

Previsão da performance competitiva de um Pokémon baseada em suas características

Tiago Barradas
tiago.figueiredo@fgv.br

26 de junho de 2023

Conteúdo

1	Introdução	2
1.1	Problema	2
1.2	Dados	2
2	Métodos	3
2.1	Pré-processamento	3
2.2	Ajuste e Avaliação	4
3	Resultados	6
3.1	Análise Exploratória de Dados	6
3.2	Ajuste do modelo escolhido	8
4	Conclusão	10
4.1	Limitações	10
4.2	Direções futuras	10
4.3	O que aprendemos	10

1 Introdução

Pokémon é uma franquia que se passa em um mundo fantasioso, onde humanos podem se tornar treinadores de criaturas fantásticas chamadas de Pokémon, uma abreviação de *Pocket Monsters*. Nesse mundo, um dos principais esportes é a "batalha Pokémon" onde dois ou mais treinadores escolhem alguns de seus Pokémon para lutarem entre si, até que todas as criaturas de seus oponentes sejam nocauteadas.

Essa franquia multimídia possui os maiores ganhos financeiros do mercado, com ganhos estimados de mais de 90 bilhões de dólares acumulados desde sua criação em 1995, ultrapassando nomes famosos como Mickey Mouse, Star Wars, e Marvel. Apesar de ter se iniciado como uma série de jogos, sua influência se expandiu para diversas áreas, desde desenhos animados, jogos de carta colecionáveis, e livros de mangá, até inúmeras formas de mercadorias como pelúcias e bonecos.

A franquia é fortemente centrada em seus jogos, com os conteúdos presentes em seus outros ramos sendo baseados nas criaturas e personagens que se originaram neles. Esses jogos são popularmente divididos em "gerações", sendo a atual geração a nona, que foi começada pelos jogos "Pokémon Scarlet" e "Pokémon Violet" em novembro de 2022. Cada jogo que inicia uma "geração" nova contém, em média, mais de 100 criaturas inéditas que expandem o ecossistema existente de Pokémon utilizáveis por treinadores, cada um deles sendo únicos em suas características e atributos de combate.

1.1 Problema

A *Pokémon Company*, empresa dona da franquia, organiza anualmente diversos torneios competitivos de batalhas entre treinadores. O maior deles, *Pokémon World Championships*, ocorre uma vez ao ano, e oferece um total de 50 mil dólares em prêmios para seus 16 competidores. Nesses torneios, cada treinador monta um time composto por 6 Pokémon distintos, e em cada batalha eles escolhem 4 dentre esses 6 para serem usados no determinado combate. Nesse contexto, é importante que os competidores escolhem Pokémon fortes, com atributos que dêem o máximo de poder para a sua equipe.

Com o passar do tempo, a comunidade de jogadores ganha um bom entendimento de quais Pokémon são os mais fortes, mas como dezenas de novas criaturas são criadas em cada "geração", tentar avaliar a performance de todas elas de uma vez é um trabalho que demanda considerável esforço. Seria interessante, então, se existisse algum tipo de modelo capaz de prever se um novo Pokémon será forte competitivamente com base apenas nas estatísticas de sua espécie.

A fim de auxiliar esses jogadores que competem em batalhas envolvendo prêmios financeiros, esse projeto se propõe a criar um modelo capaz de dizer quais Pokémon, dado uma determinada lista de seus atributos, terão um potencial forte para conquistar uma boa performance nesses campeonatos.

1.2 Dados

Os dados que serão utilizados para a realização do *fitting* do modelo são uma coleção de arquivos .csv encontrada no Kaggle contendo dados de 1010 Pokémon extraídos através de *web-scraping* do site *Smogon*, que é mantido por uma enorme comunidade de milhares de fãs da franquia. Esse site definiu diversos *tiers* de força para caracterizar cada Pokémon, e aquelas de mais interesse para nosso estudo são:

- **Uber:** A tier que contém os Pokémons mais fortes, que são banidos em todos os outros formatos por causarem problemas de balanceamento. Em uma batalha no formato Uber, todo e qualquer Pokémon é permitido.
- **Overused (OU):** Essa tier consiste dos Pokémons usados com frequência pela comunidade. Eles são fortes, mas balanceados, gerando um cenário competitivo saudável. Pokémons Uber são banidos em batalhas no formato OU. O site da *Smogon* detalha o procedimento estatístico usado para definir um Pokémon como OU.

Existem outras quatro principais tiers abaixo das citadas acima: Underused (UU), Rarely Used (RU), Never Used (NU), e Perfectly Useless (PU), e outras tiers ainda menos importantes presentes entre elas. Entretanto, os Pokémons pertencentes a elas não são muito relevantes em campeonatos. A filosofia que a *Smogon* utiliza para a classificação dos Pokémons é baseada no fato de que, quanto mais forte uma espécie for, mais comumente usada ela será em cenários competitivos onde ambos os jogadores querem vencer. É esperado, portanto, que o tier de um Pokémon seja diretamente relacionado com a "qualidade" de seus atributos de combate.

Cada espécie de Pokémon possui inúmeras características. Eles possuem 6 *stats* base que são levados em conta para os cálculos de dano causado e recebido durante combate: HP (Hit Points), ATK (Attack), DEF (Defense), SPA (Special Attack), SPD (Special Defense) e SPE (Speed), e possuem um ou dois tipos elementais, que definem suas fraquezas e resistências (por exemplo, um Pokémon de tipo Fogo recebe dano aumentado de ataques de tipo Água). Além disso, cada espécie pode ter de uma a três *Abilities*, que em geral são bônus que podem ser ativados em situações específicas durante a batalha. Outro aspecto bem importante é o chamado *learnset* de um Pokémon, uma lista que determina que *moves* (técnicas de ataque específicas) aquela espécie pode aprender. Eles também possuem estatísticas como peso e altura.

Como existem, atualmente, 298 *Abilities* e 933 *moves*, e o *learnset* de cada espécie contém uma seleção de, em média, dezenas de *Moves*, acredito que seja fora do escopo dessa matéria lidar com esse tipo de informação. Por isso, se torna inviável usar esses atributos com o meu conhecimento atual de modelagem estatística. Assim, meu modelo irá usar apenas as estatísticas dos *stats*, tipos, peso, e altura como variáveis preditoras para o tier de um Pokémon. Como discutido anteriormente, essa classificação possui uma forte correlação com o nível de poder e potencial competitivo da espécie, que é o fator que realmente queremos observar. De forma mais objetiva, estamos interessados em descobrir se a tier do Pokémon será ao menos UUBL, ou se será abaixo. A tier UUBL é reservada para aqueles Pokémons que não são utilizados o suficiente para se enquadrarem como OU, mas são fortes demais comparados aos Pokémons da tier diretamente abaixo, UU. Assim, diremos que o Pokémon será forte competitivamente se sua tier é ao menos tão alta quanto UUBL.

2 Métodos

2.1 Pré-processamento

Antes de começar a modelagem, é preciso lidar com o formato em que os dados estão salvos. Inicialmente, temos três arquivos .csv, 'pokemon_format.csv' (que relaciona o id do Pokémon com sua tier), 'pokemon_type.csv' (que relaciona o id do Pokémon com seus tipos), e 'pokemon.csv' (que descreve os seis *stats* de cada Pokémon). Além disso, é importante notar que os ids descritos nesses arquivos não são únicos para cada espécie de Pokémon, pois eles também levam em conta em que geração aquele Pokémon está sendo considerado. Isso acontece pois a

Smogon contém dados sobre o cenário competitivo em cada uma das gerações, e as espécies de Pokémon que foram originadas em gerações antigas podem ter sofrido mudanças conforme mais jogos foram lançados (por exemplo, a sexta geração adicionou o tipo Fada, e diversos Pokémon antigos ganharam esse tipo quando anteriormente eram apenas tipo Normal).

Como queremos um modelo capaz de realizar previsões especificamente sobre o cenário competitivo atual, podemos filtrar os dados para manter apenas aqueles que sejam referentes aos Pokémon que estão em seu estado mais atual possível. Realizando as operações necessárias e juntando os dados em uma única tabela, temos o seguinte resultado:

	formats	gen	name	hp	atk	def	spa	spd	spe	weight	height	isNonstandard	dex_number	alias	Type 1	Type 2	is_Strong
0	LC	ss	Bulbasaur	45	49	49	65	65	45	6.9	0.7	Standard	1.0	bulbasaur	Grass	Poison	0
1	NFE	ss	Ivysaur	60	62	63	80	80	60	13.0	1.0	Standard	2.0	ivysaur	Grass	Poison	0
2	UU	ss	Venusaur	80	82	83	100	100	80	100.0	2.0	Standard	3.0	venusaur	Grass	Poison	0
3	LC	sv	Charmander	39	52	43	60	50	65	8.5	0.6	Standard	4.0	charmander	Fire	None	0
4	NFE	sv	Charmeleon	58	64	58	80	65	80	19.0	1.1	Standard	5.0	charmeleon	Fire	None	0
...
975	OU	sv	Iron Valiant	74	130	90	120	60	116	35.0	1.4	Standard	1006.0	iron-valiant	Fairy	Fighting	1
976	Uber	sv	Koraidon	100	135	115	85	100	135	303.0	2.5	Standard	1007.0	koraidon	Fighting	Dragon	1
977	Uber	sv	Miraidon	100	85	100	135	115	135	240.0	3.5	Standard	1008.0	miraidon	Electric	Dragon	1
978	OU	sv	Walking Wake	99	83	91	125	83	109	280.0	3.5	Standard	1009.0	walking-wake	Water	Dragon	1
979	UUBL	sv	Iron Leaves	90	130	88	70	108	104	125.0	1.5	Standard	1010.0	iron-leaves	Grass	Psychic	1

Figura 1: Dados organizados em uma única tabela

O conjunto de dados completo, após as filtrações de pré-processamento realizadas, é composto por 980 Pokémon distintos, dos quais 155 são considerados fortes.

Antes de começar a explorar os dados, devemos dividi-los em um conjunto de testes e um conjunto de treino. Escolhi utilizar 80% dos dados para o treinamento do modelo, e os outros 20% para testar sua performance. Dentre esses 20%, separei metade deles para servirem como um conjunto de validação, que irá auxiliar no processo de escolha de modelos. O processo de análise exploratória dos 784 Pokémon do conjunto de treino será mais detalhado na seção 3.1.

2.2 Ajuste e Avaliação

Para esse trabalho, utilizarei majoritariamente ajustes por Máxima Verossimilhança, de forma frequentista. Como medidas de avaliação, usarei AIC e log-verossimilhança. Além disso, usarei a acurácia, true positive rate, true negative rate, e precisão avaliados no conjunto de validação.

Inicialmente, tentei ajustar um modelo bem simples, que leva em conta apenas os *stats* do Pokémon, com uma relação linear: $is_Strong \sim hp + atk + def + spa + spd + spe$. Todos os preditores se mostraram significativos (sem conter 0 em seus intervalos de confiança de 95%). Depois disso, fui aumentando a complexidade do modelo, adicionando outras variáveis e verificando possíveis relações polinomiais.

Após realizar diversas modelagens, montei a seguinte tabela com as diversas estatísticas de avaliação:

	aic	log_likelihood	accuracy	true_positive_rate	true_negative_rate	precision
is_Strong ~ hp + atk + Def + spa + spd + spe + Type2 + height + weight	435.458248	-191.729124	0.959184	0.769231	0.988235	0.909091
is_Strong ~ hp + atk + Def + spa + spd + spe + Type2 + heightweight	433.143153	-191.571576	0.959184	0.769231	0.988235	0.909091
is_Strong ~ hp + atk + Def + spa + spd + spe + Type2	432.319371	-192.159686	0.959184	0.769231	0.988235	0.909091
is_Strong ~ hp + atk + Def + spa + spe + Type2 + heightweight	440.322102	-196.161051	0.938776	0.615385	0.988235	0.888889
is_Strong ~ hp + atk + Def + spa + spd + spe + height*weight	430.26954	-206.13477	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + atk + Def + spa + spd + spe + height + weight	429.158332	-206.579166	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + np.power(hp, 2) + atk + Def + np.power(Def, 2) + spa + spd + spe	428.620437	-206.310218	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + atk + np.power(atk, 2) + Def + spa + spd + spe	427.988571	-206.994285	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + atk + Def + spa + spd + spe + heightweight	426.779331	-206.389665	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + atk + Def + spa + spd + spe	425.998434	-206.999217	0.938776	0.692308	0.976471	0.818182
is_Strong ~ hp + atk + Def + spa + spd + spe + is_Normal	415.777271	-200.888636	0.928571	0.692308	0.964706	0.75
is_Strong ~ hp + atk + Def + spa + spd + spe + is_Normal + is_Dragon + is_Psychic	410.089802	-196.044901	0.928571	0.615385	0.976471	0.8
is_Strong ~ hp + atk + Def + spa + spd + spe + is_Normal + is_Psychic	408.372978	-196.186489	0.928571	0.615385	0.976471	0.8
is_Strong ~ hp + atk + Def + spa + spd + spe + Type1 + Type2	426.257643	-172.128821	0.918367	0.692308	0.952941	0.692308
is_Strong ~ hp + atk + Def + spa + spd + spe + Type1	429.050342	-191.525171	0.897959	0.538462	0.952941	0.636364
is_Strong ~ hp + atk + Def + spa + spd + spe + Type1*Type2	596.761948	-99.380974	0.887755	0.307692	0.976471	0.666667
is_Strong ~ hp + atk + Def + spa + spd + spe + Type1:Type2	NaN	NaN	0.836735	0.230769	0.929412	0.333333

Figura 2: Tabela contendo cada modelo e suas estatísticas de avaliação

Durante o processo de exploração de modelos, tentei representar na modelagem algumas características que eu havia notado no processo de exploração de dados. Dentre eles, está o fato de que os tipos Normal, Psíquico e Dragão eram os que tinham a maior relação com a força do Pokémon. Por isso, testei em alguns modelos a utilização de variáveis dummy que indicam se o Pokémon pertence a um desses três tipos, ao invés de usar as variáveis Type diretamente. Esses modelos foram os que obtiveram os menores AIC, mas as estatísticas providas do conjunto de validação (acurácia, TPR, TNR e precisão) não foram tão boas.

Tentei, também, adicionar uma relação quadrática para os *stats* de HP e DEF, para tentar "penalizar" aqueles Pokémons que contêm esses atributos com valores muito altos. Tive essa ideia pois foi possível notar nos gráficos plotados durante a análise exploratória que aqueles Pokémons cujos HP e DEF eram muito altos não eram fortes. Entretanto, esses modelos não mostraram nenhum tipo de ganho quando comparado aos modelos sem relações polinomiais, obtendo um AIC maior sem melhorar as estatísticas de validação. Ao observar os coeficientes das variáveis preditores, notei que os dois coeficientes quadráticas não tinham garantia de serem significativos, além de terem valores muito próximos de zero (o coeficiente do HP quadrático foi 0.00002, enquanto o da DEF quadrática foi -0.0001).

É possível notar que aqueles modelos que contêm os seis *stats* e que incluem a variável Type2 (que indica qual é o segundo tipo do Pokémon, caso ele tenha um) são os que obtiveram as maiores log-verossimilhanças. Nos modelos que contêm tanto o Type1 quanto o Type2, e uma relação multiplicativa entre os dois, podemos ver que a performance no conjunto de validação piora consideravelmente, mas a log-verossimilhança é a melhor por uma margem considerável. É intrigante o fato de que adicionar o Type2 melhora o modelo na validação, mas adicionar o Type1 possui o efeito contrário, e não consegui pensar em alguma possível explicação para esse fenômeno.

Como minha prioridade para esse modelo é a capacidade de prever se um novo Pokémon será forte competitivamente, eu considerarei mais fortemente a performance dos modelos na estatística de Acurácia e True Positive Rate. Como desempate, usarei o valor AIC do modelo, que leva em conta tanto a verossimilhança quanto a quantidade de variáveis preditivas do modelo. Dessa forma, escolho como modelo final o seguinte:

$$is_Strong \sim hp + atk + def + spa + spd + spe + Type2$$

3 Resultados

3.1 Análise Exploratória de Dados

Antes de começar a modelagem, é interessante explorar os dados de treino para tentar identificar possíveis relações entre as variáveis preditoras. Inicialmente, queria observar uma scatter matrix dentre os oito atributos numéricos presentes nos dados: os seis *stats*, o peso, e a altura.

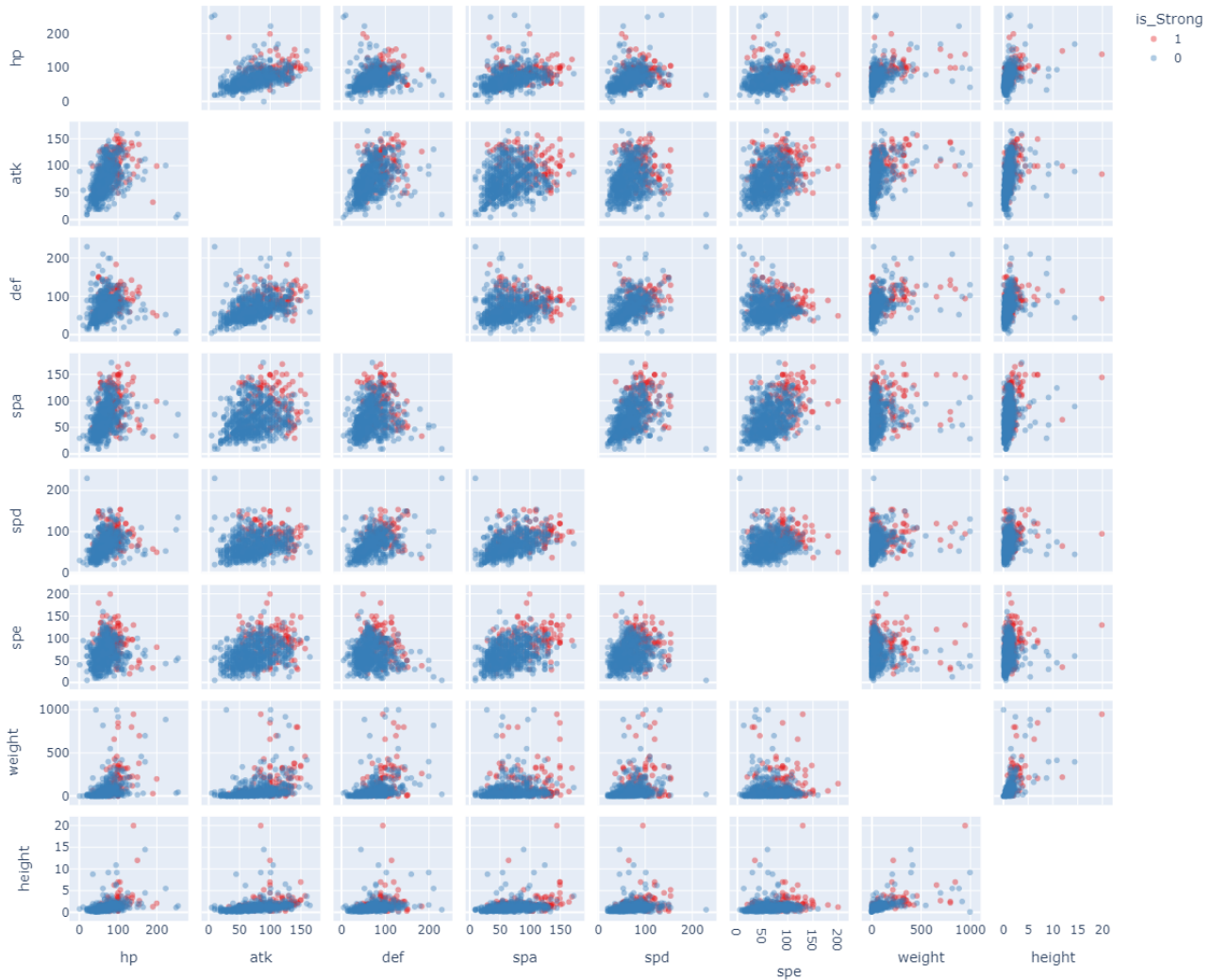


Figura 3: Scatter matrix das variáveis preditoras numéricas

É possível notar que os 6 *stats* possuem todos uma certa relação uns com os outros. Isso ocorre pois, como uma forma de balanceamento para os jogos, os *stats* de um Pokémon não costumam ser muito discrepantes entre si, salvo raras exceções. Também é claro que há uma relação entre ter atributos de combate elevados e o seu tier, os pontos com cores azuis se concentram majoritariamente nas áreas de menores valores nos gráficos.

Observando os gráficos, parece que o peso e a altura de um Pokémon não são um indicador muito bom sobre sua força. Para investigar isso melhor, irei plotar um box plot agrupado pela variável `is_Strong`:

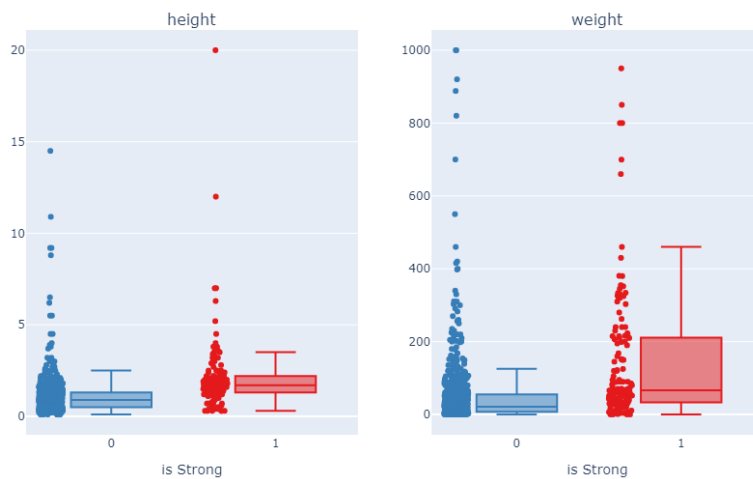


Figura 4: Scatter matrix das variáveis predictoras numéricas

Parece que essas duas variáveis possuem sim uma relação com a tier do Pokémon. Aqueles Pokémons mais fortes são, em média, mais altos e mais pesados do que aqueles que são fracos, mas parece que o peso tem uma relação muito mais forte do que a altura. Isso faz sentido, pois os Pokémons podem passar por um processo chamado "evolução", onde eles se transformam em uma outra espécie, que é mais forte do que a anterior. Esse Pokémon evoluído costuma ser maior e mais pesado do que sua forma anterior, então acredito que essa seja a origem da relação sendo exposta no gráfico acima.

Agora, quero investigar esse mesmo tipo de relação mas para cada *stat* individualmente. Além disso, também precisamos checar se há alguma relação entre os tipos de um Pokémon e sua força. Para isso, verificarei a porcentagem de Pokémons de cada tipo que são fortes, comparado ao total de Pokémons pertencentes àquele tipo.

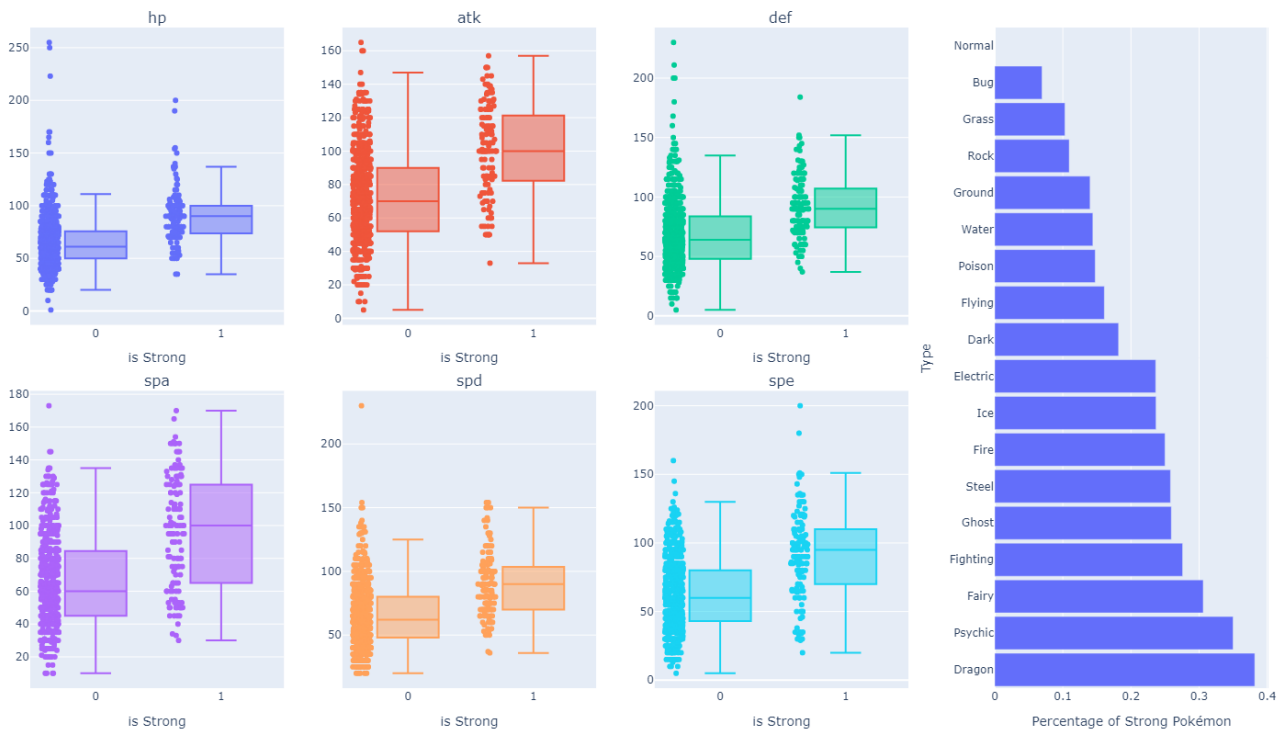


Figura 5: Exploração de como os *stats* e tipos dos Pokémons estão relacionados com sua força

Conforme o esperado, Pokémons mais fortes possuem *stats* mais elevados. Entretanto, podemos notar no caso de HP e DEF, que aqueles Pokémons com valores extremamente altos são considerados fracos. Isso indica que, esses Pokémons muito focados em atributos defensivos provavelmente são incapazes em outros aspectos importantes de combate, como em suas capacidades ofensivas.

Também é interessante notar que certos tipos estão, em média, mais presentes em Pokémons fortes do que outros. O tipo Normal, por exemplo, é composto apenas por Pokémons considerados fracos, enquanto os tipos Dragão e Psíquico são compostos por uma boa quantia de Pokémons fortes.

3.2 Ajuste do modelo escolhido

Ao analisar mais a fundo os resultados obtidos no modelo escolhido, $is_Strong \sim hp + atk + def + spa + spd + spe + Type2$, temos as seguintes informações:

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-33.6668	3.05e+04	-0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Dark]	18.9002	3.05e+04	0.001	1.000	-5.97e+04	5.97e+04
Type2[T.Dragon]	19.7499	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Electric]	18.9851	3.05e+04	0.001	1.000	-5.97e+04	5.97e+04
Type2[T.Fairy]	20.8603	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Fighting]	19.4992	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Fire]	19.7466	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Flying]	18.8183	3.05e+04	0.001	1.000	-5.97e+04	5.97e+04
Type2[T.Ghost]	21.2013	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Grass]	17.8891	3.05e+04	0.001	1.000	-5.97e+04	5.97e+04
Type2[T.Ground]	19.1237	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Ice]	19.4268	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.None]	19.1099	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Normal]	-2.0905	3.6e+04	-5.8e-05	1.000	-7.06e+04	7.06e+04
Type2[T.Poison]	19.7604	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Psychic]	19.8839	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Rock]	19.9676	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Steel]	19.8335	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
Type2[T.Water]	20.2095	3.05e+04	0.001	0.999	-5.97e+04	5.97e+04
hp	0.0309	0.005	5.764	0.000	0.020	0.041
atk	0.0162	0.006	2.873	0.004	0.005	0.027
Def	0.0299	0.006	5.008	0.000	0.018	0.042
spa	0.0156	0.005	3.113	0.002	0.006	0.025
spd	0.0187	0.006	3.066	0.002	0.007	0.031
spe	0.0436	0.006	7.065	0.000	0.032	0.056
=====						
AIC: 432.31937120804514						
Validation Accuracy: 0.9591836734693877						
True Positive Rate: 0.7692307692307693						
True Negative Rate: 0.9882352941176471						
Precision: 0.9090909090909091						

Figura 6: Tabela com coeficientes e estatísticas de avaliação do modelo final escolhido

Apesar da adição do Type2 melhorar consideravelmente a performance do modelo, é possível notar que todos os coeficientes referentes aos tipos possuem erros padrões enormes e seus intervalos de confiança contêm 0. Isso indica que não é possível ter a garantia de que essa variável preditora é significativa.

Quanto aos valores obtidos por máxima verossimilhança para eles, também tive surpresas. Pelo

mostrado na análise exploratória, imaginava que o coeficiente referente ao tipo Dragão seria o mais alto dentre os tipos, mas diversos outros tipos, incluindo Pedra e Água, que possuíam uma porcentagem baixa de Pokémons fortes, tiveram coeficientes maiores. Como era de se esperar, o coeficiente do tipo Normal é negativo, pois não há Pokémons fortes pertencentes a esse tipo.

Os *stats* que se mostraram mais relevantes para essa predição foram Speed, HP, e Defesa. Tanto HP quanto DEF me surpreenderam, acreditava que uma situação contrária iria acontecer, com valores mais elevados desses *stats* sendo prejudiciais à força do Pokémon. A Speed ser a mais relevante faz muito sentido, pois de fato é um atributo extremamente valioso em batalhas, definindo qual dos dois Pokémons irá atacar primeiro.

Agora, é preciso analisar a performance desse modelo no conjunto de testes:

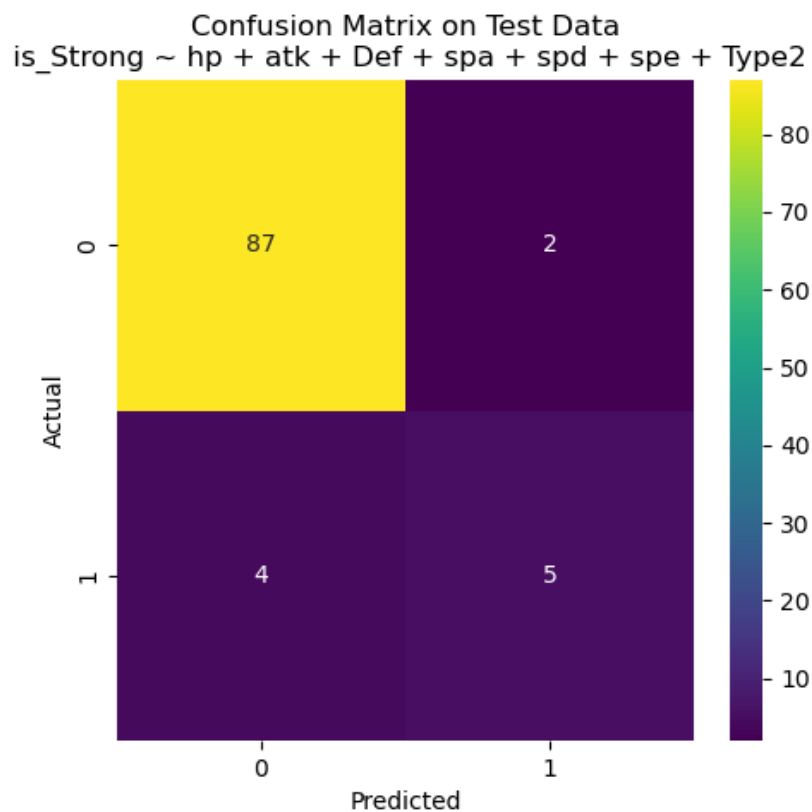


Figura 7: Matriz de confusão das predições do modelo escolhido no conjunto de testes.

Accuracy:	0.938775
True Positive Rate:	0.555556
True Negative Rate:	0.977528
Precision:	0.714286

Tabela 1: Estatísticas de teste do modelo escolhido

Apesar da acurácia ter sido boa, a True Positive Rate teve uma grande decaída quando comparada àquela obtida usando o conjunto de validação, e a precisão também se mostrou consideravelmente baixa. Parece que o modelo escolhido, apesar de se sair bem no conjunto de validação, não é um bom generalizador para o problema em questão.

4 Conclusão

4.1 Limitações

Acredito que o que possa ter causado a discrepância de performance do modelo entre o conjunto de testes e o conjunto de validação é o fato de que ambos esses conjuntos possuírem poucos Pokémons fortes. No dataset inteiro, temos que apenas 15,8% dos pontos são Pokémons considerados fortes. Na divisão aleatória feita para o conjunto de treino, teste, e validação, tivemos as seguintes proporções de Pokémons fortes:

Treino:	0.169643
Teste:	0.091837
Validação:	0.132653

Tabela 2: Proporções de Pokémons fortes em cada conjunto

Há uma diferença clara entre os percentuais de Pokémons fortes em cada um dos conjuntos. O conjunto de treino concentrou a maioria deles, enquanto o conjunto de teste acabou ficando com muitos poucos, tendo apenas 9 Pokémons fortes. Essa quantidade baixa faz com que os resultados estatísticos do modelo quando avaliado nesse conjunto tenham uma variância maior, se tornando menos confiáveis em geral.

Acredito que, se a seleção da partição dos conjuntos tivesse sido feita de forma mais inteligente, respeitando as proporções de pontos fracos e fortes em cada um dos conjuntos, a diferença de performance entre os dados de validação e teste não seria tão alta. Entretanto, mesmo com a proporção sendo correta, ainda assim teríamos uma quantidade baixa de Pokémons fortes nesses conjuntos, totalizando menos de 30 em cada um deles, o que não é um cenário ideal para a avaliação do modelo.

4.2 Direções futuras

Possíveis direções futuras para esse projeto incluem a utilização de técnicas de avaliação de modelo mais propícias para quando o conjunto de teste e validação possuem poucos dados pertencentes à classe que queremos prever, e a exploração dos atributos como *learnsets* e *abilities* que muito possivelmente iriam requerer modelos mais robustos para serem utilizados.

É importante mencionar que, conforme mais e mais Pokémons forem criados em jogos futuros, a quantidade de dados disponíveis para o treinamento e avaliação do modelo irão aumentar, o que irá diminuir algumas das limitações discutidas anteriormente.

4.3 O que aprendemos

O principal ponto de aprendizado que tive nesse trabalho foi entender a importância de separar os conjuntos de treino, teste, e validação corretamente. Além disso, poder praticar a modelagem estatística em um formato mais formal como o desse relatório foi um processo interessante, e acredito que as críticas que irei receber sobre ele serão bem construtivas para meu entendimento sobre como escrever em um formato acadêmico.