Predicting Winning Teams in League of Legends Using Random Forest Classifier

Légrádi Bence

December 8, 2025

1 Abstract

This project aims to predict the winning team in League of Legends matches using a Random Forest Classifier. By analyzing various in-game statistics, we identify key features that influence match outcomes and evaluate the model's performance.

2 Introduction

League of Legends (LoL) is a popular multiplayer online battle arena (MOBA) game where two teams compete to destroy each other's base. Predicting the winning team based on in-game statistics can provide insights into game dynamics and strategies. This study utilizes a Random Forest Classifier to analyze match data and predict outcomes. The dataset was obtained from Kaggle [1].

3 Data Preprocessing

The dataset contains various features related to in-game statistics, such as kills, characters and objectives taken. The target variable is the winning team, encoded as 1 for Team 1 and 2 for Team 2. The id columns were removed, because those number do not contain any useful information for the model. First we look at some statistics about are data. As shown in Figure 1, the distribution of winning teams is relatively balanced, with Team 1 winning 51.2% of the matches and Team 2 winning 48.8%. This balance is crucial for training a fair model. We also examined the lenght of the games. Figure 2 shows the distribution of game lengths, which appears to be averagely around 30-40 minutes.

4 Model Training

We split the dataset into training and testing sets, using 80% of the data for training and 20% for testing. A Random Forest Classifier was trained on the

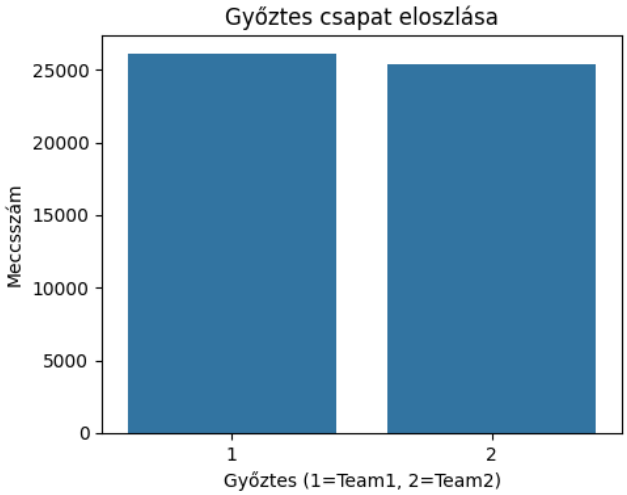

Figure 1: Winning Team Distribution

training set, and its performance was evaluated on the test set using accuracy as the primary metric. The model achieved an accuracy of 97% on the test set, indicating a strong ability to predict the winning team based on in-game statistics.

5 Feature Importance

To understand which features contributed most to the model's predictions, we analyzed the feature importance scores provided by the Random Forest Classifier. The top fifteen most important features were: As we can see in Figure 3, the total number of tower and inhibitor kills are highly influential in determining the winning team. It is not a surprise, as taking down these objectives is crucial for winning a match, because towers provide a base defense throughout the game, and inhibitors, when destroyed, allow the opposing team to spawn stronger minions that can push lanes more effectively. So when a team manages to destroy more towers and inhibitors, they gain a significant strategic advantage that often leads to victory. The other important thing is, we cannot see any champion picks in the most important features, which means that the outcome of the game is more influenced by in-game actions and strategies rather than the specific champions chosen by the players, which means the champions are balanced enough to not influence the outcome.

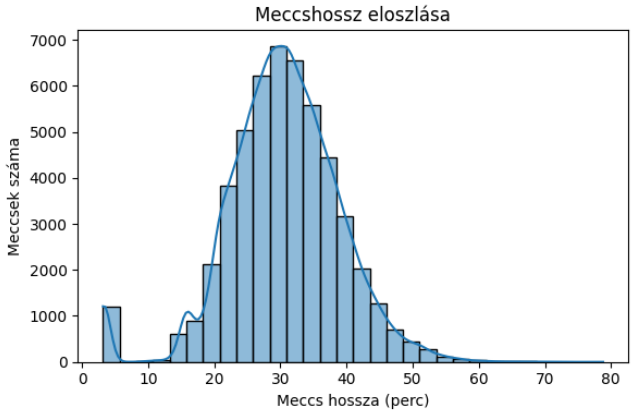

Figure 2: Game Length Distribution

6 Evaluation Metrics

To further evaluate the model's performance, we calculated additional metrics such as Confusion Matrix and ROC Curve. Figure 4 shows the confusion matrix, indicating that the model correctly predicted 1944 out of 2000 matches, with only 56 misclassifications. The ROC Curve in Figure 5 demonstrates the model's ability to distinguish between the two classes, with an Area Under the Curve (AUC) of 0.99, indicating excellent performance.

7 Conclusion

The Random Forest Classifier effectively predicts the winning team in League of Legends matches based on in-game statistics. Key features such as tower and inhibitor kills play a significant role in determining match outcomes. The model's high accuracy and strong evaluation metrics suggest its potential for practical applications in game analysis and strategy development. Future work could explore additional features and alternative modeling techniques to further enhance prediction accuracy.

References

[1] Mitchell J. (lol) league of legends ranked games. `https://www.kaggle.com/datasets/datasnaek/league-of-legends`, 2025. Accessed: 2025-11-18.

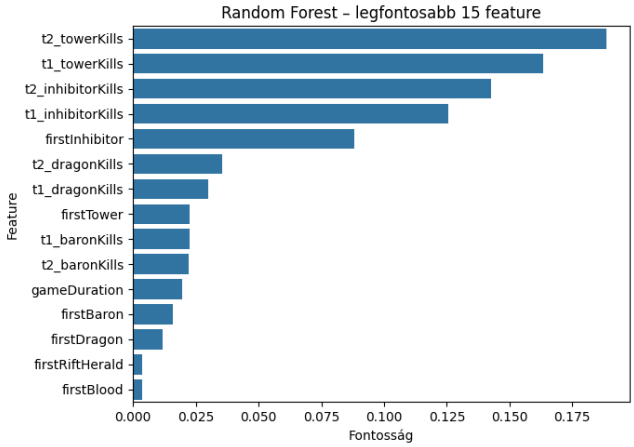

Figure 3: Top 15 Feature Importances

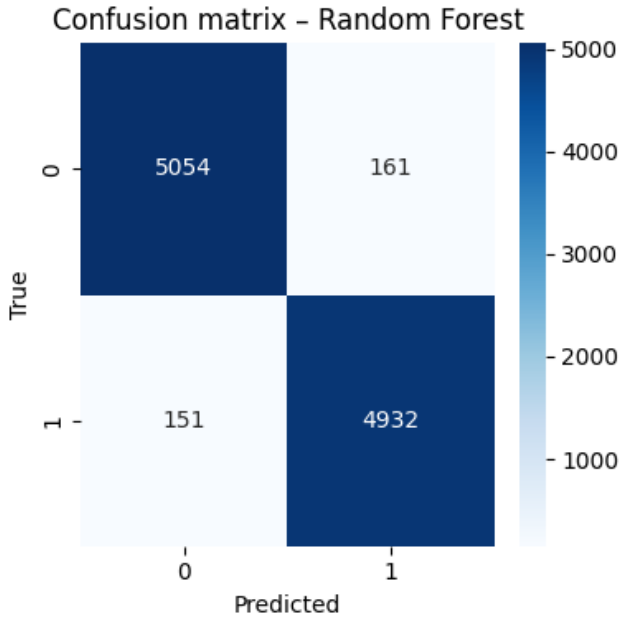

Figure 4: Confusion Matrix

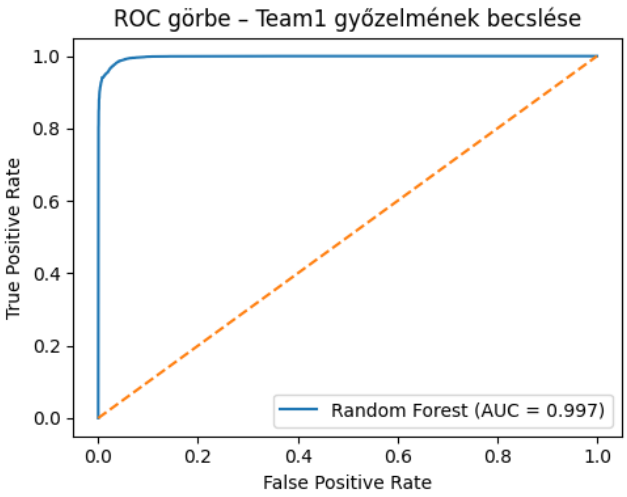

Figure 5: ROC Curve