**Names:** Abdallah Mustafa - ID 3730765

Mohammad Mahboub - ID

**Course:** Machine Learning & Data mining

**Topic:** Programming Project Report

**Date:** 25/10/2022

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

**Used Language: Python**

**External Libraires: None**

**Preprocessing Steps:**

* **Read data sets from the files and store the data in two dimensional arrays.**
* **Remove all the lines that contain missing values.**
* **Move the 1st column to last column for Mushroom and letter data sets.**

**1st Algorithm (Decision Tree)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Name** | **Mushroom** | **Car** | **Ecoli** | **Letter** | **Cancer** |
| **Average accuracy** | **98.82%** | **74.53%** |  |  |  |
| **Standard deviation** |  |  |  |  |  |

Code Implementation:

**2nd Algorithm (Boost ID3 Tree)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Name** | **Mushroom** | **Car** | **Ecoli** | **letter** | **cancer** |
| **Average accuracy** | **50.08%** | **46.7%** |  |  |  |
| **Standard deviation** |  |  |  |  |  |

**3rd Algorithm (Neural Network)**

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Name** | **Mushroom** | **Car** | **Ecoli** | **letter** | **cancer** |
| **Average accuracy** | **50.08%** | **46.7%** |  |  |  |
| **Standard deviation** |  |  |  |  |  |

**4th Algorithm (Naïve Bayes)**

**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Name** | **Mushroom** | **Car** | **Ecoli** | **letter** | **cancer** |
| **Average accuracy** | **50.08%** | **46.7%** |  |  |  |
| **Standard deviation** |  |  |  |  |  |