

In-Class Group Exercise: Defining the Problem for an ML System

1 Objective

This exercise will help students **define a business problem** and determine how an **ML system** could be applied as a solution. Each group will be assigned a **real-world scenario** and will work together to **assess feasibility, value, and potential risks** before jumping into model development.

The purpose of this exercise is to help each group start thinking about:

- How to clearly define a business problem and relate it to an ML solution.
- How to justify why ML could be an appropriate solution.
- Identify key challenges, stakeholders, and success metrics.

2 Instructions

1. **Form Groups:** The class will be divided into groups of **4–5 students**.
2. **Assign a Scenario:** Each group will be assigned **one of the six scenarios** below.
3. **Discussion (15 minutes):** The group will work through the **key questions** at the end of this document.
4. **Class Discussion:** Each group will briefly **present their problem definition, approach, and considerations** to the class.

3 Group Assignments & Scenarios

- **Group 1: Customer Churn Prediction for a Subscription Service** A streaming service (like Netflix or Spotify) wants to **reduce customer churn** by identifying users who are likely to cancel their subscription. The goal is to proactively offer personalized incentives or marketing strategies to retain high-risk users.

- **Group 2: Fraud Detection in Online Transactions** A financial services company wants to build an ML system to **detect fraudulent transactions** in real time and prevent financial losses. The challenge is to balance fraud detection with minimizing false positives that could inconvenience legitimate customers.
- **Group 3: Personalized Product Recommendations** An e-commerce company wants to improve sales by **offering personalized product recommendations** to customers based on their browsing and purchase history. The goal is to enhance user experience and increase the likelihood of repeat purchases.
- **Group 4: Predictive Maintenance for Manufacturing Equipment** A manufacturing company wants to use ML to **predict when machines will fail** to reduce downtime and maintenance costs. The system should provide early warnings so that preventive maintenance can be scheduled efficiently.
- **Group 5: Resume Screening for Hiring** A large company wants to use ML to **automate resume screening** and identify top candidates for job openings. The system should filter applicants based on relevant skills and experience while minimizing bias.
- **Group 6: Demand Forecasting for Retail Inventory** A retail chain wants to use ML to **predict future product demand** to optimize inventory management and reduce stockouts. The system should adapt to seasonal trends, promotions, and unexpected demand fluctuations.

4 Discussion Questions

All groups should discuss and address the following key areas:

1. Stakeholder Engagement

- Who would the key stakeholders be for this project?
- What questions would you ask these stakeholders to ensure you understand the business problem?
- What assumptions might be uncovered during these discussions?

2. Evaluating ML Suitability

- Do you think ML is a suitable solution? What factors would you consider to determine if ML is appropriate for this problem?
- Provide an example of an alternative, non-ML approach that could be considered. What are the limitations of this approach compared to an ML approach?

3. Define Performance Metrics

- Define three performance metrics for the ML system. Include at least:

- **One technical metric** (e.g., a model performance metric like RMSE, precision, recall).
- **One system performance metric** (e.g., response time, uptime, scalability).
- **One business metric** (e.g., customer retention, fraud reduction, cost savings).
- Explain why each of these metrics is important for evaluating the success of the ML system.

4. Understanding Value and Feasibility

- What's the potential value of the ML system to the organization? Consider both:
 - **Tangible benefits** (e.g., cost savings, revenue increase, efficiency improvements).
 - **Intangible benefits** (e.g., improved customer satisfaction, brand reputation).
- List some key **technical requirements** that would be helpful to understand early on before developing the solution (e.g., data, infrastructure).
- What gaps might exist, and how would you address them?

5. The Iterative Process

- Describe why the development of this ML system would be an **iterative process**.
- Provide an example of something that could **change during development** (e.g., a performance metric, a technical requirement).
- How would you **manage this change** to avoid scope creep?

5 Class Discussion & Wrap-Up

After the **15-minute discussion**, each group will:

- **Present their scenario** and describe how ML could solve the problem.
- **Explain their answers** to the key discussion questions.
- **Highlight challenges or trade-offs** they identified.

The instructor will **facilitate a discussion**, comparing different groups' thought processes and reinforcing key takeaways from the exercise.

6 Why This Exercise?

This in-class exercise encourages **critical thinking** and **problem formulation** before building an ML system. It aligns with **Chapter 2: Before We Build**, ensuring students:

- Understand the importance of **defining the problem first**

- Learn to **justify when ML is appropriate** (and when it isn't)
- Consider **stakeholders, risks, and data availability**
- Improve **collaboration and communication skills**