NBA Salary Predictor

David Keyer

(Theme Pending)

Background

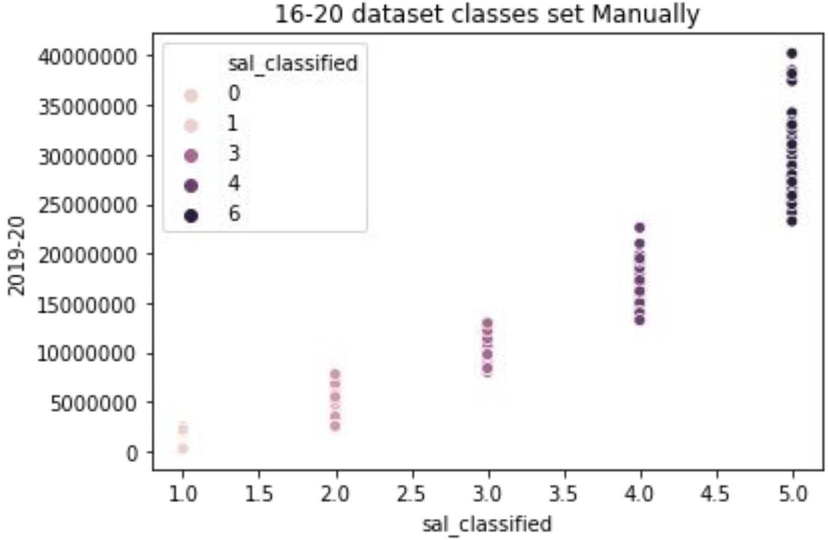
- NBA Contracts
  - For mainstays, very lucrative ⇒ almost $10,000,000 in 2016-17 among players in the league for five years
- TV Deal
  - Close to $25 Billion
  - Why is that relevant to player contracts
- As time goes on…
  - Bigger player contracts

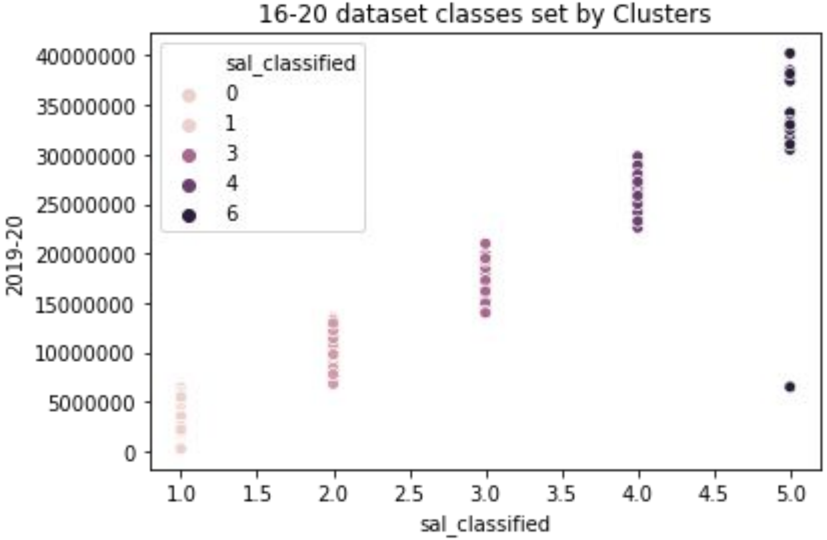
Approach & Process

- EDA
  - Dropping nulls
  - Exploring trends through visualizations
- Decisions
  - Merged statistics from five years with salary -- Why?
  - Worked primarily with three datasets -- Why?
- Originally started as a regression problem
  - Evolved into a classification problem

Modeling

- Datasets used
  - 2017, 2018, 2020
- Defining Classes
- Performance
  - Baseline of each model
  - Best model on each dataset
- Trends?
  - Are more predicted above or below their true class?
  - Reason for upwards trend?
- What can we expect going forward?


16-20 dataset classes set Manually
sal_classified
0
1
3
4
6
2019-20
40000000
35000000
30000000
25000000
20000000
15000000
10000000
5000000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
sal_classified


16-20 dataset classes set by Clusters
sal_classified
0
1
3
4
6
2019-20
40000000
35000000
30000000
25000000
20000000
15000000
10000000
5000000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
sal_classified

Strengths and Weaknesses

Strengths:

- Increasing predictive power -- 60% not bad/improves on baseline
- When misclassifying, real-world implications

Weaknesses:

- Amount of data
- Could improve on 60%

Recommendations/Conclusions

- Confusion matrices, more overpaid or underpaid?
- If someone in contract year, can use to see if they should get more or less than current salary
- Likewise, front office management can consider it before bringing in a player
- Expectation: As money increases, accuracy should rise, be able to better predict which class a player is in -- he will be getting more what he “deserves”

Next Steps

- See how model generalizes to the future
- Uncertainty -- pandemic
- Timeseries
- If salaries keep rising, will it be easier to predict salary class?