

A 94% model can still be useless.

D8A.ACADEMY/RESOURCES/BASELINE-MODEL

THE RULE

dumb answer first	before any model
value = the gap	model minus baseline
no gap, no model	← ship the rule.

THE DUMB ANSWERS

majority class	classification
predict the mean	regression
tomorrow = today	brutal to beat

IN SKLEARN

DummyClassifier	majority, two lines
DummyRegressor	mean, same idea
same CV split	← or it is fake.

THE ACCURACY TRAP

94% legit rows	"legit" scores 0.94
accuracy hides it	precision + recall
PR-AUC	the imbalance-proof score

BETTER BASELINES

the current rule	what ops does today
one tiny model	logreg, 3 features
seasonal naive	= same day last week

GOTCHAS

report both	model + baseline
0.94 vs 0.93	← worth the infra?
baselines drift	re-score at retrain

WORTH MEMORIZING

01 The two-line reality check

Run the dummy through the same split as the real model. Then compare.

```
from sklearn.dummy import DummyClassifier

dumb = DummyClassifier(strategy="most_frequent")
cross_val_score(dumb, X, y, cv=5).mean()
# 0.94  <- fraud is 6% of rows

cross_val_score(model, X, y, cv=5).mean()
# 0.94  <- the model learned nothing
```

→ Same 0.94, zero value. On imbalanced data, switch the scoring to recall or PR-AUC: the dummy collapses to 0 there and the comparison becomes honest.

02 Trap: which baseline for which task

The dumb answer changes shape with the problem. Pick the right one.

TASK	BASELINE	BEATS MORE MODELS THAN
churn, fraud, spam	majority class + recall	you would hope
price, demand	mean, or median	linear reg on bad features
forecasting	same day last week	most first Prophets

→ Interview phrasing: I never present a model score alone. I present the gap over a named baseline, on the same split, in the metric the business feels.