**Digital Assessment 2**

**DATA MINING AND ANALYSIS**

Date: 15 September, 2024

Name: Anuj Parihar

Registration Number: 21BBS0162

**Link to Assessment Codebase and Dataset**:

<https://github.com/BearTS/data-mining-assignments/tree/main/Lab/DA%202>

**Question 1.**

Collect the data set consists of 50 observations about patient enrolment in diet maintenance based on gender, weight, BMI etc (minimum 7 features). Implement a model that will recommend a strict diet is necessary or not for a patient using the naïve Bayes classification algorithm. (50x7)

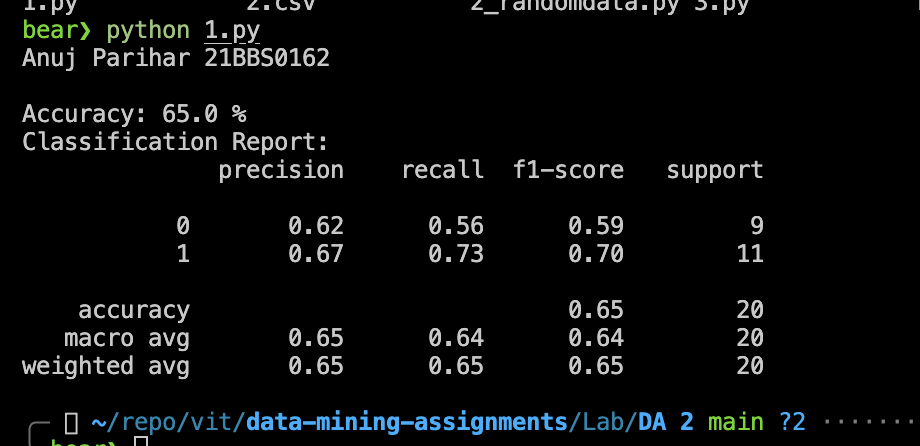
*Aim*

To implement the Naïve Bayes Classification using a diet maintenance dataset and to find out the performance of the model.

*Sample Input***:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | Age | Weight | Height | BMI | Cholesterol | PhysicalActivity | StrictDietRequired |
| Male | 37 | 77 | 156 | 30 | Normal | High | Yes |
| Female | 45 | 93 | 165 | 30.1 | High | Low | Yes |
| Male | 63 | 93 | 175 | 24.4 | High | High | Yes |
| Male | 53 | 69 | 151 | 23.3 | Normal | Low | No |
| Male | 29 | 79 | 150 | 31.9 | High | Low | Yes |
| Female | 55 | 60 | 161 | 31.9 | Normal | High | Yes |
| Male | 33 | 77 | 154 | 32.8 | Normal | Low | Yes |
| Male | 50 | 74 | 186 | 33.6 | High | High | Yes |
| Male | 34 | 88 | 181 | 26.9 | Normal | Low | No |
| Female | 27 | 82 | 158 | 26.8 | High | Low | No |
| Male | 33 | 50 | 190 | 31.7 | High | Moderate | Yes |
| Male | 42 | 76 | 184 | 29.2 | Normal | Moderate | No |
| Male | 59 | 62 | 168 | 30.1 | Normal | High | Yes |

*Output***:**

****

**Question 2:**

Implement K-means method of clustering. Use the patient details data set to classify into 3

clusters such as a person is normal, healthy and weak. A person/patient must be clustered as

any one of normal/healthy or weak based on his/her input values. (100 rows)

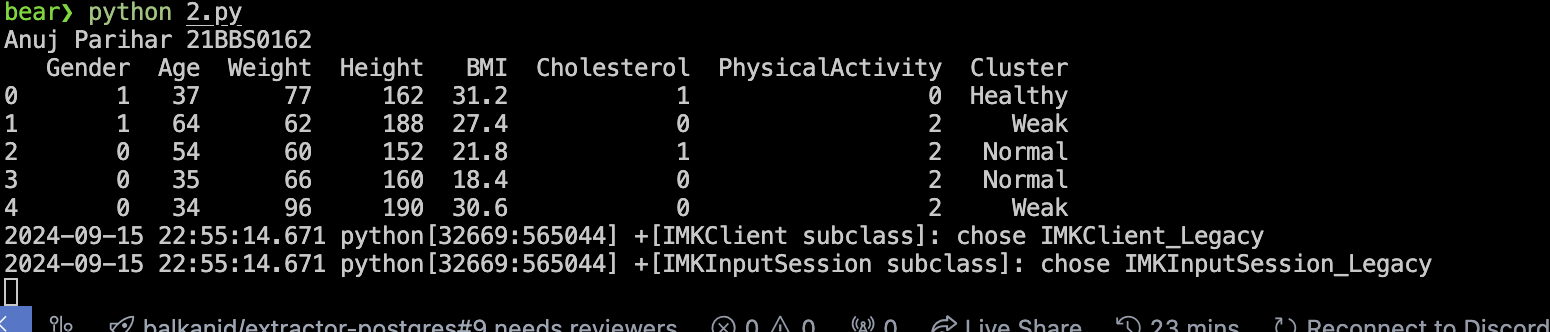
*Aim*

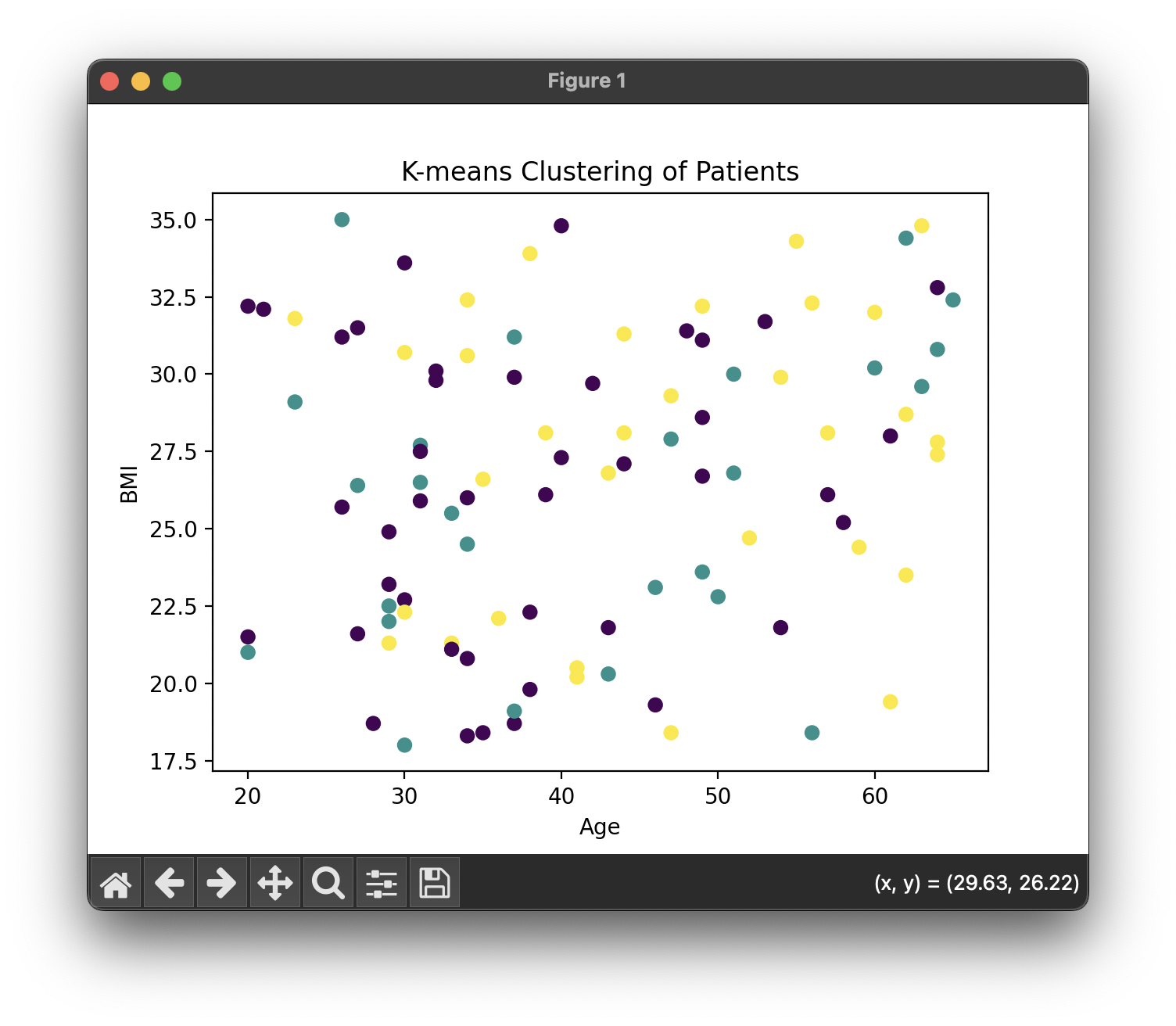
To cluster patient details using K-means clustering with k=3. The clusters are normal, healthy or weak

*Sample Input*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | Age | Weight | Height | BMI | Cholesterol | PhysicalActivity | Cluster |
| 1 | 54 | 88 | 154 | 28.2 | 1 | 1 | Weak |
| 0 | 32 | 85 | 154 | 23.2 | 0 | 1 | Weak |
| 0 | 50 | 67 | 182 | 18.7 | 0 | 2 | Healthy |
| 1 | 25 | 72 | 164 | 33.2 | 0 | 1 | Normal |
| 1 | 27 | 89 | 179 | 27.2 | 1 | 2 | Weak |
| 1 | 32 | 60 | 164 | 31.8 | 1 | 2 | Normal |
| 1 | 35 | 80 | 157 | 19.7 | 0 | 2 | Normal |
| 0 | 56 | 52 | 151 | 18.3 | 1 | 0 | Normal |
| 1 | 63 | 93 | 179 | 29.3 | 1 | 2 | Weak |
| 1 | 61 | 65 | 186 | 31.7 | 0 | 1 | Healthy |
| 1 | 57 | 64 | 175 | 34.5 | 1 | 2 | Healthy |
| 1 | 45 | 83 | 158 | 20.9 | 1 | 2 | Weak |
| 1 | 63 | 92 | 187 | 25 | 1 | 0 | Weak |
| 1 | 25 | 66 | 169 | 20.4 | 0 | 0 | Normal |
| 1 | 23 | 95 | 187 | 31.6 | 0 | 2 | Weak |
| 1 | 52 | 92 | 172 | 22 | 1 | 1 | Weak |
| 0 | 27 | 98 | 162 | 21.5 | 1 | 0 | Weak |

*Output:*





*Results:*

The data set has been clustered into 3 groups namely normal, healthy and weak and it has

been used to predict and cluster future data

**Question 3**

The ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking. Consider a dataset of 50 rows “Road transport records” with the attributes “Road ID”, “Length”, Numberof\_Bends”, “Trafficvolume” and “AccidentRisk”. Implement the same to the dataset to recommend the decision tree to classify the data

*Aim:*

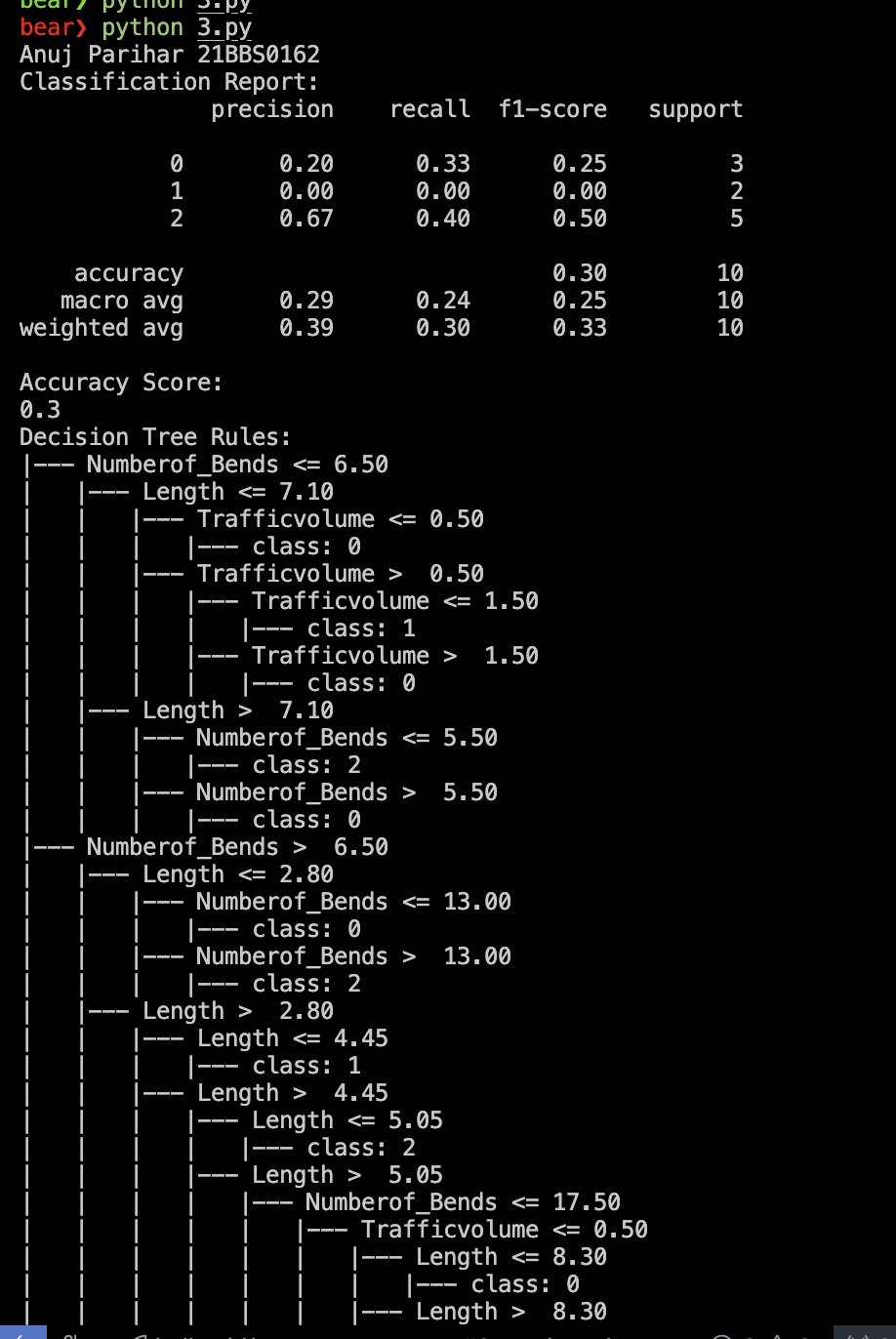
To build a decision tree for road transport records with various attributes for classification,

hence measuring its performance and building the tree

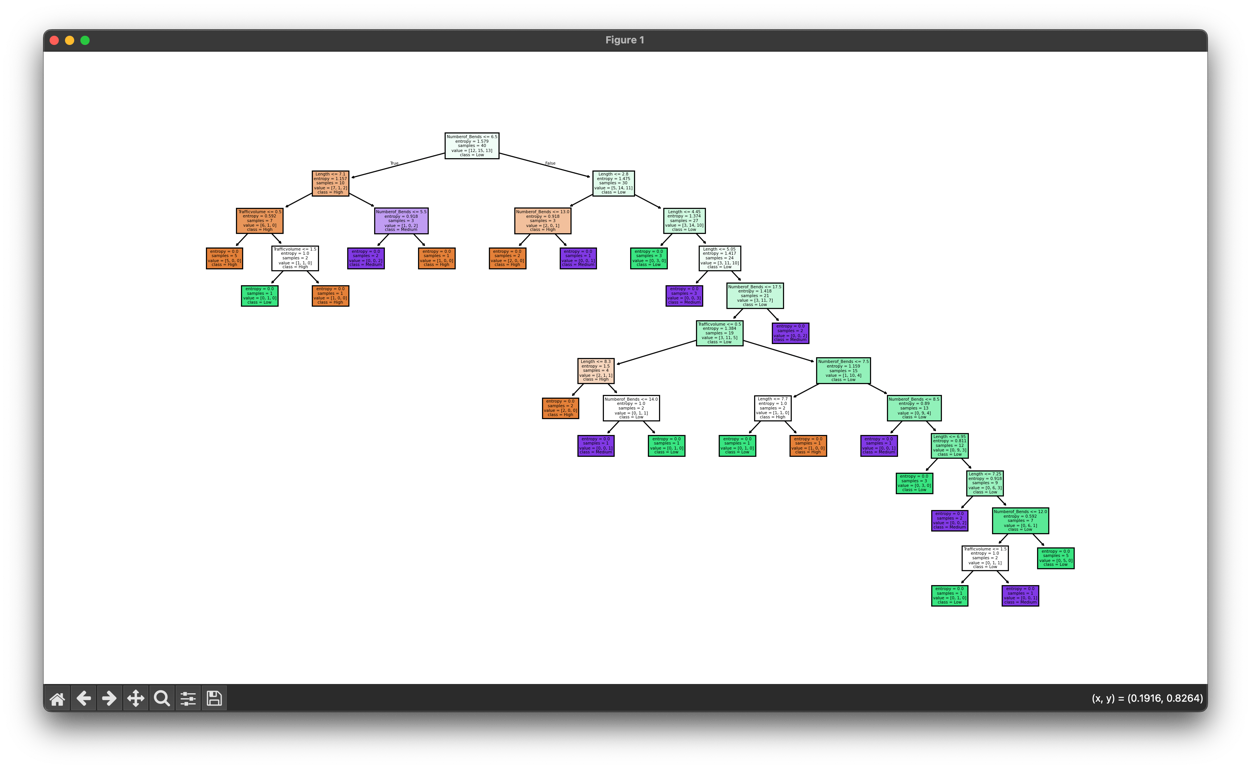
*Sample Input:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Road ID | Length | Numberof\_Bends | Trafficvolume | AccidentRisk |
| 101 | 9.8 | 5 | High | Medium |
| 102 | 8.6 | 17 | Medium | Low |
| 103 | 6.7 | 14 | Medium | Low |
| 104 | 8.5 | 17 | Medium | Low |
| 105 | 9.1 | 7 | Low | High |
| 106 | 2.8 | 3 | Low | Low |
| 107 | 9.9 | 11 | High | Medium |
| 108 | 6.2 | 8 | Low | Medium |
| 109 | 8.1 | 14 | Low | Low |
| 110 | 4.8 | 9 | Medium | Medium |
| 111 | 2.2 | 15 | Low | Medium |
| 112 | 6.9 | 9 | Low | Low |
| 113 | 3.6 | 9 | Low | Low |
| 114 | 6.7 | 19 | Medium | Medium |
| 115 | 7.5 | 13 | Low | Low |
| 116 | 6.3 | 7 | Low | Low |
| 117 | 6.1 | 6 | High | High |
| 118 | 3.1 | 16 | Low | Medium |
| 119 | 7 | 10 | Medium | Medium |

*Output:*

**

**

**

*Result:*

The decision tree is built and evaluated successfully.