Solutions

Marking: **M1** a correct method, **A1** the correct answer, **B1** an independent mark.

Each answer shows the steps, then the **result**.

2.1

accuracy $(36 + 148)/200 = \mathbf{0.92}$

precision $36/46 \approx \mathbf{0.783}$

recall $36/42 \approx \mathbf{0.857}$

2.2

$F1 = 2(0.783)(0.857)/(0.783 + 0.857) \approx \mathbf{0.818}$

baseline $158/200 = \mathbf{0.79}$

2.3

sum 4.58

mean **0.916**

the scores are close (0.900–0.935), so the result is stable across splits

2.4

the penalty date exists only after the work is late: it is the answer

IDs were given in outcome order, so the ID number encodes the label

3.1

(a) the scaling and feature choice used the test rows; the test set was used to choose among 50 models

(b) the test rows helped to build and choose the model, so the score includes luck and information it would not have on new patients

(c) split first: train, validation, test; fit scaling and feature selection on the training data only, inside each cross-validation fold; choose the model on validation, then test once and report that score

3.2

(a) recall: a missed good applicant is lost for good, while extra people only cost a quick check

(b) precision: each false positive costs an expensive interview

(c) predicting “no” for everyone already gives 98% accuracy

3.3

(a) records of the same student can be in both training and test, so the model recognises the student, not the pattern, and the test score is too high

(b) split by student: all of a student’s records in one set, so the test measures performance on new students

(c) a test on later data: train on 2023, test on 2024, because habits and rules change over time

(d) the score on new students and on a later year, with the baseline beside it; precision and recall, the errors that matter, and the limits of the data

exercises.lol exercises.com.cn