NBA player salary prediction based on statistics

Ivan Luburić sw-13-2018

1. Motivation

Every basketball player wants to play in the best league(NBA). To get there, they need to upgrade their statistics in each field. According to their performance, they will get more or less money. Also players would like to know how much money do they deserve, to be prepared when signing new contract. There are also examples that player who plays excellent doesn't have big salary, so we want to find what does determine who gets paid the most.

2. Research questions

According to statistics of each player, this reseach tries to calculate salary that player deserves. In dataset we have two tables, one with salary and name of players in one season, and other one with players statistics in more seasons. Combining this tables we can see what determines salary amount. We have all informations about player: 3 point percentage, number of assistances, steals, blocks,...and many other.

3. Related work

Other researchers tried to predict salary according to NBA team payroll and team payroll as proportion of total NBA salary payroll, but I think that player own performance is crucial to their amount of salary, especially points per game, minutes played and assists.

4. Methodology

For this experiment, I will try various of tehnics for prediction. First and the simplest is linear regression with normal equasion, finding thetas as parameters of equasion. Before that I preprocessed data, replaced null value with 0 beacause it is mostly percentage that I excluded later, after that used label encoding for categorical values. After linear rigression I tried Ridge regression with changing alpha factor. Also I tried ElsticNet as combined Ridge and Lasso regression. And finally one non-parametric method kNN regression.

5. Discussion

For evaluation I used RMSE function. After testing all these methods I came with following results:

- Linear 4413151.49
- Ridge 4291190.77
- ElasticNet 4472134.09
- kNN for k=13 got best of 4838102.09

So we find that Ridge regression has best results after preproccesing and cutting features. But in all 4 methods we see that have big number of features but not too many examples, so we have big error in each one.

6. References

https://betterprogramming.pub/predicting-nba-player-salary-with-data-science-c5702caa3f2e
https://www.kaggle.com/code/rikdifos/nba-players-salary-prediction
https://www.nba.com/