**Problem 1:**



Table 1: Sowing Tree performance before and after Clustering

Problem 2:

**Data Set Name: CAR**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 1728 | **Area:** | N/A |
| **Attribute Characteristics:** | Categorical | **Number of Attributes:** | 6 | **Date Donated** | 1997-06-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | No | **Number of Web Hits:** | 612695 |

**Attribute Information:**

Class Values:   
unacc, acc, good, vgood   
Attributes:   
buying: vhigh, high, med, low.   
maint: vhigh, high, med, low.   
doors: 2, 3, 4, 5more.   
persons: 2, 4, more.   
lug\_boot: small, med, big.   
safety: low, med, high.

Step 1: Generate Decision Tree Using J48 without ACC class value

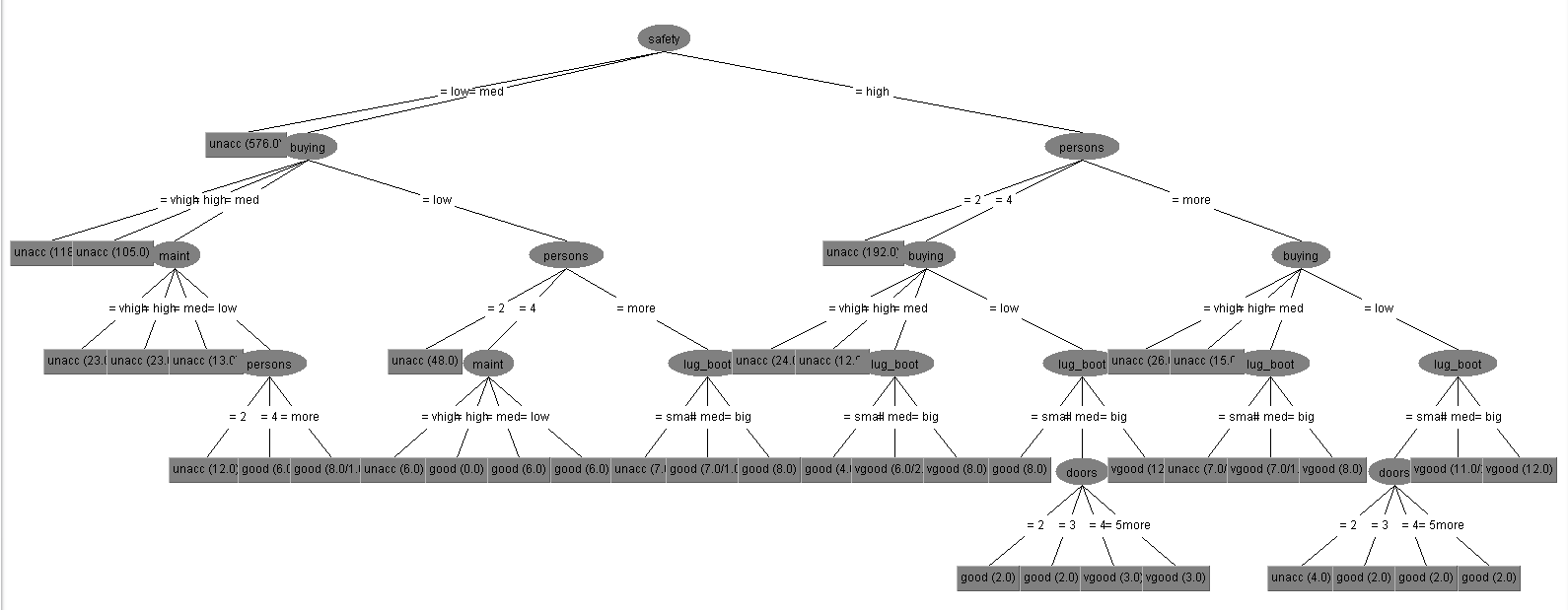


Figure 1: Decision tree 1

Number of Leaves: **40**

Size of the tree: **56**

Correctly Classified **98.8095 %**

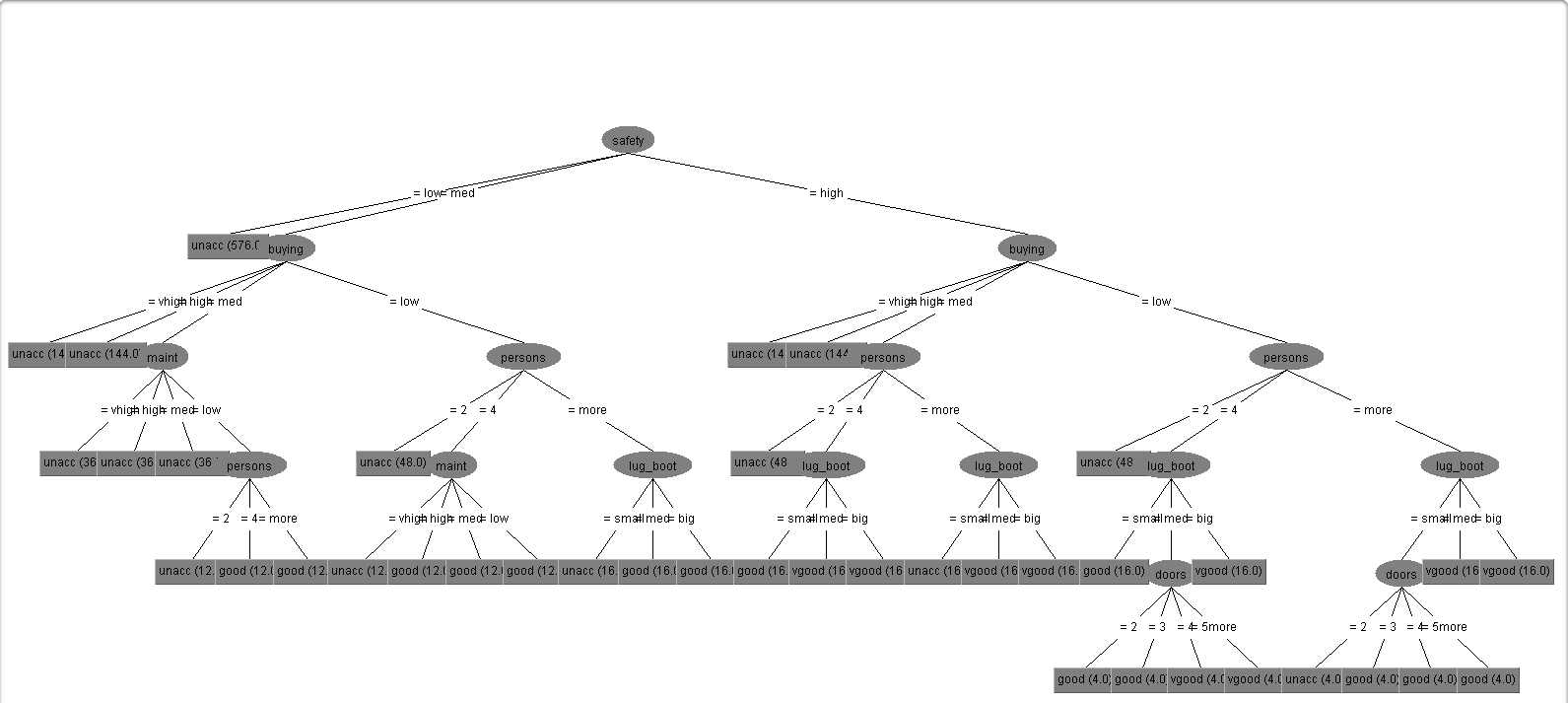
Step 2: Use Test Set for testing data and find the new class value and generate new tree 

Figure 1: Decision tree 2

Number of Leaves: **39**

Size of the tree: **55**

Correctly Classified **99.8843 %**

**Comment:** This technique is working on my selected dataset. It reduces Leaves and Size of tree, also increase accuracy.

I write a program using Weka API that classify test data and save as a new dataset. Code is available in GitHub1

[1] https://github.com/amitbd1508/J48-Using\_Weka