Predicting NBA player's salaries

Patrick Lomp
Artti Raasuke

UNIVERSITY OF TARTU
Institute of Computer Science

Introduction

The National Basketball Association is a men's professional basketball league in North America.
We found NBA to be intriguing because it has been growing as a league and so are their salaries. NBA has been growing quicker compared to other major leagues.

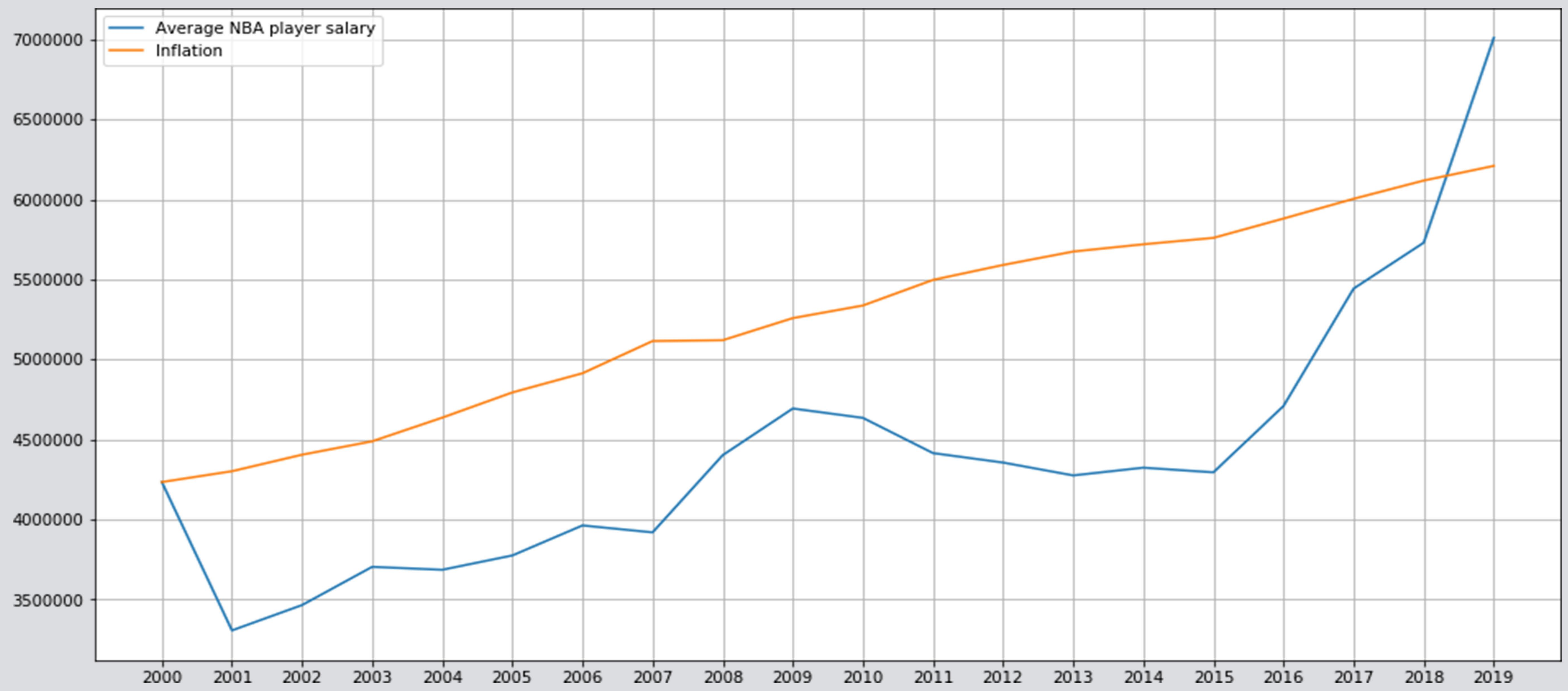
Aim

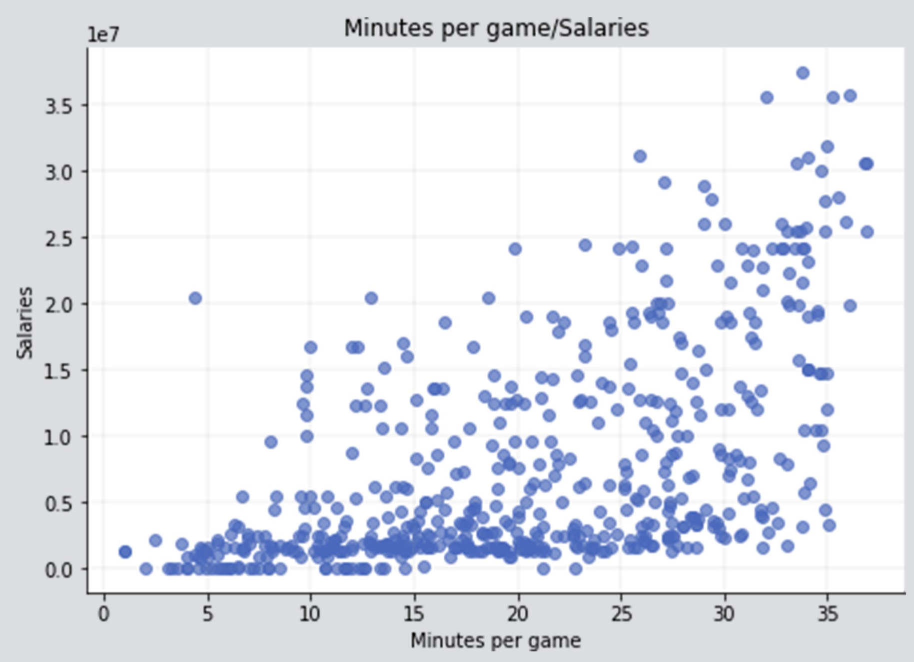
Give valuable insight how NBA salaries have been growing year-to-year.

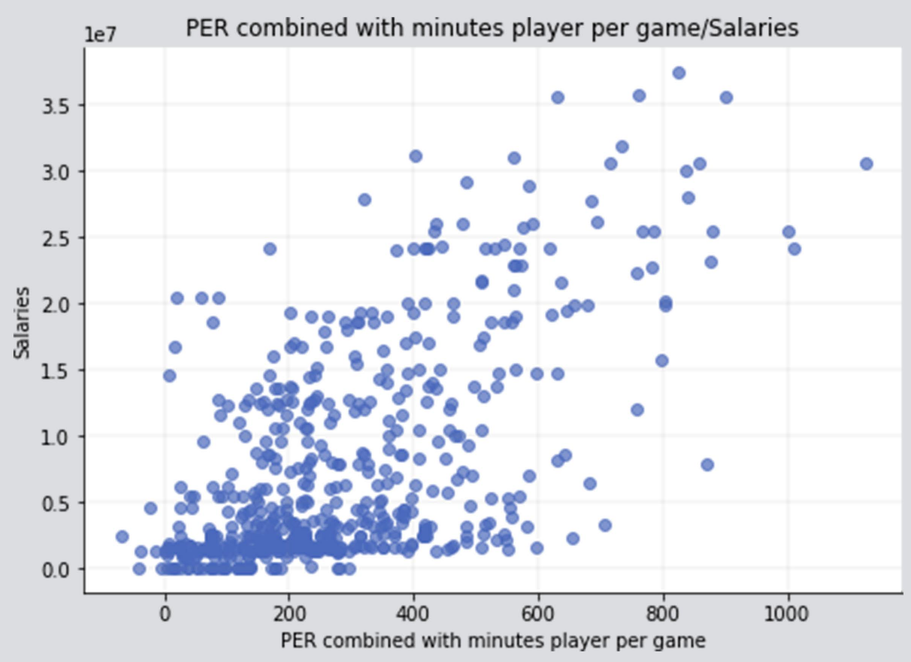
Make a model to predict player's earnings based on their statistics.

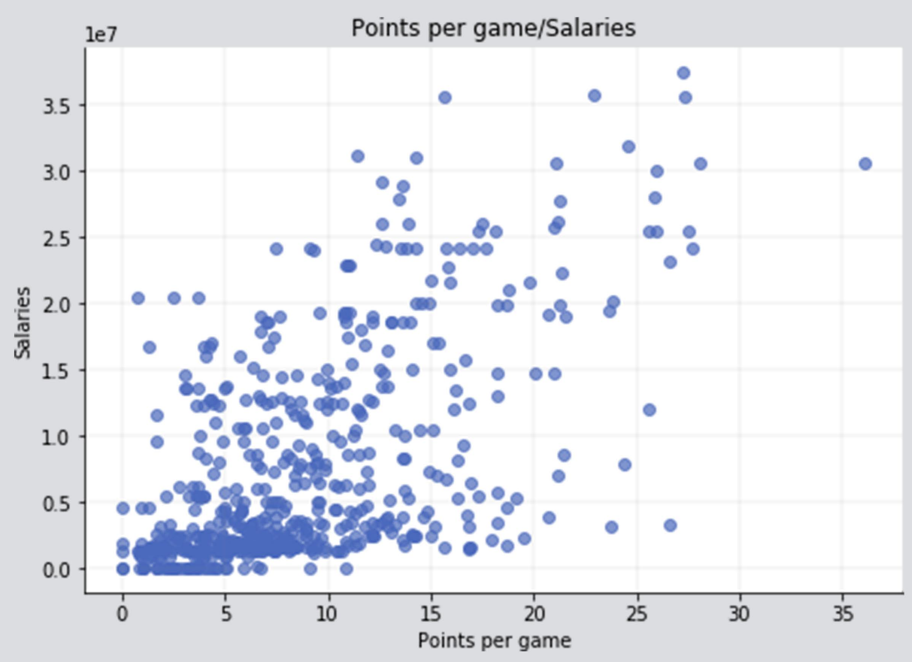
Data

One dataset with NBA player's salaries from 2000-2019.
Another dataset with performance statistics which contains variables such as: position, player efficiency rating, points, assists, rebounds. Dataset has over 40 variables.









Process

Most of our variables were numerical. For the features that were categorical we created dummy variables.

At first we predicted just salary, but because we used data from 2000-2019 we had to factor in inflation. We decided that the best approach would be to predict salary percentage of overall team cap instead.

Because the value we were predicting was continuous we also needed to use models that take in continuous values.

We tried different variations of random forest, linear regression, ridge regression and lasso method. We found lasso and ridge regression to be most accurate.

Ridge regression

Because we were predicting quantity value and we had less than 100K samples ridge regression seemed to be a fitting estimator for the job. We tried ridge regression with using different solvers such as: 'svd', 'cholesky', 'lsqr', 'sparse_cg'z. Using Ridge regression with with LSQR solver we got mean absolute error of ~10%.

Lasso method

Although Lasso and RidgeRegression are similar models,we found Lasso to be more accurate than ridge regression, which could be because Lasso gets rid of irrelevant features whereas ridge regression minimizes their impact.

Results & Conclusions

- We optained best results on our dataset with using Lasso and Ridge Regression. With Lasso we had mean accuracy error of ~2.5% and with ridge regression ~10%.
- Predicting player salary percentage of team salary cap yielded better results than just predicting salary.
- Using as many features as possible might not give the best results.
- Looking at different star players and comparing their salaries to what our models predicted Giannis Antetokounmpo seemed to be vastly underpaid whereas Chris Paul overpaid.

Patrick Lomp
patricklomp98@gmail.com

Artti Raasuke
artti.raasuke@gmail.com

TARTU ÜLIKOOL · UNIVERSITAS TARTUENSIS · 1632