

게임데이터를 이용한 승패예측 및 세분화된 변수 중요도 도출 기법

Predicting win-loss using game data and deriving the importance of
subdivided variables

오민지* · 최은선* · Som Akhamixay Oui** · 조완섭**†

충북대학교 대학원 빅데이터협동과정*, 충북대학교 경영정보학과**

요약

정보통신기술의 발달과 더불어 게임 산업이 성장하면서 유저의 게임데이터는 다양한 플레이 및 옵션에 따라 초 단위로 기록되며 방대한 양의 게임데이터를 빅데이터 기반으로 분석할 수 있게 되었다. 비즈니스와 결합하여 다양한 분야에서 수익창출을 위한 새로운 가치를 발견하는 것에 빅데이터를 활용하고 있지만, 게임 산업에서의 빅데이터 활용은 미흡한 실정이다. 본 연구에서는 리그오브레전드의 게임데이터를 이용하여 라인 별 승패예측모형을 구축한 뒤 세분화 된 라인의 특성을 반영한 변수 중요도를 도출하여 일반 게임 유저가 승률을 올리기 위해 전적검색사이트를 이용하여 사전에 팀 구성원에 대한 정보를 제공받을 수 있도록 한다.

■ 중심어 : 딥러닝, 게임데이터, AOS(Aeon Of Srtife), 빅데이터, XGboost, 앙상블

Abstract

With the development in the IT industry and the growth in the game industry, user's game data is recorded in seconds according to various plays and options, and a vast amount of game data can be analyzed based on Bigdata. Combined with business, Bigdata is used to discover new values for profit creation in various fields, but it is utilized in the game industry in insufficient ways. In this study, considering the characteristics of the subdivided lines, we constructed a win-loss prediction model for each line using the game data of League of Legends, and derived the importance of variables. This study can contribute to planning of strategies for general game users to get information about team members in advance and increase the win rate by using the record search sites.

■ Keyword : Deep Learning, Game Data, AOS (Aeon Of Srtife), Big Data, XGboost, Ensemble

I. 서론

한국 콘텐츠진흥원의 2019 대한민국 게임백서에 따르면, 2018년 국내 게임 시장 규모는 14조 2,902억 원으로 전년 대비 8.7% 증가하였다. 2009년부터 2018년까지의 10년 동안 국내 게임 산업은 꾸준히 성장해왔으며 강하나 등(2019)은 2020년까지 국내 게임 산업이 꾸준한 성장세를 보일 것으로 전망한다[1].

게임 산업의 성장과 더불어 정보통신기술의 발달로 인해, 방대한 양의 게임 데이터가 생성 및 저장되고 있으며 게임 전략 및 환경분석 등 다양한 영역에서 게임 데이터를 활용한 연구가 발전하고 있다. 특히, 게임 데이터를 통한 게임 전략 분석은 프로게임 수준에서 승패에 영향을 미치는 변인들의 분석을 통해 분석가에게 게임의 흐름을 제공하고 프로팀의 전략 수립을 가능하게 해주면서 새로운 가치를 창출한다.

게임 데이터 기반의 전략 분석은 프로게임 뿐만 아니라 일반 게임 사용자 수준에서도 새로운 가치를 가져다 줄 수 있다. 빅데이터를 저장 및 분석할 수 있는 기술이 발전하면서, 일상 속 여가인 게임에서 생성된 데이터를 실시간으로 도출하고 이를 기록 및 저장하는 것이 가능해졌으며 온라인 게임 유저(user)가 이를 열람할 수 있게 되었다. 게임 내의 환경적 요소와 상호작용을 하며 다양한 플레이 및 옵션에 따라 유저의 다양한 게임 데이터가 초 단위로 기록 되므로, 일반 게임 유저의 데이터를 사용하여 빅데이터를 기반으로 하는 정교한 분석이 가능해졌다. 또한 코로나19 확산이 장기화 되면서 많은 사람들이 실내에서 여가 생활을 즐기게 되어 PC게임 일일 이용자 수가 전년 대비 46% 증가하였고[2], e스포츠 시청자의 75%를 차지하는 MZ세대가 게임 내에서 다른 사람들과 경쟁하여 승리하는 것을 중요하게 여기면서 일반 게임 유저 데이터의 분석이 새로운 가치를 지닐 가능성이 생겨났다[3]. 또한 게임 이용 중에 발생하는 데이터의 분석을 통해 새로운 가

치를 게임개발에 환류하여 적용하는 경우, 일반 유저의 사용성 증대 및 사용자 충성도 향상 등의 부가적인 이익을 기대하게 하여, 분석을 통한 상업이익의 효과를 가져다 줄 가능성이 있다[4]. 하지만 게임을 실제로 플레이하는 일반 유저로부터 발생한 데이터를 승패예측에 활용하거나 이를 분석한 사례는 매우 제한적이다.

본 연구에서는 온라인 게임의 장르 중 RTS(real-time strategy)의 특별한 형태인 AOS(aeon of strife) 장르 게임인 리그오브레전드(League of Legends) 유저의 실제 게임 데이터를 수집하여 라인 별 승패예측모형을 구축한 뒤 XGBoost와 SHAP를 사용하여 라인 별로 세분화된 변수 중요도를 도출한다. 또한 기존에 제시된 평가지표를 바탕으로 도출된 라인 별 변수 중요도가 실제 라인의 특성을 잘 반영하는지 확인한다. 승패예측 모형 설계 시 랜덤포레스트, Gradient Boosting Classifier, XGBoost를 이용하여 모형을 설계한 후 각 모형 간의 예측 성능 비교를 통해 최종 예측모형을 결정한다. 기계학습 기법으로 게임 데이터를 분석한 본 연구의 결과가 향후 승패예측에 활용되고 부가적인 이익을 창출하여 게임 데이터의 활용성을 높이는데 기여할 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 제2장에서 관련 연구를 설명하고, 제3장에서 분석에 사용된 데이터를 기술한다. 제4장에서 랜덤포레스트, Gradient Boosting Classifier, XGBoost를 이용한 분석모형을 설계한 뒤 제5장에서 데이터 분석 및 결과를 보여준다. 제6장에서 연구내용을 요약하고 연구의 시사점 및 향후과제를 기술한다.

II. 관련연구

온라인 게임 분야에서 변수중요도를 추출하여 승패예측에 활용하는 연구가 활발히 진행되고 있다. 승패예측모형을 설계하기 위해 기존 스포츠

산업에서의 승패예측과 관련한 연구를 토대로 온라인 게임 분야에서도 통계기법 및 기계학습 기법을 사용한 승패예측 모델이 제안된 바 있다.

2.1 스포츠 승부 예측

스포츠 승부예측모델의 예측성능을 강화하기 위해 통계학적 모델과 기계학습 모델을 결합한 새로운 앙상블 모델을 제안하여 단일모델들의 경기 예측결과와 실제 경기 결과를 병합한 데이터로부터 최종예측모델을 학습하여 경기 승부예측을 수행하였다[5]. 또한 데이터마ining을 활용하여 한국 프로야구 구단의 승패예측에 영향을 미치는 집단별 주요 요인을 도출하여 집단에 맞춤형 승률 상향 전략을 제안한 연구가 진행된 바 있다[6]. 승패 예측을 위해 야구 데이터와 득점/실점과의 상관관계가 높은 세이버 메트릭스 지표를 이용하여 딥러닝 기반의 승패예측모델이 제안되기도 하였다[7].

2.2 게임 데이터 분석

온라인 게임 산업에서도 방대한 양의 게임 데이터를 이용한 연구들이 진행되었다. 용혜련 등(2015)은 데이터마ining 기법을 적용하여 게임 데이터를 분석하였으며, 유저의 게임수준에 영향을 미치는 요인을 도출하였다[4]. 리그오브레전드 게임을 시작하기 전 챔피언 선정에 있어 팀 기반의 전략을 수립하도록 머신러닝 기법을 활용한 승패 예측모델이 제안되기도 하였다[8]. 구지민 등(2017)은 판별분석, 인공신경망, SVM 등 다양한 머신러닝 기법을 사용하여 리그오브레전드 게임 내 포지션의 특성을 반영한 인자점수로 각 라인의 챔피언 특성별 평가지표와 승패예측모형을 제안했다[9]. 강하나 등(2019)은 배틀그라운드 게임을 중심으로 다양한 변수와 좌표값을 포함하는 실제 게임데이터를 사용하여 게임전략 분석 및 유저행동 패턴을 파악하고 게임 내 비정상적인 활동을 탐지하였다[1].

III. 리그오브레전드 게임 데이터 설명

3.1 분석데이터의 수집 방법

본 연구에서는 라이엇게임즈에서 제공하는 API를 사용하여 리그오브레전드 게임 데이터를 수집하고 분석을 실시하였다. Python과 API를 사용하여 데이터를 수집한 뒤 MySQL 서버에 저장하였으며, 전처리와 분석은 R을 사용했다. 분석 데이터는 리그오브레전드 한국서버의 10.16.1 패치 버전까지 플레이 된 5 × 5 개인 랭크 게임으로 소환사의 협곡 맵에서 진행된 게임을 선택하였으며, 실버 랭크에 속한 선수들의 실제 게임 데이터를 사용하였다. 또한 수집된 데이터 중 개인랭크 게임이 아닌 데이터를 삭제하였다. 이러한 기준을 만족하는 데이터로 n = 214개의 게임 데이터가 선택되었다. <표 1>은 사용할 데이터에 명시된 라인 별 역할을 정리한 표이다.

<표 1> 라인 별 역할

라인	역할
탑	탱킹과 이니쉬 및 스플릿
정글	시야 장악, 정글링 및 전체적인 라인 관리
미드	순간적 딜로 상대 암살 및 타 라인 관여
원거리딜러	지속적인 딜 담당
서포터	시야 장악 및 원거리딜러를 보조

탑의 역할 중 탱킹(Tanking)이란 생존력이 강한 챔피언을 선택하여 높은 체력과 방어력을 바탕으로 공격 시 선봉에 서는 것을 의미한다. 탑은 탱킹이란 특징을 바탕으로 이니시에이팅(Initiating)을 진행하기도 하는데, 이니시에이팅이란 상대방에게 싸움을 먼저 걸어 전투를 이끌어내는 전투 개시능력이다. 또한 탑은 강한 생존력과 순간이동 스킬을 바탕으로 스플릿(Spilt push)을 진행하기도 한다. 이는, 나뉘어져서 상대 진영 방향을 공격한다는 것을 의미하며 타 라인에서 다른 라이너들이 대치 시 탑라이너가 다른 라인을 혼자

상대하는 것을 의미한다.

정글의 정글링이란 미니언을 먹고 성장하는 타 라이너들과는 달리 정글 몹으로 성장을 하는데, 정글몹을 사냥 하는 것을 정글링이라 한다.

3.2 변수의 선정

리그오브레전드 게임 내 소환사의 협곡 맵은 5 × 5 게임으로 5명의 플레이어가 한 팀이 되어 상대방의 넥서스를 부수면 승리하는 게임이다. <그림 1>의 좌측 상단은 탑라인으로 탑라이너가 게임을 진행하며 중간은 미드라인으로 미드라이너가 게임을 진행한다. 우측하단은 바텀이라고 부르는 라인으로 원거리딜러와 서포터가 게임을 진행한다. 정글은 정해진 라인이 없이 맵의 모든 곳을 이동하며 정글몬스터와 중립몬스터를 처치하고, 팀원들을 보조한다.



<그림 1> 라인 별 역할

게임이 끝나면 API를 사용하여 유저의 플레이가 기록된 게임 데이터를 가져올 수 있으며, 한번의 게임 데이터에는 총 184개의 변수가 포함되어 있다. 또한 적합한 변수 도출에 참고하기 위해 라이엇에서 제공하는 캐리레이팅을 참고하였다. 캐리 레이팅은 각 라인 별 주요 역할을 반영하여 선수의 실력을 가장 잘 나타낼 수 있는 대표적인

지표이며 <표 2>와 같다.

<표 2> 라이엇이 제공하는 캐리레이팅

라인	지표	설명
탑	D(T)PM	(적 챔피언에게 가한 데미지 + 적 챔피언에게 받은 데미지) ÷ 경기 시간
정글	KA%	(킬 수 + 어시스트 수) ÷ 총 팀 킬 수
미드	DPM	적 챔피언에게 가한 데미지 ÷ 경기 시간
원거리 딜러	DPM	적 챔피언에게 가한 데미지 ÷ 경기 시간
서포터	APG	어시스트 수

<표 3> 선택된 변수 및 설명

변수 이름 변수 유형	설명
lane (factor)	챔피언의 해당 라인 (탑 = 0, 정글 = 1, 미드 = 2, 원거리딜러 = 3, 서포터 = 4)
kills (numeric)	상대를 죽인 횟수
deaths (numeric)	상대에게 죽은 횟수
assists (numeric)	킬 관여 횟수
totalDamageDealt ToChampions (numeric)	상대방에게 가한 총 피해량
totalHeal (numeric)	총 치유량
totalDamageTaken (numeric)	총 받은 피해량
DealtToTurrets (numeric)	포탑에 가한 총 피해량
champLevel (numeric)	챔피언 레벨
visionScore (numeric)	시야점수

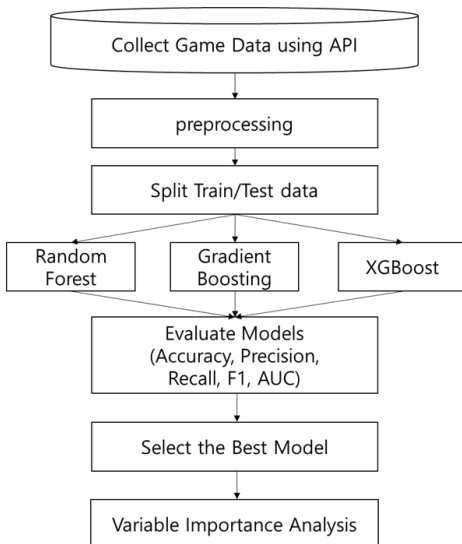
라인별 역할과 기존의 캐리레이팅을 참고하여 9개의 연속형 변수와 라인을 나타내는 1개의 요인변수를 정의하였다. <표 3>은 그 결과 선택된 변수의 유형과 설명이다.

IV. 분석모형 설계

본 장에서는 분석 프로세스를 기술한 뒤, 데이터마이닝 방법 중 승패예측 및 변수중요도 도출에 사용되는 랜덤포레스트, Gradient Boosting Classifier, XGBoost 알고리즘을 설명한다.

4.1 분석 프로세스

본 연구는 <그림 2>와 같이 API를 사용해 수집한 게임 데이터를 전처리 과정에서 라인 별로 분리한 뒤 랜덤포레스트, Gradient Boosting, XGBoost 예측모형을 구축하여 5가지 성능평가 지표를 통해 최적의 승패예측모형을 선별하고 세분화된 변수중요도를 도출한다.



<그림 2> 승패예측모형 및 변수중요도 도출 시스템

4.2 라인 별 승패예측모형

랜덤포레스트는 의사결정나무 기법의 한 종류로, 개별 의사결정나무의 결과를 결합하여 모델을 생성한다. 부스팅(boosting)기법과 다르게 서로 상관성이 없는 의사결정나무들을 통합하기 때문에 학습데이터의 노이즈에 대해 강건한(robust) 장점이 있다[10]. Gradient Boosting은 약한 예측 모델들을 결합해 강한 예측 모델을 만드는 알고리즘으로서, 회귀와 분류 모두에 사용될 수 있다. 여러 개의 개별 의사결정나무를 결합하기 때문에 과적합을 방지할 수 있는 장점이 있다[11]. 또한 개별 의사결정나무의 예측오차가 높다 하더라도 여러 개의 의사결정나무를 결합한 최종 예측 모델은 예측 오차가 낮다[12].

XGBoost는 약한 분류기를 순차적으로 개선해 나감으로써 보다 강력한 분류기를 생성하는 트리 모형에 그래디언트 부스팅(gradient boosting)기법을 적용한 앙상블 알고리즘 중의 하나이다[13]. 또한 트리모형을 병렬로 연산하기 때문에 학습과 분류가 빠르며, 다양한 하이퍼파라미터를 제공하므로 유연성이 좋다는 장점이 있다[14].

4.3 라인 별 변수중요도 도출

의사결정나무기반의 앙상블 모형은 많은 장점이 있음에도 불구하고 블랙박스 모형이기 때문에 설명변수와 반응변수의 설명력을 확보하기 어려운 단점이 있다. 예측성능에 중요한 역할을 하는 변수를 추정하여 모형의 해석력을 높이기 위해 변수중요도(Variable Importance)를 도출한다. 변수중요도는 는 수학적, 통계적 모델을 활용하여 목표변수에 대한 입력변수의 상대적인 중요도를 0 ~ +1까지 나타낸다[15]. 보다 정확한 승패예측을 위해 변수의 중요도가 낮은 변수는 입력변수에서 제거될 수 있으며, 게임 전략 수립 시 변수중요도가 높은 변수를 중심으로 효율적인 전략 수립이 가능하다.

V. 데이터 분석 및 결과

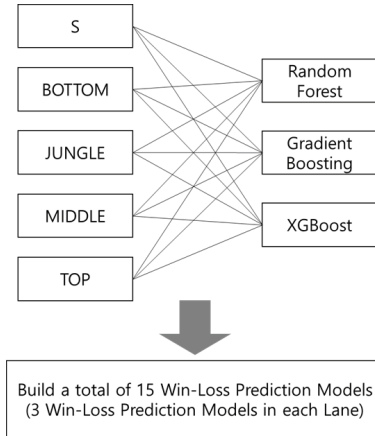
5.1 라인 별 승패예측 결과

5 × 5 개인 랭크 게임으로 진행된 214개의 플레이에서 수집한 2140개의 데이터를 라인 별로 분리하여 총 428개의 행을 가진 5개의 데이터셋으로 나누었다. 라인 별로 분리된 데이터셋은 <그림 3>과 같이 세 가지의 승패예측모델을 구축하는데 사용되었으며, 예측모델마다 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score, AUC를 <표 4>와 같이 비교하였다. 승패예측모델의 성능을 비교한 결과, XGBoost 모형이 정밀도, 재현율, F1 Score, AUC 평가지표에서 성능이 뛰어난 것을 확인하였다.

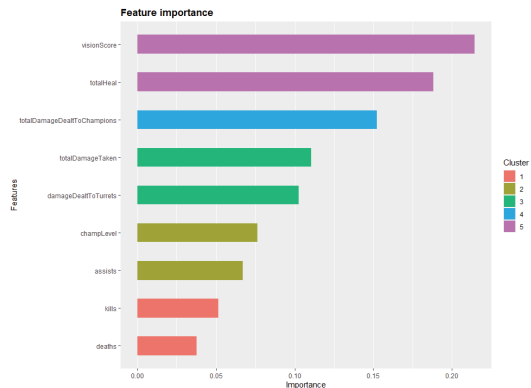
5.2. 라인 별 변수중요도 도출

라인 별 승패예측모형 중 성능평가가 우수한 것으로 나타난 XGBoost를 이용하여 라인 별 변수중요도를 도출하였다. XGBoost로 도출된 변수중요도는 SHAP를 통해 도출된 변수중요도와 비교하여 결과를 검증한다. SHAP는 머신 러닝 모델에서 각 특징의 중요성과 설명을 계산하기 위해 적용하는 일반적인 기술 중 하나이다. XGboost에서는 모델이 학습한 내용을 바탕으로 ‘weight’, ‘gain’, ‘coverage’에 따른 입력 특징들의 상대적 중요도를 계산한다. <그림 4>는 XGboost를 통해 도출된 라인 별 유저에 영향을 미치는 변수중요도를 나타내었다.

<그림 4>의 좌측에 따라 5개의 Cluster가 생성되었으며 이는 각각 라인과 매칭된다. 또한 5개의 Cluster는 앞서 명시한 라인 별 역할의 특징과 유사하게 나타났다.



<그림 3> 라인 별 승패예측모델 구축



<그림 4> 라인 별 변수중요도

<표 4> 라인 별 승패예측모델 성능결과

정확도 (Accuracy)			
lane	Random Forest	Gradient Boosting	XGBoost
서포터	0.822	0.775	0.813
원거리딜러	0.878	0.887	0.887
정글	0.831	0.822	0.803
미드	0.785	0.794	0.803
탑	0.719	0.738	0.757

정밀도 (Precision)			
lane	Random Forest	Gradient Boosting	XGBoost
서포터	0.807	0.764	0.792
원거리딜러	0.846	0.850	0.850
정글	0.809	0.806	0.781
미드	0.727	0.760	0.764
탑	0.744	0.766	0.787
재현율 (Recall)			
lane	Random Forest	Gradient Boosting	XGBoost
서포터	0.823	0.764	0.823
원거리딜러	0.825	0.850	0.850
정글	0.894	0.877	0.877
미드	0.833	0.791	0.812
탑	0.660	0.679	0.698
F1 Score			
lane	Random Forest	Gradient Boosting	XGBoost
서포터	0.815	0.764	0.807
원거리딜러	0.835	0.850	0.850
정글	0.850	0.84	0.826
미드	0.776	0.775	0.787
탑	0.700	0.720	0.740
AUC (Area Under Curve)			
lane	Random Forest	Gradient Boosting	XGBoost
서포터	0.822	0.775	0.813
원거리딜러	0.867	0.880	0.880
정글	0.827	0.818	0.798
미드	0.789	0.794	0.804
탑	0.710	0.730	0.756

서포터의 경우 시야장악과 총 치유량의 변수 중요도가 높게 나타났으며, 미드는 상대챔피언에 대한 딜량의 변수 중요도가 높게 나타났다. 탑은 총 받은 데미지와 포탑에 가한 데미지의 변수중 중요도가 높게 나타났으며, 정글의 경우 챔피언 레벨과 어시스트의 변수중중요도가 높게 나타났다. 원거리딜러의 경우에는 킬과 데스가 높게 나타났는데, 이는 상대방을 많이 죽이고, 자신은 잘 죽지 않는 원거리딜러의 특성이 반영된 것으로 보인다. <그림 4>에 명시된 Cluster 별 해당되는 라인인은 <표 5>와 같다. 또한 <표 6>은 도출된 변수 중요도의 상위 5개 변수에 대한 중요도를 나타내었다. 도출된 결과를 검증하기 위하여 SHAP를 사용하였으며, 그 결과는 <그림 5>와 같다.

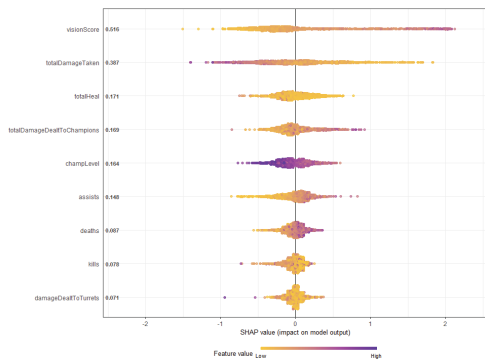
시야점수, 총 받은 데미지, 총 치유량, 상대챔피언에 대한 딜량, 챔피언레벨, 어시스트, 데스, 킬, 포탑에 가한 데미지로 나타났다. SHAP Value에서 시야점수가 가장 높게 나온 이유는 시야장악을 통해 상대방보다 정보전에서 우위를 점하는 게임규칙과 관련이 있다. 다만 Feature Value가 왼쪽으로 치우친 경향이 있는데 이는 전체 라인 중 정글과 서포터 위주로 시야 장악이 진행되기 때문이다. 총 받은 피해량의 SHAP Value가 높은 이유도 계속되는 교전으로 인한 딜 교환 때문이다. <표7>은 SHAP Value 중 상위 5개 요인을 나타내었다.

<표 5> 라인 별 역할

Cluster	해당 라인
1	원거리딜러
2	정글
3	탑
4	미드
5	서포터

〈표 6〉 라인 별 역할

Feature	Gain	Cover	Frequency
vision Score	0.2144009	0.0976063	0.1052316
totalHeal	0.1881747	0.1667395	0.1514698
total Damage DealtTo Champions	0.1524166	0.1546666	0.1415047
total Damage Taken	0.1104143	0.1288034	0.1330343
damage DealtTo Turrets	0.1024768	0.1590831	0.1506726



〈그림 5〉 SHAP를 이용한 결과 검증

〈표 7〉 상위 5개 SHAP Value

Feature	SAHP VALUE
vision Score	0.51553048
total Damage Taken	0.38738725
totalHeal	0.17060076
total Damage DealtTo Champions	0.16939976
champ Level	0.16390669

VI. 결론

본 연구에서는 리그오브레전드의 게임데이터를 이용하여 라인 별 승패예측모형을 구축한 뒤 세분화 된 라인의 특성을 반영한 변수 중요도를 도출하였다. 랜덤포레스트, Gradient Boosting Classifier, XGBoost를 활용한 승패예측모형 중 XGBoost를 사용한 승패예측모형의 성능이 우수한 것으로 드러났다. 최종 선택된 승패예측모형인 XGBoost 모형을 통해 도출된 변수중요도는 각 라인 별 특징을 잘 반영하였다. 또한, XGboost와 SHAP를 비교한 결과 공통적으로 시야점수의 중요성이 높게 나왔다. 이를 통해 시야장악이 정글과 서포터 뿐만 아니라 다른 라인에서도 중요시 되어야 한다는 점을 도출 할 수 있었다. 이와 같은 분석 결과를 통해 일반 게임 사용자가 승률을 올리기 위해 전적검색사이트를 이용하여 사전에 팀 구성원에 대한 정보를 제공받아 전략을 구상하는 것에 기여할 수 있다.

본 연구는 리그오브레전드의 여러 랭크 중 실버랭크에 한정하여 분석하였으며, 다양한 변수를 사용하지 못했다는 점에서 한계를 가진다. 게임 데이터의 특성상 데이터 수집 과정에서 게임 사용자의 플레이 특성이 반영된 데이터가 수집된다. 본 연구는 제한된 수의 게임 사용자로부터 데이터를 수집하였으므로 다양한 변수가 반영된 분석 결과를 도출하는 데 어려움이 있었다. 따라서 다수의 게임 사용자로부터 수집한 게임 데이터로 분석을 수행한다면 데이터의 편향성을 개선할 수 있다.

향후 연구에서는 다양한 변수가 반영될 수 있도록 프로 선수들의 데이터를 수집하거나 다수의 일반 게임 유저로부터 데이터를 수집한 뒤 라인 별로 세분화한 뒤 승패예측모형 및 라인 별 변수 중요도를 개선할 예정이다.

참 고 문 헌

[1] 강하나, 용혜련, 황현석. (2019). 빅데이터를 활용한 게임 전략 및 유저 행동 패턴 분석: 배틀그라운드 게임을 중심으로. 한국게임학회논문지, 19(4), 27-35.

[2] 김한준. “코로나19 장기화 조짐... 게임산업에 미치는 영향은?” ZDNet Korea, 2020.06.25.

[3] 이지효. “게임에 돈 쓰는 MZ세대... 왜 e스포츠에 열광하나” 한국경제, 2020.10.21.

[4] 용혜련, 김도진, 황현석. (2015). Data Mining을 이용한 전략시뮬레이션 게임 데이터 분석. 한국게임학회 논문지, 15(4), 59-68.

[5] 이재현, 이수원. (2020). 앙상블 기법을 통한 잉글리시 프리미어리그 경기결과 예측. 정보처리학회논문지, 9(5), 161-168.

[6] 김원종, 최연식, 유동희. (2018). 데이터 마이닝을 활용한 한국 프로야구 구단의 승패예측과 승률 향상을 위한 전략 도출 연구. 한국스포츠산업경영학회지, 23(3), 88-104.

[7] 노연석, 최재현. (2017). 기계학습을 활용한 프로야구 승부예측에 관한 연구. 한국IT정책경영학회 논문지, 9(1), 335-338.

[8] Cheolgi Kim, Soowon Lee. (2019). Predicting Win-Loss of League of Legends at the Start of the Game using Machine Learning Techniques. ISSAT DSBFI. 28-30.

[9] 구지민, 김재희. (2017). 게임 데이터를 이용한 지표 개발과 승패예측모형 설계. 한국데이터정보과학회지, 28(2), 237-250.

[10] 박준영, 채명수, 정성관. (2016). 실시간 범죄 예측을 위한 랜덤포레스트 알고리즘 기반의 범죄 유형 분류모델 및 모니터링 인터페이스 디자인 요소 제안. 정보과학회 컴퓨터의 실제 논문지, 22(9), 455-460.

[11] 박은충, 김수연, 손승우, 박서영, 이두희. (2019). Support Vector Machine과 Gradient Boosting Machine 기반의 풍력발전량 예측 알고리즘 개발

과 전력거래소 풍력발전량 예측 대회를 통한 알고리즘 성능 검증. 전기학회논문지, 68(12), 1694-1703.

[12] D. Lee, H. Shin, and R. Baldick. (2018). Bivariate Probabilistic Wind Power and Real-Time Price Forecasting and Their Applications to Wind Power Bidding Strategy Development. IEEE Transactions on Power Systems, 33(6), 6087-6097.

[13] Chen, T. and C. Guestrin(2016). XGBoost: A Scalable Tree Boosting System. KDD'16, 785-794.

[14] Jae-Young Oh, Do-Hyeon Ham, Yong-Geon Lee, Gi-bak Kim (2019). Power Generation Prediction using XGBoost with Hyperparameter Optimization, The Korean Institute of Electrical Engineers, 112-113.

[15] 정진화, 채영태. (2017). 건물유형별 에너지소비 예측성능 향상을 위한 변수중요도 및 기계학습모델 평가. 한국건축환경설비학회 논문집, 11(6), 586-598.

저 자 소 개

오 민 지(Min-Ji Oh)



- 2017년: 이화여자대학교 교육공학과 (학사)
- 2020년~현재: 충북대학교 빅데이터협동과정 석사
- 관심분야: 빅데이터, 머신러닝

최 은 선(Eun-Seon Choi)



- 2019년: 충북대학교 경영정보학과 (학사)
- 2019년~현재: 충북대학교 빅데이터협동과정 석사
- 관심분야: 빅데이터, 머신러닝



오 이(Som Akhamixay Oui)

- 2018년: National University of Laos, Computer Science (학사)
- 2018년~현재: 충북대학교 빅데이터협동과정 석사
- 관심분야: 빅데이터, 머신러닝



조 완 섭(Wan-Sup Cho)

- 1996년: KAIST 전산학과 (박사)
- 1997년~현재: 충북대학교 경영정보학과 교수
- 관심분야: 데이터베이스, 빅데이터, 인공지능, 블록체인, 데이터거버넌스