Saranga Rajagopalan
June 27, 2022

Analysis of Pokemon Dataset and Classification of Legendary Pokemon

Problem: Is it possible to build a classification model to identify legendary pokemon?

After my analysis of the Pokemon Dataset, I conclude that a classification model to identify legendary pokemon with high accuracy can be built after the correction of the following discrepancies I discovered during my analysis:

1. Data Completeness

1.1 Problem: Features had many null values.
The following features contained many null values:

- *percentage_male,*
- *type2*
- *height_m*
- *weight_kg*

1.2 Problem: Non-homogenous data types within features
The *capture_rate* feature contains object(str) values as well as int64 values.
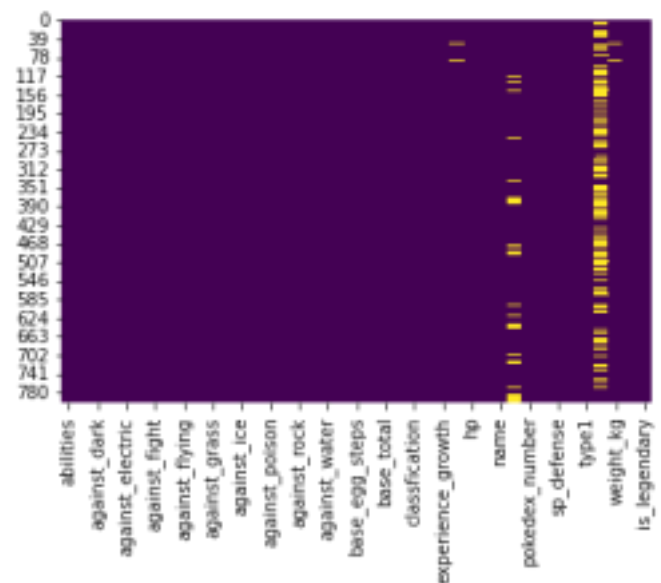

1.2 Recommendations:
- Collect more data on the height and weight of different pokemon.
- Add a category for genderless pokemon, as the majority of null values from. *percentage_male* are from genderless pokemon.
- Collect more data on pokemon with a valid *type2.*
- Reformat values within features to be the same across pokemon for consistency.

2. Data Relevance

2.1 Problem: Features in the dataset were not significant to determining if a pokemon is legendary
The following features were removed because they did not affect the the legendary status of the pokemon:
- *abilities*

- *classification*
- *name*
- *type1*
- *type2*

2.2 Recommendations:
- Collect more relevant statistical information to better the results of classification **3.**

## Multicollinearity of Predictors

3.1 Problem: Multicollinearity is suggested by the dataset
The following factors suggest multicollinearity in the dataset:
- Correlation Matrices
  - The correlation matrices show a high correlation between the following independent variables:
    - *base_total - attack, defense, sp_attack, sp_defense, speed*
- Variance Inflation Factor
  - The values for *base_total, defense, sp_attack, sp_defense,* and *speed* are infinity, and the values for *base_eggs_steps* and *base_happiness* are very large (*pokedex_number* can be ignored)

18	attack	inf
19	base_egg_steps	12.820566
20	base_happiness	20.207886
21	base_total	inf
22	capture_rate	5.174524
23	defense	inf
24	experience_growth	53.508518
25	height_m	4.545480
26	hp	inf
27	pokedex_number	209.208980
28	sp_attack	inf
29	sp_defense	inf
30	speed	inf
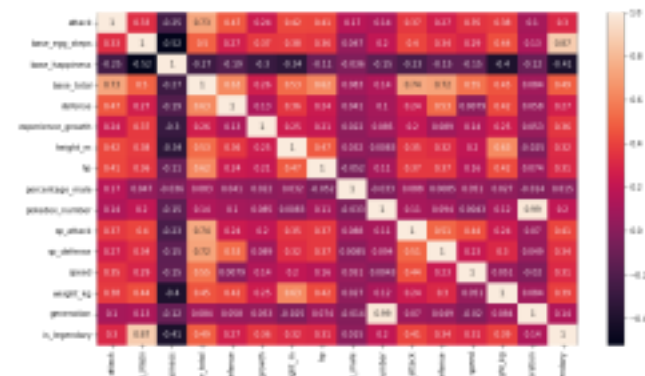
3.2 Recommendations:
- Combine the defense and attack features into their own respective features ○ For example, add the pokemon's attack and special attack stats to create a feature such as *total_attack_power*
  - For example, add the pokemon's defense and special defense stats to create a feature such as *total_defense_power*