

이재빈, 장성훈, 오준혁

이정후 선수 WAR 및 연봉 예측

SDA 1기 발표

CONTENTS

CHAPTER
01

연구 배경 및 목적

CHAPTER
02

선행 연구

CHAPTER
03

데이터 및 모델 소개

CHAPTER
04

결과 분석

CHAPTER
05

결론 및 앞으로의 방향

연구 배경 및 목적

CHAPTER 01

연구 배경

WAR이란?

선수의 종합적 가치를
나타내는 핵심 성과지표

WAR 중요성

WAR은 대체 선수 대비
승리 기여도의 의미로
연봉 산정의 중요한 지표
MLB: 1WAR당 연봉 800만달러
(약 116억)~900만달러(약 130
억)로 간주

연구 목적

MLB 진출 이후 성과 불확실성

- KBO 최고 타자 → MLB 전환 선수
 - 2024 시즌 부상으로 성과 판단 어려움
 - 2025 시즌 팀 내 타율 1위, WAR 2.4 기록
-

장기 계약과 옵트아웃 의사결정

- 6년 1.13억 달러 계약
 - 2027 시즌 후 옵트아웃 조항
 - 미래 성과 기반 가치 재평가 필요
-

WAR 기반 선수 가치 예측

- 2026 시즌 WAR 예측
- 성과 범위 및 불확실성 분석
- 옵트아웃 경제성 판단 근거 제시



선행 연구

CHAPTER 02

선행 연구

MLB projection LSTM

- 2023년 5개의 딥러닝 LSTM모델과 GRU, BiLSTM, AT-LSTM을 이용해서 mlb 타자의 홈런 개수를 예측하는 연구를 수행함
- 직전 5년간 타자의 스탯으로 다음 해 홈런 개수 예측
- 전통적 ML, FC layer, existing projection system보다 더 낮은 MSE 달성
- 나아가야 할 방향으로 데이터 부족 극복, 30홈런 이상의 발생 빈도가 드문 스탯을 잘 예측하는 방법 고안

문헌 출처: Sun, H.-C., Lin, T.-Y., & Tsai, Y.-L. (2022). Performance prediction in Major League Baseball by Long Short-Term Memory Networks. arXiv.

GDP(병살타)

- oWAR(공격 지표만 고려한 WAR) 예측 수행
- 도루성공/실패 횟수와 GDP(병살타 개수)가 큰 영향을 미침 확인

문헌 출처: Kerr, N. (2024). An analysis of the key components of WAR in baseball (Undergraduate thesis, University of Lynchburg). University of Lynchburg Digital Showcase.

선행 연구

OAA

- 수비 능력을 정량화한 OAA(Outs Above Average) 활용
- 타구 판단 능력과 스피드, 수비 범위 반영
- 이정후는 외야수이므로 외야수만 필터링하여 분석 정확도 확보

ArmStrength

- 이정후의 중견수→우익수 포지션 이동 고려
- 우익수는 송구 능력이 핵심 수비 가치
- 연도별 변동성과 노이즈가 큰 Arm_value는 제외
- 연도별 변동이 안정적인 ArmStrength 사용

KBO→MLB

- KBO 기록을 MLB 기준으로 변환 필요
- 리그 수준 차이 보정 적용
- 이정후의 KBO 2023시즌 타격 지표 변환 사용

데이터터 및 모델 소개

CHAPTER 03

데이터 및 모델 소개

데이터셋 구성

- pybaseball MLB 외야수 데이터
- 2019-2025, 100PA 이상
- 타격, 수비, 주루 지표 포함

사용 피처

- wRC+, BsR, OAA, GDP, HardHit%
K%, PA, Age, ArmStrength

데이터 및 피처 구성

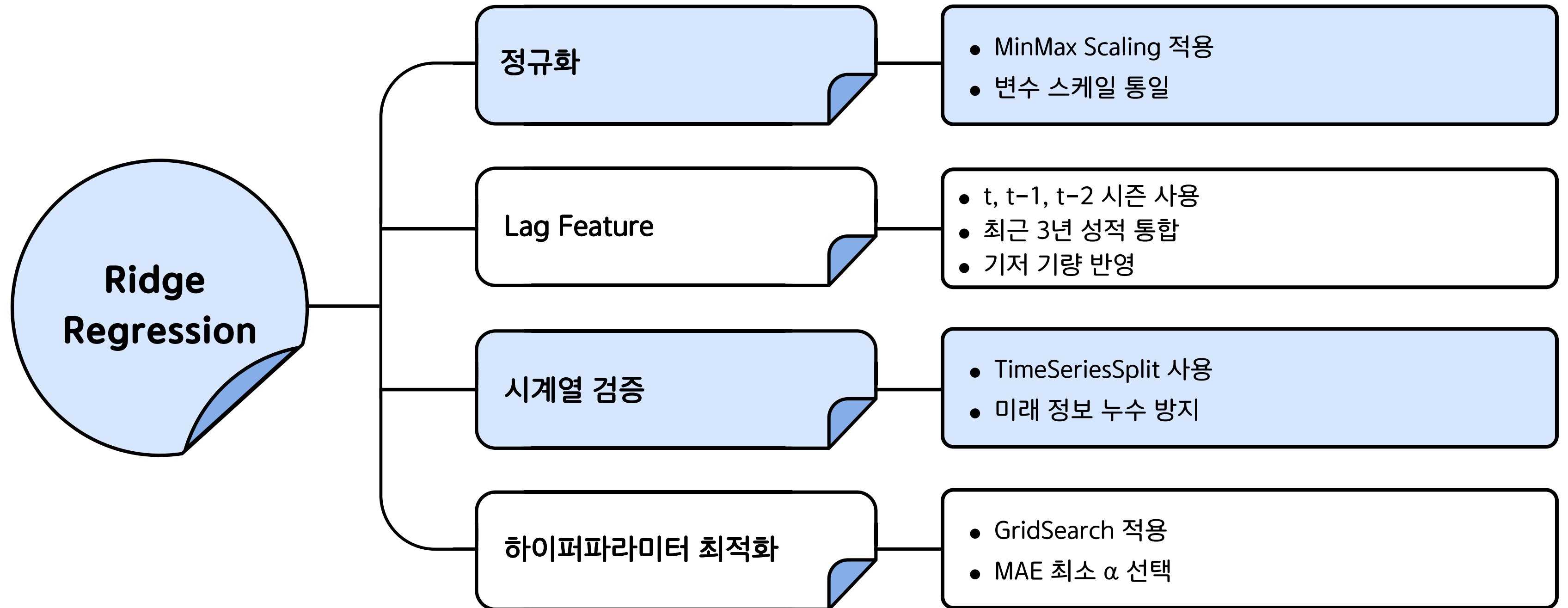
예측 목표

- 다음 시즌 WAR 예측
- 직전 3년 성적 기반

양상블 전략

- 10개 Random Seed 학습
- 예측 평균으로 분산 감소
- 안정적 WAR 추정

데이터 및 모델 소개



데이터 및 모델 소개

01

데이터 정규화

- MinMaxScaler로 0-1 스케일 변환
- 피처 간 값 범위 차이 제거

02

결측치 보정

- arm_strength 결측 발생
- 선수별 연도 평균값으로 대체

03

시계열 텐서 변환

- 입력 구조: (샘플, 3년, 9피처)
- 연도 간 성과 흐름 반영

①

②

시계열 LSTM 모델

LSTM은 시간 순서가 있는 데이터를 학습하여 과거 정보로 미래 값을 예측하는 순환신경망(RNN) 기반 딥러닝 모델이다. 특히 시계열 데이터에서 과거 값의 영향이 오래 지속되는 경우에도 장기 의존성(long-term dependency)을 효과적으로 학습할 수 있다.

04

슬라이딩 윈도우

- 3년 연속 시퀀스 생성
- 다음 해 WAR 매칭

05

신경망 구조

- LSTM(64) → Dropout(0.2) → LSTM(32)
- 시계열 패턴 학습

06

딥러닝 앙상블

- 10개 Random Seed 학습
- 예측 평균으로 분산 감소

결과 분석

CHAPTER 04

결과 분석

①

RIDGE REGRESSION 결과

②

```
=== Test Performance (last 2 years of input season t) ===  
MAE: 1.2933422469693543  
R2 : 0.16773356666303674  
=== 3YR Prediction ===  
Lee Jung-hoo 2026 WAR prediction: 2.383
```

01

WAR 예측 결과

- 직전 3년 성적 기반 예측
2026 WAR = 2.38

02

모델 성능 해석

- MAE = 1.29
- $R^2 = 0.17$
→ WAR이라는 지표의 구조적 특성 때문에 낮게
나옴

결과 분석

①

시계열 LSTM 결과

②

이정후 선수 2026 WAR 예측 리포트

▶ 1. 모델 성능 검증 (Model Performance)

- 평균 결정계수(R^2): 0.420 (1에 가까울수록 좋음)
- 평균 손실값(MSE): 0.0132 (0에 가까울수록 좋음)
- [판단] 데이터 패턴이 복잡하여 설명력이 다소 낮습니다.

▶ 2. 2026시즌 성적 예측

- 앙상블(10회) 결과: [np.float32(2.23), np.float32(2.51), np.float32(2.62), np.float32(2.13), ...]
- 최종 예상 WAR: 2.25 (± 0.23)

▶ 3. 옵트아웃 의사결정 (Decision Making)

- 예상 시장 가치: \$17.99M (약 234억 원)
- 잔여 연봉(2028): \$20.5M
- 차액(Diff): \$-2.51M

[결론] 옵트아웃 '보류/잔류'

01

LSTM WAR 예측 범위

- 10-seed 앙상블 결과
- 95% 신뢰구간 내에서 2026 WAR는 1.79~2.71으로 형성

02

모델 타당성 검증

- 예측 설명력
- $R^2 = 0.420$
- WAR 예측 오차
- ± 0.23 수준

결과 분석

01

MLB projection system과의 비교

- 현재 메이저리그에서 사용되는 projection 시스템의 예측 결과와 비교하였을 때
- 대부분의 시스템이 예측한 이정후 선수의 WAR 수치가 우리 모델의 95% 신뢰 구간 내에 완벽하게 포함되는 결과를 보였다

2026	FGDC	PROJ	27	141	609	10	68	64	9	7.7%	10.7%	.133	.291	.271	.333	.404	.322		108	-0.1	5.9	-4.9	2.2
2026	Steamer	PROJ	27	110	468	8	52	51	8	7.7%	11.4%	.137	.294	.273	.334	.410	.324		110	0.5	6.0	-2.3	2.0
2026	ZiPS	PROJ	27	129	547	9	62	56	7	7.7%	10.1%	.129	.287	.270	.332	.398	.319		107	-1.2	3.2	0.0	2.2
2026	ZiPS DC	PROJ	27	141	609	10	69	62	8	7.7%	10.1%	.129	.287	.270	.332	.398	.319		107	-1.3	3.6	-7.2	1.7
2026	ATC	PROJ	27	135	559	8	64	54	9	7.5%	11.3%	.127	.288	.266	.326	.393	.314		100	0.3	0.5	-7.4	1.2
2026	THE BAT	PROJ	27	135	559	9	67	57	8	7.4%	11.2%	.133	.293	.271	.329	.404	.319		107	-0.3	4.1	-8.1	1.5
2026	THE BAT X	PROJ	27	135	559	7	65	54	8	7.4%	11.3%	.122	.292	.268	.327	.391	.313		102	-0.2	1.4	-8.1	1.2
2026	OOPSY	PROJ	27	141	609	12	83	66	10	8.9%	13.0%	.148	.297	.272	.342	.419	.332		115	0.5	11.5	-3.5	2.9

결과 분석

①

시계열 LSTM 낙관/비관 모델 결과

②

```
Optimistic 모델 학습 중 ( $\tau=0.8$ )...  
Pessimistic 모델 학습 중 ( $\tau=0.2$ )...
```

✓ OPTIMISTIC vs PESSIMISTIC 결과

TRAIN 성능:

```
Optimistic ( $\tau=0.8$ ):  $R^2 = 0.3605$ ,  $MSE=0.0141$   
Baseline (MSE):  $R^2 = 0.4204$ ,  $MSE=0.0132$   
Pessimistic ( $\tau=0.2$ ):  $R^2 = 0.1507$ ,  $MSE=0.0187$ 
```

이정후 2026 WAR 예측:

```
Optimistic: 3.08 WAR  
Baseline: 2.25 WAR  
Pessimistic: 1.57 WAR  
차이 (Opt - Pes): 1.52 WAR
```

01

Expectile Loss 기반 LSTM 모델

- 기존 LSTM 손실함수(MSE)를 Expectile Loss로 변경하여 낙관·비관 모델을 설계
- 새로운 하이퍼파라미터 τ 를 도입해서 낙관적 모델과 비관적 모델을 디자인
- 오차 방향별 패널티 차등 적용

02

오차함수(loss func)

- 낙관 모델 손실 함수
 $0.8 (\text{실제값} - \text{예측값})^2$ (실제값 > 예측값)
 $0.2 (\text{실제값} - \text{예측값})^2$ (실제값 < 예측값)
- 비관 모델 손실 함수
 $0.2 (\text{실제값} - \text{예측값})^2$ (실제값 > 예측값)
 $0.8 (\text{실제값} - \text{예측값})^2$ (실제값 < 예측값)

결과 분석



- 10- seed 앙상블 낙관 모델이 3.08, 베이스라인이 2.25, 비관 모델이 1.57로 예측
- 100번의 몬테카를로 드롭아웃 기법을 이용한 시뮬레이션의 결과, 낙관 모델은 2.69~3.56, 베이스라인이 1.52~2.20, 비관 모델이 1.29~1.86

Optimistic 모델 학습 중 ($\tau=0.8$)...
Pessimistic 모델 학습 중 ($\tau=0.2$)...

=====

✅ OPTIMISTIC vs PESSIMISTIC 결과

=====

TRAIN 성능:

Optimistic ($\tau=0.8$): $R^2=0.3605$, $MSE=0.0141$

Baseline (MSE): $R^2=0.4204$, $MSE=0.0132$

Pessimistic ($\tau=0.2$): $R^2=0.1507$, $MSE=0.0187$

이정후 2026 WAR 예측:

Optimistic: 3.08 WAR

Baseline: 2.25 WAR

Pessimistic: 1.57 WAR

차이 (Opt - Pes): 1.52 WAR

PREDICTION DISTRIBUTION (100 MC samples):

Optimistic: Mean=3.14, 10-90%=[2.69, 3.56], std= 0.36

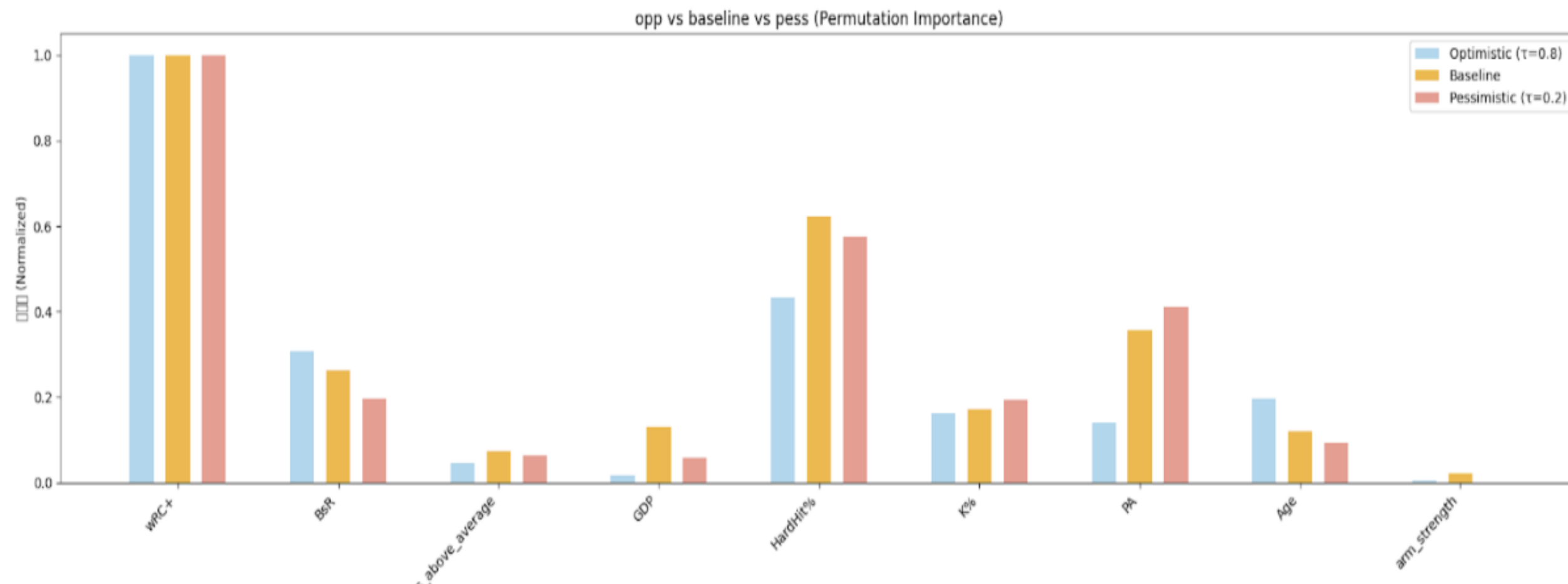
Baseline: Mean=1.88, 10-90%=[1.52, 2.20], std= 0.25

Pessimistic: Mean=1.57, 10-90%=[1.29, 1.86], std= 0.23

=====

결과 분석

모델별 WAR 변수 중요도(PERMUTATION IMPORTANCE)



그래프 분석

- 공통 핵심 변수: wRC+, HardHit%
- 낙관 모델 특징: BsR, Age (추가 성장 동력)
- 비관 모델 특징: PA, OAA, K% (하락 방어 지표)
- 베이스라인 특징: HardHit%, GDP

결론 및 앞으로의 방향

CHAPTER 05

결론 및 앞으로의 방향

모델 예측 WAR 수준

- 예측 WAR 1.8-2.7 수준

.....>

옵트아웃 가능성 낮음

중요 변수 분석

- wRC+
- HardHit%

.....>

WAR 결정 주요 요인

모델 구조 확장 가능

- 타자 WAR 예측

.....>

투수 WAR 예측 확장

Q&A

허창민, 구본주, 한종업

이정후 선수 최적 타순 배정 전략

SDA 1기 발표

CONTENTS

CHAPTER

01

문제 정의

CHAPTER

02

데이터 전처리 및 방법

CHAPTER

03

결과 분석

CHAPTER

04

한계

CHAPTER

05

결론

문제 정의

CHAPTER 01

문제 정의

1

핵심 질문

**이정후를 어느 타순에 배치해야
팀 기대 득점이 개선되는가?**

-> 선수평가가 아닌, 팀 성과 최적화의 관점
팀의 성과는 월별 상태로 분해하고, 상태별 운영 규칙으로 연결

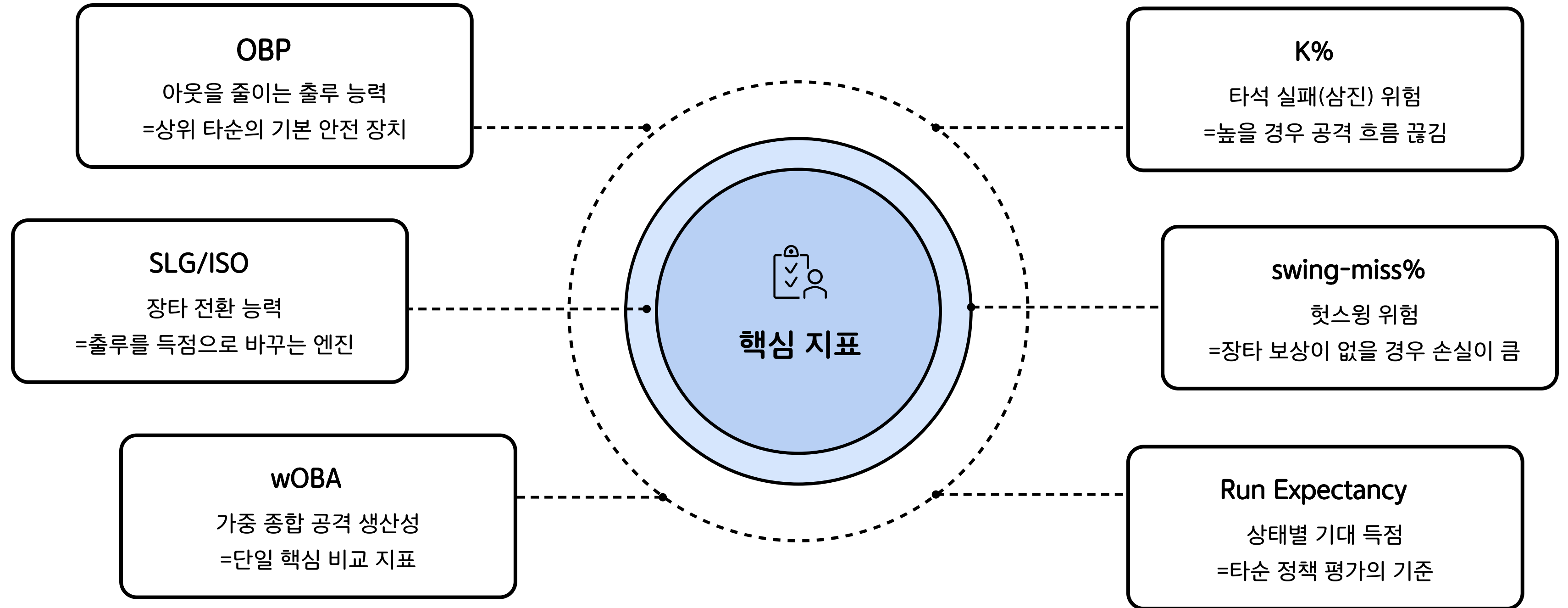
2

분석 질문

현상 설명 -> 운영 규칙

- RQ1. SF 최저승률 월 vs 동월 최고팀
- RQ2. 이정후 부진 월의 유사 고성과 선수
- RQ3. 2008~2025 SF 최고월과 현재 간극
- RQ4. 이정후 유사+꾸준한 선수(롤모델 후보)
- RQ5. SF 최저타율 월의 동군집 상위팀
- RQ6. C2 vs C4 격차와 타순 임계치(승격 조건)

야구 지표 정리



데이터 전처리 및 방법

CHAPTER 02

데이터 전처리 및 방법 설명

분석단위	샘플수	기간	설명
팀-월	3120	2008~2025	팀 공격지표 + 승률
선수-월	11412	2008~2025	선수 공격지표 + DEC 잠재벡터
SF 2025	6	2025	핵심 의사결정 대상
Lee 2025	4	2025	타순 최적화 대상

1 데이터 및 전처리

전처리: 결측치 중앙값 대체, Yeo-Johnson 왜도 보정, 표준화
분석 단위: 월별 단면 정보 기반, 시간흐름 (roll3, delta) 함께 반영
팀 성과 변화와 선수 상태 변화를 같은 축에서 교차 해석

대상	DEC latent Sil	KM latent Sil	DEC raw Sil	KM raw Sil	DEC-KM ARI
팀	0.7938	0.7946	0.0328	0.0993	0.9922
선수	0.7911	0.7912	0.0249	0.0819	0.9971

2 DEC와 대조 모델

DEC: 오토인코더 기반 잠재공간에서 군집 경계 정제; 비선형 상호작용 확인
해석: 잠재공간 분리력은 충분, DEC-KMeans ARI가 높아 라벨구조 유사
=> DEC 단독 주장 대신 대조 모델 결과를 함께 고려

스왑	최상위 k	composite_score
Team latent	8	4.432
Player latent	9	4.261
Team raw	6	5.165
Player raw	7	4.297

3 k 민감도 분석

군집 수 선택의 임의성 줄이기 위한 k스왑 수행
지표: silhouette, Davies-Bouldin, Calinski-Harabasz 결합 composite score

결과 분석

CHAPTER 03

군집 프로파일링: 성과를 만드는 구조

표 3.1: 팀 군집 프로파일(상·하위 3개)

group	cluster	cluster_name	n	win_pct	obp	slg	woba	iso	k_pct	swing_miss
상위 3개	2	고출루-장타 주도형	498	0.5834	0.3371	0.4517	0.3381	0.1883	21.260	24.428
상위 3개	1	고출루-장타 주도형	523	0.5547	0.3414	0.4335	0.3382	0.1636	18.114	20.679
상위 3개	0	고변동 파워형	339	0.5222	0.3163	0.4112	0.3138	0.1672	22.895	25.379
하위 3개	3	저생산 정체형	491	0.4027	0.2955	0.3707	0.2906	0.1398	23.989	26.408
하위 3개	5	저생산 정체형	418	0.4406	0.3054	0.3677	0.2968	0.1254	19.743	21.854
하위 3개	4	혼합 밸런스형	278	0.4882	0.3176	0.4032	0.3135	0.1599	22.998	25.921

표 3.2: 선수 군집 프로파일(상·하위 3개)

group	cluster	cluster_name	n	woba	obp	slg	iso	k_pct	swing_miss
상위 3개	2	완성형 코어 타자	1500	0.4121	0.3957	0.5798	0.2628	19.059	23.219
상위 3개	1	완성형 코어 타자	1281	0.4035	0.3947	0.5587	0.2501	17.075	21.335
상위 3개	6	중간 밸런스형	1033	0.3569	0.3401	0.4966	0.1973	17.017	22.061
하위 3개	0	저효율 미스매치형	1392	0.2833	0.2887	0.3587	0.1354	21.818	24.528
하위 3개	8	중간 밸런스형	1553	0.2888	0.3008	0.3563	0.1023	13.735	16.021
하위 3개	4	저효율 미스매치형	1008	0.2959	0.2961	0.3841	0.1480	24.370	27.208

01

팀 군집 프로파일링

- 팀 상위 군집: OBP, SLG 동시 우위
- 팀 하위 군집: 두 지표 동시 약세 또는
리스크 지표(K%, swing-miss%) 높음

02

선수 군집 프로파일링

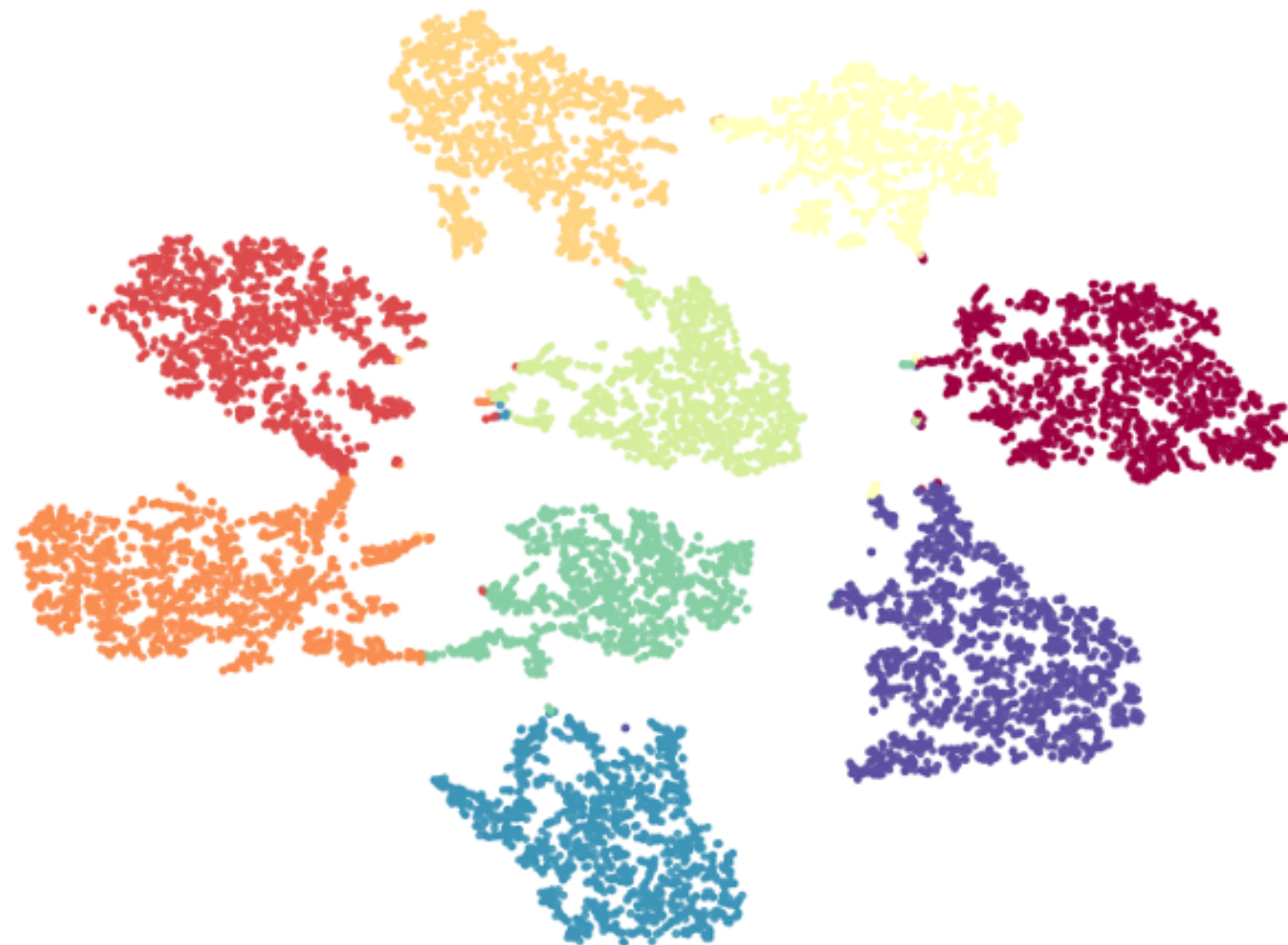
- 팀 상위 군집과 동일
- 상위 성과는 한 지표의 일시적 상승이
아니라 지표 조합의 균형에서 나옴

실행 제안: SF는 상위타선 출루 복원과 중위타선 장타 전환을 분리하지 않고 동시 관리 해야 한다.

RQ1: SF 최저 승률 월 vs 동월 최고팀

구분	월	군집	승률	OBP	SLG	wOBA
SF 최저승률 월	Jul	C3	0.3750	0.3070	0.3810	0.3000
동월 리그 최고팀	Jul	C2	0.7080	0.3230	0.4720	0.3410

t-SNE on Player DEC latent



01

SF 최저 승률 월 vs 동월 최고팀

- 동월 최고팀은 SF 대비 승률 +0.333, OBP+0.016, SLG +0.091, wOBA +0.041

02

선수 DEC t-SNE

- 의미: SF 저성과 월은 특정 선수의 부진보다 공격 체인(출루-장타연결)의 약화
- 실행 제안: 동월 상위팀 벤치마크를 바탕으로 출루 하단선과 장타 전환률을 동시 KPI로 운영
- 동월 상위 3팀 요약
- BOS: 승률0.708, OBP 0.323, SLG 0.472, wOBA 0.341
- MIL: 승률0.708, OBP 0.331, SLG 0.413, wOBA 0.324
- TOR: 승률0.692, OBP 0.363, SLG 0.476, wOBA 0.362

RQ2: 이정후 부진 월의 유사 고성과 선수

기준	이정후 월	이정후 군집	이정후 wOBA	유사 고성과 선수	비교선수 wOBA
종합 부진점수	May	8	0.2680	Flores, Wilmer	0.3370
wOBA 최저	Jun	7	0.2540	Edwards, Xavier	0.3520

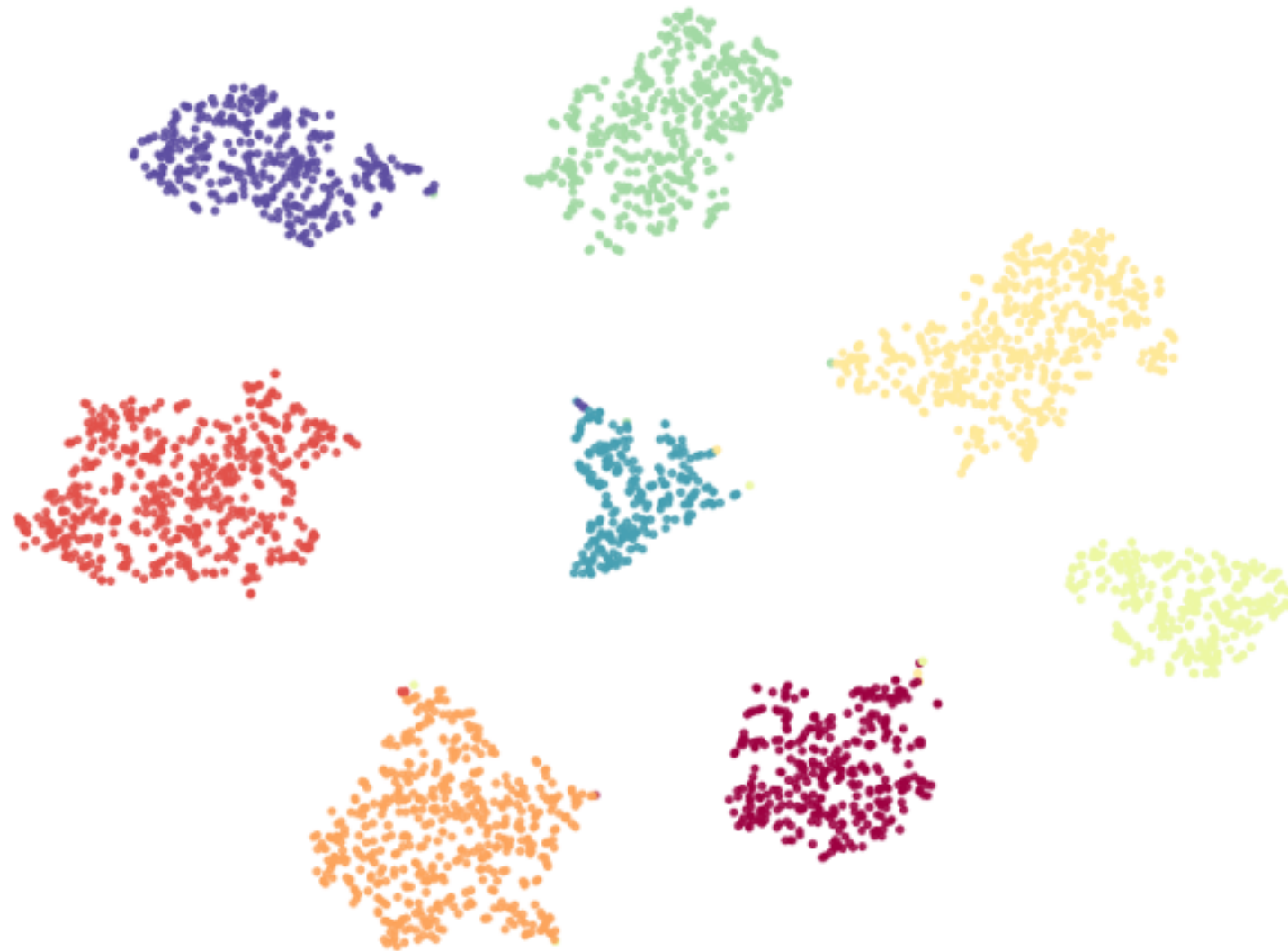


그림 3.2: 팀 DEC t-SNE

해석: 부진월을 두 기준(종합점수, wOBA 최저)으로 교차해도 같은 군집 내 고성과 비교군이 존재
-> 군집 자체보다 실행 품질 차이가 성과를 가름
실행 제안: 군집 내부 상향 전략 (존관리, 카운트 운영, 약한 스윙 제거)을 먼저 적용

RQ3: SF역사적 최고 월과 현재 간극

① SF 역사적 최고 월 대비 2025 평균 격차

metric	delta(best_minus_2025_mean)
win_pct	0.2720
obp	0.0447
slg	0.0822
woba	0.0457
iso	0.0448
k_percent	-1.683
swing_miss_percent	-1.333

01

해석

간극은 단일 지표가 아니라 다지표 결합에서 발생
->단일 지표 처방으로는 과거 상단 구조 복원 어려움

02

실행 제안

- 1) OBP 바닥선 회복
- 2) ISO/SLG 동반 상승
- 3) 리스크 지표 안정화

RQ4: 이정후 유사+꾸준한 선수

player_name	months	woba_mean	woba_std	obp_mean	slg_mean	dist	score
Hoerner, Nico	16	0.3225	0.0236	0.3447	0.3879	2.303	1.131
Arraez, Luis	20	0.3306	0.0543	0.3554	0.4017	0.9947	1.070
Perdomo, Geraldo	9	0.3599	0.0411	0.3841	0.4377	4.408	1.062
Schanuel, Nolan	8	0.3332	0.0403	0.3591	0.3972	2.612	0.9940

01

해석

dist(유사도)와 woba_std(안정성)를 함께 보면
"닭았지만 불안정한 선수"를 걸러낼 수 있음

02

실행 제안

롤모델은 결과 숫자보다
의사결정 루틴(볼카운트 대응, 2스트라이크 전략)
을 벤치마킹해야 재현성이 높음

RQ5: SF 최저 타율 월의 동군집 상위팀

구분	월	군집	BA	SLG	wOBA	승률
SF 최저타율 월	Jun	C4	0.2240	0.3610	0.3000	0.4810
동군집 최고팀	Jun	C4	0.2530	0.4350	0.3250	0.5930

01

해석

같은 군집 내에서도 승률 +0.112, SLG +0.074,
wOBA +0.025의 차이 존재
->군집은 운명 라벨 X 상태 라벨 O
같은 군집 내에서도 실행 전략으로 상단이동 가능

02

실행 제안

저타율 구간에서는
단타 증가보다는 장타 전환(ISO/SLG 회복)
을 우선 관리

RQ6: C2 vs C4 차이와 통계 검정

표 3.8: C2 vs C4 차이(단순 평균)

metric	C2_minus_C4
woba	0.1161
obp	0.0996
slg	0.1957
ba	0.0810
iso	0.1148
k_percent	-5.311
bb_percent	3.604
swing_miss_percent	-3.989

표 3.9: C2 vs C4 차이(부트스트랩 CI 및 순열검정)

comparison	metric	mean_diff	ci95_low	ci95_high	perm_pvalue
Player C2-C4	woba	0.1161	0.1135	0.1189	0.0007
Player C2-C4	obp	0.0996	0.0969	0.1024	0.0007
Player C2-C4	slg	0.1957	0.1900	0.2018	0.0007
Player C2-C4	iso	0.1148	0.1100	0.1199	0.0007
Player C2-C4	k_percent	-5.311	-5.679	-4.888	0.0007
Player C2-C4	swing_miss_percent	-3.989	-4.390	-3.542	0.0007

01

해석

C2는 C4 대비 생산성 우위, 리스크 우위를 동시에
보임

-> C2=목표 상태, C4=교정 상태 정의

02

검정 해석

mean_diff 방향과 신뢰구간, p-value가
일치하면 시각화 인상 이상의 근거를 확보한 것

SF 월별 핵심 코멘트

타순별 기대 득점(기준 6번 대비)

slot	PA	runs_vs_avg	delta_vs_6
1.000	730.000	-17.855	-2.446
2.000	710.000	-17.366	-1.957
3.000	690.000	-16.877	-1.468
4.000	670.000	-16.388	-0.9784
5.000	650.000	-15.898	-0.4892
6.000	630.000	-15.409	0.0000
7.000	610.000	-14.920	0.4892
8.000	590.000	-14.431	0.9784
9.000	570.000	-13.942	1.468

01

Mar/Apr

강점=win_pct_pct(0.900)

약점=slg_pct(0.333).우선개입지표=slg_pct,
유지지표=win_pct_pct

04

Jul

강점=bb_percent_pct(0.467),

약점=win_pct_pct(0.233). 우선개입지표
=win_pct_pct, 유지지표=bb_percent_pct

02

May

강점=bb_percent_pct(0.567),

약점=slg_pct(0.200). 우선개입지표=slg_pct,
유지지표=bb_percent_pct

05

Aug

강점=slg_pct(0.700),

약점=bb_percent_pct(0.367). 우선개입지표
=bb_percent_pct, 유지지표=slg_pct

03

Jun

강점=bb_percent_pct(1.000),약

점=slg_pct(0.133). 우선개입지표=slg_pct,
유지지표=bb_percent_pct

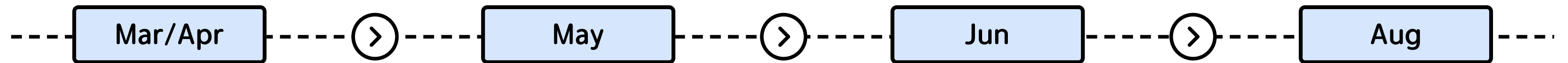
06

Sep/Oct

강점=bb_percent_pct(0.933),

약점=slg_pct(0.467). 우선개입지표=slg_pct,
유지지표=bb_percent_pct

이정후 월별 핵심 코멘트



강점=ba_pct(0.923)
약점=bb_percent_pct(0.468)
코칭우선순위=bb_percent_pct
타순테스트조건=ba_pct유지

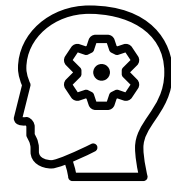
강점=ba_pct(0.233)
약점=bb_percent_pct(0.078)
코칭우선순위=bb_percent_pct
타순테스트조건=ba_pct유지

강점=bb_percent_pct(0.891)
약점=ba_pct(0.009)
코칭우선순위=ba_pct 보완
타순테스트조건
=bb_percent_pct유지

강점=ba_pct(0.819)
약점=bb_percent_pct(0.150)
코칭우선순위=bb_percent_pct
타순테스트조건=ba_pct유지

통합 해석: 운영 프레임

① 운영 프레임



팀: C3/C4 체류를 줄이고 C2형 구조로 이동(출루-장타 동시 관리)
선수: 이정후는 6번에서 시작하고, 임계치 충족 시 2번을 2주 단위로 테스트
평가: 단기 급등락보다 6-8주 지속성 기준 판단

1 권고 근거

- 1) 저성과는 단일 지표 하락이 아닌, 다지표 동시 악화로 드러났음
- 2) 동군집/동월 비교에서 상단과 하단이 갈리므로 "라벨 변경"을 기다리기 보다 라벨 내부 실행 개선

2 팀 운영 체크리스트 (권고)

경기 전: 상위타순 OBP 하단선과 중위타순 ISO/SLG 하단선을 동시에 점검
경기 후: K%와 swing-miss%를 실패 지표로 모니터링(두 축 이상 동시 악화 시 즉시 조정)

3 이정후 운영 체크리스트(권고)

기본: 6번 유지(손실 방어)
승격: OBP, wOBA 임계치가 6-8주 유지되면 2번을 2주 테스트
회귀: 임계치 이탈 또는 득점 연결성 악화 시 6번 복귀

결과 분석 요약

"팀은 구조를 복원하고, 선수 운영은 규칙으로 관리한다."

팀: 출루-장타-리스크 지표를 패키지로 개선해 C2형 구조로 이동

이정후: 6번에서 시작해 임계치 충족 시 2번을 테스트하는 단계형 정책이 기대값-리스크 균형이 가장 좋음

RQ1-RQ6 비교에서 반복적으로 확인된 패턴 (동월 벤치마크, 동군집 상단 사례, C2-C4 격차, Run Expectancy 시나리오)를 통합

개별 표만의 해석에서 오는 과해석을 줄임

한계

CHAPTER 07

한계와 해석 원칙 및 보완

한계

데이터 범위의 한계

변수 구성의 한계

모델링의 한계

통계 해석의 한계

운영 적용의 한계

해석 원칙 및 보완

월별 결론="방향성 판단"

미시 메커니즘 설명 제한적; 거시 지표 중심 적용

인과 효과 X; 우선순위 제시에 강점

인과 증명 X; 순위와 방향 중심 해석

비정량 요소의 존재; 단계형 적용 전제

실제 운영에 유의미 하지만, 적용 범위와 해석 경계를 지킬 때 가장 큰 가치를 가짐

결론

CHAPTER 08

분석 질문별 요약 결론

01

SF 최저 승률 월 vs 동월 최고 팀

SF 2025최저 승률 월은 동월 최고팀 대비
출루,장타,종합생산성이 동시에 낮은
구조적 저성과 구간으로 확인

04

이정후와 유사+꾸준한 선수

유사도(dist)와 안정성(wOBA 표준편차)를
함께 보는 방식으로
"닭았지만 불안정한 선수"를 걸러내고 벤치마킹

02

이정후 부진 월의 유사 고성과 선수

이정후 부진 월에도 동일 군집 내 고성과
유사 선수의 존재: 군집 자체보다
군집 내부 실행 품질이 성과를 가름

05

SF 최저 타율 월의 동군집 상위팀

같은 군집 안에서도 상단이동 가능 확인
군집은 운명이 아니라
상태 코드임

03

SF 역사적 최고 월과 현재 간극

SF 역사적 최고 월과 현재의 간극은
단일 지표가 아니라,
다지표 동시 약화에서 발생

06

C2 vs C4 차이와 통계 검정

C2는 C4 대비 생산성, 리스크 측면 우위;
C2= 목표 상태, C4=교정 상태

이정후 선수 타순 최종 권고

조건	권고 타순	의미
현재 상태	6번	하방 리스크 제어, 기대손실 최소화
임계치 충족 (6~8주)	2번 테스트	추가 타석의 기대이익 검증
테스트 성공	2번 고정	팀 득점 연결성 개선 시 상향 고정
테스트 실패/회귀	6번 복귀	손실 확산 방지

타순=상태 기반 규칙으로 운영
단기 변동성에 휘둘리지 않으면서도, 상향 가능성이 열릴 때 빠르게 테스트 가능

SF 운영: C3/C4 체류를 줄이고 C2형 구조로 이동하는것; OBP와 ISO 동시에 관리
손실 방어->안정화->상향 테스트->고정 의 순서로 실패 비용 감소, 성과 상향 확보

핵심 한 줄 정리

"이정후는 6번에서 시작하고,
조건이 충족되면 2번을 테스트 한다."

Q&A

THANK YOU
감사합니다