

# Technical Assignment

Machine Learning Research Software Engineer



## Objective

We would like you to demonstrate how you reason about data science problems and present your technical results to a general audience. Please choose **only one** of the two data science challenges below and prototype a solution in Python. You are encouraged to use any tools you are familiar with, such as Pandas, Scikit-learn, Keras, etc. to build this solution.

The goal of this activity is for you to impress us with your data science and software engineering skills in a hands-on setting. Please note that we don't expect you to achieve a specific score on a target metric, but rather that you show us the way you would design a complete solution, detailing the different stages of your end-to-end pipeline and the reasoning behind your decisions in a very open setting. Expect to discuss your results and methods in our next conversation.

## Challenges

On your first week at Mind Foundry, you are invited to take up a data science engagement that just started. The customer has provided an initial data sample alongside a description of what they would like to achieve at the end of their three-month engagement, which is summarised below.

As part of the engagement, we always seek to get our customers excited about their projects and aim to produce a first prototype model within the first week. Your task is to design, implement, evaluate and present this prototype for your choice of **only one** of the customers below.

### Option 1: Forecasting solar energy in the UK

The Solar Mesh is a company in the energetic sector that manages electrical energy supply. To guarantee they will meet demand, they require to know in advance how much energy they will gather from other sources, such as solar energy, within a 3-day horizon at a 30-min cadence. They have provided an initial data sample that contains half-hourly approximate measurements of produced solar energy for the last year, which they can use to fit a model to forecast solar energy.

Your task is to design, prototype and report performance for an end-to-end pipeline to forecast 3 days of solar energy production at a 30-min cadence from the data provided in solar.csv.

### Option 2: Predict heart disease

Oxheart, a fellow Oxford University company, is working closely with a healthcare provider to identify which factors drive a certain heart disease. As part of a collaboration, they would like us to build a Machine Learning model to identify whether a patient will develop this disease from a set of anonymised physiological attributes, as provided in heart.csv.

Your task is to design, prototype and report performance for an end-to-end pipeline to classify whether a patient will develop this heart disease given the data.

## Deliverables

Please submit a Jupyter notebook before your interview where you document each stage of your solution for **only one** of the cases above. This should include a working prototype as well as any appropriate comments that helps a general audience understand the main components, whilst remaining scientifically motivated.

Please submit your solution and any accompanying documents to [jane.williams@mindfoundry.ai](mailto:jane.williams@mindfoundry.ai) before the day of your interview.

As this is only an initial prototype, please bear in mind you should focus on designing a complete pipeline that convinces the customer that you can deliver a modular, end-to-end solution that meets their success criteria, rather than only focusing on achieving good performance on a specific metric. Nonetheless, reasoning around performance metrics is welcome and beneficial.

As such, please ensure that your notebook documents:

- Each stage of your end-to-end pipeline
- The justification behind your modelling decisions
- The metrics of success for the task

During your interview, we would like you to walk us through your notebook, describing not only what's on the notebook, but also to briefly touch on which other exciting extensions you could add within a 3-month engagement for these hypothetical engagements.