BitWith Bite

AI & MACHINE LEARNING WORKSHEET

Building a Full Pipeline

Chapter: AI & Machine Learning · Level ★★☆ · Time: 30 min

SCORE ___ / 20

Name _____ Class _____ Date _____

AFTER THIS WORKSHEET YOU CAN

✓ ColumnTransformer for mixed data ✓ Combine numeric + categorical ✓ Fit once, predict anywhere ✓ Grid-search a pipeline

Section A · Concept Check • BEGINNER 4 × 1 = 4

1 Different preprocessing per column type uses:

◯ A. ColumnTransformer ◯ B. RowSplitter ◯ C. TypeCaster ◯ D. MixEncoder

2 A pipeline guarantees test data gets:

◯ A. identical preprocessing ◯ B. new labels ◯ C. no scaling ◯ D. random order

3 Pipeline step params are tuned as:

◯ A. step__param ◯ B. param.step ◯ C. step->param ◯ D. p(step)

4 pipe.predict runs preprocessing:

◯ A. automatically ◯ B. never ◯ C. manually first ◯ D. twice

Section B · Problem Solving • INTERMEDIATE 2 + 3×3 = 11

5 _____Transformer routes columns to the right prep.

6 Numeric columns get a scaler; categorical get an _____.

7 Sketch a ColumnTransformer for [age, salary] + [city].

8 Why grid-search the whole pipeline, not just the model?

9 What breaks if you scale before splitting?

Section C · Challenge • CHALLENGE 5

10 Name the two big wins of pipelines.

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 Column 6 encoder 7 = ColumnTransformer([('num', StandardScaler(), ['age','salary']), ('cat', OneHotEncoder(), ['city'])]) 8 = CV folds re-fit preprocessing per split

- no leakage 9 = Test stats leak into training - overly optimistic scores | 10 = No leakage + reproducible single object to deploy