Reinforcement Learning with Proximal Policy Optimization for Strategic Betting in Counter-Strike 2 Esports

Nathan Ho
Drexel University
Philadelphia, PA, USA
nlh55@drexel.edu

Matthew Protacio
Drexel University
Philadelphia, PA, USA
mp456@drexel.edu

Alexey Kuraev
Drexel University
Philadelphia, PA, USA
ak789@drexel.edu

Abstract

This project explores the application of Proximal Policy Optimization (PPO), a reinforcement learning algorithm, to develop an intelligent betting agent for esports matches. We evaluate the agent's performance and decision-making in a simulated betting environment using historical match data.

ACM Reference Format:
Nathan Ho, Matthew Protacio, and Alexey Kuraev. 2025. Reinforcement Learning with Proximal Policy Optimization for Strategic Betting in Counter-Strike 2 Esports. In . ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Sports betting is a widely practiced recreational activity in which individuals place wagers on specific outcomes of sporting events. Bet types range broadly, including predicting specific occurrences during a game or determining the ultimate winner of a match. This paper specifically explores *moneyline* bets, where wagers are placed solely on the final result of a contest.

Despite its popularity, sports betting remains underrepresented as a quantitative research domain, partly due to its association with gambling, which contributes to limited academic inquiry and systematic study. Predicting winners accurately poses considerable challenges, as match outcomes are inherently stochastic due to significant variability in team performance and individual player dynamics.

Effective betting strategies often hinge on identifying edges, which involve detecting discrepancies between bookmaker odds and bettors' valuations. Precisely computing match odds from fundamental analyses demands extensive computational resources. However, by considering established bookmakers' odds as a reliable proxy for the market's perceived fair value, we circumvent extensive calculations while still gaining actionable insights.

Applying these concepts to professional esports introduces unique complexities. State representation in video games can lead to state explosion due to numerous exogenous variables. Moreover, continual game updates and modifications create unstable environments where strategies effective in one period may become obsolete in the next.

Nonetheless, the esports title *Counter-Strike 2 (CS2)* offers specific advantages for quantitative modeling. Economic management significantly influences gameplay outcomes, with team bankroll serving as a critical predictive indicator. Within CS2, team economics are determined by several well-defined factors, including weapon and equipment purchases, income from winning rounds, and earnings from eliminating opponents. Matches typically span best-of-13 rounds, allowing temporal economic analysis across discrete intervals.

While raw economic indicators provide foundational insight, their direct application can suffer from excessive noise and limited predictive value. To enhance signal strength, we compute more sophisticated financial metrics, including teams' Return on Investment (ROI), implied probability derived from bookmaker odds, ROI based on betting odds, and Cost Per Kill (CPK). Interpreting a CS2 team as an evolving market asset allows us to aggregate performance metrics over time, offering valuable temporal context for predictive modeling.

This paper applies reinforcement learning—specifically, Proximal Policy Optimization (PPO)—to leverage these economic indicators for betting decisions. We further explore reward function formulations beyond simple correctness, investigating the impact of Expected Value (EV) calculations and the Kelly Criterion on betting efficacy. In doing so, we examine the limitations of traditional betting methodologies, exploring whether integrating financial and probabilistic principles yields improved betting outcomes in esports scenarios.

2 Methodology

2.1 Data Collection via Web Scraping

The dataset used to train the Proximal Policy Optimization (PPO) agent was constructed by scraping esports match data from two primary sources: *https://bo3.gg* for detailed Counter-Strike 2 (CS2) match statistics, and *https://www.oddsportal.com* for corresponding betting odds. The scraping process was implemented using Python scripts utilizing the libraries *Playwright* and *BeautifulSoup.*

Match Data (bo3.gg): Utilized *Playwright* for automated browser control to navigate and load dynamically-rendered web pages, ensuring complete data retrieval. Extracted structured JSON data containing round-level statistics, including economic state, round results, map data, player statistics, and team performance metrics.

Betting Odds Data (oddsportal.com): Leveraged *Playwright* to handle interactive and dynamically updated odds information, which required simulating user interaction to reveal hidden or

Conference'17, Washington, DC, USA
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

paginated odds. Employed *BeautifulSoup* to parse HTML content efficiently, extracting structured odds data including opening, closing, and intermediate odds offered by various bookmakers.

The scraping scripts were developed with careful adherence to ethical web-scraping practices, employing randomized delays and respectful request frequencies to avoid overwhelming the servers. Additionally, extracted data was systematically stored in JSON files for ease of processing and reproducibility. The final compiled dataset provided a comprehensive representation of match states and betting odds, enabling robust feature extraction for the PPO model training pipeline. Scraped data was later analyzed for feature distribution and building a dictionary of winners for all games found.

2.2 Feature Engineering

After web scraping, a large collection of JSON formatted match rounds is accumulated.

An example of the raw JSON match data structure is shown below:

The raw JSON match data is preprocessed to serve as a state input for the PPO model. Each round in a game is transformed into a normalized *PyTorch* tensor. This ensures efficient and stable model training. As mentioned above, we aimed to convert raw team economic status into meaningful signals. These economic metrics help to capture team performance dynamics. These are defined as follows:

Delta Econ for Both Teams - Captures the absolute economic change between the starting and ending bankroll of a team within a round:

$$\Delta\text{Econ} = \text{Final Economic Value} - \text{Initial Economic Value} \tag{1}$$

ROI based on Team Econ - Measures the relative financial gain or loss by comparing final economic status to initial investment:

$$\text{ROI} = \frac{\text{Final Economic Value} - \text{Initial Economic Value}}{\text{Initial Economic Value}} \tag{2}$$

ROI based on Odds (Odds ROI) evaluates the expected profitability relative to bookmaker odds:

$$\text{Odds ROI} = (\text{Decimal Odds} \times \text{Win Probability}) - 1 \tag{3}$$

The **Implied Probability from Odds** represents bookmakers' implicit estimation of event outcomes:

$$\text{Implied Probability} = \frac{1}{\text{Decimal Odds}} \tag{4}$$

Finally, **Cost Per Kill (CPK)** quantifies a team's economic efficiency regarding combat effectiveness:

$$\text{CPK} = \frac{\text{Economic Investment per Round}}{\text{Number of Kills per Round}} \tag{5}$$

The raw JSON match data obtained required preprocessing to serve as input for the Proximal Policy Optimization (PPO) model. Each round in a game was transformed into normalized PyTorch tensors, ensuring efficient and stable model training. Numerical features, such as team economic status, round outcomes, and individual player performance statistics, were normalized using min-max scaling or standardization to ensure consistency across varying scales.

Features were selected based on their predictive value regarding match outcomes and their ability to encapsulate meaningful temporal and economic context. Specifically, selected features included team bankroll, Return on Investment (ROI), implied probability derived from betting odds, ROI based on odds, and Cost Per Kill (CPK). This targeted selection aimed to reduce dimensionality, improve signal-to-noise ratios, and enhance the model's capacity to generalize from historical data to future betting scenarios.