Code Report

The following report describes the implementation of a decision tree algorithm for classification tasks. The code is written in Python and relies heavily on pandas and Numpy libraries for data manipulation and analysis.

Decision Tree Algorithm

This script builds a decision tree using training data and predicts labels for test data using the built decision tree. The decision tree is built using the ID3 algorithm, and the gain ratio is used as the splitting criterion.

Dependencies

pandas NumPy math sys logging Function Documentation read_data(train_path: str, test_path: str) -> tuple Reads in train and test data from specified file paths.

The code is structured as follows:

Import necessary libraries and modules Define functions for reading in data, calculating entropy and gain ratio, determining the optimal feature for splitting data, building the decision tree, predicting labels for test data, and outputting results Define a function to run the program If the script is being run as the main program, call the run_program function using command line arguments

Additional Information

the code uses entropy instead of Gini index, I tried running both but the results were about the same. I'm curious how you managed to achieve such high accuracy in your instructions example. In addition, I added types and documentation using Sourcery AI