한국빅데이터학회지
제5권 제2호, 2020, pp. 77-84 https://doi.org/10.36498/kbigdt.2020.5.2.77

# 인공지능 모델에 따른 한국 프로야구의 승패 예측 분석에 관한 연구

## A Study on the Win-Loss Prediction Analysis of Korean Professional Baseball by Artificial Intelligence Model

김태훈[1] · 임성원[1] · 고진광[1†] · 이재학[2]

순천대학교 컴퓨터공학과[1], 송원대학교 전기전자공학과[2]

**요 약**

본 연구에서는 인공지능 모델에 따른 한국 프로야구의 승패 예측 분석에 관한 연구를 했다. 승리할 팀과 해당 팀의 최종 리그 순위를 예측했고, 사용자의 편의를 위해 웹사이트도 구축했다. 각 1・3・5이닝 별로 가장 정확도가 높으면서도 오차가 적은 모델을 최적 모델로 선정해 승・패 결과를 예측했고, 이를 토대로 순위표를 작성했다. 결과표는 2020년 개막인 5월 5일부터 8월 30일까지의 예측 결과를 바탕으로 작성했다. 기아타이거즈가 아닌 다른 구단끼리의 경기는 실제 결과를 사용했다. 머신러닝 모델은 KNN과 AdaBoost가 최적 모델로 선정되었으며, 실제 순위와 비교해 본 결과, 경기가 진행될수록, 예측 결과의 순위 오차가 점점 작아지는 것을 확인했다. 딥러닝 모델은 89%의 정확도를 기록했고, 머신러닝 모델과 마찬가지로 경기를 진행할수록 예측 결과 순위 오차가 작아지는 것을 확인했다. 실험 결과는 한국 프로야구 승・패 결과 예측뿐 아니라 다양한 분야에서 사용할 수 있을 것으로 사료된다. 방송국에서 야구 경기를 중계하는 중 이닝별로 인공지능 알고리즘이 예상한 승・패 여부를 중계화면에 띄울 수 있다. 시청자들에게 새로운 흥미를 일으킬 수 있을 것이고, 나아가 구단의 감독들이 이닝마다 데이터를 분석해 경기 중 유동적으로 승리하기 위한 전략을 세울 수 있을 것으로 기대된다.

■ 중심어 : 머신러닝 모델(KNN과 AdaBoost), 딥러닝 모델, 프로야구, 승패 예측

**Abstract**

In this study, we conducted a study on the win-loss predicton analysis of korean professional baseball by artificial intelligence models. Based on the model, we predicted the winner as well as each team's final rank in the league. Additionally, we developed a website for viewers' understanding. In each game's first, third, and fifth inning, we analyze to select the best model that performs the highest accuracy and minimizes errors. Based on the result, we generate the rankings. We used the predicted data started from May 5, the season's opening day, to August 30, 2020 to generate the rankings. In the games which Kia Tigers did not play, however, we used actual games' results in the data. KNN and AdaBoost selected the most optimized machine learning model. As a result, we observe a decreasing trend of the predicted results' ranking error as the season progresses. The deep learning model recorded 89% of the model accuracy. It provides the same result of decreasing ranking error trends of the predicted results that we observe in the machine learning model. We estimate that this study's result applies to future KBO predictions as well as other fields. We expect broadcasting enhancements by posting the predicted winning percentage per inning which is generated

2020년 11월 30일 접수; 2020년 12월 09일 수정본 접수; 2020년 12월 11일 게재 확정.

† 교신저자 (kjg@scnu.ac.kr)

by AI algorism. We expect this will bring new interest to the KBO fans. Furthermore, the prediction generated at each inning would provide insights to teams so that they can analyze data and come up with successful strategies.

■ Keyword : Machine Learning Model(KNN and AdaBoost), Deep Learning Model, Professional Baseball, Win-Loss Prediction

## Ⅰ. 서 론

대한민국에서 가장 인기 있는 스포츠는 야구다. 2008년 베이징 올림픽(Beijing Olympic)에서 전승하며 금메달을 딴 이후 KBO(Korea Baseball Organization)는 폭풍 성장하게 됐고, 2019년 기준 평균 관중 수는 1만 명을 넘겼다. 그에 따라 방영권료(Television Money)와 스폰서(Sponsor) 금액은 야구를 따라올 스포츠가 없게 됐다. 2020년 현재 10개 구단의 KBO 리그와 11개 구단이 북구 리그(Northern League)와 남부 리그(Southern League)로 나눠 진행하고 있는 KBO 퓨처스리그(Futures League)가 있다.

야구는 데이터 싸움으로 알려져 있다. 경기 기록이 승패의 큰 영향을 미치는 야구의 특성상 데이터를 적극적으로 활용하는 것이 좋은 성적으로 이어진다. 2011년 개봉한 영화 머니볼(Money Ball)의 실제 주인공인 오클랜드 애슬레틱스(Oakland Athletics)의 윌리엄 라마 빈(William Lamar Beane), 애칭인 빌리 빈(Billy Beane) 단장은 경기 데이터를 바탕으로 선수들을 영입하고, 그 결과 2002년 메이저리그(Major League) 20연승을 달성했다.

야구경기는 각 선수들의 역할이 명확히 구분되어 있고, 다른 종목에 비해 객관적인 경기기록 분석이 보다 용이한 특성 때문에 다양한 학문분야의 연구자들이 야구경기 기록자료를 이용하여 팀의 경기력 분석 및 승패 또는 성적순위 예측 등 다양한 주제에 연구적 관심을 보이고 있다[5].

국내에서도 데이터마이닝 기법과 야구 데이터를 이용한 한국프로야구 경기의 승패 예측 모형 제안[2][3][4], 프로야구 포스트시즌 진출 예측을 위한 통계적 모형 비교[5], 혼합형 기계 학습 모델을 이용한 프로야구 승패 예측 시스템[6], 기계학습 기법을 이용한 한국프로야구 승패 예측 모델[7][8][9] 등 기계학습 기법을 이용하여 야구 경기 결과를 예측하기 위한 연구가 활발하게 진행되고 있다[7][8][9]. 하지만 아직 최적의 결과를 얻지 못하고 있고, 이처럼 승패 예측이 어려운 이유는 많은 경기 기록들 중 승패 예측에 영향을 주는 요소의 선별이 어렵고, 예측에 사용된 자료들 간의 중복 요인으로 인해 학습 모델이 좋은 성능을 보이지 못하고 있기 때문이다[6]. 딥 러닝 신경망은 기계학습 분야 중의 하나로 대량의 데이터를 학습하기 위해 다단계의 신경망을 구성하여 데이터를 학습하는 기술을 의미한다[7][8][9][10]. 딥 러닝 기법은 최근에는 데이터의 급격한 증가로 인해 고도의 학습 기술이 필요하게 되면서 다양한 분야에서 활용되고 있다[7][8][9][10].

본 논문에서는 승패 예측에 영향을 주는 요소의 최적화를 위해 세이버메트릭스(Saber Metrics) 등을 사용하였고, 인공지능(Artificial Intelligence) 분야의 머신러닝(Machine Learning) 및 딥러닝(Deep Learning) 알고리즘을 이용해 경기 결과 예측을 위한 최적모델을 선정했다.

경기의 1・3・5 이닝 데이터를 이용해 승리팀을 예측하여 실제 경기 결과와 정확도를 비교했다. 나아가 리그 순위도 예측했다. 입력값을 이용해 결과를 반환하는 플라스크(Flask) 기반

웹페이지도 구현했다.

실험 데이터는 KBO 2016년 1월부터 2020년 8월 30일까지 기아 타이거즈(KIA Tigers)의 경기 데이터를 이용했다. 예측순위 결과표의 기아 타이거즈가 아닌 구단끼리의 경기 결과는 실제 경기 결과로 사용했다.

## Ⅱ. 인공지능 모델을 이용한 한국프로야구의 승패 예측

### 2.1 실험에 이용된 머신러닝, 딥러닝 모델

머신러닝 모델은 에이다 부스트(Adaptive Boosting, AdaBoost), K-최근접 이웃(K-Nearest Neighbor, KNN), 결정 트리(Decision Tree), 랜덤 포레스트(Random Forest), 그라디언트 부스팅(Gradient Boosting)을 선정했고, 딥러닝 모델은 텐서플로(Tensorflow) 라이브러리를 이용해 커스터마이징(Customizing) 모델을 설계했다.

2016년부터 2019년까지의 경기 데이터는 훈련 데이터(Training Data)로, 2020년도 데이터는 테스트 데이터(Test Data)로 사용했다.

심층 신경망(Deep Neural Network)은 입력층, 은닉층(Hidden Layer), 출력층으로 구성된다[2]. 입력층의 활성 함수(Activation Function)는 정규화 선형 유닛(Rectifier Linear Unit, ReLU)을 사용했고, 출력층은 시그모이드(Sigmoid)를 사용했다.

### 2.2 플라스크를 이용한 웹페이지 구축

그림 1은. 웹페이지 레이아웃(Layout) 및 실행 화면이다. 웹페이지(Web Page)는 경기통계를 입력받고, 승패 결과를 예측한다. 프로그래밍 언어는 파이썬(Python)을 이용했기 때문에 웹페이지를 플라스크로 구현했다.

사용할 머신러닝 또는 딥러닝 모델과 경기 이닝을 선택하고, 홈런, 안타, 삼진, 볼넷, 병살, 실책, 점수와 홈/어웨이 여부를 입력하면 서버로 전송된다. 선택한 모델은 플라스크 서버에서 입력값을 받아 승·패 여부를 반환한다. 반환한 값은 웹페이지로 전송돼 기아타이거즈와 상대팀의 결과를 표시한다.

웹페이지에서 SLG(Sluggiing Percentage)는 장타율을 뜻하고, OBP(On Base Percentage)는 출루율을 뜻한다. BAT는 타율을 뜻하고, OPS (On-base Plus Slugging)는 SLG와 OBP를 더한 값이다.

SLG, OBP, BAT, OPS(On-base Plus Slugging)를 구하기 위해서는 타석에 들어가서 타격을 한 횟수를 뜻하는 타수가 필요한데, 입력받는 곳에서는 타수를 일정한 값으로 넣기 어려우므로, 웹페이지에서는 SLG, OBP, BAT, OPS를 제외한 입력값을 통해 예측하도록 구현하였다.


〈Fig. 1〉 웹페이지 레이아웃(Layout) 및 실행 화면

## Ⅲ. 실험 결과

### 3.1 이닝별 데이터셋 구성

Fig. 2와 같이 기아타이거즈 공식 홈페이지에 게시된 이닝별 데이터를 파이썬(Python)의 셀레니움(Selenium) 프레임워크를 사용해 웹 크롤링(Web Crawling)했다. 크롤링한 데이터는 하나의 *.CSV(Comma Separated Value) 파일로 통합했다.

경기 날짜와 상대 팀, 기아타이거즈와 상대 팀의 이닝별 삼진, 볼넷, 실책, 안타, 홈런, 병살타, 점수, 홈/어웨이 여부, 승패 결과로 구성했다.


KIA 타자 기록

| 라인업 | 1회 | 2회 | 3회 | 4회 | 5회 | 6회 | 7회 | 8회 | 9회 | 타수 | 안타 | 타점 | 득점 | 타율 시즌 | 타율 5게임 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 二 최정원 | 2땅 | | 유땅 | | 포번 | 4구 | | 2땅 | | 4 | 0 | 0 | 0 | 0.111 | 0.000 |
| 우 이정원 | 4구 | | 중안 | | 좌안 | 2땅 | | | 삼진 | 4 | 2 | 1 | 1 | 0.227 | 0.500 |
| - 황대인 | 좌안 | | | | | | | | | 1 | 1 | 0 | 0 | 0.276 | 0.583 |

〈Fig. 2〉 기아 타이거즈 공식 홈페이지 경기통계[1]

### 3.2 데이터 분석

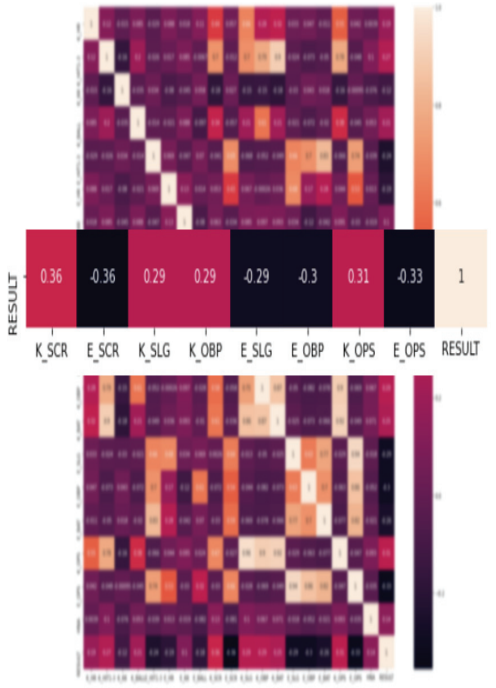
타자와 관련된 파라미터(Parameter)는 장타율이, 투수와 관련된 파라미터는 선발투수의 승률과 볼넷/이닝이 통계적으로 유의하게 나타났다[2]. 이후 세이버메트릭스(Saber Metrics)를 통해 검증된 장타율, 출루율인 SLG, OBP 파라미터를 만들었다. 이에 1, 3, 5이닝 별로 BAT, OPS 파라미터를 추가하여 Fig. 3과 같이 데이터셋을 수정했다. Fig. 4는 완성된 데이터셋의 파라미터들이다.

| Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|
| K_SLG | K_OBP | K_BAT | E_SLG | E_OBP | E_BAT | K_OPS | E_OPS | HNA | RESULT |
| 0.35 | 0.216667 | 0.15 | 0.65 | 0.386667 | 0.28 | 0.566667 | 1.036667 | 0 | 0 |
| 0.236667 | 0.196667 | 0.196667 | 0.216667 | 0.166667 | 0.166667 | 0.433333 | 0.383333 | 0 | 1 |
| 0.233333 | 0.233333 | 0.233333 | 0.2 | 0.266667 | 0.3 | 0.466667 | 0.466667 | 1 | 1 |
| 0.066667 | 0.066667 | 0.066667 | 0.13 | 0.16 | 0.15 | 0.133333 | 0.29 | 1 | 0 |
| 0.4 | 0.36 | 0.326667 | 0.386667 | 0.413333 | 0.313333 | 0.76 | 0.8 | 0 | 0 |
| 0.25 | 0.266667 | 0.216667 | 0.35 | 0.266667 | 0.266667 | 0.516667 | 0.616667 | 0 | 1 |
| 0.43 | 0.323333 | 0.26 | 0.418095 | 0.326111 | 0.309048 | 0.753333 | 0.744206 | 0 | 0 |
| 0.57 | 0.31 | 0.18 | 0.38 | 0.347619 | 0.28 | 0.88 | 0.727619 | 0 | 1 |

〈Fig. 3〉 추가된 컬럼 데이터셋


〈Fig. 4〉 데이터셋 파라미터 목록

위 데이터셋을 이용해 Fig. 5와 같이 히트맵(Heat Map)[11]을 그렸고, 승패와 파라미터들의 상관관계를 파악했다.




〈Fig. 5〉 히트맵

승패에 가장 영향을 끼쳤던 파라미터는 해당 이닝까지의 점수(K_SCR, E_SCR)다. 장타율(K_SLG, E_SLG)과, 출루율(K_OBP, E_OBP), OPS(K_OPS, E_OPS) 파라미터는 그 뒤를 따랐다.

### 3.3 이닝별 최적모델 선정

예측 결과, 1이닝에서는 KNN이, 3・5이닝에서는 Ada Boost가 훈련정확도와 테스트정확도가 가장 높았고, 오차도 가장 낮았다. 따라서 Ada Boost를 최적 모델로 선정했다.

〈표 1〉 모델별 훈련 정확도와 테스트 정확도(1이닝)

| Model | Train Accuracy[%] | Test Accuracy[%] |
|---|---|---|
| **KNN** | **63.39** | **61.17** |
| GB | 62.34 | 58.82 |
| AB | 61.29 | 55.29 |
| DT | 64.27 | 61.17 |
| Proposed Model | 58.32 | 57.16 |

〈표 2〉 모델별 훈련 정확도와 테스트 정확도(3이닝)

| Model | Train Accuracy[%] | Test Accuracy[%] |
|---|---|---|
| KNN | 72.15 | 67.05 |
| GB | 80.91 | 65.88 |
| AB | 73.73 | 71.76 |
| DT | 72.32 | 64.70 |
| **Proposed Model** | **76.32** | **75.19** |

〈표 3〉 모델별 훈련 정확도와 테스트 정확도(5이닝)

| Model | Train Accuracy[%] | Test Accuracy[%] |
|---|---|---|
| KNN | 77.75 | 72.94 |
| GB | 78.98 | 77.64 |
| AB | 82.83 | 81.17 |
| DT | 85.28 | 78.82 |
| **Proposed Model** | **89.48** | **89.41** |

딥러닝 텐서플로 라이브러리를 이용해 설계한 커스터마이징 모델의 훈련정확도는 89.48%, 테스트정확도는 89.41%를 보였다.

Fig. 6을 보면 경기를 진행할수록 정확도가 상승한다. 1이닝을 제외하면 딥러닝 모델이 대체로 머신러닝 모델보다 높은 정확도를 보였다.

| Date | Enemy | 승/패 | 실제 승/패 |
|---|---|---|---|
| 5.5 | 키움 | 패 | 패 |
| 5.6 | 키움 | 승 | 패 |
| 5.7 | 키움 | 패 | 승 |
| 5.8 | 삼성 | 패 | 패 |
| 5.9 | 삼성 | 패 | 패 |
| 5.10 | 삼성 | 패 | 승 |
| 5.11 | 한화 | 패 | 승 |
| 5.12 | 한화 | 승 | 승 |
| 5.13 | 한화 | 패 | 패 |
| 5.14 | 두산 | 승 | 패 |

〈Fig. 6〉 KNN을 이용한 최종결과

검증을 위해 Fig. 7과 같이 2020년도 실제 경기의 통계로 예측 결과와 실제 결과를 비교했다.

제안한 모델은 가장 정확도가 높았고, 실제 순위와 1순위 차이를 보였다. 다른 모델들도 이닝별로 정확도가 더 상승했다.

| | 실제 순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | NC | 56 | 34 | 0.62222 |
| 2 | 키움 | 59 | 40 | 0.59596 |
| 3 | LG | 55 | 40 | 0.57895 |
| 4 | 두산 | 52 | 41 | 0.55914 |
| 5 | KT | 49 | 43 | 0.53261 |
| 6 | 롯데 | 47 | 43 | 0.52222 |
| **7** | **KIA** | **48** | **45** | **0.51613** |
| 8 | 삼성 | 43 | 51 | 0.45745 |
| 9 | SK | 32 | 63 | 0.33684 |
| 10 | 한화 | 26 | 67 | 0.27957 |

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | NC | 57 | 35 | 0.61957 |
| 2 | 키움 | 61 | 40 | 0.60396 |
| 3 | KT | 53 | 41 | 0.56383 |
| 4 | LG | 54 | 43 | 0.5567 |
| 5 | 두산 | 52 | 43 | 0.54737 |
| 6 | 롯데 | 50 | 42 | 0.54348 |
| **7** | **KIA** | **45** | **42** | **0.51724** |
| 8 | 삼성 | 49 | 47 | 0.51042 |
| 9 | SK | 35 | 62 | 0.36082 |
| 10 | 한화 | 30 | 65 | 0.31579 |

〈Fig. 7〉 딥러닝 모델의 실제 순위와 예측순위

## Ⅳ. 결 론

본 연구에서는 인공지능 모델에 따른 한국 프로야구의 승패 예측 분석에 관한 연구를 했다. 승리할 팀과 해당 팀의 최종 리그 순위를 예측했고, 사용자의 편의를 위해 웹사이트도 구축했다.

각 1・3・5이닝 별로 가장 정확도가 높으면서도 오차가 적은 모델을 최적 모델로 선정해 승・패 결과를 예측했고, 이를 토대로 순위표를 작성했다. 결과표는 2020년 개막인 5월 5일부터 8월 30일까지의 예측 결과를 바탕으로 작성했다. 기아타이거즈가 아닌 다른 구단끼리의 경기는 실제 결과를 사용했다.

머신러닝 모델은 KNN과 AdaBoost가 최적 모델로 선정되었으며, 실제 순위와 비교해본 결과, 경기가 진행될수록, 예측 결과의 순위 오차가 점점 작아지는 것을 확인했다.

딥러닝 모델은 89%의 정확도를 기록했고, 머신러닝 모델과 마찬가지로 경기를 진행할수록 예측 결과 순위 오차가 작아지는 것을 확인했다.

본 논문의 실험 결과는 한국 프로야구 승・패 결과 예측뿐 아니라 다양한 분야에서 사용할 수 있을 것으로 사료된다. 방송국에서 야구 경기를 중계하는 중 이닝별로 인공지능 알고리즘이 예상한 승・패 여부를 중계화면에 띄울 수 있다. 시청자들에게 새로운 흥미를 일으킬 수 있을 것이고, 나아가 구단의 감독들이 이닝마다 데이터

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | 키움 | 64 | 37 | 0.633 |
| 2 | NC | 55 | 37 | 0.597 |
| **3** | **KIA** | **48** | **45** | **0.516** |
| 4 | 롯데 | 51 | 41 | 0.554 |
| 5 | 두산 | 53 | 42 | 0.557 |
| 6 | KT | 49 | 45 | 0.521 |
| 7 | LG | 49 | 48 | 0.505 |
| 8 | 삼성 | 46 | 50 | 0.479 |
| 9 | SK | 31 | 66 | 0.319 |
| 10 | 한화 | 30 | 65 | 0.315 |

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | NC | 56 | 36 | 0.608 |
| 2 | 키움 | 61 | 40 | 0.603 |
| 3 | 롯데 | 50 | 42 | 0.543 |
| **4** | **KIA** | **52** | **51** | **0.504** |
| 5 | LG | 50 | 47 | 0.515 |
| 6 | KT | 49 | 45 | 0.521 |
| 7 | 두산 | 49 | 46 | 0.515 |
| 8 | 삼성 | 44 | 52 | 0.458 |
| 9 | SK | 35 | 62 | 0.360 |
| 10 | 한화 | 30 | 65 | 0.315 |

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | NC | 57 | 35 | 0.619 |
| 2 | 키움 | 60 | 41 | 0.594 |
| 3 | 두산 | 53 | 42 | 0.557 |
| 4 | KT | 51 | 43 | 0.542 |
| 5 | LG | 52 | 45 | 0.536 |
| **6** | **KIA** | **49** | **44** | **0.526** |
| 7 | 롯데 | 48 | 44 | 0.521 |
| 8 | 삼성 | 44 | 52 | 0.458 |
| 9 | SK | 33 | 64 | 0.340 |
| 10 | 한화 | 29 | 66 | 0.305 |

〈Fig. 8〉 최적 머신러닝 모델의 예측순위

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | 키움 | 63 | 38 | 0.623 |
| 2 | NC | 56 | 36 | 0.608 |
| 3 | 두산 | 52 | 43 | 0.547 |
| 4 | 롯데 | 50 | 42 | 0.543 |
| 5 | LG | 51 | 46 | 0.525 |
| 6 | KT | 47 | 47 | 0.5 |
| 7 | 삼성 | 46 | 50 | 0.479 |
| **8** | **KIA** | **46** | **47** | **0.494** |
| 9 | SK | 34 | 63 | 0.350 |
| 10 | 한화 | 31 | 64 | 0.326 |

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | 키움 | 61 | 36 | 0.628 |
| 2 | NC | 57 | 35 | 0.619 |
| 3 | LG | 54 | 43 | 0.556 |
| 4 | 두산 | 52 | 43 | 0.547 |
| 5 | KT | 51 | 43 | 0.542 |
| 6 | 롯데 | 49 | 43 | 0.532 |
| **7** | **KIA** | **48** | **45** | **0.516** |
| 8 | 삼성 | 43 | 53 | 0.447 |
| 9 | SK | 36 | 61 | 0.371 |
| 10 | 한화 | 30 | 65 | 0.315 |

| | 예측순위 | 승 | 패 | 승률 |
|---|---|---|---|---|
| 1 | NC | 57 | 35 | 0.619 |
| 2 | 키움 | 61 | 40 | 0.603 |
| 3 | KT | 53 | 41 | 0.563 |
| 4 | LG | 54 | 43 | 0.556 |
| 5 | 두산 | 52 | 43 | 0.547 |
| 6 | 롯데 | 50 | 42 | 0.543 |
| **7** | **KIA** | **45** | **42** | **0.517** |
| 8 | 삼성 | 49 | 47 | 0.510 |
| 9 | SK | 35 | 62 | 0.360 |
| 10 | 한화 | 30 | 65 | 0.315 |

〈Fig. 9〉 제안하는 모델의 예측순위

를 분석해 경기 중 유동적으로 승리하기 위한 전략을 세울 수 있을 것으로 기대된다. 추 후 야구의 승패를 보다 더 정확히 예측할 수 있도록 승패에 영향을 주는 데이터를 다양하게 보완하여 연구할 필요가 있다.

## 참 고 문 헌

[1] https://tigers.co.kr/game/schedule/view?type=major&gameKey=20201031NCHT0&gameDate=20201031

[2] Younhak Oh, Han Kim, Jaesub Yun and Jong-Seok

Lee, “Using Data Mining Techniques to Predict Win-Loss in Korean Professional Baseball Games”, Korean Institute of Industrial Engineersm, Vol. 40, No. 1, pp. 8-17, 2014.

[3] 오광모, 이장택, “데이터마이닝을 이용한 한국 프로야구 선수들의 연봉에 관한 모형연구”, 한국스포츠사회학회지, Vol. 16, No 2. pp.2-310, 2003.

[4] Miljkovic, D., “The use of data mining for basketball matches outcomes prediction” pp.309-312, 2010.

[5] 채진석, 조은형, 엄한주, “프로야구 포스트시즌 진출 예측을 위한 통계적 모형 비교”, 한국체육측정평가학회지, Vol 12, No. 1, pp.33-48, 2010.

[6] 홍석미, 정경숙, 정태충, “혼합형 기계 학습 모델을 이용한 프로야구 승패 예측 시스템”, 한국정보과학회, Vol. 9, No. 6, pp.693-698, 2003.

[7] 서영진, 문형우, 우용태, “기계학습 기법을 이용한 한국프로야구 승패 예측 모델”, 한국컴퓨터정보학회, Vol 24, No 2, pp.17-24, 2019.

[8] 노언석, 최재현, “기계학습을 활용한 프로야구 승부예측에 관한 연구”, 한국IT정책경영학회논문지, Vol. 9, No. 1, pp.335-338, 2017.

[9] Eonseok No, “A Study of KBO Professional Baseball Game Prediction using Artificial Neural Networks”, Thesis, p.5, 2017.

[10] Sung Eun Moon, Soo Beom Jang, Jung Huk Lee, Jong Seok Lee, “Machine Learning and Deep Learning Technology Trends”, Journal of korea institute of communication sciences, Vol. 33, No. 10, pp.49-56, 2016.

[11] 김도엽, 장주용, “얼굴 특장점 검출을 위한 적분 회귀 네트워크”, 한국방송·미디어공학회, Vol. 24, No. 4, pp.564-572, 2019.

## 저 자 소 개

**김 태 훈(Tae-Hun Kim)**
·2021년 순천대학교 컴퓨터공학과 졸업(공학사)
·관심분야: 딥러닝, 머신러닝, 빅데이터

**임 성 원(Seong-Won Lim)**
·2021년 순천대학교 컴퓨터공학과 졸업(공학사)
·관심분야: 딥러닝, 머신러닝, 빅데이터

**고 진 광(Jin-Gwang Koh)**
·1997년: 홍익대학교 컴퓨터공학과 (이학박사)
·2014년: 죠지아공대(GIT) 방문교수
·1988년~현재: 순천대학교 컴퓨터공학과 교수
·관심분야: 데이터베이스, USN

**이 재 학(Jae-Hak Lee)**
·2005년: 중앙대학교 전기공학과 (공학박사)
·2017년~현재: 송원대학교 전기전자공학과 교수
·관심분야: IoT Control, ICT, 인공지능