Snapp!

Senior Machine Learning Engineer

Ba Salaam!

Congratulations on reaching this stage in the hiring process for the Senior Machine Learning Engineer position!

We have designed an assignment to assess your approach to the proposed problem. This will allow us to evaluate your technical skills, problem-solving abilities, and familiarity with machine learning, data science, and MLOps practices.

The following sections describe the assignment in greater detail.

Introduction

Here at Snapp!, we do more than 4 million daily rides. The data contains many patterns and relations, and we are committed to leveraging data to drive intelligent decision-making, add value to the business, and improve user experiences (Passengers and Drivers).

Every day, passengers make many requests, some converted to actual rides. The passenger or the driver may cancel the rest for various reasons beyond the scope of this brief text. The problem you will deal with is one of them, which is about predicting the possibility of finishing a requested ride by a driver.

Dataset Description

In this portion, you are tasked with developing a predictive model to predict whether a given request will transform into a finished ride (label). To achieve this, you can employ data found in the `train.csv` file for your training phase and `test.csv` to predict labels for this dataset with your trained models. The table below lists the columns of the data, as well as some descriptions of them.

Column Name	Description
ID	A unique identifier (ID) for each ride
UserID	A unique identifier for each driver
Created_at	The created timestamp of the ride
Origin	A unique identifier for the origin location
Destination	A unique identifier for the destination location

Snapp!

Time	The duration of the ride in minutes
Income	The price of the ride
Comment	Previous comments of passengers for a driver
Label (IsFinished)	The target variable that you need to predict

Task Description

The primary goal of this task is to predict the label associated with each ride. We expect you to demonstrate your proficiency in data preprocessing, feature engineering, model selection, evaluation, and version control using Git.

Deliverables for this component might include:
- Exploratory data analysis (EDA) report
- Feature Engineering & Feature Selection
- Model (Models) Selection and Training, and Model Evaluation (including code, documentation, and any assumptions you have made). Please make sure that at least one of the trained models is a neural network.
- Labels for data in `test.csv`
- Error Analysis
- Model Interpretability
- Model Deployment (Creating a web framework, containerizing the application, …)
- A `README.md` file as runbook of the project

Version Control

Throughout the task, use Git to track changes in your project. Commit your code regularly with meaningful commit messages that describe the changes made. Push your final code, documentation, and additional files to a Git repository.

Documentation

Provide detailed explanations of each step during the data preprocessing, feature engineering, model selection, and evaluation phases. Your documentation should be clear, concise, and easy to follow.

Snapp!

Submission Requirements

Your submission should include:

- A Git repository URL containing all the code used for data preprocessing, feature engineering, model training, and evaluation.
- A detailed report documenting your approach, findings, and justifications for the decisions made.
- A detailed description of your approach to model deployment, including code if applicable.
- A README.md file with instructions on how to set up and run your code.
- You can add your report as .md files to your Git repository
- Please include all of your analysis and implementations, and do not exclude them based on their irrational or unexpectable results

Due to your limited time, we understand that you may only deliver some of the requested items as thoroughly as possible; however, we strongly encourage you to respond to each requested item.

Thank you very much for your time and efforts in performing this assignment.
We look forward to seeing your approach and solutions.

Good Luck!