**Practical Documentation**

**Practical No: 1**

**Linear Regression**

**Name of algorithm:** Linear Regression model

**Dataset:** House pricing

**No of rows:** 200

**No of columns:** 2

**Name of columns:** Area and Price

**Dataset description:** Dataset consists of two columns i.e Area and price. This is a linear dataset. As Area increases, the price also increases. Per 50 square feet 50 thousand rupees increases.

**Mean squared error:** 9.255967954593205e-30

**Mean absolute percentage error:** 8.161215917896343e-17

**Accuracy:** 100.0

**Practical No: 2**

**Linear Regression**

**Name of algorithm:** Linear Regression model

**Dataset:** Advertising

**No of rows:** 200

**No of columns:** 4

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

**Mean squared error:** 3.938015220480282

**Mean absolute percentage error:** 0.17222599691973411

**Accuracy:** 99.82777400308026

**Practical No: 3**

**Decision Tree**

**Name of algorithm:** Decision Tree

**Dataset:** Diabetes

**No of rows:** 768

**No of columns:** 9

**Name of columns:** pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Functions, Age, Outcome.

**Mean squared error:** 0.2683982683982684

**Mean absolute error:** 0.2683982683982684

**Accuracy:** 73.16017316017316

**Confusion Matrix:**

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

**For gini (Depth and accuracy):**

| **Depth** | **3** | **5** | **10** | **15** | **20** |
| --- | --- | --- | --- | --- | --- |
| **Accuracy** | **72.727** | **73.395** | **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

**Dataset:** Diabetes

**No of rows:** 768

**No of columns:** 9

**Name of columns:** pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Functions, Age, Outcome.

**Mean squared error:** 0.23671497584541062

**Mean absolute error:** 0.23671497584541062

**Accuracy:** 76.32850241545893

**Confusion Matrix:**

| 118 | 16 |
| --- | --- |
| 33 | 40 |

**Practical No: 5**

**K-means Algorithm**

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

**Dataset:** Diabetes

**No of rows:** 768

**No of columns:** 9

**Name of columns:** pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Functions, Age, Outcome.

**Columns used :** Age and BloodPressure

**Optimal k using elbow method :** 3

**Practical No: 6**

**Artificial Neural Network**

**Name of algorithm:** Artificial Neural Network

**Dataset:** Diabetes

**No of rows:** 768

**No of columns:** 9

**Name of columns:** pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Functions, Age, Outcome.

**Activation functions:**

**1st hidden layer :** Relu

**2nd hidden layer :** Relu

**Output Layer :** Sigmoid

**No. of epochs:** 100

**Training accuracy:**76.54%

**Testing accuracy:** 71.86%

**Confusion matrix:**

| 125 | 32 |
| --- | --- |
| 33 | 41 |

**Practical No: 7**

**Hierarchical Clustering**

**Approach: Agglomerative (Bottom up)**

**Name of algorithm:** Hierarchical Clustering

**Dataset:** Diabetes

**No of rows:** 768

**No of columns:** 9

**Name of columns:** pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Functions, Age, Outcome.

**Columns used :** Age and BloodPressure

**Optimal no. of clusters depends on dendrogram:** 3

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