

PROJECT KICKOFF

Your machine learning project

Choosing a question, finding data, and planning honest evaluation

Today's key line of code

```
question -> data -> baseline  
-> models -> honest test -> report
```

What makes a good project

1

A clear question

One sentence: what you predict or discover, and for whom it would be useful.

2

Data you can get

Enough rows (a few hundred at least), legally and ethically obtained, usable offline.

3

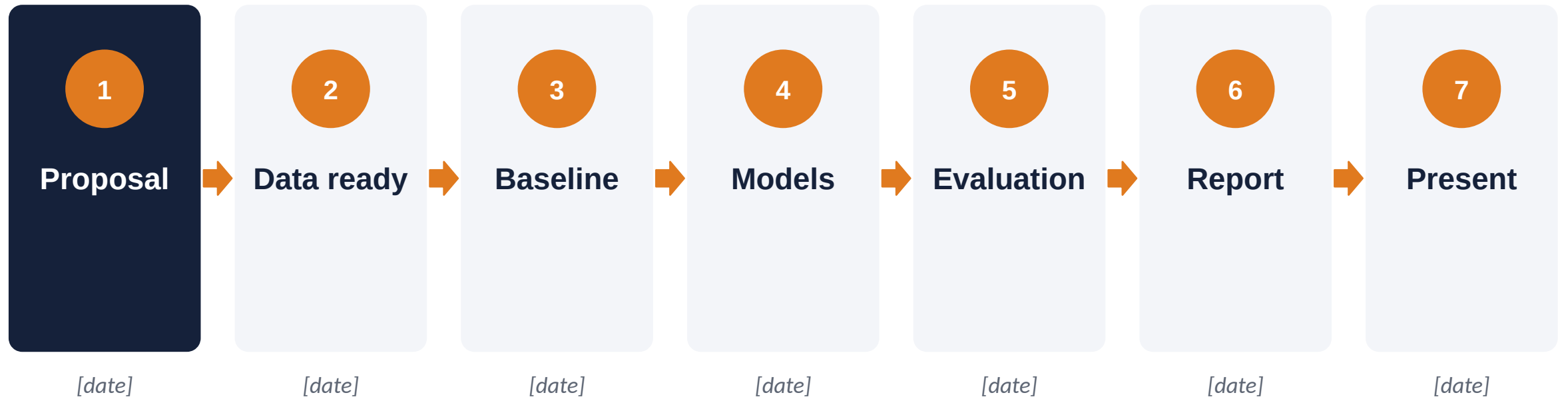
An honest measure

A metric that matches the cost of mistakes, and a baseline to beat.

Project ideas that work offline

Idea	Data	Type
Bangla or Banglish SMS spam	Messages you collect with permission, anonymised	Classification
Wine, cancer or digits deep-dive	Built into scikit-learn	Classification
Diabetes progression	Built into scikit-learn	Regression
Student study habits	A short anonymous survey of classmates	Regression or clustering
Local prices or weather	Records you obtain as a CSV file [check the source]	Regression
Your own idea	Discuss with the instructor first	Any

Milestones



Each milestone is a short check-in. Problems found early are easy to fix.

The proposal: one page

Section	What to write
Title and team	Names and student IDs
Question	One sentence: what you predict or discover
Data	Source, size, features, label; how you get it; privacy plan
Type	Classification, regression or clustering
Evaluation	Main metric and why; baseline; split and cross-validation plan
Risks	What might go wrong, and your backup plan

How the project is graded

Part	Weight	What earns marks
Proposal	[__%]	Clear question, realistic data plan
Method	[__%]	Correct split, baseline, pipeline, cross-validation
Results	[__%]	Honest test score, right metric, error analysis
Report	[__%]	Clear, one structure, limits stated
Presentation	[__%]	Explains choices and results in 5 minutes

Pitfalls to avoid

1

Too big

One clear question beats five vague ones. Shrink the scope early.

2

No baseline

A score means nothing without a simple guess to compare with.

3

Peeking at the test set

Choose everything with cross-validation; test once, at the end.

Ground rules

1

Data ethics

Permission, anonymisation, safe storage. No personal data in submitted files.

2

AI tools

[Course AI policy]. You must be able to explain every line of your code.

3

Teamwork

[Team size and rules]. Every member presents part of the work.

Next step: your proposal

Proposal due: [date]

1 Choose

Pick a question and check that the data really exists and is usable offline.

2 Explore

Load the data and make two charts. Note the size, the label and the class balance.

3 Write

Write the one-page proposal using the template. Submit by [date].

Submit on [LMS / Google Classroom]: your one-page proposal as a PDF or Word file, named Proposal_<your ID>.