13.6 Evaluation and leakage

Name: _____ Class: _____ Date: _____ **Total: 24 marks**

KEY WORDS confusion matrix • precision • recall

Objective

Derive every metric from the matrix and design a leakage-proof split.

1 Key knowledge and worked examples

■ The matrix

Test 200: TP 36, FP 10, FN 6, TN 148 – accuracy 0.92, precision 0.783, recall 0.857, F1 0.818; base rate 21% ✓.

■ Worked example: the folds

5-fold CV: 0.905, 0.925, 0.915, 0.900, 0.935 – mean 0.916, sd 0.013 ✓.

2 Practice

2.1 What does spending the test set mean? [3]

2.2 Give the accuracy of the trivial always-on-time predictor. [3]

2.3 Name the three plumbing leakages this unit lists. [3]

exercises.lol exercises.com.cn

3 Exam-style questions

3.1 A competitor reports: TP 40, FP 20, FN 2, TN 138 on the same 200-record test set.

Compute accuracy, precision, recall, and F1 for the competitor. Compare with the course model on each metric and state which model you would deploy for catching lates, with the reason. State what additional resampling evidence you would demand before trusting either. [8]

3.2 Design the leakage defence for a pipeline that scales features, imputes missing values, and selects features.

State the wrong order (all three before splitting) and what leaks at each step. Give the correct per-fold order. Explain the midnight status-code example in this vocabulary. State why fold metrics are averaged over per-fold counts, not over fold accuracies alone when folds are unequal. [7]