POKEMON PROJECT

	#	Name	Type 1	Type 2	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	1	False
1	2	Ivysaur	Grass	Poison	60	62	63	80	80	60	1	False
2	3	Venusaur	Grass	Poison	80	82	83	100	100	80	1	False
3	3	VenusaurMega Venusaur	Grass	Poison	80	100	123	122	120	80	1	False
4	4	Charmander	Fire	NaN	39	52	43	60	50	65	1	False
5	5	Charmeleon	Fire	NaN	58	64	58	80	65	80	1	False
6	6	Charizard	Fire	Flying	78	84	78	109	85	100	1	False
7	6	CharizardMega Charizard X	Fire	Dragon	78	130	111	130	85	100	1	False
8	6	CharizardMega Charizard Y	Fire	Flying	78	104	78	159	115	100	1	False
9	7	Squirtle	Water	NaN	44	48	65	50	64	43	1	False

The dataset

The dataset that I used contains various attributes about Pokemon such as their HP, their speed while attacking, their speed while defending and their legendary status. The problem statement was what factors are the most important in predicting the legendary status of a particular Pokemon for all the Pokemon in the dataset.

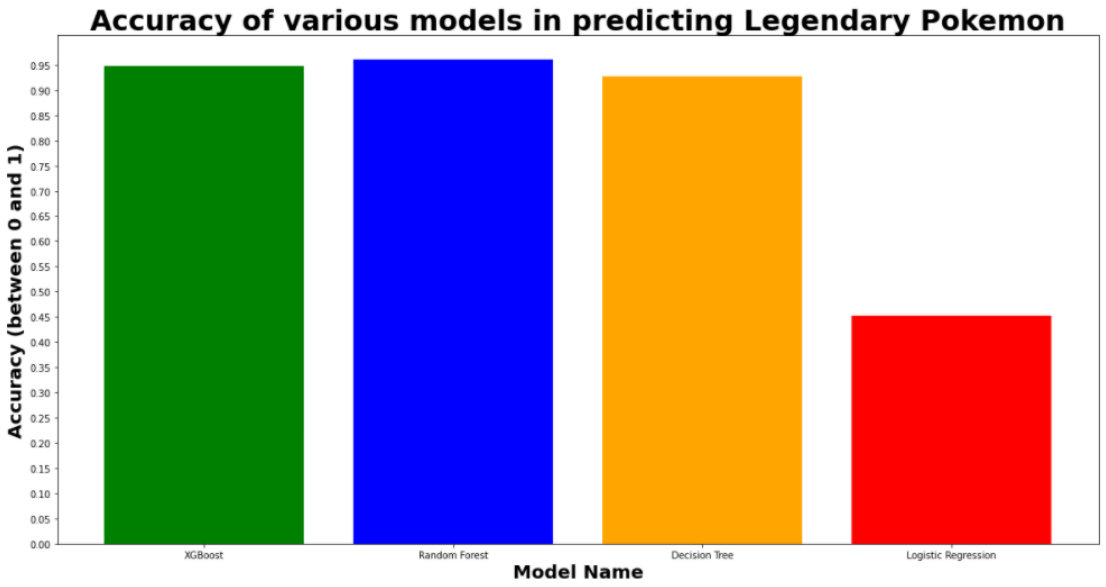
I first got a summary of the dataset using the head and tail features. Once I did that,
I checked whether there are any missing values for any of the Pokemon. I found out that 386 Pokemon in the dataset contained missing Type 2 values. I then researched and found out that a Pokemon's missing Type 2 value can be replaced with its Type 1 value. I then checked to see whether a particular Type 1 value is associated with a particular Type 2 value for every Type 1 value. If I found an association, I replaced the missing Type 2 values (based on what a particular Pokemon's Type 1 value was). If I didn't find an association, I replaced the missing Type 2 value with the Type 1 value for that particular Pokemon. I then converted the dataset into machine readable format by mapping false and true to 0 and 1 respectively for the legendary column and using the get_dummies feature on columns that contained non-numeric values. I then used StandardScaler to standardize the values in such a way that the mean is 0 and the standard deviation is 1. I then used MinMaxScaler to ensure that all the values are between 0 and 1 and removed duplicate rows from the dataset. I then removed rows in which values for any particular column were more than 3 standard deviations away from the mean of that particular column. I then used Pearson's correlation method to find the most important factors in predicting the legendary status of a particular Pokemon.

	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary	Type 1 Bug	Type 1 Dark	...	Type 2 Ghost	Type 2 Grass	Type 2 Ground
HP	1.000000	0.524889	0.384482	0.424042	0.431564	0.221272	0.098883	0.288180	-0.144390	-0.008302	...	-0.052136	-0.086774	0.073116
Attack	0.524889	1.000000	0.481025	0.385004	0.303980	0.375177	0.062193	0.295207	-0.074551	0.089843	...	0.027564	-0.022134	0.087435
Defense	0.384482	0.481025	1.000000	0.288500	0.550280	0.088885	0.073089	0.248378	-0.041457	-0.012984	...	0.051581	0.083417	0.099988
Sp. Atk	0.424042	0.385004	0.288500	1.000000	0.551685	0.443798	0.051117	0.403541	-0.161228	0.020830	...	0.007485	-0.054014	-0.086785
Sp. Def	0.431564	0.303980	0.550280	0.551685	1.000000	0.295556	0.051167	0.319438	-0.094682	-0.007491	...	0.040520	0.070982	-0.073000
Speed	0.221272	0.375177	0.088885	0.443798	0.295556	1.000000	-0.018933	0.297437	-0.088858	0.082414	...	-0.087574	-0.090908	-0.103040
Generation	0.098883	0.062193	0.073089	0.051117	0.051167	-0.018933	1.000000	0.103447	-0.015123	0.085253	...	0.118822	0.112971	-0.085559
Legendary	0.288180	0.295207	0.248378	0.403541	0.319438	0.297437	0.103447	1.000000	-0.080398	-0.000036	...	0.003832	-0.057781	-0.021788
Type 1 Bug	-0.144390	-0.074551	-0.041457	-0.161228	-0.094682	-0.088858	-0.015123	-0.080398	1.000000	-0.083153	...	-0.006981	0.087407	-0.038289
Type 1 Dark	-0.008302	0.089843	-0.012984	0.020830	-0.007491	0.082414	0.085253	-0.000036	-0.083153	1.000000	...	0.070855	-0.045371	-0.052414
Type 1 Dragon	0.127071	0.208033	0.102288	0.132087	0.138885	0.100461	0.063849	0.221871	-0.082083	-0.041881	...	-0.027809	-0.044603	0.090171
Type 1 Electric	-0.088957	-0.071071	-0.051084	0.148584	0.031302	0.148800	-0.010592	0.028800	-0.075919	-0.051190	...	0.007894	-0.027881	-0.083010
Type 1 Fairy	0.042881	-0.079011	-0.034135	0.023331	0.058490	-0.108548	0.057785	-0.001229	-0.044910	-0.030282	...	-0.020116	-0.032285	-0.037273
Type 1 Fighting	0.020829	0.117195	-0.042133	-0.114553	-0.042827	-0.010998	0.003054	-0.050454	-0.058776	-0.039832	...	-0.028328	-0.042227	-0.048782
Type 1 Fire	0.021075	0.051882	-0.045128	0.148116	0.001624	0.081805	-0.017490	0.015151	-0.082138	-0.055384	...	-0.038792	-0.059011	-0.002005
Type 1 Flying	0.010878	0.001917	-0.015109	0.053582	0.005783	0.090383	0.094188	0.128974	-0.022276	-0.015020	...	-0.009978	-0.016004	-0.018488
Type 1 Ghost	-0.107812	-0.042159	0.080917	0.012724	0.028588	-0.058343	0.114822	-0.051415	-0.059898	-0.040388	...	-0.028829	0.823518	-0.049711
Type 1 Grass	-0.002971	-0.048752	-0.013381	0.081644	-0.000289	-0.085374	0.002538	-0.028148	-0.097548	-0.065774	...	-0.043894	-0.070082	-0.081877
Type 1 Ground	0.051423	0.100889	0.078850	-0.120807	-0.088893	-0.033984	-0.022812	0.027052	-0.083153	-0.042582	...	0.070855	-0.045371	0.310258
Type 1 Ice	0.024566	-0.034492	-0.043728	0.036488	0.010681	-0.012311	0.011478	-0.013431	-0.052878	-0.035653	...	0.034681	-0.037988	0.054856
Type 1 Normal	0.075812	-0.053248	-0.157047	-0.197388	-0.122983	0.059207	-0.053587	-0.088218	-0.115082	-0.077583	...	-0.051539	-0.044523	-0.078734
Type 1 Poison	-0.002015	-0.020853	-0.022303	-0.070354	-0.048003	-0.028874	-0.094925	-0.051415	-0.059898	-0.040388	...	-0.028829	-0.043031	0.008877
Type 1 Psychic	-0.003082	-0.116148	-0.070201	0.177800	0.128977	0.057908	0.025714	0.088217	-0.078830	-0.053018	...	0.005408	-0.030428	-0.085259
Type 1 Rock	-0.025789	0.123302	0.219285	-0.058838	0.027725	-0.095055	0.015761	0.008788	-0.074070	-0.049944	...	-0.033178	0.001738	0.083430
Type 1 Steel	-0.028393	0.062348	0.256850	-0.012581	0.078511	-0.081571	0.078838	0.078884	-0.054101	-0.036479	...	0.204251	-0.038888	-0.044901
Type 1 Water	0.042951	-0.083189	-0.016881	0.017803	-0.005148	-0.028338	-0.119848	-0.059443	-0.123303	-0.083140	...	0.001285	-0.034216	0.056981
Type 2 Bug	-0.103941	-0.108470	-0.073717	-0.157332	-0.142836	-0.090501	0.012483	-0.042095	0.433188	-0.033088	...	-0.021986	-0.035231	-0.040700
Type 2 Dark	0.024848	0.147787	-0.023419	0.045558	-0.042421	0.071393	0.041382	-0.024393	-0.080998	0.308115	...	-0.027323	-0.043823	-0.050628
Type 2 Dragon	0.018888	0.099609	0.037139	0.082730	0.052497	0.051050	0.107359	0.059509	-0.080998	0.083044	...	-0.027323	-0.043823	-0.050628
Type 2 Electric	-0.022910	-0.055850	-0.110885	0.049497	-0.023036	0.110029	-0.032388	-0.003429	-0.019072	-0.043995	...	-0.029228	-0.046876	-0.054153
Type 2 Fairy	-0.025587	-0.093484	0.012347	0.019876	0.078897	-0.040740	0.023483	0.016112	-0.054101	-0.036479	...	-0.024233	-0.038888	-0.044901
Type 2 Fighting	0.080779	0.198927	0.000020	-0.079314	0.016707	0.043849	0.041774	0.005194	-0.054632	0.007018	...	-0.033594	-0.053882	-0.062246
Type 2 Fire	-0.032834	-0.009380	-0.080198	0.093038	-0.001071	0.054158	-0.028543	0.011710	-0.028575	0.042425	...	-0.031904	-0.051172	-0.059115
Type 2 Flying	0.083829	-0.039464	-0.126868	-0.073578	-0.041031	0.155740	-0.054013	0.032338	-0.010997	-0.026887	...	-0.071690	-0.114983	-0.132833
Type 2 Ghost	-0.052136	0.027564	0.051581	0.007485	0.040520	-0.087574	0.118822	0.003832	-0.006981	0.070855	...	1.000000	-0.030141	-0.034819
Type 2 Grass	-0.086774	-0.022134	0.083417	-0.054014	0.070982	-0.090908	0.112971	-0.057781	0.087407	-0.045371	...	-0.030141	1.000000	-0.055847
Type 2 Ground	0.073116	0.087435	0.099988	-0.086785	-0.073000	-0.103040	-0.085559	-0.021788	-0.038289	-0.052414	...	-0.034819	-0.055847	1.000000
Type 2 Ice	0.084563	0.045687	0.008149	0.071899	0.073530	0.006404	0.083449	0.016112	-0.054101	0.041188	...	-0.024233	-0.038888	-0.044901
Type 2 Normal	-0.013207	-0.059376	-0.048236	0.038829	-0.017512	0.055990	0.118029	-0.019122	-0.022276	-0.015020	...	-0.009978	-0.016004	-0.018488
Type 2 Poison	-0.080344	-0.100181	-0.085480	-0.023051	-0.058701	-0.074880	-0.125113	-0.073438	0.077211	-0.071325	...	-0.047382	-0.075998	-0.087793
Type 2 Psychic	0.047742	-0.104868	0.030441	0.183075	0.162943	0.025404	-0.019100	0.115292	-0.092075	-0.013828	...	-0.041243	-0.066150	-0.076419
Type 2 Rock	0.014074	0.087185	0.207238	-0.125088	-0.100209	-0.193189	0.020351	-0.043218	0.008443	-0.033948	...	-0.022552	-0.036171	-0.041788
Type 2 Steel	-0.028602	0.082884	0.205676	0.032228	0.108453	-0.051598	0.063812	0.011598	0.128185	0.038511	...	-0.025299	-0.040578	-0.046877
Type 2 Water	-0.069891	-0.081198	0.003716	0.004427	-0.067787	-0.023988	-0.078712	-0.028148	-0.081275	-0.065774	...	-0.043894	-0.070082	-0.080961

I found out that the 5 best predictors of whether a pokemon will be considered legendary or not are its Speed Attack, its Speed Defense, its Speed, its Attack and its HP respectively.

I then split the Pokemon dataset into training data and testing data. I then created the models to use in order to predict a Pokemon's legendary status. The models I used were decision trees, random forest, XGBoost, and logistic regression. I trained these models using the training dataset. After training, each model predicted whether a particular pokemon should be classified as legendary or

not for every Pokemon in the testing dataset. Then the accuracy of each model was determined by comparing the predicted legendary status for every Pokemon in the testing dataset with the actual legendary status for every Pokemon in the testing dataset.



The Random Forest model with a maximum depth of 100 had the highest accuracy out of all the models at 96.05%, whereas the logistic regression had the lowest accuracy out of all the models at 45.22%.The XGBoost model had the next highest accuracy at 94.74%, followed by the decision tree classifier at 92.76%.