

	<p style="text-align: center;">Hope Foundation's Finolex Academy of Management and Technology, Ratnagiri Department of Information Technology</p>		
Subject name	Business Intelligence Lab		Subject Code: ITL602
Class	TE IT	Semester – VI (CBCGS)	Academic year: 2018-19 (FH 2019)
Name of Student			QUIZ Score :
Roll No		Assignment/Experiment No:	05
Title:	Implementation of Decision tree classifier using Java.		

1. Course objectives applicable: LO4

2. Course outcomes applicable: LO4

3. Learning Objectives:

1. To learn construction of Decision Tree.
2. To learn the splitting criteria in Decision tree construction.
3. To develop a program for Decision tree induction.

4. Practical applications of the assignment/experiment: Construction of Decision tree

5. Prerequisites:

1. Decision Tree algorithm theoretically.
2. Java/Python programming

6. Hardware Requirements:

1. PC with minimum 2 GB RAM

7. Software Requirements:

1. Windows / Linux
2. JDK 8.0 / Python 3.6

8. Viva Questions (if any): (Online Quiz will be taken separately batch-wise)

1. What is Information Gain?
2. How the attribute selection is done as splitting criteria?
3. What are the internal and leaf nodes of a decision tree?
4. How decision tree can be used for predictions?

9. Experiment/Assignment Evaluation:

Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A <u>or</u> any other relevant method.)		6
2	Neatness/presentation		2
3	Punctuality		2
Date of performance (DOP)		Total marks obtained	10
Date of checking (DOC)		Signature of teacher	

10. Theory: <<handwritten work>>

The ID3 algorithm begins with the original set as the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set and calculates the entropy (or information gain) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain) value.

Recursion on a subset may stop in one of these cases:

- every element in the subset belongs to the same class (+ or -), then the node is turned into a leaf and labelled with the class of the examples
- there are no more attributes to be selected, but the examples still do not belong to the same class (some are + and some are -), then the node is turned into a leaf and labelled with the most common class of the examples in the subset
- there are no examples in the subset, this happens when no example in the parent set was found to be matching a specific value of the selected attribute, for example if there was no example with age ≥ 100 . Then a leaf is created, and labelled with the most common class of the examples in the parent set.

Throughout the algorithm, the decision tree is constructed with each non-terminal node representing the selected attribute on which the data was split, and terminal nodes representing the class label of the final subset of this branch.

Create classification Model using decision tree:

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

11. Performance Steps:

1. Implement a Java/Python Program to perform simulation of decision tree for the given data set.
2. Run the program and test for the tuple (Rain, Mild, High, Weak)

12. Results:

<< Add the hard-copy of the source code and output screen shots >>

13. Learning Outcomes Achieved

1. Understanding of ID3 algorithm (Used for building decision tree)
2. Understanding of calculations of Entropy and Information gain.
3. Implementation of ID3 algorithm in program.

14. Conclusion:

1. **Applications of the studied technique in industry:** Decision tree algorithm implementation.
2. **Engineering Relevance:** Software development in Data mining systems
3. **Skills Developed:** Understanding of decision tree construction and its application

15. References:

- [1] Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3rd Edition.
- [2] P. N. Tan, M. Steinbach, Vipin Kumar, "Introduction to Data Mining", Pearson Education.
- [3] Michael Berry and Gordon Linoff, "Data Mining Techniques", 2nd Edition Wiley Publications
- [4] https://en.wikipedia.org/wiki/Decision_tree_learning
- [5] <https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/>