CHAPTER I
INTRODUCTION

1.1 Research Background

Esports, short for electronic sports, has grown immensely in the past decade with the advent of online streaming media platforms such as YouTube and Twitch. Though esports and competition in video games had been around since its inception, it was not until the early 2000s where South Korea first licensed professional esports athletes, bringing mainstream media attention towards esports as well as classifying it as a legitimate sport. Esports has also managed to find itself in international sporting events, having been part of the Asian Games since 2022 China and the Southeast Asian Games since 2019. Esports encompasses a wide variety of video games and virtual sports, though as of 2024, by far the largest and most popular esports is the video game League of Legends (LoL), amassing a peak of 6.8 million viewers during its annual world championship, with a total prize pool of 2 million US dollars.

League of Legends was developed in 2009 by Riot Games. In the game, two teams consisting of five players compete against each other to simultaneously attack and defend their side's "Nexus". The team who destroys their opposing Nexus first wins. To achieve this, players pick one of over 160 champions in the game, each with their own distinct roles. They can accumulate gold to buy items and experience to level up through killing opposing players, neutral monsters and minions, or getting objectives scattered throughout the map.

Professional tournaments began in 2011 with the Season 1 World Championship at Jönköping, Sweden. As of 2024, League of Legends esports is divided into four main leagues: the LoL Championship Series (LCS) in North America, the LoL European Championship (LEC) in Europe, the LoL Championship Korea (LCK) in South Korea, and the LoL Pro League (LPL) in China. These four leagues make up the "major regions", while other regions like South America and Southeast Asia are relegated to their own minor leagues. Every region goes through their own season every year, with the best performing teams from each region getting sent to the world championship, commonly referred to as Worlds, at the end of the season to compete.

With such a large, robust, and structured professional circuit, many teams and organizations have poured resources into molding and improving their roster. This includes coaches, managers, and analysts. Like regular sports, esports teams have utilized sports analytics to give them a competitive edge over others by helping inform players and the coaching staff with strategies, decision-making, and roster moves for their season. One way for data to provide value is through predicting match outcomes, giving casual and professional players and staff alike insight into perceived strengths, weaknesses, and other features that contribute to what goes into winning a match.

Previous research on predicting match outcomes in eSports, particularly in League of Legends, has established a strong foundation for predictive modeling. These studies have successfully demonstrated the viability of various machine learning algorithms, such as Logistic Regression, Random Forests, and neural networks, for this prediction task (Ani et al, 2019). Another key focus has been feature engineering, with researchers identifying and evaluating the importance of specific in-game metrics to build predictive models (Omar et al, 2024).

However, much of this research has relied on data from non-professional ranked games, where gameplay dynamics are less coordinated, and player skill levels are more variable. While these efforts have contributed significantly to the development of predictive frameworks, they often lack the context and complexity inherent in professional esports matches.

Furthermore, while these efforts confirm the general feasibility of match prediction, a research gap often persists in the direct, methodological comparison between different model complexities, especially within the unique context of the professional esports ecosystem. Many studies tend to focus on applying a singular, often highly complex, model to maximize a performance metric, with less investigation into whether that complexity is necessary. This leaves a crucial question under-explored: Is the predictive signal within the highly structured professional dataset fundamentally linear, or does it contain intricate, non-linear interactions that can only be captured by more sophisticated ensemble methods? This study directly addresses this gap by performing a rigorous, side-by-side comparison of tuned linear and non-linear models to determine the most effective and appropriately complex approach for this domain.

1.2 Research Problem

The core objective of this study is to determine the most effective and appropriately complex supervised machine learning approach for predicting outcomes in professional League of Legends esports. While previous research has often focused on applying a singular, complex algorithm to maximize predictive accuracy, a research gap exists in the systematic, head-to-head comparison of different model architectures within the unique professional ecosystem. This study aims to fill that gap by investigating a critical question: Is the intricate, non-linear pattern recognition of ensemble models like Random Forest and XGBoost necessary, or can a simpler, more interpretable linear model achieve comparable performance on professional match data?

Furthermore, this research addresses a second, related gap concerning the applicability of high-elo ranked data as a proxy for professional play. To investigate this, the optimal model architecture identified from the professional data analysis will also be trained and evaluated on a dataset of Korean Challenger-tier ranked games. By comparing the final performance and, more importantly, the feature importances of the models trained on these two distinct datasets, this study seeks to identify any meaningful differences in the statistical patterns that define victory in each environment. This dual analysis will thereby provide not only a clear methodological benchmark for predicting professional games but also valuable insights into the transferability of predictive signals between the ranked and professional ecosystems.

1.3 Research Objectives

The objectives of this research are:

1. To develop and evaluate a supervised machine learning model capable of predicting match outcomes in professional League of Legends esports.

2. To compare the performances of linear machine learning algorithms and complex machine learning algorithms in predicting match outcomes in professional League of Legends esports.
3. To compare the differences in results between models trained with professional match data and public ranked match data.

1.4 Research Scope

The scope of data covered in this research will only be for major region games in the 2020 to 2024 Season, those games being from:

1. LEC (European League)
2. LCS (North American League)
3. LCK (South Korean League)

The decision for only using major region games is because of data accessibility and quality of the games. Though the Chinese league, the LPL, is also considered a major league, the accessibility of data from the league is very limited compared to other regions, therefore LPL teams and match data will not be considered for this research.

The data for public matches will be taken from the South Korean servers and will only include matches in the higher elo ranks (Master, Grandmaster, Challenger) during the 2025 season. The South Korean server is widely accepted by the community to be the most competitive server, and selecting the highest ranks ensures representation of the most skilled players not in professional play and it also reduces the statistical noise and unpredictable gameplay common in lower ranks.

1.5 Research Advantages

One of the main advantages of this research is its focus on professional-level esports data, offering insights that are directly applicable to the competitive scene of League of Legends. By utilizing match statistics from the highest tiers of play, the study captures patterns and dynamics that are more structured, strategic, and performance-driven. All the features will also be aligned with the end goal of predicting professional games.Furthermore, the study's primary contribution is its direct, head-to-head comparison of a simple, interpretable linear model against more complex ensemble methods. A significant advantage of this approach is the potential to determine if predictive power must come at the cost of interpretability.