

The background features a green field with a white grid pattern. Various baseball-related items are scattered around: a white pinstriped jersey with blue trim and a blue 'H' logo on the left; a white pinstriped pant leg on the left; a white baseball with red stitching at the top; a red baseball bat with a blue handle at the top; a red catcher's mask at the top right; a brown leather glove on the right; a red catcher's chest protector at the bottom right; a blue and white baseball cap with a black 'X' logo at the bottom left; and a white baseball with blue stitching at the bottom. A large white rectangular box with a blue border is centered, containing the title text.

# ***Korean Baseball Match Prediction Project***

**도원결의**



# Table of contents



01

## 프로젝트 과정

프로젝트 진행 과정 요약



02

## 내일 우리팀은?

기존 데이터 학습을 통해  
경기 결과 예측 모델 생성



03

## 애가 우리팀이었다면!

선수 데이터 학습을 통한  
나만의 팀 시뮬레이션



04

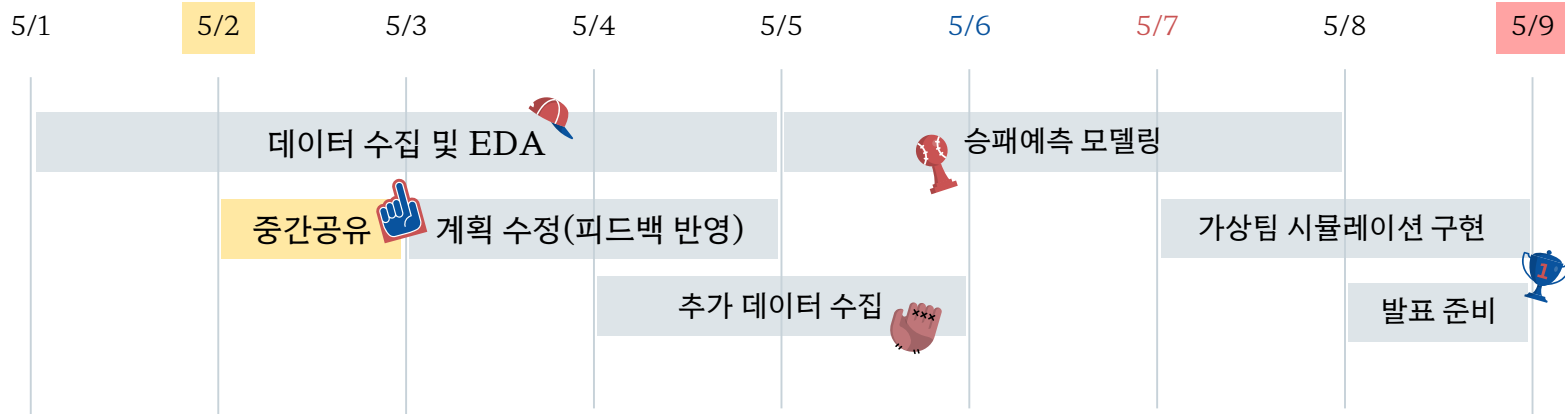
## 회고

프로젝트 회고 및 차후 수정 계획



01





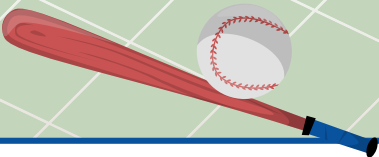
# 프로젝트 과정

# 프로젝트 진행 과정



중간 피드백 후,

-  계획보다 데이터 수집 및 EDA 기간 증가
-  가설 구체화 이후 승패예측 모델링 형태 변화



02

내일 우리팀은?

# 승패예측모델 - 데이터 수집

대상 : KBO 정규시즌 기록

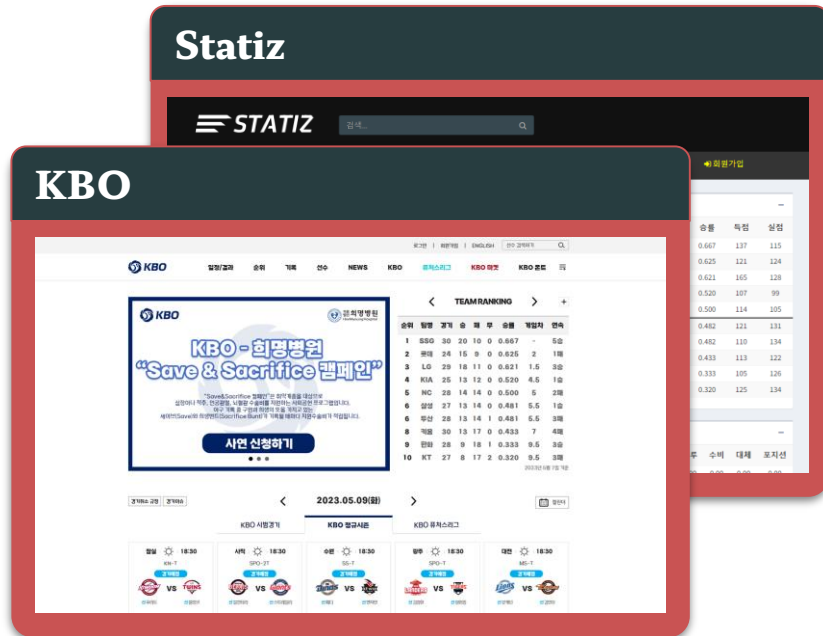
기간 : 2017 ~ 2023.04

출처 :

- 기본 : KBO 공식 홈페이지
- 추가 : Statiz

내용 :

- 시즌별 팀 종합 기록
- 시즌별 선수 종합 기록
- 날짜별 경기 기록



# 승패예측모델 - 피쳐 설명



## 경기 기본 정보

'날짜', '구장', '팀'



## 팀 성적 데이터를 기반으로 머신러닝을 통해 예측한 값

'승률' : 2023년의 팀별 예측 승률

'홈/원정\_승률' : 구장에 따라 홈/원정 경기 예측 승률

'상대승률' : 각 팀마다 상대팀과의 예측 승률



## 선수 개별 데이터 수집 후 전처리한 데이터

### 타자 기록

'WAR' : 대체선수 대비 승리기여도(타자)

'wRC+' : 조정 득점 창출력

'AVG' 타율

'OBP' 출루율

'OPS' 출루율+장타율

### 개인 투수 기록

'WAR\_1' : 대체선수 대비 승리기여도(투수)

'ERA' 투수 평균자책점

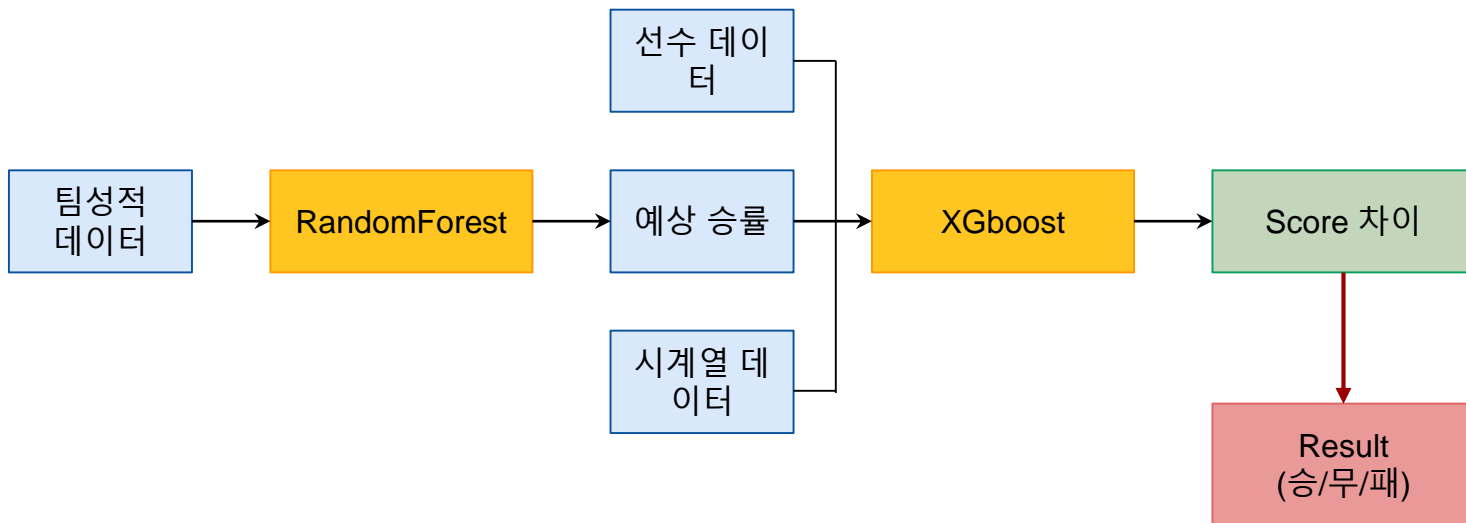
'WHIP' 이닝당 출루허용률

'FIP+' 조정 수비 무관 평균자책점

'WPA' 승리 확률 기여도

다중공선성 문제로 SLG(장타율), FIP, ERA+는 제거



## 승패예측모델 - 모델 설명





# 승패예측모델 - 결과

## 오늘의 경기 예측

 VS   
키움 LG  
선 후라도 선 플렛코



WINNER

 VS   
SSG KIA  
선 김광현 선 양현종





WINNER

 VS   
NC KT  
선 페디 선 벤자민




WINNER

 VS   
삼성 한화  
선 류캐년 선 김민우



WINNER

 VS   
두산 롯데  
선 알칸타라 선 스트레일리



WINNER

# 승패예측모델 - 결과

## 이번시즌 승패 예측



종합	승	무	패
키움	87	1	50
SSG	83	6	49
LG	83	1	53
KIA	82	3	48
NC	80	1	55
두산	70	2	64
KT	68	5	62
삼성	48	3	84
롯데	40	1	91
한화	24	3	109

03

애가 우리팀이었다면!



LG vs KIWOOM

| awayWin: 785 | homeWin: 215

away team -----	home team
Chang-ki Hong 홍창기	Jun-wan Kim 김준완
Hae-min Park 박해민	Jun-tae Park 박준태
Hyun Soo Kim 김현수	Jung-hoo Lee 이정후
Eun-seong Chae 채은성	Hye-seong Kim 김혜성
Ji-hwan Oh 오지환	Yasiel Puig 푸이그
Bo-gyeong Moon 문보경	Tae-jin Kim 김태진
Sung-ju Moon 문성주	Ji-young Lee 이지영
Kang-nam Yoo 유강남	Whee Jip Kim 김휘집
Geon-chang Seo 서건창	Sung-moon Song 송성문
Casey Kelly 켈리	Tyler Eppler 애플러

# 가상 팀 시뮬레이션

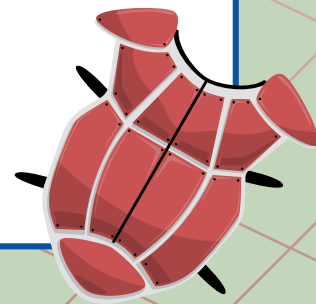


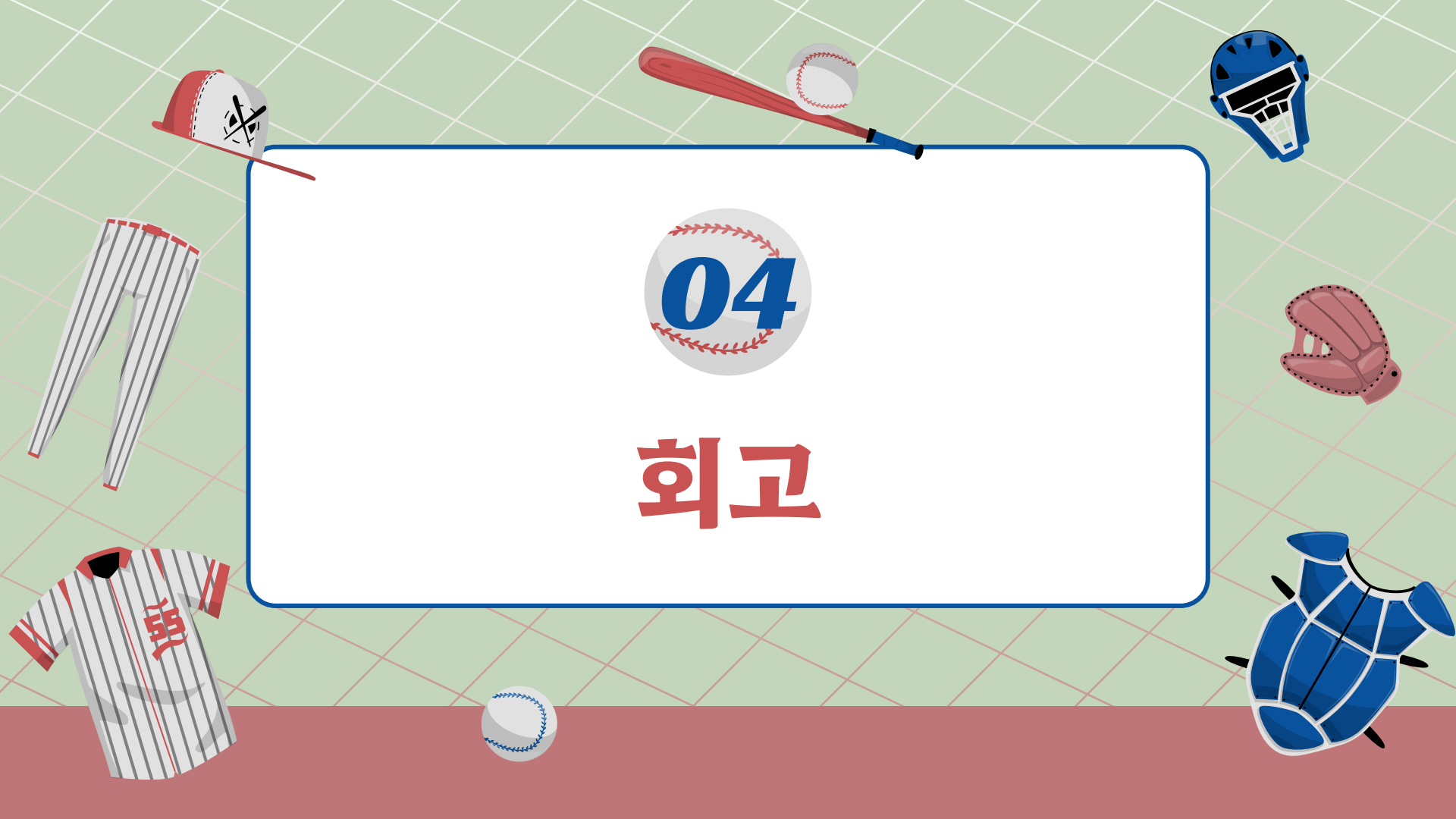
수집한 데이터 중 2022 시즌 선수 종합 기록을 활용한 가상 시뮬레이션 실행

사용데이터 :

- 타자의 타석수와 진루관련 데이터
- 투수의 상대타자수 및 진루관련 데이터
- 리그 전체 타석수 및 진루관련 데이터

선발 투수 및 1-9번 타자 입력 후 해당 데이터로 도출되는 단순한 확률로 결과 출력





04

회고

# 추가 수정 계획



선수 평가  
- 자연어처리 추가



경기데이터  
학습량 확대



가상 팀 시뮬레이션  
→ 개별 프로젝트화

# 개별 회고

## 혜원

과정 처음부터 끝까지 프로젝트  
범위 설정에 애를 먹었다 😞  
디테일한 가설 설정의 중요성..  
그치만 재미있었다☑



## 하진

아픈데도 고생많으셨죠...  
아프지마 도토잠보..



## 상빈

관심있는 분야로 분석해서  
지금껏 프로젝트 중에 가장  
재밌었다



# Resources

## 출처

- 2016~2018 한국프로야구 세이버 메트릭스 지표 분석.pdf
- Use\_of\_Machine\_Learning\_and\_Deep\_Learning\_to\_Predi.pdf
- 기계학습 기법을 이용한 한국프로야구 승패 예측 모델(학위논문).pdf
- 기계학습 기법을 이용한 한국프로야구 승패 예측 모델(학회논문).pdf
- 인공지능 모델에 따른 한국 프로야구의 승패 예측 분석에 관한 연구.pdf
- 한국프로야구 투수의 가치 있는 통계지표 산출 WAR과의 상관성을 중심으로.pdf

## 참고자료

- <https://github.com/yanismiraoui/Analyzing-sports-commentary-in-order-to-automatically-recognize-events-and-extract-insights>
- <https://github.com/numeristical/resources>
- <https://github.com/riceluxs1t/KBOPrediction>
- <https://github.com/junsooshin/baseballsimulator>



# ***Thanks!***

CREDITS:

This presentation template was created by **Slidesgo**,  
including icons by **Flaticon**, infographics and images by **Freepik**

