

# Case Study: Student Outcome Intelligence Platform

## Context

You are joining as data scientist/engineer at a large Norwegian university with 35,000 students across 8 faculties.

The university generate data across several distinct operational system. To give you enough context to prepare – without prescribing your solution – here is the description of what those systems typically contain:

### **Student Information System**

The core administration record – enrolment status, programme, faculty, grades, personal details, graduation and withdraw history. The university system of records for who a student is.

### **Learning Management System**

The digital learning platform – login activity, course access, assignment submission, discussion forum participation, video watch time, quiz attempt. High-frequency behavioural data.

### **ERP System**

Financial aid records, tuition payments status, scholarship awards, outstanding balances. Separate from the student system – often a different vendor entirely.

### **Physical Campus System**

Library access logs, WiFi presence data, building entry events. A signal of physical campus engagement independent of academic records.

These systems do not talk to each other. Data formats, update frequencies and identifier different across all of them. Some export file, some have API, some emit real time events.

## The Goal

The university want to identify students at risk of dropping out before it becomes visible in any official record – early enough for an academic advisor to intervene.

Right now, it is not possible. The signals exist across separate systems, but nobody see all of them together. A student can be missing lectures, disengaging online, and falling behind financially – and each system see only its own fragment. By the time a grade report confirms the problem, the window to act has usually closed.

Your task is to design and build the platform that closes this gap; get the data in from four systems, make it trustworthy, use it to predict risk and deliver those predictions to advisors each semester in time to act.

The platform handles personal data about real students. It needs to be accurate, explainable, fair, and auditable. It also needs to keep working correctly long after launch.

### What to factor in

- The four systems were built independently, for different purposes, over many years. Their data is inconsistent and their schema change without notice.
- Students change status over time – they take leave, switch programme, withdraw and return. A platform that only holds current state will build incorrect history and produce unreliable predictions.
- Predictions must be made using only information that existed at the time of prediction. Using future data during training produces a model that cannot work in production.
- The student population is diverse. Engagement patterns differ across groups for reasons unrelated to dropout risk. A model that does not account for this will systematically flag the wrong students.
- At some point a student or regulator will ask why that student was flagged red and who has access to their data. That question must be answerable from what you built.

## Your task

Prepare a 20–30 minute presentation where you describe how you would design and build this platform.

We assume you will spend approximately 2–3 hours preparing your response.

Focus on the key decisions needed to make the solution work in practice, rather than covering every detail. We are not looking for a single “correct” solution, but for how you approach the problem and the choices you make along the way.

You may use diagrams where helpful. No coding is required.

Be explicit about:

- assumptions you make
- trade-offs in your design