

WQD7006
Machine Learning for Data Science

Decision Tree, Random Forest, Bagging

Use Diabetes dataset to answer the questions.

1. We would like to perform classification for diabetes diagnosis. Perform classification using J48.
 - i. How many instances are there? 768
 - ii. How did the classifier perform? Provide the necessary results.
 - iii. Provide the decision tree. (Copy this from WEKA)

2. Remove skin and pedi from the dataset and re-run J48. Did the accuracy improve? No

3. Run the same classification as in (1) using Random Forest. Compare the performance with J48 using the necessary results. 76

4. You would like to see if Random Forest can be further enhanced by reducing its variance. Run bagging with Random Forest. Is the classification accuracy improved? Yes 76.56

5. To make a more conclusive statement, run the significance test for all three classifiers. Report your findings. experimenter