

**UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG**  
**DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO**  
**CURSO DE SISTEMAS DE INFORMAÇÃO**

**GILLIARD GABRIEL RODRIGUES**

**APRENDIZADO DE MÁQUINA APLICADO À CLASSIFICAÇÃO DE TIPOS DE  
POKÉMONS**

**Belo Horizonte**  
**2022**

GILLIARD GABRIEL RODRIGUES

**APRENDIZADO DE MÁQUINA APLICADO À CLASSIFICAÇÃO DE TIPOS DE  
POKÉMONS**

Trabalho apresentado para a Disciplina  
Mineração de Dados, pelo Curso de  
Sistemas de Informação da Universidade  
Federal de Minas Gerais, ministrada pelo  
Prof. Wagner Meira Júnior.

Belo Horizonte

2022

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>3</b>
<b>2 MOTIVAÇÃO .....</b>	<b>4</b>
<b>3 OBJETIVO .....</b>	<b>5</b>
<b>3.1 Geral .....</b>	<b>5</b>
<b>3.2 Específicos .....</b>	<b>5</b>
<b>4 METODOLOGIA .....</b>	<b>6</b>
<b>5 DESENVOLVIMENTO .....</b>	<b>7</b>
<b>5.1 Entendimento dos dados.....</b>	<b>7</b>
<b>5.2 Análise exploratória dos dados .....</b>	<b>7</b>
<b>5.3 Preparação dos dados .....</b>	<b>10</b>
<b>5.4 Modelagem.....</b>	<b>11</b>
<b>6 RESULTADOS/CONCLUSÕES .....</b>	<b>12</b>

## 1 INTRODUÇÃO

Pokémon é uma franquia muito popular de jogos desenvolvidos pela Game Freak e publicados pela Nintendo desde 1996 – quando ocorreu o lançamento de Pokémon Red & Blue (GAME FREAK, 2022). Desde então o número de pokémons só aumentou com o decorrer dos lançamentos e na 8ª geração dos jogos (*Pokémon Legends Arceus*) chegou a mais de 1000 pokémons e 18 tipos possíveis.

Uma questão interessante que existe desde os primórdios da franquia é se o tipo de um pokémon possui relação com seus demais atributos ou apenas influencia em sua capacidade de aprender determinados ataques.

Nesse contexto, o presente trabalho irá apresentar uma análise exploratória em cima dos pokémons da primeira à oitava geração dos jogos, levantando e explorando perguntas, descobrindo curiosidades sobre os monstros e finalizando com a aplicação de aprendizado de máquina supervisionado para classificar o tipo primário de um pokémon baseado nas demais informações disponíveis sobre ele.

## 2 MOTIVAÇÃO

A motivação por trás deste trabalho está pautada na curiosidade e vontade de responder perguntas como:

- Existem quantos pokémons de cada tipo?
- Existem quantos lendários de cada tipo? E míticos?
- Qual a proporção de pokémons que possuem apenas um tipo e aqueles que possuem dois?
- Quais as combinações de tipos mais comuns?
- Qual tipo possui, em média, os pokémons mais fortes? E os mais fracos?
- Quais são os 10 pokémons mais resistentes a outros tipos? E os 10 menos resistentes?
- Quais os 10 pokémons mais fortes em termos de soma dos atributos base?
- Qual geração possui, em média, os pokémons mais fortes? E os mais fracos?
- Quais os 5 pokémons mais fortes em cada atributo?
- É possível classificar o tipo de um pokémon baseado nas demais características disponíveis sobre ele?

A análise exploratória dos dados será guiada por essas perguntas e espera-se minerar curiosidades interessantes.

### 3 OBJETIVO

#### 3.1 Geral

O objetivo aqui é fazer uma análise exploratória sobre os dados de todos os pokémons da primeira à oitava geração, respondendo às perguntas citadas anteriormente e construindo um modelo que preveja de forma satisfatória o tipo de um pokémon a partir de suas demais características.

#### 3.2 Específicos

As tarefas específicas podem ser divididas em:

- Extração dos dados, que virão de uma base em formato .csv, retirada do *Kaggle*;
- Limpeza dos dados, pois a base pode conter informações faltantes ou dados desnecessários para o nosso objetivo;
- Análise exploratória dos dados, a fim de obter *insights* sobre as diversas informações disponíveis sobre os pokémons e decidir quais utilizar no modelo.
- Preparação dos dados, de forma a deixá-los da melhor forma para o modelo de classificação.
- Aplicação do algoritmo de classificação *Decision Trees*;
- Análise dos resultados, partindo de métricas de avaliação e da visualização da árvore de decisões gerada.

## 4 METODOLOGIA

A metodologia foi inspirada no CRISP-DM, ou seja, dividida em: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação.

Uma parte da aplicação dessa metodologia pode ser vista através do [link](#) para o *Google Colab*, que apresenta o entendimento dos dados, a preparação dos dados, a modelagem e a avaliação, com os *scripts* já rodados e os resultados já disponíveis.

## 5 DESENVOLVIMENTO

### 5.1 Entendimento dos dados

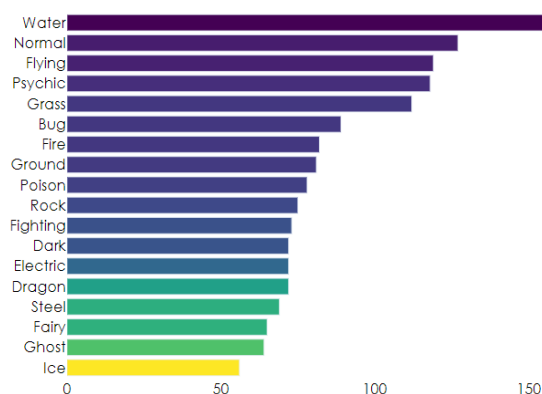
Como recurso disponível, temos uma base de dados, retirada do Kaggle, que pode ser encontrada no seguinte [link](#) e está em formato csv. Ela contém as seguintes informações sobre cada pokémons: nº na Pokédex, nome, geração, classificação, habilidades, peso e altura, tipo primário e secundário, pontos de ataque, defesa, velocidade, ataque especial e defesa especial, efetividade de outros tipos contra ele, taxa de captura, se é lendário, mítico ou mega evolução, número de passos necessários para chocar um ovo da espécie e valor base de felicidade.

### 5.2 Análise exploratória dos dados

No que diz respeito às questões levantadas anteriormente, descobriu-se que a proporção de pokémons de um ou dois tipos é bem equilibrada (47,1% x 52,9%) e também que água, normal, voador, psíquico e grama são os tipos mais comuns; psíquico é o tipo com maior número de lendários e míticos; não existe pokémon lendário do tipo inseto, mas por outro lado existem vários lendários do tipo dragão ou voador.

Figura 1 – Existem quantos pokémons de cada tipo?

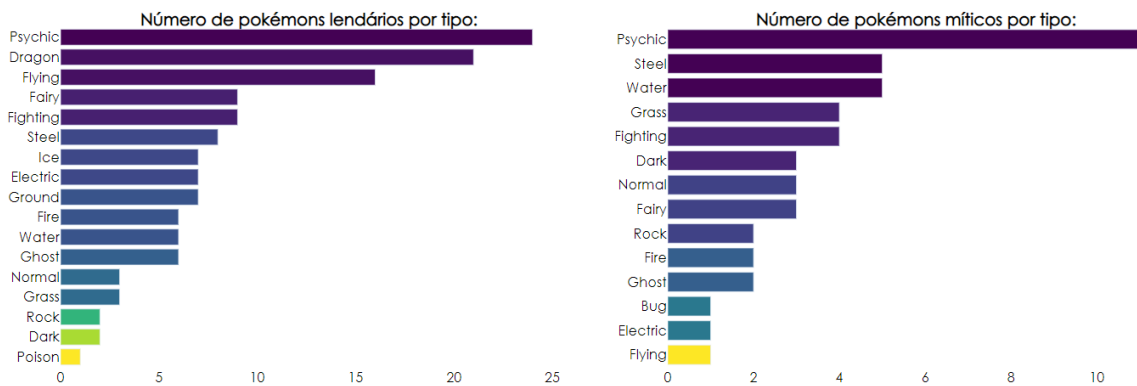
Número de pokémons por tipo:



Fonte: Google Colab.



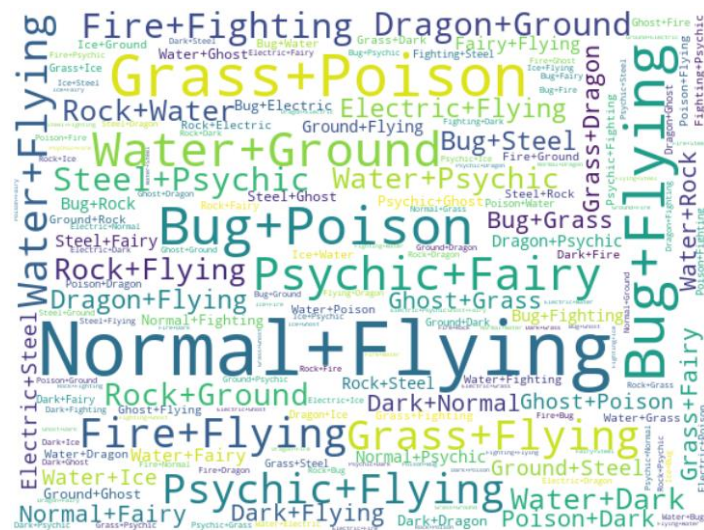
Figura 2 – Existem quantos lendários por tipo? E míticos?



Fonte: Google Colab.

Através da nuvem de palavras a seguir é possível ver uma curiosidade interessante: a combinação de tipos mais comum é normal + voador. Outras combinações bem populares são grama + venenoso e inseto + voador.

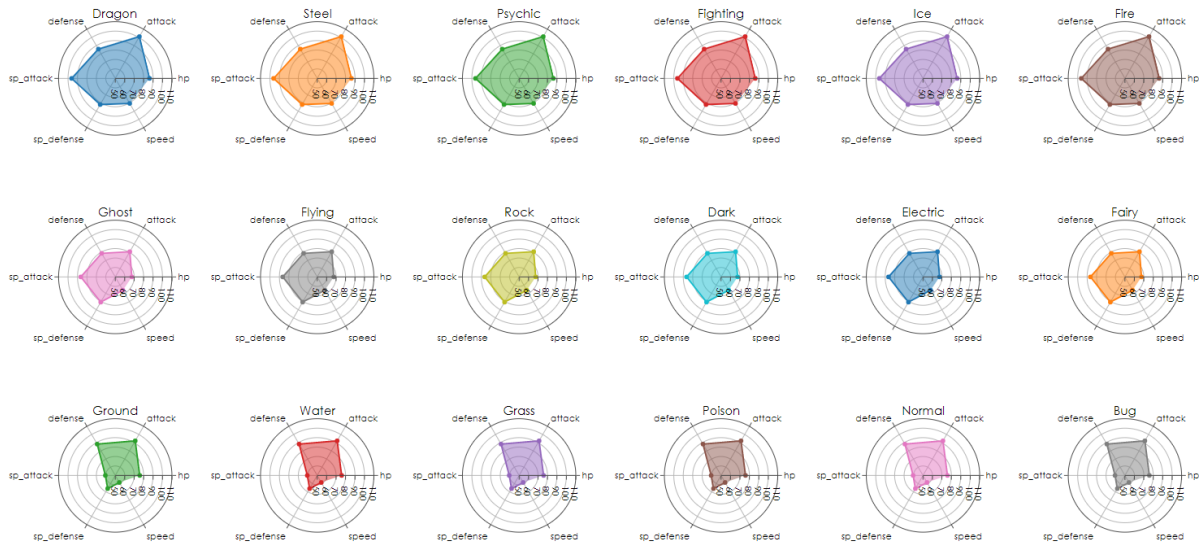
Figura 3 – Quais as combinações de tipos mais comuns?



Fonte: Google Colab.

Através da figura abaixo é possível ter uma noção dos valores médios de cada atributo base para os diferentes tipos de Pokémon. Os gráficos estão ordenados do mais forte, que se revelou ser o tipo dragão, para o mais fraco (inseto).

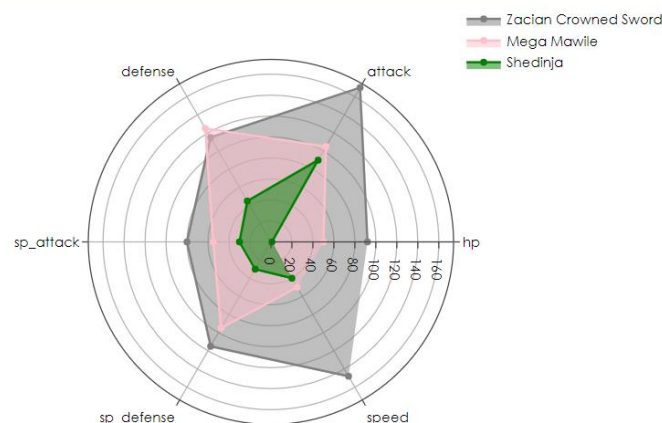
Figura 4 – Qual tipo possui, em média, os pokémons mais fortes? E os mais fracos?



Fonte: Google Colab.

Quanto aos dez pokémons mais/menos resistentes a tipos, o ranking detalhado pode ser encontrado no link do Google Colab. Aqui será apresentado apenas uma curiosidade sobre os 3 pokémons mais resistentes: como podemos ver no gráfico abaixo, aquele com mais resistência a tipos é o Shedinja, seguido pelo Zacian na sua forma 'Crowned Sword' e o Mega Mawile. No caso do Shedinja, vale lembrar que se trata do pokémon com o menor HP de todos, mas em contrapartida conta com uma habilidade (Wonder Guard) que garante que ele só receba dano através de golpes super-efetivos, Entry Hazzards (Spikes, Stealth Rock), efeitos climáticos (Sandstorm, Hail) ou status (Burn, Poison, Badly Poison).

Figura 5 – Comparativo dos 3 pokémons mais resistentes a outros tipos



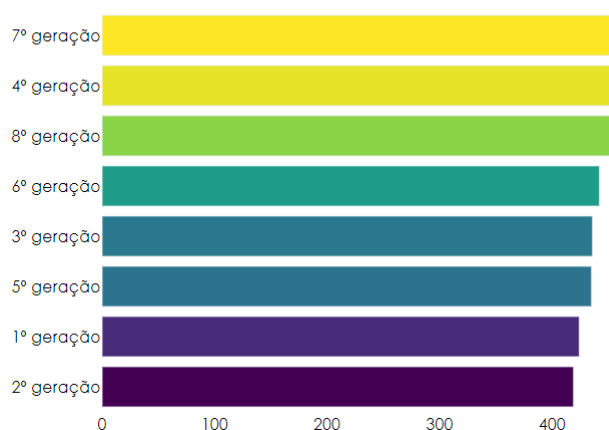
Fonte: Google Colab.

No que diz respeito à soma dos atributos base, temos que os três pokémons mais fortes são três mega evoluções: Mega Mewtwo X, Mega Rayquaza e Mega Mewtwo Y, que de fato possuem uma soma de atributos muito elevada (780). Os demais podem ser vistos no *notebook* que contém as implementações.

Através do gráfico a seguir, é possível ver que a geração que possui os pokémons mais fortes em termos de atributos base é a 7ª, enquanto que a detentora dos mais fracos é a 2ª. Outro fato que podemos notar é que as gerações não possuem médias de força muito distantes umas das outras, já que em toda geração é possível encontrar tanto pokémons fracos quanto fortes.

Figura 6 – Qual tipo possui, em média, os pokémons mais fortes? E os mais fracos?

Ranking de força por geração (em termos de atributos base):



Fonte: Google Colab.

Mais detalhes sobre as diversas questões levantadas podem ser encontrados no Google Colab.

### 5.3 Preparação dos dados

A fim de preparar os dados para aplicação do modelo, os dados desnecessários foram removidos, os dados categóricos a serem utilizados foram convertidos para tipos numéricos e os valores nulos, preenchidos.

## 5.4 Modelagem

Uma matriz de correlação foi utilizada para analisar o comportamento das diversas características em relação aos tipos possíveis de pokémons e as seguintes *features* foram escolhidas: os atributos base, as efetividades de cada tipo, se é lendário ou não, o número de passos necessários para chocar um ovo da espécie e os tipos secundários mais frequentes.

A técnica escolhida foi a de Árvores de Decisão, pois lida bem com muitas classes, possui uma alta explicabilidade e a lógica por trás é conveniente para algumas *features* escolhidas. As respostas às questões de interesse e os resultados da aplicação do modelo podem ser encontrados na próxima seção.

## 6 RESULTADOS/CONCLUSÕES

A configuração do modelo que melhor performou dentre as possibilidades testadas pelo *GridSearchCV* foi com altura máxima igual a 13, utilizando o índice de gini como critério de impureza, obtendo uma acurácia de  $\approx 98,43\%$  com os dados de treino. Para os dados de teste, a acurácia do modelo foi de  $\approx 91,79\%$ , o que já é bem satisfatório!

As demais métricas podem ser vistas na figura abaixo.

Figura 7 – Relatório de classificação

	precision	recall	f1-score	support
Bug	0.75	1.00	0.86	12
Dark	1.00	0.57	0.73	14
Dragon	0.90	1.00	0.95	9
Electric	1.00	0.90	0.95	10
Fairy	0.86	1.00	0.92	6
Fighting	1.00	0.89	0.94	9
Fire	0.85	0.92	0.88	12
Flying	1.00	0.33	0.50	3
Ghost	1.00	0.83	0.91	6
Grass	1.00	1.00	1.00	20
Ground	0.73	1.00	0.85	11
Ice	0.75	0.75	0.75	8
Normal	1.00	1.00	1.00	23
Poison	0.82	1.00	0.90	9
Psychic	1.00	1.00	1.00	16
Rock	0.89	1.00	0.94	8
Steel	1.00	0.62	0.77	8
Water	1.00	1.00	1.00	23
accuracy			0.92	207
macro avg	0.92	0.88	0.88	207
weighted avg	0.93	0.92	0.91	207

Fonte: Google Colab.

Em geral, todos os tipos primários tiveram precisão, revocação e f1-score elevados, conseguindo até prever corretamente 100% dos casos para alguns tipos, fazendo com que o modelo tenha alcançado um resultado interessante.

Além das métricas acima, para fins de explicabilidade do modelo, também foi gerada uma imagem *png* contendo a árvore de decisões utilizada por ele e pode ser visto pelo seguinte [link](#). Ao dar *zoom* nela, é possível visualizar detalhadamente quais foram as decisões tomadas.