Final Project Proposal

November 2020

Introduction to Pokémon and the Dataset: The "Pokémon With Stats" data set is a set that contains 721 Pokémon along each individual ID number, type, secondary type, and combat statistics. These 721 Pokémon span across 7 generations, compiling the entire history throughout the Pokémon video games. In the Pokémon games, depending on the situation, we may want a well-balanced team, a more defensive minded team or an offensive minded team. We hope that we can explore this dataset in a way that will help us with our team building strategies.

Clustering: What is it that makes Pokémon similar to one another? Is a Pokémon's classifier attack strength, defense strength, type? In this section we will examine the ratio of defensive and offensive statistics of Pokémon. This ratio will allow us to define a scale of offensive and defensive capabilities of selected Pokémon, allowing us to determine whether or not a Pokémon would be better suited protecting your gyms, or with you challenging to capture other gyms. A gym is a collection of other Pokémon trainers who you must defeat in order to capture the gym. The main goal of Pokémon is to capture as many gyms as possible.

Dimensionality Reduction: Which particular stats/other data does not correlate to the results that we're trying to find? When creating our combat team, we must observe and interpret the data that we are given in order to find the best possible result. Using Principal Component Analysis(PCA) we can determine whether or not certain points in our dataset skew the final interpretation, resulting in a clean, easy to read solution.

Regression/Classification: Are Electric types the fastest type of Pokémon? Do Psychic types have the best special attack? In order to construct our ideal Pokémon team, we must ask these questions in order to build a team that will fit what build we are looking for. For instance, if we want to create a speedy team, we might look into specific Pokémon types in order to fill out our roster. We want to find if there is any meaningful correlations between specific Pokémon types and stats. We can take subsets of our "Pokémon With Stats" dataset, broken up by type, and perform a regression analysis on each subset in order to see which stats are most strongly correlated to which types. If we find this to be the case, we want to test to see if we can predict a Pokémon's type by looking at its individual stats through classification.