

Introducing ‘Rapidminer’, a Big Data Analytics Tool for Football Data Analysis

축구 데이터 분석을 위한 빅데이터 분석 도구 RapidMiner 소개

Kyounggu Jung¹, Kyounggho Choi²

정경구¹, 최경호²

¹ Assistant Professor, Dept. of Sports Coaching(Soccer Major), Jeonju University, Rep. of Korea, 11663jkg@jj.ac.kr

² Professor, Department of Radiological Science, Jeonju University, Rep. of Korea, ckh414@jj.ac.kr

Corresponding author: Kyounggho Choi

Abstract: Recently, in the fields of health care, finance, and insurance, the importance of big data has already been recognized and various utilization measures such as big data opening systems are being proposed or operated. However, in the sports field, while recognizing the value and necessity of big data, it is still not active in direct execution for utilization. This is thought to be the primary cause of the difficulty in using the tool for big data analysis. Therefore, in this study, RapidMiner, a big data analysis tool, was introduced to help perform soccer data analysis more easily. In particular, by introducing examples of data handle and decision tree analysis among RapidMiner's various functions, it is expected that soccer-related researchers will help eliminate fear of big data analysis.

Keywords: Big Data, RapidMiner, Data Mining, Soccer Data, Decision Tree Analysis

요약: 최근 들어 보건의료, 금융, 보험 등의 분야에서는 이미 빅데이터의 중요성을 인식하고 그 활용에 대한 고민을 통해 빅데이터 개방시스템 등 다양한 활용방안을 제시하거나 운영 중이다. 그러나 스포츠 분야에서는 빅데이터의 가치와 필요성에 대해서는 인식하고 있으면서도, 활용을 위한 직접 실행에 대해서는 아직도 활발한 편이 아니다. 이는 빅데이터 분석을 위한 도구의 활용에 어려움이 있는 점에 일차적인 원인이 있다고 생각된다. 이에 본 연구에서는 축구 데이터 분석을 보다 쉽게 수행하는데 도움을 주고자, 빅데이터 분석 도구인 RapidMiner를 소개하였다. 특히 RapidMiner의 다양한 기능 중에서 데이터 핸들 및 의사결정나무 분석의 예를 소개함으로써, 축구 관련 연구자들이 빅데이터 분석에 대한 두려움을 없애는 데 도움이 될 것으로 기대해 본다.

핵심어: 빅데이터, 래피드마이너, 데이터 마이닝, 축구 데이터, 의사결정나무 분석

Received: November 17, 2022; 1st Review Result: December 31, 2022; 2nd Review Result: January 31, 2023
Accepted: February 28, 2023

1. 서론

축구를 포함한 모든 스포츠는 본질적으로 불확실성(uncertainty)이라는 특수성을 갖기 때문에, 자료가 발생하는 상황적 맥락과 자료의 특성에 대한 충분한 이해가 필요하다[1]. 최근 들어 컴퓨팅 기술의 비약적인 발전에 힘입어, 모든 영역에서 데이터가 폭발적으로 증가하는 빅데이터 시대가 도래하였다. 이는 스포츠 분야에서도 예외가 아니며, 수많은 데이터가 쏟아지는 스포츠 경기에서 엄청난 양의 자료를 기록하고, 분석, 의미 있는 결과물을 도출하는데 센싱(censing) 기술과 데이터 송수신 기술이 활용되고, 비디오 영상을 인식하고 자동으로 기록하는 인공지능 시스템이 등장하는 등 빅데이터와 인공지능 기술 적용의 필요성이 더욱 강조되고 있다[2][3]. 2000년대 초반에 빅데이터라는 용어가 등장한 이래 스포츠 영역에서의 빅데이터 활용 범주는 주로 선수의 경기력 향상을 위한 분석, 경기 결과 예측과 전략 수립 그리고 훈련 적을 위한 선수의 신체적 기능 상태 측정 등이 주를 이루었다[4]. 그러나 최근에 이르러 스포츠 분야에서의 빅데이터 분석 연구는 텍스트 마이닝(text mining), 소셜네트워크(social network), 인공지능망을 활용한 분류(classification)연구 등으로까지 확대되고 있다[5]. 이러한 배경에는 콘텐츠 애플리케이션의 보급 확대, 새로운 플랫폼의 보급, 네트워크 활용 증대, 각종 기기(device)의 발전 등과 같은 몇 가지 IT환경 패러다임의 변화를 꼽지 않을 수 없다[6].

스포츠 분야에서 각종 빅데이터 분석의 결과를 활용하는 가장 근본적인 이유 중의 하나는 시합(match)에서 승리하는데 필요한 정보(information)를 도출하는 데 있다. 정보는 우리 팀의 현재 상황을 파악하는 데에도 도움을 주지만, 상대편을 이해하고 맞춤 전략을 수립하는 데에도 필요하다. 빅데이터 분석을 위해서는 상기 언급된 다양한 IT 관련 기술도 필요하지만, 합목적적이고도 정확한 측정을 통하여 수집된 객관적인 자료(기록)가 더욱 중요하다. 이를 활용한 빅데이터 분석에서 가장 활발하게 이용되고 있는 분석 방법의 하나가 데이터 마이닝이다. 이를 위해 현재까지 다양한 데이터 마이닝 도구가 개발·보급되고 있다. 예컨대, Python, KNIME, ORANGE, R, RapidMiner 등이 그 예인데, [7]에 따르면 성능 면에서는 R이 좋지만, 사용자 편의성 측면에서 볼 때 RapidMiner가 비교적 우수한 것으로 평가되었다. RapidMiner는 독일 도르트문트 대학의 인공지능 연구진에 의해 YALE(yet another learning environment)이라는 이름의 오픈소스(open source) 프로젝트로 2001년에 최초로 개발되었다. RapidMiner의 장점은 다양한 데이터 형식을 연결할 수 있으며, 데이터 전처리, 모델개발, 시각화까지의 데이터 분석과정이 하나의 플랫폼에서 수행되게 되어 있어 초보자들도 매우 쉽게 배울 수 있는 빅데이터 분석 도구이다. 그러함에도 불구하고 RapidMiner를 특정 영역에서 활용할 수 있도록 소개한 연구는 [8] 등을 제외하고는 많지 않다. 국내의 경우 RapidMiner와 관련하여 단행본으로 발간된 몇 권의 도서는 있지만, 축구 분야에서 이를 활용한 연구는 없는 실정이다. 이는 RapidMiner가 아직 널리 보급되지 못한 측면도 있지만, 축구 관련 연구자들에게 쉽게 소개된 선행연구가 없기도 한 때문으로 생각된다. 오늘날 프로축구에서는 GPS 웨어러블 장비 및 IoT를 활용하여 데이터를 수집하고 있다[9]. 수집된 데이터를 정보화하기 위해서는 프로축구 분야의 다양한 관계자(연구원, 지도자, 개발자 등)들도 쉽게 분석할 수 있는 도구가 필요하다. 그런데 아직까지 이를 소개하고 있는 연구는 미미한 실정이다. 이에 본 연구에서는 이에 대한 도움을 줄 수 있도록, 기술연구그룹(TSG : technical study group)이 2020 K리그1 경기에 대해서 작성한 TSG기술보고서 데이터를 이용하여

RapidMiner를 활용한 빅데이터 분석 방법을 소개하고자 한다. 향후 프로축구 분야에서는 ICT의 발전에 더욱 힘입어 영상트래킹이나 GPS를 이용한 보다 다양한 활동형태의 기록 등이 데이터 형태로 제공될 것으로 예상되는 바, 본 연구는 이를 위한 빅데이터 분석 도구를 소개한다는 점에서 의의가 있다고 사료된다.

2. 연구방법

2.1 스포츠와 빅데이터

빅데이터에 대한 정의는 학자마다 조금씩 차이를 보이고 있지만, 방대한 양(volume), 실시간으로 이루어지는 처리 속도(velocity), 자료구성 및 형태의 다양성(variety)의 3V에 최근 들어서는 가치(value)와 진실성(veracity)을 포함하는 5V로 정의하는 데에 이견이 없는 듯하다[10]. 스포츠 분야에서 빅데이터 및 분석 기술을 적용하여 현상을 설명하거나 경기 중 나타나는 사건을 분석하고자 하는 노력 또한 지속해서 증가하는 추세이다[5]. 스포츠 분야에서 빅데이터를 적용한 예시를 [1]은 [표 1]과 같이 정리하였다.

[표 1] 스포츠 분야에서 빅데이터 적용 예

[Table 1] Example of Applying Big Data in the Field of Sports

특성	설명	스포츠 분야 예시
양	정보의 양이 제타바이트 시대로 진입	운동선수의 실시간 tracking 자료, 실시간 스포츠 경기 스트리밍에서 댓글 등
속도	데이터 생성 및 이동 그리고 처리 속도 증가	선수의 이동 거리 및 속도, 스포츠 소비자들의 실시간 상호작용 자료 등
다양성	데이터 유형의 다양성	팀 단위 경기 운영 네트워크 정보, 스포츠 팬들의 SNS 게시물 등
가치	분석 결과의 유의미한 가치	스포츠가 환경오염에 미치는 영향분석, 열악한 환경 조건이 스포츠 팬과 선수들에게 어떤 영향을 미칠 수 있는지 분석 등
진실성	수집된 자료의 진실성, 분석 결과의 정확성	스포츠 소비자 행동 예측, 스포츠 경기분석, 스포츠 조직의 운영 효율성 증진 등

2.2 분석자료

K리그의 경우 2018년부터 K리그 기술보고서(technical report)가 작성·공개하고 있다. K리그 기술보고서는 대한축구협회 산하 기술위원회 소속인 기술연구그룹(TSG: technical study group)에서 매년 발간하는 보고서이다. 이 TSG보고서에는 ‘스프린트 횟수’, ‘최고속력’ 등 다른 데서는 찾아보기 힘든 귀한 데이터들이 많다. 본 연구에서는 2020 K리그1 경기에 대해서 작성한 TSG기술보고서 데이터를 활용하여 수행된 선행연구[11] 중 [표 2]의 변수를 각각 종속변수와 독립변수로 하였다. [표 2]에 ‘최고속력’이나 ‘분 당 뒀 거리’ 등의 측정변수를 포함하지 않은 이유는, 대구FC와 수원삼성 팀의 경우 이 자료가 제시되어 있지 않기 때문이다. 본 연구의 목적이 RapidMiner 빅데이터 분석도구를 이용한 축구데이터 분석방법을 소개하는 데 있으므로, 제시되지 않은 변수는 제외하였다.

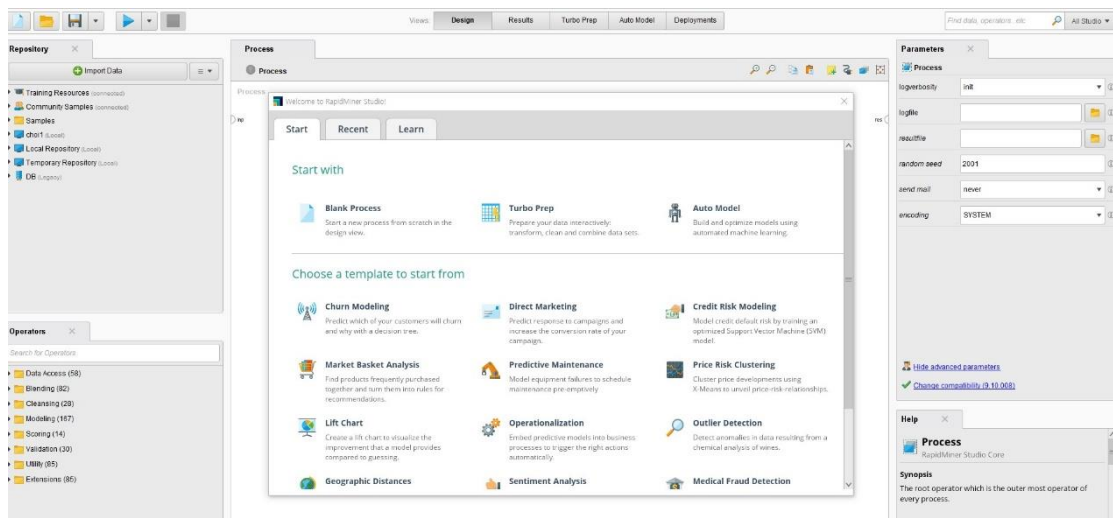
[표 2] 분석변수

[Table 2] Analysis variables

구분	측정변수
종속변수	승점
독립변수	슈팅 대비 득점 전환율(%), 평균 볼 점유율(%), 득점 당 유효슈팅(회), 경기당 유효슈팅(회), 득점 당 슈팅(회), 경기 당 슈팅(회), 총 패스 평균(회), 패스 성공률(%), 경기 당 키페스(회), 크로스 시도(회), 골키퍼 선방율(%), 태클, 인터셉트, 클리어링, 드리블, 탈압박, 공수간격(m)

2.3 분석도구

RapidMiner는 데이터 분석 및 다양한 데이터 마이닝 기법을 지원하는 분석도구이다[12]. 이를 위해 500개 이상의 오퍼레이터를 제공하는데, 기본 오퍼레이터 이외에도 웹마이닝(web mining), 텍스트마이닝(text mining), 시계열 데이터 분석(time series analysis) 등 다양한 특화된 오퍼레이터를 제공한다. 그뿐만 아니라, Oracle, IBM DB2, Microsoft SQL Server, MySQL, PostgreSQL, Access 등의 데이터베이스와 CSV, Excel, SPSS 등 다양한 데이터 포맷을 바로 읽어 처리할 수 있도록 지원하고 있다. RapidMiner의 설치는 <http://docs.rapidminer.com/studio/installation/>을 통해 무료로 내려받을 수 있다[13]. [그림 1]은 RapidMiner v. 9.10의 처음 화면이다.



[그림 1] RapidMiner 시작 화면

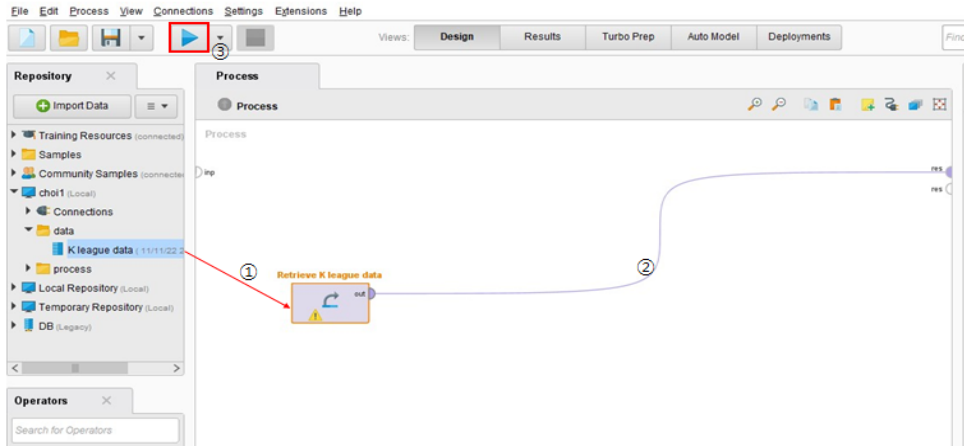
[Fig. 1] RapidMiner First Page

3. RapidMiner를 이용한 분석 소개

3.1 데이터 핸들

RapidMiner를 이용하여 빅데이터 분석을 하려면 우선 분석 대상 자료를 시스템에 불러들여야 한다. 다양한 포맷의 데이터를 불러들일 수 있지만, 본 연구에서는 가장

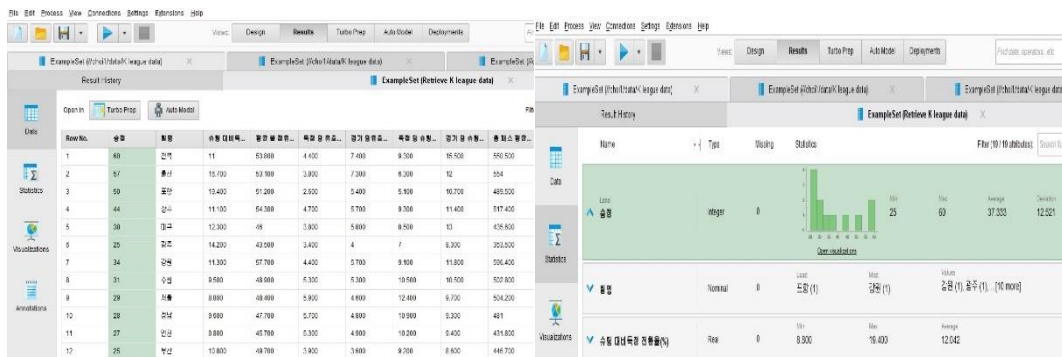
범용적으로 사용되는 Excel 자료를 이용하였다. 분석을 위하여 Process 창에 분석데이터 불러오는 방법은 여러 가지 있다. 그 중에서 'Import Data 기능'을 이용하여 외부에 저장된 자료를 불러들인 후, 왼쪽 마우스를 누른 채 'K league data'를 Process 창으로 끌고 오면(①) [그림 2]와 같이 데이터가 불러들여진다.



[그림 2] 'Import Data' 기능을 이용하여 데이터 불러오기

[Fig. 2] Importing Data Using the 'Import Data' Function

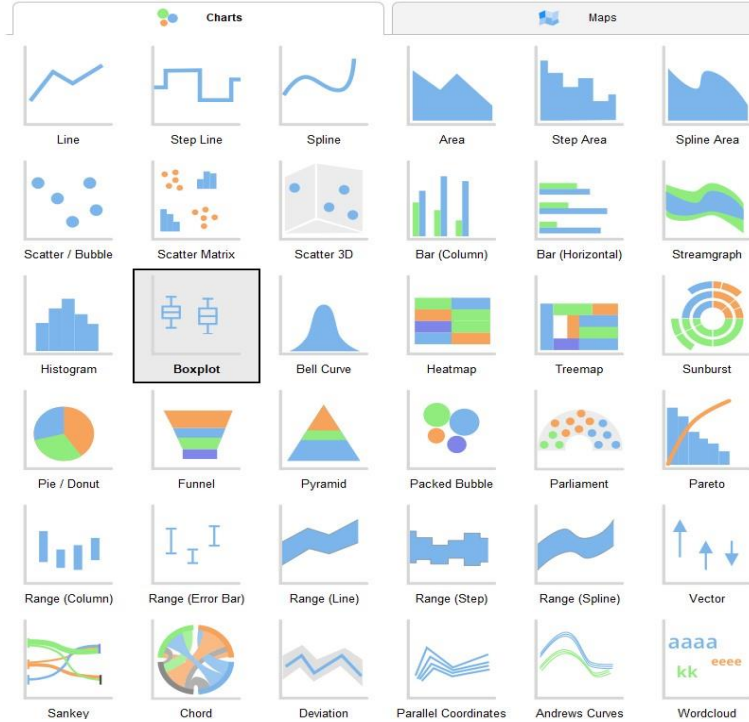
Process 창에 있는 Retrieve K league data 오퍼레이터의 out 포트를 res 포트에 연결하고(②) 실행 버튼(③)을 누른다. 그러면 [그림 3]과 같이 입력 데이터와 함께 좌측에 'Statistics'와 'Visualization' 등이 나타난다. 이때 'Statistics'를 누르면 [그림 4]와 같이 입력된 변수의 기술통계량(descriptive statistics)이 손쉽게 구해짐을 확인할 수 있다. [그림 4]는 입력된 변수 중 종속변수로 지정된 '승점'에 대한 기술통계량이다. RapidMiner에서는 종속변수를 레이블(label) 그리고 독립변수를 일반속성(regular)라고 부르는데, 'edit' 기능을 이용하여 수정할 수 있다.



[그림 3] '실행' 버튼을 누른 후 화면 [그림 4] 'Statistics'를 누른 후 기술통계량

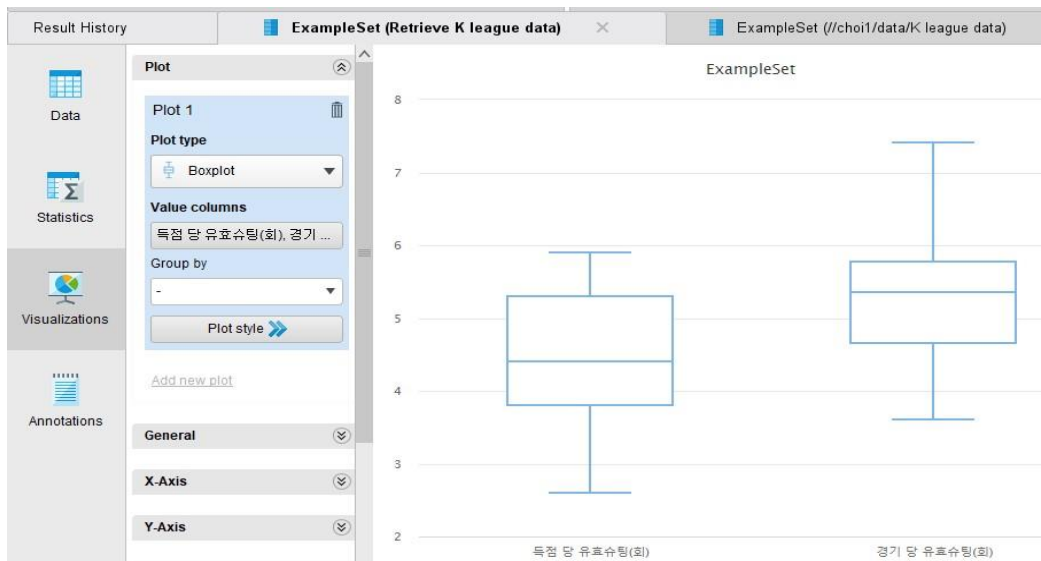
[Fig. 3] Screen After Pressing the 'Run' Button [Fig. 4] Descriptive Statistics After Pressing 'Statistics'

다음으로 'Visualization'을 누르면 [그림 5]와 같이 다양한 형태의 그림을 확인해 볼 수 있는 옵션이 주어지는데, 이때 'Plot type'을 선 'Boxplot'으로 선택하고 '득점당 유효슈팅'과 '경기당 유효슈팅' 변수에 대하여 상자그림을 그려보면 [그림 6]과 같다. 이를 통하여 임의의 두 변수 간의 비교 등 통계적인 해석을 수행해 볼 수 있다.



[그림 5] 'Visualization'에서 제공하는 그림 타입

[Fig. 5] Picture Types Provided by 'Visualization'



[그림 6] Boxplot 결과

[Fig. 6] Boxplot Results

3.2 의사결정나무 분석

데이터 특성에 따른 데이터마이닝 방법론을 정리해 보면 [표 3]과 같은데, 의사결정나무는 분류와 회귀나무(classification, regression tree, CART) 와 동의어로 사용되고 있다. 이러한 의사결정나무 분석은 자료 탐색과 모형화라는 특징을 지니고 있다[14].

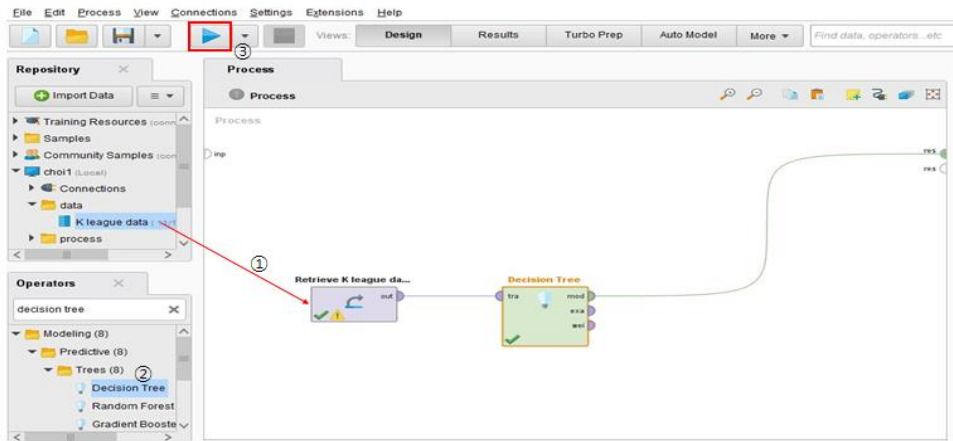
[표 3] 데이터 특성에 따른 데이터마이닝 방법론

[Table 3] Data Mining Methodology according to Data Characteristics

구분	지도		비지도
	연속형 변수	범주형 변수	무변수
연속형 예측변수	선형회귀 신경망 K-최근접이웃, 앙상블	로지스틱 회귀 신경망 판별분석	주성분분석 군집분석 협업 필터링
범주형 예측변수	선형회귀 신경망 분류와 회귀나무 앙상블	K-최근접 이웃 앙상블, 신경망 분류와 회귀나무 로지스틱회귀 나이브 베이즈	연관성 규칙 협업 필터링

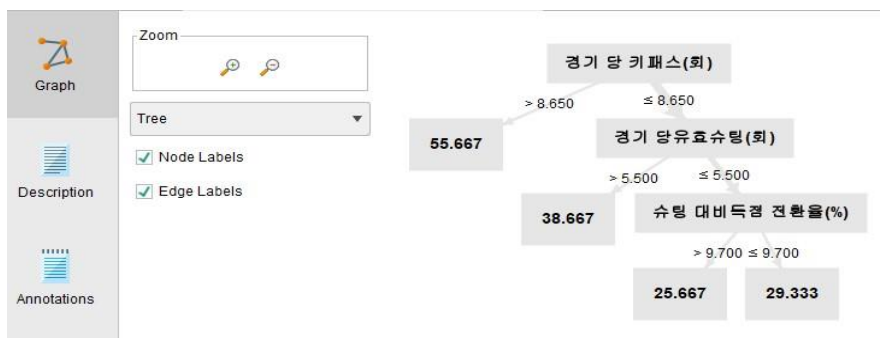
의사결정나무는 여러 개의 질문을 통해서 관심 있는 반응변수인 레이블의 값을 분류하거나 예측하는 방법으로, 이 방법은 신경망이나 판별분석 등과 달리 분류 및 예측에 적용되는 규칙을 명확하게 설명할 수 있는 장점이 있다. 의사결정나무의 장점을 정리해 보면 다음과 같다. 첫째, 이해하기 쉬운 규칙을 생성시켜준다. 둘째, 규칙 만드는 능력이 뛰어나다. 셋째, 분류(예측)작업이 쉽다. 넷째, 질적변수와 양적변수를 모두 취급할 수 있다. 다섯째, 가장 좋은 설명변수를 명확히 알아낸다. 여섯째, 특이값(outlier)에 덜 민감하다[15].

RapidMiner를 이용하여 의사결정나무 분석을 수행하는 과정을 살펴보자. 먼저 [그림 7]처럼 ‘K league data’를 Process 창으로 끌어온 후(①), 오퍼레이터 창에서 ‘Decision Tree’를 검색하여 더블클릭하면 Process 창으로 옮겨진다(②). 이때 한 가지 주의할 점은 레이블 변수가 ‘승점’처럼 양적 자료일 때에는 ‘Parameters’ 창에서 ‘criterion’을 ‘least_square’로 변환 선택해 주어야 한다. 이후 실행(③)을 누르면 [그림 8]의 의사결정나무 분석 결과를 얻게 된다. 1차 분할은 ‘경기 당 키패스(회)’ 변수임을 알 수 있다. 물론 [그림 8]의 결과는 RapidMiner를 이용하여 의사결정나무 분석을 수행하는 과정을 보여주기 위한 예이기 때문에 그 자체로 통계적인 의미를 갖지는 않는다. 한편, 의사결정나무 분석을 이용하여 수행할 수 있는 통계적인 방법으로 예측이 있는데, 이는 ‘Apply Model’ 오퍼레이터를 이용하면 된다. 다음으로 생성된 의사결정나무 모형이 얼마나 잘 맞는지를 평가해 볼 수도 있는데, 이는 ‘Performance’ 오퍼레이터를 이용하면 된다. 이상의 내용을 실제로 수행해 보는 일은 그리 어려운 작업은 아니다. 그뿐만 아니라 [16] 등의 관련 서적 또한 많이 출판되어 있는바, [17] 등의 선행연구와 함께 참고한다면 축구 관련 연구자들에게 많은 도움이 될 것으로 확신한다.



[그림 7] 'Decision Tree'를 추가하고 'res' 포트에 연결

[Fig. 7] Add a 'Decision Tree' and Connect it to the 'res' Port



[그림 8] '의사결정나무' 분석 결과

[Fig. 8] 'Decision Tree' Analysis Result

4. 요약 및 결론

4차 산업혁명이라는 용어는 이제 진부하게 느껴질 정도로 우리 생활 속 깊은 곳까지 침투하였다. 이에 따라 인공지능, 가상현실, 증강현실, 로봇, 사물인터넷, 빅데이터 등 4차 산업혁명 관련 기술들이 모든 영역에서 중심으로 자리 잡아 가고 있다. 이는 스포츠 영역도 예외가 아니어서, 다양한 형태로 스포츠와 첨단 기술 간의 융합이 이루어지고 있다. 축구의 경우 선수들의 유니폼에 부착된 센서 장비를 이용하여 경기 중 특정 선수의 '스프린트 횟수', '최고속력', '분 당 뛴 거리' 등의 데이터를 얻게 된다. 이러한 양적 데이터뿐만 아니라 동영상자료 등 비정형 데이터도 수집되는 등 그야말로 빅데이터를 얻을 수 있게 되었다. 수집된 데이터는 다양한 분석 방법을 이용하여 우리 팀과 상대 팀을 분석함으로써 승리를 위한 전략 수립에 중요한 정보로 활용될 수 있다.

다양한 스포츠 종목 중에서 데이터 활용의 빈도가 가장 높은 것으로 알려진 종목은 야구이다. 그러나 최근 들어 축구 또한 야구만큼 데이터에 기반한 통계적 수치에 의해 훨씬 과학적으로 분석될 수 있다. 과학적으로 분석된 정보는 선수 개인과 팀 전술 그리고 경기력을 객관적으로 나타내는데 기초자료가 된다. 이 점에서 축구관련자들도

통계를 포함한 빅데이터 분석을 할 수 있는 능력배양의 필요성이 제기된다. 이와 관련하여 최근 ‘스포츠 정보학’이라는 영역이 대학의 체육계열 학과에서 개설되고 있다. 스포츠 정보학이란 스포츠가 이루어지는 현장에서 경기나 선수기록 등을 통해서 발생하는 원시 자료(raw data)를 컴퓨터 등의 하드웨어와 빅데이터 분석 도구와 같은 소프트웨어를 활용하여 스포츠 관련 데이터, 정보, 지식 등을 생산하고 관리하며 활용하는 학문으로 정의해 볼 수 있다. 즉, 축구를 포함한 스포츠 분야에서도 빅데이터 분석 능력배양의 필요성이 요구되는 시대가 도래한 것이다.

본 연구에서는 축구 관련 유용한 정보를 얻기 위하여 빅데이터 분석을 수행하는 현장의 축구관련자들을 포함하여 통계나 컴퓨터 관련 초보자들도 더욱 손쉽게 수행할 수 있는 도구인 RapidMiner를 소개하였다. 소개과정에서 사용되는 데이터는 대한축구협회 산하 기술위원회 소속인 기술연구그룹(TSG: technical study group)에서 매년 발간하는 보고인 TSG보고서 데이터이고, 데이터 핸들 및 의사결정나무 분석의 예를 이용하여 RapidMiner 사용법을 설명하였다. 본 연구에서 소개한 RapidMiner는 GUI 기반 소프트웨어로, 데이터 마이닝 및 예측분석 워크플로우를 생성 및 실행할 수 있다. 수행과정에서 드래그앤드롭(drag and drop) 방식으로 수백 가지 기능을 제공하기 때문에 비전문가도 머신러닝 모델을 간편하게 개발할 수 있는 빅데이터 분석 도구이다. 본 연구가 빅데이터 분석이 필요하지만, 관련 분야 전문지식이 부족하여 어려움을 겪고 있는 축구 관련 관계자들에게 다소라도 도움이 되는 계기가 되기를 기대해 본다.

References

- [1] S. W. Hwang, H. J. Yang, S. H. Hwang, S. H., Issues of Sports Big-data Utilization and Future Applicability, Korean Journal of Convergence Science, (2022), Vol.11, No.5, pp.53-70.
Available from: <http://www.riss.kr/link?id=A108165186>
- [2] S. Kang, Trends in the use of big data and artificial intelligence in the sports field, Convergence Security Journal, (2022), Vol.22, No.2, pp.115-120.
Available from: <http://www.riss.kr/link?id=A108196291>
- [3] X. Xie, Real-time monitoring of big data sports teaching data based on complex embedded system, Microprocessors and Microsystems, (2022), Vol.81, pp.1-15.
DOI: <http://dx.doi.org/10.1016/j.micpro.2021.104181>
- [4] S. Kang, S., Utilization and prospect of big data analysis of sports contents, Convergence Security Journal, (2019), Vol.19, No.1, pp.121-126.
Available from: <http://www.riss.kr/link?id=A106112242>
- [5] H. Choi, H. The current status of sports big data analysis researches in Korea, The Korean Journal of Measurement and Evaluation in Physical and Sport Science, (2020), Vol.22, No.2, pp.63-69.
DOI: <http://dx.doi.org/10.21797/ksme.2020.22.2.006>
- [6] J. H. Cho, Utilization and prospect of sport big data, Journal of Measurement and Evaluation in Physical and Sport Science, (2012), Vol.14, No.3, pp.1-11.
DOI: <http://dx.doi.org/10.21797/ksme.2012.14.3.001>
- [7] C/ Hemlata, P. Gulia, Comprehensive study of open-source big data mining tools, International Journal of Artificial Intelligence and Knowledge Discovery, (2016), Vol.6, No.1, pp.15-18.
Available from: https://www.researchgate.net/publication/322628934_Comprehensive_Study_of_Open-Source_Big_Data_Mining_Tools
- [8] P. Ristoski, C. Bizer, H. Paulheim, Mining the web of linked data with RapidMiner, Journal of Web Semantics, (2015),

Vol.35, No.3, pp.142-151.

DOI: <http://dx.doi.org/10.1016/j.websem.2015.06.004>

- [9] Y. Zhou, X. Chen, Simulation of sports big data system based on Markov model and IoT system, *Microprocessors and Microsystems*, (2021), Vol.80, pp.1-5.
DOI: <http://dx.doi.org/10.1016/j.micpro.2020.103525>
- [10] G. Shah, A. Shah, M. Shah, Panacea of challenges in real-world application of big data analytics in healthcare sector, *Journal of Data, Informatics and Management*, (2019), Vol.1, No.3, pp.107-116.
DOI: <http://dx.doi.org/10.1007/s42488-019-00010-1>
- [11] K. Choi, K. G. Jung, Exploring K league 1 athletic performance factors using cluster analysis, *The Korea Journal of Sports Science*, (2022), Vol.31, No.2, pp.699-709.
DOI: <http://dx.doi.org/10.35159/kjss.2022.4.31.2.699>
- [12] M. Hofmann, R. Klinkenberg, *RapidMiner: Data mining use cases and business analytics applications*, CRC Press, (2013)
- [13] Y. S. Kim, C. K. Lee, *Data analysis methodology using RapidMiner(Basic)*, Chaos Press, , (2016)
Available from: <https://product.kyobobook.co.kr/detail/S000001630886>
- [14] G. Y. Park, H. J. Lee, Application of decision tree analysis for supporting tourism policy decision making : Empirical analysis of tourism development policy priorities, *Journal of Tourism Sciences*, (2019), Vol.43, No.5, pp.11-28.
DOI: <http://dx.doi.org/10.17086/JTS.2019.43.5.11.28>
- [15] B. Kim, H. Bae, K. H. Seok, D. H. Cho, K. Choi, *Data Science using RapidMiner: Basic*, Life and Power Press, Gyeonggi-do, (2020)
Available from: <https://product.kyobobook.co.kr/detail/S000001075960>
- [16] V. Kotu, B. Deshpande, *Predictive analytics and data mining: Concepts and practice with RapidMiner*, Morgan Kaufmann-Elsevier, (2014)
Available from: https://www.amazon.com/Predictive-Analytics-Data-Mining-RapidMiner/dp/0128014601/ref=sr_1_1?crd=3JFGH930KBVLW&keywords=predictive+analytics+and+data+mining&qid=1668218859&s=books&prefix=predictive+analytics+and+data+mining%2Cstripbooks-intl-ship%2C254&sr=1-1
- [17] A. Naik, L. Samant, Correlation review of classification algorithm using data mining tool: WEAK, RapidMiner, Tanagra, Orange and Knime, *Procedia Computer Science*, (2016), Vol.85, pp.662-668.
DOI: <http://dx.doi.org/10.1016/j.procs.2016.05.251>