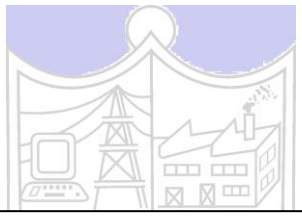


Experiment No.8

Title: Data Classification using Decision Tree Algorithm (ID3)

Batch: **Roll No.:** **Experiment No.: 7**

Aim: Exploration of data classification using Decision Tree algorithm (ID3) on a Sample Dataset

Resources needed: Any programming language, any data source (RDBMS/Excel/CSV)

Theory:

ID3 algorithm, stands for Iterative Dichotomiser 3, is a classification algorithm that follows a greedy approach of building a decision tree by selecting a best attribute that yields maximum Information Gain (IG) or minimum Entropy (H). Additionally, Decision trees are a supervised machine learning technique used for both classification and regression tasks. ID3, specifically, is a simple yet effective algorithm for building decision trees based on the concept of information gain.

In this experiment, we will use the ID3 algorithm to build a decision tree based on a IRIS data and illustrate how we can use this procedure to make a decision on an action classification of data set.

Algorithm:

1. Calculate entropy for dataset.
2. For each attribute/feature.
   2.1. Calculate entropy for all its categorical values.
   2.2. Calculate information gain for the feature.
3. Find the feature with maximum information gain.
4. Repeat it until we get the desired tree.

Procedure / Approach /Algorithm / Activity Diagram:

1. Computes the ID3 algorithm to select an attribute subset that best predicts class labels
2. Use Decision Tree Classifiers to classify the Sample Data(e.g.IRIS Sample Data).
3. Manual write should take his own data so that it could be possible to find best class labels as well as do the classification.

Results: (Program printout with output / Document printout as per the format)

Questions:

1. What are the data filtering techniques available? Explain in brief.

2. What do you mean by sampling data set? How sampling done in data science?

Outcomes:

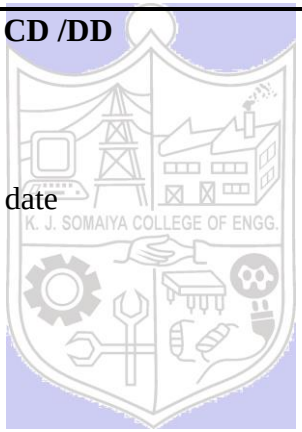
_____
_____
_____
_____

Conclusion: (Conclusion to be based on the objectives and outcomes achieved)

_____
_____
_____
_____
_____

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of faculty in-charge with date



_____

References:

Books/ Journals/ Websites:

1. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition