

# Candidate instructions: Hired/Not-Hired classifier for Pokemon Trainers

You are asked to give advice for teams of Pokemon Trainers all across Pokemon World who are looking for new trainers to add to their team.

Trainers fall into one of 8 classes (*Curmudgeon*, *Doctor*, *Dragon Tamer*, *Engineer*, *Nurse*, *Pokemon Ranger*, *Scientist* and *Skier*). In each region of Pokemon world, and for each Trainer Class, there's a distinct team that's hiring (regions are *Kanto*, *Johto*, *Hoenn*, *Sinnoh*, *Unova*, *Kalos*, *Alola* and the *Sevii Islands*). A team can be uniquely identified by trainer class and region.

Given data from job applications for Pokemon Trainers:

1. Which team (i.e. combination of region and trainer class) is the most competitive (meaning the most difficult to get hired into)?
2. Train a model that predicts whether candidates will get hired or not.
3. If you were a team hiring a Pokemon Trainer, how would you use this data to make the hiring process easier?

Please aim to spend no more than 3 hours on the entire exercise and explain your thought process and any choices you make, such as how to deal with missing values and which columns to use, as well as how you evaluate your model.

Please submit all the files necessary to run your code (for example a Jupyter notebook with Python code or an R markdown file), together with any instructions on how to run it, including which packages must be installed.

## Data

The applications of aspiring Pokemon Trainers can be found in the file `pokemon_trainer_application_data.csv`

## Columns

- Target column: `hired` (0 or 1, where 1 means "hired")
- Columns indicating team are a combination of region in Pokemon world and Trainer class: (`PokemonWorldRegion`, `PokemonTrainerClass`)

Furthermore, there are many more columns, of which you should decide whether you want to use them or not. Some have to do with a trainer's previous education or training experience, or which Pokemon they have won badges with (for instance `GymBadge1Pokemon`). Other columns may contain details of the position for which they're applying (for instance

PositionForTrainingPokemon which states which Pokemon they will be training), or details about the application itself (for instance ApplicationType or ApplyDate).

You don't need to know anything about Pokemon to do this exercise, but for your information, there are three types of Pokemon-specific names. These are present in the following columns:

Region names appear in columns: PokemonWorldRegion

Trainer classes appear in columns: PokemonTrainerClass

Pokemon names appear in columns: CurrentPokemonTraining,

PositionForTrainingPokemon, GymBadge1Pokemon, ..., GymBadge4Pokemon. A list of Pokemon names can be found in the file pokemon\_names.csv