

## 불쾌지수가 외야수의 경기 기록 데이터에 미치는 영향

김세민<sup>1</sup> · 신좌철<sup>2\*</sup>

### The Effect of Discomfort Index on Outfielder's Game Record Data

Semin Kim<sup>1</sup> · Chwa-Cheol Shin<sup>2\*</sup>

<sup>1</sup>Adjunct Professor, Department of Computer Education, Jeonju National University of Education, Jeonju, 55101 Republic of Korea

<sup>2\*</sup>Assistant professor, Department of Innovation and Convergence, Hoseo University, Chongcheong-namdo, 31499, Republic of Korea

#### 요 약

본 연구에서는 빅 데이터 분석 방법을 활용하여 스포츠 기록과 기상 데이터의 상관관계를 분석하였다. 이를 위하여 API와 크롤링으로 데이터를 수집하고, 가공한 다음 이것을 토대로 통계를 낸 후에 데이터 시각화를 진행하였다. 본 연구의 대상으로는 2019년 KBO 리그에 출전한 외야수 중에서 규정타석에 진입한 선수를 대상으로 하였다. 또한 기상 데이터는 불쾌지수를 활용하였으며 70이상과 70미만을 기준으로 하여 분석하였다. 연구 결과로는 투수가 개입하는 기록인 각종 타격지표에서는 불쾌지수가 높을수록 외야수의 기록이 향상하였으나, 투수가 개입하지 않는 도루, 볼넷, 투구 수, 도루 성공률, 타석 당 투구 수, 경기 당 투구 수 등의 기록에서는 외야수가 투수를 힘들게 하였음을 알 수 있었다. 본 연구를 통하여 스포츠 데이터 산업의 발전과 야구 선수, 구단, 코칭스태프의 경기력에 도움이 될 것으로 기대한다.

#### ABSTRACT

In this study, the correlation between sports records and weather data was analyzed using the big data analysis method. To this end, data was collected by API and crawling, data was processed, statistics were performed, and data visualization was performed. The subject of this study was a player who entered the regular at-bat among outfielders in the 2019 KBO League. In addition, meteorological data were analyzed by using the unpleasant index and above 70 and below 70. As a result of the study, in the various hitting indicators, which are the records that pitchers intervene, the higher the unpleasant index, the better the outfielder's record, but pitchers, walks, pitches, pitching success rates, pitches per turn, pitches per game. From the records of the back, it was found that the outfielder made the pitcher difficult. It is expected that this study will help the development of the sports data industry and the performance of baseball players, baseball teams, and coaching staff.

**키워드** : 빅 데이터 분석, 스포츠 데이터, 야구 기록, 불쾌지수

**Keywords** : Big Data Analysis, Sports Data, Baseball Records, Discomfort Index

Received 28 June 2020, Revised 30 June 2020, Accepted 7 July 2020

\* Corresponding Author Chwa-Cheol Shin(E-mail:ccshin@hoseo.edu, Tel:+82-42-540-9658)

Assistant professor, Department of Innovation and Convergence, Hoseo University, Asan, 31499 Republic of Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.8.978>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

현 시대의 인류는 삶의 질을 높이는 것을 추구하고 있으며, 이를 위하여 건강과 여가 관련 콘텐츠들이 주목을 받고 있다. 스포츠 산업은 IT 엔터테인먼트와 건강관련 산업에서 많은 접점이 있으므로 빅데이터 시장에서 부가가치 창출에 대한 기대가 매우 큰 유망한 분야이다. 또한 스포츠에 참여하는 선수, 구단, 방송 매체 등은 더 좋은 경기력을 발휘하기 위하여 날이 갈수록 데이터에 많은 관심을 표하고 있다. 또한 데이터는 스포츠팬들의 여러 가지 궁금증을 해소하여주고, 새로운 콘텐츠를 제공하는 역할을 하기도 한다[1].

이에 상당히 많은 수의 프로스포츠 구단은 효율성과 정확성을 기초로 한 데이터를 경기력 향상에 활용하려는 노력을 기울이고 있다. 야구에서는 세이버메트릭스(Sabermetrics)가 있고, 농구에서는 APBR메트릭스(APBRmetrics)가 그 예이다. 그러나 이와 같은 2차 기록들은 경기장에서 벌어지는 이벤트에 기반하여 작성되는 경우가 많다. 경기장에서는 경기 이외에 많은 환경 변수가 도사리고 있으며, 특히 야구에서는 야외에서 경기하는 경우가 많으므로 자연이 주는 환경 변수를 전혀 무시할 수 없다[2].

이에 본 연구에서는 KBO 리그의 규정 타석에 진입한 외야수의 주루플레이 데이터와 타석플레이 데이터를 중심으로 볼배지수와 상관계수를 분석하고자 한다. 외야수를 연구 주제로 선정한 이유는 다른 포지션과는 달리 팀 내에서 주전 경쟁이 치열하고 스트레스를 많이 받는 포지션이다. 또한 외야수는 공격력과 주루플레이가 좋아야 하는 것은 물론이고, 수비에서도 한 번의 실수가 실점으로 직결되는 확률이 높은 만큼 부담을 많이 가지는 선수들이다. 가장 큰 이유로는 외야수들은 운동장에서 가장 넓은 구역을 담당하고 있으며, 추울 때는 넓은 곳에서 외롭게 바람을 맞으며 서 있고, 더울 때는 그늘 한 점 없는 피약벌에서 땀을 흘리며 뛰어다녀야 한다. 따라서 본 연구에서는 볼배지수가 외야수들에게 어떤 영향을 주는지 의문을 가질 수 있었다.

본 논문의 구조는 다음과 같다. 2장의 관련 연구에 이어, 3장에서는 연구 절차에 대하여 기술하고, 4장에서는 연구 결과를 기술한 후, 5장에서 결론 및 제언으로 맺는다.

## II. 관련 연구

### 2.1. 스포츠 과학과 데이터 과학

스포츠 분야에서 데이터를 경기력을 높이는데 활용 한 예가 몇 년 전부터 증가하고 있다.

2014년 브라질 FIFA 월드컵 우승팀인 독일 대표팀은 빅데이터 전문 회사인 SAP와 제휴하여 팀의 경기 내용부터 선수의 건강 및 생활 데이터까지 분석하여 팀을 관리하였다[3].

물론 야구에서도 데이터를 활용하는 추세로 진행되고 있다. 야구는 종목의 특성 때문에 많은 정형 데이터가 산출된다. 매 경기마다 누적 기록인 홈런, 안타, 타점, 도루, 삼진, 승리, 세이브, 홀드 등이 축적되고 있으며 이를 바탕으로 비율 기록인 타율, 출루율, 장타율, 평균자책점, 승률 등이 갱신된다[4]. 또한 야구에서는 정형 데이터뿐만 아니라 투구 궤적, 타구 발사각도, 투구 자세 등 영상에서 출력되는 비정형 데이터를 트랙맨(Trackman)이라는 소프트웨어를 활용하여 정형 데이터 형태로 출력하여 경기력 향상을 위하여 노력하고 있다. 또한 야구 기록을 과학적으로 통계분석한 데이터인 세이버메트릭스(Sabermetrics)를 활용하기도 하고, IoT 장치를 활용하여 상황 정보 데이터를 활용하는 등의 노력을 하고 있다. 이를 위하여 각 구단에서는 데이터 사이언티스트를 적극 고용하여 활용하고 있다[5-7].

### 2.2. 빅 데이터 분석 연구 모형

본 연구에서는 SAS(Statistical Modification System)사에서 고안한 SEMMA(Sampling Exploration Modification Modeling Assesment)모형을 데이터 분석 방법론으로 활용하였다. 데이터를 분석하는 단계는 분석-설계-수집-구현-검증의 단계를 거치며 각 단계에 적용하는 SEMMA 모형은 표 1과 같다[8].

Table. 1 SEMMA Model for Big Data Analysis

Level	Contents	Application
Sampling	Utilization of Statistics Technique, Data Sampling	Requirement Analysis, Design, Data Collection
Explore	Data exploration and analysis	Requirement Analysis, Design
Modify	Quantification, Standardization,	Requirement Analysis, Design,

Level	Contents	Application
	Translation, Grouping etc.	Data Collection, Development
Modeling	Classification Clustering, Modeling etc.	Design, Development, Verification
Assesment	Model Validation, Cross Analysis, Feedback etc.	Design, Development, Verification

### 2.3. 기상 데이터 활용

인간의 삶과 문화는 기후와 날씨에 크게 영향을 받는다. 그래서 많은 연구자들이 기후 특성이 인간의 생활에 어떻게 영향을 주는지 연구하였다.[9] 특히 야구는 대표적인 야외 스포츠 종목이므로 극히 일부분의 돐 구장을 제외하고는 기상 조건에 크게 영향을 받는다.

야외에서 경기를 하는 경우가 많은 야구 선수들도 기상 조건에 따라 경기력에 영향을 받는다. 특히 기온과 습도 등은 야수나 투수들의 경기력에 영향을 줄 수 있다 [5, 10].

본 연구에서 연구하려는 외야수 포지션의 선수들은 경기에 임할 때 다른 포지션의 선수보다 넓은 거리를 뛰어다녀야 한다. 또한 그늘 한 점 없는 곳에서 직사광선을 받으며 경기를 하고 있다. 대한민국의 여름은 글러브

하나만 들고 뛰어다녀도 무더운 계절이다. 우리나라에서는 날씨가 더운 6월에서 9월 사이에 불쾌지수를 측정하여 발표한다. 불쾌지수는 우리 생활 속에서 가장 많이 언급하는 대표적인 생활 기상 지수이다. 따라서 대표적인 야외 스포츠이자 멘탈 스포츠라고 불리우기도 하는 야구에서 불쾌지수는 중요한 변수가 될 수 있음에 주목하였다. 불쾌지수는 70 이상에서부터 불쾌감을 느끼는 사람이 나타나기 시작하며, 80 이상에서는 반 이상의 사람이 불쾌감을 나타내고, 83 이상에서는 대부분의 사람이 불쾌감을 느낀다고 한다[11, 12].

## III. 연구 절차

### 3.1. 분석 대상 데이터 수집 및 생성

본 단계에서는 분석하고자 하는 대상에 해당하는 데이터를 수집하였다. 본 연구에서 필요한 데이터는 API와 Python Crawling을 활용하여 스탯티즈(그림 1, <http://www.statiz.co.kr>)에서 규정 타석에 진입한 외야수의 일일 경기기록을 수집하였으며, 기상자료개방포털(그림 2, <https://data.kma.go.kr/>)에서 각 구장별 불쾌지수 데이터를 수집하여 CSV파일로 변환하였다.

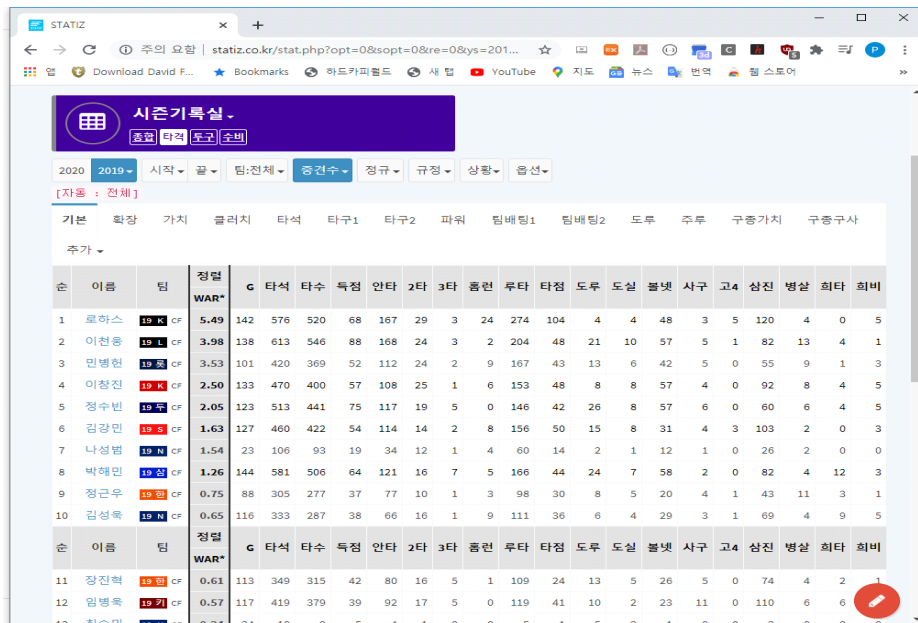


Fig. 1 Statiz Official Website

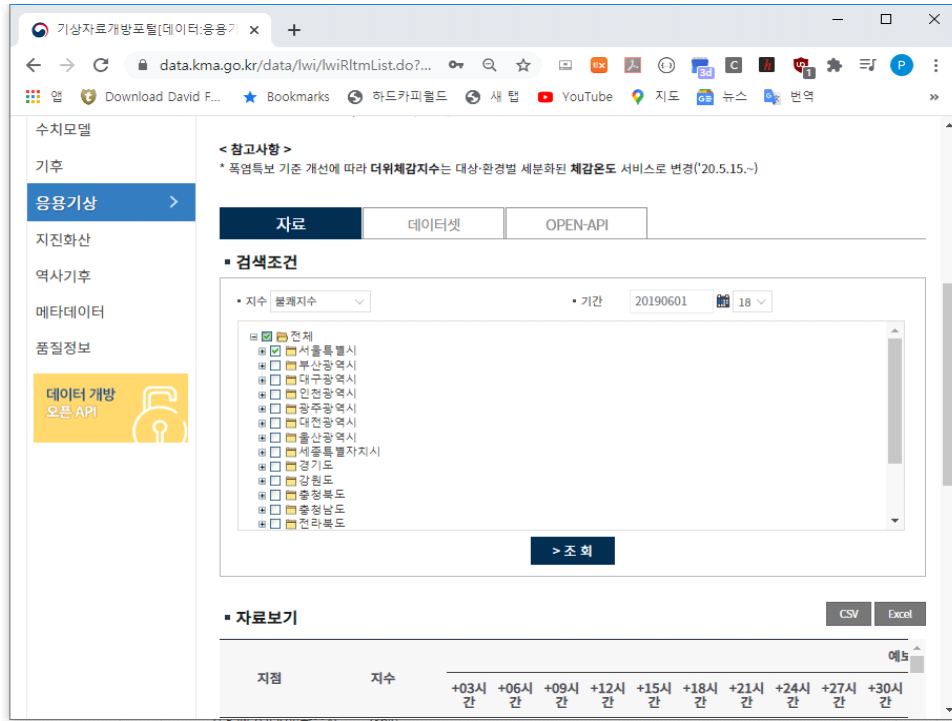


Fig. 2 Korea Weather Data Open Portal Site

### 3.2. 분석 데이터 탐색

본 단계에서는 데이터를 수량화하여 다양한 통계량을 나타내려고 데이터를 탐색한 후 선별하고 전처리하였다. 따라서 본 단계에서 연구 대상을 정하여야 한다. 본 연구에서 대상이 되는 포지션은 2019년 KBO리그의 정규시즌의 규정 타석에 진입한 외야수를 대상으로 하였다. 그 이유는 규정 타석에 진입한 외야수들은 대체적으로 한 시즌 동안 꾸준히 경기에 출전한 선수들이어서 데이터가 일정할 것이고, 그렇지 못한 선수들은 어느 한 시기만 출전한 선수들이기 때문이다.

또한 블래지수의 기준을 70으로 설정하였다. 블래지수의 정의에서는 70부터 블래지수를 느끼는 사람이 있다고 하였지만, 외야수들은 그늘 한 점 없는 장소를 뛰어다닌다는 특성이 있음을 감안하였다. 또한 80 이상의 블래지수가 측정된 날이 많지 않기도 하였다.

이를 선별하여 그림 3과 같이 외야수 기록이 있는 CSV파일에 블래지수 데이터를 입력하였다.

### 3.3. 분석 데이터 수정 및 변환

본 단계에서는 데이터를 수량화 할 수 있게 하기 위하

여 다양한 형태로 변환하여야 하는 단계이며 데이터를 그룹별로 분류해야 하는 단계이다.

또한 스맷티즈에서 제공하지 않은 도루 성공률, 타석당 투구수 등을 계산하기 위하여 수식을 입력하였다.

### 3.4. 모델 구축

본 단계에서는 분석하고자 하는 데이터를 분류/선별하고 활용할 모델을 구축하는 단계이다.

본 연구에서는 블래지수와 관련이 있을 것 같은 데이터를 선별하고자 하였다. 따라서 그라운드에서 있는 시간으로 관련짓기 위하여 도루 기록을 기반으로 하여 주루플레이에 관련한 기록을 분석하였고, 볼넷과 투구수 기록을 기반으로 하여 타석플레이에 관련된 기록을 분석하였다. 이 기록들은 숫자의 크기가 클수록 그라운드에서 있는 시간이 길어지게 되는 데이터이다. 참고 데이터는 각종 타격 지표 중 비율 기록(타율, 출루율, 장타율, OPS)을 제시하였다. 타격 비율 기록을 참고기록으로 정한 이유는 투수와 관련이 있는 기록이기 때문이다. 야구에서는 시즌이 진행될수록 투수가 지쳐서 기록이 떨어지기 때문에 타격기록이 올라가는 경우가 많다. 본

Fig. 3 Data Exploring and Input

연구에서도 표 2와 같이 불패지수가 70이상일 때 타격 비율 기록이 좋아지는 경향이 있었다.

3.5. 모델 평가 및 검증

본 단계에서는 모델을 검증하고 교차분석 하기 위하여 시각화를 진행하였다. 전체적인 연구 절차는 그림 4와 같다.

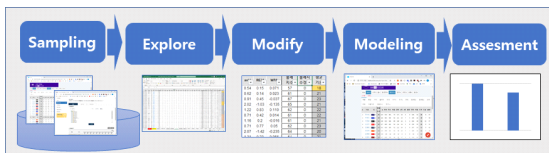


Fig. 4 SEMMA Model

IV. 연구 결과

4.1. 타격에 대한 참고기록

외야수의 타격 참고기록은 타율, 출루율, 장타율, OPS로 제시하였다. 홈런 등의 장타 기록을 제시하지 않은 이유는 장타율 기록은 홈런, 3루타, 2루타 등의 장타 기록이 포함되어야 올라가는 기록이기 때문이다. 표 2에서는 불패지수 70 미만과 70 이상으로 분류하여 타격 비율 기록을 제시하였다.

표 2에 제시된 바와 같이 불패지수가 70 이상이었을 때 타율, 출루율, 장타율, OPS 등의 타격 기록이 더 좋았다. 이는 불패지수가 높아야 타격이 좋다고 해석하기 보다는 투수가 지칠 시기가 여름이기에 타격이 전반적으로 좋아지기 때문이다. 즉, 외야수만의 기록이 아니라 투수가 개입된 기록이라고 볼 수 있다.

Table. 2 Batting Records of Outfielder

Discomfort Index	70 under	70 upper
Batting Average	.291	.305
On-Base Percentage	.358	.367
Slugging Average	.408	.450
On-Base Percentage plus Slugging Average	.766	.816

4.2. 주루플레이 기록

불패지수가 외야수의 주루플레이에 미치는 영향을 알아보았다. 주루플레이 기록은 투수와 관련이 크지 않다. 왜냐하면 도루는 주자의 의지가 개입하는 기록이기 때문이다. 주력이 느린 선수는 포수, 지명타자, 1루수에 상대적으로 많이 존재하며, 이들은 도루를 통하여 달리기를 하는 경우가 상대적으로 적지만, 외야수는 상대적으로 주력이 빠른 선수들이 많기 때문에 각 팀에서 주루플레이에 요구를 많이 받는다. 주루플레이는 불패지수를

측정할 정도의 무더운 여름에 체력이 소모되는 큰 요소 중에 하나이다. 그림 5를 보는 바와 같이 볼패지수가 70 이상일 때 각종 타격 지표는 상승하지만, 경기당 평균 도루 개수는 줄어드는 것을 알 수 있으며, 그림 6을 보는 바와 같이 도루 성공률도 낮아지는 것을 알 수 있다. 야구경기에서 도루 성공률이 70% 이상이어야 그 날 경기에 도움이 된다. 볼패지수가 70 미만일 때 도루 성공률이 70에 가까우므로 상대적으로 경기에 도움이 되었다는 것을 알 수 있다.

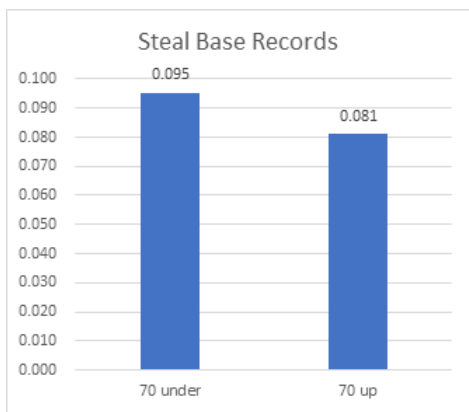


Fig. 5 Steal Base Records

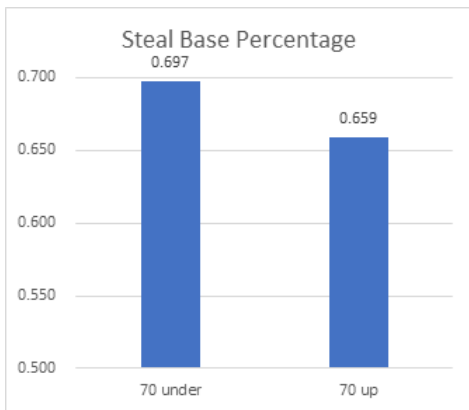


Fig. 6 Steal Base Percentage

#### 4.3. 타석플레이 관련 기록

볼패지수가 외야수의 타석플레이에 미치는 영향을 알아보았다. 타석에서 투수의 공을 많이 상대할수록 타석에서 있는 시간이 길어진다. 타석에서 공을 상대하는 타석플레이 관련 기록 역시 투수와 관련이 적다. 왜냐하면 초

구에 공을 칠지, 지켜볼지는 타자의 의지가 개입되는 기록이기 때문이다. 그림 7을 보는 바와 같이 타석 당 투구 수는 볼패지수가 70 미만이었을 때 70 이상일 때보다 투수가 공을 더 던지게 하였으며, 그림 8을 보는 바와 같이 경기 당 투구 수도 볼패지수가 70 미만이었을 때 70 이상일 때보다 투수가 공을 더 던지게 하였다. 큰 차이가 아니라고 볼 수도 있으나 한 게임 최소 아웃카운트인 27명을 곱하면 볼패지수가 70 이상일 때 상대편 투수가 게임당 공을 1개 이상 덜 던지게 한다. 공 1개로도 아웃카운트를 얻을 수 있고, 날이 더울 때 투수가 경기력이 떨어진다는 것을 볼 때 투수에게 상당히 도움이 되는 상황이다.

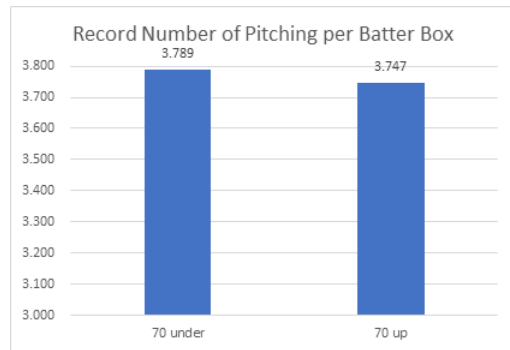


Fig. 7 Record Number of Pitching per Batter Box

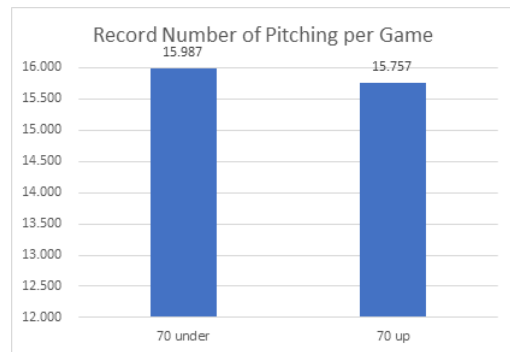


Fig. 8 Record Number of Pitching per Game

## V. 결론 및 제언

본 연구에서는 야구 경기에서 볼패지수가 외야수의 경기력에 미치는 영향을 분석하였다. 투수가 개입되지 않는 기록으로 분석하기 위하여 도루를 중심으로 하는

주루플레이와 투구 수와 볼넷을 중심으로 하는 타석플레이를 기반으로 하였다. 실제로 각종 타격 비율 기록은 투수의 경기력이 개입되기 때문에 본 연구에서는 볼패 지수가 높아졌는데 오히려 기록이 향상된 모습을 볼 수 있었다. 따라서 타격 비율지표는 본 연구에서 의미를 두지 않았다. 그러나 도루, 볼넷, 투구 수가 기반된 기록들은 볼패지수와 관련이 있었다.

본 연구에서는 규정타석에 진입한 외야수를 대상으로 한 이유는 한 시즌을 꾸준히 뛰는 선수를 대상으로 하여야 변화를 알 수 있기 때문이며, 80 이상의 볼패지수가 측정된 경우가 많지 않았으며, 선수들이 운동을 하는 상황이기 때문에 70 이상의 볼패지수를 대상으로 하였다.

본 연구를 통하여 스포츠 데이터를 활용하는 산업이 단지 경기기록만을 분석하는 것을 넘어서 외부데이터를 활용하는데 기반 기술 모델이 될 것이라 기대할 수 있다. 또한 1차 기록이라 불리는 누적 및 비율기록, 2차 기록이라 불리는 세이버메트릭스에 이어 3차 기록이라 부를 수 있는 외부 환경 데이터 적용 기록이 주목받을 수 있을 것이라 기대된다. 아울러 구단과 코칭스태프가 기록을 참고하여 적절한 선수기용과 육성에도 도움을 주고, 선수들도 본인의 데이터에 맞추어 몸 상태를 점검할 수 있으므로 경기력이 향상될 것으로 기대할 수 있다.

향후 연구과제로는 각 포지션 별로 상관관계를 분석하는 것이며, 기상뿐만 아니라 여러 가지 인문 지리 요소를 대상으로 경기력을 분석하는 것이다.

## REFERENCES

[ 1 ] S. A. Kang, "Utilization and Prospect of Big Data Analysis of Sports Contents," *Convergence Security Journal*, vol. 19, no. 1, pp. 121-126, Mar. 2019.

[ 2 ] Y. N. Jeon, Alavi, "A Study of The Athletic Satisfaction and Stress Factors of High School Baseball Players in The Case of Different Positions," Master's Thesis, The Graduate School of Kyungwon University, Seongnam, 2010.

[ 3 ] K. Y. Nam, and H. Jung, "A Study on the Using of Sports Big Data," *Sports Entertainment and Law*, vol. 22, no. 4, pp. 85-102, Nov. 2019.

[ 4 ] J. T. Lee, "Long Term Trends in The Korean Professional Baseball," *Journal of The Korean Data & Information Science Society*, vol. 26, no. 1, pp. 1-30, Jan. 2015.

[ 5 ] S. M. Kim, "The Effect of Daily Average Temperature on

The Batter's Performance in Baseball Game : based on Big Data Analysis," Master's Thesis, The Graduate School of Hoseo University, Asan, 2020.

[ 6 ] S. H. Lee, and H. J. Choi, "The Analysis of Pitching Result according to The Velocity and Pitch of Pitcher in That Case of Full-Counting on Major League Baseball(MLB)," *The Korea Journal of Sports Science*, vol. 28, no. 3, pp. 973-981, 2019.

[ 7 ] J. H. Hong, "The Effect of Golf Pre-Shop Routine on Club and Ball Data," Master's Thesis, The Graduate School of Choongang University, Seoul, 2019.

[ 8 ] SAS User Community [Internet]. Available: [http://mysas.co.kr/SAS\\_tiptech/j\\_eminer.asp?b\\_no=2666&gotopage=7&con=subject&keyword=&cmd=content&bd\\_no=29&gubun=/](http://mysas.co.kr/SAS_tiptech/j_eminer.asp?b_no=2666&gotopage=7&con=subject&keyword=&cmd=content&bd_no=29&gubun=/).

[ 9 ] D. W. Song, "Prediction of the Risk of Fire Occurrence According to the Weather Information Using Statistics and Data Mining Techniques," Ph.D. Dissertation, The Graduate School of Seoul National University of Science & Technology, Seoul, 2014.

[ 10 ] S. M. Kim, and K. S. You, "The Effect of Daily Average Humidity on Pitcher's Stats of Strike-Out," *Journal of Industrial Convergence*, vol. 18, no. 3, pp. 65-71, Feb. 2020.

[ 11 ] I. J. Jeon, and K. Y. Chung, "Life Weather Index Monitoring System using Wearable based Smart Cap," *The Journal of the Korea Contents Association*, vol. 9, no. 12, pp. 477-484, Dec. 2009.

[ 12 ] J. M. Kim, M. S. Kim, and K. N. Kim, "Crime Prediction Model based on Meteorological Changes and Discomfort Index," *Convergence Security Journal*, vol. 14, no. 6, pp. 89-95, Oct. 2014.



**김세민(Semin Kim)**

2006: 우석대학교 컴퓨터교육과 교육학석사  
 2010: 공주대학교 컴퓨터교육학과 교육학박사  
 2018: 한밭대학교 정보통신공학과 공학박사  
 2020: 호서대학교 스포츠과학대학원 야구학과 체육학석사  
 현재: 전주교육대학교 컴퓨터교육과 외래교수  
 ※관심분야: 소프트웨어교육, 로봇활용 및 메이커교육, 스포츠정보과학, 심리정보과학



**신좌철(Chwa-Cheol Shin)**

1990: 호서대학교 전자계산 이학사  
 1996: 호서대학교 전자계산 이학석사  
 2007: 호서대학교 컴퓨터공학 공학박사  
 현재: 호서대학교 혁신융합학부 조교수  
 ※관심분야: 컴퓨터융합교육, 인공지능 교육, 컴퓨팅 사고력, 교양교육, 소프트웨어 활용교육, 스포츠정보과학