

PREDICTING SUCCESS IN THE NBA

Group 14:

Cai Xin Qing

Yeo Ngee Chong

DATA PROBLEM

Data Problem Description



Problem

- When NBA teams draft new players, they have **many players to pick from**
- Difficult to predict player performance
- With data analysis, they will be able to **predict future success** of draftees



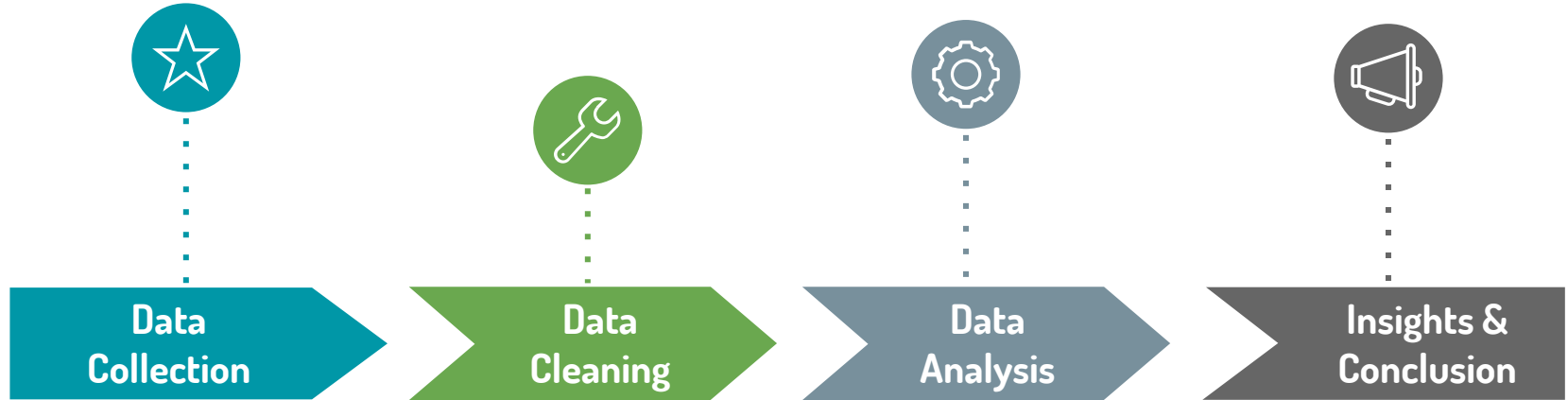
Our Motivation

- Both basketball fans
- Want to understand **what makes a successful basketball player**
- Be able to **predict** whether our favourite players will do well

Linear Regression Problem

SOLUTION

Process



- Beautiful Soup
- Basketball Reference - 1st and 2nd year NBA data
- Sports Reference - college data

DATA ACQUISITION

Data – Measures of Success



WS48
Win Shares
per 48
minutes



VORP
Value Over
Replacement
Player

PER
Player
Efficiency
Ratio



BPM
Box
Plus/Minus

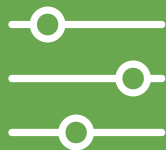


Data Cleaning



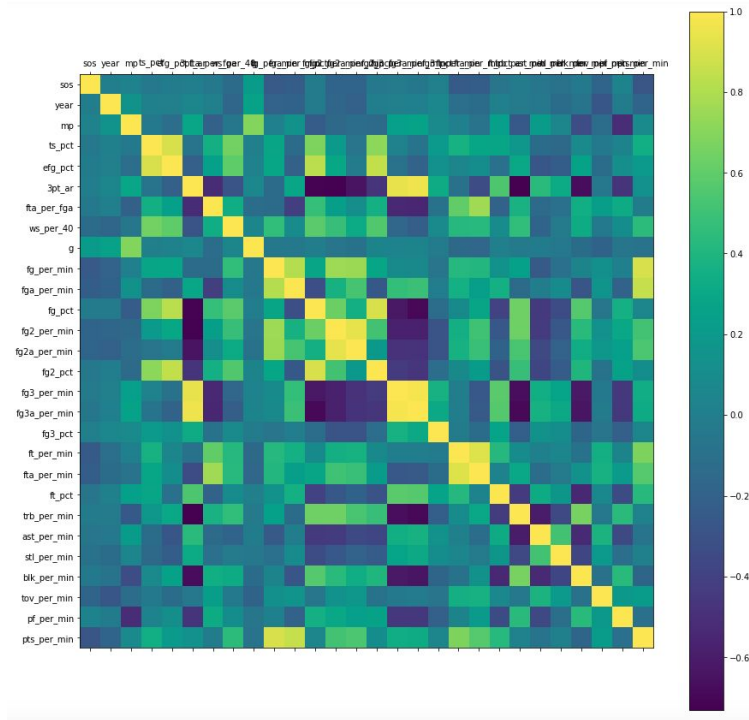
**Remove empty
data cells / rows**

Using Pandas



**Remove correlated
variables**

Reduce complexity,
improve speed



OUR ANALYSIS



Data Normalisation

- Z-Scoring
- SKLearn
- Improves ease of comparison



K Fold Cross Validation

- For estimating hyper-parameter

Using the StandardScaler preprocessing function

- Removes the mean and scaling to unit variance
- Allows comparison between data distributions

```
scaler = StandardScaler()  
comb = pd.DataFrame(scaler.fit_transform(comb))
```



Prediction Model



Data Normalisation

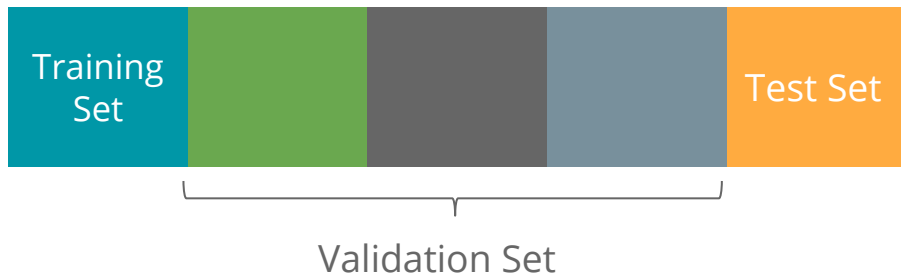
- Z-Scoring
- SKLearn
- Improves ease of comparison

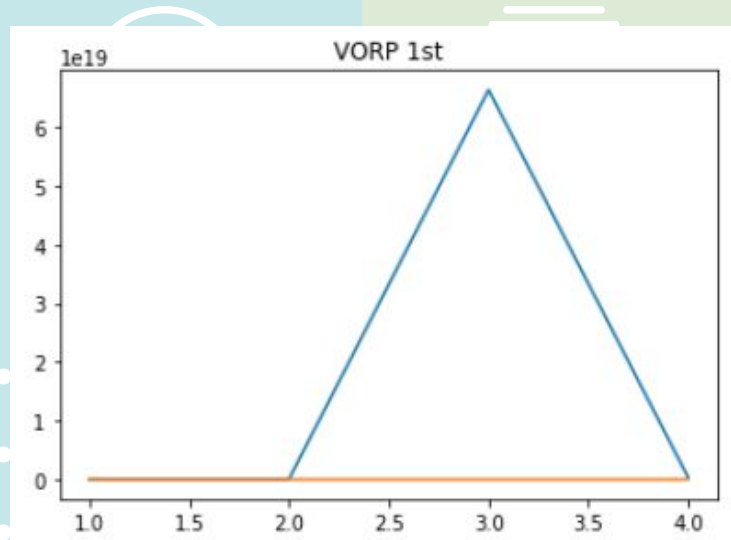


K Fold Cross Validation

- For estimating hyper-parameters

4-Fold Cross Validation





of comparison



Polynomial Degree (D)

- Find D which will minimise training data loss



Regularisation

- Balance variance and bias

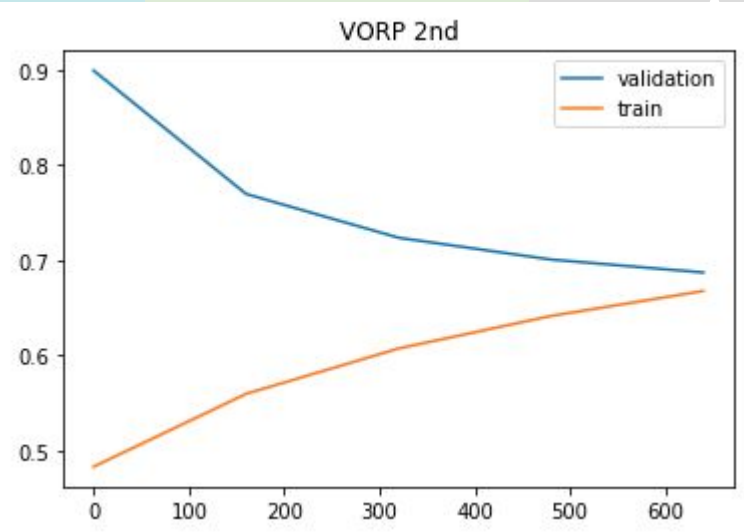


Prediction Model



Data Normalisation

- Z-Scoring
- SKLearn
- Improved
of comparison



Normal (D)

the mean
ation



Regularisation

- Balance
variance and
bias



Prediction Model



Data Normalisation

- Z-Scoring
- SKLearn
- Improves ease of comparison



K Fold Cross Validation

- For estimating hyper-parameters



Polynomial Degree (D)

- Minimise mean of evaluation error



Regularisation

- Balance variance and bias



Prediction Model

CONCLUSION

Prediction Model

- Using the hyperparameters that minimises mean of error between validation and training set
- Obtained model to **predict future performance** of new college players

Prediction Model



Win Share Results

1	Trae Young	0.04351
2	Mohamed Bamba	-0.01684
3	DeAndre Ayton	-0.03535
4	Marvin Bagley III	-0.03680

ISSUES ENCOUNTERED

Issues



Interaction effect from teammates

- **Teamwork** can improve gameplay
- Can analyse team performance in relation to players in team
- **Difficulty**: Different players have different playing time and have more/less opportunities to perform - skew data



Long runtime from scraper

- Parse the whole page to look for a link for every player
- **Inefficient**
- Could have used another data scraping framework, eg. Scrapy

Issues



Insufficient Data

- 400 data cells after data cleaning
- Statistics were not recorded before 1995, and there were missing data from some years



Inconsistent data

- Basketball rules have been constantly changing
 - Rules for handchecking and 3 point line were not implemented much longer ago

Our Contributions



Cai Xin Qing

- Presentation slides
- Data cleaning
- Data analysis



Yeo Ngee Chong

- Data Scraper
- Data cleaning
- Data analysis
- Prediction model

**THANK
YOU!**

References

- Icons8. (n.d.). [Person Icons]. *Icons8*. Retrieved from: <https://icons8.com/>
- Basketball Reference. (2018). *Basketball Reference*. Retrieved from: <https://www.basketball-reference.com/>
- Sports Reference. (2018). *Sports Reference*. Retrieved from: <https://www.sports-reference.com>