

IST565 Data Mining

HW5: Use Decision Tree to Solve a Mystery in History

In this homework assignment, you are going to use the decision tree algorithm to solve the disputed essay problem. Last week you used clustering techniques to tackle this problem.

Organize your report using the following template:

Section 1: Data preparation

You will need to separate the original data set to training and testing data for classification experiments. Describe what examples in your training and what in your test data.

Section 2: Build and tune decision tree models

First build a DT model using default setting, and then tune the parameters to see if better model can be generated. Compare these models using appropriate evaluation measures. Describe and compare the patterns learned in these models.

Section 3: Prediction

After building the classification model, apply it to the disputed papers to find out the authorship. Does the DT model reach the same conclusion as the clustering algorithms did?