Skin Condition Classification Hackathon

Welcome to the Skin Condition Classification Hackathon!

In this exciting data science mission, participants are challenged to develop a sustainable and efficient machine learning model capable of classifying skin conditions based on beauty product descriptions.

The provided dataset was labelled in advance using the power of Large Language Models (LLMs). You are required to use this labelled dataset to train one or more machine learning models for the classification task. Considering the importance of sustainability and efficiency, participants are encouraged to design the smallest possible model that can be deployed in production environments without compromising performance.

Competition Activities:

Participants will be able to:

- Train one or more machine learning models on the labelled dataset to accurately classify skin conditions.
- Optimize the model for sustainability by minimizing its size and resource consumption without sacrificing accuracy.
- Evaluate the performance of the model and compare results with fellow competitors.

End of Hackathon

At the end of the hackathon, teams will present their models, showcasing innovative solutions for sustainable and efficient skin condition classification.

Judging will be based on a combination of:

- Prediction performance
- Model size
- Resource efficiency

This will prove that impressive results can be achieved while remaining environmentally responsible.

Data Overview

- 6240 labelled samples
    - Fields:
        - Description in English (short description of about 115 words, max 822 words and min 7 words)
        - Label (skin conditions mentioned, e.g., oily, dry, normal, etc.)

Evaluation Criteria

- Performance: Accuracy of F1 score
- Sustainability: Use codecarbon to evaluate CO2 emissions

Resources Needed

- Access to a Python environment