**Data Analytics Laboratory**

**Task 3**

**Apply Decision Tree Classification technique on the given dataset**

**Introduction**

* Decision tree models are the simplest form of supervised multivariate classification models.
* A series of logical tests (generally in the form of Boolean comparisons) are applied to the sample entries and their resulting subsets in turn to arrive at a final decision.
* It is very easy to visualize the decision process in a simple flowchart to trace the rational of every assignment made by a decision tree model, making it among the most interpretable of models.
* Decision trees are flow-chart-like tree structure, Internal node denotes a test on an attribute, Branch represents an outcome of the test, Leaf nodes represent class labels or class distribution.
* Decision tree generation consists of two phases.
  + Tree construction. Partition examples recursively based on selected attributes.
  + Tree pruning. Identify and remove branches that reflect noise or outliers

**Prerequisites**

1. What is the difference between classification and clustering? Justify the decision tree algorithm is used for classification or clustering.

2. Define information gain with an example.

3. For the same example dataset considered in question number 2, calculate entropy value.

**Apply Decision Tree Algorithm on the below given golf Dataset.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Weather** | **Temperature** | **Humidity** | **Wind** | **Golf Play** |
| fine | hot | high | None | no |
| fine | hot | high | few | no |
| cloud | hot | high | none | yes |
| rain | warm | high | none | yes |
| rain | cold | medium | none | yes |
| rain | cold | medium | few | no |
| cloud | cold | medium | few | yes |
| fine | warm | high | none | no |
| fine | cold | medium | none | yes |
| rain | warm | medium | none | yes |
| fine | warm | medium | few | yes |
| cloud | warm | high | few | yes |
| cloud | hot | medium | none | yes |
| rain | warm | high | few | no |

1. Use appropriate pre-processing techniques for encoding categorical data.
2. Draw the resultant decision tree.

**Results**

The program is implemented in python and the output is observed.

**Faculty Signature**