

Decision Tree Algorithm

2025/9/25

This assignment uses the decision tree algorithm learned in class to solve a practical problem: predicting which individuals aboard the Titanic were likely to survive.

Data

You can access the Titanic data from [Kaggle](https://www.kaggle.com/c/titanic).

The screenshot shows the Kaggle competition page for 'Titanic - Machine Learning from Disaster'. At the top, there's a navigation bar with 'Overview', 'Data', 'Code', 'Models', 'Discussion', 'Leaderboard', and 'Rules'. The 'Overview' section is active, displaying a welcome message and a description of the competition. It mentions that the competition runs indefinitely with a rolling leaderboard and provides a link to the Discord channel for further discussion. The 'Description' section includes a welcome message and a brief overview of the competition's goal: to predict which passengers survived the Titanic shipwreck. On the right side, there's a sidebar with 'Competition Host' (Kaggle), 'Prizes & Awards' (Knowledge, no points or medals), 'Participation' statistics (1,402,777 Entrants, 13,607 Participants, 13,571 Teams, 46,059 Submissions), 'Tags' (Binary Classification, Tabular, Beginner, Categorization Accuracy), and a 'Table of Contents' (Description).

Algorithm implementation

As a common algorithm of machine learning, the decision tree algorithm can be implemented in various programming languages, such as Python, C++ ,and Matlab. You may refer to a 3rd party codebase (scikit-learn) of python.

What do you need to submit?

Please submit the following items together in a compressed file (e.g., .zip or .rar):

1. Source code files
2. Lab report containing:
 - The decision tree algorithm implemented
 - The features selected and the preprocessing methods applied
 - Experimental results and visual analysis

All assignments and the final project are to be submitted collectively.