

SACAIR Entelect Hackathon Challenge

Table of Contents

Introduction	2
Generate A Problem Statement (20 mins)	2
Design (1hr 40 mins)	3
Prototyping (3hrs)	4
Presentation	4
Assessment Criteria.....	5

Introduction

Welcome to the Entelect/SACAIR Hackathon Challenge!

When people think of AI and data projects, the chances are, they are mostly considering building cool AI models. In reality, there is a lot more to a data project than getting the most accurate model! Data engineers, data scientists and machine learning engineers work together with other software developers to build systems for data projects that consider the entire solution - from finding a business problem to final deployment and maintenance of an AI model.

In this Hackathon, you will be designing a solution-wide workflow to solve a business problem which you feel passionate about. The hackathon will be run in the style of a design sprint where you will pose a question which you would like to answer and then create a workflow which you would follow to implement a solution (<https://designsprintkit.withgoogle.com/methodology/overview>).

If you live in South Africa, chances are, you have strong feelings about our national power provider, Eskom.

In this hackathon, we are providing you with some Eskom data. You are expected to do the following:

- Review the data and think of a problem you would like to solve using this (and potentially other) dataset(s).
- Create a problem statement around the issue you would like to solve. Be creative – think outside the box!
- Design a system-wide data solution which considers the flow of data from its original source to the final product which is deployed and maintained.
- Prototype the data model you would like to build.
- Present your solution as an elevator pitch to the judges.

You will be assessed on the novelty, creativity and usefulness of the problem you are trying to solve, the model you build, and for telling a compelling data story.

For more information on data storytelling see these links:

- <https://powerbi.microsoft.com/en-us/data-storytelling/>
- <https://online.hbs.edu/blog/post/data-storytelling>

The following sections serve as a guide for the Hackathon with recommendations, tips and processes to follow.

Generate A Problem Statement (20 mins)

Explore the dataset and decide on a problem you would like to solve. We have given a few examples but we would like you to be Creative and come up with something unique and insightful!

Think about the following:

- Start with a pain point. Think about something which is frustrating or causes issues and pose a question around that.

- What is a question that would be most useful to answer? Ask yourself why the problem exists (use the [5 whys technique](#))
- Think about who you are solving the problem for? Is this for Eskom employees, the general public, somewhere you work, your university or school?
- What are the business, engineering, social and/or environmental benefits in solving this problem?
- Try to be very specific about your question. You do not have much time. Pick **one thing** and focus on it.
- Formulate your problem statement.

Some problem examples could be (these are intentionally vague to get you thinking!):

- How much diesel does Eskom use?
- When is the best time to perform maintenance?
- What factors affect power usage?
- How could I improve the way Eskom visualises their data?

Design (1hr 40 mins)

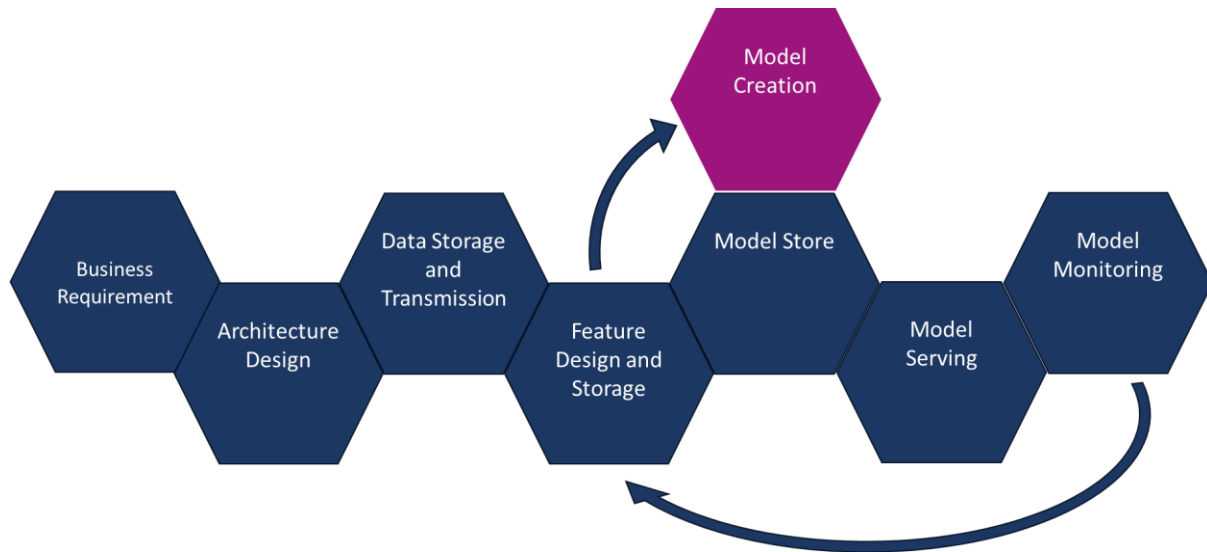
Once you have decided on a problem, ask yourself what process you would follow to try and solve the problem. Ask the questions:

- Who is this for and how do they access the solution?
- What resources do I need to solve the problem? Is this something which requires a lot of computation?
- When is this happening? How does time factor in your solution? Are you streaming live data or presenting static results in a dashboard?
- Where is the solution going to be created, hosted and maintained?
- Where is my data coming from and where should I store it?
- What additional data do I need, if any?
- How is my model going to account for changes in the data if there are any?
- How often do I need to retrain my model?
- How am I going to show my results?
- What assumptions am I making around my data and system?

Create a basic, high-level process flow from source data to visual reporting. Consider the diagram below for inspiration on different system components.

We do not expect you to build a complete architecture diagram or know exactly which software and hardware you are going to use. You are welcome to mention which technologies you are going to use (cloud resources, databases, pipelines) but we are looking for how you create a vision for the solution, not on how well you know different software.

Tip: If you would like to use a great and consistent way to show your system in different levels of detail, you may find the “Data Solution Design” section in the Dataset and Resources document useful.



Prototyping (3hrs)

Prototype some of the code required to solve the problem from a data science perspective. Consider the following:

- How data is imported.
- How data is cleaned and prepared.
- What are the dependent and independent variables?
- Find correlations and relationships between variables.
- Identify and prototype an appropriate model.
- How you will be measuring the success of your model.
- How you will visualise and report the results.

You are not expected to create the full solution which you have designed! Credit is awarded for good data science practice, making reasonable and logical assumptions, and identifying and using the correct data and models to address the problem statement.

As this is only a day-long event, you will not be judged on the accuracy of any data models. The coding/prototype will be judged on assumptions made and if your process follows good data science practice (assumptions and design are reasonable and logical and solve the problem you have posed).

Hint: We would really like you to create some visuals to show your results! Pictures speak a thousand words.

Presentation

Create a short 3-minute presentation as an elevator pitch to tell the story of your problem and the solution you designed. This should tell the data story. Talk us through the model, the assumptions and the solution.

You may create a slideshow presentation or a present your solution in another format. Credit is awarded for the content of the presentation and how you present it, not necessarily how pretty your slides are.

Assessment Criteria

You will be assessed on the following criteria:

Category	Description	Points	Team Score
Completeness of vision	Did the team consider the overall vision of the project, encompassing all elements from defining the problem to deploying and maintaining their solution? Points are awarded for considering all parts of the solution and not just the coding portion.	5	
Creativity of vision and solution	Did the team create a unique and insightful solution to their problem? Does the solution address the problem and does it make sense?	5	
Good data science	How well did the team implement good data science practices? Does their code and workflow make sense? Are all assumptions made reasonable and justifiable?	5	
Presentation	How well did the team tell a story with their data in their presentation?	5	
Model accuracy	How did the team present the results they achieved? If the team built a model, how accurate was their model and how did they ensure it is not overfitted?	3	
Collaboration and engagement	Did the team members engage with each other and with the facilitators of the hackathon?	2	
Total		25	