Classify Them All

Predicting Pokémon Types with Machine Learning

Scott Ratchford

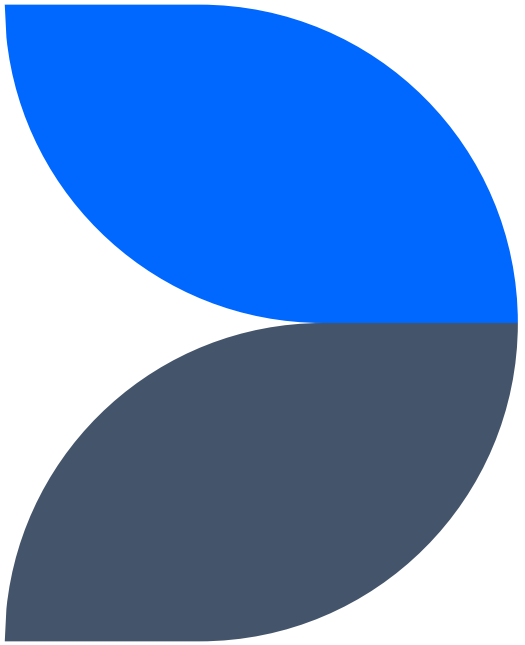
What is a Pokémon?

Pikachu

Pokémon Types

Each Pokémon species has exactly **1** or **2** types.

How can we predict Pokémon types?



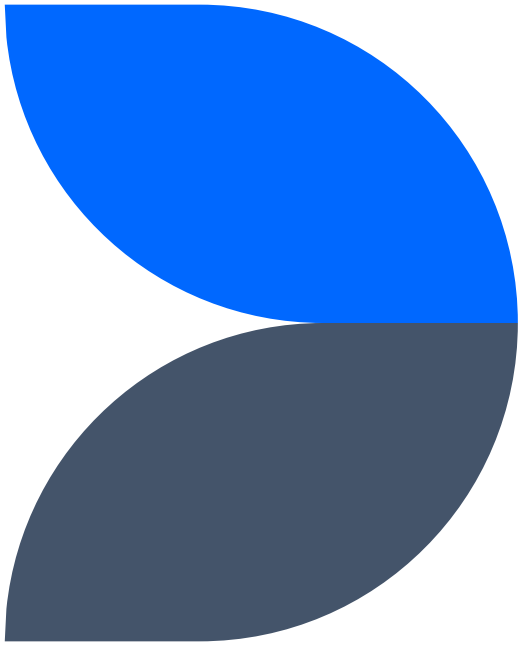
Charizard

FLYING

DRAGON

FIRE

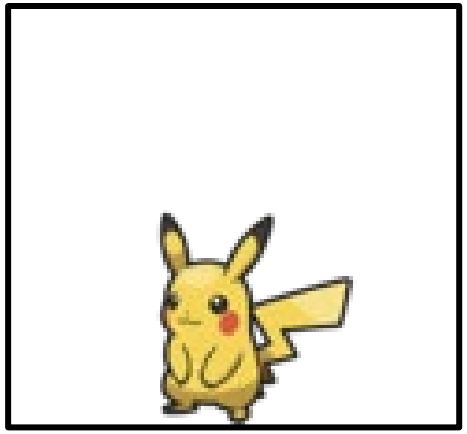
What data can we use?



Data Sets

Pokémon Images

- Sourced from Pokémon video games
- 1 image per Pokémon
- Transparent backgrounds

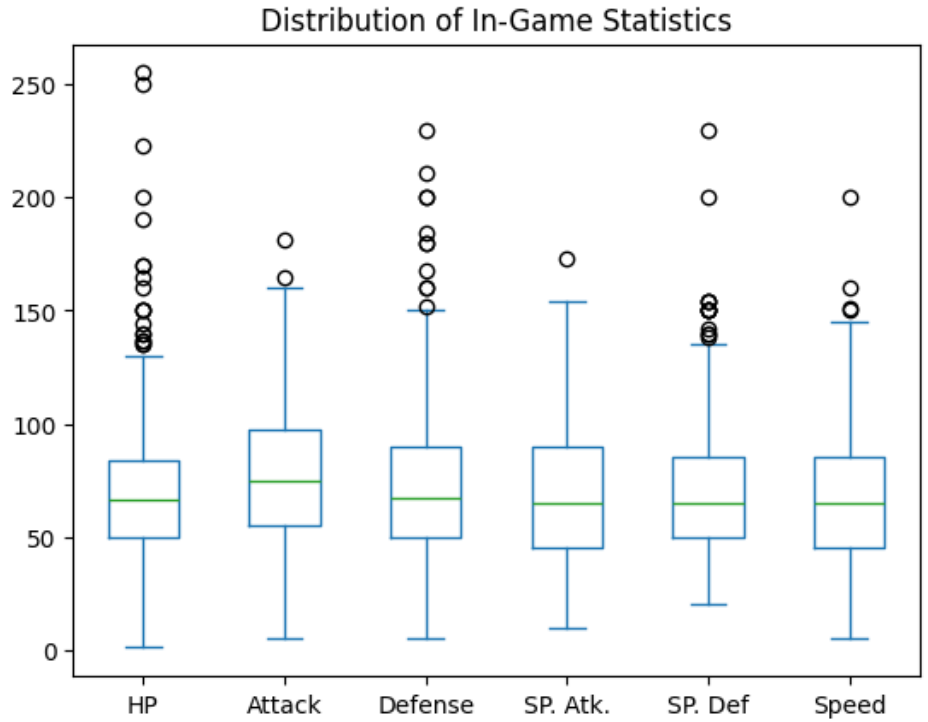


Pikachu

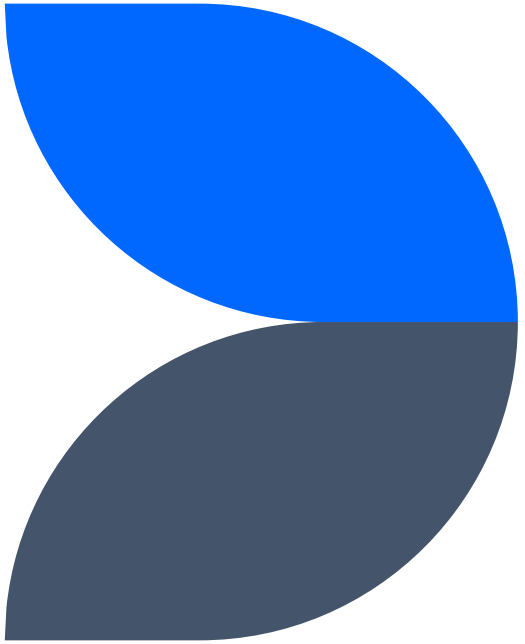
Charizard

In-Game Statistics

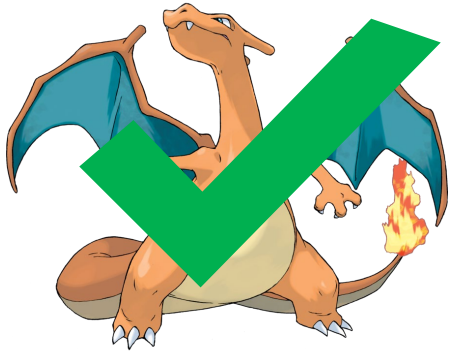
- Sourced from Pokémon video games



Data Preprocessing



1. Removed alternate forms of each Pokémon



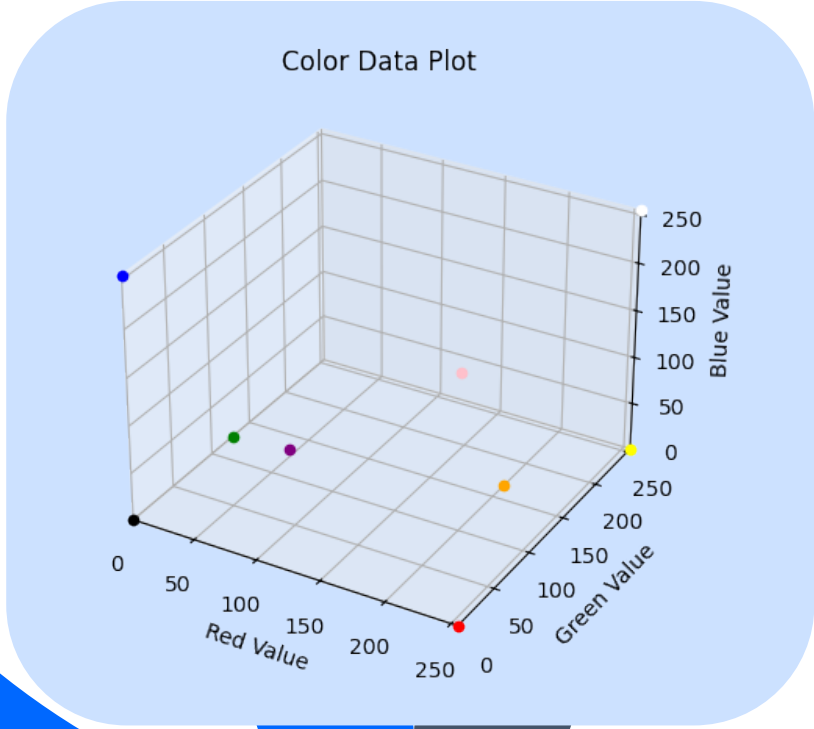
Charizard

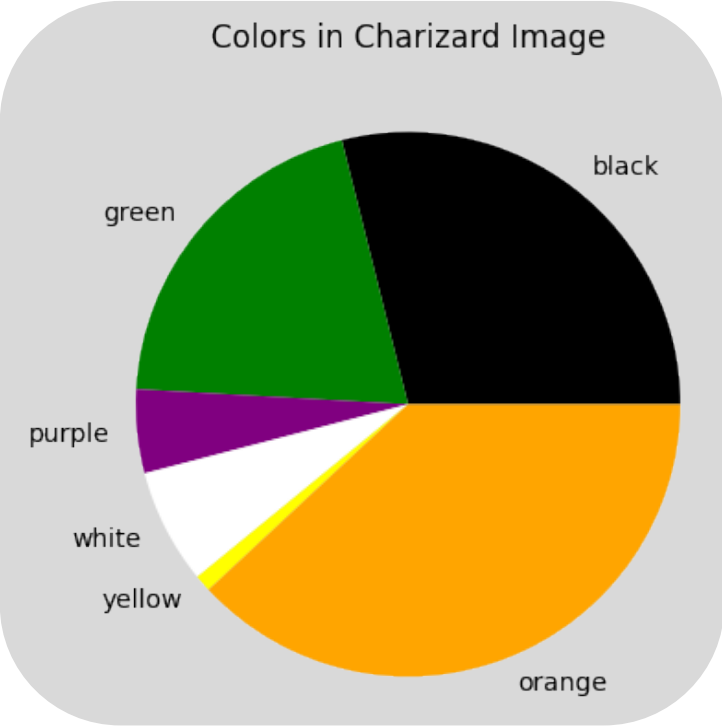
Mega Charizard X

Mega Charizard Y

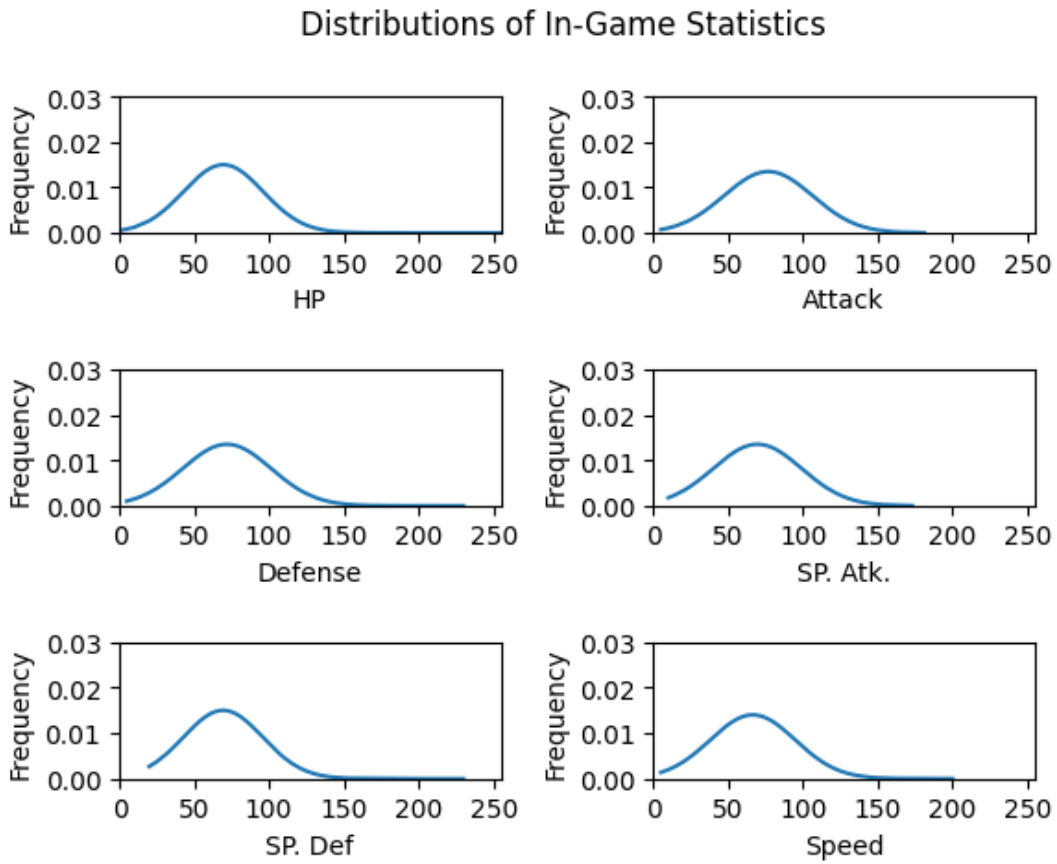
2. Quantified colors in each image

Charizard





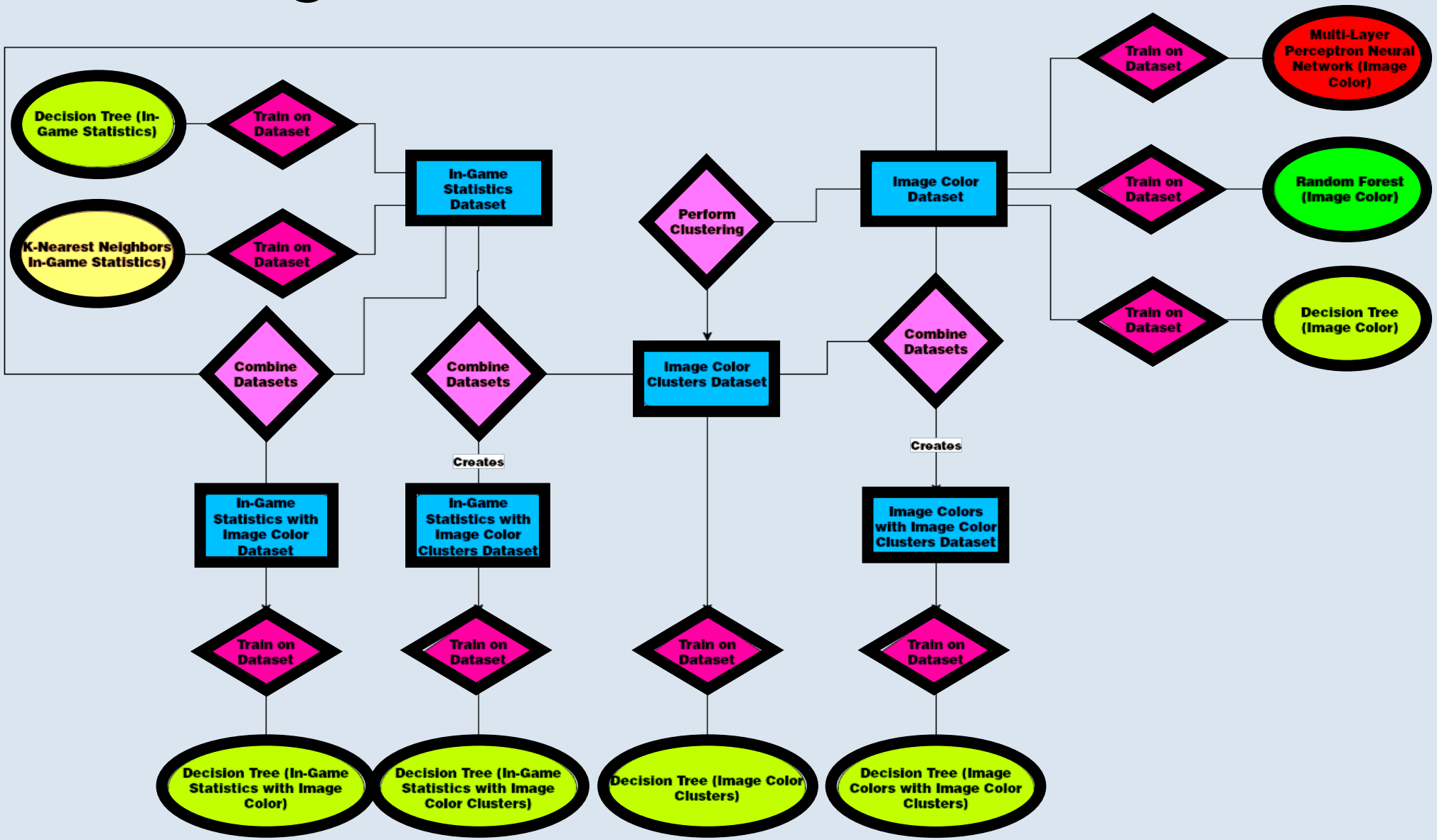
3. Scaled In-Game Statistics



4. Train-Test Split

- The same Pokémon were used in each training set and each testing set.
  - Train: 80%
  - Test: 20%

Training Models



Machine Learning Models

Data Set(s) Used	Model Type	Accuracy	
		Either Type	Both Types
In-Game Statistics	K-Nearest Neighbors	9.90%	1.04%
In-Game Statistics	Decision Tree	8.33%	1.04%
Image Colors	Decision Tree	12.50%	0.00%
Image Colors	Random Forest	7.81%	0.00%
Image Colors	Multi-layer Perceptron Neural Network	9.90%	0.00%
Image Color Clusters	Decision Tree	53.65%	8.85%
In-Game Statistics and Image Color Clusters	Decision Tree	23.44%	1.04%
Image Colors and Image Color Clusters	Decision Tree	22.92%	1.04%
In-Game Statistics and Image Colors	Decision Tree	28.13%	0.52%

Conclusions

1. In-game statistics have little to no relevance in type predictions
2. Color data are more useful when clustered into similar groups
3. Multi-label classification is significantly more difficult than single-label classification
4. Further studies could extract more data from images

Thank you

Do you have any questions?