Bajaj Finserv Health – Data Science Q2

29th April 2025

Duration: 4 hours

Problem Statement:

Develop a scalable and accurate solution to process lab reports with the objective of extracting all lab test names, their corresponding values, and reference ranges. Use the provided dataset of lab report images to build logic/model.

The logic must be implemented in Python and deployed as a Fast API service. The API should expose a POST endpoint **/get-lab-tests** that accepts an image file as input and returns the extracted lab test data in a structured JSON format.

Dataset link: **https://drive.google.com/file/d/1LzG7oJ-cqGHK9KbwXnWfkWgnQ3xi8Cr9/view?usp=sharing**

Output format:

```json
{
    "is_success": true,
    "data": [
        {
            "test_name": "HB ESTIMATION",
            "test_value": "9.4",
            "bio_reference_range": "12.0-15.0",
            "test_unit": "g/dL",
            "lab_test_out_of_range": false
        },
        {
            "test_name": "PCV (PACKED CELL VOLUME)",
            "test_value": "48.7",
            "bio_reference_range": "36.0-46.0",
            "test_unit": "%",
            "lab_test_out_of_range": true
        }
    ]
}
```

*lab_test_out_of_range** must be calculated after obtaining the bio_reference_range, it's a Boolean value describing if the test lies outside the normal range.

*is_success** (Boolean flag) - Returns True if the API call is executed successfully.

IMPORTANT NOTE

The use of LLMs such as GPT, Gemini, Claude, or any open-source LLMs as part of the final solution is strictly prohibited.

Participants must build a custom logic or model without integrating any LLM into their implementation.

Any submission violating this rule will be automatically disqualified

Best of Luck!