Footbar Data Scientist

Welcome to the Footbar Data Science Challenge! This case study aims at evaluating your capacity to build an artificial intelligence architecture, build a proper dataset, and make a good choice of technologies and libraries to build a state-of-the-art infrastructure without reinventing the wheel.

The problem: Generative AI recreating football game

In this task, you will be provided with accelerometer data capturing the gait of football players, including their acceleration, and corresponding labels indicating the actions performed during the game. Your objective is to use this data to recreate the football game by analyzing player movements and actions.

During player movement, the start and end of the gait is considered the moment when the player foot on the ground, with zero speed. This moment is also known as the **stance period**, or just **stance**. Furthermore, there is always **one action** of the player per one gait. The example of such action is shot, pass, tackle, resting, etc. For identification of such action, you will use the **acceleration** of the player. This acceleration is generated from our tracker called Meteor which embeds an accelerometer providing the 3-dimensional acceleration of the leg throughout time, at a 50Hz frequency. For this task we will provide only the **norm** of this 3-dimensional acceleration. The mentioned data can be found on this folder.

The input data are nested list of dictionaries, where each element represents one gait of the player of the session, including the norm of the acceleration and label of the action through the gait, one element is shown below:

```
{
        'norm': [24.3,5.6,7.8,10.3,22.8,…],
        'label': 'shot'
}
```

Description of some labels:
- dribble – every time player is doing with the ball except shooting/passing, even when player   receives the ball
- tackle – every time player tries to tackle the opponent, block the shot, etc.
- no action - it is considered as unknown action, or if there is nothing happened, or it was hard to decide what player did.
- shot - every time the player shoot in the current gait
- pass - if the player pass to the another player, but ball is on the ground
- cross - if the player pass, but the ball is above the ground.
- run - if the player is running or he/she is making positional defending

- walk - if the player walk - e.g after the goal, or if he is in defense and waiting for the action
- rest - when player is not moving

Because a typical football game consists of a combination of player walking, sprinting or jogging, the size of each gait is different. Furthermore, gaits are consecutive, i.e if we consider gaits as a time series then gait_1 < gait_2 < … < gait_n.

We expect the following:

1. Make a simple exploration data analysis of the input data. Can you find some differences between some different actions?

2. List of different approaches of how such game can be re-created. You can also use consider the other type of the data than we are providing.

3. Use the attached data for recreating the game. We expect the same format of the output as input. More precisely, we expect that the recreated game will be a nested list of dictionaries, where each dictionary will include two keys: norm and label. Some hints:

   - The sequence of labels should make sense, e.g. sequence of actions {shot, shot, shot, shot, shot} could be suspicious

   - Always check the gait length, e.g. be aware if the gait is not too short or too long (e.g 0.1 s of gait would be too short, and gait of length 3s would be also too long)

4. Please fully describe at least one approach you would choose (in jupyter notebook or some additional pdf):

   a. The chosen architecture/algorithm. Why the decision was made, why it makes sense, and what kind of input it assumes. If the mathematical theory for the chosen approach is too complicated, the flow chart is enough.

   b. The pre- / post-processing of the data

5. Parametrize your fitted algorithm/program for recreating the game in the following way:

    a. It will be possible to generate as many games as we want

    b. It will be possible to generate the game of any length (e.g game of 15, 20 or 60 minutes).

    c. **Bonus:** It will be possible to generate game with specific type of play - e.g more attacking game (there will be more passes, shots), defending game (more tackles, interceptions, etc.), or just normal game.

6. Send us your code as a git repository containing instructions to run your code (or invite us as a collaborators). Be sure that your code will be reproducible at any machine (python env, dev containers or docker are your friends). Furthermore, as was mentioned above, make sure that your solution can be run with one-liner, and it will generate the expected output. You may use Github but take care of not including the word Footbar, the public link to the dataset, or anything that could allow the other applicants to browse for your work!

7. Evaluation will be done on our side. What we do is that we will generate some games, and we will check if the generated game makes sense or not.