

기계학습을 활용한 게임승패 예측 및 변수중요도 산출을 통한 전략방향 도출

김용우, 김영민
한양대학교 기술경영대학원
{ywkim, yngmnkim}@hanyang.ac.kr

Predicting Game Results using Machine Learning and Deriving Strategic
Direction from Variable Importance

Yongwoo Kim, Young-Min Kim
Graduate School of Technology & Innovation Management, Hanyang University

요 약

본 연구에서는 게임 초반 10분의 데이터를 이용하여 리그오브레전드 게임의 최종승패를 랭크별로 예측하고, 구축된 승패예측 모형으로부터 변수중요도를 추출하여 승리를 위한 초반 게임운영의 방향성을 알아보았다. 그 결과 모든 랭크에서 70% 이상의 정확도로 승패를 예측할 수 있었다. 이는 경기 양상이 대부분 뒤집히지 않고 최종승패로 이어지는 것을 의미하며, 이러한 경향성은 상위 랭크로 갈수록 더욱 강하게 나타났다. 랭크와 무관하게 킬(데스)가 초반 게임에서 최종승패에 가장 큰 영향을 미치는 요소로 나타났으나, 일부 변수는 랭크에 따라 중요도 순위가 변화하였고 이는 유저가 속한 랭크에 따라 승리에 효과적인 초반 전략방향에 차이가 있음을 시사한다.

ABSTRACT

In this study, models for predicting the final result of League of Legends game were constructed for each rank using data from the first 10 minutes of the game. Variable importance was extracted from the prediction models to derive strategic direction in early phase of the game. As a result, it was possible to predict final results with over 70% accuracy in all ranks. It was found that early game advantage tends to lead to the final win and this tendency appeared stronger as it goes to challenger ranks. Kill(death) was found to be the most influential factor for win, however, there were also variables whose importance rank changed according to rank. This indicates there is a difference in the strategic direction in the early stage of the game depending on the rank.

Keywords : Machine Learning(기계학습), Classification(분류), League of Legends(리그오브레전드)

Received: May. 10. 2021 Revised: Jun. 14. 2021
Accepted: Jun. 17. 2021
Corresponding Author: Young-Min Kim (Hanyang University)
E-mail: yngmnkim@hanyang.ac.kr

© The Korea Game Society. All rights reserved. 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.

ISSN: 1598-4540 / eISSN: 2287-8211

1. 서 론

2020년 e-Sports 실태조사에 따르면, 2019년 기준 글로벌 e-Sports 산업 규모는 전년 대비 9.9% 증가하여 950만 달러에 육박한다. 2019년 기준 국내 e-Sports 산업 역시 전년 대비 22.8% 증가하였으며 산업 규모는 약 1,400억 원에 달한다[1]. 선수, 구단, 스폰서(기업), 미디어, 시청자, 일반 게이머 등 각 주체가 상호작용하면서 e-Sports도 다른 스포츠들과 마찬가지로 프로 스포츠의 형태를 갖추어나가고 있다[2]. 2018 자카르타·팔렘방 아시안게임에서는 e-Sports가 시범종목으로 선정되었으며, 오는 2022 항저우 아시안게임에서는 정식종목으로 선정되었다.

야구나 농구 같은 전통적인 스포츠에서는 이미 적극적으로 데이터를 활용하여 구단 운영이나 게임 전략 수립에 활용하고 있다[3]. e-Sports에서도 게임플레이를 통해 생성되는 데이터를 적극적으로 활용하기 시작했다. 제작사는 데이터를 분석하여 게임의 밸런스 패치에 활용하거나, 허가되지 않은 프로그램을 사용하는 이용자들을 단속하는 데에 활용하고 있다[4]. 제작사들은 게임에서 생성된 데이터 중 일부를 자사의 게임발전을 위해 공개하고 있는데, 이러한 데이터들을 이용해 생성된 다양한 지표들을 사용자들에게 전문적으로 제공하는 사이트들도 생겨났다[5,6]. 프로게이머들은 데이터를 활용하여 자신의 플레이를 점검하고 향후 승리를 위한 발판으로 활용하는 것이 이제는 일상이 되었다.

게임데이터의 활용이 게임제작사나 프로게이머들에게만 국한된 것은 아니다. 일반 유저들 또한 인터넷 사이트를 이용하여 자신이 플레이한 게임의 데이터를 확인하고 보완할 점을 모색하거나, 최신 전략을 습득하는 것이 매우 일상적인 일이 되었다. 게임 실력을 향상시키고자 하는 일반인을 대상으로 하는 사설학원 및 과외도 성행하고 있다[7]. 프로 게이머는 물론 일반 유저들도 게임의 실력향상에 많은 관심이 있다는 방증이다. 그러나 지금까지의 연구들은 대부분 프로게이머 혹은 그에 준하는 실

력을 갖춘 제한된 유저층을 연구대상으로 하였다.

본 연구에서는 리그오브레전드(League of Legends)의 랭크게임 데이터를 이용하여 랭크별 승패예측을 위한 모형을 구축한다. 초심자부터 프로 수준까지 모든 실력 대의 유저들을 대상으로 랭크별 모형을 구축한다. 게임 초반 양상이 승패로 직결되는지 확인하기 위해서 최초 10분까지의 in-game 데이터만 활용하여 최종승패를 예측하는 모형을 기계학습을 이용하여 구축한다. 또한, 구축된 승패예측 모형에서 변수중요도를 산출하여 승패에 영향을 주는 경기 초반 주요변수가 무엇인지 알아본다.

제2장에서는 관련된 선행연구를 살펴본다. 제3장에서는 분석에 사용된 데이터를 설명하고 승패예측을 위한 모형을 설계한다. 제4장에서는 분석결과를 제시하고, 제5장에서 연구내용의 결과와 시사점을 정리한다.

2. 관련 연구

승패예측에 관한 연구는 야구, 농구 등 기존 스포츠 영역에서 이미 활발히 진행됐고, 리그오브레전드의 승패예측 연구 역시 최근 들어 활발하게 진행되고 있다.

게임 시작 전, 챔피언 선택 단계에서의 데이터만으로 게임의 최종승패를 예측한 연구들이 있다. Neumann(2015)은 의사결정나무를 이용하여 [8], 김철기 외(2019)는 LSTM을 사용하여 in-game 데이터는 배제하고 경기 시작 전 챔피언 선택 단계의 데이터만을 활용하여 승패예측 모형을 구축하였다[9].

게임 시작 이후의 in-game 데이터를 이용해서 승패를 예측한 연구들도 있다. 김동욱 외(2017)는 한국 서버 최상위 0.01% 플레이어들의 랭크게임 데이터를 이용해서 승패예측 모형을 구축하였다[10]. 구지민 외(2017)는 젤린저 및 다이아몬드 랭크에 속한 상위 플레이어들의 랭크게임 데이터로 다양한 기계학습 기법들을 활용하여 승패예측모형

을 구축하였고 게임 내 포지션 특성을 반영한 새로운 지표를 제안하였다[11]. 오민지 외(2020)는 실버랭크에 속한 유저들의 랭크게임 데이터로 기계학습 기법을 활용하여 포지션별 승패예측모형을 구축하였다[12].

Ani et al.(2019)은 프로게이머들의 공식경기 데이터를 가지고 챔피언 선택 단계에서의 데이터와 in-game 데이터를 모두 이용하여 승패를 예측하였다[13].

유저의 행동변수를 통해서 해당 유저가 속한 랭크를 예측한 연구도 있다. Thompson JJ et al.(2013)은 스타크래프트2의 데이터를 이용하여 게임 내 유저들의 행동이 기록된 변수들로 유저들이 속한 랭크를 예측하였다. 랭크별로 유저들의 행동변수들이 다르게 나타나는 것을 확인하였고, 랭크에 따라서 게임플레이 패턴이 다를 수 있음을 밝혔다[14].

지금까지 랭크게임의 in-game 데이터를 이용한 승패예측 연구들은 대부분 제한적인 유저층을 대상으로 게임 시작 시부터 종료 시까지의 데이터를 모두 활용하여 승패를 예측하였다. 또한, 승패예측 자체에 초점을 맞추거나 역할군별 특성을 반영한 새로운 지표를 발굴하는 것에 집중하였다. 그러나, 본 연구는 초심자부터 프로게이머 수준까지 모든 랭크에 속한 유저들을 분석대상으로 한다.

통상적으로 리그오브레전드 한 경기당 약 30분의 시간이 소요되지만, 경기 시작부터 종료까지의 모든 데이터를 사용하지 않고 시작 후 10분까지의 데이터만을 활용하여 최종승패를 예측한다. 이상광 외(2019)는 챌린저 유저들을 대상으로 한 연구에서 게임 시작 후 10분까지의 데이터로 76.2%의 승패예측 정확도를 기록한 바 있으며, 30분까지의 데이터를 이용한 승패예측 모형에서는 78.7%의 정확도를 보였다[15]. 이는 최상위권 유저들의 게임에서 초반 10분의 영향력이 매우 크다는 것을 의미하며, 게임 초반 전략방향 탐구의 필요성을 시사한다.

본 연구에서는 게임 초반 10분 데이터만을 이용

한 승패예측 모형을 통해 모든 랭크에서 초반 중요도가 어느 정도인지 랭크별로 확인하고 초반운영의 중요성이 다른 랭크에서도 나타나는지 확인한다. 또한, 승패예측 모형으로부터 변수중요도를 추출하여 초반에 어떤 전략 방향을 취하는 것이 승리에 도움이 되는지 그리고 그것이 랭크별로 달라지는지 확인한다. 이를 통해 게임성에 대한 실무적인 함의를 제공하고, 게임을 플레이하는 유저들에게 게임운영에 필요한 실질적인 정보를 제공하고자 한다.

3. 데이터 및 모형설계

3.1 데이터 수집

본 연구에서는 세계에서 가장 인기 있는 게임 중 하나인 리그 오브 레전드(League of Legends)의 in-game 데이터를 이용한다. 리그오브레전드는 블루 팀과 레드팀에 각각 배정된 5명의 플레이어가 150여 개의 챔피언 중 하나를 선택하여 5대 5로 플레이한다. 이때, 각 팀의 플레이어는 Top, Mid, Bottom, Jungle, Support 역할군을 분담하여 수행하며, 상대 팀 진영 중심부에 있는 넥서스라는 건물을 먼저 파괴하는 팀이 승리하게 된다. 게임의 유형으로 일반 게임과 랭크 게임이 있다. 일반 게임은 랭크에 영향을 미치지 않는 게임이고, 랭크 게임은 게임의 경기 결과에 따라 반복적으로 점수를 얻거나 잃으면서 사용자는 9단계(Iron, Bronze, Silver, Gold, Platinum, Diamond, Master, Grand Master, Challenger)의 리그로 승급 혹은 강등된다. 이러한 과정을 통해서 본인의 실력과 비슷한 사용자들과 실력을 겨루게 된다.

데이터를 수집하기 위해 Riot Games에서 제공하는 API[16]를 이용하였다. 2021시즌 한국 서버 소환사의 협곡 맵에서 치러진 솔로랭크 게임을 대상으로 최하위 Iron 랭크부터 최상위 Challenger 랭크까지 각 8,800개의 게임을 추출하였다. 리그가 4단계(IV, III, II, I)로 분할되어있는 Iron부터

Diamond까지는 각 분할 별로 2,200개의 게임이 균등하게 추출되었다. 총화추출된 총 79,200개의 인스턴스가 분석에 활용되었으며, 분석에 활용된 데이터는 게임버전 11.2 및 11.3에서 플레이된 게임으로부터 생성되었다.

3.2 변수 선정

게임 종료 후 생성되는 데이터에는 총 10명의 플레이어의 행동이 timestamp에 따라서 기록되어 있다. Timestamp를 제한하여 게임 시작 후 10분까지의 in-game 데이터와 최종승패만 추출하였다. 추출된 데이터 중에서 아이템 구매, 스킬 순서, 특성 등 유저가 선택하는 캐릭터에 따라 달라지는 변수들을 제외하고 승패예측 모형에 투입할 수 있는 변수들만 선택하였다. 이때, 골드획득량은 변수로 사용될 수 있음에도 불구하고 제외되었는데, 골드획득량이 다른 모든 변수들의 결과로 나타나는 변수이기 때문이다. 상관관계가 매우 강하게 나타나는 변수들을 동시에 사용하는 경우 산출된 변수중요도를 신뢰하기 어려워진다[17].

변수들은 플레이어 단위로 저장되어 있는데, 이를 팀 단위 데이터로 통합하여 요약하는 과정을 거쳤다. 팀 단위 데이터는 플레이어 단위로 저장된 데이터를 팀 단위로 합하여 팀별 차이로 변환한 것인데 각 팀 플레이어의 골드, 킬, 데스, 어시스트, 오브젝트 획득량의 합을 팀별로 합산한 후 그 차이 값을 구한 것이다. 정의된 변수의 이름과 설명은 [Table 1]과 같으며 모두 실수형 변수이다. 실수형 변수들은 모두 0-1 정규화를 거쳤는데 data leakage¹⁾를 예방하기 위해 모델 생성 직전에 수행되었다. 전처리 및 승패예측 모형 구축을 위해 Python이 사용되었다.

모형을 생성하기 전 별도의 변수선택 작업은 거치지 않았다. 승패예측에 사용할 모델들이 모델 내에 변수선택 방법을 내재하고 있고(embedded method), 내재된 방법에 따라 중요변수를 선택하여 모델이 만들어지기 때문이다[21].

[Table 1] List of Variables

No	Variable	Description
1	Diff_LV	챔피언들의 LV합 차이
2	Diff_CS	일반 미니언 CS (Creep Score) 차이
3	Diff_jglCS	정글 미니언 CS (Creep Score) 차이
4	Diff-K	Kill 차이 (Death 차이)
5	Diff_A	Assist 차이
6	Diff_WARDplaced	설치한 와드 수 차이
7	Diff_WARDkill	제거한 와드 수 차이
8	Diff_Inhibitor	파괴한 억제기 수 차이
9	Diff_TOWERkill	파괴한 타워 수 차이
10	Diff_FirstDRAGON	첫 Dragon 획득 여부 차이
11	Diff_FirstHERALD	첫 Herald 획득 여부 차이
12	Diff_Firsttower	첫 타워 파괴 여부 차이
13	Diff_FirstBLOOD	첫 킬 여부 차이
14	FirstDragon_AIR_DRAGON	첫 획득 드래곤 종류 (바람의 드래곤)
15	FirstDragon_EARTH_DRAGON	첫 획득 드래곤 종류 (대지의드래곤)
16	FirstDragon_FIRE_DRAGON	첫 획득 드래곤 종류 (화염의 드래곤)
17	FirstDragon_WATER_DRAGON	첫 획득 드래곤 종류 (바다의 드래곤)

3.3 모형 설계

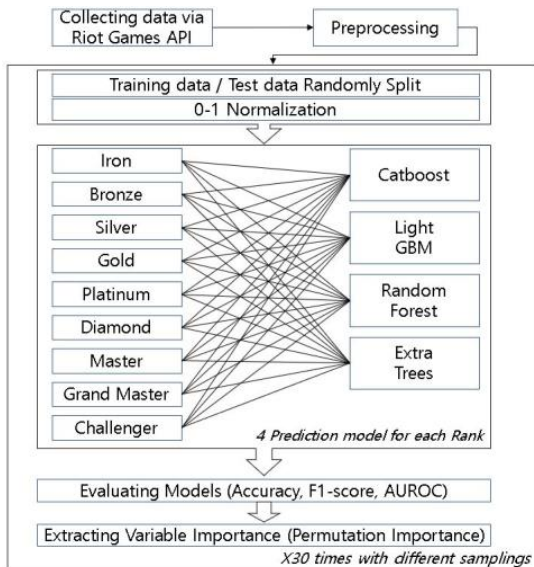
본 연구에서는 앞서 정의된 변수들을 가지고 게임의 최종결과인 승패를 예측한다. 종속변수인 승리 혹은 패배는 오직 두 가지의 값을 가지는 명목 변수이므로 이진 분류(Binary Classification)문제로 볼 수 있다. 이 분류문제를 기계학습을 통해 예측한다. 게임으로부터 생성된 로그 데이터의 특성을 감안할 때 기계학습을 사용하는 것이 바람직할 수 있다. 기계학습을 통한 모델링은 복잡한 가정이 요구되지 않아 분석이 용이하며, 통제되지 않은 환경에서 얻은 데이터나 데이터에 많은 비선형 상호작용이 있는 경우 통계적 모델보다 더욱 효과적이기 때문이다[18]. 일반적으로 성능이 뛰어나다고 알려진 부스팅 방식 기법 Light GBM(LGBM), C

1) Data leakage는 학습 데이터 세트 외부의 정보가 모델생성에 사용되는 것을 의미한다.

atboost(CAT)와 배깅 방식의 기법 RandomForest(RF), ExtraTrees(ET) 총 4개의 모델이 사용되었다[19,20].

수집된 데이터를 각각의 랭크별로 나눈 뒤, 랭크별로 각각의 모델을 구축하였다. 변수중요도 산출을 위해서 구축된 모델로부터 permutation importance를 산출하고 그 순위를 매겼다. 흔히 사용되는 impurity importance (gini index)의 경우 많은 범주를 가지는 범주형 변수나 연속형 변수의 중요도를 높게 평가하는 경향이 있지만, permutation importance는 이러한 종류의 편향으로 인해 영향을 받지 않아서 일반적으로 선호되기 때문이다[22].

모델의 성능 및 변수중요도는 데이터 및 모델 특성에 의존적이므로, 새로운 학습 및 테스트 데이터 샘플링, 승패예측 모형 구축, 변수중요도 추출 과정을 기법별로 30회 반복하여 일관성을 확보하였다. 데이터 수집부터 변수중요도 추출까지의 전 과정이 [Fig. 1]과 같이 나타나 있다.



[Fig. 1] Process of predicting game results and extracting variable importance by rank

4. 분석결과

4.1 랭크별 승패예측

게임 시작 후 10분까지의 in-game 데이터를 이용하여 랭크별로 생성된 각 모델로부터 정확도(Accuracy), AUROC(Area Under the Receiver Operating Characteristics), F1-Score(F-Score)를 30회 산출하여 평균한 결과가 [Table 2]에 나타나 있다. 3가지의 모델 평가지표 (F1-Score, AUROC, Accuracy) 간 큰 차이가 없었는데, 블루 팀 혹은 레드 팀이 승리하는 비율이 약 50%로 예측대상의 클래스가 균형 잡힌 상태이기 때문이다.

지표들의 평균값에서 모델별로 큰 차이가 발견되는 않았다. 다른 모델들보다 미세하게 성능이 우수했던 RandomForest를 기준으로 살펴보면, 랭크에 관계없이 초반 10분까지의 상황으로도 게임의 최종승패를 70% 이상의 정확도로 예측할 수 있었다. Bronze~Gold 구간에서의 정확도는 70.67%~71.76%로 거의 차이가 없었다. Platinum~Diamond 구간에서는 71.95%~72.07%, Master 76.15%, Grand Master 84.59%, Challenger의 경우 91.40%로 승패예측 정확도가 상위 랭크로 갈수록 높아졌다. 특히 Grand Master, Challenger 랭크의 경우 10분까지의 상황이 최종승패로 직결되는 경향이 매우 큰 것으로 나타났다. 이는 최상위권 랭크의 게임은 후반부에서의 역전이 빈번하게 일어나지 않는다는 것을 시사한다. 최상위권 유저들의 경우 유리함을 지속적으로 활용하여 최종 승리까지 연결 짓는 방법을 잘 알고 있는 것으로 보인다. 반면, Iron 랭크의 경우 승패예측 정확도는 74.14%로 Bronze~Gold 구간보다 정확도가 높았는데, 최하위권 유저들은 불리한 상황에서 게임을 풀어나가는 것에 익숙하지 않기 때문으로 보인다.

Silva 등(2018)은 프로게이머들의 공식경기 데이터를 가지고 Simple RNN(Recurrent Neural Network)을 이용하여 승패를 예측하였다. 0~5분의 데이터를 이용해 승패를 예측했을 때 63.91%의

[Table 2] Model Performance

Rank	Model	Accuracy	F1-Score	AUROC
		Mean	Mean	Mean
Iron	CAT	73.05%	72.15%	73.02%
	ET	73.13%	72.04%	73.08%
	LGBM	72.51%	71.50%	72.47%
	RF	74.14%	72.95%	74.08%
Bronze	CAT	70.86%	71.55%	70.84%
	ET	70.98%	71.58%	70.96%
	LGBM	70.15%	70.77%	70.13%
	RF	71.76%	72.41%	71.74%
Silver	CAT	70.22%	70.93%	70.20%
	ET	70.11%	70.64%	70.10%
	LGBM	69.96%	70.65%	69.94%
	RF	70.91%	71.42%	70.90%
Gold	CAT	70.05%	70.79%	70.02%
	ET	69.70%	70.25%	69.69%
	LGBM	68.88%	69.66%	68.85%
	RF	70.67%	71.20%	70.66%
Platinum	CAT	71.47%	72.07%	71.46%
	ET	71.34%	71.66%	71.35%
	LGBM	70.58%	71.10%	70.57%
	RF	71.95%	72.27%	71.95%
Diamond	CAT	71.17%	70.49%	71.15%
	ET	71.57%	70.69%	71.54%
	LGBM	70.59%	69.91%	70.57%
	RF	72.07%	71.15%	72.03%
Master	CAT	75.37%	73.73%	75.25%
	ET	75.47%	73.56%	75.30%
	LGBM	74.53%	72.91%	74.42%
	RF	76.15%	74.35%	75.99%
Grand Master	CAT	83.84%	83.31%	83.83%
	ET	84.39%	83.80%	84.36%
	LGBM	82.81%	82.20%	82.79%
	RF	84.59%	84.00%	84.55%
Challenger	CAT	90.86%	90.62%	90.85%
	ET	91.22%	90.96%	91.20%
	LGBM	90.51%	90.26%	90.50%
	RF	91.40%	91.17%	91.39%

예측 정확도를, 5~10분 데이터로는 68.69%의 정확도를, 10~15분 데이터로는 75.23%의 예측 정확도를 기록하였다[23]. 5분 단위의 시간 간격을 기준으로 데이터를 나누고 사용했다는 점에서 차이는 있지만, 프로게이머들이 Challenger 혹은 Grand Master에서 랭크게임을 플레이한다는 점을 고려하면 리그 경기의 역전 가능성이 랭크게임보다 훨씬 더 크다고 할 수 있다. 이러한 차이는 비슷한 실력을 가진 사람들과 무작위로 매칭되어 플레이하

는 랭크게임이 동일한 팀원들과 많은 연습을 통해 조직적인 플레이가 가능한 공식경기과 전혀 다른 특성을 가지기 때문으로 보인다.

4.2 랭크별 중요변수

랭크별로 생성된 각 모델로부터 변수중요도를 30회 산출하였다. 30회 산출된 변수중요도 순위를 모델별로 합산한 모델별 변수중요도 순위는 [Table 4]에 정리되어 있다. 모델별 순위를 평균하여 랭크별로 가장 중요한 변수가 무엇인지 비교할 수 있도록 요약하였으며 그 결과는 [Table 3]와 같다.

[Table 3] TOP5 important variables

Rank	Variable	LG BM	CA T	RF	ET	ME AN
Iron	diff_K	1	1	1	1	1
	diff_CS	2	2	3	2	2.25
	diff_LV	4	3	2	4	3.25
	diff_jglCS	3	4	4	3	3.5
	diff_A	5	5	6	6	5.5
Bronze	diff_K	1	1	1	1	1
	diff_CS	2	2	3	2	2.25
	diff_LV	4	3	2	3	3
	diff_jglCS	3	4	4	4	3.75
	diff_A	5	5	5	5	5
Silver	diff_K	1	1	1	1	1
	diff_CS	2	2	3	2	2.25
	diff_LV	4	3	2	4	3.25
	diff_jglCS	3	4	4	3	3.5
	diff_A	5	5	5	5	5
Gold	diff_K	1	1	1	1	1
	diff_CS	2	3	3	2.5	2.625
	diff_jglCS	3	2	4	2.5	2.875
	diff_LV	6	4	2	5	4.25
	diff_A	4	5	5	4	4.5
Platinum	diff_K	1	1	1	1	1
	diff_LV	2	2	2	2	2
	diff_jglCS	3	4	4	3	3.5
	diff_CS	4	3	5	5	4.25
	diff_A	5	5	6	4	5
Diamond	diff_K	1	1	1	1	1
	diff_LV	2	2	2	2	2
	diff_jglCS	4	3	6	3	4
	diff_A	5	4	3	4	4
	diff_CS	3	5	4	5	4.25

Master	diff_K	1	1	1	1	1
	diff_LV	5	2	2	4	3.25
	diff_jglCS	3	3	6	2	3.5
	diff_CS	2	4	7	3	4
	diff_A	4	5	5	5	4.75
Grand Master	diff_K	1	1	1	1	1
	diff_LV	6	2	2	5	3.75
	diff_jglCS	3	3	6	3	3.75
	diff_CS	2	5	7	2	4
	diff_A	4	4	5	4	4.25
Challenger	diff_K	1	1	1	1	1
	diff_jglCS	3	2	6	2	3.25
	diff_CS	2	4	7	3	4
	diff_A	5	5	4	5	4.75
	diff_LV	7	3	3	7	5

랭크와 무관하게 모든 모델에서 승패에 가장 영향을 가장 크게 미치는 변수는 킬(데스)였다. 이외 일반 미니언 CS, 정글 미니언 CS, 레벨이 랭크와 무관하게 공통으로 중요한 변수로 나타났다. 랭크별로 도출된 상위 5개의 중요변수에서 큰 차이가 없었는데, 이는 낮은 랭크에서 플레이하는 유저나 높은 랭크에서 플레이하는 유저나 초반 게임운영의 전체적인 방향성은 비슷하다는 것을 의미한다.

다만, 일반미니언 CS의 경우 Iron~Gold 구간에서는 2순위였으나, Platinum 랭크부터는 그 중요도가 감소하였다. 반대로 정글미니언 CS의 경우 Platinum 및 상위 랭크에서 중요도가 높게 나타났다. 이는 Platinum 및 상위 랭크에서는 상대방 진영에 소환되는 정글 미니언을 획득해 상대팀의 정글 미니언 획득을 방해하는 방식으로 격차를 벌려나가는 것이 중요해진다는 것을 의미한다.

5. 논의 및 한계점

e-Sports가 프로스포츠의 형태를 갖춰나가면서 여타 프로스포츠처럼 데이터를 이용한 다양한 승패 예측 연구가 진행되고 있다. 본 연구는 리그오브레전드 게임의 최종승패를 예측하고 최종승패에 영향을 주는 요인을 도출하여 게임을 플레이하는 유저들에게 실질적인 정보를 제공하고 게임성에 대한 실무적인 함의를 제시하고자 하는 연구이다. 리그오

브레전드 게임 초반 10분간의 데이터를 이용하여 랭크별 승패예측모형을 구축한 뒤, 변수중요도를 산출하여 초반에 취해야 할 전략 방향에 관해 기술하였다.

LightBGM, Catboost, RandomForest, Extra trees를 활용한 승패예측에서 랭크와 관계없이 최종승패가 초반에 대부분 결정지어진다는 사실이 발견되었다. 이 결과는 일반적인 통념과는 다르게 낮은 랭크에서 게임을 즐기는 일반 유저들 역시 승리를 위해서는 초반에 더욱 집중력을 발휘할 필요가 있음을 의미한다. 또한, 이러한 결과는 게임의 운영 주체에도 시사하는 바가 있다. 사람들이 스포츠에 열광하는 이유 중 하나는 예측 불가능성인데[24], 이처럼 게임의 승패가 너무 이른 시점에 결정될 경우 자칫 게임의 재미가 반감될 수 있으므로 주의할 필요가 있다. 중·후반부에 획득할 수 있는 오브젝트의 등장시점 및 영향력을 조절하거나, 초반부의 교전에서 얻을 수 있는 이득을 감소시키는 방향의 패치는 이러한 경향성을 둔화시킬 수 있을 것이다. 반대로, 게임의 초반 중요성을 현재 상태로 유지시키면서 행복 시스템을 개편하는 방안도 생각해 볼 수 있다. 판세가 다소 기울어진 게임에서 경우 지고 있는 팀의 유저들은 탈주 페널티를 받지 않기 위해서 의무적으로 플레이하게 되는데, 이때 지고 있는 팀의 유저들은 부정적인 게임경험을 할 가능성이 있다. 다소 일방적인 전개로 게임 내에서 흥미를 잃어버리는 것은 물론, 욕설이나 트롤링에도 노출되기 쉽다[25]. 또한, 짧은 플레이타임이 리그오브레전드 게임을 선호하는 이유중에 하나임을 고려하면[26], 행복조건의 완화는 유저의 부정적인 경험을 감소시킬 가능성이 있다.

승패예측 모형으로부터 변수중요도를 추출한 결과 랭크와 무관하게 비슷한 변수들이 상위권에 위치하였다. 그러나, Platinum 랭크 이상에서는 정글 미니언 CS의 중요도가 일반미니언 CS의 중요도와 같거나 높았지만 Gold 랭크 이하에서는 그 반대였다. 이러한 결과는 하위 랭크에서의 효과적인 승리 전략이 상위 랭크에서는 그만큼 효과적이지 않을

[Table 4] Variable Importance

Rank	Model	diff_K	diff_LV	diff_CS	diff_jeCS	diff_A	diff_FirstDRAGON	diff_WARDKill	diff_FirstHERALD	diff_FirstBLOOD	diff_WARDplaced	FirstDragon_WATER_DRAGON	FirstDragon_FIRE_DRAGON	diff_FirstTOWER	FirstDragon_EARTH_DRAGON	FirstDragon_AIR_DRAGON	diff_TOWERKill	diff_Inhibitor
Iron	LGBM	1	4	3	2	5	6	17	14	8	12	7	9	13	10	15.5	15.5	11
	CAT	1	4	3	2	6	5	10	14.5	8	7	9	16	11	13	17	14.5	12
	RF	1	3	4	2	5	6	7	17	8	11	9	16	13	12	15	10	14
	ET	1	2	4	3	6	5	14	12	8	7	13	16	9	11	17	15	10
Bronze	LGBM	1	4	3	2	5	6	7	10	11	8	9	12	16	14	17	13	15
	CAT	1	3	4	2	5	6	9	10	17	7	8	16	13	14	11	12	15
	RF	1	3	4	2	5	6	8	9	16	10	7	11	15	14	13	17	12
	ET	1	2	4	3	5	8	9	7	14	6	12	15	10	13	17	16	11
Silver	LGBM	1	4	3	2	5	6	7	15	8	17	16	11	14	12.5	10	9	12.5
	CAT	1	4	3	2	5	7	6	16	12	13	11	9	14	17	10	8	15
	RF	1	3	4	2	5	7	6	17	13	11	14	16	9	10	15	8	12
	ET	1	2	4	3	5	11	6	8	9	7	15	16	12	14	17	13	10
Gold	LGBM	1	6	3	2	4	5	17	7	8	11	16	12	10	14	9	15	13
	CAT	1	5	2.5	2.5	4	6	9	7	8	11	14	10	13	16	12	17	15
	RF	1	4	2	3	5	6	14	7	8	15	13	9	11	10	17	16	12
	ET	1	2	4	3	5	7	10	6	8	9	14	15	12	11	16	17	13
Platinum	LGBM	1	2	3	4	5	6	10	7	8	11	12	9	14	16	13	17	15
	CAT	1	2	3	5	4	7	8	6	12	11	9	10	13	14	17	15	16
	RF	1	2	4	3	5	7	9	6	17	15	8	10	11	13	16	14	12
	ET	1	2	4	5	6	7	8	3	14	9	11	15	13	10	16	17	12
Diamond	LGBM	1	2	4	3	5	6	8	12	7	11	10	9	14	17	16	13	15
	CAT	1	2	3	5	4	6	7	14	10	13	11	8	16	12	17	9	15
	RF	1	2	3	5	4	6	8	17	15	10	14	7	11	13	9	16	12
	ET	1	2	6	4	3	5	7	8	14	9	17	11	13	12	16	15	10
Master	LGBM	1	5	3	2	4	8	7	9	6	10	11	12	15	16.5	13	14	16.5
	CAT	1	4	2	3	5	8	6	9	7	10	11	12	14	16	13	17	15
	RF	1	2	3	4	5	6	7	9	8	10	14	11	13	15	12	17	16
	ET	1	2	6	7	5	3	9	4	11	8	15	16	12	10	13	17	14
Grand Master	LGBM	1	6	3	2	4	8	7	9	5	10	11	12	15	16	13	14	17
	CAT	1	5	3	2	4	9	7	8	6	10	12	11	15	14	16	17	13
	RF	1	2	3	5	4	6	8	9	7	10	12	13	15	16	14	11	17
	ET	1	2	6	7	5	4	9	3	11	8	10	14	16	17	13	12	15
Challenger	LGBM	1	7	3	2	5	8	6	9	4	10	14	12	17	15.5	11	13	15.5
	CAT	1	7	2	3	5	9	6	8	4	10	12	16	13	15	11	17	14
	RF	1	3	2	4	5	9	8	7	6	10	13	11	15	17	12	14	16
	ET	1	3	6	7	4	5	9	2	12	8	10	11	16	17	13	14	15

수 있음을 시사한다. 게임 운영 주체는 이러한 경향을 반영하여 전략이 급격하게 변화하는 랭크에 최초로 승급한 유저들에게 게임 팁을 제공하는 것을 고려할 수 있다.

프로게이머들의 리그 경기의 승패예측 정확도와 최상위권 유저들의 랭크게임 승패예측 정확도를 비교했을 때 차이가 있었다. 이는 랭크게임과 리그 경기의 특성 차이를 고려해야 한다는 점을 보여준다. 프로게이머와 비슷한 실력을 갖춘 Challenger 및 Grand Master에 속한 유저들이 플레이한 랭크 게임 데이터를 바탕으로 얻은 분석결과 일치라도, 그 결과를 e-Sports 경기에 적용할 때는 주의할 필요가 있다는 점을 시사한다.

본 연구는 리그오브레전드의 2021시즌 데이터를 바탕으로 이루어졌다. 따라서, 향후에 지속적인 패치가 이루어지면서 챔피언이나 아이템이 현재와 달라질 수 있으며, 그에 따라 분석결과도 달라질 수 있음을 고려해야 한다. 또한, 본 연구의 데이터는 Riot Games API를 이용하여 수집되었다. API를 통해 제공되는 데이터 제공량 및 제공되는 변수에 제한이 있어 대용량 데이터를 이용한 분석이 수행되지 못하였고 게임 내 세부적인 요소들이 고려되지 못했다. 향후 더 많은 변수를 확보하여 분석을 수행한다면 더욱 상세한 세부전략을 도출할 수 있을 것으로 기대된다. 추후 연구에서는 역전승이 일어난 경기들의 패턴을 분석하여, 지고 있는 상황에서 역할균별로 어떤 전략을 취하는 것이 승리에 도움이 되는지 밝히는 더욱 실용적인 연구로 발전시키고자 한다.

ACKNOWLEDGMENTS

This research was supported by the Ministry of Food and Drug Safety of the Republic of Korea (21163수입안516).

REFERENCES

- [1] Korea Creative Content Agency, “The 2020 Survey on the Korean e-Sports Industry”, Korea Creative Content Agency, 2020.
- [2] Sae-Sook Oh and Dae-Hoon Kim, “Analysis of the Academic Research Trend of e-sports”, *Journal of Wellness*, vol. 7, No. 2, pp.113-121, 2012.
- [3] Jung Hwan Cho, “Utilization and Prospect of Sport Big Data”, *Korean Society For Measurement And Evaluation In Physical Education And Sports Science*, vol. 14, No. 3, pp.01-11, 2012.
- [4] Yong Goo Kang and Huy Kang Kim, “Game Bot Detection Based on Action Time Interval”, *Journal of The Korea Institute of Information Security and Cryptology*, vol. 28, No. 5, pp.1153-1160, 2018.
- [5] GGtics, “Analytics/AI to help players win - YOUR.GG”, accessed Apr 13, 2021, <https://your.gg/?lang=en>.
- [6] OP.GG, “LoL Stats, Record Replay, Database, Guide - OP.GG”, accessed Apr 13, 2021, <http://www.op.gg>.
- [7] 일요서울, “사교육 시장에 등장한 ‘게임학원’, 그 배경은?”, accessed Apr 13, 2021, <http://www.ilyoseoul.co.kr/news/articleView.html?idxno=237470>.
- [8] Alexander Neumann, “Developing a Model to Predict Match Outcomes in League of Legends,” Barrett, The Honors College Thesis, Arizona State University, 2015.
- [9] Cheolgi Kim and Soowon Lee, “Predicting Win-Loss of League of Legends Using Bidirectional LSTM Embedding”, *Korea Information Processing Society*, vol. 9, No. 2, pp.61-68, 2020.
- [10] Dong-Wook Kim, Jeawon Park and Jaehyun Choi, “A study for the prediction of winning of e-Sports using machine Learning”, *Journal of The Korea Society of Information Technology Policy & Management*, vol. 9, No. 1, pp.319-325, 2017.
- [11] Jimin Ku and Jaehee Kim, “Development of game indicators and winning forecasting models with game data”, *Journal of the Korean Data and Information Science Society*, vol. 28, no. 2, pp.237-250, 2017.
- [12] Min-ji Oh, Eun-seon Choi, Som Akhamixay

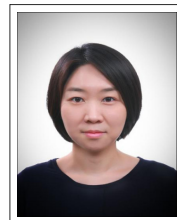
- Oui and Wan-sup Cho, “Predicting win-loss using game data and deriving the importance of subdivided variables”, The Korea Journal of BigData, vol. 5, No. 2, pp.231-240, 2020.
- [13] Ani, R., Harikumar, V., Devan, A. K., and Deepa, O. S., “Victory prediction in League of Legends using Feature Selection and Ensemble methods”, In 2019 International Conference on Intelligent Computing and Control Systems, pp. 74-77, 2018.
- [14] Thompson JJ, Blair MR, Chen L and Henrey AJ, “Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning”, PLoS ONE 8(9): e75129.
- [15] Lee, Sang-Kwang, Seung-Jin Hong, and Seong-Il Yang. “Predicting Game Outcome in Multiplayer Online Battle Arena Games”, International Conference on Information and Communication Technology Convergence, IEEE, 2019.
- [16] Riot Games, “Riot Developer Portal”, accessed Apr 13, 2021, <https://developer.riotgames.com>.
- [17] Gregorutti B., Michel B. and Saint-Pierre P, “Correlation and variable importance in random forests”, Statistics and Computing, vol. 27, pp.659-678, 2017.
- [18] Bzdok, D., Altman, N., and Krzywinski, M., “Statistics versus machine learning”, Nature methods, 15(4), pp 233 - 234, 2018.
- [19] Rich Caruana and Alexandru Niculescu-Mizil, “An empirical comparison of supervised learning algorithms”, In Proceedings of the 23rd international conference on Machine learning, pp.161-168, 2006.
- [20] Bentejac C., Csörgő, A. and Martinez-Munoz G., “A comparative analysis of gradient boosting algorithms”, Artificial Intelligence Review, vol. 54, pp.1937-1967, 2021.
- [21] Van Saeys, Iñaki Inza and Pedro Larrañaga, “A review of feature selection techniques in bioinformatics”, Bioinformatics, Volume 23, Issue 19, pp. 2507 - 2517, 2007.
- [22] Nicodemus K.K., “Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures”, Brief Bioinform, vol. 12, No. 4, pp.369-373, 2011.
- [23] Silva, A., G. Pappa and L. Chaimowicz, “Continuous Outcome Prediction of League of Legends Competitive Matches Using Recurrent Neural Networks.”, In Proceedings of SBGames, code: 188226, 2018.
- [24] Ellis Cashmore. “Making Sense of Sports”, Routledge, 2010.
- [25] Seong-Eun Seo and Chi-Yo Kim, “Recognition of the Type and Cause of <League Of Legend> Trolling”, Korea Game Society, vol. 15, No. 4, pp. 93-110, 2015.
- [26] Kyu Bok Lee and Young Jae Kim, “Analysis of Factors that Influence Users’ Preference for MOBA Game Genre: Focusing on the Game Systems of League of Legends”, Global Cultural Contents, vol. 47, pp. 107-124, 2021.



김용우 (Kim, Yongwoo)

약력 : 2018 Universität Lüneburg DataScience 석사
2021-현재 한양대학교 기술경영대학원 박사과정

관심분야 : 기계학습, 데이터마ining, e-Sports



김영민 (Kim, Young-Min)

약력 : 2010 Paris-VI 컴퓨터공학 박사
2020-현재 한양대학교 기술경영대학원 교수

관심분야 : 기계학습, 데이터마ining