

2023 AI·DS 캡스톤

머신러닝을 활용한 미국프로야구의 투수와 타자 유형별 출루 및 아웃 예측 모델

안동수 . 김지민 . 손주찬



Contents

Chapter 1. 서론

Chapter 2. 모델 개발

Chapter 3. 결론 및 활용

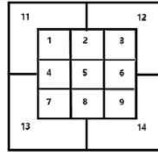
Chapter 1. 서론

서론

선행 연구의 한계

1. 경기의 복합적 상황을 반영 불가

항목	내용	속성(종류)	타입
예측결과 · 종속	event ▷ 투구결과 ▷ hit-into-play 된 타구 중 안타와 홈런만 측정	▷ 안타 : 1루타(single), 2루타(double), 3루타(triple) ▷ 홈런(homerun)	명목
	pitch_type	구종 ▷ 직구 : 패스트볼(4심), 싱커(2심), 커터, 컷패스트볼 ▷ 변화구 : 슬라이더, 커브, 너클커브, 슬로우커브, 너클, 이퍼스, 슬로브, 스위퍼 ▷ 오프스피드 : 스플릿핑거, 체인지업, 포크볼, 스크류볼	명목
예측을 위한 투입변수 · 독립	release_speed	투구속도 단위 : 마일(mph)	연속
	zone	투구존 1번~14번	명목
	stand	타자위치 좌완 / 우타	명목
	p_throws	투수주손 좌완 / 우완	명목
	Balls	볼카운트 0~3	이산
	strikes	스트라이크카운트 0~2	이산



투구 결과에 영향을 줄 수 있는 다양한 변수를 논외
 +
 예측 하려는 상황이 타격 상황에만 한정 되어 있어 단편적
 ↓
 작은 변수 하나만 달라져도 여러가지 상황이 나오는
 야구의 복잡성을 포착하지 못한 한계

2. 선수의 특성별 영향 간과

		구 질					전체	
		직구	커브	슬라이더	체인지업	싱커		
선수	S	빈도	479	241	0	187	5	912
		전체%	31.6%	1.8%	0%	7.1%	0.2%	34.5%
	P	빈도	834	47	725	123	0	1729
		전체%	31.6%	1.8%	27.5%	4.7%	0%	65.5%
전체	빈도	1313	288	725	310	5	2641	
	전체%	49.7%	10.9%	27.5%	11.7%	0.2%	100.0%	

투수 안에서도 선수의 스타일에 따라 다양한 유형이 존재
 +
 포지션별 선수의 유형에 따라 상성이 존재하기도 하고 공략법이 달라짐
 ↓
 기존 연구는 선수의 특성에서 발생하는 변수에 대해
 충분한 고려가 이루어 지지 않았다는 한계 존재

참고 문헌:

1. 조선미, 김주학, 강지연, 김상균. "머신러닝(XGBoost) 기반 미국프로야구(MLB)의 투구별 안타 홈런 예측 모델 개발." 한국체육측정평가학회지, vol. 25, no. 1, 2023, pp. 65-76.
2. 손혁. "프로야구 투수유형과 구질과의 관계." 국내석사학위논문 고려대학교 교육대학원, 2004. 서울
3. 황수웅. 불확실성(uncertainty)을 고려한 스포츠 빅데이터 분석: Bayesian 추정과 Deep Learning을 활용한 프로야구 심판의 Ball/Strike 판정 평가 모델 개발. 서울대학교 대학원. 2023

서론

연구 차별점

다양해진
입력 변수

예측 상황
다각화

선수 유형별
발생 요인 반영

기존 변수

신규 변수

변수	설명	측정 방법	단위
event	hit, strikeout, walk, error, double, triple, home run, etc.	이벤트 코드	개
pitch_type	fastball, curveball, slider, etc.	피치 유형	종류
release_speed	구속 속도	마일 퍼 HOUR	mph
spin	구속 회전	회전 수	rpm
release_height	구속 높이	피트	ft
release_angle	구속 각도	도	deg
release_velocity	구속 속도	마일 퍼 HOUR	mph
release_spin_rate	구속 회전	회전 수	rpm
release_spin_direction	구속 회전 방향	도	deg

+ 공의 상하 움직임
투구 폼
투수 익스텐션
투수 릴리스 포인트
주자 여부
점수, 이닝
아웃카운트

투구 결과에 영향을 줄 수 있는 요인을
기존 연구 보다 다각적으로 고려

기존 연구

본 연구

타격 결과 (안타, 홈런)
OR
투구 결과 (스트라이크, 볼)

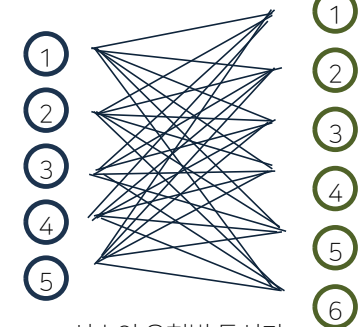
비교할 수 있는 결과가 아닌
같은 상황 안에서 경우의 수를 예측

출루
(안타, 홈런, 볼넷)
+
아웃
(삼진, 뜬공, 병살...)

야구에서 발생할 수 있는
다양한 상황을 포함한 예측

타자 유형

투수 유형



선수의 유형별 특성과
이로 인해 발생할 수 있는
상성에 따른 변수도 고려

서론

연구 필요성

출루/아웃 지표 파악의 중요성

Base Runners			2010-2015			1993-2009			1969-1992			1950-1968		
1B	2B	3B	0 outs	1 outs	2 outs	0 outs	1 outs	2 outs	0 outs	1 outs	2 outs	0 outs	1 outs	2 outs
—	—	—	0.481	0.254	0.098	0.547	0.293	0.113	0.477	0.252	0.094	0.476	0.256	0.098
1B	—	—	0.859	0.509	0.224	0.944	0.565	0.245	0.853	0.504	0.216	0.837	0.507	0.216
—	2B	—	1.100	0.664	0.319	1.175	0.723	0.349	1.102	0.678	0.325	1.094	0.680	0.330
1B	2B	—	1.437	0.884	0.429	1.562	0.966	0.471	1.476	0.902	0.435	1.472	0.927	0.441
—	—	3B	1.350	0.950	0.353	1.442	0.991	0.388	1.340	0.943	0.373	1.342	0.926	0.378
1B	—	3B	1.784	1.130	0.478	1.854	1.216	0.533	1.715	1.149	0.484	1.696	1.151	0.504
—	2B	3B	1.964	1.376	0.580	2.053	1.449	0.626	1.967	1.380	0.594	1.977	1.385	0.620
1B	2B	3B	2.292	1.541	0.752	2.390	1.635	0.815	2.343	1.545	0.752	2.315	1.540	0.747

출처: <http://www.tangotiger.net/re24.html>

- 해당 주루/아웃 상태에서 해당 이닝이 끝날 때까지 득점할 확률

동일 아웃 상황에서도 주자 유무에 따라 득점 확률의 차이

→ 아웃과 출루의 지표를 파악 = 승패에 직결되는 요소

데이터에 기반한 전략 수립



출처: <https://www.gqkorea.co.kr/2020/03/05/>

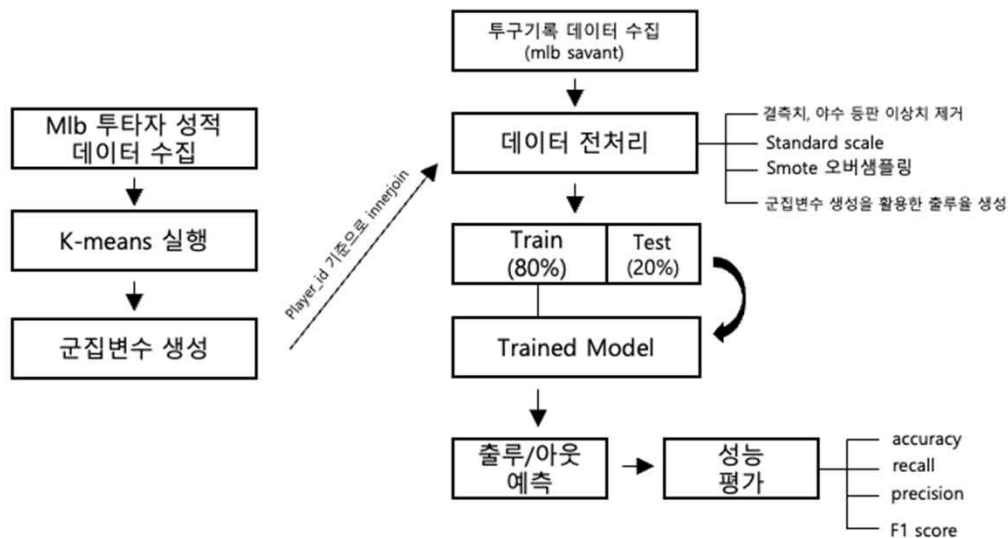
선수의 유형과 유형별 상성을 알고 있으면

타자 배치, 투수 교체 전략 등

전략 수립에 합리적인 결정 가능

서론

모델 개발 계획



Step 1. 데이터 수집

Step 2. 데이터 전처리

Step 3. K-means로 새로운 독립변수 생성

Step 4. 데이터 병합

Step 5. 변수 스케일링 & 오버샘플링

Step 6. 출루율 변수 생성

Step 7. Train set & Test set 생성

Step 8. 모델 학습

Chapter 2. 모델 개발

모델 개발

Step 1. 데이터 수집

메이저리그 공식 기록 홈페이지 "MLB Savant"에서 투구 관련 데이터를 수집



Search Results				
Rk.	Player	Pitches	Total	Pitch %
1	Cole, Gerrit RHP	3281	3281	100.0
2	Cease, Dylan RHP	3262	3262	100.0
3	Gallen, Zac RHP	3248	3248	100.0
4	Castillo, Luis RHP	3207	3207	100.0
5	Mikolas, Miles RHP	3197	3197	100.0
6	Giolito, Lucas RHP	3190	3190	100.0
7	Webb, Logan RHP	3182	3182	100.0
8	Snell, Blake LHP	3168	3168	100.0
9	Lynn, Lance RHP	3167	3167	100.0
10	Wheeler, Zack RHP	3155	3155	100.0
11	Bassitt, Chris RHP	3139	3139	100.0
12	Keller, Mitch RHP	3119	3119	100.0
13	Strider, Spencer RHP	3100	3100	100.0
14	Nola, Aaron RHP	3087	3087	100.0
15	Burnes, Corbin RHP	3079	3079	100.0

출처: <https://baseballsavant.mlb.com/>

수집된 데이터 목록 (2020 ~ 2022) → 총 421879행 데이터 수집

1. 종속 변수

- 출루, 아웃

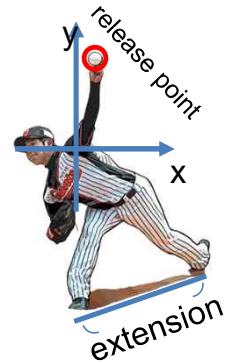
2. 독립변수

- 스트라이크, 볼

- 구속, 구종, 공의 상하 움직임

- 투구 폼, 투수의 익스텐션, 투수의 릴리스 포인트

- 투구 당시 경기 상황 (주자 여부, 점수, 이닝, 아웃카운트)



참고문헌:

SmartPitch: Applied Machine Learning for Professional Baseball Pitching Strategy. (Otremba Jr, 2022)

모델 개발

Step 2. 데이터 전처리

1). 이상치 제거

정상적인 포지션 간(투수vs 타자)의 출루 예측을 하고자 함
→ 타자가 투수로 등판한 경우를 이상치로 판단하고 제외

- 타자 마운드 등판 판단 기준.
 - 2022 년 MLB 투수 최저 구속 69.9 마일 미만
 - 2021 년 MLB 투수 최저 구속 64.7마일 미만
 - 2020 년 MLB 투수 최저 구속 65마일 미만
- 이상 위 공들은 타자가 던진 것으로 판단
→ 연도별 투수 최저 구속 미만의 투구를 이상치로 간주해 제거

2). 결측치 제거

변수들의 결측치 비율이 모두 0.3% 이하
→ 모델에 영향이 없을 것이라 판단 모든 결측치 제거

```

• release_pos_x    0.013823
• release_pos_z    0.013823
• pfx_x            0.001728
• pfx_z            0.000494
• release_spin_rate 0.328060
• release_extension 0.140209
• release_pos_y    0.013823
• spin_axis        0.328060
    
```

결측치 비율 (%)

3). 범주형 데이터 전처리

범주형 데이터를 숫자로 매핑

- 투수 주손: 우완/좌완 → 1 / 2
- 타자 주손: 우타/좌타 → 1 / 2
- 구종: 직구계열 → 1, 변화구 계열 → 2, 오프스피드 → 3
- 출루: 안타, 홈런, 볼넷, 몸에 맞은 공 → 1
- 아웃: 병살타, 삼중살, 뜬공, 땅볼 → 0

Pitch_type	Release_speed	On_1b	...	release_pos_x	release_pos_y	events
3	90.0	0	...	-2.20	6.49	1
1	95.2	0	...	-2.33	6.41	0
1	92.2	1	...	-2.13	6.48	0
3	83.7	0	...	-2.28	6.33	0

독립 변수

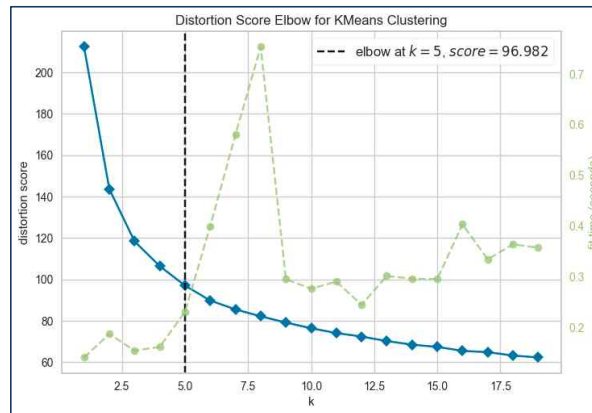
종속 변수

모델 개발

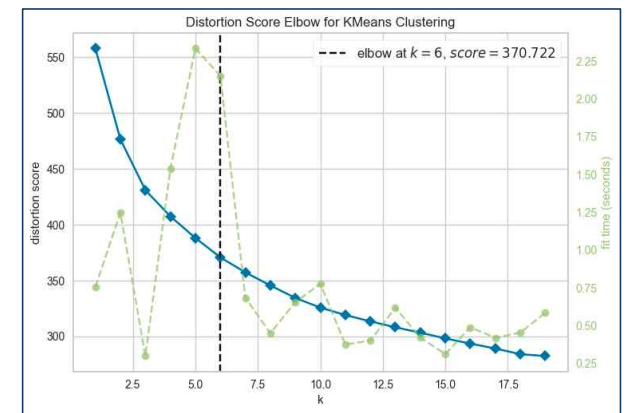
Step 3. 선수별 특성 반영을 위해 K-means 로 새로운 변수 생성

1). 선수의 성적으로 군집화 실시

1. 2020 ~ 2022년 투수와 타자의 성적 평균 데이터 사용
 - 타자: 삼진률, 볼넷 출루율, 홈런, ops, 타구의 질
 - 투수: 직구와 변화구 구속, 피안타율, 방어율, 타구 허용속도, 구종 구사율
2. "Elbow Method"를 이용한 군집 개수 선정
 - 타자 5개
 - 투수 6개
3. 군집화 실시



타자 군집 개수 선정



투수 군집 개수 선정

참고문헌:

- 1). 한국프로야구에서 타자능력의 측정 (이장택, 2014)
- 2). Investigating Major League Baseball Pitchers and Quality of Contact through Cluster Analysis"(Marcou,2020)

모델 개발

Step 3. 선수별 특성 반영을 위해 K-means 로 새로운 변수 생성

2). 군집별 특성 확인

• 타자 군집별 특성 (평균)

군집	키워드	대표 선수	홈런	삼진율	볼넷율	타율	출루율	장타율	OPS	Babip	Pull Percent	Line drive percent
1번	당겨치기, 낮은 타율, 높은 볼넷 출루율	최지만	7.06	27.9%	9.1%	0.220	0.300	0.398	0.698	0.276	42.49%	23.2%
2번	전반적 낮은 수치	쓰쓰고 요시토모	2.46	26.3%	7.5%	0.195	0.264	0.291	0.556	0.255	37.22%	22.0%
3번	낮은 삼진율 높은 라인드라이브 타율	김하성	6.66	20.3%	7.4%	0.255	0.318	0.386	0.704	0.306	38.71%	25.0%
4번	높은 타율, 타점, 홈런 핵심 타자	오타니 쇼헤이	18.6	22.3%	9.6%	0.258	0.337	0.467	0.804	0.297	39.29%	23.98%
5번	메이저리그 적응 실패, 매우 낮은 성적	린즈웨이	0.148	45.27%	3.4%	0.114	0.149	0.146	0.296	0.235	23.82%	15.7%

• 투수 군집별 특성 (평균)

군집	키워드	대표 선수	직구 구속	변화구 구속	삼진율	볼넷 허용률	방어율	허용타구 속도	포심 구사율	싱커 구사율	슬라이더 구사율	직구 구사율	땅볼유도
1번	싱커 구사, 땅볼 유도형	펠릭스 페냐	93	83	20%	8.9%	4.36	88.2	12.6	45.8	24.5	59.3	51.4
2번	빠른 포심, 탈삼진 능력	맥스 슈어저	94.3	84.2	24.6%	9.6%	4.15	88.8	51.3	9.1	22.6	56.6	39.9
3번	낮은 구속, 정교한 제구력	류현진	90.3	78.8	21.1%	8.2%	4.38	88.1	26.2	20.9	9.9	69.8	42.8
4번	메이저리그 경력 적지만 방어율 높음	조던 아모토	91.9	81.3	18%	9.9%	7.51	90.3	43.4	14.1	21.5	56.7	36.3
5번	마무리 투수, 높은 삼진율	에드윈 디아즈	93.7	84.2	25.5%	9.5%	3.95	88.3	34.6	13.1	46.5	43.7	43.2
6번	다양한 구종, 선발급 능력	오타니 쇼헤이	92.9	80.4	24.4%	9.1%	4.13	88.5	45.1	11.6	14.8	55.5	41.7

모델 개발

Step 3. 선수별 특성 반영을 위해 K-means 로 새로운 변수 생성

2). 군집별 특성 확인

• 타자 군집별 특성



선구안이 좋은 타자그룹

- 낮은 타율(0.220)
- 높은 볼넷율(9.1%)
- 높은 출루율(0.300)

볼을 잘 골라내는 능력을 가졌고
출루율이 높은 타자 그룹

Ex). 최지만 선수



컨택트형 타자 그룹

- 평균 이상의 타율(0.255)
- 낮은 삼진율 (20.3%)
- 높은 라인드라이브 타구율 (25%)

볼 컨택이 좋기 때문에
좋은 타구질을 만들어내는 타자 그룹

Ex) 김하성 선수



홈런형 파워타자 그룹

- 높은 타율 (0.258)
- 높은 타점 (58)
- 높은 홈런 수 (59)

팀내 중심타선을 담당할 정도의
파워를 가진 4번타자형 타입

Ex) 오타니 쇼헤이 선수

모델 개발

Step 3. 선수별 특성 반영을 위해 K-means 로 새로운 변수 생성

2). 군집별 특성 확인

• 투수 군집별 특성



땅볼 유도형 투수그룹

- 메이저리그 평균 구속(93마일)
- 땅볼 유도 구종인 싱커의 높은 구사율 (45.8%)
- 높은 땅볼유도 비율 (51.4%)

→ 싱커를 주로 구사하는 땅볼유도형

Ex) "땅볼 유도형" 펠릭스 페냐 선수



직구 위주 삼진형 투수그룹

- 높은 포심 구사율 (51%)
- 높은 삼진율 (24.6%)
- 가장 빠른 직구구속(94.3마일)

→ 직구 구위로 승부하는 투수
Ex) "메이저리그 최고의 포심으로 선정" 맥스 슈어저



제구형 투수그룹

- 낮은 직구속도(90마일)
- 낮은 변화구 속도(78마일)
- 낮은 볼넷 허용률(8.2%)

→ 변화구와 직구 사이에 큰 속도편차를 활용하고 정교한 제구력으로 승부
Ex) "칼제구" 류현진 선수



직구, 슬라이더 등 구위가 좋은 삼진형 투수그룹

- 적은 평균 이닝 소화(39.5이닝)
- 가장 높은 삼진율 (25.5%),
- 가장 낮은 방어율 (3.95),
- 낮은 빠른 타구 허용률 (36.8%)
- 직구 구사율 (43%)
- 슬라이더 구사율(46.5%)

→ 직구, 슬라이더를 주로 구사하는 마무리 투수 타입

Ex) "철벽마무리" 에드윈 디아즈



다양한 구종을 구사하는 에이스 선발투수그룹

- 낮은 방어율(4.13)
- 많은 평균이닝(63이닝)
- 삼진율 24%
- 모든 구종 구사비율 최소 10%이상

→ 다양한 구종을 던지는 에이스 선발 투수

Ex) "MVP" 오타니 쇼헤이

모델 개발

Step 3. 선수별 특성 반영을 위해 K-means 로 새로운 변수 생성

3). 군집 변수 생성

Player_id	p_formatted_ip	k_percent	bb_percent	...	Pitcher_cluster_label
424144	18.0	19.4	8.3	...	0
425794	154.1	19.8	6.1	...	2
425844	125.0	18.0	4.3	...	5
429722	65.1	18.8	7.	...	4

각 군집에 해당하는 선수의
고유 번호를 찾아 변수 할당

모델 개발

Step 4. 데이터 병합

Pitch_type	Release_speed	On_1b	...	Batter	Pitcher
2	84.6	1	...	542303	601713
2	84.7	1	...	518692	601713
1	91.7	1	...	645277	601713
1	90.6	0	...	542225	601713

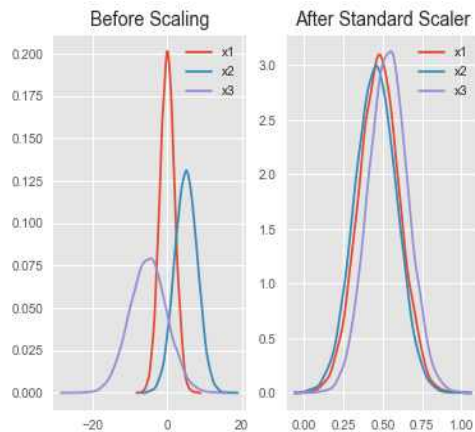
Player_id	p_formatted_ip	k_percent	bb_percent	...	Pitcher_cluster_label
424144	18.0	19.4	8.3	...	0
425794	154.1	19.8	6.1	...	2
425844	125.0	18.0	4.3	...	5
429722	65.1	18.8	7.	...	4

원 데이터 Batter, Pitcher 변수의 고유 번호와
성적 데이터의 Player_id 중 일치하는 행에 맞는 군집 번호 추가

Step 5. 스케일링

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean
 σ = Standard Deviation



변수마다 다른 데이터 스케일을 가지고 있어
보다 정확한 예측을 위해 조정 작업이 필요

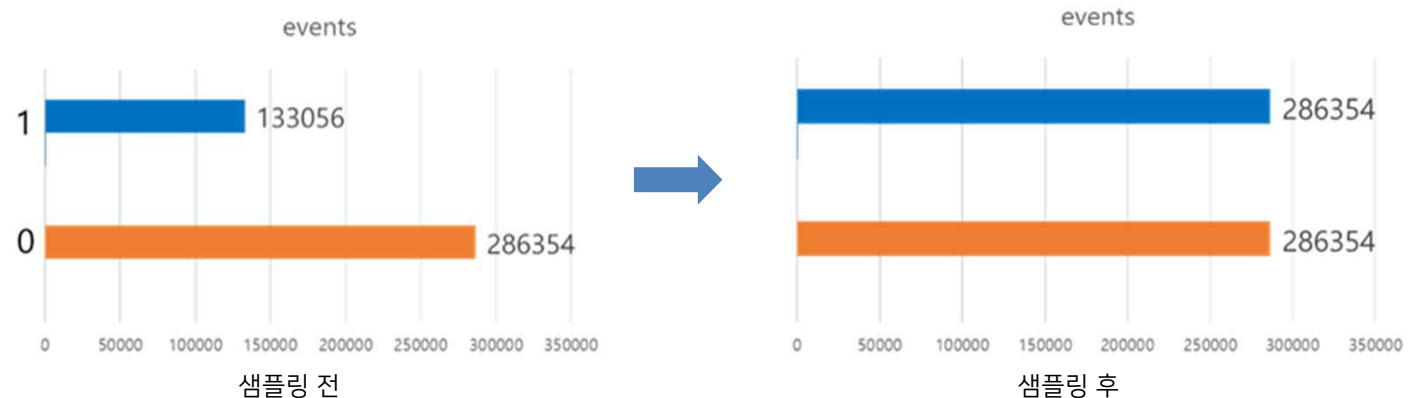


scikit-learn 라이브러리의 표준 스케일러를 사용하여
속성의 평균을 0으로, 분산을 1로 스케일링 실시

모델 개발

Step 6. Smote Oversampling

- 종속 변수 (출루 / 아웃)의 불균형 문제로 Smote Oversampling 기법 활용하여 문제 해결



Step 7. 선수 유형간 출루율 변수 생성

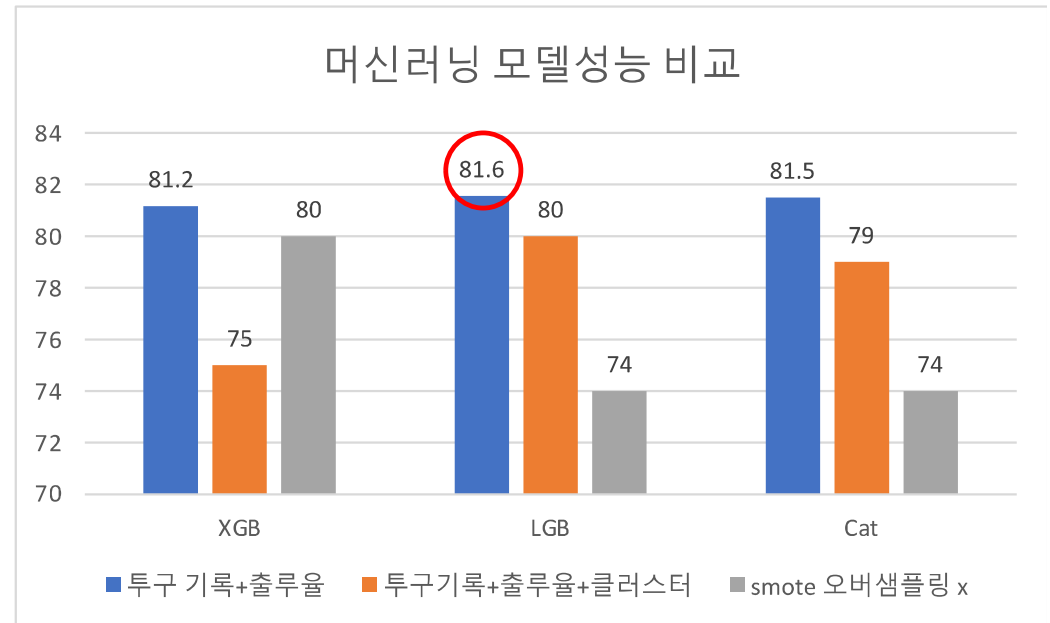
- 투수 및 타자의 각 군집별 대결 결과를 통해 출루와 아웃의 비율을 계산
→ 출루율 계산 결과를 "On_base_ratio" 라는 새로운 열에 생성하여 예측에 사용되는 변수로 활용
- 출루율 = (출루)/(투수와 타자의 맞대결 횟수)
→ Ex) P1(투수군집1) vs B1(타자군집1) = $100(\text{출루}) / 100(\text{출루}) + 300(\text{아웃}) = 0.25$

모델 개발

Step 8. 모델 학습

항목		내용	
주요 모델 파라 미터	n_estimators	300	트리개수
	num_leaves	65	트리의 리프 노드 개수
	max_depth	13	트리의 최대 깊이
	learning_rate	0.1	학습률
	colsample_bytree	0.5	트리별 특성 샘플링 비율
	reg_alpha	0	L1 정규화 파라미터
	reg_lambda	0	L2 정규화 파라미터
학습 태스크	subsample	0.5	전체 데이터 사용
	학습비	8:2	모델 학습과 테스트

- 오버샘플링으로 인한 overfitting 발생하지 않음
- 하이퍼파라미터는 randomsearch cv로 조정



참고 문헌:

- 1). Football Matches Outcomes Prediction Based on Gradient Boosting Algorithms and Football Rating System(Razali, 2021)
- 2). 머신러닝을 활용한 빅데이터 분석을 통해 KBO 타자의 OPS 예측(한정섭, 2022)

Chapter 3. 결론 및 활용

결론 및 활용

1. 팀 자원 파악, 선수 기용, 전략구상에 활용



2023 MLB 신인 투수 센가 코다이

• 선수 유형과 유형별 상성 파악

1. 팀 내에서 어떤 유형의 선수가 있는지 전력 정보 수집
 2. 현재 팀의 전력으로는 어떤 유형의 선수를 상대하는데 어려움이 있는지 파악
- 선수 유형 정보를 통해 팀 자원 및 선수 기용 전략 도출 가능

• 데이터를 기반으로 신인 투수 유형을 파악해 선수 기용에 필요한 정보 도출

- ✓ 해당 연구 결과를 통해 2023 MLB 신인 투수 분석
 - 센가 코다이 → 투수 5 번 군집 (구위가 좋아 삼진율이 가장 좋은 투수 유형) :
 - 5번 군집에 해당하는 투수들(ex. 에드윈 디아즈)의 특성을 분석해서 신인과의 대결 준비
- 신인 선수의 유형을 파악해 전략 준비

결론 및 활용

2. 선수 유형별 출루율 인사이트

1. 땅볼 유도형 투수 vs 파워형 타자:

상대적으로 땅볼 투수형은 컨택트형 타자에 강하고, 파워형 타자 그룹에 약한 면모
(컨택트형 타자에게 출루 허용 \rightarrow 0.409 / 파워형 타자에게 출루 허용 \rightarrow 0.458)



“땅볼 유도형”
펠릭스 페냐

VS



“홈런형 파워타자”
오타니 쇼헤이

2. 구위 삼진형 투수 vs 파워형 타자:

상대적으로 마무리 투수형은 컨택트형 타자에 강하고 파워형 타자 그룹에 약한 면모
(컨택트형 타자에게 출루 허용 \rightarrow 0.413 / 파워형 타자에게 출루 허용 \rightarrow 0.485)



“구위 삼진형 투수”
에드윈 디아즈

VS



“홈런형 파워타자”
오타니 쇼헤이

3. 다양한 구종 구사 가능한 투수

- 선구안형 타자에게 강한 면모: 0.362
- 홈런형 파워 타자에게 강한 면모: 0.372
- 컨택트형 타자에 약한 면모: 0.415



다양한 구종 구사 투수
오타니 쇼헤이

VS



“컨택트형 타자” 김하성



“선구안형 타자” 최지만



“홈런형 파워타자”
오타니 쇼헤이

결론 및 활용

3. 선수 개인별 분석

- 커쇼 선수 유형별 분석



- 커쇼 선수 분석 인사이트



- Input : 2020-2022 커쇼 투구 데이터 + 클러스터

- 선구안형 타자 유형 출루율: 0.232

- 컨택트형 타자 유형 출루율: 0.267

- 장타형 타자 유형 출루율: 0.157

→ 장타형 타자에게 강했으나, 컨택트형 타자에게는 상대적으로 약한 면모를 보임

→ 컨택트형 타자들에게 대응할 수 있는 투구 전략 세울 필요성

결론 및 활용

4. 경기전 예상 라인업 결과예측으로 전략 준비



• 활용 계획

1. 상대 팀 예상 라인업 선별
2. 선발 투수 투구 정보와 투수유형 및 타자 유형을 입력
3. 타격 결과 예측
4. 상대 선수별 전략 사전 준비

→ 각 유형에 해당하는 타자를 상대할 때 어느 위치에 어떤 구종을 던질 시
출루 허용이 낮음을 모델로 확인하여 선수별 전략 준비

11.8% (17.0%)	7.7% (10.2%)	5.3% (8.4%)	8.8% (15.5%)	7.7% (13.2%)	6.1% (13.3%)
7.7% (11.6%)	6.2% (8.2%)	3.3% (5.5%)	10.0% (14.2%)	8.8% (12.3%)	7.1% (12.4%)
5.4% (11.6%)	3.8% (6.7%)	4.0% (9.9%)	16.6% (24.7%)	17.7% (24.9%)	12.7% (23.8%)

결론 및 활용

5. 새로운 연구 참고 자료



새로운 방법론으로 투구기록과 투수 및 타자 군집을 활용하여 출루율을 예측하는 연구를 제시

→ 이는 스포츠 연구에 새로운 참고 자료로 활용될 수 있을 것으로 기대

References

1. 김혁주. "한국 프로야구에서 출루 능력과 장타력이 득점 생산성에 미치는 영향," 한국데이터정보과학회지, vol. 23, no. 6, pp. 1165-1174, 2012.
2. 박태신, 김재윤. "머신러닝을 활용한 KBO 외국인 투수 재계약 예측 모형," 한국데이터정보과학회지, vol. 33, no. 6, pp. 963-976, 2022, doi: 10.7465/jkdi.2022.33.6.963.
3. 손혁. "프로야구 투수유형과 구질과의 관계," 국내석사학위논문 고려대학교 교육대학원, 2004. 서울
4. 이장택. "한국프로야구에서 타자능력의 측정," 한국데이터정보과학회지, 25(2), 349-356, 2014.
5. 이승훈, 최형준. "미국 프로야구(MLB) 풀카운트 상황에서 투수의 구질, 구속 변화에 따른 투구 결과 분석," 한국체육과학회지, vol. 28, no. 3, pp. 973-981, 2019, doi: 10.35159/kjss.2019.06.28.3.973.
6. 조선미, 김주학, 강지연, 김상균. "머신러닝(XGBoost)기반 미국프로야구(MLB)의 투구별 안타 및 홈런 예측 모델 개발," 한국체육측정평가학회지, vol. 25, no. 1, 2023, pp. 65-76.
7. 조형석. "MLB 타자들의 스윙존에 따른 스윙선택 성향 분석," 국내석사학위논문 명지대학교 기록정보과학전문대학원, 2021.
8. 최영환. "4차 산업혁명형 ICT기술이 스포츠 분야에 미치는 기술 · 문화적 동향분석," 한국스포츠학회 16, no.3 (2018): 1-12.
9. 황수웅. "불확실성(uncertainty)을 고려한 스포츠 빅데이터 분석: Bayesian 추정과 Deep Learning을 활용한 프로야구 심판의 Ball/Strike 판정 평가 모델 개발," 서울대학교 대학원, 2023.
10. Albert, Jim. 'Beyond Runs Expectancy'. 1 Jan. 2015: 3 – 18.
11. Marcou, Charlie. "Investigating Major League Baseball Pitchers and Quality of Contact through Cluster Analysis" (2020). Honors Projects. 765.
12. Nathan, Alan M. "What new technologies are teaching us about the game of baseball." Proceedings of the Euromech Physics of Sports Conference. 2012.
13. Otremba Jr., Stephen Eugen. "SmartPitch: Applied Machine Learning for Professional Baseball Pitching Strategy." Massachusetts Institute of Technology, degree of Master of Engineering in Electrical Engineering and Computer Science, 2022. Available at <https://hdl.handle.net/1721.1/145144>.



THANK YOU

