April 29, 2021

REPORT 3
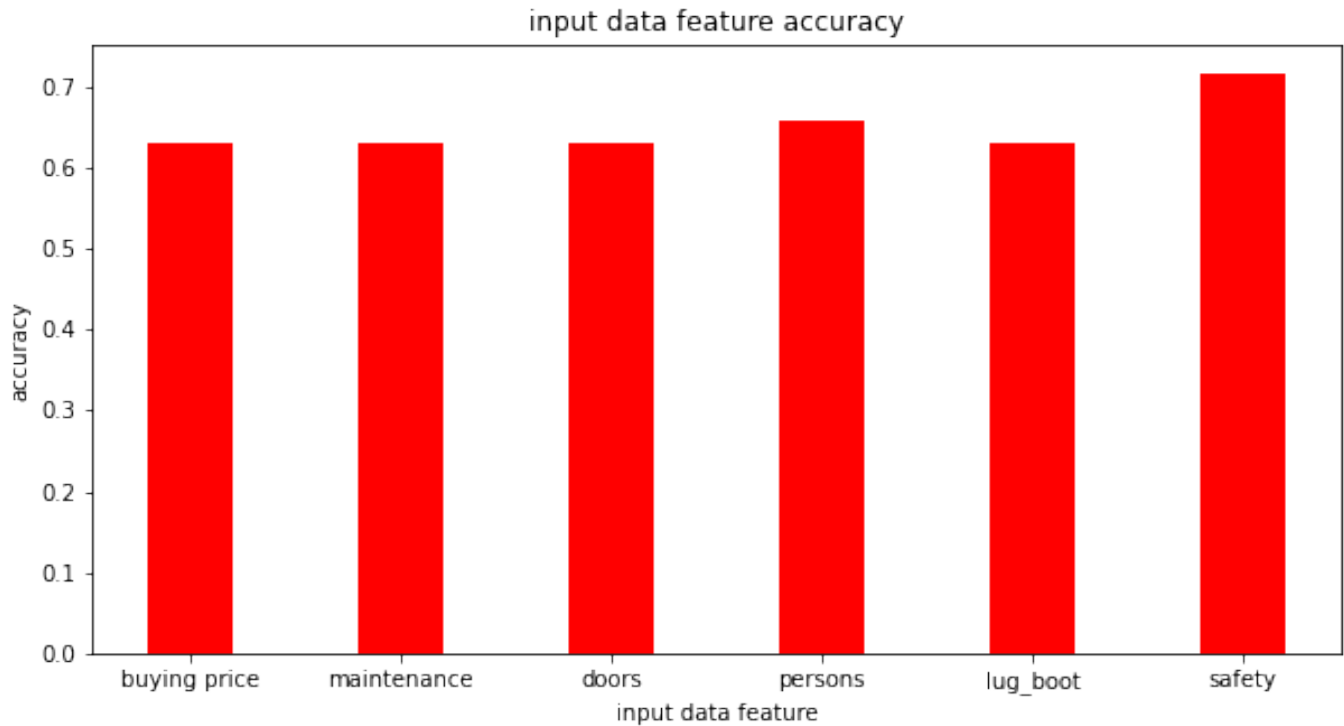
Steven Guo

cs 4300

University of Missouri-St. Louis

Click Here For Google Colab Source

1 Feature Importance

This is accuracy of all features of the dataset.

buying price rating-63.14
maintenance rating-63.14
number of doors-63.14
number of people-65.70
lug boot rating-63.14
safety rating-71.57

From what I can tell, buying price, maintenance, number of doors, and lug boot rating are all the same accuracy. They also have the lowest accuracy and I will be removing these features top to bottom. Then I will remove features with higher accuracy to try to study feature importance. The accuracy on these input features are very close together so there may not be any features to remove. The process of removing input features is this:

buying price-¿maintenance-¿number of doors-¿ lug boot rating-¿number of people-¿ safety rating

2 Removing Input Features

Base Case

```
Accuracy: 87.52%
Precision: 83.45%
Recall: 82.51%
F1-score: 0.83
```

First Input removed-buying price rating

```
Accuracy: 86.69%
Precision: 82.91%
Recall: 80.49%
F1-score: 0.82
```

Second Input removed-maintenance rating

```
Accuracy: 86.45%
Precision: 82.34%
Recall: 80.49%
F1-score: 0.81
```

Third Input removed-number of doors

```
Accuracy: 88.10%
Precision: 84.63%
Recall: 82.74%
F1-score: 0.84
```

Fourth Input removed-number of people

```
Accuracy: 72.89%
Precision: 63.47%
Recall: 62.33%
F1-score: 0.63
```

Fifth Input removed-lug boot rating

```
Accuracy: 87.52%
Precision: 83.00%
Recall: 83.18%
F1-score: 0.83
```

Sixth Input removed-safety rating

```
Accuracy: 70.33%
Precision: 60.43%
Recall: 56.50%
F1-score: 0.58
```

Results

I tested the accuracy of removing an input feature from the base dataset by removing the input column and then finding dataset's accuracy. I did this with each input feature just to test the accuracy and to see if it matches the input data feature accuracy. Unfortunately this was not the case. The lowest accuracy was the buying price according to the table, but when testing the input feature removal on the dataset the lowest accuracy was the safety rating.

3   Iteratively remove one feature at a time

buying price

```
Accuracy: 86.03%
Precision: 82.13%
Recall: 79.37%
F1-score: 0.81
```

buying price,maintenance

```
Accuracy: 87.69%
Precision: 84.94%
Recall: 80.94%
F1-score: 0.83
```

buying price,maintenance,number of doors

```
Accuracy: 85.95%
Precision: 81.65%
Recall: 79.82%
F1-score: 0.81
```

buying price,maintenance,number of doors,lug boot rating

```
Accuracy: 88.18%
Precision: 85.82%
Recall: 81.39%
F1-score: 0.84
```

buying price,maintenance,number of doors,lug boot rating,number of people

```
Accuracy: 86.86%
Precision: 83.76%
Recall: 79.82%
F1-score: 0.82
```

buying price,maintenance,number of doors,lug boot rating,number of people,safety rating

```
Accuracy: 87.77%
Precision: 85.99%
Recall: 79.82%
F1-score: 0.83
```

Results

The highest accuracy achieved was 88.18 percent and that was when buying price, maintenance, number of doors, and lug boot rating input features were removed. The lowest accuracy achieved was 85.95 and that was when buying price and maintenance was removed. This is