = LabelEncoder()
df['CO2_Category_Code'] = le_CO2_Category.fit_transform(df['CO2_Category'])
```

5. Display the new dataframe showign the column "CO2\_Category\_Code"

```
##title 5. Display the new dataframe showign the column "CO2_Category_Code"
df.head(10)
```

|   | Car        | Model      | Volume | Weight | CO2 | CO2_Category | CO2_Category_Code |
|---|------------|------------|--------|--------|-----|--------------|-------------------|
| 0 | Toyoty     | Aygo       | 1000   | 790    | 99  | Low          | 1                 |
| 1 | Mitsubishi | Space Star | 1200   | 1160   | 95  | Low          | 1                 |
| 2 | Skoda      | Citigo     | 1000   | 929    | 95  | Low          | 1                 |
| 3 | Fiat       | 500        | 900    | 865    | 90  | Low          | 1                 |
| 4 | Mini       | Cooper     | 1500   | 1140   | 105 | Medium       | 2                 |
| 5 | VW         | Up!        | 1000   | 929    | 105 | Medium       | 2                 |
| 6 | Skoda      | Fabia      | 1400   | 1109   | 90  | Low          | 1                 |
| 7 | Mercedes   | A-Class    | 1500   | 1365   | 92  | Low          | 1                 |
| 8 | Ford       | Fiesta     | 1500   | 1112   | 98  | Low          | 1                 |
| 9 | Audi       | A1         | 1600   | 1150   | 99  | Low          | 1                 |

Regression Problem

Input Variables (Volume, Weight)

Target Variable (CO2)

Prepare the Data with 75% Training and 25% Test Split

```
##@title Prepare the Data with 75% Training and 25% Test Split
```

▼ Insert other coding cells as needed

```
##@title Insert other coding cells as needed
```

▼ Classification Problem

Input Variables (Volume, Weight)

Target Variable (CO2\_Category\_Code)

▼ Prepare the Data with 75% Training and 25% Test Split

```
##@title Prepare the Data with 75% Training and 25% Test Split
```

▼ Insert other coding cells as needed

```
##@title Insert other coding cells as needed
```

[Question 1 - 5 points] For the Decision Tree Regressor, which depth (1 to 10) is giving is giving highest accuracy?

Display the depth and accuracy as a table.

▼ Question 1

```
##@title Question 1
```

[Question 2 - 5 points] Using your best regressor, can you predict the CO2 emissions for the following values?

- Volume: 1575, Weight: 1200: CO2: \_\_\_\_?
- Volume: 1800, Weight: 1400: CO2: \_\_\_\_?

▼ Question 2

```
##@title Question 2
```

[Question 3 - 5 points] For the Decision Tree Classifier, which depth (1 to 10) is giving is giving highest accuracy?

Display the depth and accuracy as a table.

▼ Question 3

```
##@title Question 3
```

[Question 4 - 5 points] Using your best classifier, can you predict the CO2\_Category for the following values? Note that you need to convert CO2\_Category\_Code back to CO2\_Category using your Label Encoder. Your answer will be either 'High' or 'Medium' or 'Low'.

- Volume: 1575, Weight: 1200: CO2\_Category: \_\_\_\_?
- Volume: 1800, Weight: 1400: CO2\_Category: \_\_\_\_?

▼ Question 4

```
##@title Question 4
```

[Question 5 - 5 points] Display the "Decision Tree Regressor" using matplotlib. (Visual Representation of your best model - model with the highest accuracy)

▼ Question 5

#@title Question 5

✓ 0s completed at 3:44 PM

