RegdNo. 322201297

Assignment No. 1

Course Code: ECAP776 **Registration Number:** 322201297

Instructions:

a. Attempt all questions given below in your own handwriting. Assignment in typed format will not be considered for evaluation.

b. The student has to complete the assignment in the allocated pages only. Any other page in case utilized shall not be considered.

Q1. What are series and dataframes of NumPy package? Brief how they are used for data handling.

[10 Marks] [CO3, L2]

Series (One-Dimensional Arrays) in Numpy:

A one-dimensional array (or vector) is a sequence of elements of the same type. It is similar to a list in python but with enhanced functionality for numerical operations.

Creation:

```
import numpy as np
one_d_array = np.array([1, 2, 3, 4, 5])
print(one_d_array)
```

Usage:

- Used for handling and manipulating linear data.
- Supports operations like indexing, slicing, and arithmetic operations.
- Example of accessing an element:

```
print(one_d_array[2])
```

Two-Dimensional Arrays in Numpy:

- A two-dimensional array (or matrix) is an array of arrays, where each sub-array has the same number of elements.
- It is similar to a table or spreadsheet.

Creation

```
two_d_array([[1,2,3], [4,5,6],[7,8,9]])
print(two_d_array)
```

Usage:

- Used for handling and manipulating tabular data.
- Supports operations like indexing, slicing, and matrix operations.
- Example of accessing an element:

```
print(two_d_array[1,2]) # Output: 6
```

Data Handling with Numpy Arrays:

1. Data cleaning
Handling missing values, removing duplicates, and transforming data.
Example:

```
cleaned_array = np.nan_to_num
                (two_d_array)
```

2. Data Transformation
Applying functions to data reshaping, and aggregating

```
mean_value=
   np.mean(two_d_array)
print(mean_value)
```

3. Data Analysis
- Performing statistical analysis and generating summary statistics.

```
reshaped_array =
two_d_array.
   reshape(1, 9)
print(reshaped_array)
```

Signature of the Student Tanya

Note:-

CO: is the Course Outcome as per your course syllabus.

L1-L6: Learning level objectives as per Revised Bloom Taxonomy (RBT).

RegdNo. 322201297

Assignment No. 1

Course Code: ECAP776 **Registration Number:** 322201297

Instructions:

a. Attempt all questions given below in your own handwriting. Assignment in typed format will not be considered for evaluation.

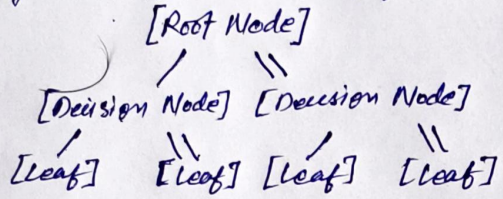
b. The student has to complete the assignment in the allocated pages only. Any other page in case utilized shall not be considered.

Q2. Give the structure of decision tree. Give the important parameters used in the decision tree algorithm.

[10 Marks] [CO4, L1]

Decision Tree Structure

A decision tree is a flowchart-like structure used for classification and regression. Here is a simplified visual representation.



- Root Node: Represents the entire dataset.
- Decision Nodes: Internal nodes where the data is split based on a feature.
- Branches: Paths that connect nodes, representing the outcome of a decision.
- Leaf Nodes: Terminal nodes that represents the final output.

Important Parameters used in Decision Tree Algorithm:

1. Criterion: Function to measure the quality of a split. Common options are Gini (Gini Impurity) and entropy (Information gain).
2. Splitter: Strategy used to choose the split at each node. Options are best split and random split.
3. Max Depth: The maximum depth of the tree. Limits the number of levels in the tree to prevent overfitting.
4. Min Samples Split: The minimum number of samples required to split an internal node. Helps control the growth of the tree.
5. Min Samples Leaf: The minimum number of samples required to be at leaf node. ~~Helps control the growth of the tree.~~ Ensures that leaf nodes have a minimum number of samples.
6. Max Features: The number of features to consider when looking for the best split. Can be an integer, float, or string (e.g. auto, sqrt, log2)
7. ~~Max Leaf Nodes: The number of features to consider when looking for the best split. Can be an integer.~~
7. Max Leaf Nodes: The maximum number of leaf nodes in the tree. Limits the number of leaf nodes to prevent overfitting.
8. Minimum Impurity Decrease.

Signature of the Student Tanya

Note:-

CO: is the Course Outcome as per your course syllabus.

L1-L6: Learning level objectives as per Revised Bloom Taxonomy (RBT).