|  |  |
| --- | --- |
| Semester | T.E. Semester VI – Computer Engineering |
| Subject | Data Warehousing and Mining |
| Subject Professor In-charge | Prof. Kavita Shirsat |
| Assisting Teachers | Prof. Kavita Shirsat |
| Laboratory | Lab 312 A |

|  |  |  |
| --- | --- | --- |
| Student Name | Deep Salunkhe | |
| Roll Number | 21102A0014 | |
| Grade and Subject  Teacher’s Signature |  |  |

|  |  |  |
| --- | --- | --- |
| Experiment Number | 03 | |
| Experiment Title | To perform the calculation of ID3 Algorithm | |
| Resources / Apparatus Required | Hardware:  Computer system | Software:  Python |
| Description | 1. **Decision Tree Algorithm:** ID3 is a popular decision tree algorithm used for classification tasks. 2. **Objective:** It aims to build a decision tree that can be used to make decisions or predictions based on input features. 3. **Entropy-Based Approach:** ID3 uses an entropy-based approach to select the best attributes for splitting the data. It calculates the Information Gain for each attribute to determine the most informative one. 4. **Information Gain:** Information Gain measures the reduction in entropy (uncertainty) achieved by splitting the data on a particular attribute. The attribute with the highest Information Gain is chosen as the splitting criterion. 5. **Entropy:** Entropy is a measure of the impurity or randomness of data. Lower entropy indicates more ordered and predictable data, while higher entropy indicates randomness. 6. **Recursive Splitting:** ID3 recursively splits the dataset into subsets based on the selected attribute until a stopping criterion is met. The splitting continues until either all data points in a subset belong to the same class or a predefined depth limit is reached. 7. **Categorical Attributes:** ID3 is designed primarily for categorical (discrete) attributes, not continuous ones. It handles discrete attribute values well. | |
| Program |  | |
| Output |  | |
| Conclusion: | ID3 employs an entropy-based approach to select the most informative attributes for splitting the dataset. It calculates Information Gain, which measures the reduction in uncertainty achieved by splitting the data based on a particular attribute. | |