project luther

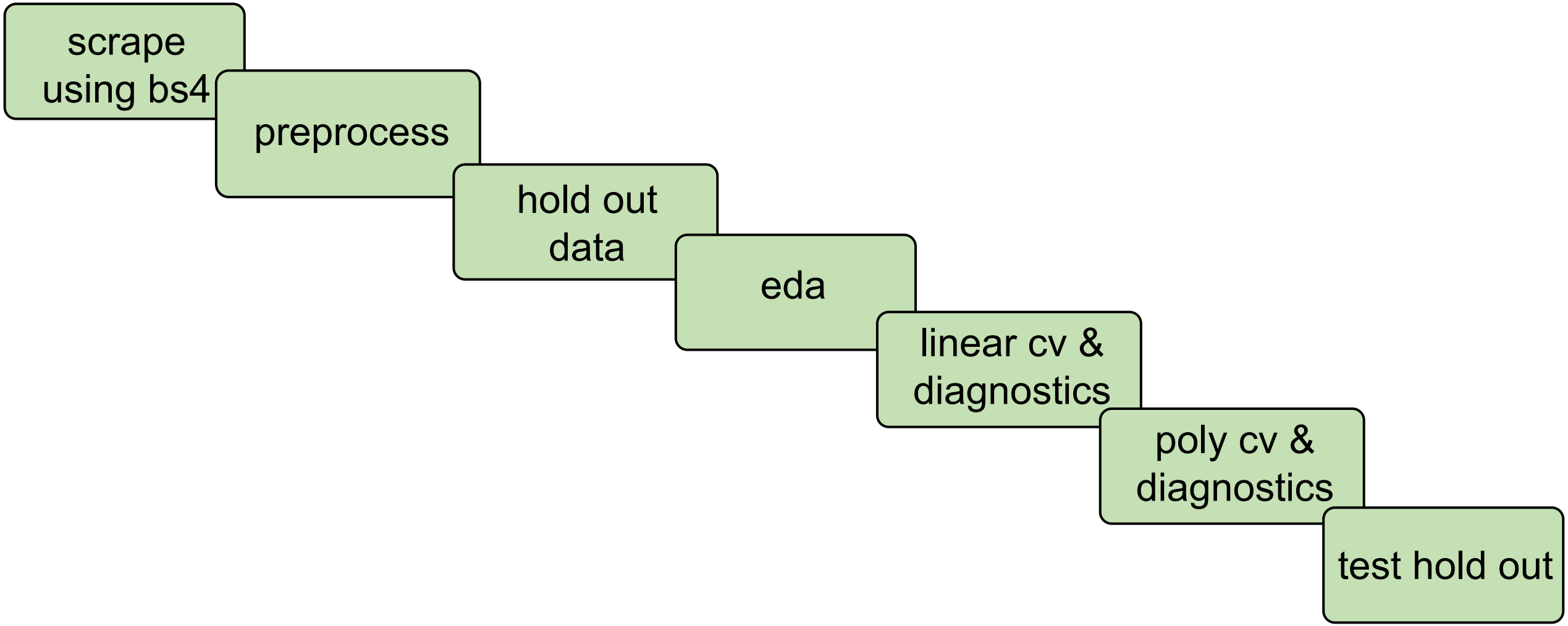
metis

andrea sorcinelli

july 20, 2018

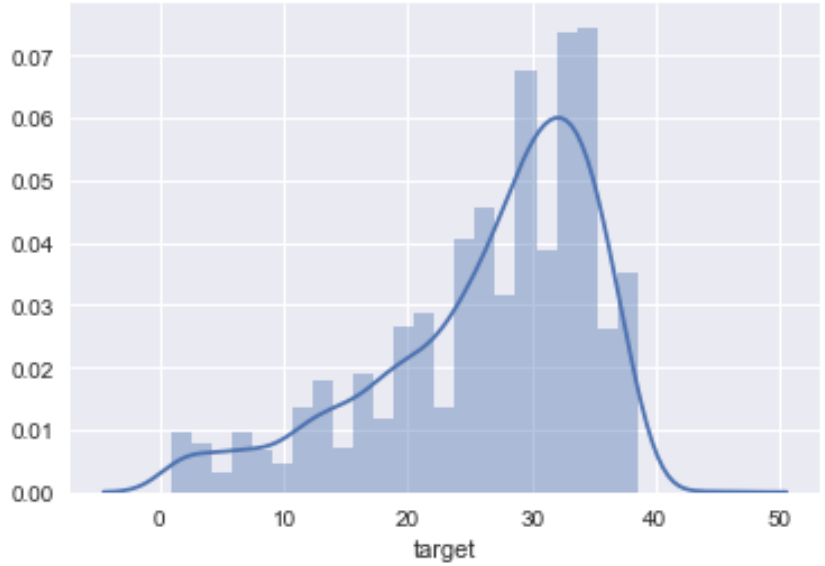
Cristiano Ronaldo leaves Real Madrid to sign with Juventus

approach



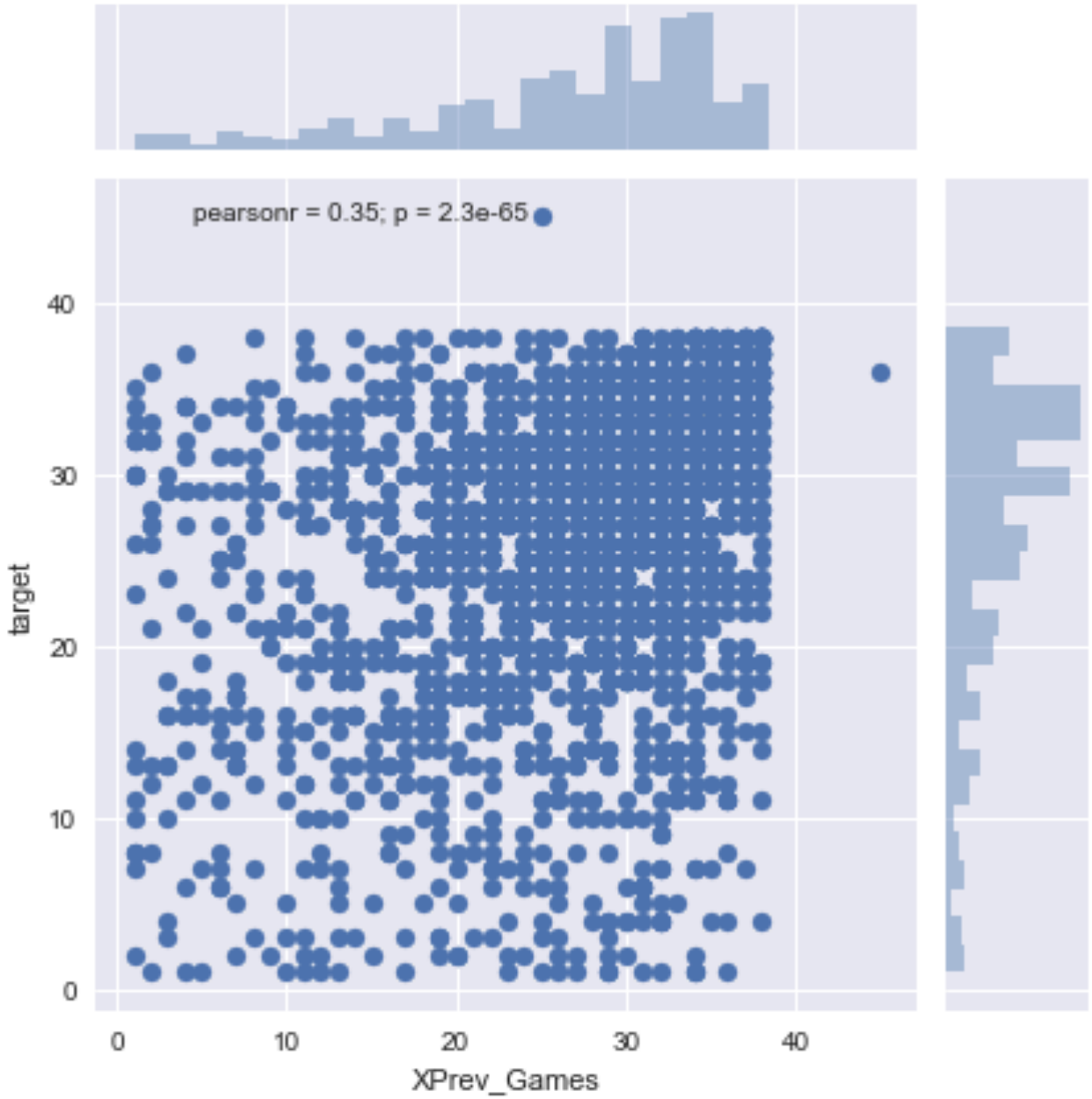
https://fbref.com/en/players/

exploratory data analysis



	XPrev_Age	XPrev_Games	XPrev_Min	XPrev_Goals	XPrev_Assists	XPrev_Fouls	XPrev_Red_Cards	target
XPrev_Age	1.000000	0.002318	0.024909	-0.090147	-0.088826	-0.153679	-0.030090	-0.171983
XPrev_Games	0.002318	1.000000	0.920993	0.291846	0.312853	0.457221	0.102006	0.349450
XPrev_Min	0.024909	0.920993	1.000000	0.243986	0.268719	0.421521	0.117597	0.353784
XPrev_Goals	-0.090147	0.291846	0.243986	1.000000	0.537613	0.293664	0.012042	0.097824
XPrev_Assists	-0.088826	0.312853	0.268719	0.537613	1.000000	0.285411	0.032485	0.134295
XPrev_Fouls	-0.153679	0.457221	0.421521	0.293664	0.285411	1.000000	0.209443	0.163076
XPrev_Red_Cards	-0.030090	0.102006	0.117597	0.012042	0.032485	0.209443	1.000000	0.067736
target	-0.171983	0.349450	0.353784	0.097824	0.134295	0.163076	0.067736	1.000000

Previous & current games



model parameters

- previous minutes
- previous games
- previous age
- previous fouls
- previous assists
- previous goals
- previous red cards

$R^2 = .129$

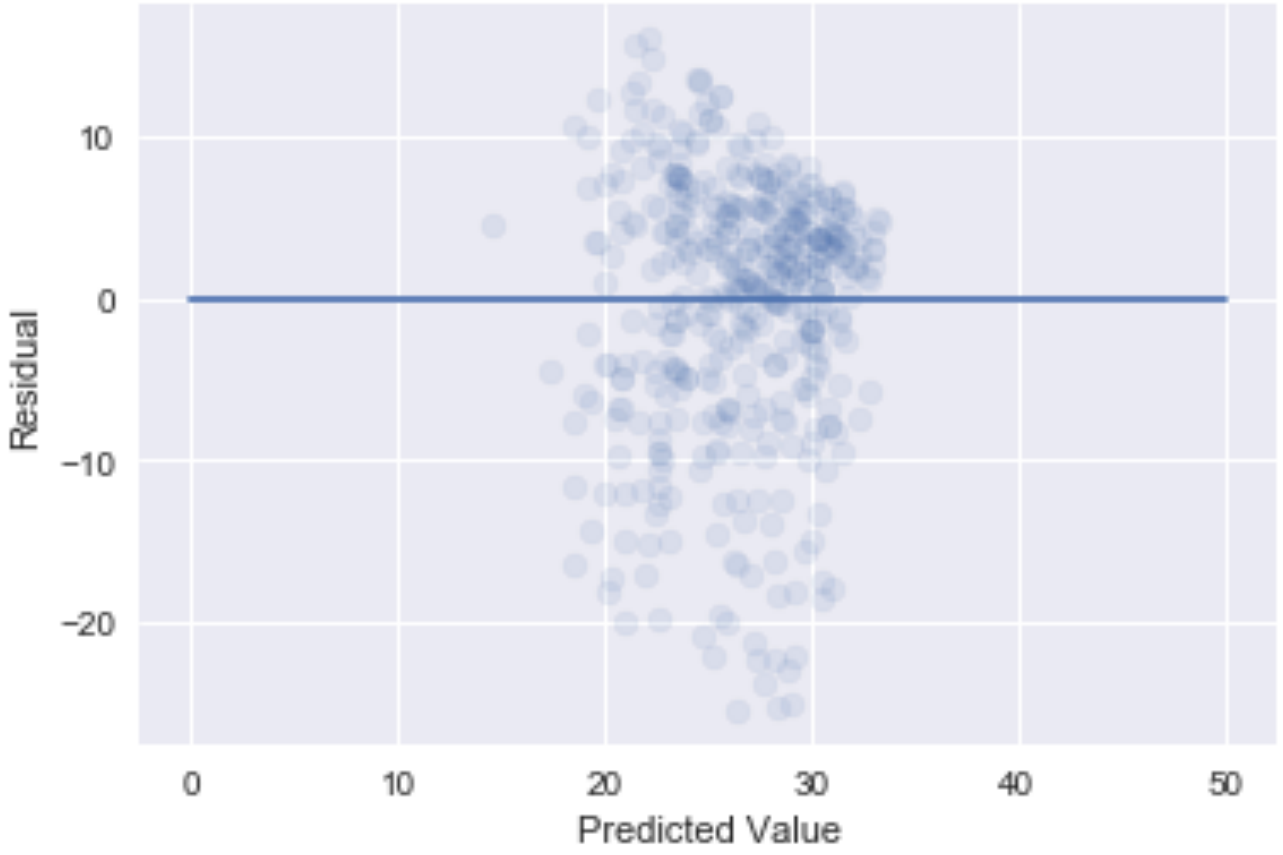
$R^2 = .160$

$R^2 = .163$

linear model with cross validation

R^2 stable around .16 but residuals aren't great

polynomial model with cross validation



R^2 improved to .18 and residuals look better

But polynomial did not perform better on final test set; R^2 for both = .12

conclusions

- age matters but not as much as how much players played previous season
- a lot of unexplained variance
- add new data?