Predicting Early Game Success in League of Legends

Lukáš Častven, Michal Kilian

2025-05-03

Our goal

League of Legends (LoL) is a 5v5 multiplayer online battle arena where teams compete to destroy the enemy Nexus. The early game is a critical phase focused on gaining initial advantages. Our primary goal is to predict the eventual winning team using only data from this limited timeframe (first 12 minutes).

Dataset

Data recorded what happened in pro League of Legends games – who got killed, when towers fell, and so on. From all that detail, we focus just on the very beginning, taking a "snapshot" of the game right at the 12-minute mark. We pulled out key things like how many kills each team had, who knocked down the first tower, and who drew "first blood" (got the very first kill of the game).

EDA

Differences in kills, dragons, and first blood seemed the strongest hints about who was ahead. But even with these clues, winning and losing games often looked very similar in these numbers. This was the first sign of SISO.

Creating models

Based on EDA (non linear relationship, multiple data types, different scales), we decided to start with a simple decision tree model. It had 66.8% accuracy, next we trained random forest which had more stable results and 68% accuracy. Finally, we trained an AdaBoost model to improve the accuracy with boosting.

Model	Accuracy	Sensitivity	Specificity	Balanced accuracy	AUC
Decision Tree	0.6689	0.7508	0.5771	0.6639	0.713
Random Forest	0.6806	0.6890	0.6712	0.6801	0.744
AdaBoost.M1	0.6812	0.7245	0.6327	0.6786	0.747

All models plateaued at around 67%-68%, and this was the second sign of SISO. While the AdaBoost model had a bit higher overall accuracy and AUC, the random forest was much more balanced, and speed of training was orders of magnitude different.

Results

We built models that can predict the winner with a accuracy around 68% using early game info. The random forest was the most reliable one we built because it gave the best balance of correct predictions.

Reason for relatively poor accuracy of our models, is that the numbers we can gather from just the first 12 minutes using our current features aren't enough to predict the whole game. Early game leads are important in games' outcomes, they are not definite predictors. Big comebacks or strategic moves after early game can still change who wins the game.