**Instructions:**

**There are a total of 6 (six) multi-part questions, with point values noted for each question.**

**If necessary, delete any missing values from the datasets**

**Please show your calculations, or the details of your program(s), for each problem. Your program(s) should be commented so that each step is clearly explained.**

**Combine all of your answers/files into a single zipped file and post the zipped file to CANVAS.**

**Problems #1** **(20 points)**

**A researcher is analyzing a dataset on heart health, referred to as the "Heart Problem" dataset. The dataset includes patient counts categorized by age (divided into quartiles) and gender, along with their heart health status (either "Problem" or "No Problem"). The table below summarizes the data:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Quartile** | **Gender** | **No Problem** | **Problem** |
| Quartile 1 | Female | 21 | 1 |
|  | Male | 36 | 23 |
| Quartile 2 | Female | 20 | 5 |
|  | Male | 29 | 26 |
| Quartile 3 | Female | 9 | 8 |
|  | Male | 14 | 37 |
| Quartile 4 | Female | 22 | 10 |
|  | Male | 12 | 27 |
| Row Total |  | 163 | 137 |

**Use the table above and Excel to classify patients with heart problems using Age Category and Gender:**

**Construct a C4.5 decision tree (one level only).**

**Must use Excel**

**Problems #2 (15 points)**

**Load the “UCI\_heart.CSV” dataset, from the raw\_data module in CANVAS, into Python (see the data dictionary at the bottom of this document). This is a dataset used for predicting “Heart\_problems” in patients. Construct a C50 model to classify Heart\_problems based only on** **the “sex”, “fbs”, “restecg”, “exang” and “Age\_quantile” variables and 25%-75% test-training splits**

**Problems #3 (15 points)**

**Load the “UCI\_heart.CSV” dataset, from the raw\_data module in CANVAS, into Python (see the data dictionary at the bottom of this document). This is a dataset used for predicting “Heart\_problems” in patients. Construct a Random Forest model to classify Heart\_problems based only on the “sex”, “fbs”, “restecg”, “exang” and “Age\_quantile” attributes** **and 25%-75% test-training splits**

**Problem #4 (15 points)**

**Use python/R to cluster (Algorithm=K-means; K=2) patients in the** **“UCI\_heart.csv” dataset using “testpbs”, “chol”, “thalach”,”exang”, attributes only.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

1. **What are the members of each cluster?**
2. **What are the coordinates for the cluster centers?**

**Problem #5 (15 points)**

**Use python/R to cluster(Algorithm=hierarchical; two clusters) patients in the “UCI\_heart.csv” dataset using “testpbs”, “chol”, “thalach”,”exang”, attributes only.**

1. **What are the members of each cluster?**
2. **What are the coordinates for the cluster centers?**

**Problem #6 (20 points)**

**Using data in the table below and Excel, construct a Neural Network with one Output Layer (z) and one Hidden Layer (A and B).** (20 points)

1. **Calculate the predicted outcome if the inputs to the input nodes are (x=1, Node 1=.4, Node 2=.8 Node 3= .8 and Node 4=.2).**
2. **Adjust the weight if the actual output is 0.8500**

**Must use Excel**

|  |  |  |
| --- | --- | --- |
| **From** | **To** | **Weight** |
| X | A | 0.5 |
| Node 1 | A | 0.6 |
| Node 2 | A | 0.8 |
| Node 3 | A | 0.6 |
| Node 4 | A | 0.2 |
| x | B | 0.7 |
| Node 1 | B | 0.9 |
| Node 2 | B | 0.8 |
| Node 3 | B | 0.4 |
| Node 4 | B | 0.2 |
| xx | z | 0.5 |
| A | z | 0.9 |
| B | z | 0.9 |

**Data Dictionary**

|  |  |  |  |
| --- | --- | --- | --- |
| age | Feature | Integer | Age |
| sex | Feature | Categorical | Gender |
| trestbps | Feature | Integer | resting blood pressure (on admission to the hospital) |
| chol | Feature | Integer | serum cholesterol |
| fbs | Feature | Categorical | fasting blood sugar > 120 mg/dl |
| restecg | Feature | Categorical | Resting electrocardiogram |
| thalach | Feature | Integer | maximum heart rate achieved |
| exang | Feature | Categorical | exercise induced angina |
| Heart\_problem | Target | Integer | diagnosis of heart disease |
| Age\_quantile | Feature | Categorical | Age quartile |