

For my final project, I used my CSV dataset of the first 500 Pokemon by ID number that I had built and used in previous assignments for this course. In particular, I wanted to see if it was possible to predict a Pokemon's primary type (water, fire, etc.) using its physical attributes (height, weight) and battle attributes (health points, attack, defense, etc.). I hypothesized that it was indeed possible to predict primary type based on the given variables and that there was a solid correlation between them.

I began by using Seaborn to create visualizations of the physical and battle attributes of the different Pokemon types. In order to do this in an organized manner, I had to first create one extra column to represent the general battle attribute total, or "strength" of a Pokemon. I had previously done this in my original Seaborn assignment earlier this year, so I simply recreated this column as well as the scatter plot I had also created in said previous assignment representing the average battle statistic total (or strength) of each of the primary Pokemon types, as I felt it would be useful to see for this assignment. My scatter plot showed decent variation, with Dragon as the "strongest" type on average and Bug as the "weakest". I repeated this process for physical attributes, creating a general "size" column by combining height and weight, and creating a bar graph representing the average size for each Pokemon type. This bar graph showed Steel to be the "biggest" type on average and Fairy to be the "smallest". I then wanted to see if there was any sort of correlation between size and

strength as a whole regardless of type, so I plotted the average size for each of the types against the average strength of each of the types as a Seaborn “regplot”. The resulting plot showed me that there was somewhat of a positive correlation between the two, meaning that bigger Pokemon tend to generally be stronger.

I began the testing of my hypothesis by creating clusters of Pokemon based on their physical and battle attributes. I used K-means clustering to break the dataset into 25 clusters based on all physical and battle attributes individually as well as the previously created size and stat total columns. I had expected to see that the clusters would each mostly contain Pokemon of the same type, however this was not the case as the clusters did not support my hypothesis. Although I did not include this data when I had initially created my dataset at the beginning of the semester, I was able to use my knowledge of Pokemon to see that the Pokemon were mostly clustered based on evolutionary stage instead of type. Although this goes against my initial hypothesis, this makes sense, as when Pokemon that can evolve do evolve they are almost always going to get bigger and stronger than their previous evolutionary stage.

I wanted to further test the validity of my hypothesis by classifying my dataset. In particular, I wanted to see how accurately a Decision Tree classifier could predict primary type given physical and battle attributes. Although I had some trouble with getting this to run initially, as there was some confusion over how the bins needed to be set up for the labels, I was able to print debug using my training and test sets and got the classifier to run with an accuracy score of 0.17. Just like clustering, this method

once again goes against my hypothesis, as such a low accuracy score means the classifier could not predict primary type very accurately given the provided attributes. However, it did provide me with some useful insight, as the visualization of the decision tree showed the most useful attributes in predicting typing to be special attack, attack, and defense. This makes sense as different Pokemon types tend to specialize in either special or non-special battle attributes.

In the end, my hypothesis was generally disproven by my testing. However, given more time, I would have loved to explore the correlation between typing and battle attributes without physical attributes, as my decision tree classifier leads me to believe there may be some correlation there. I would have also liked to explore the relationship between physical and battle attributes and evolutionary stage, as my clustering leads me to believe that there is some correlation between those variables as well.