Data Science II Final P roject Proposal

yy3019 Yuanzhi Yu, zl2860 Zongchao Liu, wk2343 Gavin Ko

3/13/2020

1. Project Title:

Statistical Pokemon Master - Predict Pokemon's Species Strength by Mutiple Factors

2. Introduction:

For people like us in their 20's, Pokemon - the imaginary creature created by The Pokemon Company, has played a siginificant role in our childhood. The variety of appearance, vivid characterisitcs and detailed design made this IP fascinating. For Pokemons, subtantial statistics like height, body weight, type and species strength are attributed to each of them.

The initial process of acquiring a pokemon in games is catching them. While some of them are tame and easy to catch, the others are more hostile and reluctant to be caged. Accordingly, a reliable catch rate estimation could help with trainer's decision making when encountered with a specific pokemon. On the other hand, a few Pokemon stands out from their peer and being classified as "lengendary", which indicates either uniqueness or being powerful.

Another intriguing factor of the Pokemon world is battling. Pokemon trainers would train their companion with different skills, and would bring them up to one versus one battle. Generally speaking, Pokemon with higher species strength are considered stronger with higher chance to win in a fight. Therefore, we want to predict Pokemons' species strength by multiple characteristics attributed to it.

3. Data Description:

1) Source

https://www.kaggle.com/alopez247/pokemon

2) Variable Description

Response

- Species strength: Sum of all the base stats (Health Points, Attack, Defense, Special Attack, Special Defense, and Speed).
- Catch_Rate: Catch Rate. The probability of cathing the pokemon.
- isLegendary: Boolean that indicates whether the Pokémon is Legendary or not.

Predictors

- Number: Pokémon ID in the Pokédex.
- Name: Name of the Pokémon.
- Type_1: Primary type.
- Type_2: Second type, in case the Pokémon has it.

- Total: Sum of all the base stats (Health Points, Attack, Defense, Special Attack, Special Defense, and Speed).
- HP: Base Health Points.
- Attack: Base Attack.
- Defense: Base Defense.
- Sp_Atk: Base Special Attack.
- Sp_Def: Base Special Defense.
- Speed: Base Speed.
- Generation: Number of the generation when the Pokémon was introduced.
- Color: Color of the Pokémon according to the Pokédex.
- hasGender: Boolean that indicates if the Pokémon can be classified as female or male.
- Pr_male: In case the Pokémon has Gender, the probability of its being male. The probability of being female is, of course, 1 minus this value.
- EggGroup1: Egg Group of the Pokémon.
- EggGroup2: Second Egg Group of the Pokémon, in case it has two.
- hasMegaEvolution: Boolean that indicates whether the Pokémon is able to Mega-evolve or not.
- Height_m: Height of the Pokémon, in meters.
- Weight_kg: Weight of the Pokémon, in kilograms.
- Body_Style: Body Style of the Pokémon according to the Pokédex.

4. Interested Questions:

- How are the features of a Pokemon going to affect it's species strength?
- Is there a specific trend/ difference between generations on general strength factors?
- Given a set of factors, can we predict whether a pokemon is lengendary?