

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/372110305>

The Final Ranking Prediction of the Korean Professional Basketball League Using Machine Learning Algorithms: A Sports Analytics Perspective

Article in *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science* · July 2023

DOI: 10.21797/ksme.2023.25.2.001

CITATIONS

2

READS

502

2 authors:



Philsoo Kim

Korea Sport Management Research Institute

66 PUBLICATIONS 105 CITATIONS

[SEE PROFILE](#)



Sang-Hyun Lee

Korea Sport Management Research Institute

32 PUBLICATIONS 35 CITATIONS

[SEE PROFILE](#)

머신러닝을 활용한 한국프로농구 정규리그 최종 순위 예측: 스포츠 애널리틱스 관점

김필수 한국스포츠경영전략연구원 · 이상현* 아주대학교

국문초록

현대 스포츠 경기분석 데이터의 축적과 다양하게 집적된 기술의 발달로 인공지능(artificial intelligence)이 대중화되면서 스포츠 측정평가의 분야에서도 애널리틱스의 발전이 가속화되고 있다. 스포츠 애널리틱스는 스포츠 조직 관련 데이터를 기반으로 스포츠 비즈니스의 성과를 높이기 위한 의사결정에 도움이 될 수 있는 일련의 기술과 방식을 의미한다. 아울러, 최근 스포츠 애널리틱스는 미래의 성과 향상이라는 목적 달성을 위해 빅데이터를 활용한 미래 예측의 영역을 포함한다. 하지만, 아직 일반적으로 사용되는 예측이라는 개념에 부합하는 연구의 발전이 어려운 실정이다. 본 연구는 스포츠 애널리틱스의 발전을 위해 기계학습을 적용하여 한국프로농구 정규리그 최종순위를 예측하고 이에 대한 체계화된 논의의 진행을 목적으로 한다. 이를 위해, 한국프로농구 2006~2007 - 2022~2023시즌의 전체 경기 데이터를 활용하여 최종 라운드를 제외한 각 라운드 종료 시점에서 해당 시즌의 결과를 예측하는 실증분석을 시행하였다. 본 연구의 분석을 위해 머신러닝 알고리즘인 서포트 벡터 머신(support vector machine, SVM), 그래디언트 부스팅(gradient boosting), 엑스트라 트리(extra tree)가 사용되었다. 분석결과, 1라운드가 종료된 시점의 평균 오차는 1.88이었으나 라운드가 진행될수록 1.47, 1.21, 0.92, 0.72로 줄어든다는 것을 확인할 수 있었다. 또한, 1라운드 종료 시점에는 서포트 벡터 머신이 1.85로 가장 낮은 오차를 보였지만, 2라운드 종료 이후부터는 엑스트라 트리가 가장 낮은 평균 오차를 기록하였다. 전반적으로 가장 높은 성능을 보인 엑스트라 트리 기준으로 4라운드 이후로는 대체로 실제 시즌 종료 후의 팀 순위와 유사한 결과를 예측하는 것으로 나타났다. 본 연구는 시즌이 종료되기 전 스포츠 경기와 관련된 빅데이터를 바탕으로 라운드별 및 시즌 종료 결과를 예측한 선도적인 연구로서 스포츠 애널리틱스 관점에서 스포츠 경기 데이터의 기계학습이 적용된 분석으로서 그 가치와 의미를 지닌다.

주요어: 스포츠 애널리틱스, KBL 정규리그, 순위예측, 머신러닝, 서포트 벡터 머신, 그래디언트 부스팅, 엑스트라 트리

I. 서론

4차산업혁명으로 대변되는 인공지능(AI)의 발전은

이제는 미래의 기술이 아닌 우리의 삶의 일부가 되었다. 알파벳(구글의 모기업)의 자회사인 딥마인드가 ‘알파고(AlphaGo)’를 개발함으로써 인공지능 기술 시대의 등장을 알렸다면, Open AI의 ChatGPT와 더불어 구글 역시 생성형 AI인 Bard를 전격적으로 공개하

* 교신저자 이상현(ich45337@hanmail.net)

경기도 수원시 영통구 월드컵로 206 아주대학교 다산관 509호

여 시범 운영하게 되면서 빅데이터를 활용한 AI가 전문가의 영역에만 국한된 것이 아닌 일반인도 보편적으로 활용할 수 있는 기술이라는 것을 전 세계가 인지할 수 있게 만들었다. 빅데이터를 활용한 AI의 보편화는 사회 전반의 많은 시스템을 바꾸었으며, 사회의 변화는 스포츠 과학 분야에도 영향을 미치고 있다.

스포츠 경기와 관련한 양질의 빅데이터가 축적되고 활용되면서 이에 기반한 인공지능과 그 하위 범주에서의 기계학습(machine learning)의 적용이 현대 스포츠 과학에서 중요하게 작용하고 있다(김필수, 이상현, 2023a; Bunker & Susnjak, 2022; Bunker & Thabtah, 2019; Haghighat, Rastegari, Nourafza, Branch, & Esfahan, 2013; Horvat & Job, 2020; Keshtkar Langaroudi & Yamaghani, 2019; Thabtah, Zhang, & Abdelhamid, 2019). 이런 변화는 스포츠 애널리틱스의 발전에도 영향을 미쳤다. 스포츠 애널리틱스는 비즈니스 애널리틱스에서 먼저 개념화된 용어로 최적의 의사결정과 솔루션을 위해 생성되는 데이터를 효율적으로 수집·분석하여 새로운 인사이트를 찾아내고 경기력으로 발현될 수 있는 예측의 영역에서 정밀성을 추구하는 과정으로 정의될 수 있다(Alamar, 2013; Apostolou & Tjortjis, 2019; Baumer, Matthews, & Nguyen, 2023; Singh, 2020; Srivastava, Chaudhary, Gupta, & Rana, 2021). 아울러, 스포츠 애널리틱스는 의사결정의 결과로 파생되는 미래를 예측하여 최적의 의사결정을 도출하는 것을 목적으로 한다(장영재, 2017; Alamar, 2013; Davenport, 2006; Fried & Mumcu, 2016; Jayal, McRobert, Oatley, & O'Donoghue, 2018; Troilo & Bouchet, 2022). 이를 위해 방대한 스포츠 빅데이터의 수집, 정제 및 처리하는 능력이 필수적이다(Davenport, 2014; Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Hung Byers, 2011).

스포츠 애널리틱스는 그 목적에 따라 큰 틀에서 '경기와 관련한 분석'(on-field analytics)과 '경기 외적의 분석'(off-field analytics)으로 구분된다. 본 연구

에서는 경기와 관련한 기록들을 분석하고 경기력과 연관된 다양한 변수들에 대해 선행적으로 예측함으로써 감독과 코치 등의 의사결정자가 팀 승리를 위한 최적의 의사결정을 할 수 있도록 경기와 관련한 분석(Srivastava et al., 2021; Troilo & Bouchet, 2022)에 초점을 맞추고자 한다. 이는 주로 팀/선수의 경기력 측면의 전력 평가, 연봉산정, 특정 시점에서 경기의 승리 가능성 및 선수 로스터에 따른 승리 가능성을 추정하고 이에 따라 최적의 의사결정과 전략적 선택을 효율화하는 것에 목적을 둔다(Alamar, 2013; Baumer, Matthews, & Nguyen, 2023).

최근 머신러닝을 적용한 경기와 관련한 분석은 경기력과 관련한 거의 모든 자료를 수집하고 축적할 수 있는 기술, 컴퓨팅 연산능력, 알고리즘의 발달에 따라 스포츠 경기결과 예측이 스포츠 애널리틱스의 주요한 연구 분야로 나타나기 시작했다(김태훈, 임성원, 고진광, 이재학, 2020; 김필수, 이상현, 2023a; 김필수, 이상현, 전성삼, 2023; 서영진, 문형우, 우용태, 2019; 예원진, 이성노, 2022; 최형준, 2020; 2022; 한정섭, 정다현, 김성준, 2022; Alonso & Babac, 2022; Bunker, & Thabtah, 2019; Horvat, Havaš, & Srpak, 2020; Huang & Li, 2021; Morgulev, Azar, & Lidor, 2018; Pantzalis & Tjortjis, 2020; Srivastava, Chaudhary, Gupta, & Rana, 2021).

머신러닝을 통한 분석의 강력한 장점 중의 하나는 널리 흩어져 있는 스포츠 데이터에 대한 편견을 배제하고 분석을 진행함으로써 기존의 이론으로는 파악하지 못한 명확한 데이터의 패턴을 찾아내는데 것이다. 이러한 흐름에서 머신러닝을 적용하여 경기결과를 예측하는 연구는 농구(김필수, 이상현, 2023a; 예원진, 이성노, 2022; Alonso & Babac, 2022; Nguyen, Nguyen, Ma, & Hu, 2022; Thabtah, Zhang, & Abdelhamid, 2019), 야구(김주학, 조선미, 강지연, 2022; 김태훈, 임성원, 고진광, 이재학, 2020; 서영진, 문형우, 우용태, 2019; Huang & Li, 2021; Koseler & Stephan, 2017; Yaseen, Marhoon, &

Saleem, 2022), 축구(김주학, 강지연, 조선미, 2020; 최형준, 2022; Baboota & Kaur, 2019; Berrar, Lopes, Davis, & Dubitzky, 2019; Herold, Goes, Nopp, Bauer, Thompson, & Meyer, 2019) 등의 프로스포츠에 더해 경륜이나 경마와 같은 스포츠베팅 산업에 속한 스포츠(김필수, 이상현, 전성삼, 2023; Hubáček, Šourek, & Železný, 2019; Stübinger, Mangold, & Knoll, 2019)까지 폭넓게 생산되고 있다.

하지만 아직 머신러닝을 적용한 스포츠 애널리틱스 연구의 발전은 제한적이다. 인공지능을 활용한 애널리틱스가 가진 강점 중 하나는 미래에 대한 예측이다. 최근 발견되는 타 분야에서 인공지능을 활용한 애널리틱스는 주식 예측, 사고 가능성 예측 등 현 상황을 바탕으로 미래 상황을 예측하고 이에 대응해 나아갈 방법을 제공하는 분야에서 활발하게 활용된다(노병준, 2021; 이창열, 2022; 이철현, 오염덕, 2022).

하지만, 국내에서 발견할 수 있는 스포츠 경기결과 혹은 시즌 성적예측을 주제로 하는 연구들은 대부분 이를 예측하기 위해 투입되는 독립변수와 종속변수(경기결과, 시즌 성적)가 같은 시점이나 구간에 측정된다(예원진, 이성노, 2022; 김필수, 이상현, 2023a). 이는 일반적으로 사용되는 예측이라는 개념보다 성적에 대한 설명의 개념이 더 적절해 보인다. 현재까지 축적된 스포츠 데이터와 이를 분석할 수 있는 머신러닝 알고리즘 및 기술 수준의 발전으로 미루어 볼 때, 사회 곳곳에서 발견되는 예측과 관련한 분석이 스포츠 분야에서 충분히 이루어질 수 있을 것으로 보인다.

경기 결과(성적)의 예측이라는 측면에서 볼 때, 예상하지 못한 요인에 의해 결과가 달라질 가능성이 큰 개별 경기와 비교하면, 한 시즌의 성적은 팀 전반의 전력에 더 많은 영향을 받을 수 있다는 점에서 머신러닝을 적용한 성적예측이 비교적 정확할 수 있다(김필수, 이상현, 2022). 이에 따라, 본 연구는 머신러닝을 활용하여 전술한 스포츠 애널리틱스의 관점을 이론적으로 조명하고, 시즌 종료에 앞서 프로스포츠팀의 성적을

예측함으로써 머신러닝을 활용한 스포츠 애널리틱스 연구의 발전 방향을 제시하고자 한다. 스포츠 애널리틱스가 미래를 예측하여 의사결정 선택에 영향을 미치는 것을 목표로 한다는 점에서 이를 충실히 반영하는 연구의 수행은 시의적절한 것으로 판단된다. 이를 달성하기 위해 본 연구에서는 국내 4대 프로스포츠 중의 하나인 한국프로농구(KBL)팀의 정규리그 주요 시점까지의 경기기록을 바탕으로 시즌 종료 후의 성적을 머신러닝 알고리즘을 활용하여 예측하고자 한다. 이 과정에서 시즌 성적을 예측하기에 적절한 머신러닝 알고리즘을 비교하여 각각이 지니는 특징을 첨예하게 분석함으로써 한국프로농구가 가지는 경기의 특성과 정합성이 좋은 머신러닝 알고리즘을 제안하고자 한다. 한국프로농구는 1997년을 시작으로 오랜 기간 리그가 운영되면서 경기기록에 대한 관리가 체계화되고 있으며, 많은 득점으로 인해 발생하는 분절된 플레이 덕분에 다양한 세부 기록을 축적해왔다. 따라서 프로농구는 머신러닝을 활용하여 경기기록을 학습하고 예측하기에 적절한 종목이라고 볼 수 있다.

아직 국내 프로스포츠를 샘플로 머신러닝을 적용하여 경기결과 혹은 시즌 성적을 예측한 연구는 찾기 힘든 실정이다. 본 연구는 시즌이 진행되는 다양한 단계에서 현재까지 축적된 경기 데이터를 과거의 기록을 바탕으로 분석하여 미래를 예측한다는 측면에서 머신러닝을 활용한 스포츠 애널리틱스의 이론적 발전 및 분석기술 향상에 의미가 있을 것으로 보인다.

II. 연구방법

1. 연구대상 자료 및 변수

본 연구를 수행하기 위해 한국프로농구(KBL) 홈페이지에서 2006~2007시즌부터 2022~2023시즌까지의 경기별 데이터를 파이썬(Python)의 Beautiful Soup, Selenium 등을 활용하여 웹 스크래핑(web

scraping)을 진행하였으며 이후 본 연구에 머신러닝 분석에 필수적인 변수를 체계적으로 정리하였다. 각각의 시즌별로 추출된 전체적인 데이터의 수는 <표 1>과 같다. 총 분석의 샘플인 9,062경기 중 2019~2020 시즌의 경우 코로나바이러스 감염증-19로 인해 시즌이 중단되게 되면서 경기 수가 다른 시즌과 달라 분석에서 제외했다. 본 연구의 머신러닝을 활용한 한국프로농구 정규리그 최종순위 예측을 위해 최종적으로는 8,636경기가 분석에 사용되었다.

표 1. 추출된 한국프로농구리그 시즌별 경기 수

시즌	경기 수	시즌	경기 수
2006~2007	540	2015~2016	540
2007~2008	540	2016~2017	540
2008~2009	540	2017~2018	540
2009~2010	538	2018~2019	540
2010~2011	538	2019~2020	426
2011~2012	540	2020~2021	540
2012~2013	540	2021~2022	540
2013~2014	540	2022~2023	540
2014~2015	540	총계	9,062

본 연구에서는 시즌이 종료된 시점에서 각 팀의 최종 순위를 보다 정밀하게 예측하기 위한 종속변수로 ‘최종 승/패 마진(승리 경기 수-패배 경기 수)’을 사용하였다. 독립변수로는 ‘홈/어웨이’ 여부나 ‘총 득점’, ‘총 실점’ 등의 다양한 경기지표를 종합적으로 사용하였다. 본 연구의 분석에 활용된 한국프로농구의 1·2차 경기지표(traditional statistics / advanced metrics)들은 Sarlis & Tjortjis (2020)가 체계적으로 정리하여 제시한 공식에 의해 계산되어 도출되었다. 이후 10개 구단의 시즌별 1~5라운드까지의 평균 지표들을 산출하여 투입하였다. 마지막으로 ‘최근 5경기 승리’, ‘최근 5경기 패배’, ‘최근 5경기 승률’은 각 라운드 마지막 경기 기준으로 설정했다. 통합적인 측면에서 본 연구의 분석에 사용된 독립/종속변수는 <표 2>와 같이 정리될 수 있다.

표 2. 분석에 사용된 변수

구분	변수
독립변수	홈/어웨이, 각 라운드 승률, 총득점, 총실점, 1쿼터 득점, 2쿼터 득점, 3쿼터 득점, 4쿼터 득점, 연장 득점, 1쿼터 실점, 2쿼터 실점, 3쿼터 실점, 4쿼터 실점, 연장 실점, 단일 선수 최다 득점, 단일 선수 최다리바운드, 단일 선수 최다어시스트, 단일 선수 최다스틸, 2점슛 성공, 2점슛 시도, 2점슛 실패, 3점슛 성공, 3점슛 시도, 3점슛 실패, 자유투 성공, 자유투 시도, 자유투 실패, 턴오버, 리바운드, 어시스트, 스틸, 블락, 파울, 속공, 턴오버, 페인트존 성공, 굿 디펜스, 자유투성공률, 2점성공률, 3점성공률, 2점슛허용, 3점슛허용, 리바운드 허용, 어시스트 허용, 스틸 허용, 블락 허용, 상대팀 파울, 상대팀 속공, 상대팀 턴오버, 상대팀 페인트존 성공, 상대팀 굿 디펜스, 상대팀 자유투성공률, 상대팀 2점성공률, 상대팀 3점성공률, EFF, 개인 반칙, PIR, Standard Tendex, Modified Tendex, TOVr, REBr, BLKr, ASTr, PRA/G, TSr, EFGr, Total_Sr, PPS, 최근 5경기 승리, 최근 5경기 패배, 최근 5경기 승률, 경기 수
종속변수	최종 승패 마진

2. 분석방법

본 연구에서 사용된 머신러닝을 위해 서포트 벡터 머신(Support Vector Machine), 그래디언트 부스팅(Gradient Boosting), 엑스트라 트리(Extra Trees)가 사용되었다. 본 연구의 종속변수는 승패가 아닌 연속적인 성격을 지니는 승패 마진이기 때문에 전반적으로 회귀 계열의 알고리즘을 사용하여 분석을 진행하였다. 본 연구에서 사용된 모든 알고리즘 기반 회귀모델의 하이퍼 파라미터는 디폴트를 유지하였으며, 예측결과와의 일관성을 위하여 random_state는 0으로 설정하였다.

본 연구의 분석을 위해 확보한 데이터를 중심으로 코로나-19에 영향이 있었던 2019~2020시즌을 제외한 2006~2007시즌부터 2021~2022시즌의 1~5라운드 평균 지표와 최종 승패 마진을 포함한 데이터를 학습 및 테스트 데이터로 활용하였다. 구체적으로, 2006~2007시즌을 테스트 데이터로, 2007~2008시즌부터 2021~2022시즌 데이터를 학습 데이터로 활용해서 이를 기반으로 2006~2007시즌의 승패 마진

을 예측했다. 이후 2007~2008시즌 데이터를 테스트 데이터로, 이를 제외한 나머지 시즌 데이터를 학습 데이터로 사용하여 2007~2008시즌의 승패 마진을 예측했다. 이 과정을 테스트 데이터가 2021~2022시즌이 될 때까지 반복한 뒤, 가장 높은 성능을 나타낸 알고리즘을 기반으로 2022~2023시즌 정규리그 최종순위를 예측했다. 본 연구를 위해 사용된 머신러닝 기법은 다음과 같다.

가. 서포트 벡터 머신

서포트 벡터 머신은 초평면을 이용하여 N 차원을 $(N-1)$ 차원으로 나누는 알고리즘으로서 선형이나 비선형 문제를 모두 해결할 수 있는 강력한 분류기로 평가받는다(Cortes & Vapnik, 1995; Noble, 2006; Pai, ChangLiao, & Lin, 2017). 서포트 벡터 머신은 과적합 가능성이 작으며, 분류와 예측 모두를 위해 사용할 수 있다는 장점이 있지만, 커널 함수에 변동 여부에 따라서 설명력이 낮아질 수 있으며, 데이터의 위치에 따라 데이터의 명확한 분류가 힘들다는 한계를 지닌다(Levandovski & Lobo, 2017; Schölkopf, Smola, & Bach, 2002). <그림 1>은 서포트 벡터 머신이 데이터를 분류하는 방법을 단순화하여 시각적으로 나타낸다고 볼 수 있다.

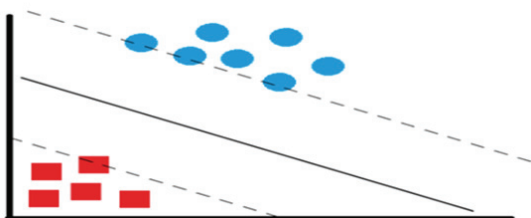


그림 1. 서포트 벡터 머신

본 연구에서는 서포트 벡터 머신으로 2022~2023시즌 순위를 예측하기 위해 sklearn.svm, LinearSVR 라이브러리를 사용하였으며, 스포츠 과학 분야의 연구

에서 서포트 벡터 머신은 기존의 연구자들이 경기 결과 예측을 위해 보편적으로 활용하여 적용하는 방법론 중의 하나라고 볼 수 있다(김필수, 이상현, 2023a; 예원진, 이성노, 2022; Ahmadalinezhad, Makrehchi, & Seward, 2019; Igiri, 2015; Pai, ChangLiao, & Lin, 2017).

나. 그래디언트 부스팅

그래디언트 부스팅은 샘플에 가중치를 할당하고 학습시키는 과정을 반복적으로 수행하여 학습시키는 앙상블 기법의 하나로 경사 하강법을 이용하여 오차를 최소화하는 기법이다. 초기에 생산한 분류기의 예측 결과와 실제 결과 간에 존재하는 차이를 계산하고 이를 다음 분류기에 적용하는 방식으로 예측결과를 도출한다. 잔차를 학습하여 모형을 개선한다는 특징 때문에 데이터 세트를 통한 예측정확도뿐만 아니라 일반화 가능성 역시 높인다는 장점을 지닌다(Arik & Pfister, 2021; Baboota & Kaur, 2019; Bradley Boehmke & Brandon Greenwell, 2020; Chen, Jhou, Lee, & Lu, 2021; Cho, Yoon, & Lee, 2018; Nguyen, Nguyen, Ma, & Hu, 2022). <그림 2>는 그래디언트 부스팅이 데이터를 분류하는 방법을 나타낸다.

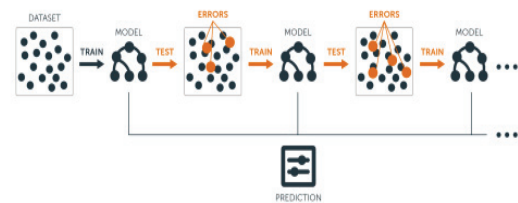


그림 2. 그래디언트 부스팅

그래디언트 부스팅은 최근 프로스포츠 분야의 경기 승패나 성과 예측을 위해 자주 사용되는 머신러닝 알고리즘 중의 하나이다(Cho, Yoon, & Lee, 2018; Phatak, Mehta, Wieland, Jamil, Connor, Bassek,

& Memmert, 2022; Wang, Qiu, & Li, 2021). 본 연구의 분석을 위해 파이썬의 sklearn.ensemble과 GradientBoostRegressor 라이브러리가 활용되었다.

다. 엑스트라 트리

엑스트라 트리는 그래디언트 부스팅과 마찬가지로 여러 개의 분류기를 동시에 사용하는 앙상블 기법의 하나로서 결정 트리를 기반으로 연산을 진행하는 알고리즘이다. 엑스트라 트리는 부트스트래핑 방식으로 샘플을 구성한 뒤 이를 여러 개의 결정 트리를 통해 분류하는 랜덤 포레스트의 변형된 형태 중 하나지만, 엑스트라 트리는 부트스트래핑을 사용하지 않고 데이터 전체를 사용한다. 전체 데이터의 특성 중 일부를 무작위로 선택한 뒤, 정보가 큰 것을 기준으로 무작위 연산을 진행하는 특성 때문에 랜덤 포레스트와 비교하면 과적합 위험이 적고 학습이 빠르다는 장점을 지닌다(Bhati & Rai, 2020; Goetz, Weber, Bloecher, Stieltjes, Meinzer, & Maier-Hein, 2014; Sharaff & Gupta, 2019; Pliakos & Vens, 2016). <그림 3>은 엑스트라 트리가 데이터를 분류하는 원리를 간략히 보여준다.

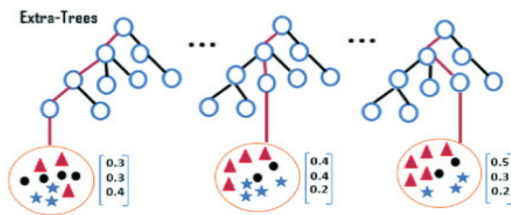


그림 3. 엑스트라 트리

이러한 종합적인 이유로 최근 스포츠 경기의 결과를 예측하기 위해 엑스트라 트리를 분석에 사용하는 연구가 늘어나기 시작했다고 볼 수 있다(Iskandaryan, Ramos, Palinggi & Trilles, 2020; Jain, Quamer, & Pamula, 2021). 본 연구에서 엑스트라 트리를 적용하기 위해 파이썬의 sklearn.ensemble, Extratrees Regressor 라이브러리를 사용했다.

III. 연구결과

본 연구에서 구축한 한국프로농구 리그의 라운드 시점의 결과를 바탕으로 6라운드가 종료된 이후 각 팀의 최종순위를 예측한 결과는 <표 3>과 같다. 분석결과, 서포트 벡터 머신, 그래디언트 부스팅, 엑스트라 트리 로 각 팀의 2022~2023시즌 최종 결과를 예측했을 때, 1~5라운드 기준 평균 팀당 오차는 1.88, 1.47, 1.21, 0.92, 0.72로 라운드가 진행될수록 오차가 낮아진다는 것을 알 수 있다. 1라운드 종료 시 시즌 최종순위를 예측할 경우 서포트 벡터 머신의 오차(278)가 가장 작았으며, 다음으로 엑스트라 트리, 그래디언트 부스팅의 순서로 결과값이 도출된다는 것을 알 수 있다.

표 3. 라운드별 예측정확도

라운드 (경기수)	서포트 벡터 머신 (팀당 오차)	그래디언트 부스팅 (팀당 오차)	엑스트라 트리 (팀당 오차)	평균 오차
1(45)	278(1.45)	288(1.92)	282(1.88)	1.88
2(90)	242(1.61)	210(1.40)	210(1.40)	1.47
3(135)	204(1.36)	174(1.16)	168(1.12)	1.21
4(180)	168(1.12)	134(0.89)	112(0.75)	0.92
5(225)	134(0.89)	100(0.67)	92(0.61)	0.72

하지만, 2라운드 종료 후 시즌 순위를 예측할 경우 그래디언트 부스팅과 엑스트라 트리의 오차가 210으로 서포트 벡터 머신(242)보다 작았다. 3~5라운드 종료 뒤에는 엑스트라 트리의 오차가 다른 알고리즘보다 작다는 것을 확인할 수 있다. <그림 4>에서는 이를 그래프로 표시하였다.

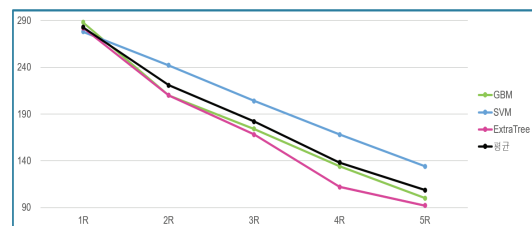


그림 4. 라운드 시점의 알고리즘별 정규시즌 결과 예측 성능

라운드 진행되어 갈수록 시즌 종료 후의 순위 예측 정확도가 높아지는 것을 확인할 수 있다. 라운드가 진행되고 실질적으로 축적되어 사용할 수 있는 정보가 많아질수록 엑스트라 트리를 통한 예측정확도가 크게 개선되는 것을 확인할 수 있다.

〈그림 5〉는 2022~2023시즌 각 라운드가 종료되었을 때 당시 팀 순위와 예측 순위를 잘 보여준다고 볼 수 있다. 1라운드 종료 시점(2022년 10월 15일 ~ 2022년 11월 10일)의 시즌 최종 예상순위는 실제 최

1라운드 종료 시점			
순위	1R 순위	1R 승률	예측 순위
1	KGC	0.800	MOBIS
2	DAYONE	0.667	DB
3	DB	0.667	DAYONE
4	SAMSUNG	0.600	KGC
5	MOBIS	0.556	SK
6	LG	0.500	LG
7	KT	0.333	SAMSUNG
8	KCC	0.333	KCC
9	SK	0.250	KT
10	KOGAS	0.222	KOGAS

2라운드 종료 시점			
순위	2R 순위	2R 승률	예측 순위
1	KGC	0.789	KGC
2	MOBIS	0.611	MOBIS
3	DAYONE	0.556	LG
4	LG	0.556	SK
5	SK	0.500	DAYONE
6	SAMSUNG	0.421	KCC
7	KOGAS	0.412	KOGAS
8	DB	0.412	DB
9	KCC	0.389	SAMSUNG
10	KT	0.333	KT

3라운드 종료 시점			
순위	3R 순위	3R 승률	예측 순위
1	KGC	0.704	KGC
2	MOBIS	0.593	MOBIS
3	LG	0.577	LG
4	SK	0.556	SK
5	DAYONE	0.481	KCC
6	KCC	0.481	DAYONE
7	KT	0.444	DB
8	KOGAS	0.407	KT
9	DB	0.407	KOGAS
10	SAMSUNG	0.357	SAMSUNG

4라운드 종료 시점			
순위	4R 순위	4R 승률	예측 순위
1	KGC	0.694	KGC
2	LG	0.639	LG
3	MOBIS	0.611	MOBIS
4	SK	0.556	SK
5	DAYONE	0.528	DAYONE
6	KT	0.444	KCC
7	DB	0.444	DB
8	KCC	0.444	KT
9	KOGAS	0.361	KOGAS
10	SAMSUNG	0.278	SAMSUNG

5라운드 종료 시점			
순위	5R 순위	5R 승률	예측 순위
1	KGC	0.739	KGC
2	LG	0.667	LG
3	MOBIS	0.600	SK
4	SK	0.600	MOBIS
5	DAYONE	0.523	DAYONE
6	KCC	0.444	KCC
7	KT	0.422	KT
8	DB	0.386	DB
9	KOGAS	0.348	KOGAS
10	SAMSUNG	0.267	SAMSUNG

정규리그 최종 순위		
순위	최종 순위	최종 승률
1	KGC	0.685
2	LG	0.667
3	SK	0.667
4	MOBIS	0.630
5	DAYONE	0.519
6	KCC	0.444
7	DB	0.407
8	KT	0.389
9	KOGAS	0.333
10	SAMSUNG	0.259

그림 5. 각 라운드 종료 시점의 팀 순위 및 예측결과

종순위와 큰 차이가 분명 존재하는 것으로 보인다. 하지만, 2라운드 종료 시점부터(2022년 11월 12일 ~ 2023년 4월 4일)의 예측 순위는 실제적인 정규리그 순위와 대부분 일치한다는 것을 실증적으로 확인할 수 가 있어 유의미한 결과로 볼 수 있다.

〈그림 6〉은 각 라운드가 종료될 때마다 예측한 최종 순위의 변화 과정을 세 머신러닝 알고리즘 중 가장 예측정확도가 높았던 엑스트라 트리로 시각화한 자료를 나타낸다. 1~2라운드 사이에 각 팀의 시즌 초반 경기력의 편차가 발생하면서 예측 순위의 변동 폭이 컸지만, 3라운드 이후에 대대로 안정적인 흐름을 이어가는 것을 확인할 수 있었다.

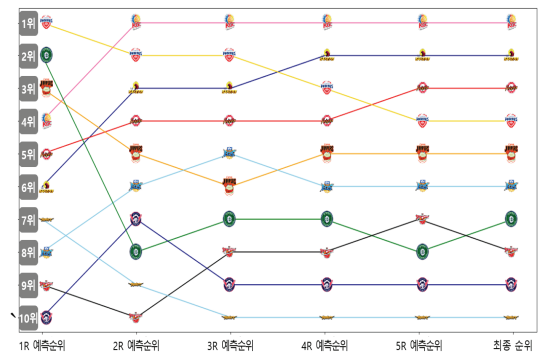


그림 6. 각 라운드 종료 시점별 정규리그 최종순위 예측 변화

IV. 논의 및 결론

본 연구는 빅데이터의 축적과 컴퓨팅 연산능력의 증대, 그리고 인공지능의 발달과 더불어 머신러닝 알고리즘의 고도화 등이 맞물려 애널리틱스가 발전하게 되면서, 스포츠 애널리틱스 관점을 기반으로 스포츠 경기결과를 예측하는 연구가 필요하다는 것을 역설하였다. 또한, 이를 달성하기 위해 머신러닝을 적용하여 한국프로농구 리그의 정규리그 순위를 예측하였다. 분석결과, 예측 시점이 시즌 종료에 가까워질수록 순위예측 가능성이 커지며, 각 예측 시점에 따라 가장 높

은 성능을 지니는 알고리즘이 달라진다는 것을 확인할 수 있다. 본 연구에서는 1라운드 종료 시점(2022년 11월 10일 기준)에는 서포트 벡터 머신의 예측 가능성이 가장 크게 나타났지만, 2라운드 종료 시점(2022년 12월 8일 기준)에는 엑스트라 트리와 그래디언트 부스팅과 같은 앙상블 기반의 알고리즘이 가장 예측정확도가 높았다고 볼 수 있다. 또한, 3라운드 종료 시점(2023년 1월 1일 기준)부터는 엑스트라 트리가 한국프로농구 정규리그 최종순위를 예측하기에 가장 적절하다는 것을 실증하였다.

먼저, 본 연구는 다음과 같은 이론적 시사점을 내재한다. 본 연구는 머신러닝을 활용하여 정규리그 시즌 종료 직전 각각의 한국프로농구팀의 정규리그 순위가 어떻게 최종적으로 수렴될 수 있는지에 대해 예측한 선도적인 연구라는 점에서 그 차별점을 지닌다. 머신러닝을 활용한 데이터 분석은 스포츠 애널리틱스의 목표 중 하나인 예측에 최적화된 방법론 중 하나라는 측면에서 스포츠 애널리틱스 학자들이 주목하고 있다(Apostolou & Tjortjis, 2019; Baumer et al., 2023; Singh, 2020; Srivastava et al., 2021). 하지만, 현재까지의 선행연구를 살펴보면 국내에서는 아직 미래 경기가 펼쳐지기 전에 현재까지 주어진 정보를 바탕으로 미래의 경기결과나 성적을 예측하는 연구가 부재하다고 본다. 일부 연구자들이 회귀분석에 기반하여 경기결과와 결정요인을 분석하거나(김필수, 이상현, 2022; 김필수, 이상현, 2023b; 장진희, 문춘걸, 2014) 머신러닝 기법을 통해 사후적으로 해당 경기에 사용되는 독립변수의 “예측” 정확도를 사용하는 경우는 존재하지만(김필수, 이상현, 2023a; 예원진, 이성노, 2022), 통상적으로 사용되는 “예측”이라는 단어에 부합하는 경기 이전에 해당 경기의 결과나 시즌이 마무리되기 전에 국한된 자료만으로 미래 결과를 예측하는 연구는 아직 국내에 존재하지 않는다. 본 연구는 이러한 측면에서 한국프로농구 데이터를 활용하여 여러 알고리즘을 비교하는 한편 다양한 라운드별 시점에

서의 기록 간의 관계를 탐색함으로써 시즌이 끝나기 전 정규리그의 최종순위를 선제적으로 예측했다는 점에서 이론적 시사점을 지닌다.

둘째, 본 연구는 여러 머신러닝 알고리즘을 비교·분석하여 한국프로농구의 맥락에서 시즌 성적을 예측하기에 적절한 알고리즘을 제시했다. 머신러닝 알고리즘은 여러 가정과 장단점을 내재하고 있으며, 이를 사용하는 선행연구에서는 연구 문제에 가장 적합한 알고리즘을 사용하고자 노력한다. 하지만, 각 알고리즘이 가진 한계점을 완전히 극복하기가 어려우므로 스포츠 경기결과 예측과 관련된 연구의 경우 한 연구 내에서도 여러 알고리즘을 비교·분석하기도 한다(김필수, 이상현, 2023a; 김필수 외, 2023; 예원진, 이성노, 2022). 본 연구에서는 본 연구의 목적이 단일경기 결과의 예측이 아닌 정규리그 종료 후의 승패 마진에 기반한 순위예측이라는 점에서 연속적인 종속변수에 대한 영향력을 설명하는 회귀 기반의 머신러닝 알고리즘을 사용하여 분석을 진행했다. 분석결과, 일반적으로 스포츠 경기결과를 분석하기 위해 많이 활용되는 서포트 벡터 머신보다 앙상블 계열의 알고리즘, 특히 엑스트라 트리가 한국프로농구의 정규리그 시즌 결과를 예측하기에 유리한 측면이 있다는 것을 확인할 수 있었다.

셋째, 본 연구에서는 분석을 시행한 시점에 따라 달라지는 머신러닝 알고리즘의 성능을 비교하였다. 앞서 기술하였듯이 머신러닝의 여러 알고리즘은 각각이 가진 가정과 연산 방법 때문에 주어진 문제의 종류에 따라 성능에 차이가 존재한다. 하지만, 본 연구의 분석 결과는 같은 문제와 알고리즘이라도 분석 시점에 따라서 성능이 달라질 수 있다는 것을 보여준다. 예를 들어, 1라운드 종료 시의 예측 성능은 서포트 벡터 머신이 가장 우수하고 그래디언트 부스팅이 가장 낮은 성능을 보여주었지만, 2라운드 종료 후 그래디언트 부스팅은 엑스트라 트리와 함께 비교군 중 가장 좋은 성능을 나타냈다. 또한, 3라운드 이후부터는 엑스트라 트리가 시즌 후 성적을 가장 잘 예측했다. 이는 같은 문제를 같

은 자료를 활용하여 해결하는 과제라 하더라도, 시점에 따라 가장 적합한 알고리즘이 달라질 수 있다는 것을 암시한다. 이러한 연구결과의 발견은 스포츠 애널리틱스의 관점에서 실질적으로 승/패를 예측하고자 하는 향후 연구에 좋은 참고자료로써 활용될 수 있을 것으로 판단된다.

이에 더해 본 연구는 다음과 같은 실무적 시사점을 지닌다. 첫째, 팀의 전략이나 선수의 조합을 재설정해서 팀의 승률을 높이는 방안을 마련하는데 본 연구의 분석결과가 사용될 수 있다. 스포츠 애널리틱스라는 분야는 본질에서 주어진 데이터를 바탕으로 미래의 경기를 비롯한 다양한 결과에 긍정적인 영향을 미치기 위해 발전해왔다. 따라서 현장 관계자들은 각 팀이 가진 상황에서 어떤 요인들이 미래 경기에 영향을 주는지를 참고하여 대비할 수 있다.

둘째, 현장 관계자들은 각 팀의 상황에 따라 적절한 포스트시즌 준비가 가능하다. 예를 들어 한국프로농구의 경우, 1·2위는 6강 플레이오프를 거치지 않고 바로 다음 라운드인 4강 플레이오프에 진출하지만 3~6위 팀들은 반드시 6강 플레이오프부터 준비해야 한다. 따라서, 정규리그 시즌이 종료된 후에 챔피언결정전까지 진출하는 과정에서 어떤 팀과 플레이오프 경기를 치르게 될 것인가에 관해 본 연구와 같이 2라운드가 종료된 시점에서 어느 정도 수렴하여 예측할 수 있다면 감독이나 코치진 및 전력 분석팀은 물론 선수들로서도 다양한 측면에서의 전술 전략과 가용 선수들을 선제적으로 준비할 수 있을 것이다.

마지막으로 스포츠 애널리틱스 분야에 대한 기여에도 불구하고 본 연구의 결과를 확장하여 발전시킬 여지가 충분하다고 본다. 첫째, 한국프로농구 종목과는 다른 특성을 내재한 다양한 프로스포츠 종목에도 이와 비슷한 유형의 연구를 수행함으로써 스포츠 애널리틱스 관점 연구의 확장과 누적에 대한 노력을 통해 연구결과의 일반화를 담보할 필요가 있다.

둘째, 중요한 것은 다양한 종속변수에 대한 예측이

라고 살펴볼 수 있다. 스포츠 애널리틱스를 실무적으로 적용하는 측면에서 프로스포츠팀은 현재 상황에서 미래 결과에 대한 예측을 정밀하게 시도해야 하고 최종적인 정규리그 시즌 순위뿐만 아니라, 개별 경기의 승리 가능성, 경기 당 점수 차이와 같이, 더욱 다양한 부분에서의 예측이 필요하다고 본다. 따라서, 향후 연구에서는 더욱 다양한 종속변수를 설정 함으로써 연구결과의 축적과 확장에 의의가 있을 것으로 기대된다.

셋째, 본 연구의 결과를 토대로 더욱 다양한 알고리즘의 정교한 예측이 필요하다. 기존 스포츠 분야의 연구에서는 여러 알고리즘을 적용해 단순하게 성능을 비교하고 그치는 경우가 대부분이었다. 하지만 본 연구에서는 알고리즘별로 시점에 따라 성능이 달라진다는 것을 확인할 수 있었다. 예를 들어, 서포트 벡터 머신의 경우 누적되는 정보의 수가 많을수록 예측력이 선형에 가까운 방식으로 높아지지만, 엑스트라 트리의 경우 특정 시점에 예측력이 높아지는 속도가 더 빨라진다는 것을 확인할 수 있었다. 즉, 같은 스포츠 경기와 목적이 라 하더라도 보유하고 있는 데이터의 속성에 따라서 가장 적합한 알고리즘의 종류가 달라져야 한다는 것을 의미한다. 향후 연구에서는 이런 점을 반영하여 체계적으로 정리한다면, 스포츠 애널리틱스 연구의 큰 진보를 이루는 데 도움이 될 것으로 판단된다.

참고문헌

- 김주학, 강지연, 조선미(2020). 머신러닝과 Kano 모델을 활용한 팬 중심 축구선수 속성 모형 개발. 한국체육측정평가학회지, 22(3), 91-104.
- 김주학, 조선미, 강지연(2022). 야구 경기 승패 예측을 위한 합성곱 신경망(CNN) 최적화 연구. 한국체육측정평가학회지, 24(4), 153-165.
- 김태훈, 임성원, 고진광, 이재학(2020). 인공지능 모델에 따른 한국 프로야구의 승패 예측 분석에 관한 연구. 한국빅데이터학회지, 5(2), 77-84.

- 김필수, 이상현(2022). 자원기반관점의 한국프로야구팀 정규 리그 승률 결정요인에 관한 연구, 한국스포츠산업경영학회지, 27(5), 16-37.
- 김필수, 이상현(2023a). 빅데이터 분석을 적용한 한국프로농구 리그 정규시즌 경기결과와 머신러닝분류모형 예측성능 비교에 관한 연구. 한국체육학회지, 62(2), 263-277.
- 김필수, 이상현(2023b). 로지스틱 회귀모형을 이용한 한국프로농구팀의 승패 결정요인 분석: 자원기반관점과 상황이론, 한국스포츠산업경영학회, 28(1), 1-15.
- 김필수, 이상현, 전성삼(2023). 머신러닝을 적용한 정류 경기 순위 예측 및 평가에 관한 연구: 2016~2022년 출주표 정보 및 경주 결과 활용, 한국스포츠산업경영학회지, 28(2), 1-19.
- 노병준(2021). 교통사고 분석 및 예측을 위한 AI 모델 개발. 대한교통학회 학술대회지, 820-838.
- 예원진, 이성노(2022). 2022 FIBA 남자농구 아시안컵 경기결과를 활용한 머신러닝 분류 모형의 예측 성능 비교. 한국체육측정평가학회지, 24(3), 53-69.
- 이창열(2022). 재난 사고 데이터 기반 위험 예측 기술 연구. Journal of the Society of Cultural Heritage Disaster Prevention, 7(2), 131-137.
- 이철현, 오염덕 (2022). AI 기반환경의 주식 시세예측을 위한 성능 비교분석 시스템. 한국컴퓨터정보학회 학술발표논문집, 30(1), 127-128.
- 장영재(2017). 스포츠 애널리틱스-산업공학의 새로운 기회. ie 매거진, 24(1), 29-35.
- 장효진, 광현, 최승희(2015). 회귀모형을 이용한 한국프로농구 승부결과 분석. 한국지능시스템학회 논문지, 25(5), 489-494.
- 최형준(2020). 국내 스포츠 빅데이터 분석 연구의 현황. 한국체육측정평가학회지, 22(2), 63-69.
- 최형준(2022). 축구의 경기 결과 예측을 위한 머신러닝 기법 비교. 한국체육측정평가학회지, 24(4), 81-91.
- 한정섭, 정다현, 김성준(2022). 머신러닝을 활용한 빅데이터 분석을 통해 KBO 타자의 OPS 예측. 차세대융합기술학회논문지, 6(1), 12-18.
- Ahmadalinezhad, M., Makrehchi, M., & Seward, N. (2019, August). Basketball lineup performance prediction using network analysis. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 519-524).
- Alamar, B. (2013). *Sports analytics. A guide for coaches, managers, and other decision makers*. Columbia University Press.
- Alonso, R. P., & Babac, M. B. (2022). Machine learning approach to predicting a basketball game outcome. *International Journal of Data Science*, 7(1), 60-77.
- Apostolou, K., & Tjortjis, C. (2019, July). Sports Analytics algorithms for performance prediction. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-4). IEEE.
- Arik, S. Ö., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 8, pp. 6679-6687).
- Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, 35(2), 741-755.
- Baumer, B. S., Matthews, G. J., & Nguyen, Q. (2023). Big ideas in sports analytics and statistical tools for their investigation. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1612.
- Berrar, D., Lopes, P., Davis, J., & Dubitzky, W. (2019). Guest editorial: special issue on machine learning for soccer. *Machine Learning*, 108, 1-7.
- Bhati, B. S., & Rai, C. S. (2020). Ensemble based approach for intrusion detection using extra tree classifier. In *Intelligent Computing in Engineering: Select Proceedings of RICE 2019* (pp. 213-220). Springer Singapore.
- Boehmke, B., & Greenwell, B. M. (2019). *Hands-on machine learning with R*. CRC press.
- Bunker, R., & Susnjak, T. (2022). The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research*, 73, 1285-1322.
- Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied computing and informatics*, 15(1), 27-33.
- Chen, W. J., Jhou, M. J., Lee, T. S., & Lu, C. J. (2021). Hybrid basketball game outcome prediction model by integrating data mining methods for the national basketball association. *Entropy*, 23(4), 477.

- Cho, Y., Yoon, J., & Lee, S. (2018). Using social network analysis and gradient boosting to develop a soccer win-lose prediction model. *Engineering Applications of Artificial Intelligence*, 72, 228-240.
- Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98-107.
- Davenport, T. H. (2014). What businesses can learn from sports analytics. *MIT Sloan Management Review*, 55(4), 10-13.
- Fried, G., & Mumcu, C. (Eds.). (2016). *Sport analytics: A data-driven approach to sport business and management*. Taylor & Francis.
- Goetz, M., Weber, C., Bloecher, J., Stieltjes, B., Meinzer, H. P., & Maier-Hein, K. (2014). Extremely randomized trees based brain tumor segmentation. *Proceeding of BRATS challenge-MICCAI*, 14, 6-11.
- Haghighat, M., Rastegari, H., Nourafza, N., Branch, N., & Esfahan, I. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: An International Journal*, 2(5), 7-12.
- Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, 14(6), 798-817.
- Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5), e1380.
- Hubáček, O., Šourek, G., & Železný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2), 783-796.
- Igiri, C. P. (2015). Support vector machine-based prediction system for a football match result. *IOSR Journal of Computer Engineering*, 17(3), 21-26.
- Iskandaryan, D., Ramos, F., Palinggi, D. A., & Trilles, S. (2020). The effect of weather in soccer results: an approach using machine learning techniques. *Applied Sciences*, 10(19), 6750.
- Jain, P. K., Quamer, W., & Pamula, R. (2021). Sports result prediction using data mining techniques in comparison with base line model. *Opsearch*, 58, 54-70.
- Jayal, A., McRobert, A., Oatley, G., & O'Donoghue, P. (2018). *Sports analytics: Analysis, visualisation and decision making in sports performance*. Routledge.
- Keshtkar Langaroudi, M., & Yamaghani, M. (2019). Sports result prediction based on machine learning and computational intelligence approaches: A survey. *Journal of Advances in Computer Engineering and Technology*, 5(1), 27-36.
- Levandoski, A., Lobo, J. (2017). *Predicting the NCAA Men's Basketball Tournament with Machine Learning*, 1-15.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5, 213-222.
- Nguyen, N. H., Nguyen, D. T. A., Ma, B., & Hu, J. (2022). The application of machine learning and deep learning in sport: predicting NBA players' performance and popularity. *Journal of Information and Telecommunication*, 6(2), 217-235.
- Noble, W. S. (2006). What is a support vector machine?. *Nature Biotechnology*, 24(12), 1565-1567.
- Pai, P. F., ChangLiao, L. H., & Lin, K. P. (2017). Analyzing basketball games by a support vector machines with decision tree model. *Neural Computing and Applications*, 28, 4159-4167.
- Pantzalis, V. C., & Tjortjis, C. (2020, July). Sports analytics for football league table and player performance prediction. In *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-8). IEEE.
- Phatak, A. A., Mehta, S., Wieland, F. G., Jamil, M., Connor, M., Bassek, M., & Memmert, D. (2022). Context is key: normalization as a novel approach to sport specific preprocessing of KPI's for match analysis in soccer. *Scientific Reports*, 12(1), 1117.

- Pliakos, K., & Vens, C. (2016). Feature induction based on extremely randomized tree paths. *Online proceedings*, 3-18.
- Sarlis, V., & Tjortjis, C. (2020). Sports analytics—Evaluation of basketball players and team performance. *Information Systems*, 93, 101562.
- Schölkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.
- Sharaff, A., & Gupta, H. (2019). Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018* (pp. 189-197). Springer Singapore.
- Singh, N. (2020). Sport analytics: A review. *learning*, 9, 11.
- Srivastava, A., Chaudhary, A., Gupta, D., & Rana, A. (2021, September). Usage of Analytics in the World of Sports. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-7). IEEE.
- Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science*, 6(1), 103-116.
- Troilo, M., & Bouchet, A. (2022). Professional Sports Organizations and Business Analytics: Monopoly Power vs Debt Financing. *Journal of Applied Sport Management*, 14(4), 4.
- Wang, L., Qiu, K., & Li, W. (2021). Sports action recognition based on GB-BP neural network and big data analysis. *Computational Intelligence and Neuroscience*, 2021.

저자정보

김필수(Philsoo Kim)

한국스포츠경영전략연구원 원장
Philsookim@naver.com

이상현(Sang Hyun Lee)

아주대학교 박사과정
ich45337@hanmail.net

논문투고일	2023년 05월 26일
심사완료일	2023년 06월 30일
게재확정일	2023년 07월 03일

Abstract

The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science. 2023, 25(2), 103-115

The Final Ranking Prediction of the Korean Professional Basketball League Using Machine Learning Algorithms: A Sports Analytics Perspective

Philsoo Kim *Korea Sports Management Research Institute* • Sang Hyun Lee *Ajou Univ.*

The advent of artificial intelligence (AI) as a result of big data accumulation and high technology has invited the word "analytics" into the sports industry. We define sports analytics as the process of pursuing better decision-making for the success of sports organizations by utilizing data from the past. This process includes a prediction of the future. Therefore, it provides analysis for stakeholders in the sports domain to help them make better decisions in their current situation. This study attempts to contribute to the field of sports analytics by predicting the results of the 2022-2023 regular season of the Korean Basketball League (KBL) teams. We collected data from the 2006-2007 to 2022-2023 season from the KBL website and analyzed it with three machine learning algorithms (Support Vector Machine, Gradient Boosting, and Extra Trees) from various points (at the end of rounds 1, 2, 3, 4, and 5). The results showed that the error of prediction decreased as the round progressed. The average error was 1.88 when the first round ended, but 0.72 when the fifth round ended. Furthermore, the Support Vector Machine showed the lowest error (1.85) among the three algorithms when the first round ended. However, Extra Trees performed the lowest error when predicting the results of the regular season after the second (1.40), third (1.12), fourth (0.75), and fifth (0.61) rounds. Overall, the predictions of Extra Trees after the third round were similar to the actual regular season results. This study established a milestone in the field of Korean sports analytics by conducting research that predicted the results of the regular season before the season actually ended.

Keywords: Sports analytics, KBL regular season, rank prediction, machine learning, support vector machine, gradient boosting, extra tree