

MODEL X

Instruction and Guide

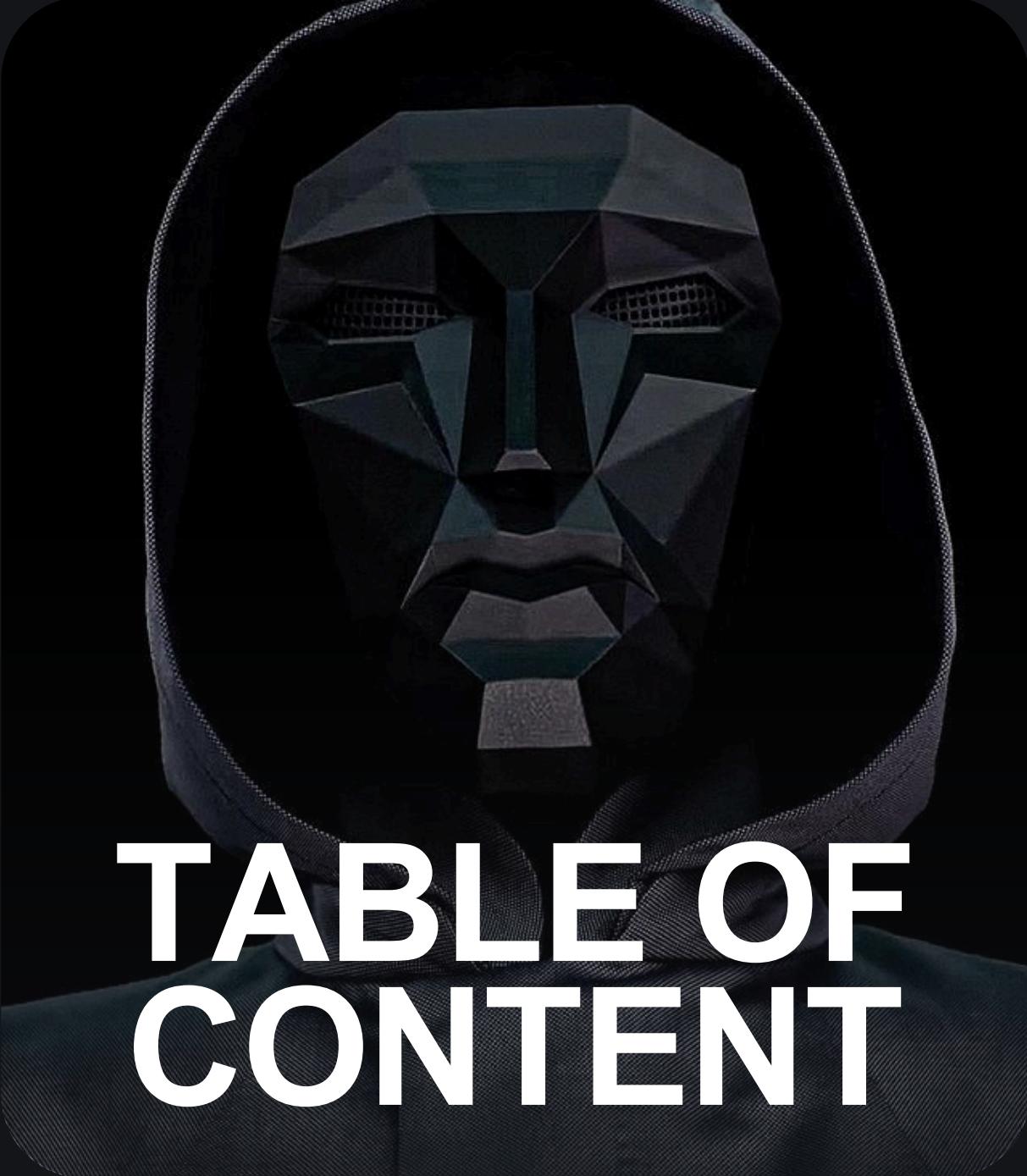


TABLE OF CONTENT

Scenario

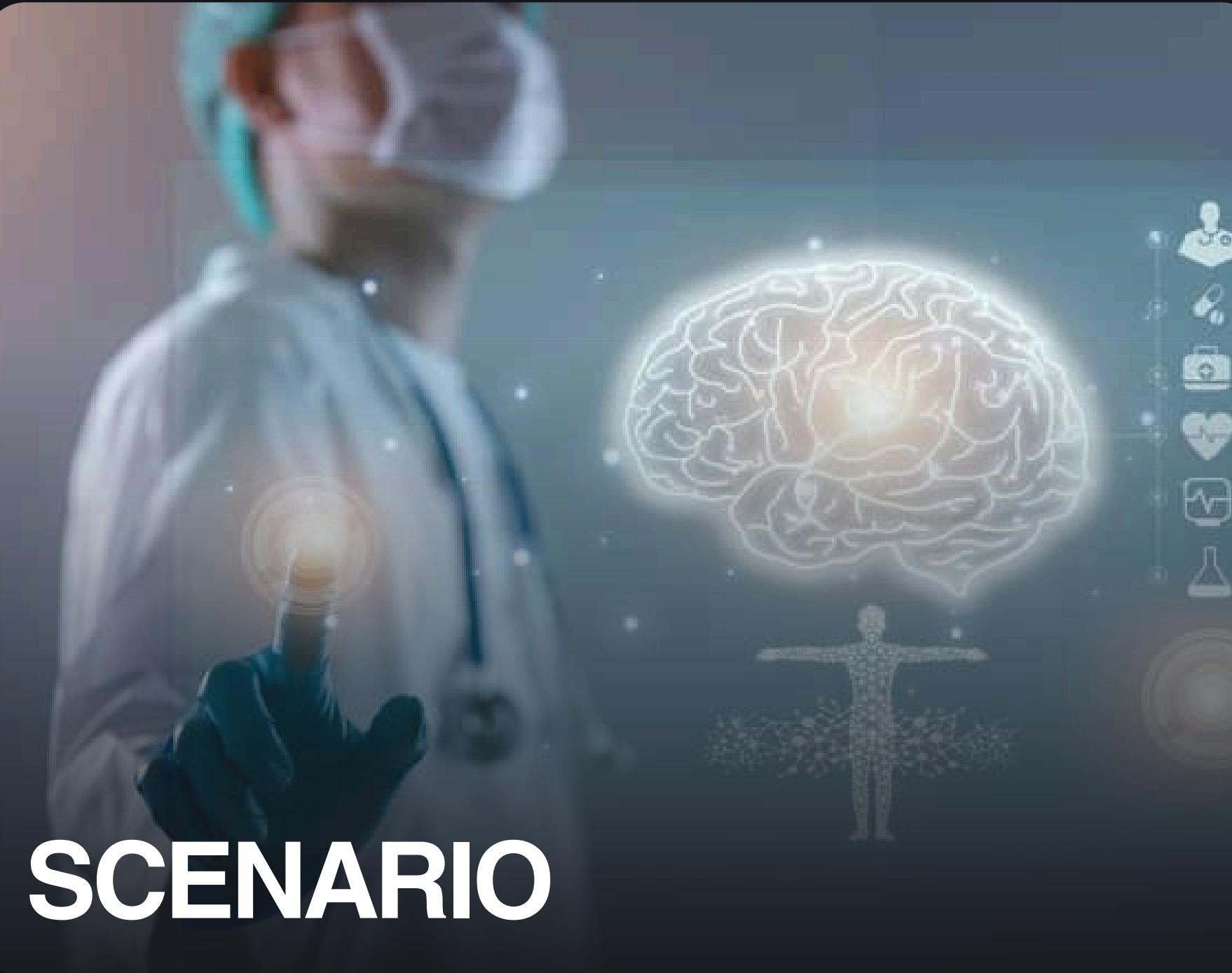
Introduction to the dataset

What you need to do

Instructions

Submission Structure

Marking Criteria



SCENARIO

- Dementia is a major and growing global health issue.
- Many risk factors are non-medical, such as lifestyle, education and social context.
- In this hackathon, you will explore how well non-medical information alone can help predict dementia risk.
- **Your goal:** build a binary classification model that predicts whether a person is at risk of dementia or not, expressed as a probability (0–100%), **using only non-medical variables** from the dataset.

INTRODUCTION TO THE DATASET



Dictionary

Companion document describing every column: variable names, meanings, value encodings and special codes . Use it to understand which fields are allowed and which must be excluded.



Dataset

Curated subset of the NACC cohort. Each row is one participant visit, with **both medical and non-medical** features plus a binary label indicating dementia vs. no dementia.

WHAT YOU NEED TO DO

Explore the non-medical data

Select a valid feature set

Feature engineering and Preprocessing

Build a binary classification model

Evaluate and improve the model

Explain what the model has learned

INSTRUCTIONS

We are not only judging final model performance but also the overall quality and thoughtfulness of your approach, including preprocessing, feature engineering and hyperparameter tuning. During the hackathon, we encourage you to use version control (for example Git) and to build and compare multiple models rather than stopping at the first one that works, as this will help you achieve a higher score.

Develop your solution in a python notebook

Use your preferred ML stack

You must not use medical/diagnostic variables as model inputs.

Do not merge with any external patient-level data sources

Use the given template for the report

Justify your every step

Provide comprehensive Exploratory Data Analysis

SUBMISSION STRUCTURE

Report

If you are using Git, you may submit only a PDF report, but it must include the link to your GitHub repository with all your code.

OR

Zip file

If you are not using Git, you must submit a ZIP file containing your full source code and the PDF report.



MARKING CRITERIA

| | |
|---|-----|
| 1. Data Preprocessing & Understanding | 15% |
| 2. Feature Engineering & Selection | 20% |
| 3. Model Development & Training | 20% |
| 4. Hyperparameter Tuning & Optimization | 15% |
| 5. Model Evaluation & Metrics | 15% |
| 6. Explainability & Insights | 10% |
| 7. Report / Presentation & Code Quality | 5% |



**LET THE GAME
BEGINS**