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.

**Task 1: General Questions**

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

I prefer Python because most predictive models and associated packages are built with it. Additionally, Python's widespread use makes it easier to collaborate with software engineers and integrate with various tools and systems.

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

When I work at Advantech, I need to get the sales orders and product information for training models through SQL.

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?

In a previous project, my colleague and I had a disagreement over the approach to handling data preprocessing for a machine learning model. They preferred using a complex feature engineering technique, while I believed a more automated and more efficient approach by establishing preprocessing modules.

To address the disagreement, I first listened carefully to their rationale and understand the benefits they saw in their approach. I then presented my perspective, focusing on how the modules could streamline the process and potentially lead to more reproducible results.

Ultimately, I convinced my colleague to adopt my approach with blending the strength of complex approach into the modules.

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

**Greatest Strengths:**

1. **Analytical Thinking**: One of my greatest strengths is my ability to approach problems analytically. I excel at breaking down complex issues into manageable components. This strength will help me contribute to creating robust models and making data-driven decisions that drive business outcomes.
2. **Proficiency with Tools and Technologies**: I have extensive experience with a range of data science tools and programming languages, including Python, R, SQL, and various machine learning libraries such as Pytorch and Scikit-Learn. This proficiency ensures that I can quickly adapt to new technologies and integrate them into our workflow, enhancing our team's productivity and the quality of our analyses.
3. **Collaboration and Communication**: I am skilled at translating complex technical findings into clear insights for stakeholders who may not have a technical background. This ability to communicate effectively will facilitate better collaboration across teams and ensure that our data-driven recommendations are understood and implemented effectively.

**Greatest Weaknesses:**

1. **Limited Experience with Certain Domain-Specific Tools:** While I am proficient with a broad range of data science tools, I have less experience with some specialized domain-specific tools or platforms that may be used in specific industries. To address this, I am committed to continuous learning and actively seeking opportunities to gain familiarity with new tools and technologies relevant to the domain.
2. **Delegation:** In the past, I’ve had challenges with delegating tasks effectively. I often prefer to take on complex tasks myself to ensure they are done to a high standard. However, I’ve been working on improving my delegation skills by building trust in my team and providing clear instructions and support, which helps in leveraging the team's strengths and ensuring timely project completion.

**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.