

Zachary Evans
Calvin Pugmire
Timothy Lencioni

Problem:

Our goal with this project is to make a predictive model that given 2 teams of pokemon names is able to predict which team will win in a competitive Gen 9 Regulation F Battle. This will involve making a merge on the pokemon given to have included their base stats and their typing. What this problem does not include are individual IV's (individual points) given to the pokemon and ev's (effort values) given to the pokemon as well as the experience of the trainer. However we are hoping to see some form of accuracy appear from this outcome.

Gathering Data:

We have 2 sources of data we are working on. One is a smaller data set with data gathered by the Pokemon Collegiate Video Game Championship League which consists of colleges nationwide competing against each other. Additionally we will be using the second link to gather an additional 5000 rows of data but to do this we will be using beautiful soup4 (a python package) to gather this data.

<https://docs.google.com/spreadsheets/d/1l81oXaNUirQnSGuFbW9bLpFPNNWW2Ejlb9Z3qjEsoal/edit#gid=174807770>
<https://replay.pokemonshowdown.com/>

Description of dataset and the process:

Taking a look at the dataset in the first link we have a table similar to this:

Timestamp	What week are you reporting for?	Format	Winning Team	Winning player's Discord Username (all lowercase)	Winning Player's 6 Pokemon Copy paste from Pokemon Showdown Replay, as shown in the image:	
9/22/2023 17:13:17	Week 1	Regulation E	Electric - Southeast	hopesc2	Iron Hands / Chien-Pao / Landorus-Therian / Ogerpon-Hearthflame / Tornadus / Flutter Mane	
Losing Team	Losing player's Discord Username (all lowercase)	Losing Player's 6 Pokemon Copy paste from Pokemon Showdown Replay, as shown in the image.	Score	Game 1 Replay (or Youtube link)	Game 2 Replay (or N/A if Youtube)	Game 3 Replay (or N/A if no third game or Youtube)
Electric - Northeast	punkcharizard	Baxcalibur / Urshifu-* / Gholdengo / Arcanine-Hisui / Roaring Moon / Torterra	2-0	https://replay.pokemonshowdown.com/gen9vgc2023regulation-1950002970-3pw22gzmhupgdqod0pa0jgx80ws8orwpw	https://replay.pokemonshowdown.com/gen9vgc2023regulation-1950003328-9no2wpv8hg94e4zrovis5nzhv00wstppw	N/A

When Looking at this table we only need to focus on 2 pieces of information, the winning team's 6 pokemon and the losing team's 6 pokemon.

After dropping the additional rows we then can expand our data so that we have each pokemon in its own row for the above example that will look like (in a 12x1 array):

Iron Hands	Chien-Pao	Landorus Therian	Ogrepon-Hearthflame	Tornadus	Flutter Mane
Baxcalibur	Urshifu-*	Gholdengo	Arcanine-Hisui	Roaring Moon	Torterra

Additionally we will add one more column containing which team won since for this dataset it is always the first team it will be 0. Afterwards we will be able to further expand each pokemon to look as follows:

Type 1	Type 2	Hp	Defense	Sp. Defense	Attack	Sp. Attack	Speed
6	7	78	78	85	84	109	100

This means to start this project we will be training with around 5500 instances (500 from CVL and 5000 from Smogon) and each instance will have 12*8 or 96 features with the results.

Machine learning model:

We are thinking since we are trying to determine a win lose dataset that we should use backpropagation to train our data set.

Discussion of plans:

Currently our plans are still focused on data cleaning and scraping. We have access to the CVL data so we are working on cleaning the data as much as possible while others are focusing on making a working web scraper for the showdown replay site. We are hoping to have both of these done this week so we can clean the larger dataset and start evaluating our model.