LEGENDS AREN'T BORN, THEY'RE MADE...

Gotta *Classify* 'Em All

By Rob Alterman
AKA Professor Oak

Background

Goal:

* **Classify Pokémon as legendary vs. non-legendary** *

Data:

- 809 unique Pokémon (generations 1-7), 80 of them considered legendary
- Web-scraped *pokemondb.net* and *serebii.net*, as well as made calls to the PokéAPI
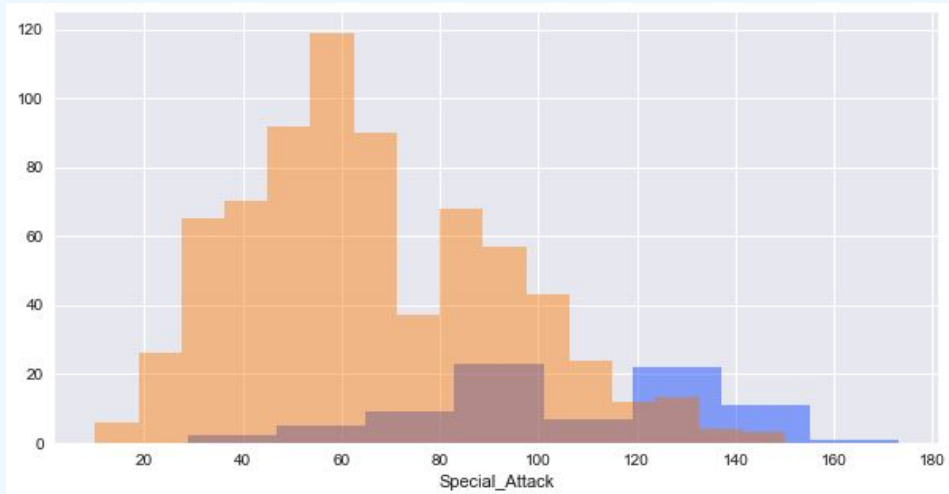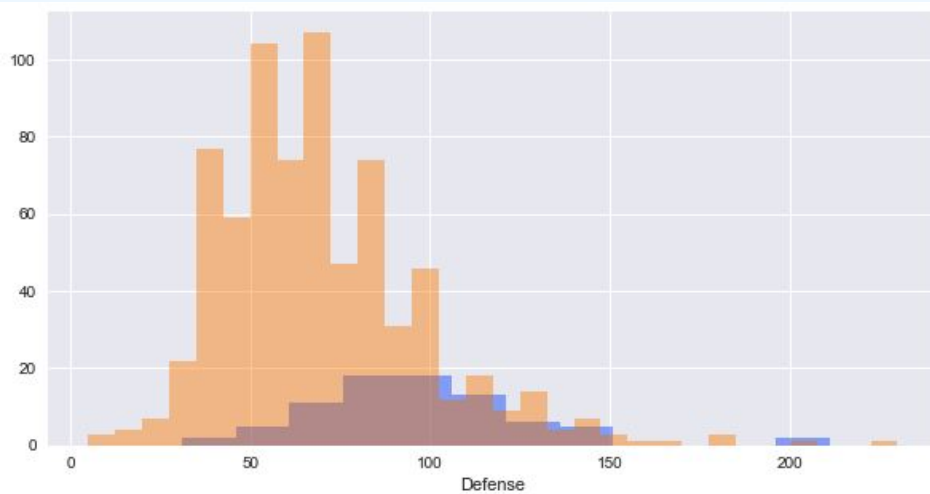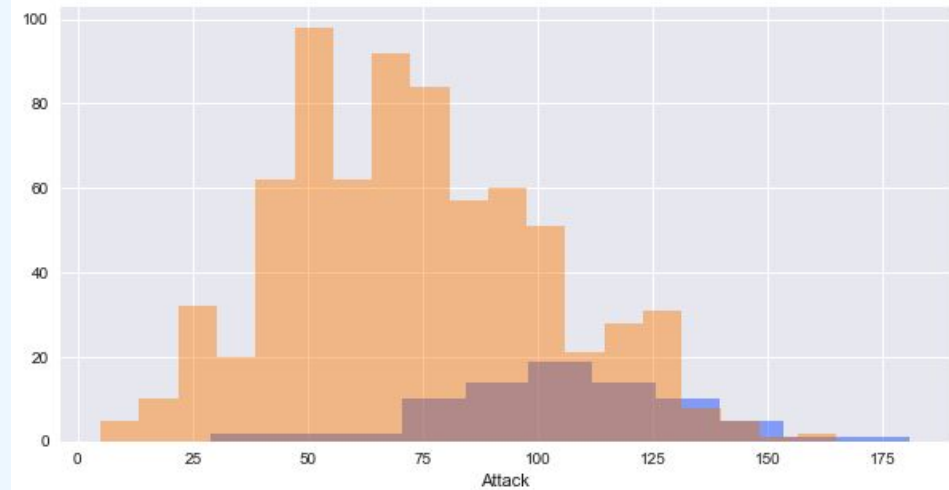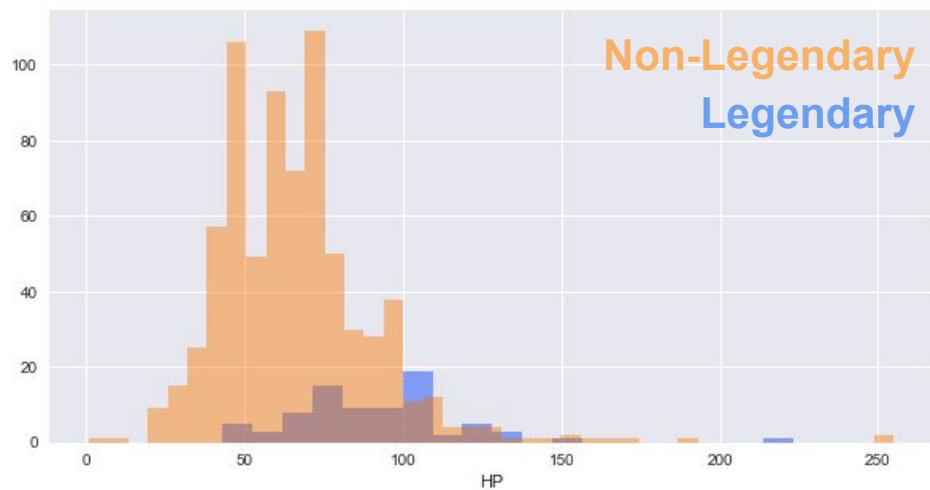
Features:

- HP, Attack, Defense, Special Attack, Special Defense, Speed, Height, Weight

Data

	Pokemon_Name	Type1	Type2	HP	Attack	Defense	Special_Attack	Special_Defense	Speed	Height_M	Weight_KG	Legendary
1	Bulbasaur	Grass	Poison	45.0	49.0	49.0	65.0	65.0	45.0	0.7	6.9	0
2	Ivysaur	Grass	Poison	60.0	62.0	63.0	80.0	80.0	60.0	1.0	13.0	0
3	Venusaur	Grass	Poison	80.0	82.0	83.0	100.0	100.0	80.0	2.0	100.0	0
4	Charmander	Fire	NaN	39.0	52.0	43.0	60.0	50.0	65.0	0.6	8.5	0
5	Charmeleon	Fire	NaN	58.0	64.0	58.0	80.0	65.0	80.0	1.1	19.0	0
6	Charizard	Fire	Flying	78.0	84.0	78.0	109.0	85.0	100.0	1.7	90.5	0
7	Squirtle	Water	NaN	44.0	48.0	65.0	50.0	64.0	43.0	0.5	9.0	0
8	Wartortle	Water	NaN	59.0	63.0	80.0	65.0	80.0	58.0	1.0	22.5	0
9	Blastoise	Water	NaN	79.0	83.0	100.0	85.0	105.0	78.0	1.6	85.5	0

Non-Legendary
Legendary
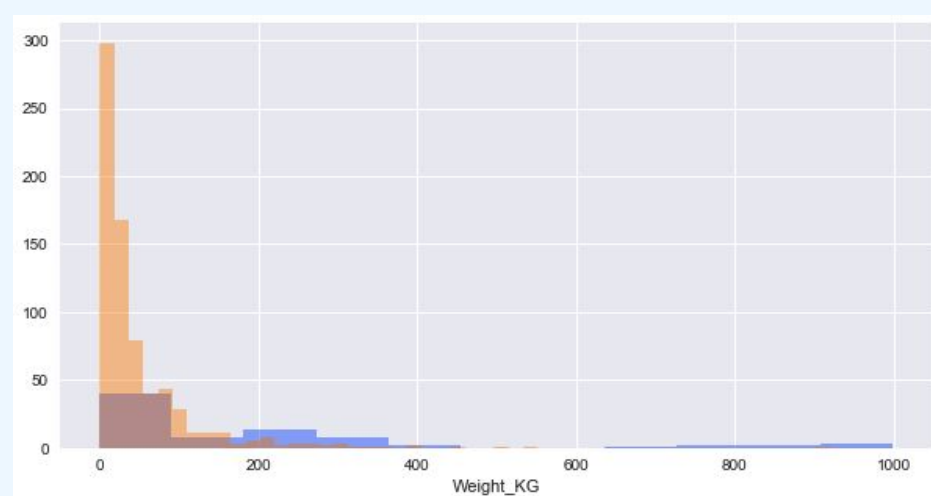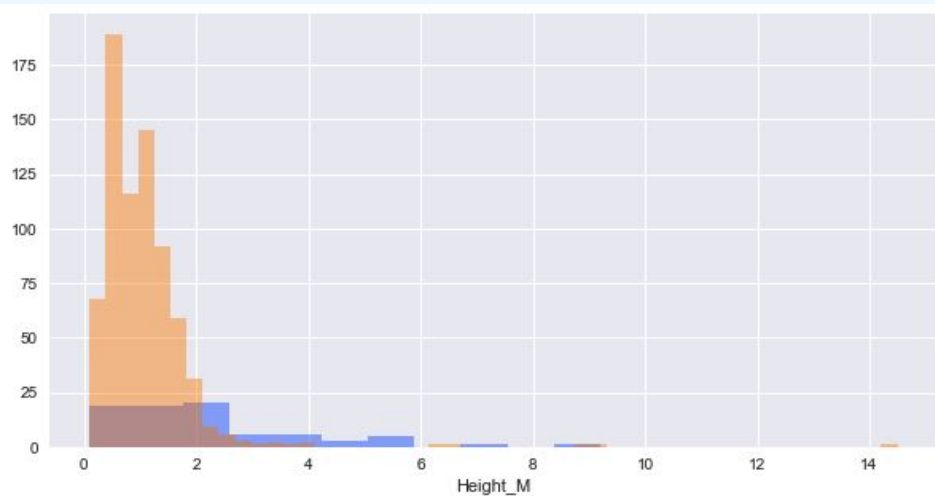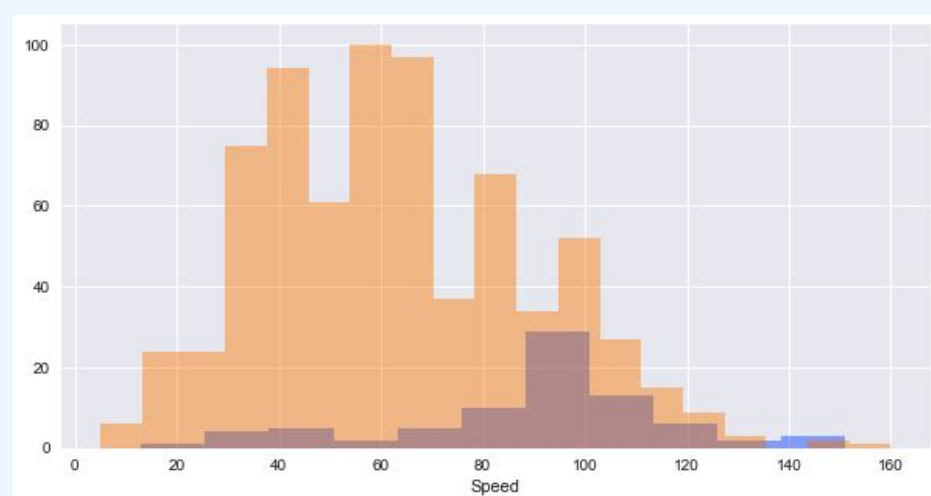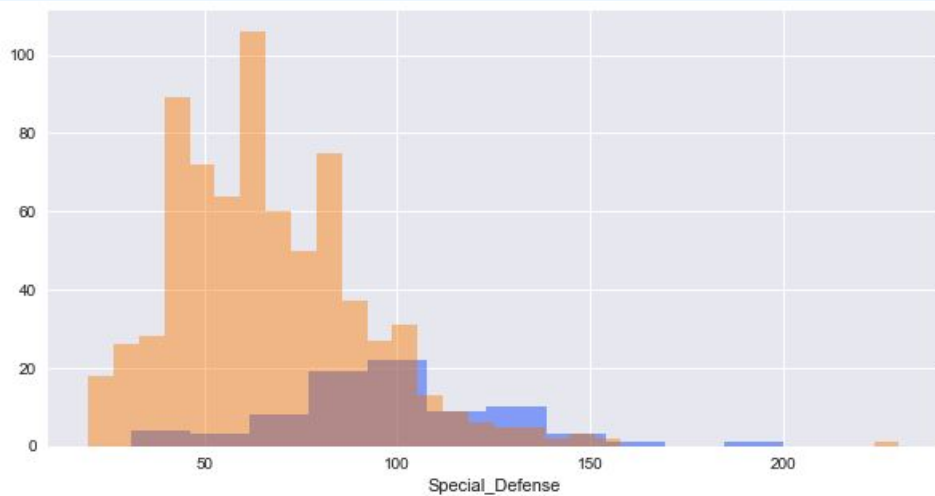
Modeling

- The aforementioned 8 features on the outcome variable of legendary or not

- Performed SMOTE on training data to account for class imbalance since only about 10% of all Pokémon in the data were legendary — synthesized more elements for the legendary class

- Ran Decision Tree, Random Forest, XGBoost, and SVM

Decision Tree

```
[[136  10]
 [  6  10]]
```

	precision	recall	f1-score	support
0	0.96	0.93	0.94	146
1	0.50	0.62	0.56	16
accuracy			0.90	162
macro avg	0.73	0.78	0.75	162
weighted avg	0.91	0.90	0.91	162

Random Forest

```
[[141   5]
 [  3  13]]
```

	precision	recall	f1-score	support
0	0.98	0.97	0.97	146
1	0.72	0.81	0.76	16
accuracy			0.95	162
macro avg	0.85	0.89	0.87	162
weighted avg	0.95	0.95	0.95	162

XGBoost 🤘

```
[[141   5]
 [  2  14]]
```

	precision	recall	f1-score	support
0	0.99	0.97	0.98	146
1	0.74	0.88	0.80	16
accuracy			0.96	162
macro avg	0.86	0.92	0.89	162
weighted avg	0.96	0.96	0.96	162

SVM

```
[[142   4]
 [  2  14]]
```

	precision	recall	f1-score	support
0	0.99	0.97	0.98	146
1	0.78	0.88	0.82	16
accuracy			0.96	162
macro avg	0.88	0.92	0.90	162
weighted avg	0.97	0.96	0.96	162

Grid Search & Final Model

- Ran Grid Search on SVM model with C values ranging from 0.001 to 1000 (moving one decimal place to the right for each value in that range) and for all possible kernel values

- Optimal parameters were C = 1000 and kernel = rbf (radial)

- Accuracy, Precision, and F1-Score all increased, while Recall dropped

Before				
[[142 4]				
[2 14]]				
	precision	recall	f1-score	support
0	0.99	0.97	0.98	146
1	0.78	0.88	0.82	16
accuracy			0.96	162
macro avg	0.88	0.92	0.90	162
weighted avg	0.97	0.96	0.96	162

After				
[[145 1]				
[4 12]]				
	precision	recall	f1-score	support
0	0.97	0.99	0.98	146
1	0.92	0.75	0.83	16
accuracy			0.97	162
macro avg	0.95	0.87	0.91	162
weighted avg	0.97	0.97	0.97	162

Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

ZUBAT :l7 ♀
TOTODILE :l7 ♂
HP: 24/ 24
TOTODILE
is confused!

|	Predicted 0	Predicted 1
Actual 0	Thinking it *is not* a legendary Pokémon when it indeed *is not*	Thinking it *is* a legendary Pokémon when it actually *is not*
Actual 1	Thinking it *is not* a legendary Pokémon when it actually **is**	Thinking it *is* a legendary Pokémon when it indeed *is*

In the case of the question at hand, I believe **precision** is preferred to **recall**: FN > FP

I rather think I **don't** have a legendary Pokémon when I actually **do**, than be disappointed when I think I have a legendary Pokémon when it really isn't one.

Feature Importance

- Because SVM does not have feature importance with the rbf kernel, I ran a grid search on my XGBoost model, since it was also very good, and calculated the feature importance on that model

- Feature importance seems to match correlations with 'legendary' from heatmap
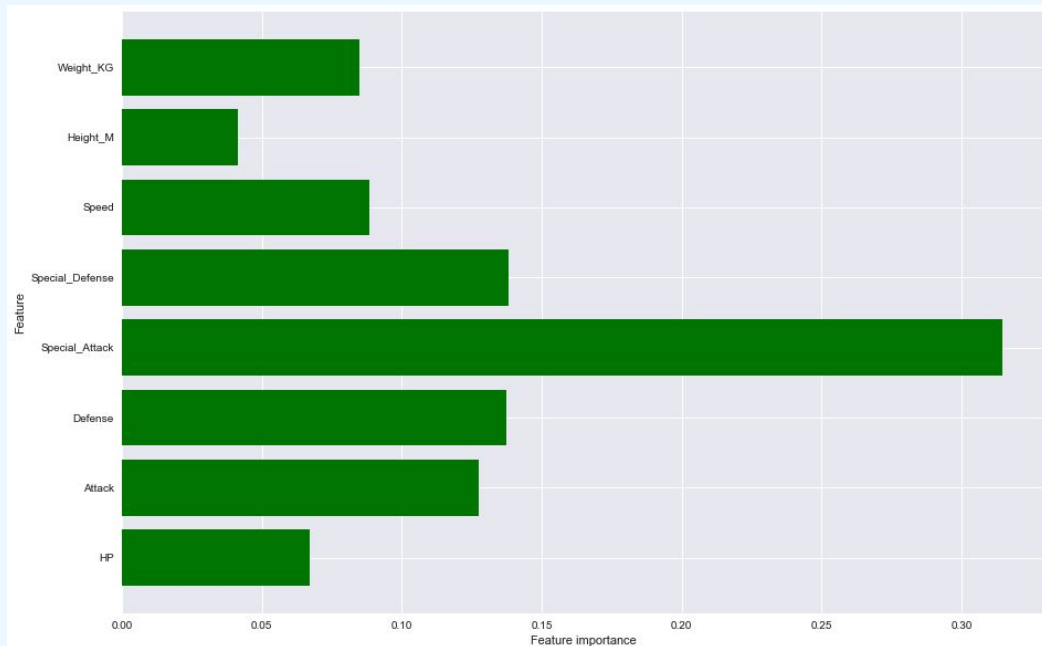
Misclassification Evaluation

Misclassified Pokémon from Test Set

	Pokemon_Name	Type1	Type2	HP	Attack	Defense	Special_Attack	Special_Defense	Speed	Height_M	Weight_KG	Legendary	label
808	Meltan	Steel	NaN	46.0	65.0	65.0	55.0	35.0	34.0	0.2	8.0	1	False Neg
720	Hoopa	Psychic	Ghost	80.0	110.0	60.0	150.0	130.0	70.0	0.5	9.0	1	False Neg
763	Tsareena	Grass	NaN	72.0	120.0	98.0	50.0	98.0	72.0	1.2	21.4	0	False Pos
793	Nihilego	Rock	Poison	109.0	53.0	47.0	127.0	131.0	103.0	1.2	55.5	1	False Neg
778	Mimikyu	Ghost	Fairy	55.0	90.0	80.0	50.0	105.0	96.0	0.2	0.7	0	False Pos
796	Xurkitree	Electric	NaN	83.0	89.0	71.0	173.0	71.0	83.0	3.8	100.0	1	False Neg

All Other Pokémon

Legendary	HP	Attack	Defense	Special_Attack	Special_Defense	Speed	Height_M	Weight_KG
0	66.279230	73.149931	68.701513	65.510316	66.676754	62.896836	1.051307	48.096286
1	92.513158	105.302632	100.315789	106.315789	98.118421	92.223684	2.228947	203.903947

That's All Folks!