

DATA MINING



데이터 마이닝 프로젝트

NBA 선수 연봉 분류 예측



단국대학교 정보통계학과 최현석, 황성진

목차



STEP 1

주제, 데이터 소개 및 수집방법

STEP 2

분석 목표

STEP 3

기대 효과

DATA MINING



주제, 데이터 소개 및 수집방법



주제 소개



NBA는 샐러리 캡(Salary Cap) 제도로 팀별 연봉 총액을 제한하고 있으며, 이에 따라 선수 연봉은 경기 성과와 구단 전략을 반영하는 지표로 작용한다. 그렇기에 본 프로젝트에서는 NBA 경기 데이터를 활용해 연봉을 예측하고자 한다.

NBA_Player_Salaries(22-23).csv

Player Name	Team	Pos	Age	GP	GS	MP	FG	FG%	3P	3P%	FT	FT%	ORB	DRB	TBR	AST	STL	BLK	TOV	PF	Total Minutes	GP	GS	MP	FG%	3P%	FT%	ORB	DRB	TBR	AST	STL	BLK	TOV	PF	Total Minutes							
J. John Wall	ATLANTA Hawks	SU	34	94	82	31.4	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
R. Russel Westbrook	MILWAUKEE BUCKS	SU	34	94	82	32.2	9.1	49.0%	4.3	13.0%	8.7	93.1%	4.7	4.4	13.2	6.4	2.4	2.1	2.7	5.2	5.8	4.4	24	1.7	11.4	7.6	13.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	
A. L. Johnson	DETROIT PISTONS	SU	34	94	82	31.4	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
K. Kevin Durant	BROOKLYN NETS	PF	34	95	83	34.5	11.8	63.6%	1.8	9.4%	4.9	40.4%	8.3	84.6%	6.1	6.1	14.5	6.1	4.1	4.3	6.7	5.7	1.7	2.4	29.7	14.7	25.9	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
D. DeMar DeRozan	CHICAGO BULLS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
E. Kemba Walker	BOSTON CELTICS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
G. Giannis Antetokounmpo	MILWAUKEE BUCKS	SU	28	91	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
F. Kyrie Irving	BROOKLYN NETS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
H. J. Redick	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
I. Iman Shumpert	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
J. Kyle Korver	ATLANTA Hawks	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
K. Khris Middleton	MILWAUKEE BUCKS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
L. L. Batts	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
M. J. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
N. N. Horford	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
O. O. Babb	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
P. P. Anderson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
Q. Q. Parker	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
R. R. Hildebrand	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
S. S. Johnson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
T. T. McLean	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
U. U. Jackson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
V. V. Jackson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
W. W. Green	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
X. X. Johnson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
Y. Y. Jackson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
Z. Z. Jackson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
A. A. Anderson	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
B. B. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
C. C. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
D. D. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
E. E. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
F. F. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
G. G. Beal	DETROIT PISTONS	SU	34	94	82	32.2	9.1	49.0%	4.3	14.4%	8.7	91.4%	4.7	4.4	13.2	6.4	2.4	2.1	2.1	25.4	4.0	44.0	14.6	4.0	82.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4			
H. H. Beal	DETROIT PISTONS	SU	34	94	82																																						



데이터 소개

데이터이름 : NBA_Player_Salaries(2022-23)

출처 : Welsh, J. Kaggle

<https://www.kaggle.com/datasets/jamiewelsh2/nba-player-salaries-2022-23-season>

수집기관 : Basketball-reference(경기기록/통계), HoopsHype(연봉)

샘플크기 : 471개 관측치, 52개 변수

수집기간 : 2022~2023 Season



수집방법

본 프로젝트에서는 Kaggle에서의 제임스 월슨 사용자가 NBA 경기 기록, 통계, 연봉 등 제공하는 사이트에서 웹 크롤링을 통해 수집한 NBA 선수 연봉 및 통계 데이터를 활용하였다.

분석 목표



[분류기준]

NBA 선수의 연봉 분류 예측 모델 구축



*모델 종류: KNN, Naïve Bayes, Decision Tree 모델 + 교차검증(CV)

Percent Cap	구간	분류	선수 의미
<2%	1	C	최소 보장 계약, 2-way, 벤치 선수 등
2-4%	2	B	로테이션~스타팅급 선수
4-7%	3		
7-10%	4		
10-13%	5	A	팀의 에이스, 스타, 올스타급 선수
13-16%	6		
16-19%	7		
>19%	8		

22-23시즌 기준 셀러리 캡(Salary Cap) = \$ 123,655,000

[변수] - 선수/성과 지표

선수 정보: 선수 이름, 나이, 포지션, 연봉, 소속 팀 등 - 총 6개 변수

기본 경기 지표: 경기수, 출전 시간, 필드골, 자유투, 2점슛, 3점슛, 파울 등 - 총 22개 변수

슛 효율성 지표: 유효 슈팅 성공률, 자유투 시도 비율 등 - 총 4개 변수

고급 비율 스탯: 리バ운드 점유율, 어시스트 비율 등 - 총 10개 변수

고급 기여도 지표: 선수 효율성지수, 승리 기여도 등 - 총 10개 변수

Percent Cap	Cluster
< 2% (Min)	1
2-4% (Vet Min)	2
4-7% (MLE, Bi-annual)	3
7-10%	4
10-13%	5
13-16%	6
16-19%	7
> 19% (Maxes)	8

Table 2 Clustering bounds used to group data, with various exceptions listed

기대 효과



① 기대 효과

1 예측된 연봉 등급을 활용한 샐러리캡 운영 전략 수립

-> 고/중/저 연봉 그룹 간 균형을 고려해 예산 내 전력 구성 지원

② 기대 효과

2 저연봉 고효율 선수 발굴

-> 예측 결과보다 실제 연봉 등급이 낮은 선수 식별

③ 기대 효과

3 연봉 결정에 영향을 주는 핵심 지표 파악

-> 경기 성과 데이터로부터 연봉 수준에 영향을 주는 주요 변수 분석

