Name: -Aarti Yogesh Pardeshi

PRN: -2030331245062

Subject: - Machine Learning

Class: -Third Year

Branch: - Computer Engineering

**Practical No.-1**

**Linear Regression**

**Name of Algorithm:** Linear Regression

**Name of Dataset:** House Pricing

**Dataset Description:**

The dataset is consisting of two columns named as Area and Price for house price prediction. The area is started with 200 sq. Ft and price is started with 10 lakhs.

* **No. of rows:** 200
* **No. of columns:** 2
* **Name of columns:** Area and Price

**Mean squared error:** 1.5866622340345284e-28

**Mean absolute percentage error:** 2.758283583646273e-16

**Accuracy:** 100.0

**Practical No.-2**

**Linear Regression**

**Name of Algorithm:** Linear Regression model

**Name of Dataset:** Advertising

**Dataset Description:**

* **No. of rows:** 200
* **No. of columns:** 4
* **Name of columns:** TV, Radio, Newspaper and Sales

**Mean squared error:** 3.938015220480285.

**Mean absolute percentage error:** 0.17222599691973425.

**Accuracy:** 99.82777400308026

**Practical No.-3**

**Decision Tree**

**Name of Algorithm:** Decision Tree Algorithm

**Name of Dataset:** Diabetics

**Dataset Description:**

* **No. of rows:** 768
* **No. of columns:** 9
* **Name of columns:** Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome

**Mean squared error:** 0.2683982683982684

**Mean absolute percentage error:** 0.2683982683982684

**Accuracy:** 73.16017316017316

**Confusion Matrix:**

|  |  |
| --- | --- |
| 145 | 12 |
| 50 | 24 |

**For Gini (Depth and Accuracy):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Depth** | 3 | 5 | 10 | 15 | 20 |
| **Accuracy** | 72.72 | 72.395 | 70.659 | 71.006 | 71.006 |

**For Entropy (Depth and Accuracy):**

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

**Practical No.-4**

**Logistic Regression**

**Name of Algorithm:** Logistic Regression Algorithm

**Name of Dataset:** Diabetics

**Dataset Description:**

* **No. of rows:** 768
* **No. of columns:** 9
* **Name of columns:** Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome

**Mean squared error: 0.2222222222222222**

**Mean absolute percentage error: 391617358901782.44**

**Accuracy: -391617358901682.44**

**Confusion Matrix:**

|  |  |
| --- | --- |
| 145 | 12 |
| 50 | 24 |

**Practical No.-5**

**K-Means Algorithm**

**Name of Algorithm:** K-Means Algorithm

**Name of Dataset:** Diabetics

**Dataset Description:**

* **No. of rows:** 768
* **No. of columns:** 9
* **Name of columns:** Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome

**Optimal k using Elbow Method:** 3

**Practical No.-6**

**Artificial Neural Network(ANN)**

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

**Name of Dataset:** Diabetics

**Dataset Description:**

* **No. of rows:** 768
* **No. of columns:** 9
* **Name of columns:** Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome

**Activation Function:**

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

**No. of Epochs: 20**

**Training Accuracy:** 66.29%

**Testing Accuracy:** 70.56%

**Confusion Matrix:**

|  |  |
| --- | --- |
| 97 | 60 |
| 31 | 43 |

**Practical No.-7**

**Hierarchical Clustering**

**Approach: Agglomerative Clustering(Bottom Up)**

**Name of Algorithm:** Hierarchical Clustering

**Name of Dataset:** Diabetics

**Dataset Description:**

* **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 clusters depends on dendrogram:** 3

**Method:** Ward method and Euclidean Distance Matrices