



Model Evaluation & Tuning

SCORE

___ / 20

Chapter: AI & Machine Learning · Level ★★★ ·

Time: 30 min

Name _____ Class _____ Date _____

AFTER THIS WORKSHEET YOU CAN

✓ Train/validation/test splits ✓ Cross-validation ✓ Precision/recall/F1 ✓ Grid vs random search

**Section A · Concept Check**

• BEGINNER

4 × 1 = 4

- 1 The test set is used:
 - ☐ A. once, at the end ☐ B. every epoch
 - ☐ C. for tuning ☐ D. never
- 2 K-fold CV averages scores over:
 - ☐ A. k splits ☐ B. k models forever
 - ☐ C. k datasets ☐ D. k epochs
- 3 Precision measures:
 - ☐ A. how many predicted positives were right
 - ☐ B. all accuracy ☐ C. speed ☐ D. recall twice
- 4 Recall matters most when:
 - ☐ A. missing positives is costly (disease)
 - ☐ B. never ☐ C. styling ☐ D. compiling

**Section B · Problem Solving**

• INTERMEDIATE

2 + 3 × 3 = 11

- 5 F1 is the harmonic mean of precision and _____.
- 6 Accuracy misleads on _____ classes.
- 7 99% accuracy on 1% fraud data – why bad?

- 8 Explain 5-fold CV.

- 9 Precision vs recall trade-off example?

**Section C · Challenge**

• CHALLENGE

5

- 10 Why never tune on the test set?

Reflection — the most useful thing I learned: _____

A ___/4 B ___/11 C ___/5 Total ___/20

Teacher's Signature

Parent's Signature

✂ ANSWER KEY – FOLD OR CUT BEFORE HANDING OUT

1-A 2-A 3-A 4-A | 5 recall 6 imbalanced 7 = Predicting all-negative gives 99% - use precision/recall 8 = Split into 5; train on 4, validate on 1, rotate; average 9 = Spam filter:

high precision avoids losing real mail; disease test: high recall avoids missed cases | 10 = You'd overfit evaluation - scores stop reflecting reality
