After Full Marks: Did the Model Really Learn?

Name: _____ Date: _____

A model gets 9 out of 10 correct by always saying NOT CAT. But there is only one cat! An overall score can hide the mistake you care about.

One example, already done

Truth: CAT plus nine NOT CAT cards. Always predict NOT CAT. Accuracy = 9/10 = 90%; cat recall = found cats / actual cats = 0/1 = 0%.

Your turn

1. New truths: CAT, CAT, NO, NO, NO, NO. Predictions: CAT, NO, CAT, NO, NO, NO. Count true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN).

2. Compute accuracy = (TP+TN)/6, precision = TP/(TP+FP), recall = TP/(TP+FN). If a denominator is zero, write “undefined,” not 100%.

3. Label three envelopes: training learns weights; validation chooses settings; final test evaluates after choices. Where would you choose a threshold? Where should duplicates of one source image go?

After Full Marks: Did the Model Really Learn?

Make it visible

Change one thing

Invent ten labels where always saying CAT gets 90% accuracy yet misses every non-cat. Which class-specific count reveals the problem?

My puzzle - make it in class

Design six truths and predictions for a parent. Keep the four confusion counts and metric calculations on an answer copy. Do not change labels after seeing the answers.

At home (with a parent)

Show a parent the 90% score, then reveal how many cats there were and ask what else should be counted.

My parent tried my puzzle: Yes / Not yet Parent initials: _____