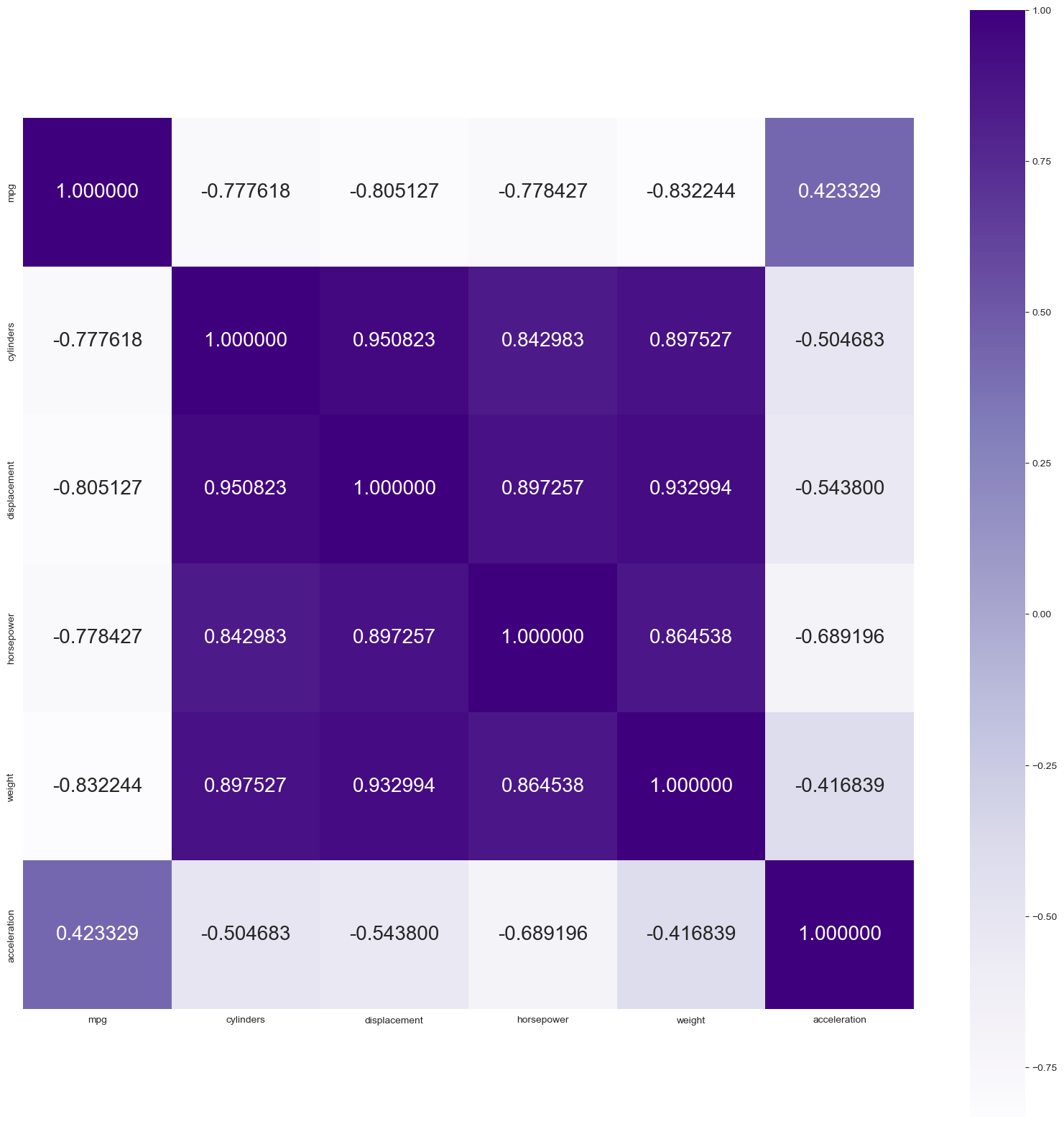
Machine Learning

Term Project

Alpay Kurt 190303020 – Samet Şenol 190303013

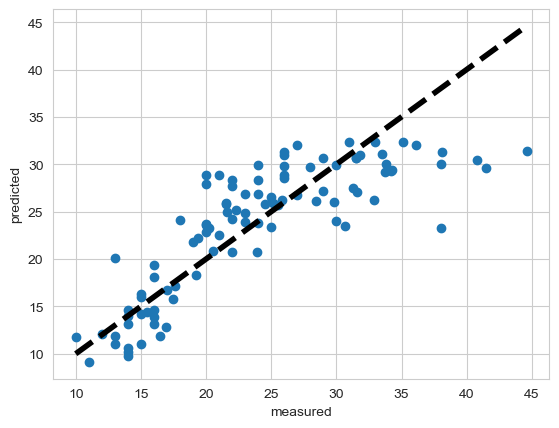
In this project, the code uses a classifier trained using the MPG.csv (Miles Per Gallon).  
The goal is to estimate which features are important for car gas and how much important they are.

In this project, we are going to show the performance metric(such as, correlation coefficient, Interception, Confusion Matrix, Accuracy Score, Precision score, R2 Score) results and related graphics for every method.

**Correlation:**our correlationshape is (6,6) and our heat map is like:

|  |  |  |
| --- | --- | --- |
| **Method** | **Precision score** | **Accuracy Score** |
| Multiple Linear Regression | 78.57% | 74.63% |
| Logistic Regression | 84.69% | 81.78% |
| Polynomials Regression | 78.57% | 74.63% |
| K-NN | 84.69% | 81.69% |
| Naive Bayes | 72.44% | 72.87% |
| Decision Trees | 83.67% | 80.31% |
| Random Forest | 84.39% | 81.39% |
| Support Vector Machine | 83.67% | 80.69% |
| K-means clustering | 89.54% | 90.04% |
| Hierarchical clustering | 45.40% | 30.16% |

***Multiple Linear Regression:***

****

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

coefficient:

[-0.31162377 -0.00470112 -0.02647071 -0.00552154 0.08056846]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Interception:

43.94699577982187

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Confusion Matrix:

[[56 8]

[13 21]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

78.57142857142857 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

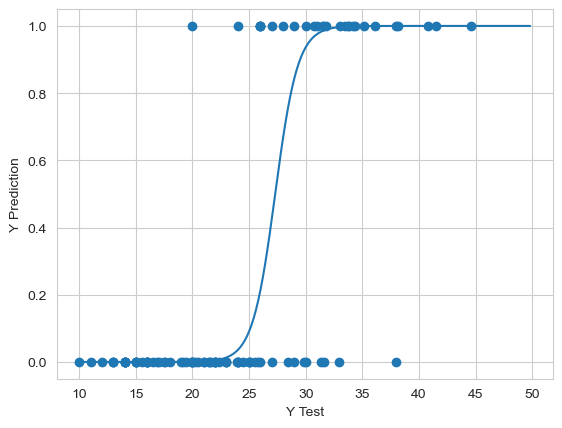
74.63235294117648 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

-0.02848575712143897

***Logistic Regression:***

****

Confusion Matric:

[[62 7]

[ 8 21]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

84.6938775510204 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

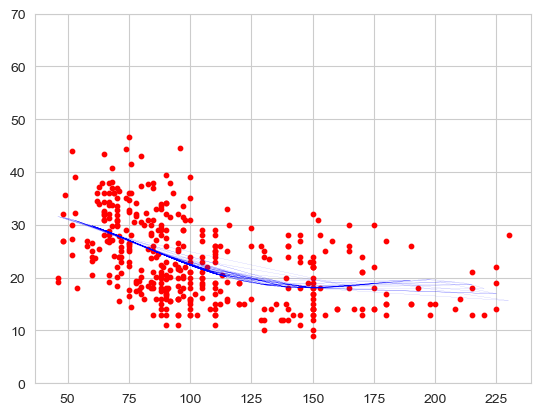
81.78571428571428 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R^2 Score:

0.2653673163418292

***Polynomial Regression:***

******

Confusion Metric:

[[56 13]

[ 8 21]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

78.57142857142857 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

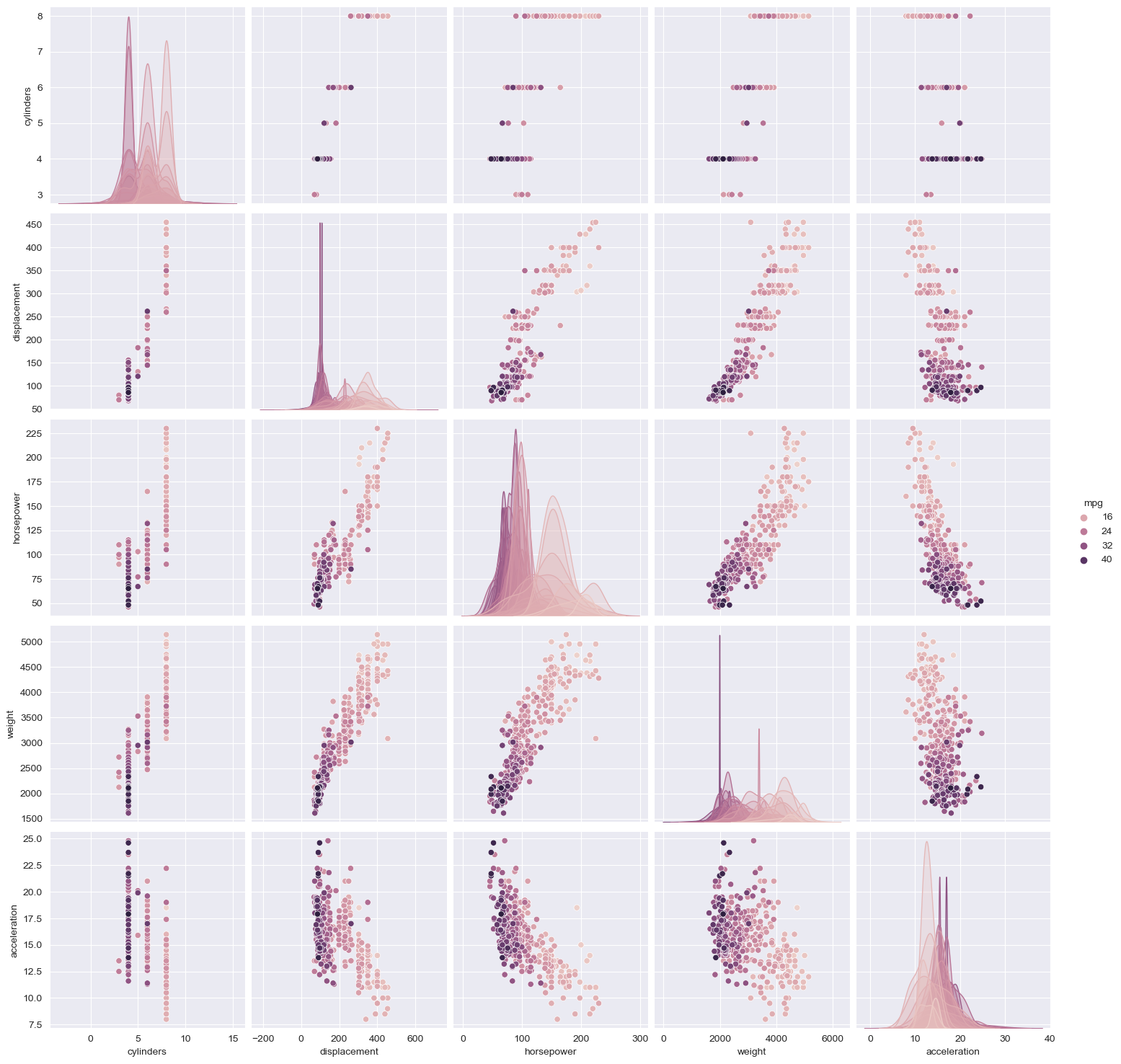
Precision score:

74.63235294117648 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

-0.02848575712143897

***K-NN:***

Confusion Matric:

[[45 24]

[ 3 26]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

72.44897959183673 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

72.875 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

-0.32233883058470747

***Naïve-Bayes:***

Confusion Matric:

[[45 24]

[ 3 26]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

72.44897959183673 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

72.875 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

-0.32233883058470747

***Decision Tree:***

Confusion Matric:

[[58 11]

[ 5 24]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

83.6734693877551 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

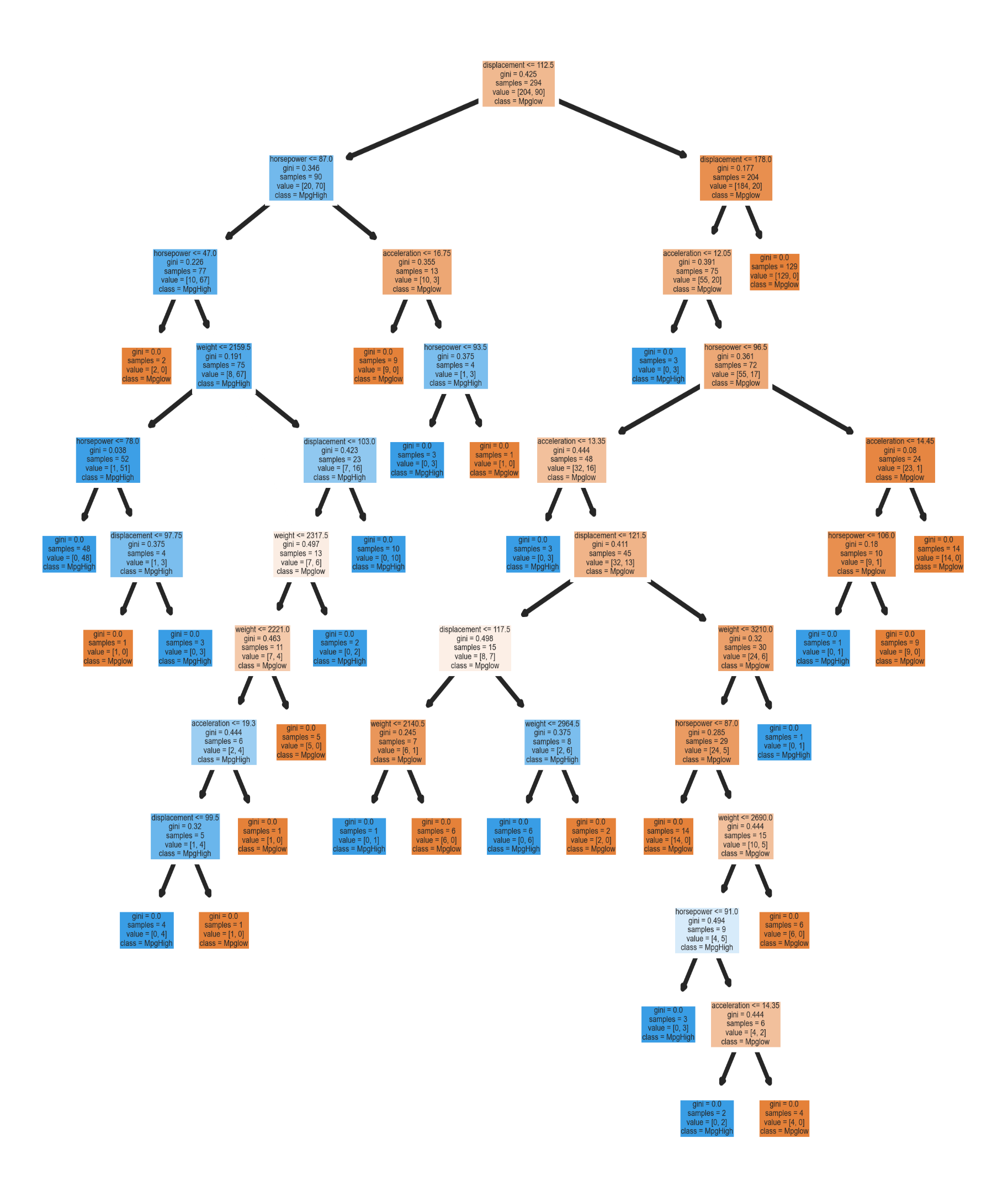
Precision score:

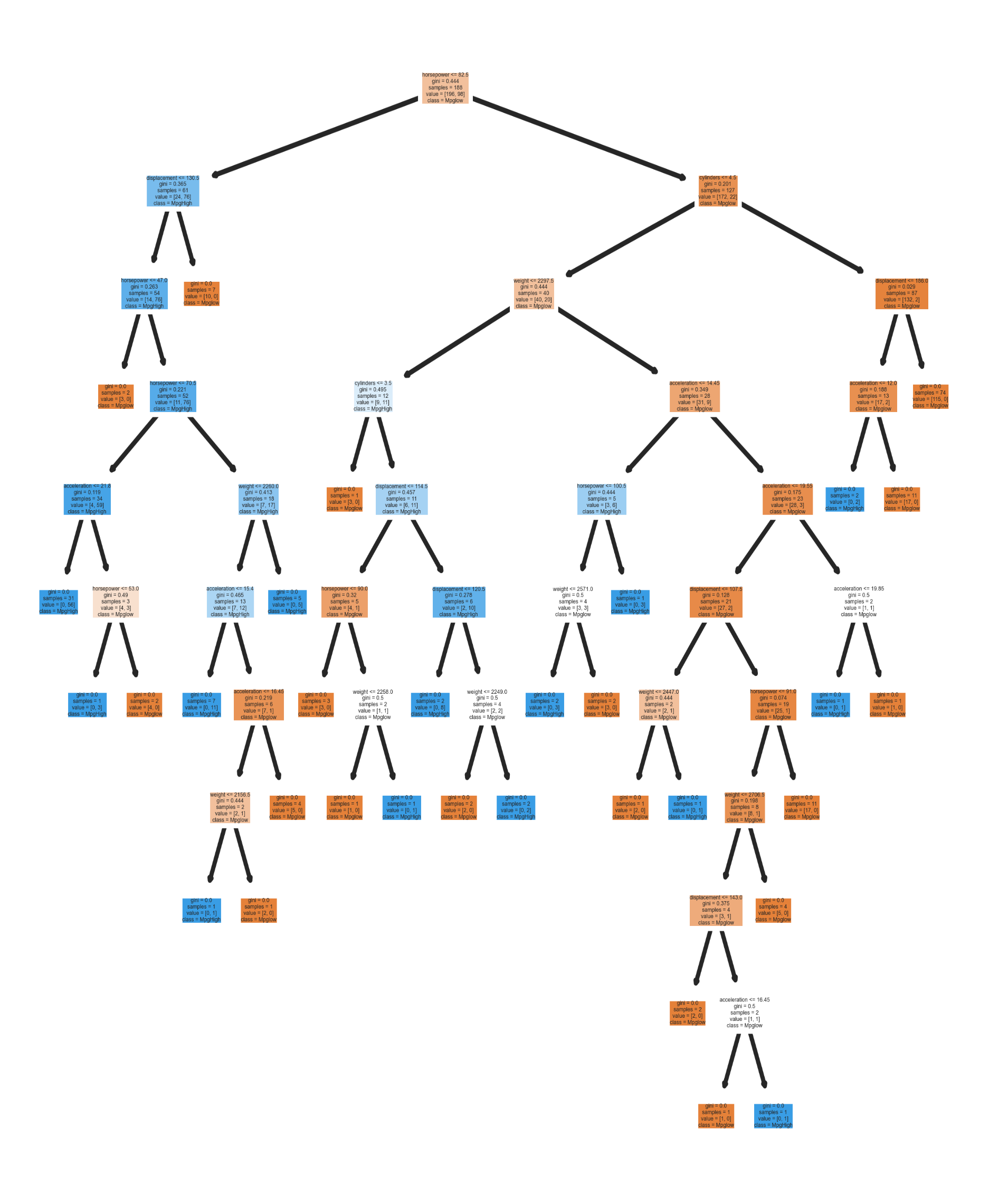
80.31746031746032 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

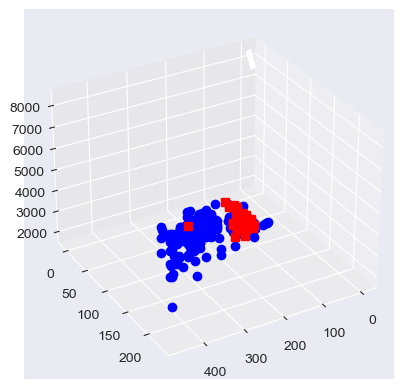
R2 Score:

0.21639180409795122



***Random Forest:***

***SVM (Support Vector Machine):***

****

Confusion Metrics:

[[62 7]

[ 9 20]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

83.6734693877551 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

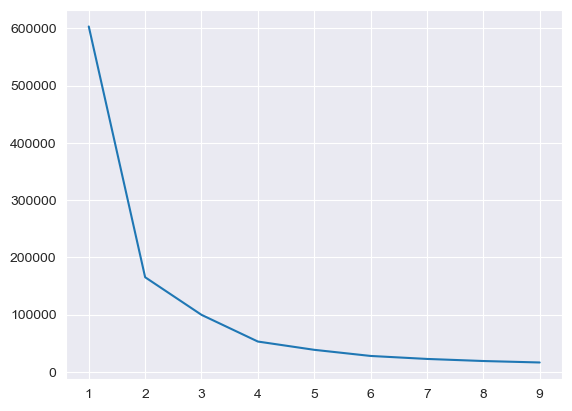
80.69900886802294 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

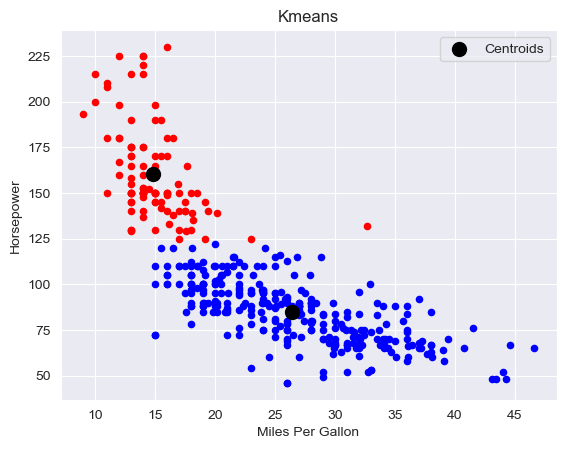
R2 Score:

0.21639180409795122

***K-Means Clustering:***

******

**As you can see from elbow, we definitely should use 2 clustering**

****

Confusion Metrics:

[[ 92 32]

[ 9 259]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

89.54081632653062 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

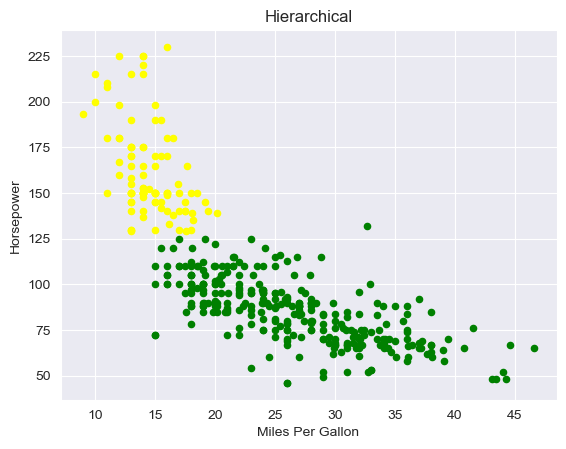
90.04627266850396 %

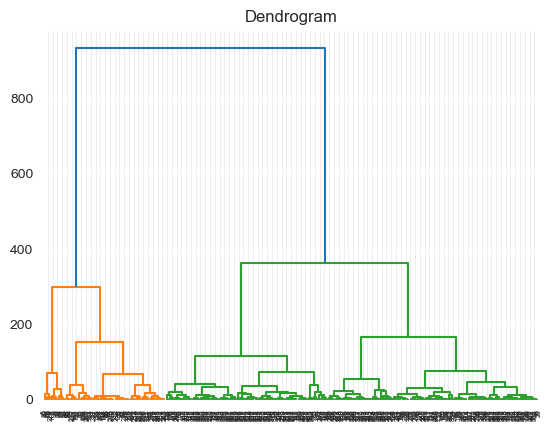
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

0.5163697640828117

***Hierarchical Clustering:***

******

******

Confusion Metrics:

[[178 97]

[117 0]]

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Accuracy Score:

45.40816326530612 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Precision score:

30.16949152542373 %

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 Score:

-1.6072416472416475