Please answer the following questions using github/gitlab/ bitbucket repo (Select the one that is a good fit for you). Finally, please provide a public repo link so the hiring manager can read your answers.

Repo: <https://github.com/ihsiukaoBerkeley/FIT-IT>

**Task 1: General Questions**

1. What is your preferred language when building predictive models and why?

My preferred language for building predictive models is Python. Python has several powerful packages for model building, such as sci-kit, TensorFlow, and PyTorch. Moreover, packages like Numpy, Pandas, Matplotlib, Seaborn, and PySpark are useful for data manipulation and visualization.

1. Provide an example of when you used SQL to extract data.

During the shrimp farm automation project, I used SQL to extract data from AWS Timestream, which stored time series sensor readings from our POC site. I mainly performed SQL queries for two tasks: visualization and experiments. When constructing a dashboard using Grafana, I ran simple queries to extract historical data points for graphs and complex queries with aggregate functions to display averages, medians, and more. In addition, I used SQL to extract data to evaluate experiment results from tuning various equipment and environmental parameters.

1. Give an example of a situation where you disagreed upon an idea or solution design with a co-worker.  How did you handle the case?

My colleague and I disagreed about which backend platform to use for a project POC. I proposed building our platform using AWS services because they provided the highest customizability. On the other hand, my colleague believed we should use an existing SaaS solution because of its user-friendly UI and minimal maintenance requirements. We met to brainstorm the pros and cons of each option and aligned the decision with the client's requirements. Ultimately, we concluded that my colleague's solution was more suitable for that specific client. We also agreed to follow the same procedure for future projects to match different clients' requirements.

1. What are your greatest strengths and weaknesses and how will these affect your performance here?

My greatest strengths are my eagerness to acquire new knowledge and my willingness to listen to others. Staying updated with important papers, SOTA models, and industry news is crucial. My eagerness to learn ensures I keep up with these developments and enables me to bring innovative solutions to the team. My willingness to listen encourages me to collaborate with others and helps bring different perspectives to my work, improving the overall quality of the project.

One of my weaknesses is that I can get tunnel vision after working on a task for too long. I may forget to occasionally take my engineering glasses off and put on other perspectives. To mitigate this, I would take regular breaks and seek fresh perspectives from my colleagues to ensure my work aligns with the goal.

**Task 2: Python model development**

**Objective**: Given the dataset, **train.csv** - the training dataset; **Exited** is the binary target and **test.csv** - the test dataset; your objective is to predict the probability of Exited, write a python script (main.py) that when run (e.g. python main.py) will output:

* a CSV file containing the following:
  + Predicted Exited from test.csv
  + Evaluation of the predictive ability of the model.(e.g. F1 Score, Confusion Matrix…etc)
* Plots that can help us visualize the classification data and the prediction curves.
* Please submit codes, explanations, and plots when finished. Try to be more **specific**, a README might be helpful.