

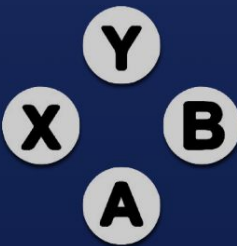
POKÉMON

Battle AI

PRESS START

Team: Izabella Freitas, Guoqing Zhang,
Tianyu Zhu, Hongyi Shen, Mary Ann Collins

The Erdős Institute
Data Science Boot Camp May 2024



START

SELECT

Project Introduction

Welcome to the world of *competitive* Pokémon!

- The “official” style of competitive battling is called Video Game Championships (VGC)
- Players ‘train’ their pokémon teams using online simulators, like *Pokémon Showdown!*
- Stakeholders:
 - Competitive Pokémon battling community, ranked and casuals players
 - Pokémon Showdown - to track changes in the metagame
 - Official Pokémon Tournaments Organizers

Question: is it possible to predict the outcome of VGC battles?



Project Goal

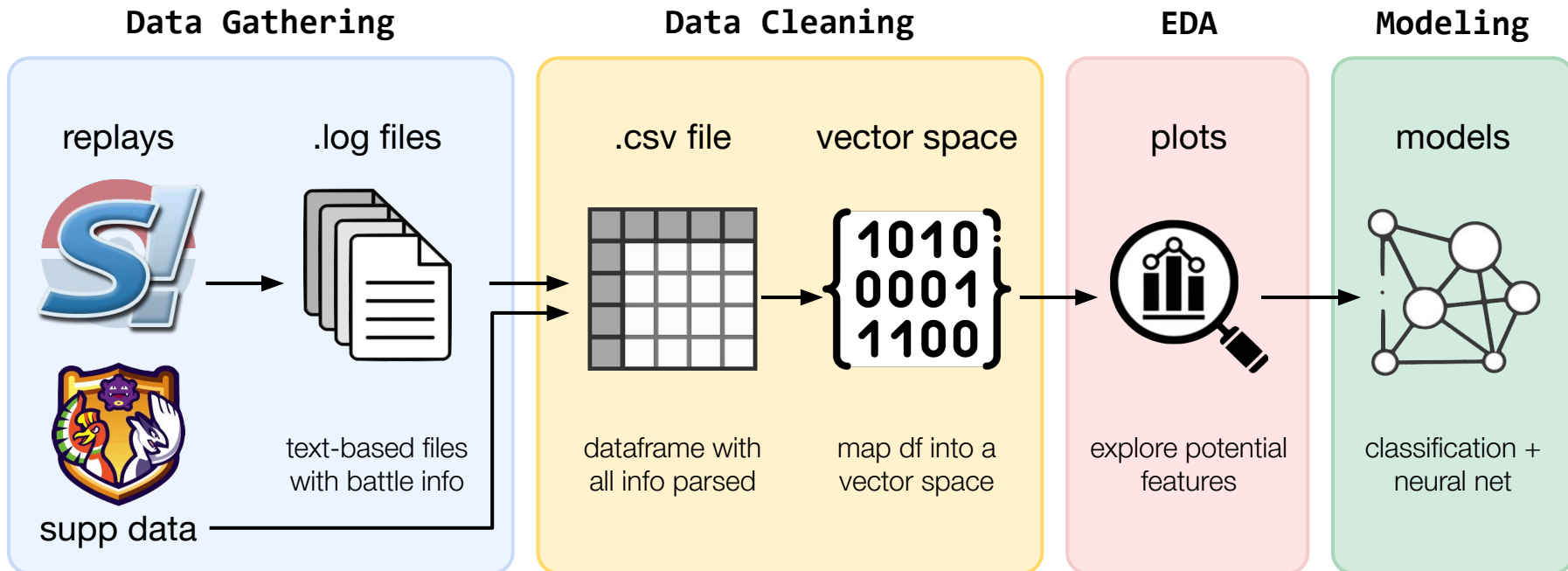
Goal: AI model to predict winners of VGC battles from *Pokémon Showdown!* replay logs

The screenshot shows a battle replay log for Turn 3. On the left, a player named 'MathHacker' has an active Pelipper (L50) and inactive Pokémon including a trainer and several others. On the right, a player named 'Breab Boy' has an active Farigiraf (L50) and inactive Pokémon including Incineroar and others. The battle field shows Conkeldurr (L50) with 49% HP and status conditions BRN and DARK, and Urshifu (L50) with 1% HP. The weather is Rain (3 or 6 turns), Electric Terrain (3 or 6 turns), and Trick Room (3 turns). A text log on the right side of the battle area contains the following text: 'The opposing Conkeldurr used Drain Punch! It's super effective! (Incineroar lost 70% of its health!)', 'Incineroar fainted!', 'The opposing Farigiraf used Hyper Voice! A critical hit! Pelipper hung on using its Focus Sash! (Pelipper lost 99% of its health!) (The opposing Farigiraf used its Throat Spray!)', 'The Throat Spray raised the opposing Farigiraf's Sp. Atk!', 'Pelipper used Hurricane! (The opposing Conkeldurr lost 51% of its health!)', and '(Rain continues to fall.)'. Annotations with blue lines point to various elements: 'weather' points to the Rain status; 'terrain' points to the Electric Terrain status; 'field' points to the Trick Room status; 'inactive pokémon' points to the inactive Pokémon list for player 1; 'active pokémon' points to the active Pokémon in the battle; '% HP' points to the HP bars; 'stat modifier' points to the 0.67x Atk and 1.5x SpA modifiers; 'status condition' points to the BRN and DARK status icons; and 'battle text logs' points to the text log on the right.

battle replay logs are **full of data!**



Workflow: overview



Data Gathering

web scraping

total: 14,474 battle.log files

replays

.log files



supp data

text-based files
with battle info

```
jj:MathHacker
jj:Brebaboy
t:1716606667
gametype|doubles
player|p1|MathHacker|picnicker|1145
player|p2|Brebaboy|linebacker|1181
teamsize|p1|6
teamsize|p2|6
gen|9
tier|[Gen 9] VGC 2024 Reg G
rated|
rule|Species Clause: Limit one of each Pokémon
rule|Item Clause: Limit one of each item
clearpoke
poke|p1|Miraidon, L50|
poke|p1|Incineroar, L50, F|
poke|p1|Ogerpon-Cornerstone, L50, F|
poke|p1|Urshifu-*, L50, M|
poke|p1|Ursaluna-Bloodmoon, L50, M|
poke|p1|Pelipper, L50, F|
poke|p2|Calyrex-Ice, L50|
poke|p2|Farigiraf, L50, M|
poke|p2|Conkeldurr, L50, F|
poke|p2|Incineroar, L50, F|
poke|p2|Iron Bundle, L50|
poke|p2|Amoonguss, L50, F|
teampreview|4
t:1716606698
start
switch|p1a: Miraidon|Miraidon, L50|100/100
switch|p1b: Incineroar|Incineroar, L50, F|100/100
switch|p2a: Farigiraf|Farigiraf, L50, M|100/100
switch|p2b: Amoonguss|Amoonguss, L50, F|100/100
t:1716606833
move|p1a: Pelipper|Wide Guard|p1a: Pelipper
-singleturn|p1a: Pelipper|Wide Guard
move|p2b: Conkeldurr|Mach Punch|p1a: Pelipper
-resisted|p1a: Pelipper
-damage|p1a: Pelipper|0 fnt
faint|p1a: Pelipper
move|p2a: Farigiraf|Psychic|p1b: Urshifu
-supereffective|p1b: Urshifu
-damage|p1b: Urshifu|0 fnt
faint|p1b: Urshifu
-weather|RainDance|[upkeep]
-damage|p2b: Conkeldurr|43/100 brn|[from] brn
upkeep
t:1716606846
switch|p1b: Miraidon|Miraidon, L50, tera:Electric|100/100
-activate|p1b: Miraidon|ability: Hadron Engine
turn|4
t:1716606878
move|p2b: Conkeldurr|Knock Off|p1b: Miraidon
-damage|p1b: Miraidon|16/100
-enditem|p1b: Miraidon|Choice Specs|[from] move: Knock Off|[of] p2b: Conkeldurr
move|p2a: Farigiraf|Hyper Voice|p1b: Miraidon
-damage|p1b: Miraidon|0 fnt
faint|p1b: Miraidon
win|Brebaboy
raw|MathHacker's rating: 1145 &rrarr; <strong>1123</strong><br />(-22 for losing)
raw|Brebaboy's rating: 1181 &rrarr; <strong>1203</strong><br />(+22 for winning)
\j*Brebaboy
\p|player|p2|
```



Data Cleaning

web scraping

replays

.log files



supp data

text-based files
with battle info

parse battle information into a dataframe

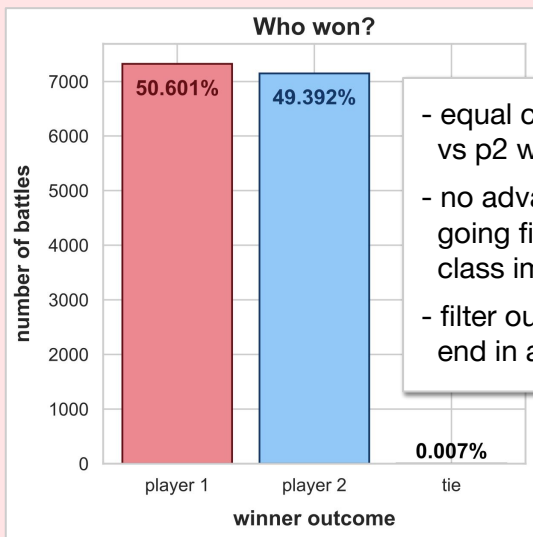
| | battle_id | turn | total_turn | rank | weather | field | condition | p1_side | p1a_form | p1a_hp | ... | p2c_status | p2c_tera | |
|--|-----------|------------|------------|------|---------|------------|-------------------|--------------|----------|--------------|-----|------------|----------|---------|
| | 0 | 2099996083 | 0 | 5 | 1643.5 | NaN | Psychic Terrain:5 | NaN | NaN | Smeargle | 100 | ... | NaN | unknown |
| | 1 | 2099996083 | 1 | 5 | 1643.5 | NaN | Psychic Terrain:4 | Trick Room:4 | NaN | Smeargle | 1 | ... | NaN | unknown |
| | 2 | 2099996083 | 2 | 5 | 1643.5 | SunnyDay:5 | Psychic Terrain:3 | Trick Room:3 | NaN | Torkoal | 100 | ... | NaN | unknown |
| | 3 | 2099996083 | 3 | 5 | 1643.5 | SunnyDay:4 | Psychic Terrain:2 | Trick Room:2 | NaN | Torkoal | 100 | ... | fnt | unknown |
| | 4 | 2099996083 | 4 | 5 | 1643.5 | SunnyDay:3 | Psychic Terrain:1 | Trick Room:1 | NaN | Torkoal | 55 | ... | fnt | unknown |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | 100019 | 2127998356 | 2 | 3 | 1464.5 | NaN | Psychic Terrain:3 | NaN | NaN | Flutter Mane | 100 | ... | fnt | unknown |
| | 100020 | 2127998356 | 3 | 3 | 1464.5 | NaN | Psychic Terrain:2 | NaN | NaN | Flutter Mane | 100 | ... | fnt | NaN |
| | 100021 | 2127999658 | 0 | 3 | 1064.0 | SunnyDay:5 | NaN | NaN | NaN | Groudon | 100 | ... | NaN | unknown |
| | 100022 | 2127999658 | 1 | 3 | 1064.0 | SunnyDay:4 | NaN | NaN | NaN | Groudon | 14 | ... | NaN | unknown |
| | 100023 | 2127999658 | 2 | 3 | 1064.0 | SunnyDay:3 | Grassy Terrain:4 | NaN | NaN | Ogerpon | 100 | ... | NaN | unknown |

100024 rows x 74 columns

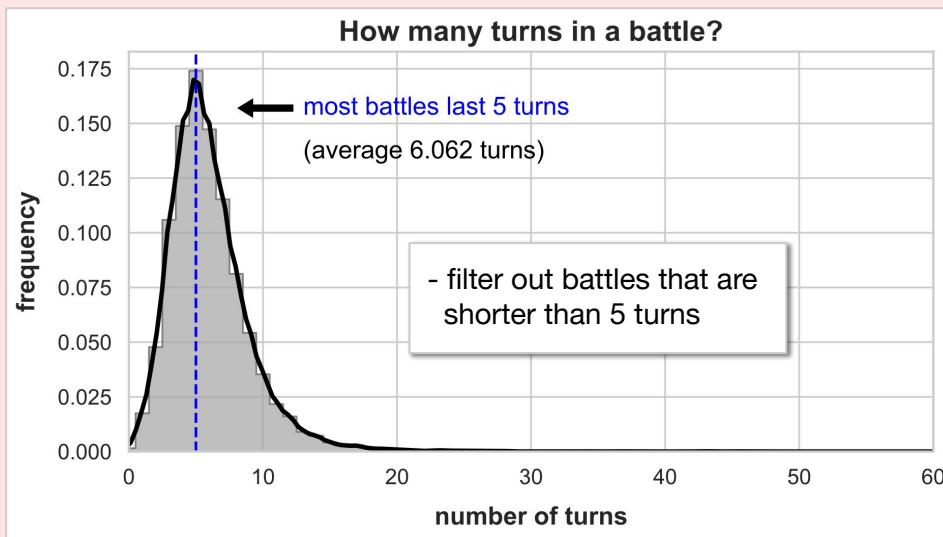
**supplemental data includes base stats and typing for all pokémon*

Exploratory Data Analysis

analyzing battle outcomes and length:



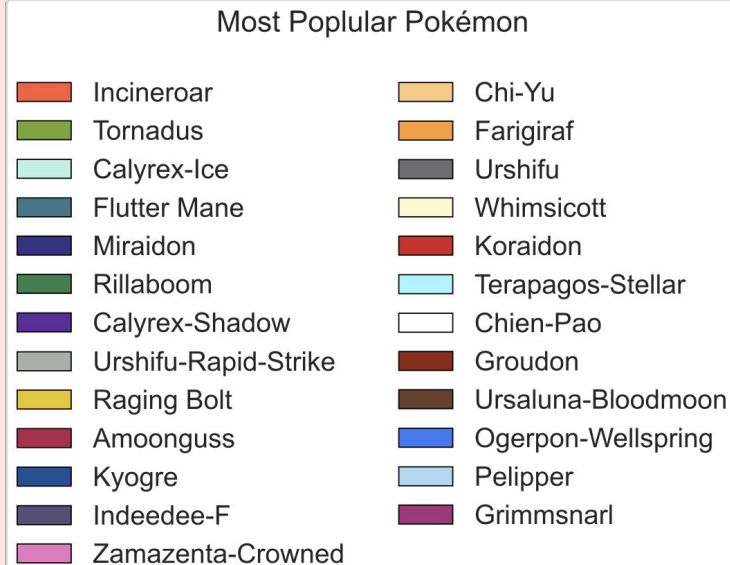
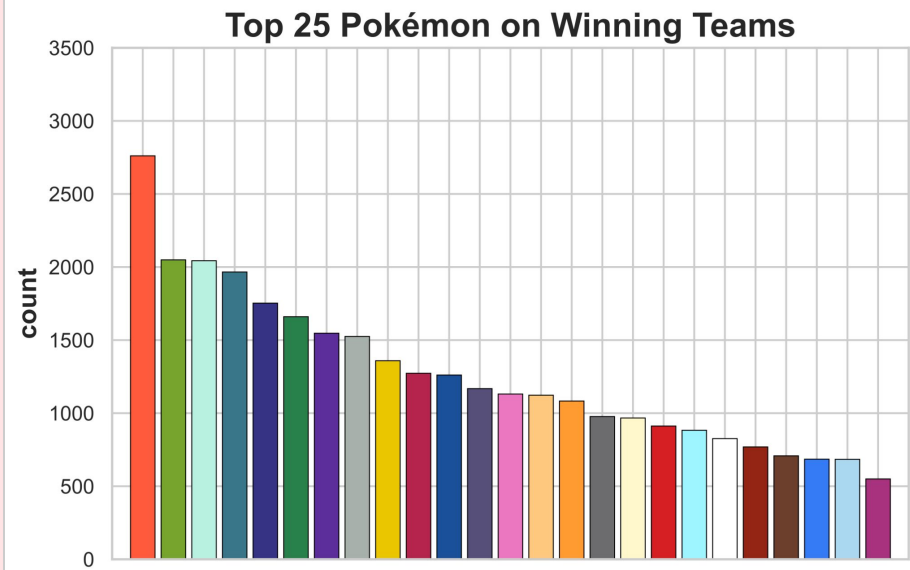
- equal chance of p1 vs p2 winning
- no advantage to going first (and no class imbalance)
- filter out battles that end in a tie



- filter out battles that are shorter than 5 turns

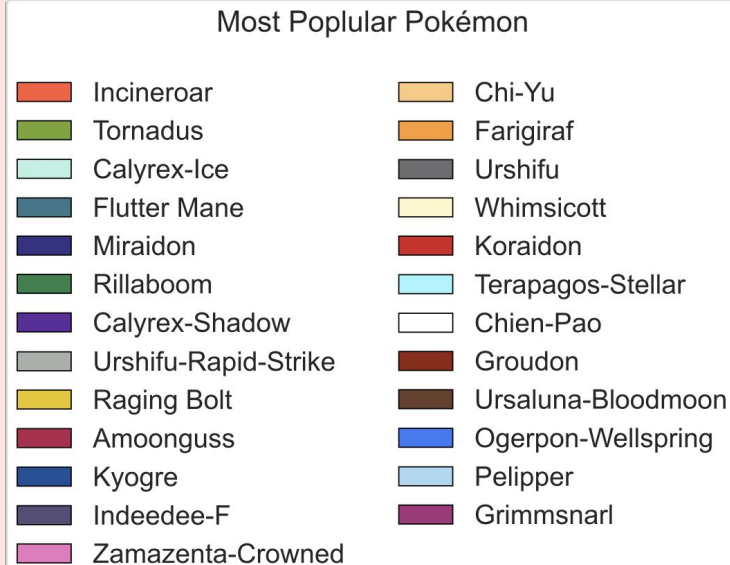
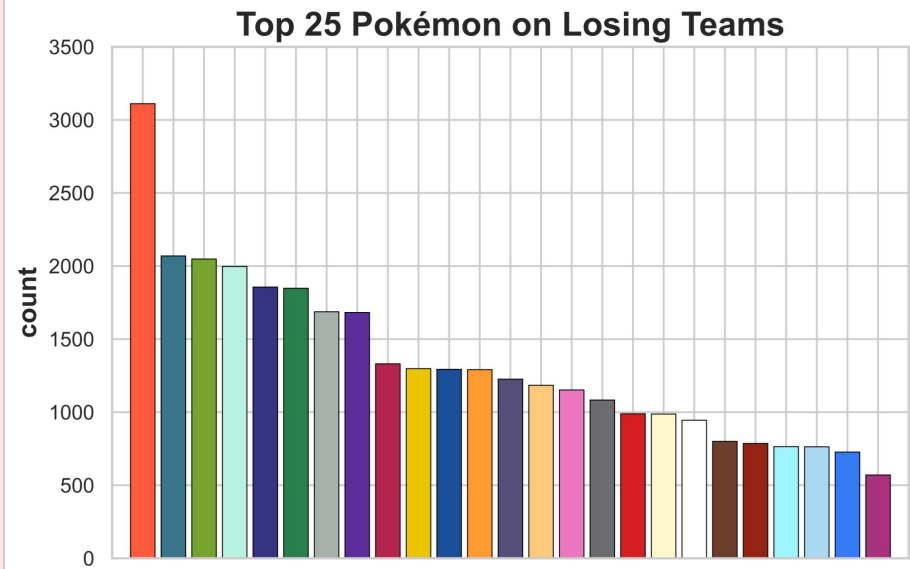
Exploratory Data Analysis

analyzing team composition: are certain pokémon more likely to win or lose?



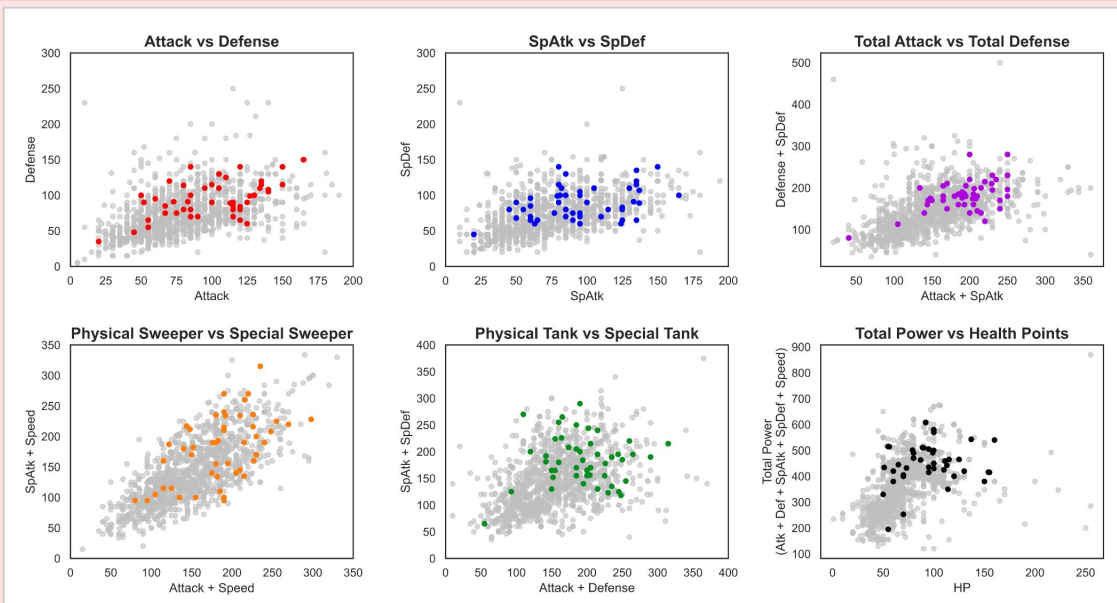
Exploratory Data Analysis

analyzing team composition: are certain pokémon more likely to win or lose?



Exploratory Data Analysis

analyzing pokémon base stats: where do the top pokémon fall?



- gray dots = stat distribution for all pokémon
- top 50 pokémon highlighted



Modeling: RandomForest Classifier

Data Input:

- focus only on the Pokémon at the start of each battle (ignoring the battle process)

Model:

- RandomForest Classifier
- ExtraTree Classifier

Accuracy: ~ 61%

Remark:

- If we consider the Pokémon on each turn and include the battle conditions, accuracy increases to 79%

```
Accuracy: 0.6117443868739205
Classification Report for RandomForestClassifier:
              precision    recall  f1-score   support

     -1         0.61         0.59         0.60         1430
     1         0.61         0.63         0.62         1465

 accuracy                0.61         2895
 macro avg              0.61         0.61         0.61         2895
 weighted avg          0.61         0.61         0.61         2895
```

```
Accuracy: 0.6082901554404145
Classification Report for ExtraTreesClassifier:
              precision    recall  f1-score   support

     -1         0.61         0.58         0.60         1430
     1         0.61         0.63         0.62         1465

 accuracy                0.61         2895
 macro avg              0.61         0.61         0.61         2895
 weighted avg          0.61         0.61         0.61         2895
```

Modeling: Classifier

Data Acquisition

All Battles Info

Names

Pokémons for
each battle

Assign
Number

Vectorized
Data

| battle_id | turn | total_turn | rank | weather | field | condition | p1_side | p1a_form | p1a_hp | ... | |
|-----------|------------|------------|------|---------|------------|-------------------|--------------|----------|----------|-----|-----|
| 0 | 2099996083 | 0 | 5 | 1643.5 | NaN | Psychic Terrain:5 | NaN | NaN | Smeargle | 100 | ... |
| 1 | 2099996083 | 1 | 5 | 1643.5 | NaN | Psychic Terrain:4 | Trick Room:4 | NaN | Smeargle | 1 | ... |
| 2 | 2099996083 | 2 | 5 | 1643.5 | SunnyDay:5 | Psychic Terrain:3 | Trick Room:3 | NaN | Torkoal | 100 | ... |
| 3 | 2099996083 | 3 | 5 | 1643.5 | SunnyDay:4 | Psychic Terrain:2 | Trick Room:2 | NaN | Torkoal | 100 | ... |
| 4 | 2099996083 | 4 | 5 | 1643.5 | SunnyDay:3 | Psychic Terrain:1 | Trick Room:1 | NaN | Torkoal | 55 | ... |

5 rows × 74 columns

```
{'Smeargle': 1,  
'Torkoal': 2,  
'Brute Bonnet': 3,  
'Tornadus': 4,  
'Tyranitar': 5,  
'Iron Valiant': 6,  
'Iron Jugulis': 7,
```

| battle_id | p1_forms | p2_forms | win | |
|-----------|------------|-------------------------------|-------------------------------|----|
| 0 | 2099996083 | [1, 13, 483, 2, 0, 0, 0, 0] | [10, 18, 483, 2, 13, 0, 0, 0] | -1 |
| 1 | 2099997604 | [3, 41, 483, 2, 13, 0, 0, 0] | [10, 18, 483, 5, 19, 0, 0, 0] | -1 |
| 2 | 2100002072 | [4, 18, 483, 5, 19, 0, 0, 0] | [47, 52, 483, 2, 13, 0, 0, 0] | 1 |
| 3 | 2100002639 | [6, 17, 483, 140, 7, 0, 0, 0] | [19, 5, 483, 10, 22, 0, 0, 0] | 1 |
| 4 | 2100002744 | [8, 13, 483, 9, 18, 0, 0, 0] | [13, 41, 483, 18, 0, 0, 0, 0] | -1 |

Modeling: Turn-based Model Selection

Data Input:

- vectorized status of turns (flattened to 2-dimension, rescaled to fed models)
- all features counted

| | output | weight | weather0 | weather1 | weather2 | weather3 | field0 |
|-------|--------|---------|----------|----------|----------|----------|--------|
| 298 | 2 | 2.53875 | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 93827 | 0 | 2.35750 | 0.0 | 5.0 | 0.0 | 0.0 | 4.0 |
| 44595 | 2 | 1.72375 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 78846 | 2 | 1.67875 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 57799 | 0 | 2.09500 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

5 rows × 494 columns ← Polyfeatures may run out of memory

Model:

- input a turn
- output prediction of the final winner

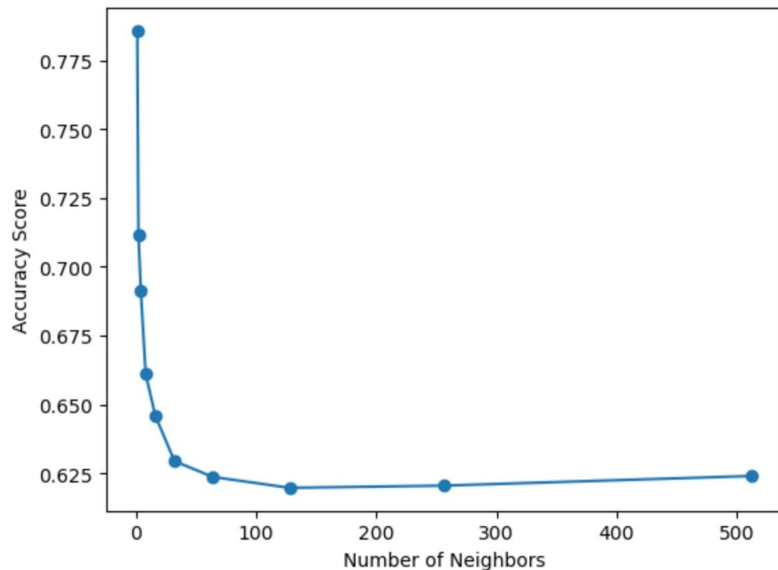
Extra Trees Classifier:

has highest accuracy 78.0%

```
lda takes 2.580268144607544 seconds to fit, with accuracy 0.6723319170207448
log takes 7.023941516876221 seconds to fit, with accuracy 0.6567358160459885
svc takes 73.52816128730774 seconds to fit, with accuracy 0.6715321169707573
qda takes 2.588124990463257 seconds to fit, with accuracy 0.6161959510122469
gnb takes 0.9691414833068848 seconds to fit, with accuracy 0.5795051237190703
dtc takes 12.424785375595093 seconds to fit, with accuracy 0.6751812046988253
rfc takes 53.95642709732056 seconds to fit, with accuracy 0.7663084228942765
etc takes 90.58779811859131 seconds to fit, with accuracy 0.7800549862534366
xbg takes 7.621344804763794 seconds to fit, with accuracy 0.7150212446888278
```

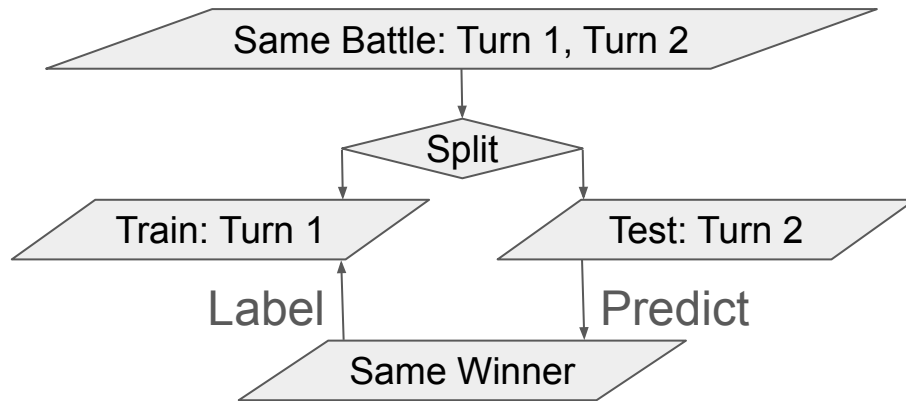
Modeling: Turn-based KNN

Issue: small k but large accuracy



(Recall: most battles last 5 turns)

Fact: underlying distribution
(stratification not preferred)



Possible Fix: split battles instead of turns



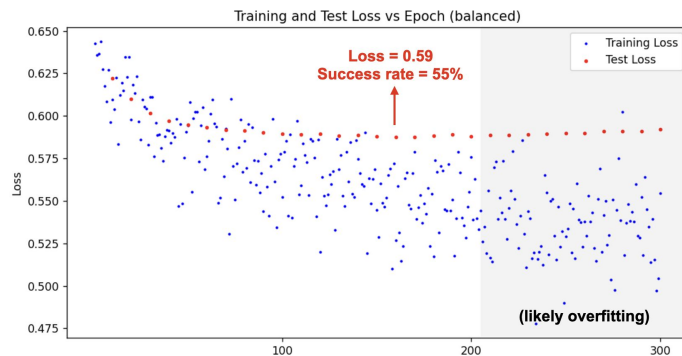
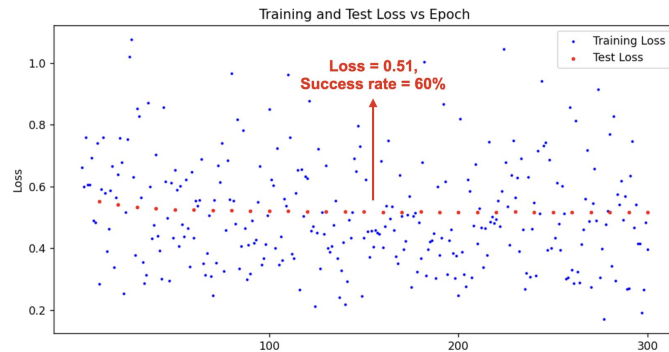
Modeling: Neural Network

Model 1: 'toy model'

- Accuracy 59-60% in predicting the correct winner
- may be biased based on the remaining total HP and number of pokémon remaining

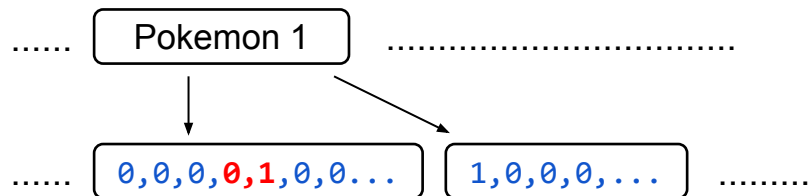
Model 2: weighted NN

- reduced influence of the total HP of pokémon remaining by decreasing the weight
- Accuracy dropped to 55% to give a correct winner prediction



Applications: analyzing team dynamics

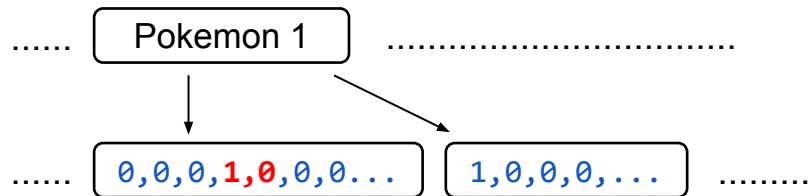
Input Vector: change pokémon



Rillaboom



Output 1



Flutter Mane



Output 2



Applications: analyzing team dynamics

Pokémon that play similar roles:



- medium speed
- moderate durability
- powerful special attacks which hits both opponent pokemons

Pokémon that play different roles:



Best Physical Attacker
under Trick Room

VS



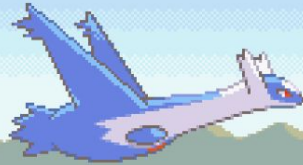
BEST support
Pokémon EVER





Pokémon

Bottle AI



Credits

We would like to thank:

Zach Hafen-Saavedra

Steven Gubkin

Alec Clott

Roman Holowinsky

and

The Pokémon Community!



○ START

○ SELECT