

## 머신러닝 알고리즘을 활용한 한국프로축구 경기결과 예측: 2020~2023시즌 K리그 경기데이터 활용

김 필 수<sup>1</sup> · 이 상 현<sup>2\*</sup> · 전 성 삼<sup>3</sup>

<sup>1</sup>한국스포츠경영전략연구원 원장

<sup>2\*</sup>한국스포츠경영전략연구원 부원장

<sup>3</sup>한국스포츠경영전략연구원 선임연구원

## Predicting the Outcomes of K-League Matches using Machine Learning Algorithms

Philsoo Kim<sup>1</sup> · Sang Hyun Lee<sup>2\*</sup> · Seong-Sam Jeon<sup>3</sup>

<sup>1</sup>Director, Korea Sport Management Research Institute, Seoul 06543, Korea

<sup>2\*</sup>Deputy Director, Korea Sport Management Research Institute, Seoul 06543, Korea

<sup>3</sup>Senior Researcher, Korea Sport Management Research Institute, Seoul 06543, Korea

### [요 약]

인공지능 기술의 발전에 따라 프로스포츠의 여러 종목에서 머신러닝을 활용하여 경기결과 예측 연구가 활발히 이루어지고 있지만, 가장 큰 산업적 규모를 지닌 프로축구에 대한 경기 예측 연구는 부족한 실정이다. 본 연구는 이러한 한계를 극복하기 위해 2020년 2월 25일부터 2023년 7월 22일까지 한국프로축구리에서 펼쳐진 762경기 데이터를 K리그 데이터 포털에서 수집한 뒤, 파이썬(Python 3.10.9)을 활용하여 6가지 머신러닝 알고리즘으로 경기 결과를 예측했다. 분석 결과, 선형회귀를 기준으로 총 144경기 중 119경기(82.6%)에 대한 경기결과를 정확히 예측하였다. 본 연구는 국내에서는 선도적으로 머신러닝을 적용하여 한국프로 축구 리그의 경기결과를 예측한 실증연구라는 측면에서 이론과 실무적 의의가 있다.

### [Abstract]

With the advancements in artificial intelligence technology, research on predicting match results using machine learning is actively conducted in various sports. However, there is lack of research on match prediction in professional football, a sport with the largest industrial scale. To overcome this limitation, this study collected data from 762 matches held in the Korean Professional Football League from February 25, 2020, to July 22, 2023, through the K-League Data Portal. Using Python 3.10.9, six machine learning algorithms were utilized to predict match results. The analysis revealed that, based on linear regression, the results of 119 matches (82.6%) of a total of 144 matches were predicted accurately. This study holds theoretical and practical significance as a pioneering empirical study in South Korea that applies machine learning to predict match results in the Korean Professional Football League.

**색인어** : 머신러닝, 빅데이터 분석, 경기결과 예측, 스포츠 애널리틱스, 한국프로축구 리그

**Keyword** : Machine Learning, Big Data Analysis, Predicting Professional Match Outcomes, Sports Analytics, K-League

<http://dx.doi.org/10.9728/dcs.2024.25.4.1027>



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.

**Received** 29 February 2024; **Revised** 25 March 2024

**Accepted** 29 March 2024

**\*Corresponding Author; Sang Hyun Lee**

**Tel:** 

**E-mail:** shscottlee@naver.com

## 1. 서론

최근 스포츠 산업과 경영 분야의 연구에서 가장 큰 변화는 인공지능 기술의 도입과 빅데이터 분석의 고도화를 통한 스포츠 애널리틱스의 발전이라고 볼 수 있다[1]-[4]. 스포츠 애널리틱스는 스포츠과학과 데이터 애널리틱스의 융합된 학문과 기술을 바탕으로 다량의 데이터를 활용하여 스포츠 관련 조직과 산업 전반에 가치를 창출할 수 있는 다양한 의사결정을 체계화하는 분야로 개념화할 수 있다.

인공지능은 물론 빅데이터와 정보기술을 활용한 스포츠 애널리틱스의 고도화는 다양한 스포츠 산업과 경영의 저변확대와 전반적인 성장에 지대한 영향을 미치고 있으며, 앞으로 더 많은 발전이 기대된다. 스포츠 현장에서는 스포츠와 데이터 애널리틱스를 기반으로 경기데이터(on-field analytics)를 활용하여 선수나 팀의 경기력 예측[5], 부상 예방[6],[7], 전략 및 전술 구상[4], 스포츠 산업의 성장과 확장에 반드시 중요한 스포츠 베팅[8]-[10] 등에 이르기까지 다양한 영역에서 활용되고 있다.

스포츠 애널리틱스 분야는 인공지능 기술이 고도화되고 빅데이터가 체계적으로 구축되면서 경기결과를 예측하고 이를 학계나 현장에서 스포츠 산업적 측면에서 적용하여 혁신하고자 꾸준히 노력해왔다[11],[12]. 최근 스포츠 관련 시장의 급격한 확대는 이러한 인공지능과 빅데이터 기술의 진일보가 결합하여 융·복합적으로 나타나는 현상으로 연구의 노력이 확장되고 있다[13]-[15].

스포츠 산업에서의 빅데이터 애널리틱스의 적용은 스포츠와 직·간접적으로 관련된 데이터를 적용하여 스포츠 조직의 전반적인 경영 성과를 높이고 효율성을 극대화하기 위해 그 활용의 폭을 넓히는 과정을 포함한다. 스포츠 애널리틱스는 구체적으로 인공지능 기술발전과 궤를 같이한다고 볼 수 있다. 기존의 스포츠 통계분석이 의사결정에 도움이 될 수 있는 보조적 자료를 제공하는 데 목적을 둔 것에 반해, 인공지능을 적용한 스포츠 애널리틱스는 최적의 스포츠 관련 의사결정을 실시간에 의사결정자가 적극적으로 참여하는 선행적 과정을 포함한다[16]-[21].

이러한 스포츠 애널리틱스의 중요성을 고려하면 빅데이터를 기반으로 빠르게 정보를 취합하고 전처리하는 것은 물론 정보처리에 적합한 빅데이터 정제기술과 인공지능의 활용이 필수적이라고 본다[22],[23]. 특히, 인공지능을 활용한 빅데이터의 활용은 경기결과 예측 분야에서 전성기를 맞이하고 있다. 이는 단순히 경기결과 예측을 통한 관객의 경기에 대한 몰입증대부터 적극적인 베팅까지 많은 분야에서 활용될 수 있다. 이 과정에서 경기결과 예측을 통한 산업적 활용은 베팅 산업, 미디어 분야에서의 새로운 시장 획득 등과 같은 산업적 활용 자체는 물론, 대중의 스포츠에 대한 적극적 관심과 몰입으로 스포츠 산업의 전체적인 규모의 증대에까지 영향을 미칠 수 있을 것으로 보인다.

스포츠 경기결과 예측의 관심은 학술적인 연구에도 그대로

반영된다. 최근 인공지능을 활용하여 경기결과를 예측하는 연구가 급격히 증가하고 있다는 점이 주목된다. 국내외의 선행 연구를 체계적으로 고찰하면 골프[24], 농구[8],[13],[14],[25]-[30], 야구[31]-[35], 축구[36]-[40],[9]와 같은 프로스포츠는 물론이고 경마[41],[42], 경륜[43],[10]과 같은 경주 스포츠에서조차 경기결과 예측과 관련한 연구가 쏟아지고 있다. 하지만, 진술한 선행연구의 체계적인 고찰에도 불구하고 아직 국내 프로축구경기 데이터를 활용하여 경기결과를 예측하는 연구는 현재까지 발견하기 힘든 실정이다.

프로축구 종목은 명실상부 세계 최대 규모의 인기 스포츠로 볼 수 있다. 실제로 2022년 기준 프로축구의 시장규모는 3.2억 불로 추정되며 지속 성장하여 2028년에는 4.0억 불의 가치로 증가할 것으로 예상된다[44]. 이는 국내에서 가장 인기 있는 프로스포츠 종목으로 알려진 프로야구의 세계적인 시장 규모가 10억 불의 규모와 비교해 볼 때 상당히 높은 수치라 볼 수 있다. 따라서, 프로축구 관계자는 물론 일반인들 역시 경기결과 예측에 높은 관심을 가지는 것으로 확인된다[9],[45],[36].

하지만, K리그(한국프로축구) 경기결과 예측에 관한 연구가 부재하다는 것은 스포츠 산업 및 현장에서의 수요와 연구의 발전 간 공백이 발생하여 존재한다는 것을 알 수 있다. 축구는 다른 팀 스포츠와 비교하면 인공지능을 도입한 경기결과 예측에 분명한 어려움이 존재한다. 경기데이터를 활용하여 경기결과를 예측하기 위해서는 경기와 관련한 양질의 충분한 데이터가 필요하다. 농구는 다른 스포츠에 비해 득점이 많이 발생한다는 특성 덕분에 득점을 예측할 수 있는 많은 데이터가 누적된다고 볼 수 있다. 또한, 야구는 플레이 간 분절이 자주 발생하기 때문에 세부적인 경기지표 대부분을 데이터로 축적할 수 있다는 장점이 존재한다. 하지만, 축구는 적은 득점과 플레이의 단절 없는 연결성, 다른 프로스포츠 종목과 비교하면 작전타임의 부재 등으로 상대적으로 분석을 위한 자료의 축적에 어려움이 존재한다[9].

아울러, K리그는 유럽의 5대 축구 리그(잉글랜드 프리미어리그, 스페인 라 리가, 독일 분데스리가, 이탈리아 세리에 A, 프랑스 리그앙)와 비교했을 때, 실증분석을 위한 충분한 데이터가 누적되어 있다고 보기 힘들다. 또한, 다른 프로축구리그에 비해 K리그는 잦은 무승부가 더 발생하는 특성을 간과할 수 없으므로 경기결과 예측이 더욱 어렵다고 볼 수 있다. 따라서, K리그 경기결과를 예측하기 위해서는 다른 종목이나 리그에서의 경기결과를 예측하는 것과는 다른 시도가 필요하다.

이에 따라 본 연구의 목적은 K리그가 내재하고 있는 경기결과 예측의 어려움을 극복하고, K리그 경기 기록을 활용하여 머신러닝 알고리즘을 활용한 경기결과 예측을 시도하고자 한다. 특히, 2차 경기지표(advanced metrics)가 충분히 축적되어 있지 않으며, 무승부가 많이 나타나는 리그의 특성을 반영하여 승/패를 예측하기 위해 분류 중심의 알고리즘을 사용한 대부분의 기존 연구와 달리 회귀 알고리즘을 중심으로 경기결과 각각의 머신러닝 알고리즘의 예측 성능을 비교

하고, 이를 바탕으로 경기결과를 예측하고자 한다.

보다 구체적으로, 선형회귀(Linear Regression), 캣부스트(Cat Boost), 랜덤 포레스트(Random Forest), 엑스트라 트리(Extra Trees), XG 부스트(XG Boost), SGD(Stochastic Gradient Descent)의 6가지 알고리즘을 활용하여 득점과 실점 차이를 예측하고, 이 중 가장 높은 예측 성능을 기록한 머신러닝 알고리즘을 기반으로 경기결과를 실질적으로 분류하고자 한다. 이를 위해 2020~2023시즌 K리그 정규리그에서 발생한 경기 관련 자료를 수집하여 분석하고 이에 관한 학술 및 실무적 시사점을 살펴보고자 한다.

## II. 연구 방법

### 2-1 연구대상

본 연구의 머신러닝 알고리즘을 활용한 K리그 경기결과 예측을 위해 2020시즌부터 2023시즌 7월 22일까지 진행된 K리그의 762경기(n=1,524)를 연구의 표본으로 설정하였다. 표 1은 연구의 분석이 진행된 기간 K리그에 소속되었던 16개 팀의 홈/원정 경기 수와 참여 시즌을 나타낸 것이다.

본 연구의 실증분석을 본격화하기 위한 자료는 K리그 데이터 포털(<https://data.kleague.com/>)에서 제공하는 데이터를 수집하였다. 수집된 K리그 경기 기록 데이터는 선수 단위로 수집된 데이터를 팀 단위로 재구성하고 평균0, 분산1의 정규

표 1. 연구대상 구단의 참가 경기 및 시즌 수

Table 1. Number of games and seasons played by the clubs in the study

Team	Home	Away	Total	Season
Gangwon FC	63	64	127	4
Gwangju FC	44	45	89	3
Gimcheon Sangmu FC	19	19	38	1
Daegu FC	62	65	127	4
Daejeon Hana Citizen	12	12	24	1
Busan I Park	13	14	27	1
Sangju Sangmu FC	13	14	27	1
FC Seoul	64	63	127	4
Seongnam FC	52	51	103	3
Suwon Samsung Bluewings	64	63	127	4
Suwon FC	52	48	100	3
Ulsan HD FC	64	63	127	4
Incheon United FC	63	64	127	4
Jeonbuk Hyundai Motors	64	63	127	4
Jeju United FC	48	52	100	3
Pohang Steelers	65	62	127	4
Total	762	762	1,524	48

분포 형태로 표준화하는 체계적인 전처리(preprocessing) 과정을 통해 본 연구에 적합한 고유 데이터 세트로 구축하여 분석에 적용되었다.

### 2-2 연구 변수

본 연구에서는 K리그 경기결과 예측을 위해 각 팀의 득점을 예측하고자 하였다. 머신러닝 회귀모델을 통해 예측된 득점을 기반으로 예상 경기결과를 “승/무/패”의 범주로 분류하였다.

먼저 본 연구에서 득점 예측을 위해 사용된 연구 변수는 기본적으로 K리그 데이터 포털에서 제공하는 양 팀의 경기 기록에 해당한다고 볼 수 있다. 총 117개의 변수가 사용되었으며, 큰 틀에서 “공격 지표”, “패스 지표”, “수비 지표”, “골키퍼 지표”로 구분하여 나눌 수 있다.

표 2. 연구대상 변수

Table 2. Variables in this study

Variables(Opp = Opponents)	
Home/Away	Successful handling aerial balls
Shots	Handling aerial balls attempts
Shots on target	Opp shots on target
Blocked shots	Opp blocked shots
Shots off target	Opp shots off target
Shots inside box	Opp shots inside box
Shots outside box	Opp shots outside box
Offsides	Opp offsides
Frees	Opp frees
Corners	Opp corners
Trows	Opp throws
Successful Dribbles	Opp successful dribbles
Dribble Attempts	Opp dribble attempts
Pass Attempts	Opp pass attempts
Accurate passes	Opp accurate passes
Key passes	Opp key passes
Opposition half accurate passes	Opp opposition half accurate passes
Opposition half pass attempts	Opp opposition half pass attempts
Central zone accurate passes	Opp central zone accurate passes
Central zone pass attempts	Opp central zone pass attempts
Own half accurate passes	Opp own half accurate passes
Own half pass attempts	Opp own half pass attempts
Accurate long passes	Opp accurate long passes
Long pass attempts	Opp long pass attempts
Accurate mid-range passes	Opp accurate mid-range passes

Mid-range pass attempts	Opp mid-range pass attempts
Accurate short passes	Opp accurate short passes
Short pass attempts	Opp short pass attempts
Accurate forward passes	Opp accurate forward passes
Forward pass attempts	Opp forward pass attempts
Accurate transverse passes	Opp accurate transverse passes
Transverse pass attempts	Opp transverse pass attempts
Accurate back passes	Opp accurate back passes
Back pass attempts	Opp back pass attempts
Accurate crosses	Opp accurate crosses
Cross attempts	Opp cross attempts
Escaping pressure	Opp escaping pressure
Ground duels won	Opp ground duels won
Ground duels	Opp ground duels
Aerial duels won	Opp aerial duels won
Aerial duels	Opp aerial duels
Tackles won	Opp tackles won
Tackles attempts	Opp tackles attempts
Clearances	Opp clearances
Interceptions	Opp interceptions
Blocking passes	Opp blocking passes
Acquiring loose balls	Opp acquiring loose balls
Blocking shots	Opp blocking shots
Ball misses	Opp ball misses
Foul	Opp foul
Fouled	Opp fouled
Yellow cards	Opp yellow cards
Red cards	Opp red cards
Catchings	Opp catchings
Punchings	Opp punchings
Successful handling aerial balls	Handling aerial balls attempts
Accurate goal kicks	Opp accurate goal kicks
Goal kick attempts	Opp goal kick attempts
Team scoring	

K리그 데이터는 최근 축구 지표 중 가장 관심이 높은 xG(Expected Goal) 값이나 xT(Expected Threat) 계수 등의 2차 경기지표(advanced stat)는 제공되지 않으나, 1차 경기지표(traditional stat)에 해당하는 패스 지표를 위치와 방향, 거리 등의 기준으로 세분화하여 제공하고 있어서 이를 본 연구의 변수로 활용하였다. 표 2는 사용된 변수를 정리하여 나타낸 것이다.

## 2-3 머신러닝 알고리즘

### 1) 선형회귀(Linear Regression)

선형회귀 모델은 수학적식으로 표현된 함수를 나타내며 하나 이상의 독립변수와 종속변수 사이의 상관관계를 의미한다 [12],[28]. 종속변수와 독립변수의 관계를 가장 잘 설명하는 상수항  $b$ 를 최적화하는 방법으로 본 연구에서는 여러 개의 독립변수를 통해 종속변수를 예측하기 위하여 다중 선형회귀 (Multiple Linear Regression) 기반 알고리즘을 사용하였다 [46].

### 2) 캣부스트(Cat Boost)

캣부스트는 “Categorical Boosting”의 줄임말로 그래디언트 부스팅 기반의 머신러닝 라이브러리로 범주형(categorical) 변수를 다루는데 특화되어 있다. 그래디언트 부스팅의 치명적인 문제점인 과적합의 문제를 개선하면서 XG Boost나 Light GBM 등에 비해 학습 속도가 빠르다는 장점을 지닌다[47].

### 3) 랜덤 포레스트(Random Forest)

랜덤 포레스트는 의사결정나무 기반의 앙상블 기법의 일종이다. 여러 개의 의사결정나무를 학습시키고 그 결과를 집계하여 최종 예측 결과를 도출하는 방식으로 작동한다. 각각의 의사결정나무는 서로 다른 데이터 샘플과 변수를 통해 학습되므로, 단일 의사결정나무 알고리즘과 비교하면 더욱 높은 성능을 나타낼 수 있다[48],[49].

### 4) 엑스트라 트리(Extra Trees)

엑스트라 트리는 랜덤 포레스트 알고리즘을 확장한 머신러닝 기법이다. 랜덤 포레스트는 부트스트랩 샘플을 사용하여 전체 특성에 대한 정보 이득을 계산하는 반면 엑스트라 트리는 훈련 세트 전체를 사용한다고 볼 수 있다. 구체적으로 엑스트라 트리 기법은 데이터 샘플의 개수와 특성을 무작위로 선정하여 그에 대한 최적은 노드(Node)를 나눈다. 이를 통해 성능이 낮아질 가능성을 배제할 수는 없지만, 과적합의 위험을 줄이고 검증 세트의 점수를 높이는 효과가 있다. 일반적으로 엑스트라 트리는 랜덤 포레스트 기법과 비교하면 무작위성이 더 크기 때문에 많은 결정 트리를 훈련해야 한다[50],[51].

### 5) XG 부스트(XG Boost)

XG 부스트는 캣부스트와 마찬가지로 그래디언트 부스팅 기반의 알고리즘 기반의 기법으로 그래디언트 부스팅 알고리즘에 병렬 학습이 지원되도록 구현한 구조이다[52]. XG 부스트는 학습 시간이 빠르고 과적합 위험이 적으면서도 대용량의 구조화된 데이터 세트에서 높은 성능을 보여 스포츠 경기 예측 관련 머신러닝 알고리즘의 적용에 적합하다고 볼 수 있다.

### 6) 확률적 경사 하강법(SGD Regressor)

확률적 경사 하강법(Stochastic Gradient Descent :SGD)

을 활용하여 정규화된 경험적 손실을 최소화하는 방식으로 학습하는 선형 모델이다[29]. 확률적 경사 하강법은 기본적인 학습 방식인 GD(Gradient Descent)가 전체 데이터를 계산하여 최적의 값을 구하는 데에 비해, 일부 데이터만을 계산하여 정확도가 낮아지더라도 계산 속도를 높이는 방법이다. SGD 회귀는 구현이 쉽고 효율적이지만 피쳐 스케일링에 민감하고 하이퍼 파라미터가 많다는 한계를 나타내기도 한다.

### 2-4 자료 처리

본 연구에서는 K리그 데이터 포털에서 공개하고 있는 경기 기록을 활용하여 각 팀의 득점을 예측하고, 이를 기반으로 전술한 각각의 6가지 머신러닝 알고리즘을 적용하여 경기결과를 선제적으로 예상하고 각 알고리즘에 대한 성능을 비교하였다. 본 연구를 위해 K리그 데이터 포털에서 제공하는 경기 데이터를 파이썬(Python 3.10.9)의 Selenium, Request, BeautifulSoup 라이브러리를 통해 수집하였다. 선수 단위로 수집된 데이터는 팀 단위로 통합하였으며 파이썬의 Scikit-Learn 라이브러리의 Standard\_scaler() 함수를 통해 평균 0, 분산 1의 정규분포 형태로 표준화하였다. 2020시즌부터 2022시즌까지 진행된 K리그의 618경기를 학습데이터로 활용하였으며, 7월 22일까지 진행된 2023시즌 144경기를 테스트 데이터로 설정하였다.

6가지 알고리즘의 성능을 비교하기 위해 본 연구에서는 머신러닝을 통해 각 팀의 “득점”을 예측하고자 회귀모델을 적용하였다. 이로 인해 분류모델에서 흔히 사용되는 F1 스코어, 민감도, 특이도 등의 평가 지표가 아닌 회귀모델 성능 평가에 적합한 결정계수(R-squared), 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE), 평균 절대 비율 오차(MAPE)를 사용하였다[8],[53]-[55].

방법론적 측면에서 제시한 6가지의 머신러닝 알고리즘 모델 중 성능이 가장 좋은 모델을 활용하여 7월 22일까지 진행된 K리그 144경기에서의 각 팀 득점을 예상하였다. 양 팀의 예상 득점을 비교하여 예상 득실차가 -0.5에서 0.5 사이의 경기는 무승부로 예측하였으며, 0.5 이상 차이가 나는 경우 예상 득점이 높은 팀이 승리 팀, 예상 득점이 낮은 팀을 패배 팀으로 예측하였다. 최종적으로 경기 예측 결과와 실제 경기 결과를 비교하여 예측 정확도를 파악하고 경기결과와 예상 득실 차에 따른 예측 정확도를 비교하고 중요 변수를 도출하였다. 본 연구를 위한 자료의 처리는 파이썬(Python 3.10.9)의 numpy, pandas, scikit-learn, eli5 라이브러리를 활용하였다.

## III. 연구 결과

### 3-1 머신러닝 모델 간 성능 비교

본 연구에는 K리그 경기 기록을 활용하여 경기결과를 예측

하기 위해 6가지 머신러닝 알고리즘 모델을 학습시킨 후 각 팀의 “득점”을 예측하였다. 선형회귀 모델은 결정계수 0.815, MAE 0.346, RMSE 0.490, MAPE 0.441로 4가지 지표에서 모두 가장 우수한 성능을 나타내었다. 결정계수 기준 캣부스트(0.733)와 엑스트라 트리(0.693)가 그 뒤를 이었다.

비교분석 결과에 따라서 선형회귀 모델의 예측 결과를 기반으로 K리그 2022~2023시즌의 경기결과를 예측하고 그 예측 정확도를 확인하였다. 표 3은 6가지 머신러닝 모델의 성능을 비교하여 그 우위에 대해 나타내고 있다.

아래의 그림 1에서는 본 연구에 적용된 각각의 머신러닝 알고리즘의 결정계수와 RMSE 수치를 비교하여 시각화하였다.

### 3-2 득점 예측 모델 중요 변수

K리그 경기에서 득점을 예측하는데 중요한 변수는 반대로 생각하면 득점을 기록하는 데 큰 영향을 미치는 변수로 볼 수 있다. 이러한 측면에서 본 연구에서는 무엇보다 투입된 121개의 변수 중 어떤 변수들이 경기결과 예측에 매우 중요하게 작용하였는지를 더욱 면밀하게 파악하기 위해 파이썬 eli5 라이브러리의 PermutationImportance를 활용하여 예측에 활용된 모델의 순열 중요도(permutation importance)를 도출하였다. 표 4는 eli5 라이브러리를 통해 도출한 상위 10개의

표 3. 머신러닝 알고리즘 간 성능 평가 지표 비교  
Table 3. Comparison of performance metrics among machine learning algorithms

ML algorithms	r2 score	MAE	RMSE	MAPE
Linear Regression	<b>0.815</b>	<b>0.346</b>	<b>0.490</b>	<b>0.441</b>
Cat Boost	0.733	0.410	0.589	0.559
Extra Trees	0.693	0.404	0.631	0.580
Random Forest	0.624	0.471	0.699	0.706
XG Boost	0.596	0.460	0.725	0.624
SGD Regression	0.567	0.586	0.750	0.795

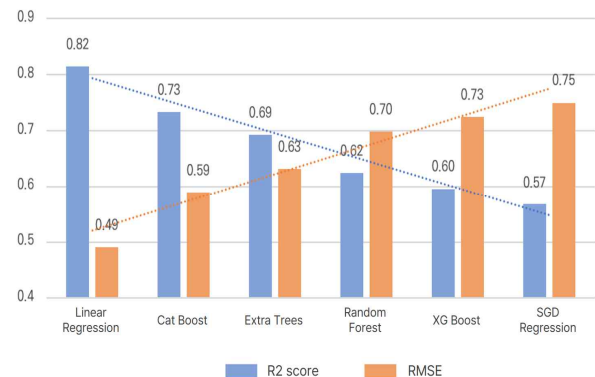


그림 1. 머신러닝 알고리즘 간 평가 지표 비교  
Fig. 1. Comparison of function among machine learning algorithms

**표 4.** K리그 2023시즌 경기결과 예측 모델의 변수 중요도  
**Table 4.** Variable importance in the K League 2023 match result prediction model

Rank	Variables	Weights
1	Shots on target	2.6471
2	Opp catchings	0.9442
3	Opp punchings	0.6830
4	Opp shots outside box	0.0060
5	Opp accurate goal kicks	0.0060
6	Frees	0.0046
7	Corners	0.0029
8	Opp accurate short passes	0.0028
9	Accurate transverse passes	0.0020
10	Opp yellow cards	0.0019

중요 변수 및 가중치를 나타낸 것이다.

본 연구에서 분석된 K리그 경기결과 예측을 위해 가장 중요한 변수로는 “유효슈팅”과 “상대 골키퍼 캐칭”, “상대 골키퍼 편칭”이 도출되었다. 유효슈팅은 골대 안을 향하는 슈팅으로 유효슈팅이 득점으로 연결되지 않도록 하는 골키퍼의 기록이 유효슈팅과 함께 가장 중요한 변수로 나타난 것으로 확인되었다 이를 포함하여 상위 10개의 변수 중요도로 “상대 박스 밖 슈팅”, “상대 골킥 성공”, “프리킥”, “코너킥”, “상대 샷패스 성공”, “횡패스 성공”이 구성되어 분석 결과로 도출되었다.

**3-3 2023시즌 K리그 경기결과 예측**

2023년 2월 25일 K리그 개막 이후 7월 22일까지 진행된 모든 경기결과를 예측하기 위해 가장 높은 머신러닝 알고리즘의 성능을 보인 선형회귀모델을 선택하여 2020~2022시즌의 데이터를 학습시켰다. 학습된 모델에 2023시즌의 경기 기록을 투입하여 각 팀의 득점을 예상하고 경기결과를 예측하였다. 본 연구의 실증분석 결과 전체 144경기 중 119경기의 실제 결과에 적중하여 82.64%의 예측 정확도로 K리그 경기 예측에 성공하였다.

실제로 승리한 58경기 중 머신러닝 알고리즘 기반의 예측은 52경기를 승리로 예측하였고 5경기는 무승부, 1경기는 패배로 예측하여 실제 승리 경기의 예측 정확도는 89.7%에 해당하였다. 무승부를 거둔 실제 41경기 중 머신러닝 알고리즘은 29경기만을 예측하고 8경기는 승리, 4경기는 패배로 예측하여 70.7%의 예측 성공률을 기록하였다. 실제로 패배한 45경기 중 머신러닝 알고리즘은 38경기를 패배로 예측하였고 1경기를 승리, 6경기를 무승부로 예측하여 84.4%의 예측 정확도를 보였다.

연구 결과, 본 연구에 적용된 머신러닝 알고리즘은 K리그 경기 예측에 있어 승리 → 패배 → 무승부 순으로 예측이 정

**표 5.** K리그 2023시즌 예측 결과 및 실제 경기결과 분포  
**Table 5.** Predicted and Actual Results Distribution in K-League 2023 Season

	Result-Win	Result-Draw	Result-Lose	Total
Pred-Win	52	8	1	61
Pred-Draw	5	29	6	40
Pred-Lose	1	4	38	43
Total	58	41	45	144

확하였으며, 특히 무승부 경기는 승리와 패배와 비교하여 예측 정확도가 큰 폭으로 하락하였다. 표 5는 K리그 2023시즌의 114경기를 예측 결과와 실제 결과에 따라 나타낸 것으로 하이라이트로 표시된 부분이 예측에 성공한 경우이다.

**IV. 논 의**

본 연구는 2020년 2월 25일부터 2023년 7월 22일까지 펼쳐진 K리그 144경기를 선형회귀, 캣부스트, 랜덤 포레스트, 엑스트라 트리, XG 부스트, SGD 회귀분석의 6가지 머신러닝 알고리즘을 적용하여 경기결과를 예측한 결과 결정계수 기준으로 선형회귀(0.815) → 캣부스트(0.733) → 엑스트라 트리(0.693) → 랜덤 포레스트(0.624) → XG 부스트(0.596) → SGD 회귀분석(0.567) 순으로 예측 성능이 높은 것으로 나타났다. 이를 바탕으로 선형회귀 기반 머신러닝 알고리즘을 기준으로 실제 예측을 경기분석에 적용해 본 결과, 총 144경기 중 119경기(82.6%)의 결과를 정확히 예측하였다. 이 중 홈팀이 승리한 경기에 승리할 것으로 예측할 가능성이 89.7%로 가장 높았고, 패배한 경기를 패배할 것으로 예측한 경기는 84.4%로 뒤를 이었다. 마지막으로 무승부 경기를 무승부로 정확히 예측한 경기는 전체의 70.7%에 해당한다.

본 연구는 다음과 같은 이론적 시사점을 지닌다. 첫째, K리그의 경기결과를 다양한 머신러닝 알고리즘을 적용하여 실질적으로 예측한 최초의 연구로서 그 가치를 지닌다고 볼 수 있다. 전술하였듯이, 프로축구의 승패 예측은 경기데이터 축적의 한계와 빈번한 무승부의 발생으로 다른 종목에 비해 상대적으로 매우 어려우며, K리그의 경우 데이터의 내생적 한계로 인해 이러한 점이 더욱 두드러진다[9].

그러나 고도화된 다양한 머신러닝 알고리즘을 활용하여 분석한 본 연구의 예측 성능은 결정계수 기준 0.815로, 기존의 선행연구에서 프리미어리그 경기결과 예측을 시행했을 경우 기록한 정확도 0.5~0.6 수준의 예측 성능[56],[45],[36]을 상회하는 연구 결과의 수준을 도출하였다. 스포츠 애널리틱스가 상대적으로 더 발전한 타 종목이나 5대 프로축구리그와 비교 시 K리그의 경우에는 그만큼의 충분한 데이터를 축적하고 있지는 않으나, 분석 방법과 다양한 변수의 활용을 통해 기존의 선행연구에서 달성하였던 예측 수준 혹은 그 이상의 예측



성능을 확보할 수 있다는 것을 실증하였다는 점에서 학문적 의의를 지닌다.

둘째, 본 연구는 K리그의 경기 결과를 예측하는 데 여러 머신러닝 알고리즘을 반영하여 최적의 성능을 도출하였다. 전술하였듯이, 축구는 무승부가 자주 발생한다는 점 때문에 승/패 두 가지 경우 중 하나의 선택지만 예측하면 되는 타 종목 예측과 비교해 승/무/패를 각각 예측해야 하므로 더 어렵다고 할 수 있다. 특히, K리그는 다른 프로축구리그에 비해 무승부가 더 많이 발생하는 특성이 있으므로 단순한 승패 분류 방식 이외의 정교화된 접근이 필요하다. 본 연구의 머신러닝 알고리즘 적용 표본인 K리그 144경기 중 41경기가 무승부로 무려 28.5%에 해당한다. 이러한 맥락에서 K리그 경기는 프리미어 리그(PL)에 비해서도 예측이 난해하다고 볼 수 있는데, 2022~2023시즌 유럽 프리미어리그의 경우 무승부 경기는 전체 760경기 중 174경기(22.9%)에 불과하며, 경기 승패에 결정적인 설명력을 가지는 xG 등의 2차 경기지표도 부족하다.

이러한 요인으로 인해 경기결과 예측을 서포트 벡터 머신이나 로지스틱 회귀 등의 분류 기반의 알고리즘을 사용할 경우, 타 스포츠 종목보다 예측 성능이 낮은 가능성이 충분히 존재한다. 연구분석의 측면에서 K리그의 경기결과 예측을 통한 충분한 경기결과 예측 성능을 확보하는 데 어려움이 분명 존재한다고 볼 수 있다. 이러한 이유로 K리그 경기결과 예측 연구가 여전히 학계에서 부족한 것은 역시 이에 기인한 것으로 보인다.

본 연구에서는 이러한 한계점을 극복하기 위해 예상 득점 마진을 회귀분석 기반 알고리즘을 활용하여 분석하고, 이를 바탕으로 승/무/패로 다시 분류함으로써 경기결과 예측 수준을 높이기 위해 노력하였고, 그 결과 전체 경기의 82.6%에 해당하는 경기를 정확히 예측하였다. 이는 경기결과 예측을 위해 국내에서 최초로 시도된 방식으로 후속 연구의 방법론적 측면을 새롭게 고찰하는데 기여도가 있을 것으로 기대된다.

셋째, 본 연구에서는 K리그 경기데이터를 활용하여 경기결과를 예측하기에 적합한 머신러닝 알고리즘을 제안하였다. 본 연구는 선형회귀, 캣부스트, 랜덤 포레스트, 엑스트라 트리, XG 부스트, SGD 회귀분석의 6가지 머신러닝 알고리즘을 활용하여 비교·분석하고 이 중 가장 높은 성능을 보이는 선형회귀를 사용하여 경기결과를 예측하였다. 머신러닝을 통한 경기결과 예측은 분석하고자 하는 종목의 특성이나 목표 변수, 리그의 특성마다 최적화된 알고리즘에 차이가 존재한다 [8],[13],[14],[24][45]. 본 연구는 6개의 회귀 기반의 머신러닝 알고리즘을 비교하고 분석한 결과 K리그 2023시즌에 펼쳐진 경기결과를 예측하기 위해 선형회귀 기반의 머신러닝 모델이 가장 적합한 알고리즘이라는 것을 검증하였다.

아울러, 본 연구는 다음과 같은 실무적 시사점을 지닌다. 첫째, 감독과 코치진을 포함한 프로스포츠 산업의 현장 관계자들은 머신러닝 알고리즘 기반의 경기결과를 활용하여 전력상 예상되는 경기결과와 실제 경기결과 사이에 괴리가 발생하는 경우를 자세히 분석할 수 있다. 이를 통해 현 전력 이상의 성적을 거두기 위해서, 혹은 전력보다 낮은 성적이 나오지

않도록 대처하고 전략을 구성할 수 있다고 본다. 머신러닝 알고리즘을 통한 경기결과 예측은 경기지표를 통해 나타나는 전력과 경기결과 간의 관계를 데이터만을 가지고 분석함으로써 편견 없이 결과를 예측한다는 장점을 지닌다. 즉, 해당 경기에 임하는 팀의 전력을 정확하게 평가하여 경기의 결과를 예측한다고 볼 수 있다.

본 연구의 분석 결과는 K리그의 전체 144경기 중 119경기의 결과를 예측하는 높은 적중률을 보였으나 여전히 25경기(17.4%)에 대한 괴리가 발생하는 점에 주목할 필요가 있다. 이는 전력상 승리를 반드시 확보했어야 하는 경기에서 예상보다 못한 경기결과를 나타내거나, 패배해야 할 것으로 보이는 경기에서 더 나은 성적을 도출하였다는 것을 의미한다. 해당 경기에 대해 더욱 체계적으로 자세하게 분석한다면, 각 팀은 현 단계에서 내재한 전력에 비해 뛰어난 성적을 거두는 것이 가능할 것으로 보인다. 이는 현재의 경기를 위한 대처보다 팀이 가진 특성과 내재한 전력을 장기적 관점에서 효율적으로 활용하는 데 도움이 될 수 있을 것으로 보인다.

둘째, 스포츠 산업적 측면에서 마케팅 담당자를 포함한 프로스포츠 관계자들은 실시간 경기 예측 결과를 바탕으로 적절한 이벤트나 미디어 콘텐츠 자료를 준비할 수 있을 것으로 보인다. 예를 들어, 정규리그 우승을 앞둔 팀은 우승이 확정되는 경기가 언제인지를 머신러닝 알고리즘에 기반하여 선제적으로 예측하여 이에 적합한 이벤트를 준비하고 관련된 비용을 효과적으로 집행할 수 있다. 또한, 세부적으로 승/패뿐만 아니라 경기별 예상되는 득점 마진에 따른 이벤트를 준비하는 것 또한 가능하다고 본다. 경기결과가 언제나 정확하지는 않을 수 있지만, 경기결과 예측을 바탕으로 관련 행사를 준비한다면 효율적인 재정 집행이 가능할 것으로 보인다.

셋째, 스포츠 베팅 혹은 미디어와 같은 경기 자체보다 경기결과를 이용한 다양한 스포츠 산업에 활용될 수 있을 것으로 판단된다. 스포츠 베팅 관련 산업의 경우 급격한 발전에도 불구하고 국내에서는 머신러닝 알고리즘 적용을 통한 경기 예측에 관한 학술적 논의가 여전히 부족한 상황이고, 스포츠 선진국과 비교하면 그 규모 역시 매우 작다고 볼 수 있다. 하지만, 스포츠 베팅산업은 그 자체로 스포츠 산업 전반의 확장력이 여전히 실질적인 것은 물론이고, 스포츠 베팅을 통해 팬들과 대중이 해당 스포츠 종목에 더욱 관심을 가지고 적극적으로 참여할 수 있게 하는 동인이 되기도 한다[43],[10],[57],[58]. 따라서, 본 연구와 같이 체계적이고도 과학적인 머신러닝 알고리즘에 기반한 경기결과 예측에 관한 분석을 활용하여 기타 산업에 적극적으로 활용하는 것은 스포츠 산업 전반의 발전에 기여도가 충분히 있을 것으로 보인다.

## V. 결 론

본 연구는 국내에서 스포츠 산업 내 가장 큰 시장규모인 프로스포츠 종목 중 K리그 경기결과를 머신러닝 알고리즘을 활용하여 선도적으로 예측한 실증연구로서 그 이론·실무적 의

의를 지닌다. 본 연구의 결론 및 제언의 차원에서 향후 연구에서는 본 연구에서 포괄적으로 다루지 못했던 다양한 요인들을 함께 고려함으로써 연구의 전반적인 발전과 확장에 도움이 될 수 있을 것으로 판단된다.

첫째, 본 연구에서 활용하지 못한 의미 있는 변수를 추가로 고려함으로써 보다 정교한 예측이 가능할 것으로 보인다. 프리미어리그 경기결과를 예측한 선행연구[9]에 근거하면, xG 값과 같은 2차 경기지표를 활용함으로써 기존 0.54 이하였던 F1-score를 0.86까지 개선하였다. 아직 K리그 데이터베이스에는 경기력을 충분히 설명할 수 있는 xG 값과 유사한 2차 경기지표가 사용되지 않고 있다. 향후 연구에서는 이와 같은 2차 경기지표를 계량적으로 확보하여 머신러닝 알고리즘에 사용할 수 있다면, 더욱 정밀한 수준의 예측 정확도를 도출하는 것이 가능할 것으로 보인다.

둘째, 무승부에 대한 예측 정확도를 높일 필요가 존재한다. K리그에서는 유럽 5대 리그와 비교하여 정규리그 전체 경기에서 무승부 발생 비중이 높은 것으로 나타났다. 하지만, 본 연구에서는 승리(89.7%)나 패배(84.4%) 예측과 비교하면 무승부(70.7%)일 경우 경기결과 예측 적중 비율이 현저히 낮았다고 볼 수 있다. 이는 K리그에서 승/패 경우에 고려한 머신러닝 알고리즘과 무승부일 때 고려할 필요가 있는 다른 알고리즘의 적용이 필요할 가능성이 존재함을 시사한다. 후속 연구에서는 이러한 한계를 극복하여 무승부가 나타나는 경기에서도 예측 정확도를 정교하게 높이는 방법에 관해 실증연구를 설계한다면, 보다 수준 높은 분석과 결과를 도출할 수 있을 것으로 보인다.

셋째, 본 연구에서 적용된 변수의 범위보다 적은 숫자의 변수를 사용함으로써 간명성을 높이고 현장에서 활용하기에 도움이 되는 연구로 분석을 고도화할 필요가 있다. 포괄적이면서도 다양한 경기지표를 사용하는 것은 전반적인 설명력을 높이는 데 도움이 될 수 있지만, 동시에 간명성(parsimony) 역시 중요한 과제이다. 본 연구에서는 총 121개의 변수를 활용하여 F1-score 0.815를 기록하였다. 이는 예측 성능 측면에서는 가치가 충분하다고 볼 수 있지만, 현장에서 활용하기에는 다소 많은 변수의 수가 된다고 판단해 볼 수 있다. 이러한 측면에서 최근 프리미어리그 경기결과 예측에 단 16개의 변수만 활용하여 F1-score 0.86을 기록한 국내 연구는 참고할 만하다[9]. 향후 K리그 경기결과 예측에 있어 다양한 통계 기법과 머신러닝 알고리즘을 활용하여 모델에 투입되는 변수의 전반적인 숫자를 줄이면서 예측 정확성을 담보할 필요가 있다. 이러한 노력은 연구의 활용도를 극대화할 뿐만 아니라 중요하게 여겨지는 간명성은 물론, 프로스포츠 산업의 현장 관계자들이 주목해야 할 변수들에 대해 더욱 명확히 파악할 수 있는 참고자료의 기능을 제공할 수 있을 것으로 예상된다.

넷째, 본 연구에서는 2020년 이후 개편된 K리그 홈페이지에 나타난 변수들을 최대한 활용하기 위해 2020시즌부터 2023시즌 7월 22일까지 진행된 K리그의 762경기(n=1,524)만을 분석에 활용하였다. 그 결과, 샘플 수의 한계로 최근 스

포즈 경기분석에서도 그 활용도가 증가하고 있는 딥러닝을 사용하지 못한 한계점을 지닌다. 향후 연구에서는 더욱 많은 경기의 데이터를 활용함으로써 다양한 딥러닝 알고리즘을 활용하여 데이터를 분석한다면, 보다 의미 있는 진전이 이루어질 수 있을 것으로 기대한다.

## 참고문헌

- [1] D. Araújo, M. Couceiro, L. Seifert, H. Sarmento, and K. Davids, *Artificial Intelligence in Sport Performance Analysis*, Routledge, 2021.
- [2] V. Dhar, "What Is the Role of Artificial Intelligence in Sports?," *Big Data*, Vol. 5, No. 3, pp. 173-174, 2017. <https://doi.org/10.1089/big.2017.29022.vdb>
- [3] I. Ghosh, S. Ramasamy Ramamurthy, A. Chakma, and N. Roy, "Sports Analytics Review: Artificial Intelligence Applications, Emerging Technologies, and Algorithmic Perspective," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 13, No. 5, e1496, September/October 2023, <https://doi.org/10.1002/widm.1496>
- [4] E. Morgulev, O. H. Azar, and R. Lidor, "Sports Analytics and the Big-data Era," *International Journal of Data Science and Analytics*, Vol. 5, pp. 213-222, 2018. <https://doi.org/10.1007/s41060-017-0093-7>
- [5] V. Cordes and L. Olfman, "Sports Analytics: Predicting Athletic Performance with a Genetic Algorithm," *Twenty-second Americas Conference on Information Systems*, San Diego, 2016.
- [6] M. Casals and C. F. Finch, "Sports Biostatistician: A Critical Member of All Sports Science and Medicine Teams for Injury Prevention," *Injury Prevention*, Vol. 23, No. 6, pp. 423-427, 2017. <https://doi.org/10.1136/bjsports-2016-042211rep>
- [7] R. S. Sikka, M. Baer, A. Raja, M. Stuart, and M. Tompkins, "Analytics in Sports Medicine: Implications and Responsibilities that Accompany the Era of Big Data," *The Journal of Bone & Joint Surgery*, Vol. 101, No. 3, pp. 276-283, February 2019. <http://dx.doi.org/10.2106/JBJS.17.01601>
- [8] P. Kim, "Predicting the Outcome of Korean Professional Basketball Games and Applying Sports Betting Using Artificial Intelligence Algorithms," *The Korean Journal of Physical Education*, Vol. 62, No. 5, pp. 339-361, September 2023. <https://doi.org/10.23949/kjpe.2023.9.62.5.23>
- [9] P. Kim, S. S. Jeon, and S. H. Lee, "The Application of Machine Learning Algorithms to Predict English Premier League Match Results," *The Korean Journal of Physical Education*, Vol. 62, No. 4, pp. 337-353, 2023, <https://doi.org/10.23949/kjpe.2023.7.62.4.24>



- [10] P. Kim, S. S. Jeon, and S. H. Lee, "A Study on the Application of Machine Learning to Predict Keirin Competition Ranking and Sports Betting Methods," *Journal of Korea Service Management Society*, Vol. 24, No. 2, pp. 157-192, June 2023. <https://doi.org/10.15706/jksms.2023.24.2.007>
- [11] B. C. Alamar, *Sports Analytics*, Columbia University Press, 2013.
- [12] R. P. Bunker and F. Thabtah, "A Machine Learning Framework for Sport Result Prediction," *Applied Computing and Informatics*, Vol. 15, No. 1, pp. 27-33, January 2019. <https://doi.org/10.1016/j.aci.2017.09.005>
- [13] P. Kim and S. H. Lee, "The Application of Big Data Analysis in Comparison of Machine Learning Algorithms to Predict Korean Professional Basketball League Team Results," *The Korean Journal of Physical Education*, Vol. 62, No. 2, pp. 263-277, 2023. <https://doi.org/10.23949/kjpe.2023.3.62.2.19>
- [14] P. Kim and S. H. Lee, "The Final Ranking Prediction of the Korean Professional Basketball League Using Machine Learning Algorithms: A Sports Analytics Perspective," *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, Vol. 25, No. 2, pp. 1-17, 2023. <https://doi.org/10.21797/ksme.2023.25.2.001>
- [15] R. J. Puerzer, "From Scientific Baseball to Sabermetrics: Professional Baseball as a Reflection of Engineering and Management in Society," *NINE: A Journal of Baseball History and Culture*, Vol. 11, No. 1, pp. 34-48, 2002. <https://doi.org/10.1353/nin.2002.0042>
- [16] B. S. Baumer, G. J. Matthews, and Q. Nguyen, "Big Ideas in Sports Analytics and Statistical Tools for Their Investigation," *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 15, No. 6, e1612, November/December 2023. <https://doi.org/10.1002/wics.1612>
- [17] T. W. Miller, *Sports Analytics and Data Science: Winning the Game with Methods and Models*, FT Press, 2015.
- [18] L. Passfield and J. G. Hopker, "A Mine of Information: Can Sports Analytics Provide Wisdom from Your Data?," *International Journal of Sports Physiology and Performance*, Vol. 12, No. 7, pp. 851-855, 2017. <http://dx.doi.org/10.1123/ijspp.2016-0644>
- [19] V. Sarlis and C. Tjortjis, "Sports Analytics—Evaluation of Basketball Players and Team Performance," *Information Systems*, Vol. 93, 101562, November 2020. <https://doi.org/10.1016/j.is.2020.101562>
- [20] N. Singh, "Sport Analytics: A Review," *Learning*, Vol. 9, No. 1, pp. 64-69, 2020. <https://doi.org/10.2991/itmr.k.20.0831.001>
- [21] A. Srivastava, A. Chaudhary, D. Gupta, and A. Rana, "Usage of Analytics in the World of Sports," in *Proceedings of 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)[ICRITO]*, Noida, India, 2021. <https://doi.org/10.1109/ICRITO51393.2021.9596466>
- [22] T. H. Davenport, "What Businesses Can Learn from Sports Analytics," *MIT Sloan Management Review*, Vol. 55, No. 4, pp. 10-13, 2014.
- [23] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.
- [24] P. Kim and S. H. Lee, "Application of AI Machine Learning Algorithms to Predict Korea Ladies Professional Golf Association (KLPGA) Players Top 10 Ranking: A Sports Analytics Perspective," *Korean Journal of Sport Management*, Vol. 28, No. 4, pp. 51-66, August 2023. <https://doi.org/10.31308/KSSM.28.4.51>
- [25] S. W. An and Y. M. Kim, "Prediction of Key Variables Affecting NBA Playoffs Advancement: Focusing on 3 Points and Turnover Features," *Journal of Intelligence and Information Systems*, Vol. 28, No. 1, pp. 263-286, 2022. <https://doi.org/10.13088/jiis.2022.28.1.263>
- [26] Y. Z. Ni and S. N. Lee, "Comparison of Prediction Performance of Machine Learning Classification Model Using 2022 FIBA Men's Basketball Asian Cup Match Results," *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, Vol. 24, No. 3, pp. 53-69, 2022. <https://doi.org/10.21797/ksme.2022.24.3.005>
- [27] R. P. Alonso and M. B. Babac, "Machine Learning Approach to Predicting a Basketball Game Outcome," *International Journal of Data Science*, Vol. 7, No. 1, pp. 60-77, 2022. <https://doi.org/10.1504/IJDS.2022.124356>
- [28] T. Horvat, L. Havaš, and D. Srpak, "The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes," *Symmetry*, Vol. 12, No. 3, 431, 2020. <https://doi.org/10.3390/sym12030431>
- [29] N. H. Nguyen, D. T. A. Nguyen, B. Ma, and J. Hu, "The Application of Machine Learning and Deep Learning in Sport: Predicting NBA Players' Performance and Popularity," *Journal of Information and Telecommunication*, Vol. 6, No. 2, pp. 217-235, 2022. <https://doi.org/10.1080/24751839.2021.1977066>
- [30] F. Thabtah, L. Zhang, and N. Abdelhamid, "NBA Game Result Prediction Using Feature Analysis and Machine Learning," *Annals of Data Science*, Vol. 6, No. 1, pp.

- 103-116, 2019. <https://doi.org/10.1007/s40745-018-00189-x>
- [31] J. H. Kim, S. M. Cho, and J. Y. Kand, "A Study on Optimization of Convolutional Neural Network (CNN) for Win-Loss Prediction of Baseball Game," *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, Vol. 24, No. 4, pp. 153-165, 2022. <https://doi.org/10.21797/ksme.2022.24.4.013>
- [32] T. H. Kim, S. W. Lim, J. G. Koh, and J. H. Lee, "A Study on the Win-Loss Prediction Analysis of Korean Professional Baseball by Artificial Intelligence Model," *The Korean Journal of BigData*, Vol. 5, No. 2, pp. 77-84, 2020. <https://doi.org/10.36498/kbigdt.2020.5.2.77>
- [33] M. L. Huang and Y. Z. Li, "Use of Machine Learning and Deep Learning to Predict the Outcomes of Major League Baseball Matches," *Applied Sciences*, Vol. 11, No. 10, 4499, 2021. <https://doi.org/10.3390/app11104499>
- [34] K. Koseler and M. Stephan, "Machine Learning Applications in Baseball: A Systematic Literature Review," *Applied Artificial Intelligence*, Vol. 31, No. 9-10, pp. 745-763, 2017. <https://doi.org/10.1080/08839514.2018.1442991>
- [35] A. S. Yaseen, A. F. Marhoon, and S. A. Saleem, "Multi-modal Machine Learning for Major League Baseball Playoff Prediction," *Informatica*, Vol. 46, No. 6, pp. 1-9, 2022. <https://doi.org/10.31449/inf.v46i6.3864>
- [36] R. Baboota and H. Kaur, "Predictive Analysis and Modelling Football Results Using Machine Learning Approach for English Premier League," *International Journal of Forecasting*, Vol. 35, No. 2, pp. 741-755, April-June 2019. <https://doi.org/10.1016/j.ijforecast.2018.01.003>
- [37] D. Berrar, P. Lopes, J. Davis, and W. Dubitzky, "Guest Editorial: Special Issue on Machine Learning for Soccer," *Machine Learning*, Vol. 108, pp. 1-7, 2019. <https://doi.org/10.1007/s10994-018-5763-8>
- [38] A. C. Constantinou, "Dolores: A Model That Predicts Football Match Outcomes from All Over the World," *Machine Learning*, Vol. 108, No. 1, pp. 49-75, 2019. <https://doi.org/10.1007/s10994-018-5703-7>
- [39] M. Herold, F. Goes, S. Nopp, P. Bauer, C. Thompson, and T. Meyer, "Machine Learning in Men's Professional Football: Current Applications and Future Directions for Improving Attacking Play," *International Journal of Sports Science & Coaching*, Vol. 14, No. 6, pp. 798-817, 2019. <https://doi.org/10.1177/1747954119879350>
- [40] S. Khan and V. B. Kirubanand, "Comparing Machine Learning and Ensemble Learning in the Field of Football," *International Journal of Electrical and Computer Engineering*, Vol. 9, No. 5, 4321, 2019. <https://doi.org/10.11591/ijece.v9i5.pp4321-4325>
- [41] P. Borowski and M. Chlebus, *Machine Learning in the Prediction of Flat Horse Racing Results in Poland*, University of Warsaw Faculty of Economic Sciences, 2021. <https://doi.org/10.13140/RG.2.2.22254.95043>
- [42] W. C. Chung, C. Y. Chang, and C. C. Ko, "A SVM-based Committee Machine for Prediction of Hong Kong Horse Racing," in *Proceedings of 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*, Pattaya, Thailand, 2017. <https://doi.org/10.1109/UMEDIA.2017.8074091>
- [43] P. Kim, S. H. Lee, and S. S. Jeon, "A Study on the Prediction and Evaluation of Keirin Competition Rankings Using Machine Learning Application," *Korean Journal of Sport Management*, Vol. 28, No. 2, pp. 76-94, April 2023. <https://doi.org/10.31308/KSSM.28.2.76>
- [44] Research and Markets, *Football Market Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2023-2028*, IMARC Group, 2022.
- [45] H. J. Choi, "Comparison of Machine Learning Methods for a Prediction of Match Outcomes in Soccer," *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, Vol. 24, No. 4, pp. 81-91, 2022. <http://doi.org/10.21797/ksme.2022.24.4.081>
- [46] R. Bunker and T. Susnjak, "The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review," *Journal of Artificial Intelligence Research*, Vol. 73, pp. 1285-1322, December 2019. <https://doi.org/10.13140/RG.2.2.22427.62245>
- [47] M. C. Malamatinos, E. Vrochidou, and G. A. Papakostas, "On Predicting Soccer Outcomes in the Greek League Using Machine Learning," *Computers*, Vol. 11, No. 9, 133, 2022. <https://doi.org/10.3390/computers11090133>
- [48] G. Biau and E. Scornet, "A Random Forest Guided Tour," *Test*, Vol. 25, pp. 197-227, 2016. <https://doi.org/10.1007/s11749-016-0481-7>
- [49] T. Horvat and J. Job, "The Use of Machine Learning in Sport Outcome Prediction: A Review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 10, No. 5, e1380, September/October 2020. <https://doi.org/10.1002/widm.1380>
- [50] D. Iskandaryan, F. Ramos, D. A. Palinggi, and S. Trilles, "The Effect of Weather in Soccer Results: An Approach Using Machine Learning Techniques," *Applied Sciences*, Vol. 10, No. 19, 6750, 2020. <https://doi.org/10.3390/app10196750>
- [51] P. Zuccolotto, M. Sandri, and M. Manisera, "Spatial Performance Analysis in Basketball with CART, Random Forest and Extremely Randomized Trees," *Annals of*

*Operations Research*, Vol. 325, No. 1, pp. 495-519, 2023.  
<https://doi.org/10.1007/s10479-022-04784-3>

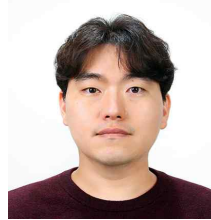
- [52] C. J. Lu, T. S. Lee, C. C. Wang, and W. J. Chen, "Improving Sports Outcome Prediction Process Using Integrating Adaptive Weighted Features and Machine Learning Techniques," *Processes*, Vol. 9, No. 9, 1563, 2021. <http://doi.org/10.3390/pr9091563>
- [53] S. D. Yun, S. A. Lee, and J. H. Kim, "Improving Team's Scoring Ability in KBO(Korea Baseball Organization) League Using Each Batting Order's Statistics and Regression Model of Machine Learning," *Korean Journal of Convergence Science*, Vol. 12, No. 11, pp. 261-275, November 2023. <https://doi.org/10.24826/KSCS.12.11.17>
- [54] I. Behravan and S. M. Razavi, "A Novel Machine Learning Method for Estimating Football Players' Value in the Transfer Market," *Soft Computing*, Vol. 25, No. 3, pp. 2499-2511, 2021. <https://doi.org/10.1007/s00500-020-05319-3>
- [55] S. Guan and X. Wang, "Optimization Analysis of Football Match Prediction Model Based on Neural Network," *Neural Computing and Applications*, Vol. 34, pp. 2525-2541, 2022. <https://doi.org/10.1007/s00521-021-05930-x>
- [56] J. H. Yi and S. W. Lee, "Prediction of English Premier League Game Using an Ensemble Technique," *KIPS Transactions on Software and Data Engineering*, Vol. 9, No. 5, pp. 161-168, 2020. <https://doi.org/10.3745/KTSDE.2020.9.5.161>
- [57] G. Buzsaki, A. Peyrache, and J. Kubie, "Emergence of Cognition from Action," in *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 79, pp. 41-50, Cold Spring Harbor Laboratory Press, 2014. <https://doi.org/10.1101/sqb.2014.79.024679>
- [58] C. Deutscher, M. Otting, S. Schneemann, and H. Scholten, "The Demand for English Premier League Soccer Betting," *Journal of Sports Economics*, Vol. 20, No. 4, pp. 556-579, 2019. <https://doi.org/10.1177/1527002518794803>



### 김필수(Philsoo Kim)

2013년 : 연세대학교 대학원  
 (경영학석사)  
 2022년 : 한양대학교 대학원  
 (스포츠산업학박사)

2014년~2016년: 한국연구재단 Global PhD Fellow  
 2015년~2017년: 경기대학교 외래교수  
 2022년~현 재: 한국스포츠경영전략연구원 원장  
 ※관심분야: 스포츠경영, 스포츠 애널리틱스, 인공지능, 프로스포츠 등



### 이상현(Sang Hyun Lee)

2011년 : 아주대학교 대학원  
 (경영학석사)  
 2024년 : 아주대학교 대학원  
 (경영학박사)

2017년~2018년: 아주경영연구소 연구원  
 2019년~2022년: G. Lab 연구원  
 2023년~현 재: 한국스포츠경영전략연구원 부원장  
 ※관심분야: 스포츠 애널리틱스, 감독 역량, 리더십, 조직문화



### 전성삼(Seong-Sam Jeon)

2021년 : 인하대학교  
 (문화콘텐츠문화경영학과 학사)  
 (스포츠과학과 부전공)

2022년~2023년: 에이치앤컨설팅 연구원  
 2024년~현 재: 한국스포츠경영전략연구원  
 데이터 사이언티스트  
 ※관심분야: 스포츠 애널리틱스, 프로스포츠, 인공지능 등