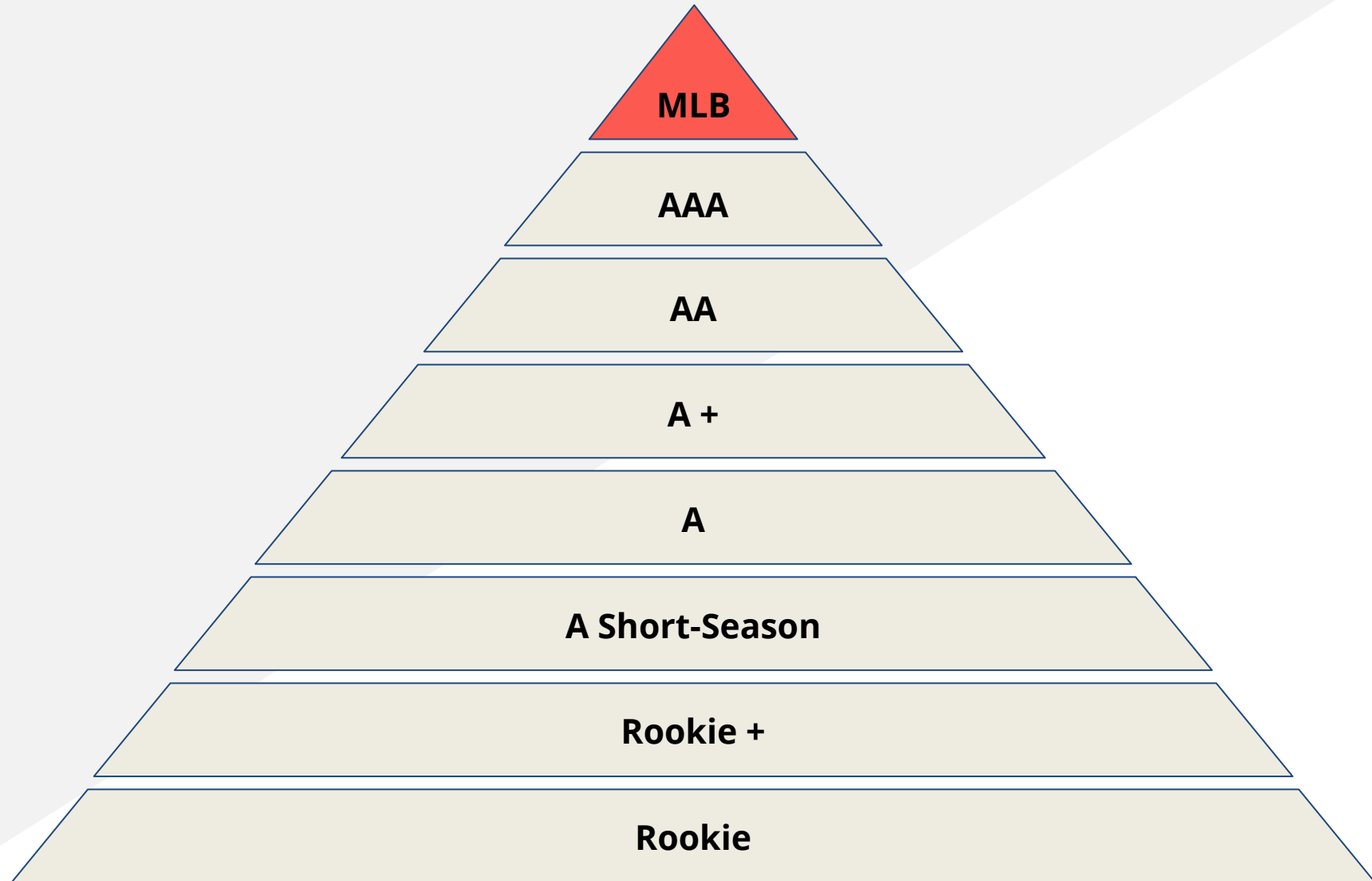


The Farm : Who gets harvested?

By Jun Soo Kim

Minor League Baseball (= Farm System)





Following the dreams...



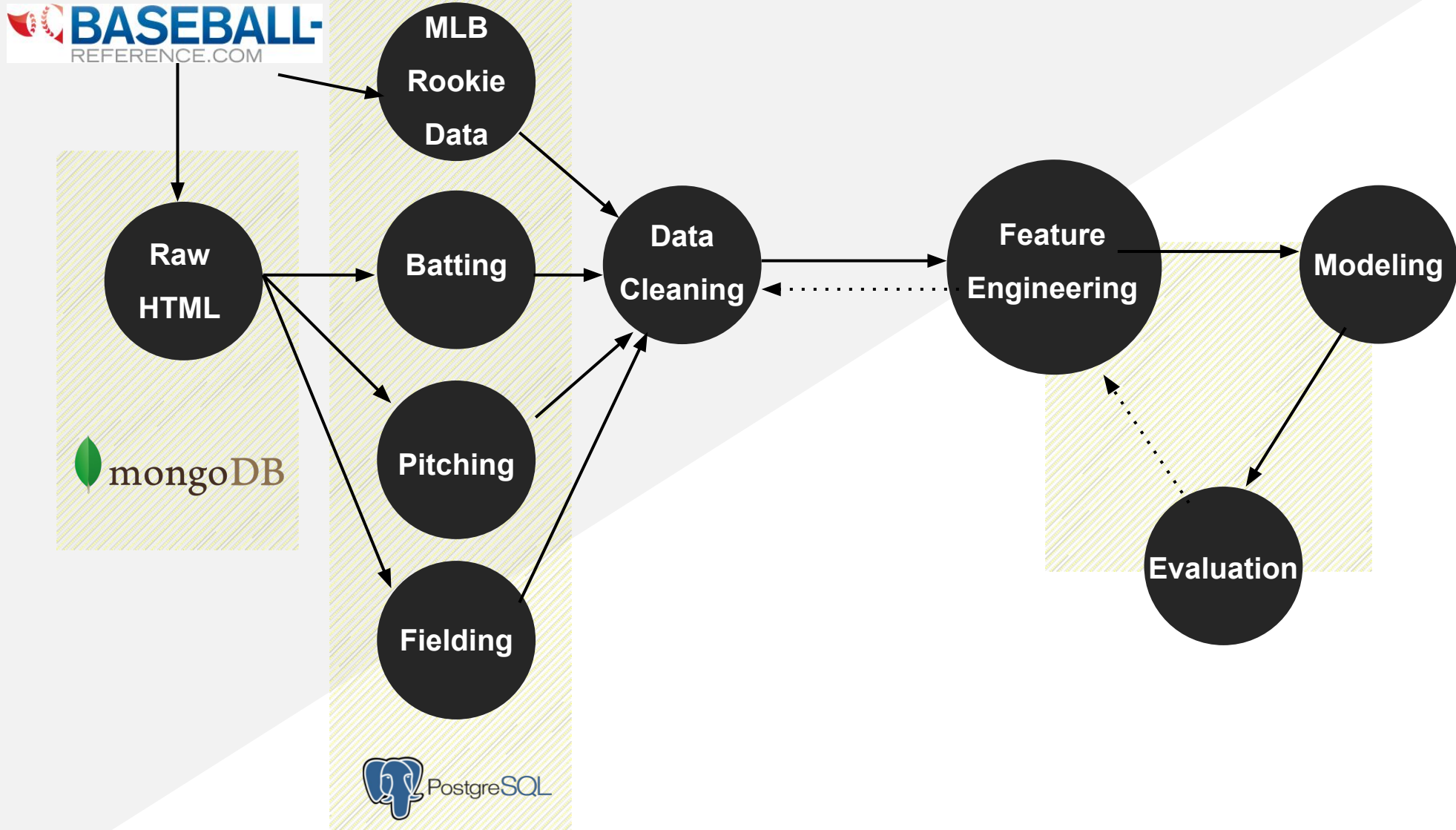
Players qualify for the minimum benefit pension plan of \$34,000 per year after **43 days** in the MLB

Questions

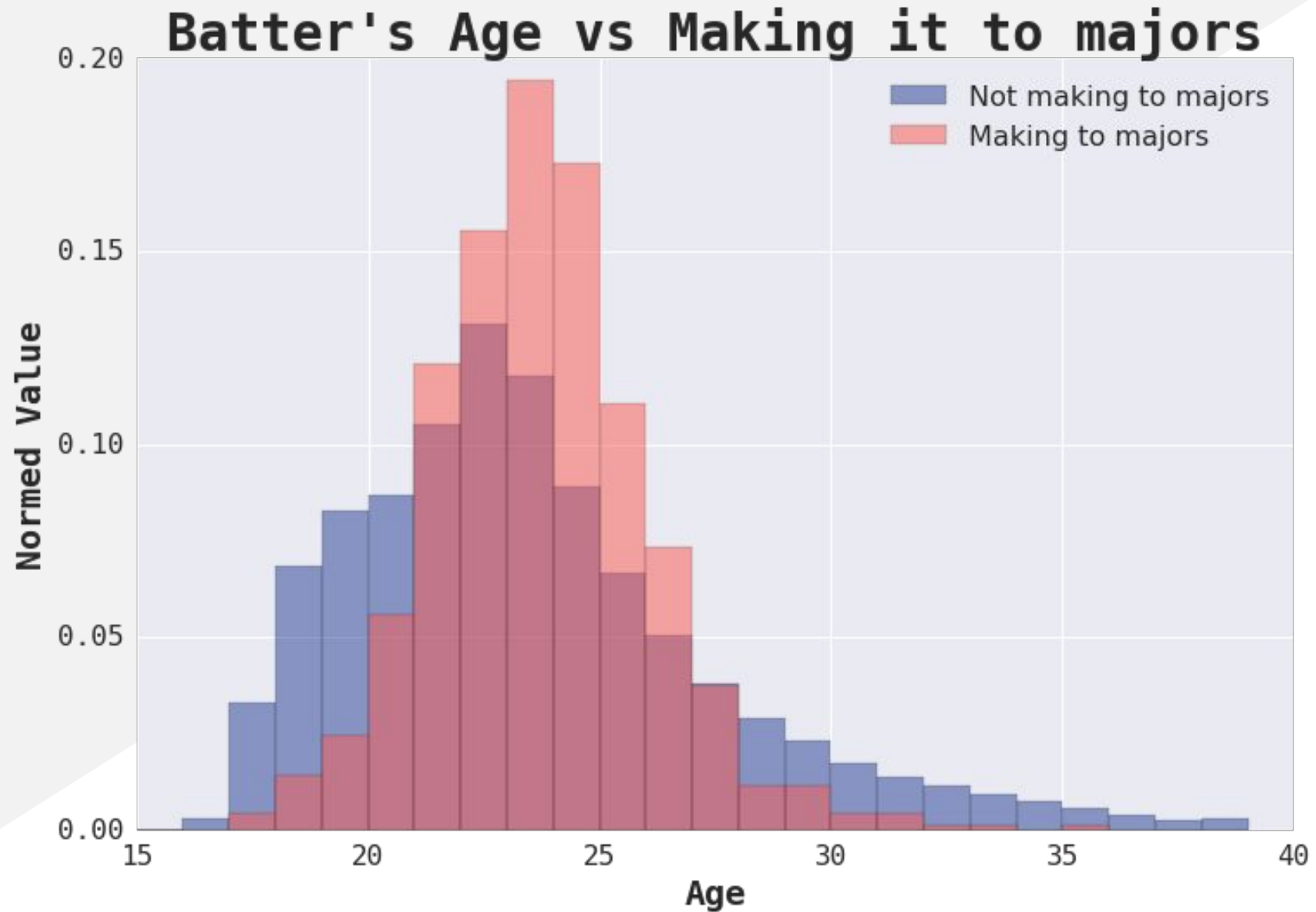
**Can I predict
a minor league player's chance
of making it to the majors next year
given the yearly stats?**

**What are
the important
features?**

Process



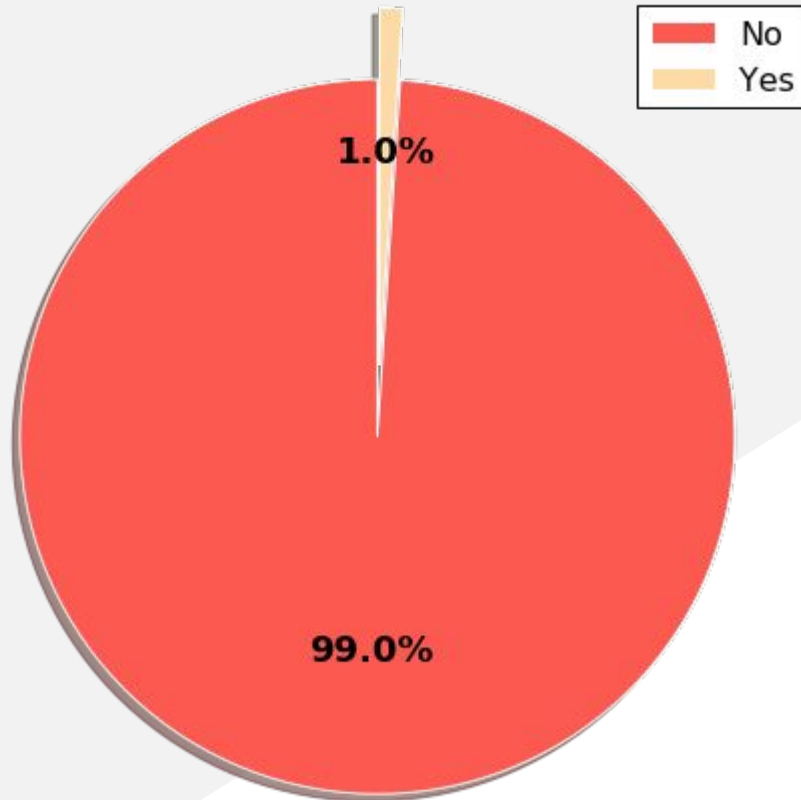
Data



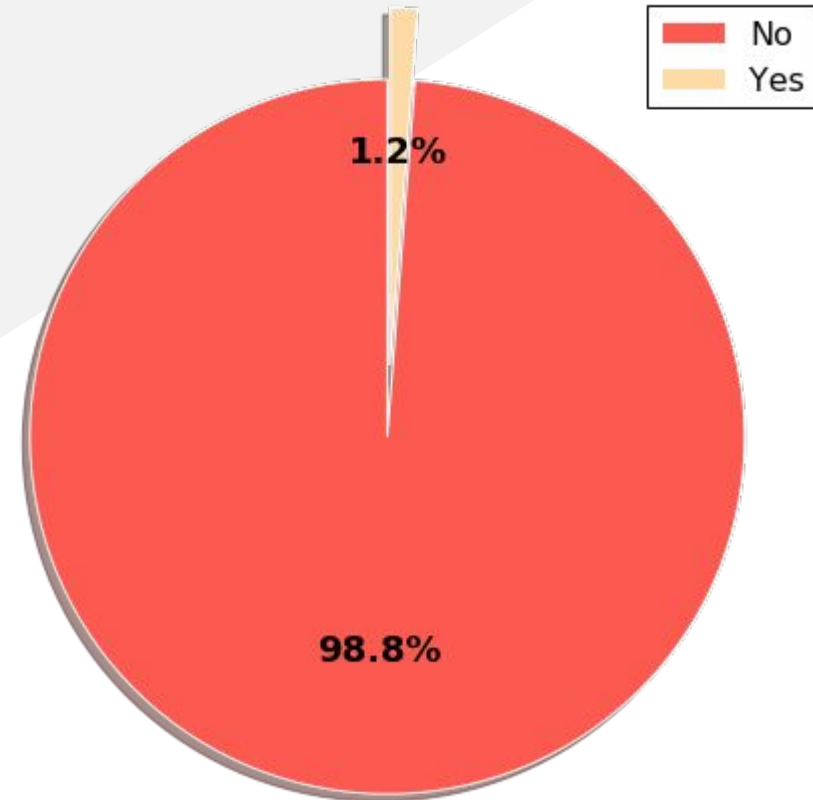
Data

**Imbalanced
Class!**

**Batters:
Promoted to majors next year?**



**Pitchers:
Promoted to majors next year?**



Modeling

Original Data (Batters); Logistic Regression

Will make it to majors next year?	Predicted Yes	Predicted No
Actual Yes	0 0.00 % (True Positive)	145 1.00 % (False Negative)
Actual No	0 0.00 % (False Positive)	13383 99.0 % (True Negative)

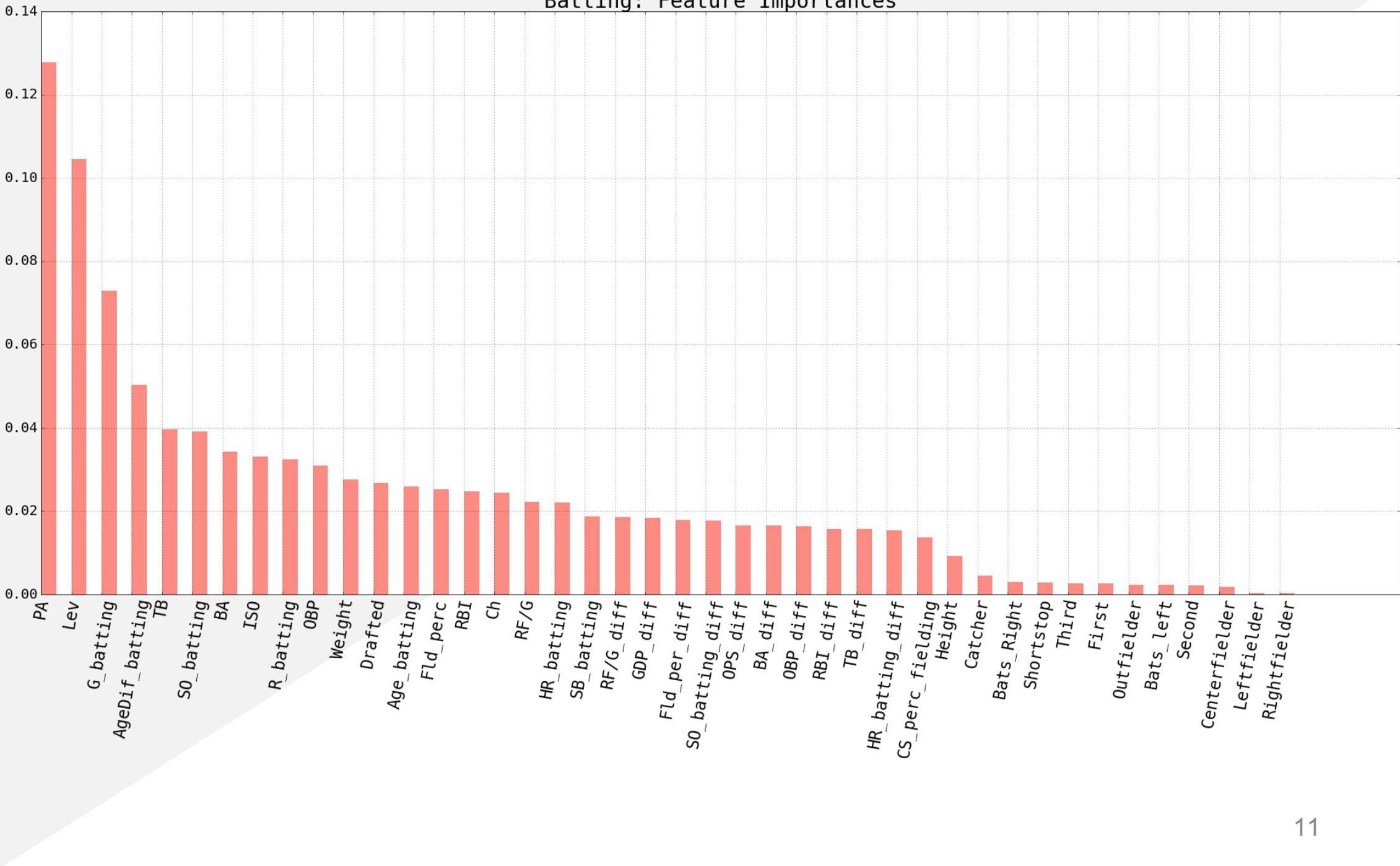
Modeling

Random Under-sampled Data (Batters; ratio = 0.5);

Logistic Regression

Will make it to majors next year?	Predicted Yes	Predicted No
Actual Yes	114 0.8 % (True Positive)	31 0.2 % (False Negative)
Actual No	1900 14 % (False Positive)	11483 85 % (True Negative)

Batting: Feature Importances



Next Steps

1.
Adjust the stats with league averages and PF (Park Factor)

2.
More feature engineering, narrowing down the features

3.
Refer to similar projects from other people, have more insights

4.
Go further, predict how WELL the player will perform in the majors

The Farm :
Who gets
harvested?

Jun Soo Kim

jun.kim8706@gmail.com

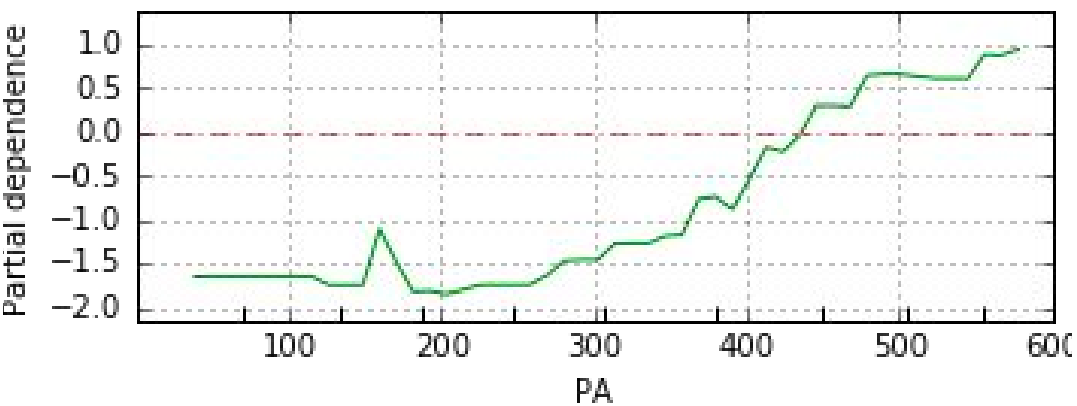
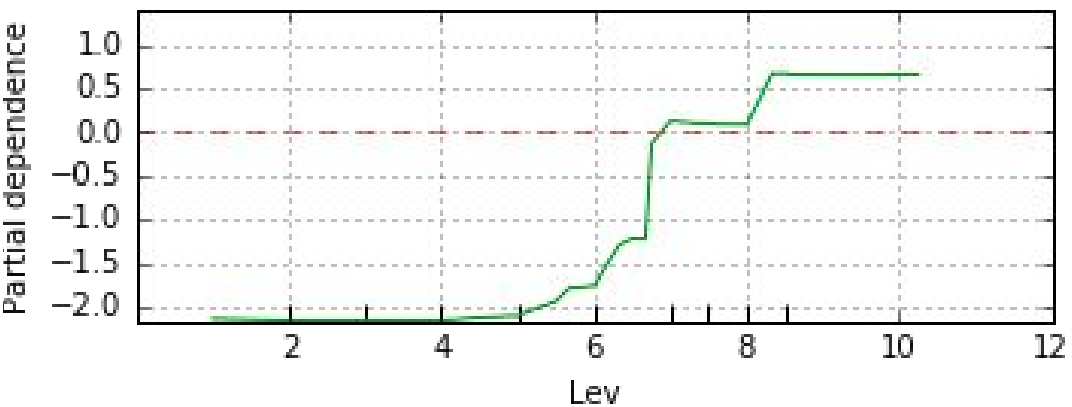
www.linkedin.com/in/jun-soo-kim

Thank you!

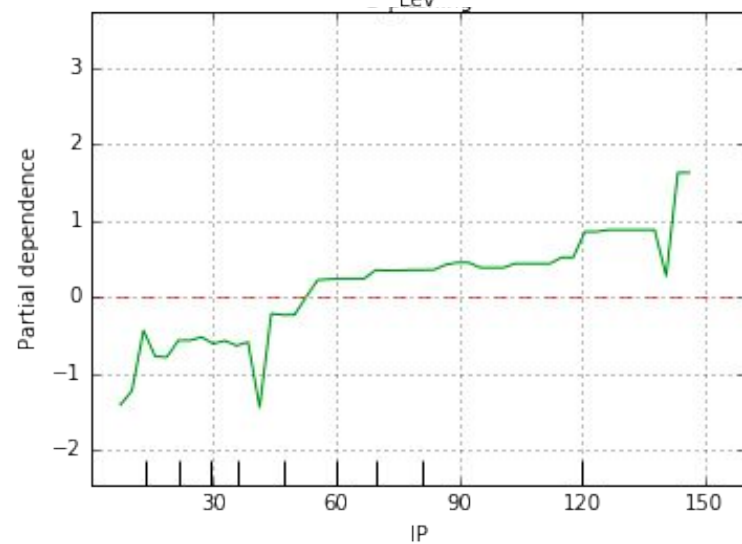
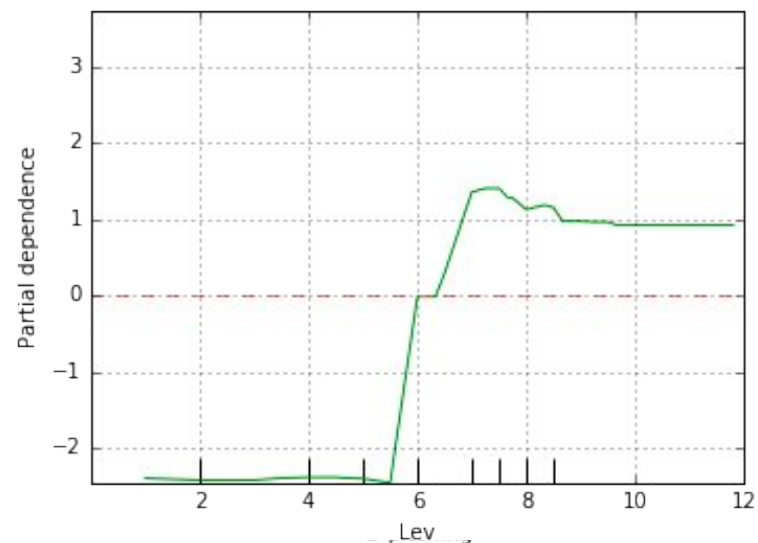
	coeff	pvals	conf_lower	conf_higher
AgeDif_batting	0.18	0.00	0.06	0.30
Age_batting	-0.41	0.00	-0.51	-0.31
BA	-2.73	0.34	-8.31	2.85
G_batting	0.01	0.14	0.00	0.02
Lev	0.76	0.00	0.60	0.91
OBP	3.77	0.09	-0.58	8.12
PA	0.00	0.00	0.00	0.01
R_batting	3.80	0.79	-23.51	31.11
SB_batting	7.22	0.04	0.40	14.04
SO_batting	-50.28	0.02	-92.66	-7.90
TB	5.78	0.00	2.10	9.46
Weight	0.01	0.01	0.00	0.02

	coeff	pvals	conf_lower	conf_higher
Lev	0.93	0.00	0.76	1.09
AgeDif_pitching	0.42	0.00	0.29	0.55
Age_pitching	-0.57	0.00	-0.68	-0.45
BF	-0.02	0.00	-0.04	-0.01
HR9	0.05	0.76	-0.28	0.38
IP	0.12	0.00	0.06	0.17
SO/W	-0.01	0.75	-0.08	0.05
SO9	0.13	0.00	0.08	0.18
WHIP	-0.75	0.01	-1.33	-0.18
Weight	0.01	0.00	0.01	0.02

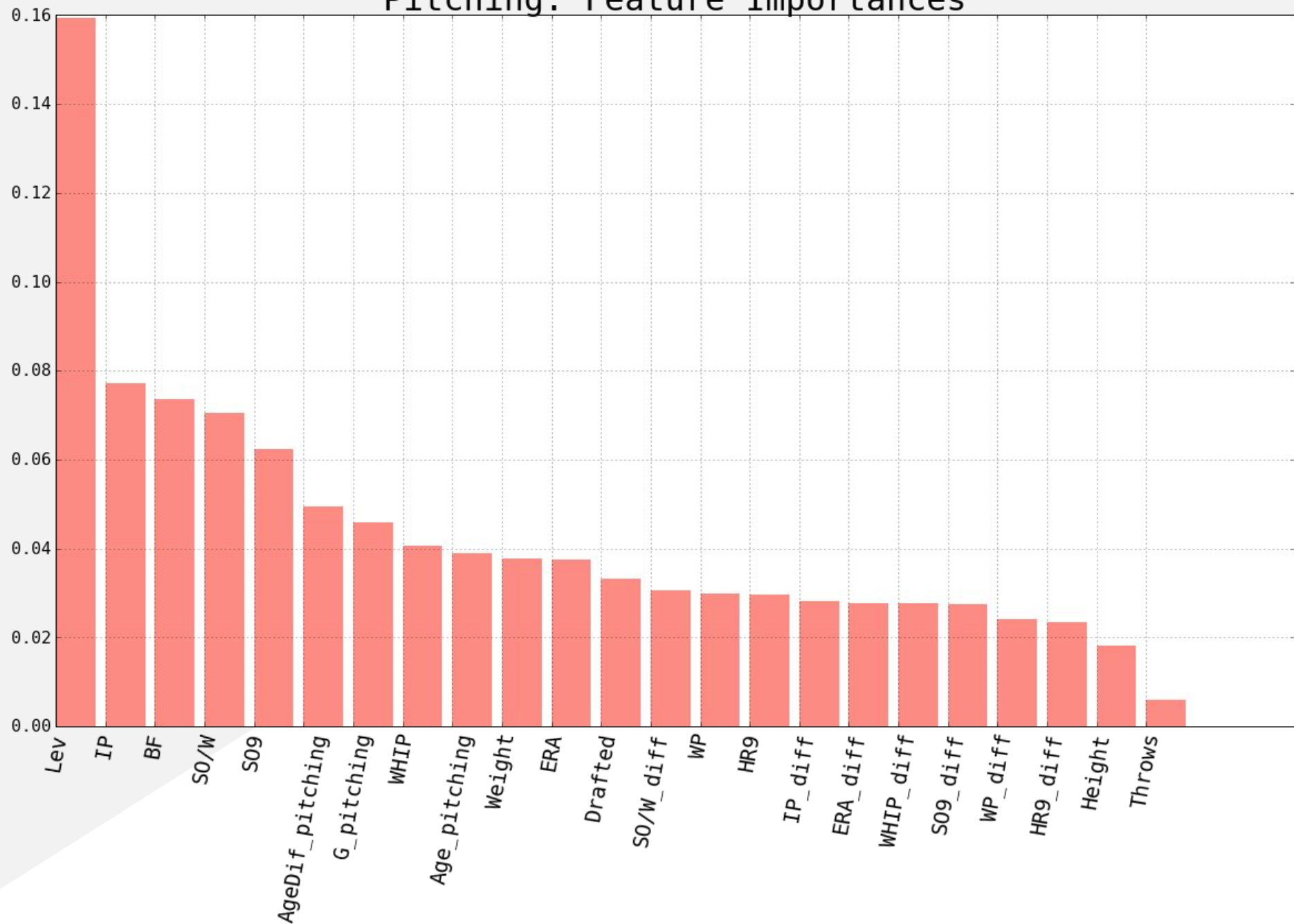
Batters



Pitchers



Pitching: Feature Importances



Data

No clear boundary!

