**ML Practicals**

Name: Sheetal Dunda Wanghare

Branch: Computer Engineering

Year: Third Year

PRN no. 2030331245049

**Practical no.1**

**Name of Algorithm**: Linear Regression

**Name of Dataset**: House pricing

• **No. of rows**: 200

• **No. of columns**: 2

• **Name of columns**: Area and Price

**Mean Square Error:** 1.4557113237678404e-28

**Mean Absolute Percentage Error** 1.5875984671137725e-16

**Accuracy**: 100.0

**Practical no.2**

**Name of Algorithm**: Linear Regression

**Name of Dataset**: Advertising

• **No. of rows**: 200

• **No. of columns**: 4

• **Name of columns**: TV,Radio,Newspaper and Sales

**Mean Square Error**: 3.938015220480285

**Mean Absolute Error**: 0.17222599691973425

**Accuracy**: 99.82777400308026

**Practical no.3**

**Name of Algorithm**: Decision tree

**Name of Dataset**: Diabetes

• **No. of rows**: 768

• **No. of columns**: 9

• **Name of columns**:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI,

DiabetesPedigreeFunction,Age,Outcome.

**Mean Square Error**: 0.2683982683982684

**Mean Absolute Error**: 0.2683982683982684

**Accuracy**: 72.72727272727273

**Confusion matrix**:

|  |  |
| --- | --- |
| 144 | 13 |
| 50 | 24 |

* **For Entropy:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Depth | 3 | 5 | 10 | 15 | 20 |
| Accuracy | 73.160 | 70.659 | 67.881 | 67.881 | 67.881 |

* **For Gini:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Depth | 3 | 5 | 10 | 15 | 20 |
| Accuracy | 72.727 | 72.395 | 70.659 | 71.006 | 71.006 |

**Practical no.4**

**Name of Algorithm**: Logistic Regression

**Name of Dataset**: Diabetes

• **No. of rows**: 768

• **No. of columns**: 9

• **Name of columns**:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI,

DiabetesPedigreeFunction,Age, Outcome.

**Mean Square Error**: 0.22077922077922077

**Mean Absolute Error**: 0.22077922077922077

**Accuracy**: 77.92207792207793

**Confusion Matrix**:

|  |  |
| --- | --- |
| 141 | 16 |
| 35 | 39 |

**Practical no.5**

**Name of Algorithm**: K-means Clustering algorithm

**Name of Dataset**: Diabetes

• **No. of rows**: 768

• **No. of columns**: 9

• **Name of columns**:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI,

DiabetesPedigreeFunction,Age, Outcome.

**Columns used**: Age and BloodPressure

**Optimal no. of k using elbow method**:3

**Practical no.6**

**Name of Algorithm**: Artificial Neural Network (ANN)

**Name of Dataset**: Diabetes

• **No. of rows**: 768

• **No. of columns**: 9

• **Name of columns**:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI,

DiabetesPedigreeFunction,Age, Outcome.

**Activation functions:**

* **1st Hidden layer:** Relu
* **2nd Hidden layer:** Relu
* **Output layer:** Sigmoid

**No.of epochs:** 200

**Training Accuracy:**77.84%

**Testing Accuracy:**65.80%

**Confusion Matrix:**

|  |  |
| --- | --- |
| 100 | 57 |
| 22 | 52 |

**Practical no.7**

**Name of Algorithm**: Hierarchical clustering

**Approach used:** Agglomerative clustering

**Name of Dataset**: Diabetes

• **No. of rows**: 768

• **No. of columns**: 9

• **Name of columns**:

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI,

DiabetesPedigreeFunction,Age, Outcome.

**Columns used**: Age and BloodPressure

**Optimal number of clusters using dendrogram**:3