Data Mining Project Proposal

Group Members: Yuet Ho, Mir Shahiduzzaman, Abby Fontana, Ari Markos, and Tanvir Ahmed

**Data Set:** [Complete Pokemon Dataset (Updated 16.04.21)](https://www.kaggle.com/mariotormo/complete-pokemon-dataset-updated-090420)

Our data set is a complete set of data on the 1028 Pokemon, which are creatures in a popular video game series. These creatures come in various shapes, sizes, and types and our data including Pokedex number, English name, German name, Japanese name, which generation the Pokemon was introduced in, whether the Pokemon is legendary, sub-legendary, or mythical or not, species, and much more. Overall, there are 52 different features available for each pokemon. These different features along with the size of our data set should be enough to do analysis on.

**Problem:**

We want to utilize this data to study the behavior and quality of each Pokemon based on typing. There are 18 Pokemon types (Normal, Fire, Water, Grass, Flying, Fighting, Poison, Electric, Ground, Rock, Psychic, Ice, Bug, Ghost, Steel, Dragon, Dark and Fairy) and all of them have different movesets and type advantages/disadvantages. For example, grass is weak against fire and fire is weak against water and so on. Each individual pokemon also has different stats such as hp (health points), attack, defense, and more. Our goal will be to find the strongest type of Pokemon by looking at all of these factors in order to help players consider which typings to pick in game.

**Proposed Methods:**

We will look at the different stat values and other data points to try to determine what makes a strong Pokemon. After we have figured that out, we will use different models such as naive classifier, decision trees, etc. and see which model works the best for our data set. Using the model that has the best results, we will run it with all of the relevant features and find out which Pokemon type is the strongest.