

# 캡스톤 디자인 #1

머신러닝, 딥러닝을 활용한  
스포츠 승부예측 모델

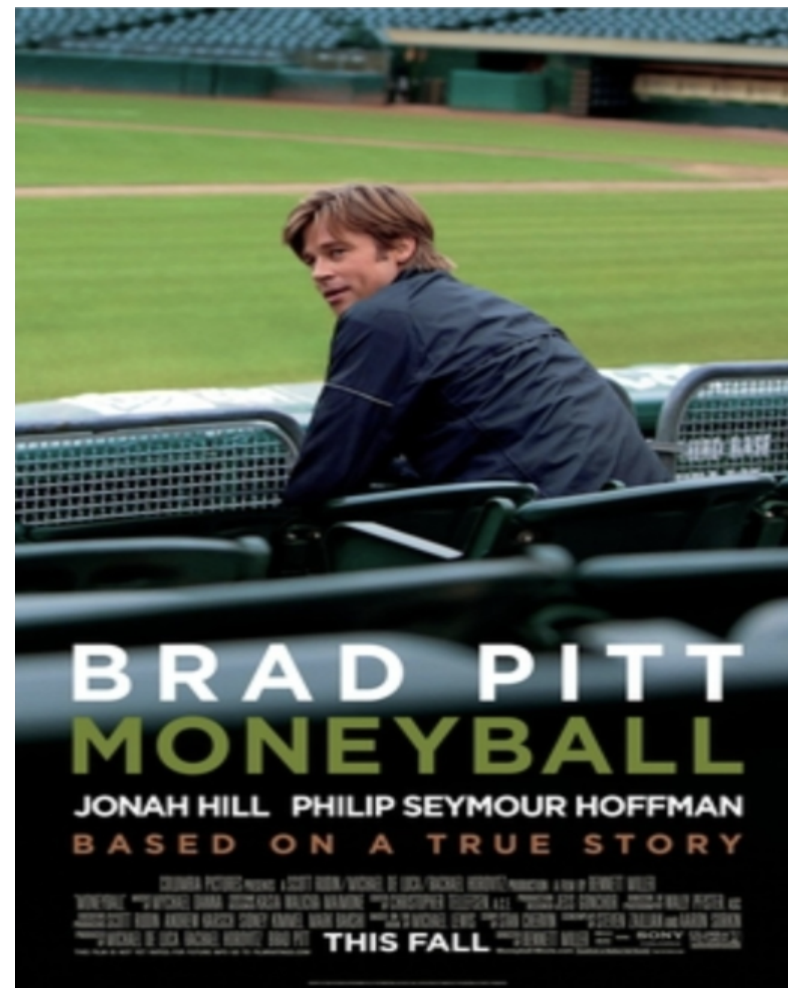


# CONTENS

- 1 주제선정 개요
- 2 기존 문제점
- 3 사용 기법
- 4 타당성
- 5 기대효과

## CONTENS 01

# 주제선정 개요



통계 시스템을 기반으로 한 선수 영입  
데이터, 통계, 경제학적인 수치과정을 가지고 팀을 변화한 사례



스포츠에서 데이터 분석관련 요구의 증가

CONTENS 01

# 주제선정 개요



스쿼드, 선수개인 스탯  
인간의 직감



경기 외적 변수  
통계적 수치



# 기존 문제점



기본적인 야구 데이터(타율, 삼진, 홈런)를 이용하여  
야구 경기의 승패나 승률을 예측



가장 중요한 지표는 **득점과 실점!**  
타율, 삼진, 홈런 등과 같은 기본적인 데이터는  
실제 득점이나 실점과의 상관관계가 정확한 예측이 어렵다

## 기존 문제점



기존 연구에서는 상대팀에 따라 **징크스**와 같은 상대적인 전력의 차이가 발생하는 경향을 고려하지 못함



기존 연구에서 팀별 데이터를 전체적으로 학습하기 때문에 투자 비용, 감독의 전술과 같은 **팀별 특성**을 고려하지 못함

## 사용 기법 (EDA)

Beautifulsoup

Python을 활용해 웹 사이트의 데이터를 추출  
경기기록과 관련된 데이터 추출에 사용

OpenAPI

OpenAPI를 제공하는 사이트를 활용하여  
경기기록과 관련된 데이터 추출에 사용

# 사용 기법 (ML)

Logistic  
Regression

승,패로 나뉜 이진분류 문제에 적합  
데이터의 특성들을 파악하여 회귀예측에 용의

RandomForest  
Classifier

여러 결정 트리의 예측을 평균내어  
최종으로 승패를 예측함에 용의

추후 진행과정에서 다양한 모델을 사용할 예정!



CONTENS 03

# 사용 기법 (DL)

DeepLearning

Dataset의 특성 확인 후  
CNN계열의 모델 or RNN계열의 모델 선택 예정

# 타당성

F1 Score

정밀도와 재현율의 조화평균  
0과 1사이의 값을 가지며 1에 가까운 모델을 선정

Tensorboard

Tensorboard 시각화를 통해  
학습과정에서 최적의 epoch를 결정

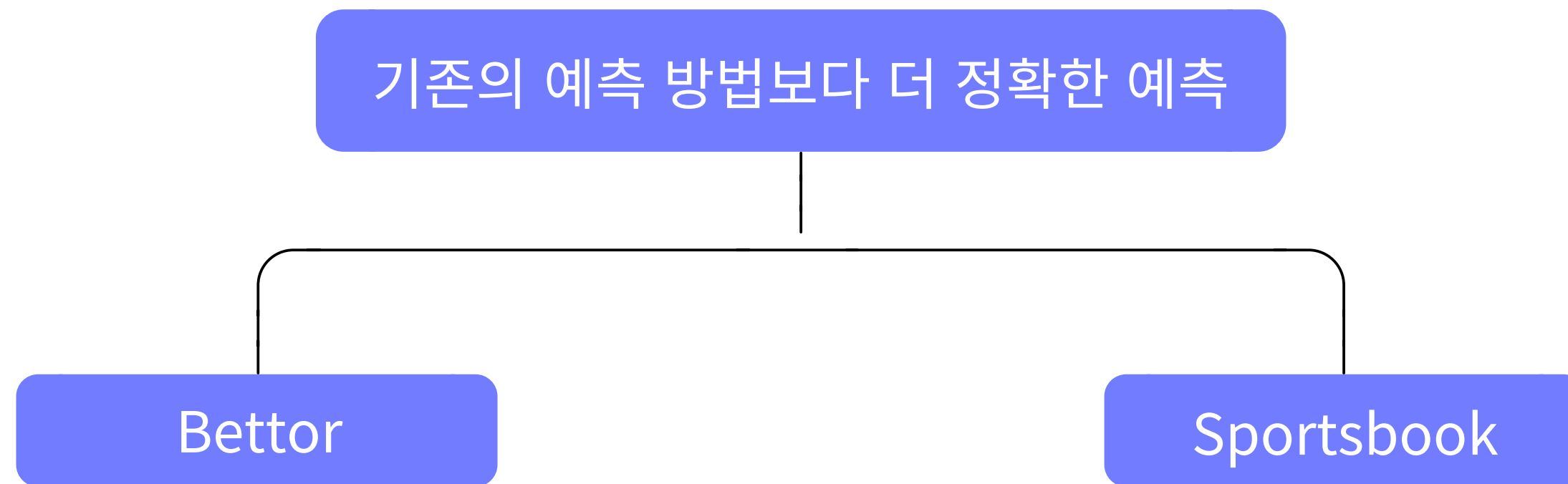
CONTENS 04

# 타당성

New  
Dataset

시즌이 종료되면 종료된 시즌의 데이터를  
train dataset에 추가하여 학습에 반영한다

# 기대효과



기존의 방법보다 더 정확한 예측 결과를 바탕으로  
알맞은 베팅 전략을 통해 더 많은 수익 창출 가능

더 정확해진 예측 결과를 구독 서비스를 통해  
베팅하는 사람들에게 제공하여 수익 극대화

**THANK  
YOU**

