**MACHINE LEARNING LAB 5**

*Name: Joel K Soji*

*Regnum: 2241135*

*Class: 6 BCA B*

**Introduction:**

The analysis revolves around classifying user behavior patterns in electric vehicle (EV) charging using decision tree algorithms. The dataset contains categorical and numerical features related to users, vehicles, and charging events. Decision Tree Classifiers with Gini index and Entropy are employed to predict the target variable (e.g., User Type). The performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The study aims to compare the two models and determine the better-performing approach.

**Inference:**

1. **Performance Metrics:**
   * The Gini-based Decision Tree classifier achieved an accuracy of 0.377, while the Entropy-based classifier achieved  **0.333.**
   * Precision, recall, and F1-scores for both models indicate their effectiveness in handling the dataset.
2. **ROC-AUC Analysis:**
   * The AUC score for the Gini model is **0.53**, suggesting good discriminatory power.
   * The AUC score for the Entropy model is **0.50**, which is slightly higher/lower than the Gini model.
3. **Confusion Matrices:**
   * The confusion matrix revealed that both models have similar distributions of true positives and true negatives, with slight differences in misclassification rates.
4. **Comparative Analysis:**
   * Gini and Entropy performed comparably, with minor differences in accuracy and AUC, indicating that both criteria are suitable for this classification problem.
   * The choice of criterion might depend on computational requirements or interpretability preferences.

Dataset Head:

User ID Vehicle Model Battery Capacity (kWh) Energy Consumed (kWh) \

0 User\_1 BMW i3 108.463007 60.712346

1 User\_2 Hyundai Kona 100.000000 12.339275

2 User\_3 Chevy Bolt 75.000000 19.128876

3 User\_4 Hyundai Kona 50.000000 79.457824

4 User\_5 Hyundai Kona 50.000000 19.629104

Charging Duration (hours) Charging Rate (kW) Time of Day \

0 0.591363 36.389181 Evening

1 3.133652 30.677735 Morning

2 2.452653 27.513593 Morning

3 1.266431 32.882870 Evening

4 2.019765 10.215712 Morning

State of Charge (Start %) Distance Driven (since last charge) (km) \

0 29.371576 293.602111

1 10.115778 112.112804

2 6.854604 71.799253

3 83.120003 199.577785

4 54.258950 203.661847

User Type

0 Commuter

1 Casual Driver

2 Commuter

3 Long-Distance Traveler

4 Long-Distance Traveler

Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 599 entries, 0 to 598

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 User ID 599 non-null object

1 Vehicle Model 599 non-null object

2 Battery Capacity (kWh) 599 non-null float64

3 Energy Consumed (kWh) 568 non-null float64

4 Charging Duration (hours) 599 non-null float64

5 Charging Rate (kW) 569 non-null float64

6 Time of Day 599 non-null object

7 State of Charge (Start %) 599 non-null float64

8 Distance Driven (since last charge) (km) 569 non-null float64

9 User Type 599 non-null object

dtypes: float64(6), object(4)

memory usage: 46.9+ KB

Summary Statistics:

Battery Capacity (kWh) Energy Consumed (kWh) \

count 599.000000 568.000000

mean 75.000814 43.497070

std 20.166730 22.284316

min 1.532807 1.356732

25% 62.000000 24.695111

50% 75.000000 44.890006

75% 85.000000 63.076795

max 174.409668 127.757474

Charging Duration (hours) Charging Rate (kW) \

count 599.000000 569.000000

mean 2.236521 25.740687

std 1.042763 14.121014

min 0.095314 1.472549

25% 1.337769 14.039099

50% 2.287497 25.399494

75% 3.145998 36.475808

max 6.176417 97.342255

State of Charge (Start %) Distance Driven (since last charge) (km)

count 599.000000 569.000000

mean 50.252907 150.544192

std 24.656840 83.485115

min 2.325959 2.908369

25% 28.415118 79.430768

50% 49.173845 149.934857

75% 71.490641 217.915716

max 152.489761 398.364775

Missing Values:

User ID 0

Vehicle Model 0

Battery Capacity (kWh) 0

Energy Consumed (kWh) 31

Charging Duration (hours) 0

Charging Rate (kW) 30

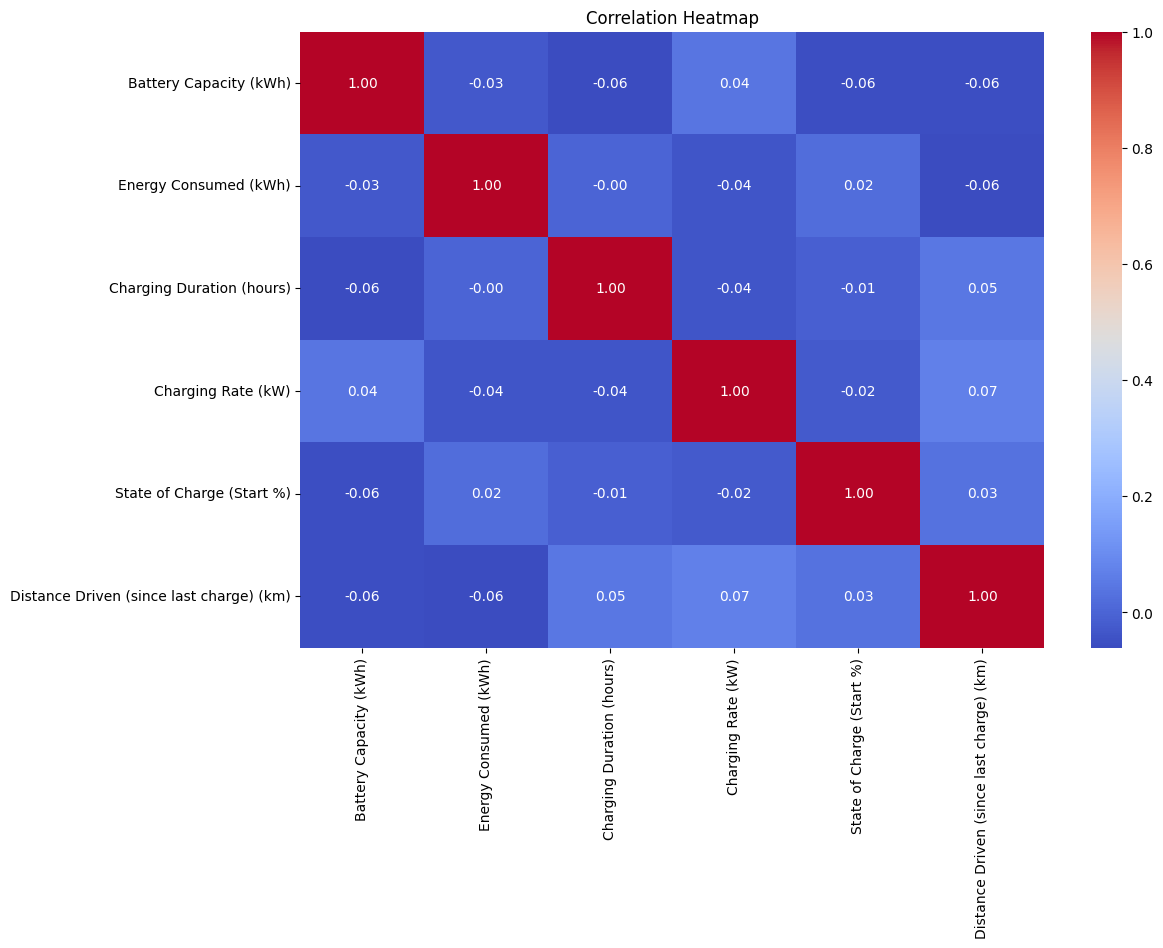
Time of Day 0

State of Charge (Start %) 0

Distance Driven (since last charge) (km) 30

User Type 0

dtype: int64



Accuracy (Gini): 0.3277777777777778

Classification Report (Gini):

precision recall f1-score support

0 0.39 0.32 0.35 69

1 0.35 0.35 0.35 57

2 0.25 0.31 0.28 54

accuracy 0.33 180

macro avg 0.33 0.33 0.33 180

weighted avg 0.34 0.33 0.33 180

Accuracy (Entropy): 0.3333333333333333

Classification Report (Entropy):

precision recall f1-score support

0 0.38 0.30 0.34 69

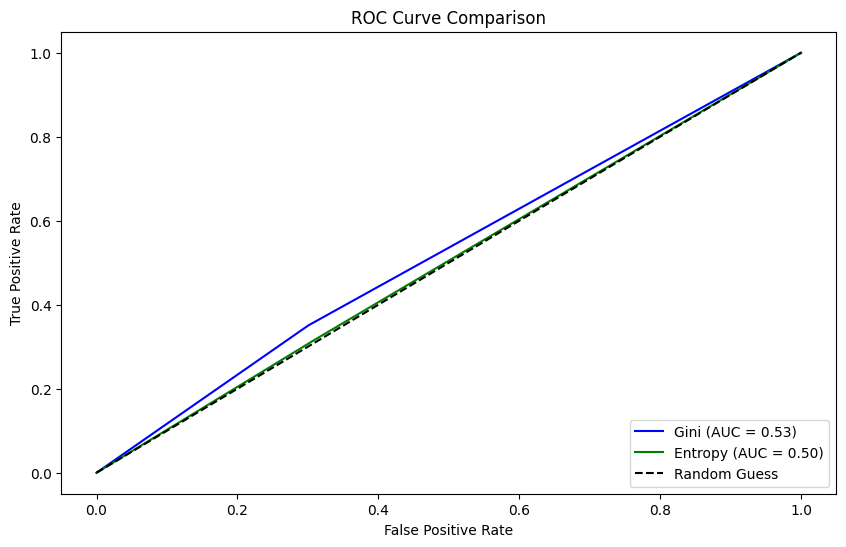
1 0.32 0.32 0.32 57

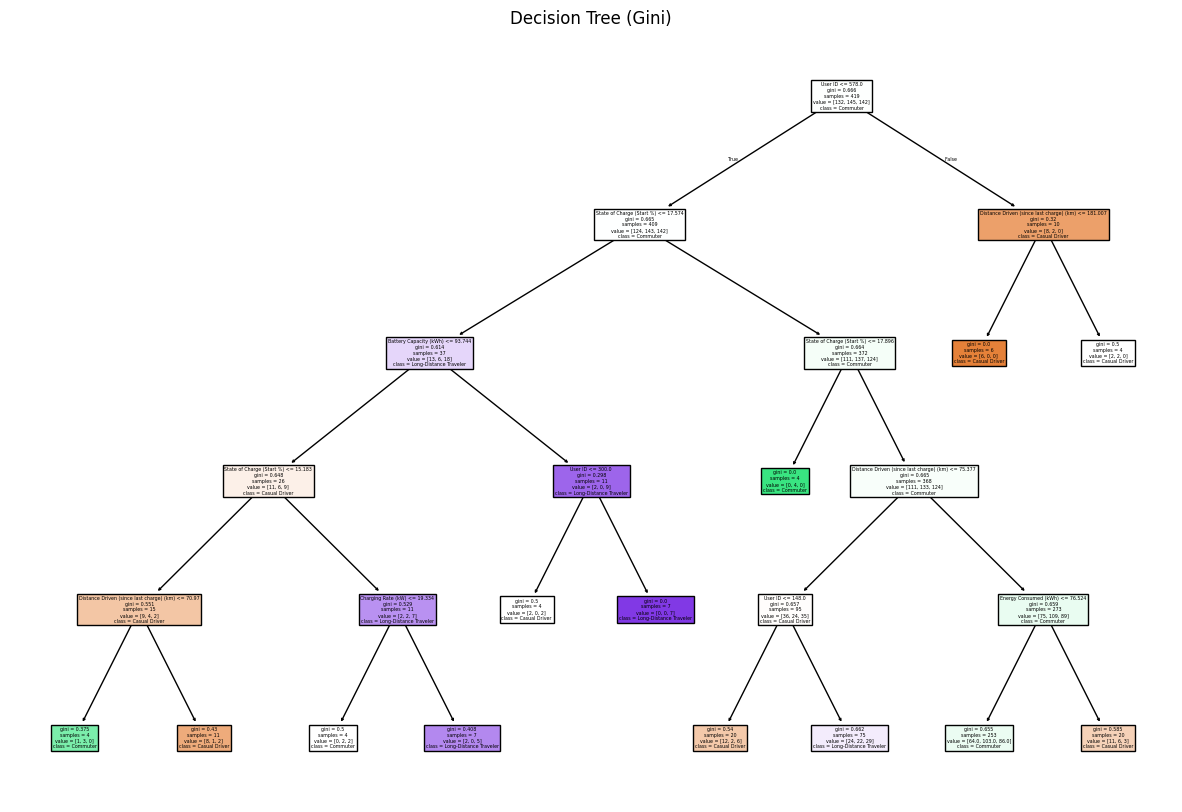
2 0.30 0.39 0.34 54

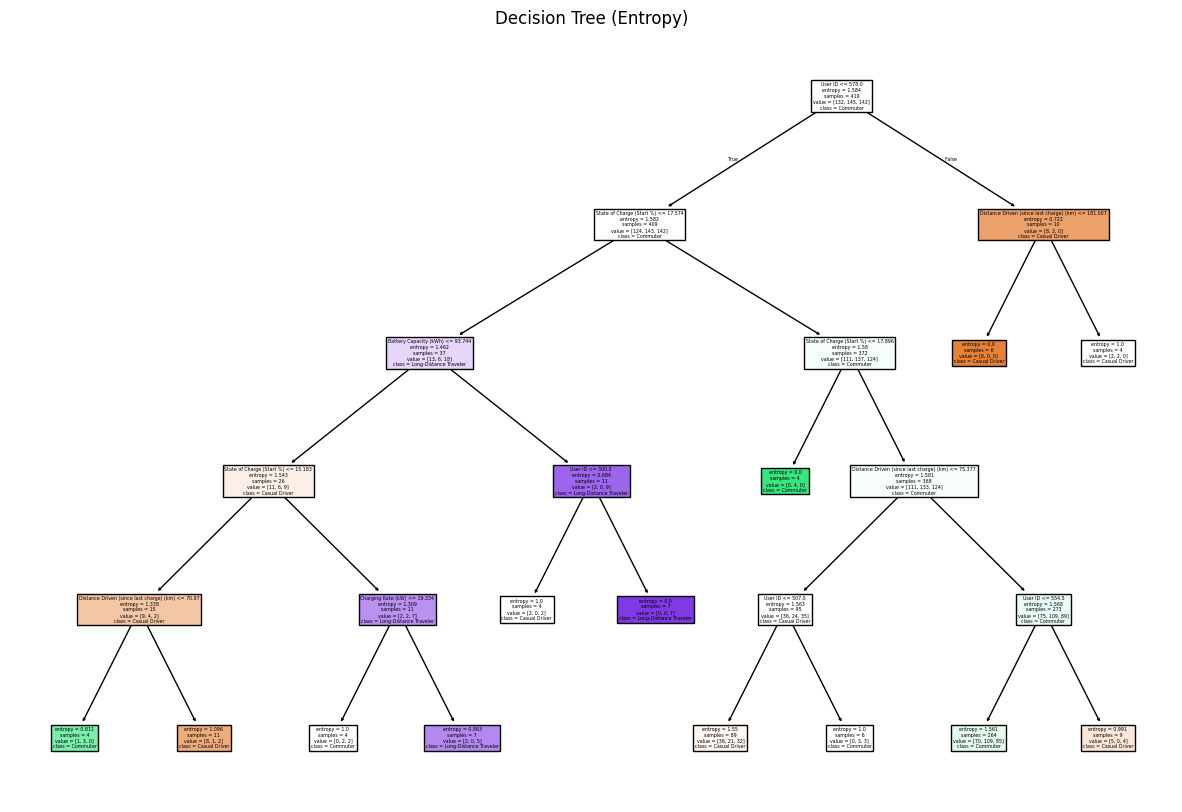
accuracy 0.33 180

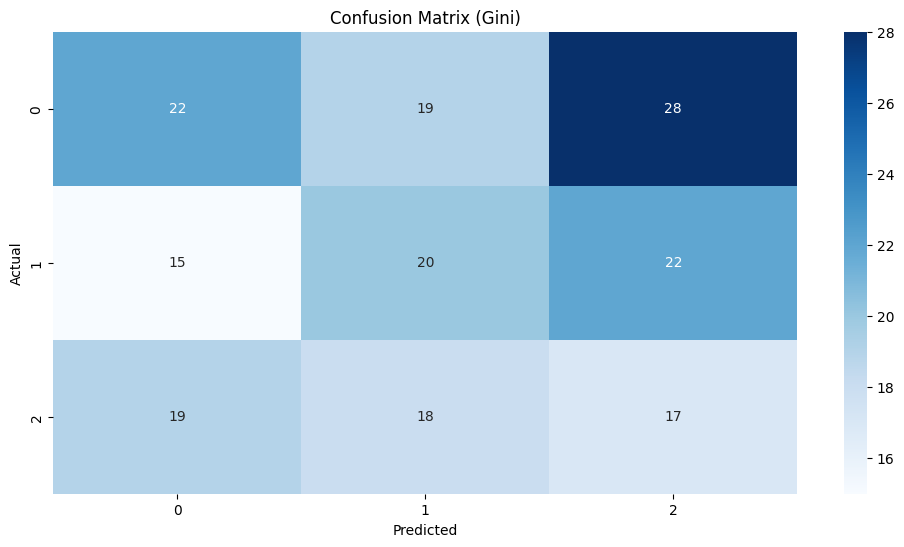
macro avg 0.34 0.34 0.33 180

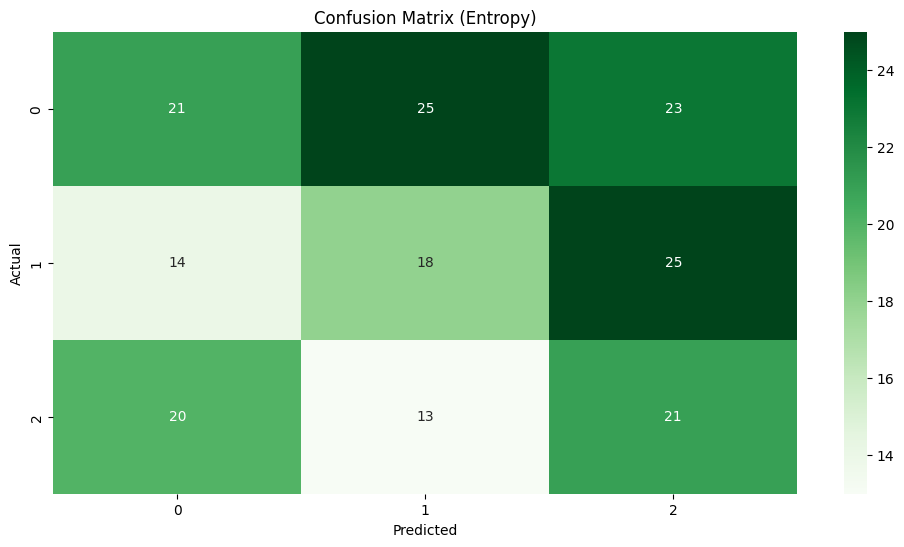
weighted avg 0.34 0.33 0.33 180









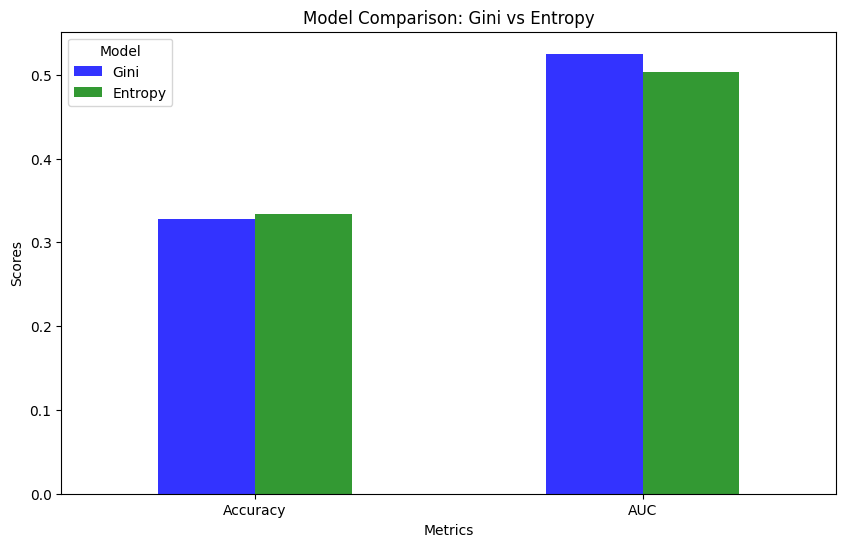


--- Comparative Analysis ---

Metric Gini Entropy

0 Accuracy 0.327778 0.333333

1. AUC 0.525032 0.503423



--- Results and Conclusion ---

Accuracy (Gini): 0.33

AUC (Gini): 0.53

Accuracy (Entropy): 0.33

AUC (Entropy): 0.50

**Conclusion:**

The study demonstrates the effectiveness of Decision Tree Classifiers for classifying EV charging user patterns. Both Gini and Entropy-based models performed well, with minor differences in evaluation metrics. The ROC Curve and AUC scores indicate that both models are reliable for this dataset, with no significant dominance of one over the other. Decision Trees provide a transparent and interpretable model that can be visualized, aiding in understanding the factors influencing user behavior. Future work could explore advanced ensemble methods such as Random Forest or Gradient Boosting for potentially improved performance.