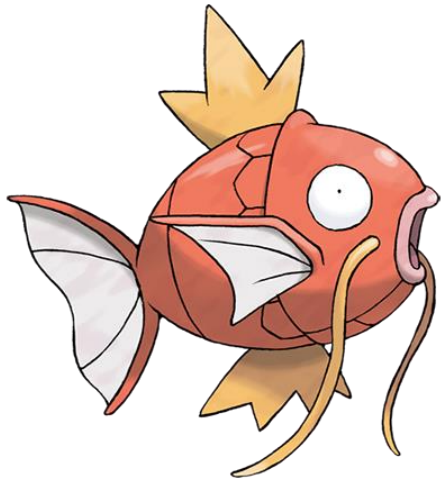
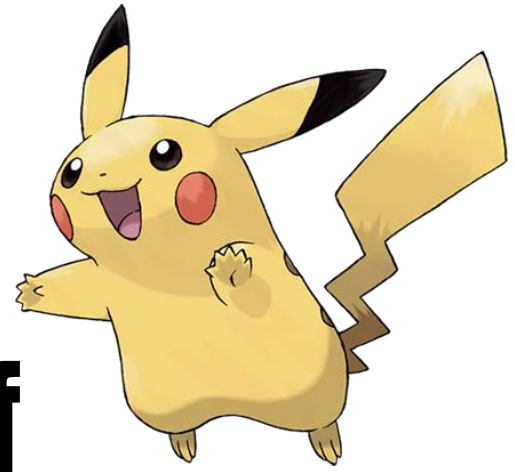
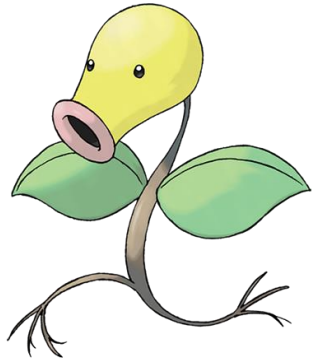




Predicting the Spawn of Pokémon in Pokémon GO

Travis Roundy, Justin Olson, Anna Yudina, Taylor Thomas

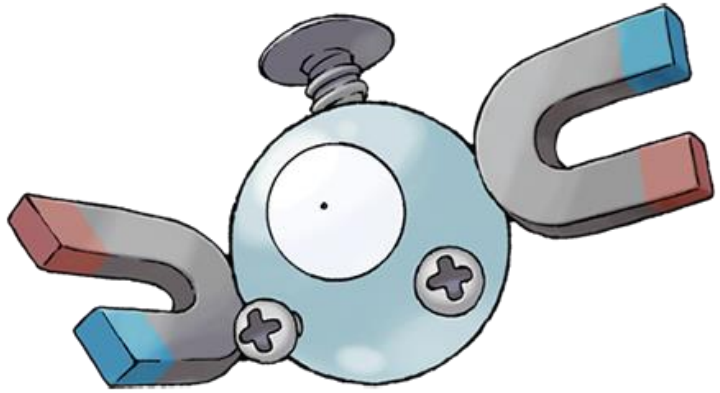


Pokémon GO

- Mobile game for iOS and Android
- 150 Pokémon to catch to date
- GPS oriented

Our Tools

- Vowpal Wabbit
  - Train/Test Data
- Python
  - Parse CSV into VW
  - JSON



Our Dataset

pokemonId	latitude	longitude	appeared	_id	cellId_90r	cellId_180	cellId_370	cellId_730	cellId_146	cellId_292	cellId_585	appeared	appeared	appeared	appeared	appeared	appeared	appeared	terrainTyp	clos
16	20.52575	-97.4608	2016-09-0	NTgxMDk:	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	57	dummy_d	8	8	2016	14	F
133	20.5237	-97.4612	2016-09-0	OTQ1NDg	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	57	dummy_d	8	8	2016	14	F
16	38.90359	-77.1998	2016-09-0	NTQ0OTQ	9.92E+18	9.92E+18	9.92E+18	9.92E+18	9.92E+18	9.92E+18	9.92E+18	night	5	57	dummy_d	8	8	2016	13	F
13	47.6659	-122.313	2016-09-0	NTU2MTU	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	night	5	56	dummy_d	8	8	2016	0	T
133	47.66645	-122.312	2016-09-0	MTY2ODg	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	night	5	56	dummy_d	8	8	2016	0	T
21	-31.955	115.8536	2016-09-0	MTA4NTIv	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	night	5	55	dummy_d	8	8	2016	13	F
66	-31.9542	115.852	2016-09-0	NzMxNzg:	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	night	5	55	dummy_d	8	8	2016	13	F
27	26.23526	-98.1976	2016-09-0	MTIzNDcy	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	night	5	55	dummy_d	8	8	2016	13	F
35	20.52555	-97.4588	2016-09-0	MzcwMjg(	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	55	dummy_d	8	8	2016	14	F
19	32.92856	-84.3403	2016-09-0	NzU5MDI(	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	night	5	54	dummy_d	8	8	2016	8	F
116	32.93065	-84.3399	2016-09-0	OTMyMjg:	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	night	5	53	dummy_d	8	8	2016	8	F
74	32.94365	-84.3344	2016-09-0	MTQwNDg	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	9.87E+18	night	5	53	dummy_d	8	8	2016	8	F
16	26.23555	-98.1972	2016-09-0	MTU5OTc(	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	night	5	53	dummy_d	8	8	2016	13	F
19	20.52577	-97.4602	2016-09-0	MTM4NDc	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	53	dummy_d	8	8	2016	14	F
19	26.23603	-98.1969	2016-09-0	NDExMTc:	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	night	5	53	dummy_d	8	8	2016	13	F
19	47.66433	-122.313	2016-09-0	MTA4NjQ(	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	6.09E+18	night	5	52	dummy_d	8	8	2016	0	T
16	20.52649	-97.4607	2016-09-0	MTM4NjE(	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	51	dummy_d	8	8	2016	14	F
13	53.61142	-113.37	2016-09-0	MTc3MTQ	6.03E+18	6.03E+18	6.03E+18	6.03E+18	6.03E+18	6.03E+18	6.03E+18	night	5	51	dummy_d	8	8	2016	12	F
32	20.52571	-97.4595	2016-09-0	MjMwNDk	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	9.65E+18	night	5	51	dummy_d	8	8	2016	14	F
129	-31.9543	115.8516	2016-09-0	Mzc5MDU	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	3.04E+18	night	5	51	dummy_d	8	8	2016	13	F
23	26.33186	-81.8145	2016-09-0	MTcxOTM	9.86E+18	9.86E+18	9.86E+18	9.86E+18	9.86E+18	9.86E+18	9.86E+18	night	5	50	dummy_d	8	8	2016	13	T
75	26.23553	-98.1977	2016-09-0	NDYwNzM	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	9.68E+18	night	5	50	dummy_d	8	8	2016	13	F

https://www.kaggle.com/semioniy/predictemall

Our Dataset: Classes and Features

- Pokémon ID’s from 1 – 150

- Include:
  - [latitude", "longitude", "appearedHour", "appearedMinute", "terrainType", "closeToWater", "city", "weather", "temperature", "windSpeed", "windBearing", "pressure", "weatherIcon","urban", "suburban", "midurban", "rural", "cooc_1" - "cooc_151"]

- Ignore:
  - ["_id“,"appearedLocalTime","pokestopDistanceKm","gymDistanceKm", "class"]

Our Dataset: The Data

- 296,022 Data Points
  - 52,114 Pidgey
  - 39,637 Rattata
  - 27,367 Weedle
    - **Total 40.24% of the data**
- Obvious skewing

Testing our Data

- 4 Tests:
  - 23 Classes with 3000 data points each
  - 27 Classes with 2000 data points each
  - 45 Classes with 1000 data points each
  - 150 Classes with up to 5000 Data points each
- Training/Test data randomized from single VW file

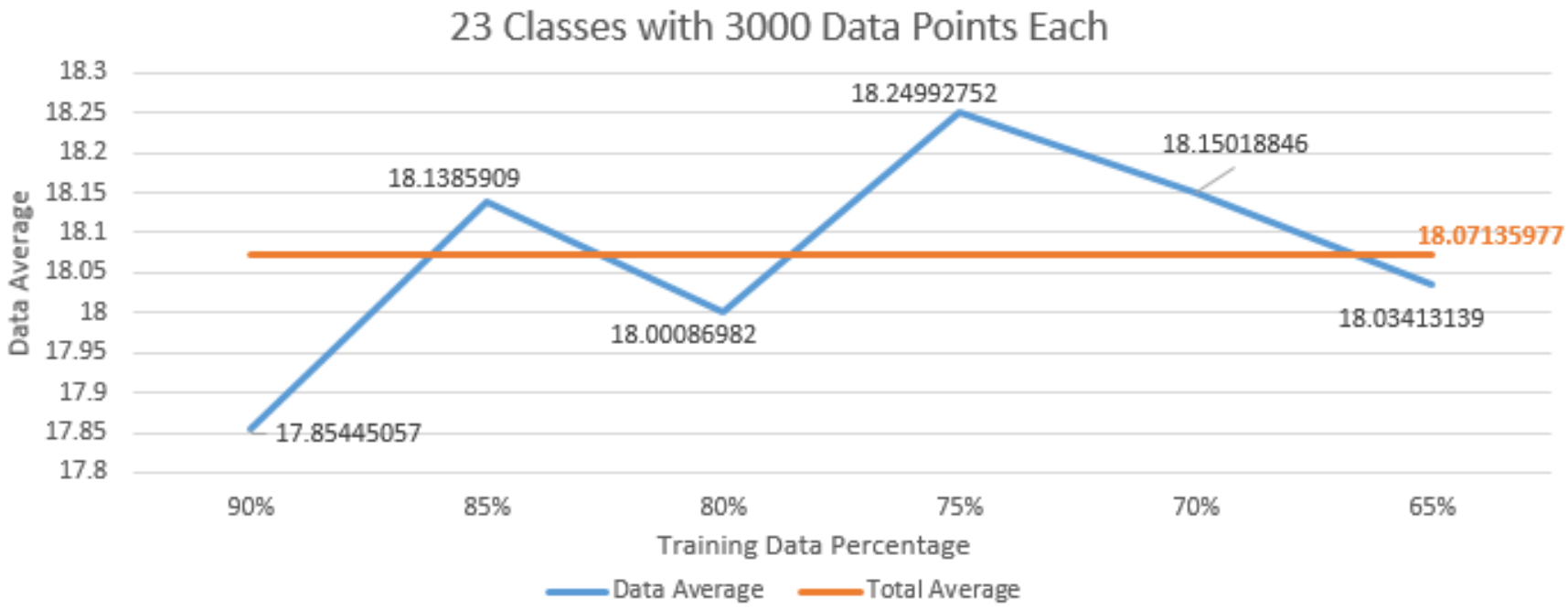
Testing our Data

- Training Script:
  - 50 Passes
  - 0.2 Learning Rate
  - ooa of either 23, 27, 45, or 150
  - Logistic Loss Function
- 6 Training Sizes:
  - 90%, 85%, 80%, 75%, 70%, 65%
- 5 Runs Per Training Size

Test 1: Limiting to Classes with >3000 Points

- 23 Classes
- 3000 Instances of each class
  - Baseline of 4.348%

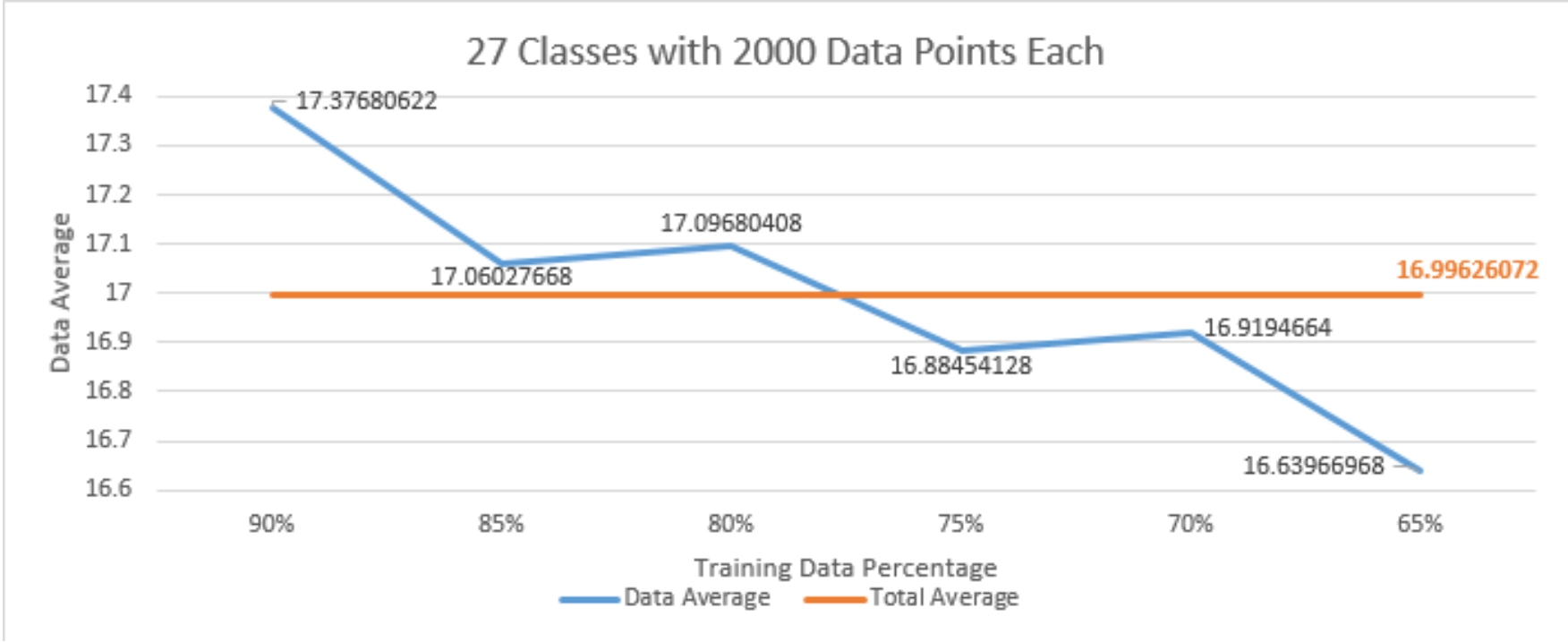
- 18.071% Accuracy



Test 2: Limiting to Classes with >2000 Points

- 27 Classes
- 2000 Instances of each class
  - Baseline of 3.074%

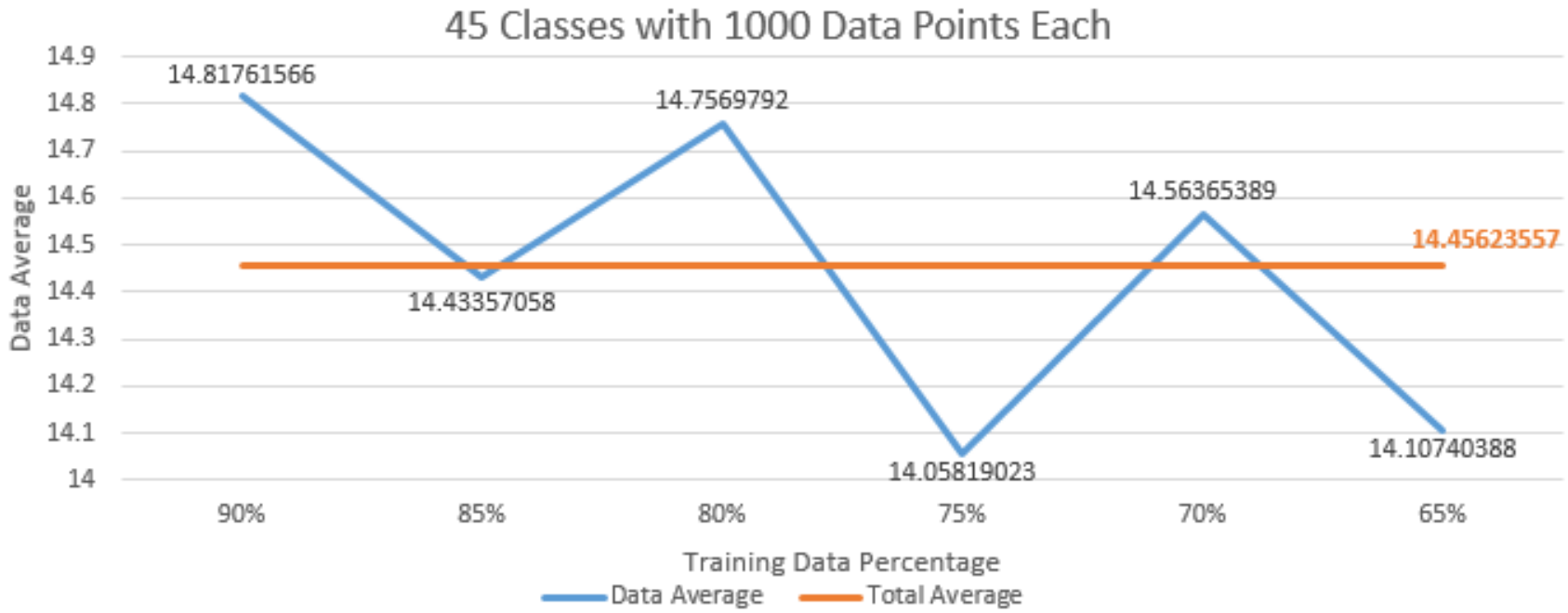
- 16.996% Accuracy



Test 3: Limiting to Classes with >1000 Points

- 45 Classes
- 1000 Instances of each class
  - Baseline of 2.222%

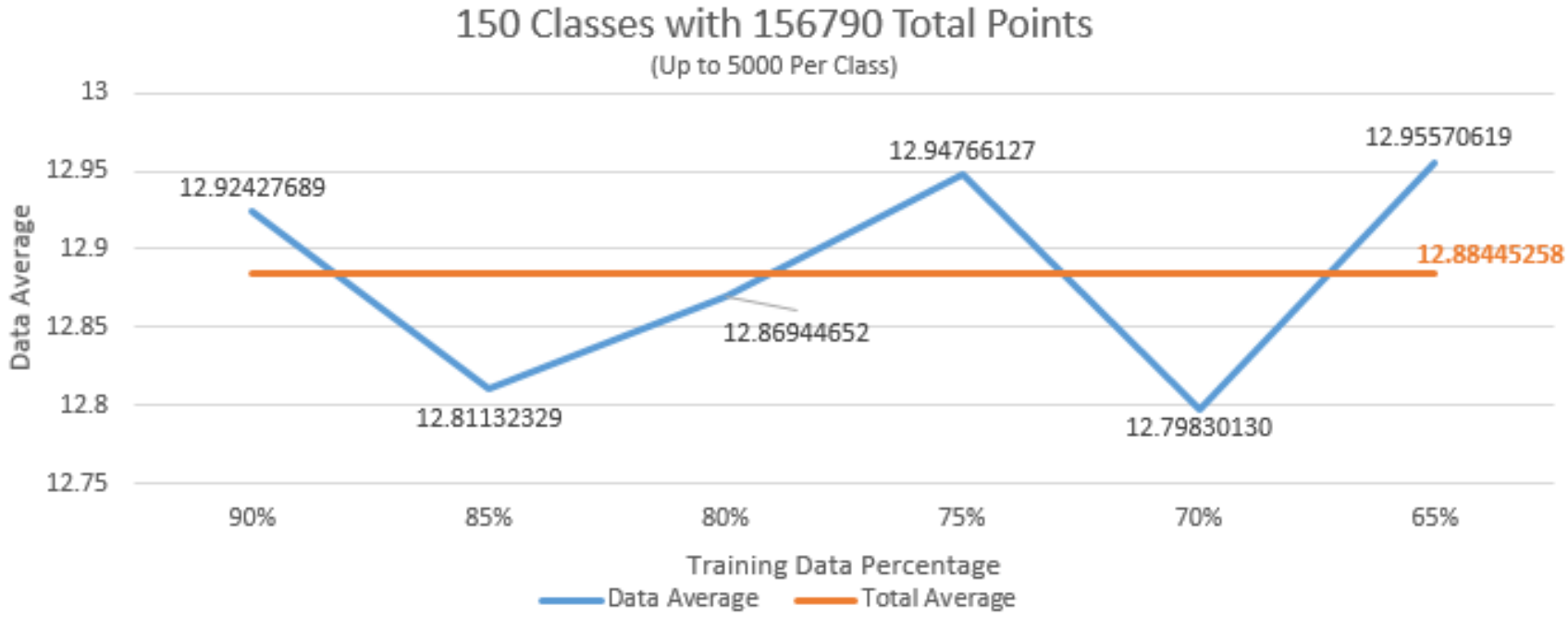
- 14.456% Accuracy



Test 4: All Classes with up to 5000 Points

- 150 Classes
- Up to 5000 Instances of each class, 156,790 total points
  - Baseline of 3.190%

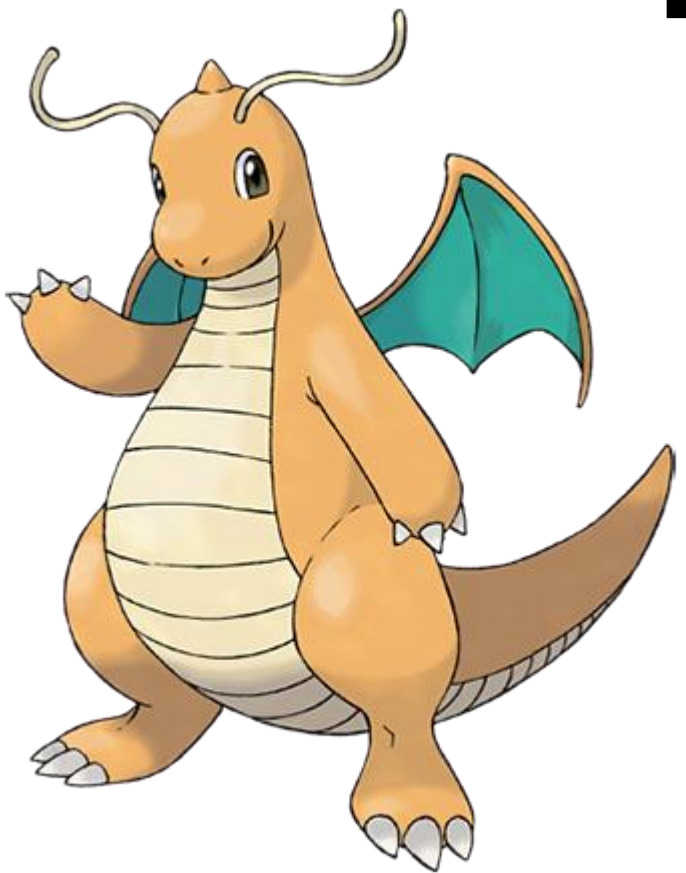
- 12.884% Accuracy



Future Work and Conclusion

- Determine Best/Worst features
  - Raise accuracy of classification
- Run more tests with different data points
- Gain more data for lacking Pokémon
- Run same tests with various loss functions

- Overall, we gained between 9% and 14% increases in accuracy between the baseline and our results



Thank you!

Any Questions?

Images from Pokemon.com

- http://assets22.pokemon.com/assets/cms2/img/pokedex/full/001.png
  - 001 changed to the corresponding Pokémon ID to get images.