Haritha Menon

Hxm180019

ML- Report for assignment 2

B)

MONKS-1

Avg Training error: 0.0725

Avg Test error: 0.075

Graph-

X- axis – Depth of Tree

Y-axis -Training/Test Error

A close up of a map

Description automatically generated

MONKS-2

Avg Training error: 0.168

Avg Test error: 0.243

Graph-

X- axis – Depth of Tree

Y-axis -Training/Test Error

A close up of a map

Description automatically generated

MONKS-3

Avg Training error: 0.046

Avg Test error: 0.093

Graph-

X- axis – Depth of Tree

Y-axis -Training/Test Error

A close up of a map

Description automatically generated

C)

MONKS-1

Depth =1

A close up of a logo

Description automatically generated

A picture containing clock, drawing

Description automatically generated

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 216 | 0 |
| True Negatives | 108 | 108 |

Depth =3

A screenshot of a cell phone

Description automatically generated

A close up of text on a white background

Description automatically generated

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 144 | 72 |
| True Negatives | 0 | 216 |

Depth =5

A close up of text on a white background

Description automatically generated

![A close up of text on a white background

Description automatically generated]()

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 156 | 60 |
| True Negatives | 12 | 204 |

D)

Decision Tree and Confusion Matrix using scikit-learn

MONKS-1

Depth =1

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 216 | 0 |
| True Negatives | 108 | 108 |

Depth =3

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 144 | 72 |
| True Negatives | 0 | 216 |

Depth =5

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 168 | 48 |
| True Negatives | 24 | 192 |

E)

I used [Occupancy Detection Dataset](https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+) from UCI repository

Attribute Information:

date time year-month-day hour: minute: second

Temperature, in Celsius

Relative Humidity, %

Light, in Lux

CO2, in ppm

Humidity Ratio, Derived quantity from temperature and relative humidity

Occupancy, 0 or 1,

0 for not occupied

1 for occupied status

Occupancy dataset (ID3 implementation)

Depth =1

A close up of a logo

Description automatically generated

A picture containing object, clock, drawing

Description automatically generated

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1615 | 78 |
| True Negatives | 0 | 972 |

Depth =3

A screenshot of a cell phone

Description automatically generated

A close up of text on a white background

Description automatically generated

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1615 | 78 |
| True Negatives | 0 | 972 |

Depth =5

A close up of text on a white background

Description automatically generated

A close up of a logo

Description automatically generated

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1626 | 67 |
| True Negatives | 134 | 38 |

Occupancy dataset (SKLEARN i)

Depth =1

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1639 | 54 |
| True Negatives | 3 | 969 |

Depth =3

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1637 | 56 |
| True Negatives | 1 | 971 |

Depth =5

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | Predicted Positives | Predicted Negatives |
| True Positives | 1641 | 52 |
| True Negatives | 122 | 850 |

ID3 -

* Implements using ‘=’
* The left side of the tree has the data points with same values as split.
* The right side of the tree has the data points with values not the same as split.

SKLearn’s-

* Implements using ‘<=’
* The left side of the tree has the data points with same values as split value.
* The right side of the tree has the data points with values greater than split value.

Conclusion -

The result is approximately similar as the values in confusion matrix is very similar.