Session 2 Handout: What the Numbers Mean

FAST-05 Applied Data Science and ML for Business Users · 7formations · Instructor: Wadood Chaudhary · All examples fictional

Why this session matters

"Our model is 95% accurate" is the most common claim in AI sales, and often the least useful. When the thing you're looking for is rare (fraud, defects, churn), a model can be highly accurate and nearly worthless. This session teaches four boxes, three metrics, and the one question that matters: what does each kind of error cost us?

Key concepts

1. The four boxes (confusion matrix)

	Actually fraud	Actually legitimate
Flagged	True positive (TP): caught	False positive (FP): false alarm
Not flagged	False negative (FN): missed	True negative (TN): correctly passed

2. Three metrics in plain words

Metric	Formula	Plain question
Accuracy	(TP + TN) / all	How often is it right overall?
Precision	TP / flagged	When it flags, how often is it right?
Recall	TP / actual positives	Of all the real cases, how many did it catch?

Base rate matters. At Pioneer Valley Mutual, 20 of 1,000 claims are fraud (2%). Flagging nothing is 98% accurate and catches zero fraud.

3. Errors have costs

- **False positive:** review time (≈$60), delayed payment, an upset honest customer.
- **False negative:** fraud paid (≈$8,000 average in the lab).

Neither error is "worse" in general; it depends on the business. A missed cancer screening and a missed marketing lead are very different false negatives.

4. Thresholds are business decisions

Lowering the threshold flags more: recall rises, precision falls, reviews increase. Raising it does the reverse. The model doesn't choose; **you** do, based on costs and capacity.

Worked example: Pioneer Valley Mutual

Threshold	Flagged	TP	FN	Precision	Recall	Accuracy	Cost (flags × $60 + FN × $8,000)
0.7	50	14	6	28%	70%	95.8%	$51,000
0.6	80	17	3	21%	85%	93.4%	$28,800
0.5	120	18	2	15%	90%	89.6%	$23,200
None	0	0	20	n/a	0%	98.0%	$160,000

Decision: adjusters can review 80/month, so 0.6 fits capacity at less than 60% of the 0.7 cost. The most accurate option (flag nothing) is the most expensive.

Explaining it: "At 0.6 the model catches about 17 of every 20 frauds. About 1 in 5 flagged claims turns out to be fraud, so most reviews clear honest customers quickly."

Checklist: reading a metric claim

- ☐ What's the base rate of the thing being predicted?
- ☐ What are precision and recall, not just accuracy?
- ☐ At what threshold?
- ☐ What does each error type cost us?
- ☐ Can we handle the volume of flags?

Do and don't

Do	Don't
Ask for precision and recall at your threshold	Accept "95% accurate" alone
Put a dollar or time cost on each error	Assume all errors are equal
Factor in review capacity	Pick the lowest-cost threshold you can't staff
Explain with counts ("14 of 20")	Explain with formulas to non-experts
Revisit the threshold as costs change	Set it once and forget it

Try this at work

1. Find a flag or score your team acts on.
2. Estimate the cost of a false positive and a false negative.
3. Ask the owner what threshold is used and why.

Glossary

Term	Meaning
Confusion matrix	Table of TP, FP, FN, TN
Base rate	How common the positive case is
Precision	Share of flags that are correct
Recall	Share of real positives caught
False positive / negative	Wrong flag / missed case
Threshold	Score cut-off for action

Key takeaways

- Accuracy can hide failure on rare events.
- Precision and recall answer different business questions.
- Choose the threshold by cost and capacity.