Machine Learning in Production / AI Engineering Midterm, Spring 2023

Christian Kaestner and Eunsuk Kang

Name: _____

Andrew ID: _____

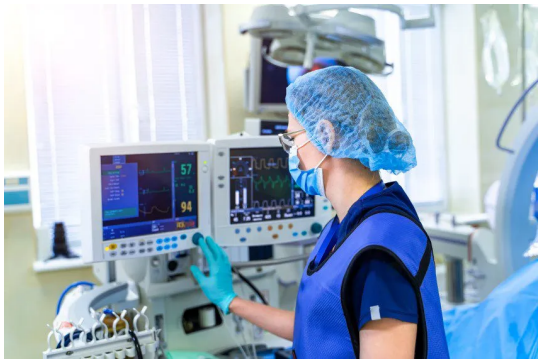
Instructions:

- Not including this cover sheet, your exam should have **9** pages. Make sure you are not missing any pages.
- All questions in this midterm refer to the scenario on Page 2. Answers are graded in the context of the scenario; **generic answers that do not relate to the scenario will not receive full credit.**
- The exam has a maximum score of **57** points. The point value of each problem is indicated. We allocated approximately one point per minute.
- **Please write legibly**. We are unlikely to be able to grade your solution if we can't read it.
- We give an amount of space commensurate with what we expect you to need for each question. We use horizontal lines to suggest where to not use the full page. You may exceed those limits if it is clear where to find the rest of your answer. However, we strongly recommend writing concise, careful answers; short and specific is much better than long, vague, or rambling.
- **Do NOT write anything you want us to grade on the back of pages.** We will scan the exam and will not look at the back sides.
- This is a **closed book exam**; no books or electronics allowed. You may refer to 6 sheets of notes (handwritten or typed, both sides).

Scenario

Predictive maintenance is an approach to maintenance that uses data analysis and machine learning algorithms to predict when maintenance is needed on equipment before it fails. The goal is to prevent unplanned downtime, reduce maintenance costs, and extend the lifespan of the equipment.

You are employed at a large consulting firm (think Accenture, PwC, IBM) working with various clients to develop AI solutions for their problems. You have received a substantial contract from UPMC (a deep-pocketed and highly profitable non-profit healthcare company operating 40 hospitals) to develop a predictive maintenance strategy for medical equipment, such as imaging machines, ventilators, and dialysis machines. UPMC hopes that by detecting and repairing issues before they result in equipment failure, the predictive maintenance system could reduce unplanned downtime, improve patient outcomes, and extend the lifespan of the usually very expensive medical equipment. At the same time, good predictions could reduce the cost of unnecessary maintenance.



UPMC provides access to log files from various machines (both historic and live) and they have logs of past equipment maintenance and replacement. Your predictive maintenance software can be integrated with their IT and electronic medical record systems in the hospitals. For some common equipment like ventilators and dialysis machines, UPMC owns thousands of devices; for more expensive ones like MRI scanners, they have less than a hundred but maintenance logs for about 15 years. In addition, UMPC commits to hiring technical staff that can take pictures of devices and parts regularly going forward.

You plan to start with relatively simple regression models based on age and usage patterns of devices but plan to also explore more advanced models that predict maintenance problems from images.

The goal is not just to develop models, but to actually deploy a solution that could be used to order replacement parts, to issue work orders for maintenance staff, to inform operators about possible accuracy issues in operation, and to provide comprehensive reports. Administrators who have experience with maintenance in the various departments are available to oversee the system.

Your team currently consists of three fairly experienced data scientists and one software engineer and one database expert, who all know little about medical devices. You can pull in help from across the large consulting company. You can also try to work with UPMC employees though they tend to be usually busy and overworked and have little patience for this modern AI voodoo. UPMC's lawyers have already asked to be involved at some point.

GPT3 and ChatGPT have substantially contributed to developing this scenario.

Question 1: Goals and Telemetry [13 points]

(a) [7 points] Your first task is to identify goals that your predictive maintenance system should achieve. Identify a *user goal* from the perspective of the maintenance staff at UPMC. State (1) the goal (2) the measure (3) data to be collected for the measure, and (4) operationalization.

User goal:

Measure:

Data to collect (existing or additional):

Operationalization:

(writing below this line is allowed but discouraged)

(b) [6 points] You plan to evaluate how the model does *in production*. In particular, you would like to see how often the model recommends maintenance unnecessarily (false positive). Design a measure and suggest what data to collect and how to operationalize the measure with telemetry. The measure can be an approximation, but must be plausible within the realism of the scenario.

Measure:

Data to collect (what and how):

Operationalization:

(writing below this line is allowed but discouraged)

Question 2: Trade-offs [16 points]

You are considering how to deploy the predictive maintenance models, either completely in the cloud as a service to the hospitals or on a dedicated machine in the maintenance department of every hospital.

(a) [6 points] Identify and rank two qualities that are important for the decision in this scenario and one quality of little importance (no measure required for any of them). Provide a brief justification of why they are important or not important:

Quality 1 (most important):

Quality 2 (second most important)**:**

Quality 3 (low importance):

Justification:

(writing below this line is allowed but discouraged)

(b) [6 points] Make a recommendation with a brief justification of how to deploy the models, considering the tradeoffs between the qualities. Refer explicitly to the important qualities identified previously. If you are missing information to make that decision, describe what information you would need and how you would make a recommendation with it.

(c) [4 points] You are concerned that technical deployment decisions are influenced by groupthink in your organization. Describe a possible problem that can plausibly arise from groupthink in the scenario and suggest an intervention.

Problem:

Intervention:

Question 3: Model and Data Quality [8 points]

(a) [4 points] Early experiments with machine learning create a model that seems rather mediocre in terms of accuracy. Briefly discuss how you could use the ideas of *capabilities* to get a more nuanced understanding of the weaknesses of the model. Ensure that your answer demonstrates an understanding of the concept and relates to the defect prediction scenario.

(b) [4 points] Provide a plausible concrete example of concept drift in the scenario (i.e., decision boundary changes; not data drift) that may degrade the accuracy of your models in production over time. Indicate how you can manage the drift (a “solution”).

Example:

Solution:

(writing below this line is allowed but discouraged)

Question 4: Risks and Mitigation [20 points]

If maintenance orders are not issued in time to prevent equipment failure, it could cost your customer thousands (if not millions) of dollars to replace broken equipment. Before deploying a new ML product that uses images of equipment to predict a failure, you'd like to identify and address potential hazards in your system, focusing on the following requirement: *The maintenance orders are issued in a timely manner for equipment that is at risk of a failure.*

(a) [2 points] Name two "**machine components**" (in the world vs machine sense) relevant to the scenario and requirement.

-
-

(b) [3 points] State one **software specification** that is necessary for the system to satisfy the above requirement.

(c) [3 points] State one **environmental assumption** that is necessary for the system to satisfy the above requirement.

(writing below this line is allowed but discouraged)

(d) [12 points] Describe two plausible failure scenarios that may result in a violation of the above requirement. These scenarios would correspond to paths in a fault tree (you do not need to draw the tree). At least one of these scenarios should involve a fault or a mistake in an ML component. For the first scenario describe a mitigation strategy. The mitigation should be at the system level, outside of the ML component (i.e., not just "train a more accurate model" or “use an ensemble model”). Your answer must identify whether the strategy eliminates or reduces the likelihood of the requirement violation and briefly explain how.

Failure Scenario 1:

Mitigation for Scenario 1
Check one: ☐ *eliminate scenario* **or** ☐ *reduces the likelihood of violation*
Description:

Failure Scenario 2: