

## SIEMENS - Machine Learning Engineer

### Scenario-based Assignment: E2E Predictive Maintenance in Manufacturing

**Objective:** The purpose of this assignment is to understand better your problem solving, and technical skills, and your thought process, and how in general you would go about performing some critical tasks in the role, which may help us to evaluate better your overall skill set.

**Scenario Background:** You are working for a manufacturing company that aims to implement predictive maintenance using machine learning techniques. The goal is to predict potential breakdowns or maintenance needs in machinery to optimize operations and reduce downtime.

**Assignment Overview:** Here we use a simulated engine degradation under different combinations of operational conditions and modes provided by NASA (find the data and details: [CMAPSS-Jet-Engine-Simulated-Data](#)). The objective is to predict the number of remaining operational cycles before failure.

The candidate's primary task is to outline a high-level plan on how they would approach building an algorithmic predictive maintenance solution. While the primary objective of the task is to discuss namely the thought process, problem solving approach, planning, nevertheless, if you want to go beyond the theoretical description i.e., analyzing and transforming the data, develop a baseline model, and serve, or propose a deployment and monitoring strategy, feel free to do so as it may give us a better sense of technical. In case you opt to go deep dive and implement, feel free to host it publicly in GitHub and share your progress prior to our in-person meeting.

Please note we would not expect the candidate to spend so much time on the task completion. Either case, the high-level discussion, or the high-level discussion plus hands-on, we do not need you to prepare any slides or prepare much content beforehand; your time is mutually respected. Spending time to understand the problem as much as possible and be able to explain the underlying steps and provide reasons for the decisions at the various steps in the product lifecycle are of utmost important and is thus our primary objective.

To facilitate the interactive discussions, and assist you in the preparation, in the following we provide main and important points, but not limited to, that we would like to address with you in our meeting.

## Key Tasks and Deliverables:

- Problem Formulation:

How do you go around formulating the problem?

How do you define the problem objective?

In what aspects if make sense to engage with stakeholders?

How do you decide on your data preparation and model training?

- Data Understanding and Preparation:

Describe how you would explore and clean the dataset.

What considerations would you consider regarding missing values, outliers, or data quality issues? Explain how you would transform the dataset into a suitable format for machine learning, highlighting any feature engineering steps you might take.

- Model Selection and Development:

Propose a machine learning model(s) suitable for predictive maintenance in this scenario.

Explain why you chose this model(s) and how it aligns with the problem statement.

Outline the steps you would take to train, validate, and select the model.

- Solution Implementation:

You may describe how you would decide codebase structure.

Is that critical to follow a certain programming guidelines or principle?

Would you consider a separate environment for the development and production?

How do we serve the solution? How about backend and frontend communications?

Do we need to consider writing tests in our pipeline? If yes, what components?

What considerations would you take to improve code stability and efficiency?

- Deployment, Maintenance and Automation:

Briefly discuss your approach to deploying this model into a production environment and ideas to automate the end-to-end pipeline as much as it is required and possible.

How do you decide how and where to deploy i.e., cloud, edge? Containerization?

How would you optimize model inference in production?

What considerations would you take for scalability, reliability, and performance?

How would you monitor the deployed model's performance? How about data?

What metrics or tools would you use for monitoring the solution, and why?

How do you go about automating the whole pipeline, e.g. data preparation, modelling, deployment (to facilitate model improvement, monitoring, and continuous improvement)?

Good Luck! 🍀