COMP 472 Artificial Intelligence   Summer 2023

Mini project 1

Due date: as indicated on Moodle

You can work in a group of 2 minimum or 3 maximum. One submission per group.

Decision Tree Classifier with Entropy-Based Splitting

Use the dataset provided in the book section 19.3 to develop a decision tree classifier that uses entropy as a measure to split nodes and make decisions. The program should be able to build a decision tree from training data and use it to classify unseen instances based on their feature values.

Features and Requirements:

Decision Tree Construction: Implement a function that builds a decision tree classifier from a given training dataset. The decision tree should use entropy as the criterion to determine the best feature to split on at each node.

Entropy Calculation: Implement a function to calculate the entropy of a given set of instances. Use the entropy formula ($H(X) = -\Sigma P(x) * log2(P(x))$) to measure the entropy, where x represents the class labels of the instances.

Splitting Criteria: Implement a function to evaluate the potential splits based on different features and calculate the information gain or reduction in entropy for each split. Select the feature with the highest information gain as the splitting criterion at each node.

Classification: Implement a function that uses the constructed decision tree to classify unseen instances based on their feature values. Traverse the decision tree by evaluating the feature conditions at each node and reaching the appropriate leaf node for classification.

User Interface: Design a user-friendly interface to interact with the program. Allow users to provide input datasets, specify parameters, and visualize the constructed decision tree.

Documentation and Presentation: Document your program's functionality, and code structure. You will give a demo to your TA. The TA's will post time slots on Moodle for the demos. All group members must be present during the demo. You must submit the code on Moodle once you are done as a zip archive. The name of the archive must be of the following format:
<Student1 first name Student 1 last name> <Student 1 ID>-<Student2 first name Student 2 last name> <Student 2 ID> - <Student3 first name Student 3 last name> <Student 3 ID>.zip

You can code this project either in Java or Python.