

Win-Loss Prediction Using AOS Game User Data

Ye-Ji Kim*, Jung-Hye Min*

*Student, Dept. of Computer Science, Inha Technical College, Incheon, Korea

*Professor, Dept. of Computer Science, Inha Technical College, Incheon, Korea

[Abstract]

E-sports, a burgeoning facet of modern sports culture, has achieved global prominence. Particularly, Aeon of Strife (AOS) games, emblematic of E-sports, blend individual player prowess with team dynamics to significantly influence outcomes. This study aggregates and analyzes real user gameplay data using statistical techniques. Furthermore, it develops and tests win-loss prediction models through machine learning, leveraging a substantial dataset of 1,149,950 individual data points and 230,234 team data points. These models, employing five machine learning algorithms, demonstrate an average accuracy of 80% for individual and 95% for team predictions. The findings not only provide insights beneficial to game developers for enhancing game operations but also offer strategic guidance to general users. Notably, the team-based model outperformed the individual-based model, suggesting its superior predictive capability.

▶ **Key words:** Machine Learning, Win/Loss Prediction, Data Mining, Game Data

[요 약]

현대 사회의 새로운 스포츠로 정의되는 e-스포츠는 세계적으로 많은 사랑을 받는 스포츠로 자리매김했다. 그 중, E-sports를 대표하는 AOS(Aeon of Strife) 장르의 게임은 플레이어 개인과 팀의 운영이 승패를 좌우하는 요소가 된다는 특징을 가진다. 본 논문은 실제 유저들의 게임 데이터를 수집하고 데이터를 통계적 기법으로 분석하여 정보를 제공한다. 또한, 수집한 데이터를 활용하여 머신러닝 기법을 이용하여 승패 예측 모델을 설계하고 실험한다. 5개의 머신러닝 알고리즘이 사용되었고, 평균적으로 개인 데이터 모형에서는 Accuracy 80%, 팀 데이터 모형에서는 Accuracy 95%의 성능을 보인다. 본 연구에서 모형 설계 시 사용된 데이터는 개인 데이터 1,149,950건, 팀 데이터 230,234건으로 규모가 크고 일반 유저들의 플레이 성격을 잘 반영하고 있기 때문에 개발사의 게임 운영이나 일반 유저의 전략 수립 등에 도움이 될 것으로 기대한다. 실험 결과, 개인 데이터 모형과 팀 데이터 모형을 비교하였을 때, 팀 단위 모형의 성능이 상대적으로 매우 좋게 나타났다.

▶ **주제어:** 머신러닝, 승패 예측, 데이터 마이닝, 게임 데이터

-
- First Author: Ye-Ji Kim, Corresponding Author: Jung-Hye Min
 - *Ye-Ji Kim (yeji4268@naver.com), Dept. of Computer Science, Inha Technical College
 - *Jung-Hye Min (jhmin@inhac.ac.kr), Dept. of Computer Science, Inha Technical College
 - Received: 2023. 11. 13, Revised: 2023. 12. 11, Accepted: 2023. 12. 12.

I. Introduction

최근, 게임 시장의 규모는 꾸준히 성장하고 있다. 2021년 국내 게임 시장 매출액은 전년 대비 11.2% 증가한 20조 9,913억 원이고[1], 전 세계 게임 시장 규모는 258조 원으로 성장하는 추세를 보인다[2]. 본 연구는 많은 인기를 끌고 있는 리그 오브 레전드 게임 데이터를 머신러닝 기술을 이용해 분석한다.

게임 데이터를 분석하는 것은 게임 회사, 게임 유저, 게임 시장에 걸쳐 긍정적인 효과를 가져온다. 게임 회사의 입장에서 데이터 분석을 통해 승패에 많이 관여하는 부분들에 집중하고, 실제로 그 지표들을 게임의 운영에 반영할 수 있다. 유저의 입장에서는 분석 결과를 참고하여 게임 플레이 시, 집중해야 하는 부분을 파악하고 이를 통해 게임 전략 수립이 가능하다. 이는, 유저에게 게임의 재미를 올려 주는 요소로 작용할 수 있다. 실제로 E-sports 중계에서는 분석된 데이터 기반으로 중계하며, 시청자들은 분석 데이터를 보고 응원하는 팀을 고르거나 예측을 해보기도 하면서 시청의 재미를 느끼게 된다. 이를 통해 시청자들은 게임을 직접 플레이하지 않고도 게임을 즐길 수 있다[3]. 이는, 게임으로의 유입으로 이어져 전체 게임 산업의 규모 성장에 도움이 된다. 또한, E-sports 구단에서도 데이터 분석을 사용할 수 있다. 데이터 분석을 통해 현재 경기력을 평가하고, 해당 평가를 바탕으로 전략 수립이 가능하다.

연구 대상인 리그 오브 레전드는 라이엇게임즈(RiotGames)에서 개발, 서비스하고 있는 AOS(Aeon of Strife) 장르 게임이다. AOS 장르 게임은 정해진 맵에서 레벨과 스킬을 올리고 아이템을 갖추어 상대방의 진영을 파괴하는 실시간 공성 게임이다. 실시간으로 게임이 진행되는만큼 순간의 선택이 전략이 되어 승패에 많은 영향을 미친다. 이러한 이유로 E-sports 대회 중 가장 많은 시청자 수를 기록할 정도로 많은 사랑을 받으며 E-sports의 규모 성장을 불러오기도 했다.

본 연구는 일반 플레이어의 데이터를 사용한다. 그 이유는 프로게이머들의 데이터는 게임 개발사나 일반 플레이어들에게 활용되기 쉽지 않기 때문이다. 그렇기 때문에 본 연구는 기존 연구보다 더 넓은 범위의 대상에게 적용 가능하다는 점에서 활용성을 기대해 볼 수 있다. 또한, 데이터의 규모가 개인 데이터는 1,149,950건, 팀 데이터는 230,234건으로 데이터의 양이 충분하기 때문에 본 연구에서 사용되는 데이터는 일반 사용자를 잘 대표하고 있다고 할 수 있다. 그렇기 때문에 본 연구에서 제안한 모형은 일반 사용자를 포함한 넓은 활용 범위를 가지게 될 것이다.

본 연구의 구성은 다음과 같다. 2장에서는 국내 관련 연구 동향 분석과 사용한 알고리즘에 대한 설명을 진행한다. 추가로, 게임 데이터 분석을 위해 필요한 용어 설명도 포함 되어 있다. 3장에서는 데이터를 수집하고, 전처리하는 과정이 소개되어 있다. 그리고 4장에서는 수집된 데이터를 사용하여 해당 데이터를 시각화하는 방식으로 EDA(Exploratory Data Analysis)을 진행했다. 다양한 관점의 시각화에 따라, 게임 동향에 대한 정보를 제공하였다. 5장에서는 데이터를 개인 데이터와 팀 데이터로 나누어 각각의 승패 예측 모형을 설계하고 실험하였다. 또한, 다양한 머신러닝 알고리즘에 따른 성능 비교를 포함하였다. 마지막 6장에서는 전체적인 연구 결과를 정리하고, 승리를 위해 집중해야 할 요인들에 대해 언급했다.

II. Preliminaries

1. Related Works

스포츠에 대한 관중의 관심이 늘어남에 따라 승패 예측에 관한 연구도 활발히 진행되고 있다. 승패 예측에 관한 연구는 스포츠뿐만 아니라 승패 개념이 존재하는 다양한 분야에 적용 가능하기 때문에 가치를 가진다. 승패 예측의 대표적인 연구 기법은 기존의 데이터를 예측 모형에 학습시키고 해당 모형을 사용해 승패를 예측하는 모형을 만드는 것이다. 그렇기 때문에 예측 결과에 큰 영향을 미치는 요인들에 대해 분석하고 보다 정확한 결과를 도출하는 모형을 만드는 것이 승패 예측 연구의 핵심이 될 수 있다. 기존의 승패 예측 관련 연구들을 살펴보면, 다양한 기법을 적용하고 실험하는 것으로 모형의 정확도를 개선한다. 농구 경기 결과를 활용한 승패 예측 연구에서는 KNN, Decision Tree, Logistic Regression, SVM, Random Forest 알고리즘을 사용한 모형을 구현하고 각각의 정확도를 비교하는 실험을 진행했다. 그 결과, Random Forest 모형의 훈련 정확도가 가장 높게 나타났다. 그러나, 데이터의 샘플 개수가 충분하지 않아 테스트 정확도와 훈련 정확도 사이의 차이가 발생했고, 이를 과적합이라고 추정했다[4]. 이러한 연구 결과를 통해, 충분한 양의 데이터 샘플을 확보하는 것은 승패 예측 모형을 구현하는 과정에서 중요하다는 것을 알 수 있다.

본 연구의 목표인 게임의 승패 예측에 관한 연구는 크게 두 가지 관점으로 나뉜다. 첫 번째는 시간 관점에서의 연구이다. 시간 관점에서 진행된 연구는 게임의 데이터를 진행 시간 단위로 나누어 분석하고, 승패 예측 모형을 설계

하는 것이다[5]. 두 번째는 전체 게임 관점에서의 연구이다. 진행 시간과 상관없이 게임의 현재 시점에서의 데이터를 사용하는 것이다. 본 연구는 전체 게임 관점에서의 연구를 채택하였다. 해당 시점의 관련 연구에서는 요인 분석 기법을 이용해 승패에 영향을 미치는 게임 속 요인들을 분석하고 판별 분석과 인공신경망을 이용해 승패 예측 모델을 설계하였다. 그 결과, 판별 분석이 인공신경망에 비해 상대적으로 좋은 성능을 보이는 것을 확인했다[6]. 추가로, 대부분의 관련 연구에서는 e-sports에서의 게임의 승패를 예측하는 것을 목표로 삼고 프로게이머나 최상위권 티어의 데이터를 사용한 것을 확인할 수 있었다. 반면, 본 연구에서는 일반 사용자들에게도 적용 가능한 승패 예측 모델을 만들고자 하기 때문에 일반 사용자의 데이터를 활용하여 연구의 차별화를 두고자 한다.

2. Machine Learning Algorithm

정형 데이터를 사용하는 경우, 딥러닝 계열 알고리즘보다는 트리 기반 알고리즘이 성능이 높다고 알려져 있다[7]. 따라서, 본 연구에서도 해당 알고리즘을 사용하였고, 비교를 위해 부스팅 계열 알고리즘을 함께 사용하였다. 사용하는 알고리즘과 평가 지표에 대한 설명은 다음과 같다.

- Random Forest Classifier

분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 하나로, 훈련 과정에서 구성한 다수의 결정 트리를 랜덤하게 학습시켜 분류 또는 회귀의 결과 도출에 사용한다.

- Gradient Boosting Classifier

약한 학습기의 오류 데이터에 가중치를 부여하면서 부스팅을 수행하는 대표적인 알고리즘이다. 가중치 업데이트에 경사 하강법을 이용하여 최적화된 결과를 얻는다.

- XGB Classifier

뛰어난 예측 성능, 과적합 규제, 교차 검증 내장으로 다양한 분석에 사용되는 XGB는 잔차로부터 훈련한 트리를 추가하여 약한 학습기를 강력한 학습기로 바꾸는 알고리즘이다.

- Logistic Regression

독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는 데 사용하는 기법으로, 일반적인 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수 간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용하는 알고리즘이다.

- LightGBM

트리 기반의 학습 알고리즘으로, 다른 부스팅 알고리즘과 달리 트리의 수직 확장으로 동작하는 Leaf-wise 알고

리즘이다. 빠른 속도와 적은 자원 사용량으로 널리 사용되고 있다. 파라미터에 따라 성능 차이가 발생하기 때문에 하이퍼파라미터 튜닝이 모델의 성능에 큰 영향을 미치는 알고리즘이다.

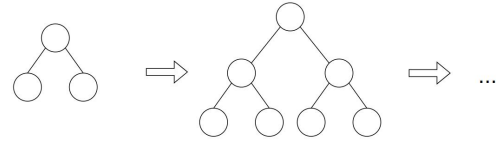


Fig. 1. Level-wise

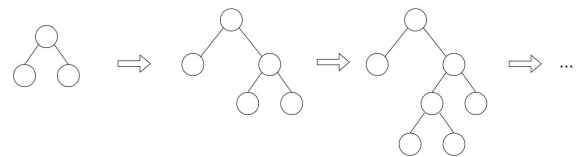


Fig. 2. Leaf-wise

- 모델 성능 평가 지표

본 연구에서는 구축한 모델의 성능을 혼동행렬의 4가지 지표로 평가한다. Accuracy는 전체 데이터 중에서 올바르게 분류된 데이터의 비율로, 분류 모델 성능 평가 시 가장 대표적으로 사용되는 지표이다. Recall은 실제 Positive 데이터에서 Positive로 맞게 분류된 데이터의 비율을 말한다. Precision은 Positive로 예측된 데이터 중 실제로 Positive인 데이터의 비율이다. 분류 시, Positive (승리)를 더욱 정밀하게 예측하기 위한 지표이다. Recall, Precision은 Trade-off 관계이기 때문에 이들을 조화 평균한 F1-Score를 사용한다.

3. Description of Terminology

Table 1은 리그 오브 레전드 게임 데이터에서 사용되는 용어의 설명이다.

Table 1. Description of game terminology

Terminology	Description
Level	Experience points obtained by the user each time they play the game. This scale allows for the assessment of the player's passion and effort invested in the game.
Tier	In games, tiers and ranks are used to represent skill levels. In League of Legends, the hierarchy is structured as Iron, Bronze, Silver, Gold, Platinum, Diamond, Master, Grandmaster, and Challenger in order.
Position	Position refers to the attacking role that a player takes on in the game, divided into Top, Mid, Bottom, Utility, and Jungle.

Dragon	It is a monster that spawns based on the game's duration. It appears in the southeast of the map (Summoner's Rift). Upon defeating the dragon, the team receives a buff based on the dragon's element.
Rift Herald	An epic monster that spawns in the northwest of the map (Summoner's Rift)
Baron	It is a neutral monster that spawns in the northwest of the map (Summoner's Rift). Upon defeating it, the team gains positive effects.
Minion	They are soldiers that move along the lanes and attack the towers or nexus of the opposing team. Upon defeating them, you can gain gold and experience.
Tower	Each lane has three structures, and there are two structures in front of the Nexus.
Inhibitor	Structures that exist around the Nexus, one in each lane.
View Score	The score that represents the extent to which a player contributes to securing the team's vision or disrupting the enemy team's vision.
Kill	Eliminating enemy team players. Upon player elimination, player gains gold.

III. Data Collection & Preprocessing

승패 예측 모형 설계를 위해서는 필요한 데이터를 수집하고 전처리하는 과정이 필요하다. 본 연구에서는 개발사에서 제공하는 API를 사용해 데이터를 수집하고, 수집된 결과를 전처리하여 모형 설계에 사용될 데이터를 확보했다.

1. Data Collection

모형 설계에 사용되는 데이터는 일반 플레이어들의 실제 게임 데이터이다. 해당 데이터의 수집 범위를 정하기 위해 게임 데이터 수집에 앞서 계정 데이터를 먼저 수집하고, 사용자 티어와 레벨 간의 관계를 확인한 뒤, 범위를 결정하는 과정을 진행했다.

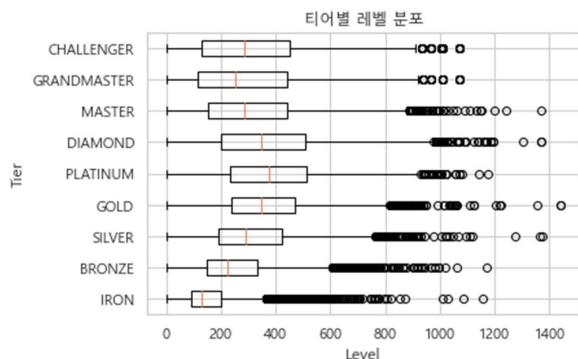


Fig. 3. Level Distribution by Tier

Fig. 3은 수집된 데이터 중 아이언부터 챌린저 티어까지 티어별 5,000명의 계정 레벨을 시각화한 결과이다. 아이언부터 플래티넘 티어까지는 티어가 상승할수록 사용자 레벨의 중앙값이 비례하여 올라가는 형태를 확인할 수 있다. 그러나, 다이아몬드부터 챌린저 티어까지는 이러한 경향을 따르지 않기 때문에 수집 대상에서 제외하였다. 따라서, 데이터 수집은 리그오브레전드 한국 서버의 5*5 랭크 게임, 소환사의 협곡 맵에서 플레이된 아이언부터 플래티넘 티어의 게임을 대상으로 진행하였다.

리그 오브 레전드의 개발사인 RiotGames에서는 유저의 게임 데이터를 API(Application Programming Interface) 형식으로 제공하고 있다[8]. 따라서, 본 연구에서는 Python 언어를 사용해 개발사로부터 제공되는 여러 개의 API를 연결하여 필요한 데이터를 수집했다.

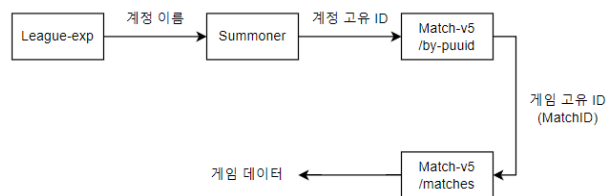


Fig. 4. API Usage Order

Fig. 4는 데이터 수집에 사용된 API의 사용 순서이다. 호출한 결과를 다음 순서의 API의 입력으로 사용하여 최종적으로 게임 데이터를 얻는 방식을 이용했다. 총 4개의 API를 거쳐 마지막 단계에서 결과로 받은 JSON 타입의 데이터는 Pandas의 DataFrame 타입으로 변환되어[9] CSV(Comma Separated Values) 파일 형태로 저장되게 하였다. 이러한 과정을 거쳐 각 티어별로 25,000건, 즉, 총 125,000건의 게임 단위 데이터를 수집할 수 있었다.

2. Data Preprocessing

수집된 데이터를 승패예측모형에 사용하기 위해서는 전처리 과정이 필요하다. Table 2는 각 데이터에 진행한 전처리 항목을 정리한 표이다.

Table 2. Pre-processing List

Data	Preprocessing
User(Account) Data	Missing Value Handling: Delete records with missing values
	Creating Derived Variables : Total number of games(win+loss)
	Extracting Data of the Top 5000 Players (Sorting Criteria: Total Number of Games)

In-game Data: Individual	Missing Value Handling: Delete records with missing values
	Splitting Game-level Data into Individual Data for Each of the 10 Players
In-game Data: Team	Missing Value Handling: Delete records with missing values
	Creating Derived Variables Related to Objectives: Baron, Dragon, Inhibitor, Herald, Tower
	Creating Team-level Derived Variables: Kills, Deaths, Total Damage Dealt, Minions Killed

리그 오브 레전드 게임 내에서는 승패에 영향을 미치는 랜덤한 요소가 거의 존재하지 않는다. 이러한 특성상, 이상치가 눈에 띄게 나타나지 않고 데이터의 분포가 특정 범위 내에 위치하기 때문에 예측 모델의 입력에 사용될 데이터에 정규화 작업을 하는 경우, 성능 개선의 효과가 크게 나타나지 않을 것으로 추측했다. 이러한 가설을 검증하기 위해 이후 모델 설계 단계에서 해당 항목을 실험하여 확인했다.

3장에서는 승패 예측 모형을 설계하기 위해 필요한 데이터를 구축하는 과정을 진행했다. 계정의 티어와 레벨 간의 관계 시각화를 통해 수집 범위를 결정하고, 개발사에서 제공하는 API를 사용해 수집 범위에 맞는 데이터를 수집했다. 수집 결과, 레코드 1개당 한 게임에 참여한 10명의 게임 정보를 담고 있는 125,000건의 데이터가 수집되었다. 추가로, 수집된 데이터를 개인 데이터와 팀 데이터로 나누어 전처리하였고 그 결과, 개인 데이터는 1,149,950건, 팀 데이터는 230,234건의 데이터가 구축되었다.

Table 3. Volume of Data

Data	Description	Size
Game Data	Comprising Data for 10 Individuals	120,000
Team Data	Comprising Data for 5 Individuals (Forming a Team of 5)	240,000
Individual Data	Comprising Data for 1 Individual	1,200,000

IV. Game Data Analysis

승패 예측 모형을 생성하기 전, 수집된 데이터를 파악하기 위한 EDA(Exploratory Data Analysis)를 진행했다. 사용자의 티어와 각 속성 간의 관계나 포지션별 속성 간의 관계 등의 시각화를 통해 승패 예측 모형의 성능에 영향을 줄 수 있는 속성들을 파악하였다.

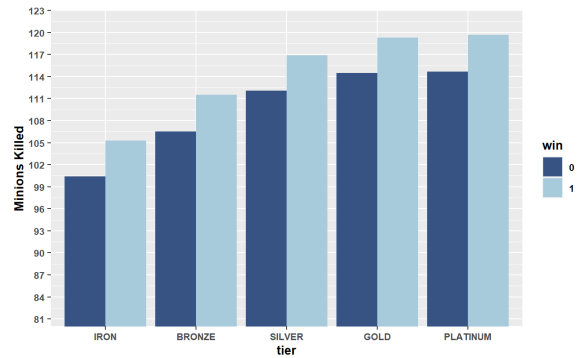


Fig. 5. Minion Kills by Tier

Fig. 5는 승패에 따라 티어별로 미니언의 처치(kill) 수 차이를 확인할 수 있는 그래프이다. 티어가 올라가는 만큼, 평균적인 미니언 처치 수가 상승하는 경향을 보인다. 추가로, 게임에서 승리했을 경우(1)가 패배했을 경우(0)보다 전체적인 수치가 높은 것을 확인할 수 있는데, 이는 미니언 처치와 승리가 양의 관계에 있다고 해석할 수 있다.

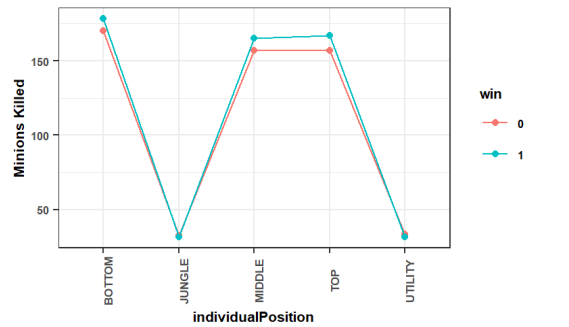


Fig. 6. Minion Kills by Position

Fig. 6은 승패에 따른 포지션별 미니언 처치 수를 나타내는 그래프이다. 승리한 경우(1)와 패배한 경우(0)의 차이는 JUNGLE, UTILITY 포지션을 제외한 3개의 포지션에서 나타났다. 그 이유는, 리그 오브 레전드의 각 포지션에 따른 역할 차이에서 나온다. 그에 따라 미니언 처치가 요구되지 않는 JUNGLE, UTILITY 포지션의 경우, 미니언을 처치한다고 해도 승리에 유의미한 영향을 주지 않는다는 것이다.

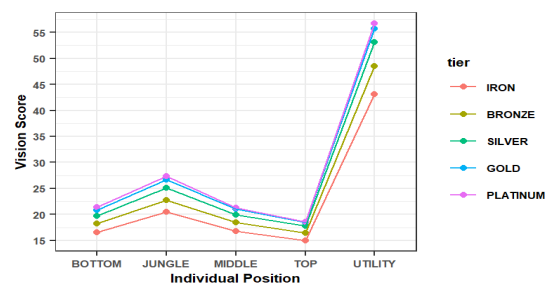


Fig. 7. Distribution of Vision Scores by Position

Fig. 7은 포지션별로 시야 점수를 티어에 따라 구분한 그래프이다. x축은 티어, y축은 승리한 게임의 시야 점수 (Vision Score) 평균값이다. 위 표를 보면, 승리를 위해 시야 점수가 가장 많이 요구되는 포지션은 UTILITY 포지션으로 나타났다.

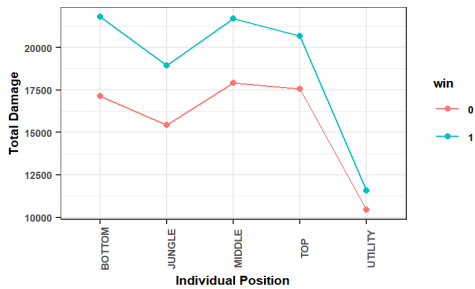


Fig. 8. Damage Dealt by Position

Fig. 8은 포지션별 가한 데미지를 승패에 따라 구분한 그래프이다. TOP, UTILITY를 제외한 3개의 포지션에서 승리했을 경우 가한 데미지와 패배했을 경우의 차이가 가장 크게 나타났다. 그에 비해 UTILITY 포지션에서는 가장 작은 차이를 보였다.

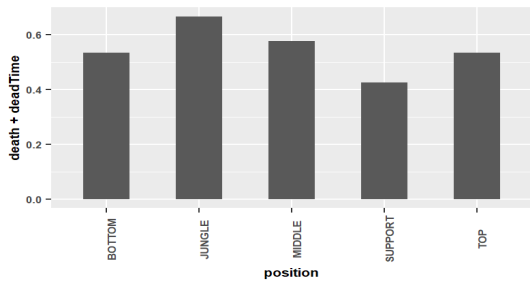


Fig. 9. The Impact of Positional Deaths on Victory or Defeat

Fig. 9는 죽음에 관련된 변수인 죽은 횟수와 죽어있던 시간의 중요도를 합산하여 승패에 미치는 영향을 나타낸 그래프이다. JUNGLE 포지션의 죽음이 승패에 가장 큰 영향을 미치는 것으로 나타났다.

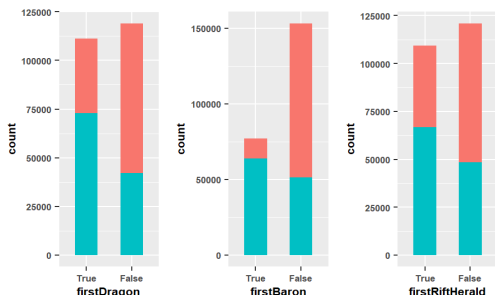


Fig. 10. Victory and First Objective Takedown Status (Blue : Win, Red : Loss)

Fig. 10은 첫 드래곤, 바론, 협곡의 전령을 처치했는지의 여부에 따른 승패 횟수를 시각화 한 그래프이다. 세 항목 모두 처음 등장하는 오브젝트를 처치하는 경우, 승리하는 비율이 높았다. 이것으로 보아, 첫 오브젝트를 처치하는 것은 승리할 확률이 높다고 볼 수 있다.

탐색적 데이터 분석(EDA) 결과, 개인 데이터와 팀 데이터 각각에서 예측 모형에 영향을 줄 수 있는 속성들과 그들 간의 관계, 각 포지션에게 요구되는 역할 등을 파악할 수 있었다.

V. Win-loss Prediction Model

리그 오브 레전드는 5명이 한 팀을 이루어 두 팀이 승리를 놓고 경쟁하는 전략 게임이다. 따라서, 본 연구의 목표인 승패 예측에서는 팀 단위 데이터와 팀을 이루는 개인의 데이터로 나누어 모형을 설계하는 것이 핵심이 될 수 있다. 개인 단위 모형과 팀 단위 모형의 구축을 통해 개인과 팀에게 승리를 위하여 요구되는 속성을 파악하고 승패를 예측하는 과정을 진행했다. 사용된 알고리즘은 LGBM, XGBoost, Random Forest, Gradient Boost Classifier, Logistic Regression 으로, 해당 알고리즘의 설명은 2장 2절에 설명되어 있다. 입력으로는 API를 통해 게임 진행 중에도 실시간으로 수집 가능한 항목으로 구성되어 있는 개인 단위 데이터와 팀 단위 데이터를 사용한다. 설계된 모형의 출력은 승패의 여부를 나타내므로, 분류 문제에 해당한다.

1. Win-loss Prediction Model: Individual Data

팀 단위 게임에서 개인의 플레이는 팀 단위 플레이와 비교했을 때 상대적으로 중요도가 떨어져 보일 수 있지만, 게임에서의 팀 운영은 개인의 플레이가 모여서 이루어지기 때문에 주목할 만하다. 또한, 일반 플레이어의 경우, 대체로 랜덤한 플레이어들과 팀으로 매칭되어 게임을 플레이 하기 때문에 개인의 전략 수립도 팀 전략 수립만큼 승리에 영향을 미칠 수 있다. 이런 점을 고려하여 개인 데이터를 사용하여 개인의 관점에서 승패를 예측하는 모형을 설계하는 작업을 진행했다.

1.1 Design of Prediction Model

- 사용 데이터

Table 4. Variables of Player data

Variable Name	Data Type	Description
tier	category	Player Tier
duration	int	Game Duration
assists	int	Assist Count
deaths	int	Death Count
individualPosition	category	In-game Position
kills	int	Player Kill Count
firstBloodAssist	bool	First Assist
firstBloodKill	bool	First Kill
firstTowerAssist	bool	First Tower Destruction
firstTowerKill	bool	First Tower Destruction
totalDamageDealtToChampions	int	Total Damage Dealt to Champions (Players),
totalHeal	int	Total Healing
totalMinionKilled	int	Total Minion Kills
visionScore	int	Vision Score
win	bool	Victory Status

Table 4는 플레이어 개인의 데이터를 설명한 표이다.
-모델별 성능 비교

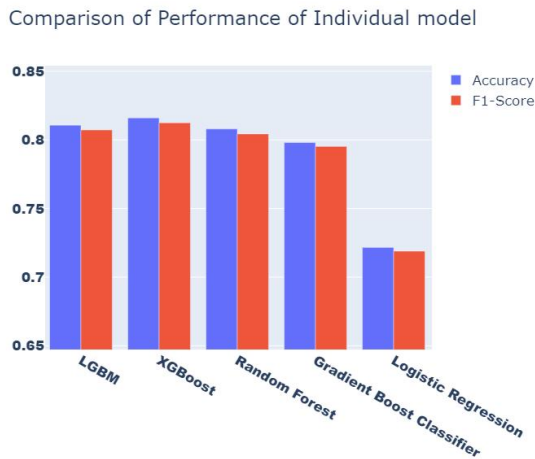


Fig. 11. Comparison of Performance of Player Models

Fig. 11는 개인 단위 데이터를 사용해 단일 모델을 구축하여 승패를 예측하였을 때의 모델의 정확도를 나타낸 결과이다. 상위 2개 모델은 XGBoost, LGBM으로 나타났으며, 각각 Accuracy 기준 0.8160, 0.8107로 두 모델을 제외한 3개의 알고리즘과 비교하였을 때, 상대적으로 성능이 좋은 것을 확인할 수 있었다.

예측 모형 설계에서 사용되는 데이터에 정규화 작업을 진행하는 것은 모형 성능의 향상을 가져온다고 알려져 있

다. 이런 현상이 본 연구에서 설계하는 모형에 적용되는지 확인하기 위해 Standard Scaler를 사용해 정규화를 진행했다. 모델에 적용한 결과, XGBoost 모델 기준 Accuracy 0.8160으로, 단일 모델과 비교했을 때, 큰 차이를 보이지 않았다. 모델에 사용된 데이터에서 주목할 만한 이상치가 나타나지 않았고, 값이 특정 범위 내에 분포하고 있기 때문에 정규화 작업으로 큰 효과를 보지 못했다.

추가로, 단일 모델에서 상대적으로 좋은 성능을 보였던 상위 2개의 모델을 Voting 방식으로 앙상블 하여 모형을 설계했다. 그 결과, Accuracy 기준 0.8182로, 큰 성능 향상을 보이지는 않았다.

1.2 What Variable Is Most Important?

각 특성이 승리에 기여하는 정도를 확인하기 위해 python의 shap 라이브러리를 사용하여 시각화 작업을 진행했다[10]. shap 라이브러리는 속성 간의 상호작용에서 변하는 상황을 이해하고, 어떻게 될 것인지 예측하기도 하며, 순간의 의사 결정에서 어떻게 결정하는 것이 더 득이 되는지를 계산하는 게임 이론을 구현한 라이브러리이다. Fig. 12는 shap 라이브러리를 사용하여 각 속성 값의 shapely value를 시각화한 것이다.

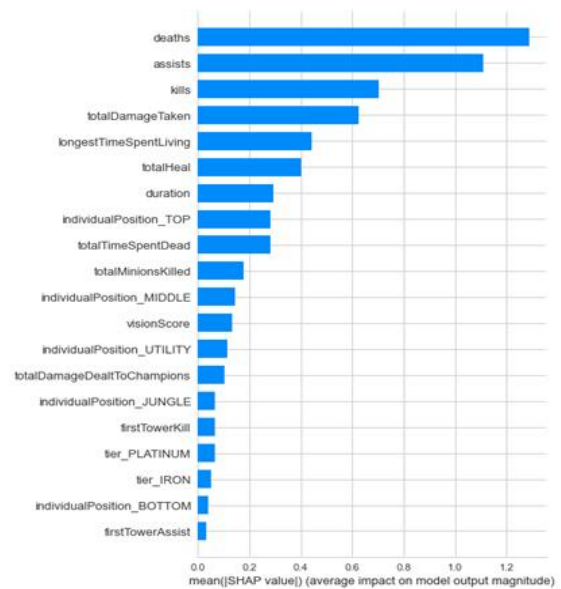


Fig. 12. Shapely value of Player data variables

플레이어가 게임 내에서 맡은 포지션에 따라 각 특성의 기여도가 달라질 수 있지만, 공통적으로 승리에 영향을 미치는 특성을 확인할 수 있었다. 분석 결과, 상위 특성 3개는 deaths(죽은 횟수), assists(처치 기여 수), kills(처치 수)로 나타났다. 게임 내에서 플레이어 개인은 다른 플레이어를 처치해 kills(처치 수)를 늘리는 것보다 deaths(죽은 횟수)를

줄이는 것이 승리에 좋은 영향을 미친다는 것이다. 이는, 일반 플레이어 간의 게임에서는 한 플레이어의 활약이 팀의 승리를 가져오지 못한다는 것을 의미하기도 한다.

2. Win-loss Prediction Model : Team Data

AOS 장르 게임에서는 팀 단위의 좋은 전략이 팀의 승리를 가져온다. 본 연구에서는 구축된 팀 데이터를 사용해 승패 예측 모델을 설계함으로써 팀의 전략에 도움이 될 속성을 파악하고 승패를 예측하는 것에 활용될 수 있도록 한다.

2.1 Design of Prediction Model

- 사용 데이터

Table 5. Variables of Team data

Variable Name	Data Type	Description
team	category	Team Classification
firstBaron	bool	First Baron Kill
baron	int	Baron Kills Count
firstDragon	bool	First Dragon Kill
dragon	int	Dragon Kills Count
firstInhibitor	bool	First Inhibitor Kill
inhibitor	int	Inhibitor Kills Count
firstRiftHerald	bool	First RiftHerald Kill
riftHerald	int	RiftHerald Kills Count
firstTower	bool	First Tower Kill
tower	int	Tower Kills Count
kills	int	Total Player Kills
deaths	int	Total Deaths
totalDamageDealtToChampions	int	Total Damage Dealt
totalMinionsKilled	int	Total Minions Killed
win	bool	Victory Status

- 모델별 성능 비교

Comparison of Performance of Team model

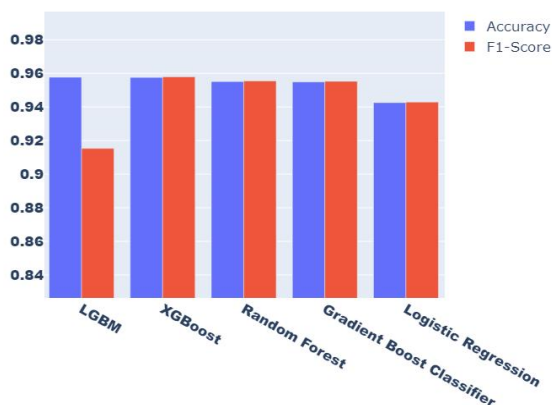


Fig. 13. Comparison of Performance of Team model

Fig. 13은 전처리한 데이터를 사용하여 단일 모델을 구축한 결과이다. 구축한 5개의 모델 모두 Accuracy 0.94 이상의 성능을 보였다.

앞서 개인 예측 모형에서 테스트 했던 것과 같이 팀 모형 설계에도 동일하게 정규화 과정을 진행했다. 결과로는 LGBM 모델의 Accuracy 기준 0.4978로, 성능이 크게 떨어지는 것을 확인했다. 이는, 이미 데이터가 정규성을 충족한 상태이기 때문에 추측된다.

마지막으로, 단일 모델 설계에서 상대적으로 좋은 성능을 보인 2개의 모델을 사용해 앙상블 기법을 이용했다. 그 결과, Accuracy 0.9585의 성능을 확인할 수 있었다. 앞서 구축된 5개의 단일 모델은 모두 Accuracy 0.94 이상의 좋은 성능을 보였지만, 모델의 안정적인 성능을 위해서는 앙상블 기법을 활용한 모델을 사용할 수 있다.

2.2 What Variable Is Most Important?

Fig. 14는 shap 라이브러리를 활용해 팀 단위 데이터의 승리 기여도 특성을 시각화한 것이다.

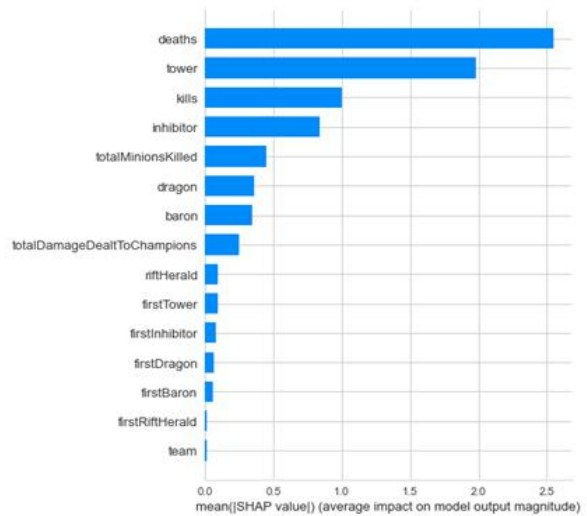


Fig. 14. Shapley Value of Team Data Variables

팀 예측 모형에서는 기여도 상위 3개 특성으로 deaths (죽은 횟수), tower(타워 처치 수), kills(처치 수)를 확인할 수 있었다. 여기서 주목할 점은, 팀 예측 모형의 속성 기여도에서는 개인 예측 모형의 경우와 달리 tower(타워 처치 수)의 높은 기여도를 보인다는 것이다. 이는 각 공격 경로에 있는 타워를 파괴하고 최종적으로 상대 진영의 넥서스를 파괴하면 승리하는 승리 조건과의 높은 연관성으로 인한 결과로 추측된다.

VI. Conclusions

본 연구에서는 리그 오브 레전드 API를 사용하여 수집한 데이터를 탐색적 데이터 분석(EDA)한 뒤, 승패 예측 모형을 생성했다. 탐색적 데이터 분석 작업에서는 수집한 데이터의 시각화를 통해 승패에 따른 여러 특성들의 변화를 확인하고 예측 모형에서 사용될 데이터의 변수 선택에 활용했다. 수집된 데이터는 승패 예측 모형을 설계하는 데에 사용되었는데, 구축된 두 모델을 비교했을 때, 모델의 성능 면에서 차이를 확인할 수 있었다. 이는 사용한 데이터가 프로 리그 게임의 데이터가 아닌 일반 유저의 데이터이기 때문으로 추측한다. 일반 플레이어들의 게임은 프로 리그 게임과 달리 실력이 평균화되어 있지 않고, 전략을 미리 정하지 않은 상태에서 게임에 참여하는 특성을 가지고 있다. 따라서, 프로 리그 게임에서는 한 플레이어의 활약이 팀의 승리를 가져오는 경우가 흔하게 존재하지만, 이것이 일반 유저들의 게임까지 적용되지는 않는다는 것이다. 이러한 이유가 있기 때문에 개인 단위 모델에서는 승패를 예측하는 과정에서 상대적으로 정확도가 높기 어려울 것으로 추측한다.

본 연구를 통해 얻은 결론은 다음과 같다.

- JUNGLE 포지션의 죽음이 승패에 가장 큰 영향을 미친다.
- 첫 오브젝트 처치 여부는 승리에 유의미한 영향을 준다.
- 리그 오브 레전드에서 승패를 예측할 때는 개인 단위 데이터보다 팀 단위 데이터를 사용하는 것이 더 좋은 성능을 기대할 수 있다.
- 두 가지 모델 모두 XGBoost, LGBM 알고리즘 성능이 가장 좋았다.
- 개인의 플레이에서 승패에 영향을 미치는 상위 3개 속성은 다음과 같다.
 - deaths(죽음 횟수)
 - assists(처치 기여 횟수)
 - kills(플레이어 처치 횟수)
- 팀의 플레이에서 승패에 영향을 미치는 상위 3개 속성은 다음과 같다.
 - deaths(죽음 횟수)
 - tower(타워 처치 횟수)
 - kills(플레이어 처치 횟수)

본 연구는 일반 플레이어들의 데이터를 1,000,000건 이상 사용하였기 때문에 대표성이 있고, 신뢰도가 높아 활용 가치가 있다고 생각한다. 개인과 팀의 데이터를 분리하여 모형을 설계함으로써 본 게임의 사용자, 게임 개발사 등에

활용될 수 있을 것으로 기대한다. 게임 개발사는 승패에 따른 지표를 분석하고 그 결과를 게임 운영을 위한 업데이트에 반영하는 것에서 긍정적인 효과를 볼 수 있다. 개인 예측 모형에서 사용 챔피언(캐릭터) 별 총 치유량, 입힌 데미지 등의 기여도를 파악하여 챔피언(캐릭터) 밸런스 조정이 가능하고, 팀 예측 모형에서는 승패에 영향을 미치는 오브젝트를 파악하고, 원하는 방향으로 조절을 하는 데에 분석 결과를 참고 자료로 사용될 수 있다. 또한, 게임의 플레이어 측면에서는 승리를 위해 집중해야 하는 속성을 확인하여 플레이 전략을 세울 수 있다는 점에서 본 연구의 분석 결과와 예측 모형은 게임의 재미를 느낄 수 있는 요소로 작용할 수 있다. 그러나, 개인 예측 모형의 성능이 팀 예측 모형의 성능만큼 높지 않기 때문에 적극적인 활용에는 한계가 있을 수 있다. 그러므로, 보다 나은 예측을 위해서는 사용된 속성 이외의 속성을 추가로 수집하여 모형을 실험하는 것이 필요해 보인다. 추후 더 높은 정확도를 가진 모형의 개발로 승패 예측 모형이 적극적으로 활용되는 것을 기대한다.

REFERENCES

- [1] <https://www.kocis.go.kr/koreanet/view.do?seq=1043662>
- [2] <https://www.joseilbo.com/news/htmls/2023/09/20230906496893.html>
- [3] Jae-Woo Lee, Jong-Won Lee, and Hyo-Nam Kim, "A study on the Revitalization of e-Sports by Analyzing e-Sports Events," Proceedings of the Korean Society of Computer Information Conference , pp. 53-55, 2021.
- [4] Yuan-Zhen NI, and Seong-No 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. DOI :10.21797/ksme.2022.24.3.005
- [5] Lee J.M.. Designing a Win-Loss Prediction Model for League of Legends Using Machine Learning and Providing Metrics Master's Thesis form Ajou University in South Korea, 2, 2021.
- [6] Ku J.M. and Kim J.H., Development of game indicators and winning forecasting models with game data, Journal of the Korean Data & Information Science Society, 28, 2, 237-250, 3, 2017, DOI : 10.7465/jkdi.2017.28.2.237
- [7] Leo Grinsztajn, Edouard Oyallon, Gael Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," Advances in Neural Information Processing I. Introduction Systems 35, pp. 507-520, 2022, DOI : 10.48550/arXiv.2207.08815

- [8] League of Legend Developers Web Site, Available : [https://
developer.riotgames.com/](https://developer.riotgames.com/)
- [9] Pandas library : <https://pandas.pydata.org/>
- [10] Shap library : <https://shap.readthedocs.io/>

Authors



Ye-Ji Kim received the A.S. degree in Department of Computer Science from Inha Technical College in 2023 and is currently a B.S. in Department of Computer Science at Inha Technical College, Incheon, Korea.

Her research interests include data science, artificial intelligence, data analysis.



Jung-Hye Min received the B.S. degree in mathematics from Ewha Women's University, Seoul, Korea, in 1995, and the M.E. and Ph.D. degrees in computer science and engineering from the Pennsylvania State

University, U.S.A. in 2003 and 2005, respectively. From 2005 to 2021 she was a Principal Enginner with Samsung Research, Samsung Electronics, Seoul, Korea. Dr. Min joined the faculty of the Department of Computer Science at Inha Technical College, Inchoen, Korea in 2022 and is currently an assistant Professor. Her research interests include image enhancement, image style transfer, and pattern recognition.