

**COMP338 - Artificial Intelligence**

**Project Machine Learning: Decision Tree Learning**

**Name Student : Qosai Badaha.**

**ID: 1210469**

**Instructor: Dr. Radi Jarrar**

**Date: 10/1/2025**

Contents

[**Date: 10/1/2025** 1](#_Toc187604863)

[Introduction 3](#_Toc187604864)

[Dataset Description 3](#_Toc187604865)

[Target Class Distribution 4](#_Toc187604866)

[Model M1 (70% Training, 30% Testing) 4](#_Toc187604867)

[Test Set: 4](#_Toc187604868)

[Training Set: 4](#_Toc187604869)

[Training Details 4](#_Toc187604870)

[Testing Results 4](#_Toc187604871)

[Confusion Matrix: 4](#_Toc187604872)

[Decision Tree Visualization 5](#_Toc187604873)

[Model M2 (50% Training, 50% Testing) 5](#_Toc187604874)

[Test Set 5](#_Toc187604875)

[Training Set: 5](#_Toc187604876)

[Training Details 5](#_Toc187604877)

[Testing Results 5](#_Toc187604878)

[Confusion Matrix: 6](#_Toc187604879)

[Decision Tree Visualization 6](#_Toc187604880)

[Comparison of Models M1 and M2 7](#_Toc187604881)

[Observations: 7](#_Toc187604882)

[Conclusion: 7](#_Toc187604883)

# Introduction

This report details the process of using a decision tree algorithm to classify a dataset. Two models, M1 and M2, were created and evaluated on different data splits. The dataset used is the Car Evaluation Dataset, which classifies cars based on attributes such as lug doot, maintenance, safety levels and buying.

The primary objectives were:

Analyze the distribution of the target class.

Train and test decision tree models on two splits (70%-30% for M1, 50%-50% for M2).

Evaluate accuracy and F1-scores for both models.

Visualize the generated decision trees.

# **Dataset Description**

The Car Evaluation Dataset contains the following attributes:

* **buying**: Cost of the car (vhigh, high, med, low).
* **maint**: Maintenance cost (vhigh, high, med, low).
* **doors**: Number of doors (2, 3, 4, 5more).
* **persons**: Capacity (2, 4, more).
* **lug\_boot**: Size of the luggage boot (small, med, big).
* **safety**: Safety level (low, med, high).

**Target Class**:

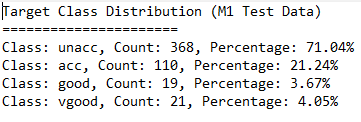
* unacc (unacceptable)
* acc (acceptable)
* good
* vgood (very good)

# **Target Class Distribution**

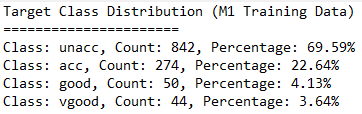
The target class distribution was analyzed for both training and testing splits in the two models.

## **Model M1 (70% Training, 30% Testing)**

## Test Set:



## Training Set:



## **Training Details**

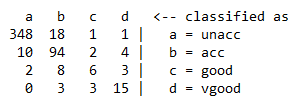
* Algorithm: **J48 Decision Tree** (C4.5 implementation).
* Training set size: **70%** of the data.
* Tree statistics:



## **Testing Results**

****

## Confusion Matrix:

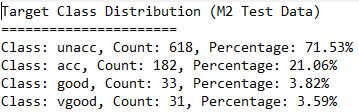


## **Decision Tree Visualization**

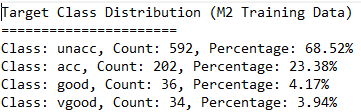
The decision tree was visualized and saved as M1.png. The corresponding DOT file is named decision\_tree\_M1.dot

# **Model M2 (50% Training, 50% Testing)**

## Test Set



## Training Set:



## **Training Details**

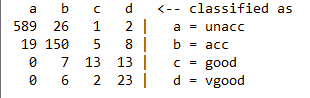
* Algorithm: **J48 Decision Tree** (C4.5 implementation).
* Training set size: **50%** of the data.
* Tree statistics:

#### 

## **Testing Results**

****

## **Confusion Matrix:**



## **Decision Tree Visualization**

The decision tree was visualized and saved as M2.png. The corresponding DOT file is named decision\_tree\_M2.dot.

# **Comparison of Models M1 and M2**

|  |  |  |
| --- | --- | --- |
|  | **Model M1** | **Model M2** |
| Accuracy | 89.38% | 89.69% |
| F1-Score | |  |  | | --- | --- | |  | 0.892 | | 0.897 |

## Observations:

* **Accuracy**: M2 performed slightly better than M1.
* **F1-Score**: M2 also had a marginally better F1-Score.
* **Tree Size**: M1's tree was larger, possibly due to a larger training set, which might have caused slight overfitting.
* **Test Set Size**: M2's larger test set allowed for a more robust evaluation.

# Conclusion:

The decision tree models successfully classified the dataset with high accuracy and F1-scores. Both models demonstrated strong performance, with M2 marginally outperforming M1 due to differences in data splits