

Machine Learning 1 - Practice Problems 2.5

Universidad de La Sabana
Santiago Toledo-Cortés
Facultad de Ingeniería 2025-2

1. Dataset Preparation

Obtain the dataset Wine dataset, split it into training (70%) and test sets (30%).

2. Decision Tree Classifier

Train a decision tree classifier and calculate a confidence interval for its accuracy on the test dataset using a Binomial distribution.

3. Cross-validation with Bootstrap

Perform cross-validation with bootstrap (30 repetitions) and calculate a confidence interval for its accuracy using a Student's t distribution.

4. Random Forest vs. Decision Tree

Consider an additional random forest classifier and compare its performance with the decision tree classifier using cross-validation (30 repetitions) with bootstrap and a t-test.

5. Classifier Comparison using McNemar's Test

Compare the two classifiers using McNemar's test.