Overview

🏆 Cardiovascular Disease Prediction Competition

📌 Overview

Cardiovascular diseases (CVDs) remain one of the leading causes of death globally. Early and accurate prediction is critical for effective prevention and treatment strategies. This week's challenge builds on your previous work, where you developed machine learning models for heart disease prediction. Now, you will **implement an Artificial Neural Network (ANN) model** using the **same clinical and demographic dataset** to predict the likelihood of heart disease.

📝 Description

Participants are required to design, implement, and optimize an **ANN-based predictive model** to classify individuals as either at risk (1) or not at risk (0) of cardiovascular disease. The objective is to build a model that not only achieves high performance but also generalizes well to unseen data.

📊 Evaluation

Submissions will be evaluated using the **F1 score**, a metric that balances precision and recall. This ensures that both false positives and false negatives are minimized—crucial for real-world clinical applications.

Each ground-truth label is either:

- **0** – No Cardiovascular Disease
- **1** – Cardiovascular Disease Present

📤 Submitting

- Participants must submit a CSV file containing two columns:
    - `id`: Unique identifier for each instance
    - `cardio`: Predicted label (0 or 1)
- Format example:

```csv
id,cardio
1001,0
1002,1
1003,0
```

⏳ Timeline

- **April 5, 2025** – Competition Starts
- **April 11, 2025** – Submission Deadline (11:59 PM WAT)

💾 Code Requirements

All teams must submit their notebooks to their **team's GitHub repository** for verification.

📊 Evaluation

Submissions in this competition will be evaluated based on their ability to accurately predict the presence or absence of cardiovascular disease.

🏆 Scoring Metric

The competition will use the **F1 Score** as the primary evaluation metric. The F1 Score is the harmonic mean of precision and recall, providing a balance between false positives and false negatives.

📌 Label Information

Each ground-truth label is a **binary value**:

- `0` - No cardiovascular disease
- `1` - Cardiovascular disease present

🏅 Leaderboard Calculation

- The **Public Leaderboard** score is based on a subset of the test data.
- The **Private Leaderboard** score is computed on the full test dataset and will determine the final rankings at the end of the competition.

The team with the highest **F1 Score** on the **Private Leaderboard** will be declared the winner.

Good luck! 🚀

Dataset Description

📂 Cardiovascular Disease Dataset

The dataset consists of **70,000** patient records, including **11 features** and **1 target variable**. The data was collected during medical examinations.

📊 Data Description

The dataset contains three types of input features:

- **Objective**: Factual patient information.
- **Examination**: Results from medical tests.
- **Subjective**: Self-reported patient information.

🔍 Features

Feature	Type	Column Name	Data Type	Description
Age	Objective	`age`	int (days)	Patient's age in days
Height	Objective	`height`	int (cm)	Patient's height in centimeters
Weight	Objective	`weight`	float (kg)	Patient's weight in kilograms
Gender	Objective	`gender`	categorical	1: Female, 2: Male

Systolic Blood Pressure	Examination	`ap_hi`	int	Systolic blood pressure value
Diastolic Blood Pressure	Examination	`ap_lo`	int	Diastolic blood pressure value
Cholesterol	Examination	`cholesterol`	categorical	1: Normal, 2: Above normal, 3: Well above normal
Glucose	Examination	`gluc`	categorical	1: Normal, 2: Above normal, 3: Well above normal
Smoking	Subjective	`smoke`	binary	0: No, 1: Yes
Alcohol Intake	Subjective	`alco`	binary	0: No, 1: Yes
Physical Activity	Subjective	`active`	binary	0: No, 1: Yes
Cardiovascular Disease	Target Variable	`cardio`	binary	0: No Disease, 1: Disease Present

All values were recorded at the time of the medical examination.

📌 Dataset Source

The dataset is publicly available on **Kaggle**:

Cardiovascular Disease Dataset