



# Pokémon Battle Predictor

PROJECT 4- GROUP 2

Ryan Johnson, Daniel Hall, Robert Maysent | 9/20/2023

## Introduction

Pokémon, a globally recognized brand, invites you to step into a captivating universe where trainers embark on extraordinary journeys to capture and engage in epic battles with creatures nestled within iconic Poké Balls. With its rich history and enduring legacy, Pokémon has not only conquered the realm of animated series and video games but has also captured the hearts of millions worldwide.

## Data

We used one dataset which we got from Kaggle that had everything we needed, such as columns with their stats, abilities, as well as columns for each type their effective against with the numeric value of effective they are. Everything in this dataset for the most part was clean, only data cleaning that was needed was dropping unnecessary columns that weren't needed and keeping all numerical valued columns to help with our process of making our model.

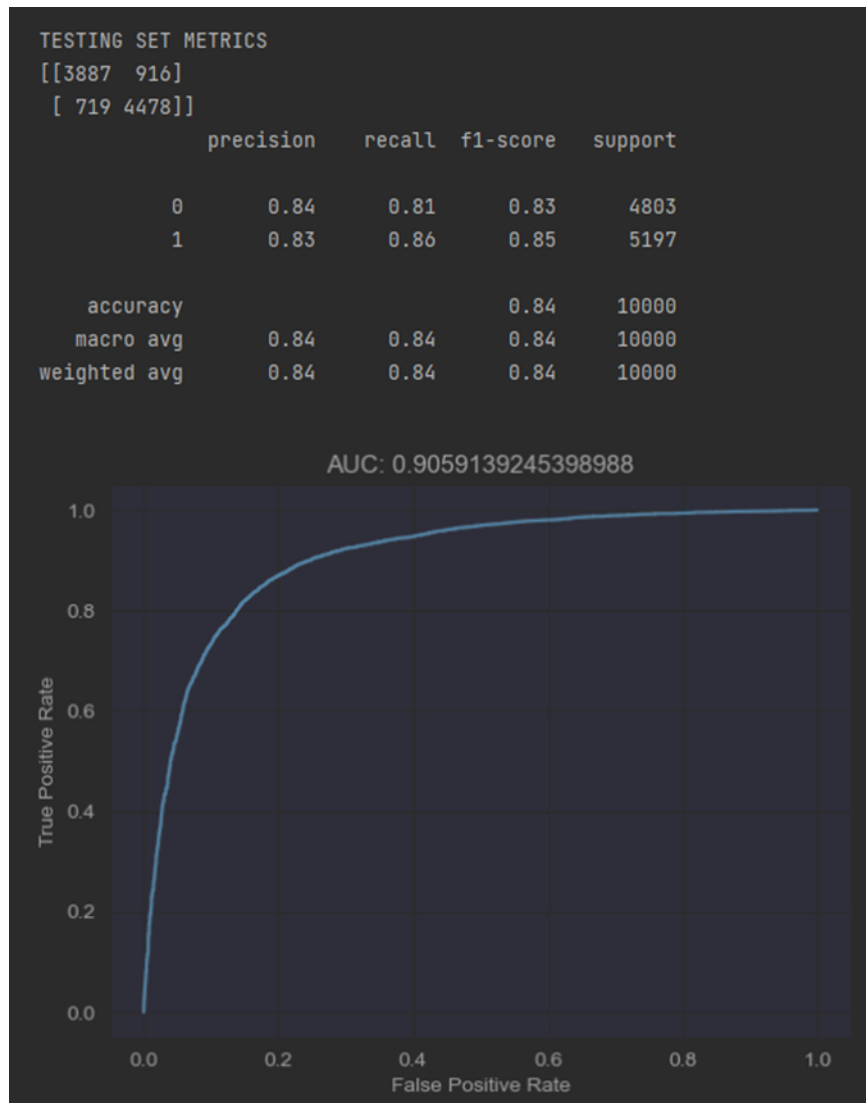
The Complete Pokémon Dataset | Kaggle

<https://www.kaggle.com/datasets/rounakbanik/pokemon>

## What Does Our Model Do?

We chose Pokémon battles for our project due to the high number of numeric fields such as attack, defense, speed, etc., additionally the effectiveness of the types of Pokémon against the others was listed in a decimal. This lent itself to creating a good machine learning model, however it still took some development. When we initially sourced our data there were also the categorical fields for types of Pokémon included which seemed at the start to be a contributing factor to who would win in a battle, so we set out to dummify that data which led to 36 new columns 18 for each of the 2 types a Pokémon can possibly have. However, upon testing the various models we found a few things that seemed like noise by checking the feature importance on our models, so we removed columns such as experience scale, steps to hatch egg of the Pokémon, the Pokémon's generation it came from, if the Pokémon was legendary, and all 36 of our dummified type columns. I believe the type's importance was low because we already had the effectiveness conveyed by decimal. Once our data was cleaned to its final state we tried a linear regression, which failed abysmally due to the fact there is no 1-1 correlation, as well as a random forest, light gradient boost machine, and XGBoost model. With the returned

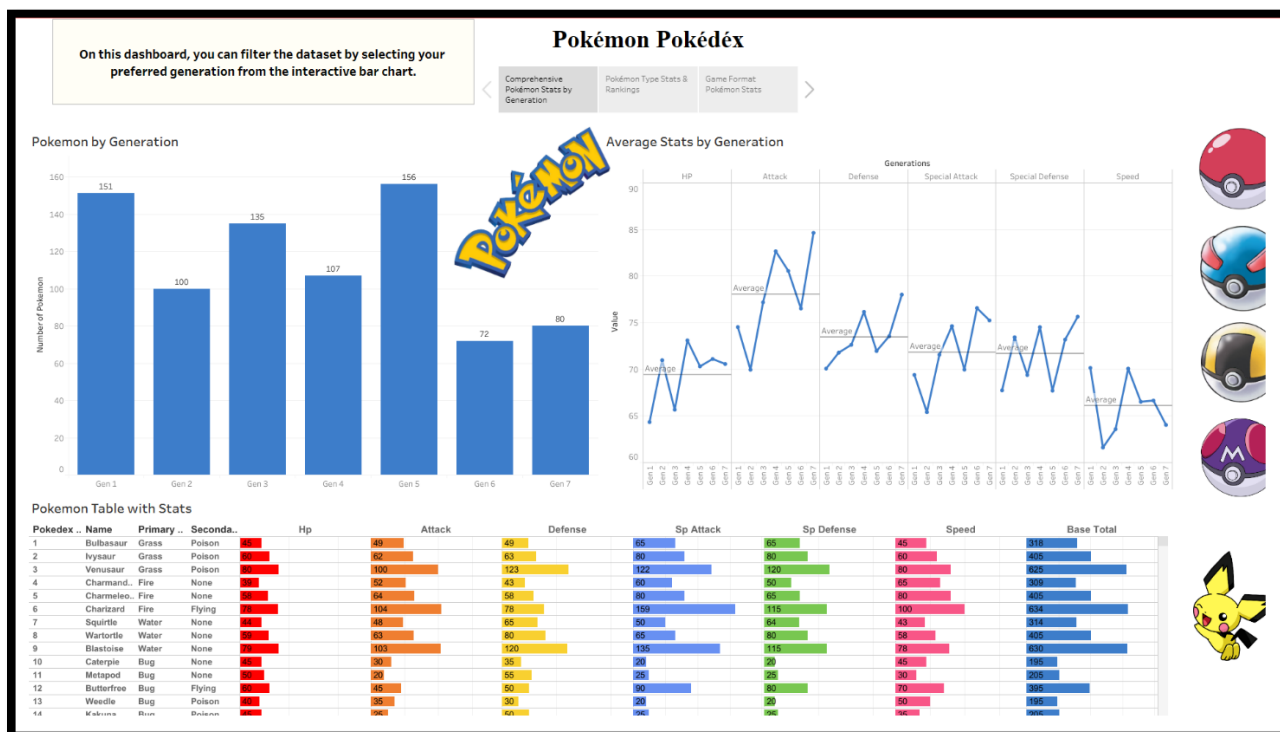
scores we landed on the XGBoost model as it had the least over-fitting and best ROC curve with a value over .9. However with our model there are some limitations that have to do with the intangibles of Pokémon battles, some Pokémon have move-sets that take advantage of being high speed to limit their opponents options and these kinds of Pokémon are not represented well with our data and will lose to most Pokémon even if they obviously shouldn't. Additionally we have an upper limit for our input of Pokémon IDs in the battle page of 801, if you input anything beyond that it breaks the model, however even with a max tag on the input element in HTML it does not stop users from typing in something like 999. This is unable to be blocked without some further work in jQuery.



## Visualizations

Our journey led us to the creation of an engaging and informative Tableau story, a vibrant addition to our website's dedicated "Pokémon Pokédex" page. This dynamic feature boasts not one, not two, but three meticulously crafted dashboards, all geared towards providing users with a wealth of knowledge about Pokémon, presented in a manner that is not only easily digestible but also suitable for individuals of all ages.

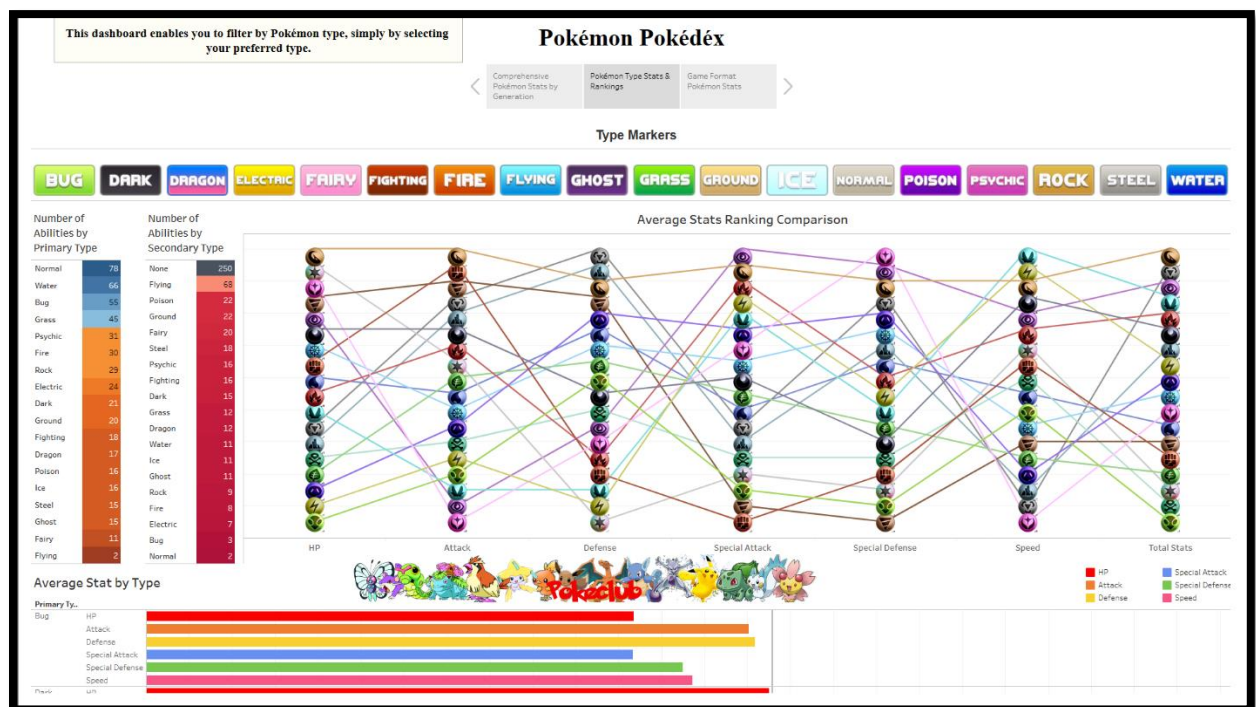
The inaugural dashboard, our statistical marvel, stands as a testament to our commitment to delivering valuable insights about the Pokémon universe. Within its digital confines, users can embark on an exploration of the Pokémon world by generation. This means not only can they discover the sheer number of Pokémon within each generation but also gain an understanding of the average statistics that define these enchanting creatures. Furthermore, a comprehensive Pokémon table graces this dashboard, offering users a treasure trove of information about their favorite creatures. What sets this dashboard apart is its versatility; users have the power to filter the Pokémon data according to their preferred generation, ensuring a customized and enriching experience that caters to Pokémon enthusiasts of all backgrounds and age groups. With this feature, we've successfully bridged the gap between the intricate world of Pokémon statistics and accessibility, making it a welcoming resource for everyone.



Moving on to our second dashboard, we ventured deeper into the fascinating realm of Pokémon abilities. This dashboard was meticulously designed to shed light on the myriad of abilities that Pokémon may possess, providing valuable insights into their prevalence and characteristics.

At its core, this dashboard delivered a comprehensive breakdown of the number of abilities based on both primary and secondary types. By categorizing abilities in this manner, users gained a holistic understanding of the distribution of these traits across the Pokémon world. Moreover, our dashboard went a step further by presenting average statistics based on the specific ability types. This allowed users to discern how certain abilities might impact a Pokémon's overall stats, providing a deeper appreciation for the intricacies of these creatures.

But the story didn't end there. Our commitment to delivering valuable information knew no bounds, and so, we also included an analysis of average statistics based on Pokémon types. This multifaceted approach ensured that users not only grasped the significance of abilities but also how they interacted with the broader framework of Pokémon typings. In doing so, our second dashboard transformed into a treasure trove of knowledge, revealing the nuanced relationships between abilities, types, and stats in the Pokémon world.

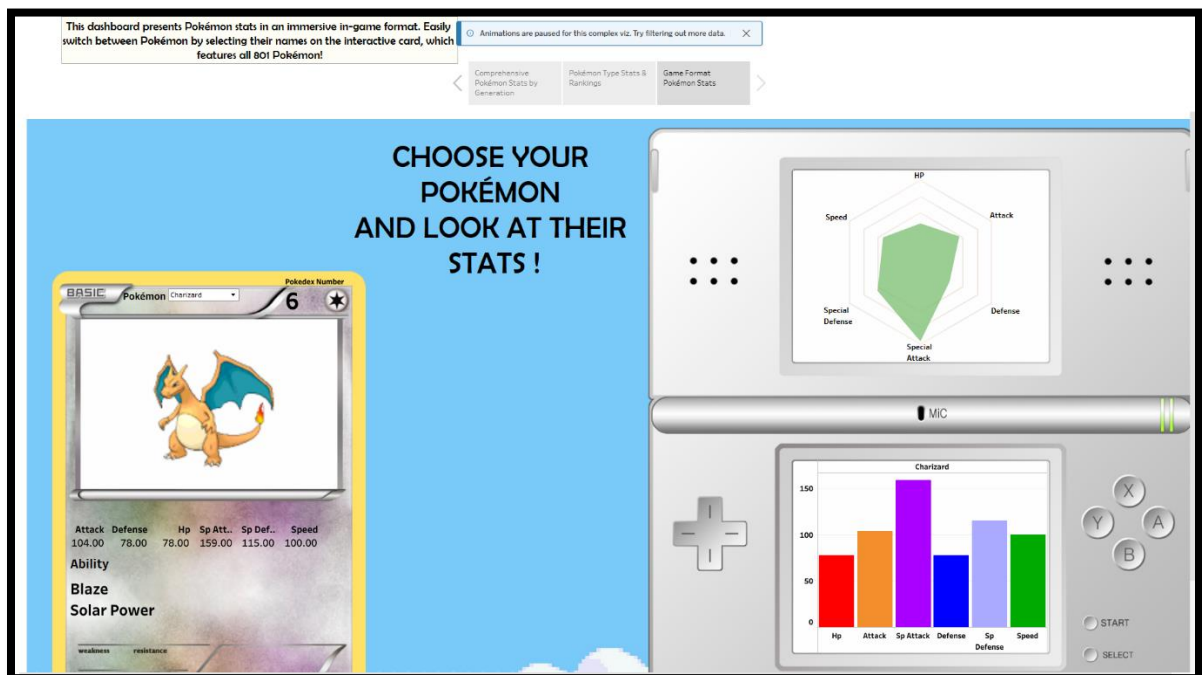


Our third and final dashboard was a true testament to our dedication to recreating the immersive experience of Pokémon games in a digital format. Designed to resonate with fans of all ages, this dashboard seamlessly combined the nostalgia of collecting Pokémon cards with the interactive appeal of the Nintendo DS games.

At its heart, this dashboard brought the magic of Pokémon cards to life by presenting each Pokémon in a format mirroring the cherished trading cards. Users could explore intricately detailed Pokémon cards, complete with comprehensive statistics and abilities, evoking the same excitement as holding physical cards in their hands. It was a digital homage to the timeless tradition of collecting and trading Pokémon cards.

To amplify the gaming experience, we introduced a Nintendo DS emulator that truly immersed users in the Pokémon world. The upper screen of the emulator featured a radar chart, faithfully replicating the in-game presentation of Pokémon stats. This visual representation allowed users to assess and compare a Pokémon's strengths and weaknesses, just as they would in their favorite video games.

On the lower screen, a dynamic bar chart provided a detailed breakdown of individual Pokémon stats, offering a comprehensive view of their capabilities. This feature was not just informative but also engaging, enhancing users' understanding of each Pokémon's unique attributes.



## Limitations

One such limitation revolved around the representation of various Pokémon types within our dataset, which occasionally led to unexpected and fascinating outcomes in our battle predictions.

A striking illustration of this phenomenon was the surprising victory of Magikarp, a Pokémon typically known for its vulnerability and unassuming nature, over the formidable Rayquaza, a legendary dragon with awe-inspiring power. This apparent anomaly served as a vivid reminder of the intricate nuances that govern Pokémon battles, where type advantages, strategic maneuvers, and situational factors can occasionally defy conventional expectations.

In response to these intriguing challenges, we took a proactive step to enhance the interpretability of our predictive model. To provide users with a more comprehensive view of the battles, we introduced a percentage chance indicator. This valuable addition quantifies the model's assessment of the probability that the predicted outcome aligns with the actual battle result.

## Summary

Our overarching mission has always been to empower users with invaluable insights into the vast and intricate world of Pokémon, leveraging the power of data utilization as our guiding star. As we bring our journey to a close, it's worth emphasizing that there are myriad pathways to achieving success within the Pokémon universe, whether you prefer to rely on the advanced analytics of machine learning models, engage in thorough data research, or simply trust your instincts when it comes to winning battles and making strategic decisions.

Whether your quest is to assemble an unbeatable team capable of triumphing over the formidable Elite Four, uncover the most favorable type matchups across various generations, or simply immerse yourself in the wonder of your favorite Pokémon, we sincerely hope that you consider our resource as an invaluable addition to your toolkit.

As you embark on your Pokémon adventures, armed with newfound knowledge and tools, remember that the world of Pokémon is one of endless discovery and boundless

excitement. So, with this newfound insight and resource at your disposal, it's time to step out into the wild, explore diverse regions, and embark on the thrilling journey of "catching 'em all!" The Pokémon world is waiting, and we're here to support you every step of the way.

## Works Cited

The Complete Pokémon Dataset | Kaggle

<https://www.kaggle.com/datasets/rounakbanik/pokemon>

Pokémon Encyclopedia

<https://bulbapedia.bulbagarden.net/wiki/Stat>