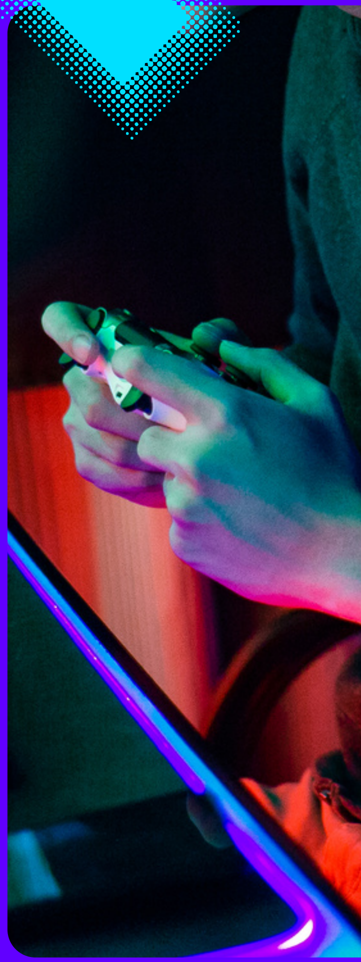
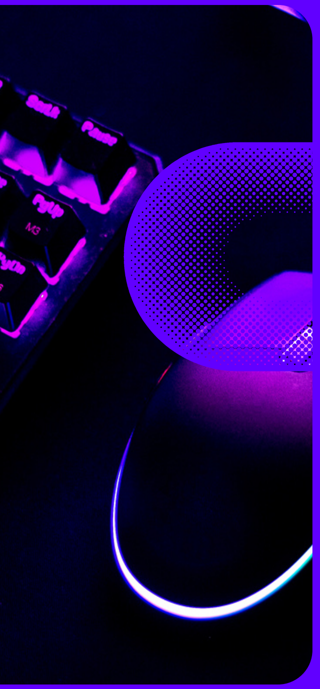
League of Legends Winning Team Prediction





Wenya Cai, Hao Duong, Vivi Li,
Alisa Liu, Seunghyun Park, Xinran Yu

TABLE OF CONTENTS

TEAM MEMBERS

Hao Duong

Business Analyst & Strategist

GitHub ID: haohaoduong

Vivi Li

Business Analyst & PM

GitHub ID: viviliyatong

Seunghyun Park

Data Scientist (Predictive Model)

GitHub ID: SeunghyunParkk

Xinran Yu

Data Scientist (Predictive Model)

GitHub ID: xinranyu228

Wenya Cai

Data Scientist (Causal Analysis)

GitHub ID: WenyaCai

Alisa Liu

Data Scientist (Causal Analysis)

GitHub ID: 18ZL107

GitHub repository: McGill-MMA-EnterpriseAnalytics/League-of-Legends

LEAGUE OF LEGENDS INTRO

In the game, two teams (blue and red) of five players battle in player-versus-player combat, each team occupying and defending their half of the map. Each of the ten players controls a character (champion) with unique abilities.

During a match, the characters become more powerful by collecting experience points, earning gold, and purchasing items to defeat the opposing team. In League's main mode, a team wins by pushing through to the enemy base and destroying their "Nexus".

champions

Nexus (blue)

INDUSTRY OVERVIEW

eSports is reportedly a $1.45bn industry (in 2020) and is predicted to grow to $6.75bn by 2030

eSport teams make the majority of their money from **sponsorships**, **merchandising**, and **league payments (2.23 million U.S. dollars in 2023 league)**

PROBLEM STATEMENT

Objective: to demonstrate the **value** and **applicability** of predictive analytics in competitive Esports, leveraging a unique dataset to predict the outcome of a game within the first 10 minutes

1

What are some strategies for gaining advantages, winning just 10 mins into the game?

2

How can brands assess team's performance for sponshorships ?

3

How do company calculate their ROI based on teams' performance indicators ?

DATASET OVERVIEW

This dataset is recorded during the first 10min of the game.

There are **19 features per team** (38 in total) collected during 10min in-game. This includes kills, deaths, gold earned, experience points collected, champion level, etc.

For example,
blueFirstBlood: First kill of the game. 1 if the blue team did the first kill, 0 otherwise.
blueTotalJungleMinionsKilled: Blue team total jungle monsters killed.
blueTowersDestroyed: Number of structures destroyed by the blue team (towers...).
(same for red team)

blueWins is the target variable. A value of 1 means the blue team has won. 0 otherwise. (balanced)

PREDICTIVE MODEL

EDA & Preprocessing

- *Variables Distribution*
- *Variables Collinearity*
- *Remove unhelpful variables*
- *Standardization*
- *Train, test split*

Feature Engineering

New Features:

- *Helpful*
- *JunglePercentage*
- *WardsRemaining*

Modelling

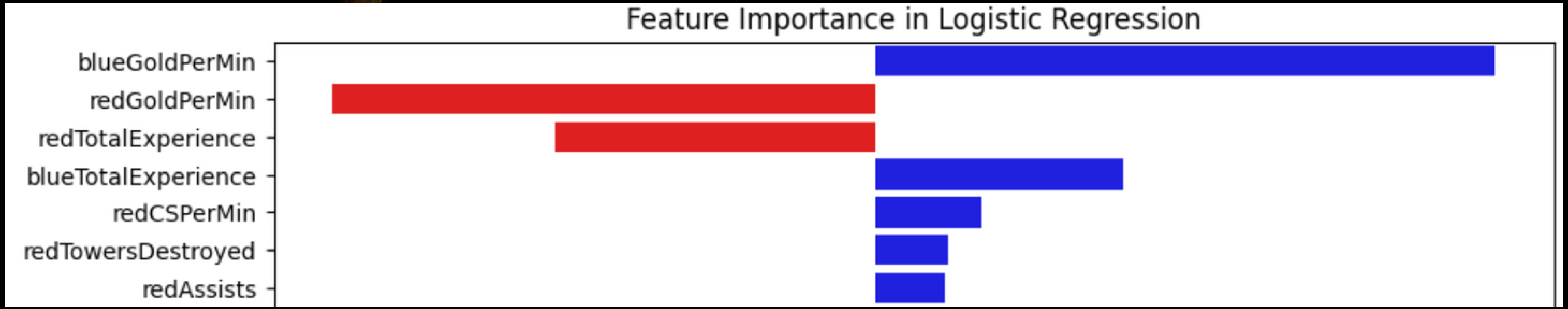
- *Logistic Regression (73.3%)*
- *Decision Tree (71%)*
- *Random Forest (71.8%)*
- *LightGBM (72.2%)*
- *XGBoost (71.8%)*
- *Gradient Boosting Classifier (72.6%)*

Model Tuning

- *C: 100*
- *Penalty: l2*

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.73	0.73	983
1	0.73	0.74	0.73	993
accuracy			0.73	1976
macro avg	0.73	0.73	0.73	1976
weighted avg	0.73	0.73	0.73	1976



From Predictions to Causal Inference

Machine Learning

$E[Y|X]$

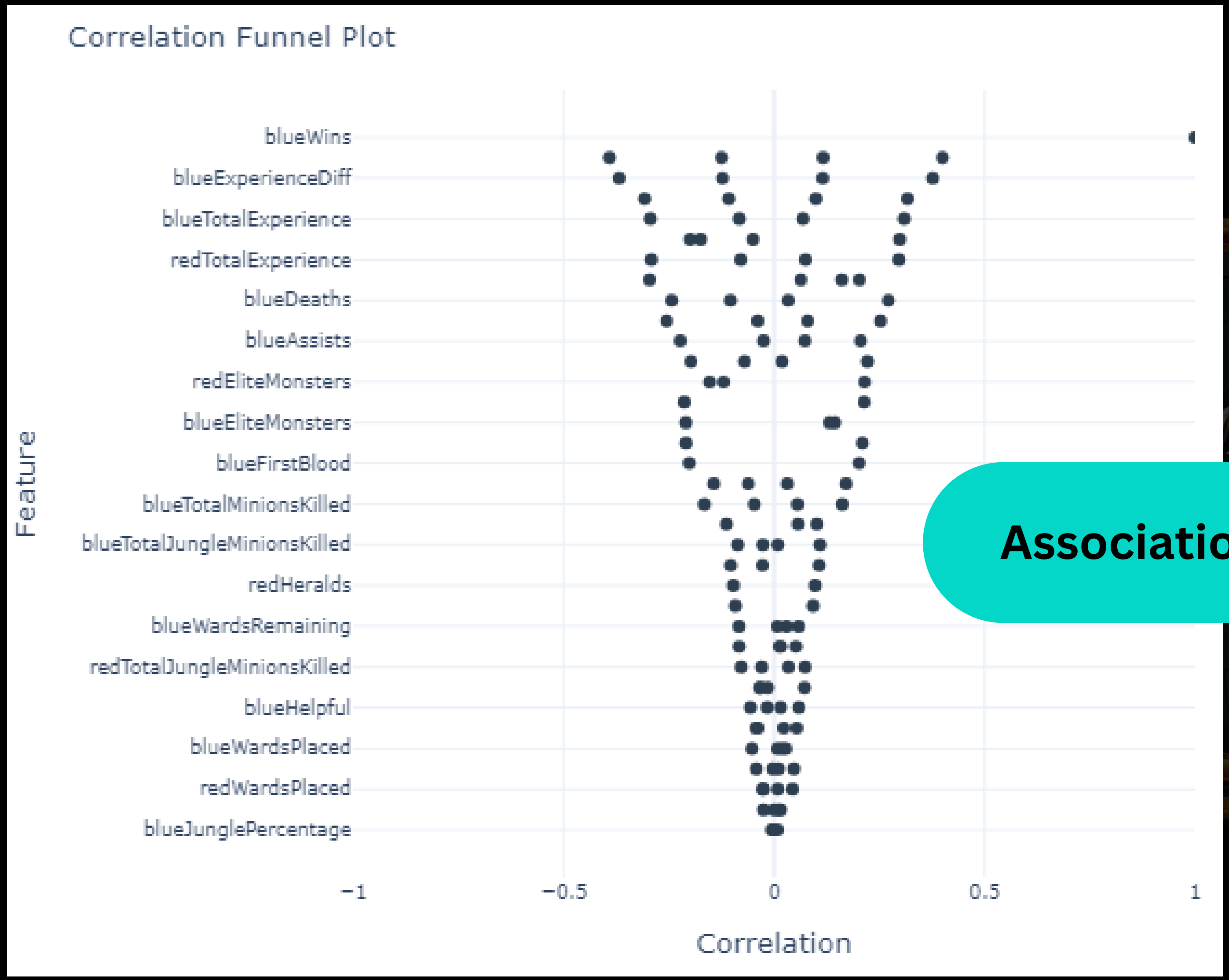
" Given these actions we take in early games, what is our chance of winning?"

Causal Inference

$E[Y|X,T]$

" Can taking this action lead to winning?"

Use Case Statement



- *Gold Difference has highest correlations with winning, especially when the difference is high(corr 0.4)*

Association → Causation

"Does having more gold give you a competitive edge?"

Parameter

- Treatment: high/low gold diff
- outcome: Winning

Modeling Selection - Xlearner()

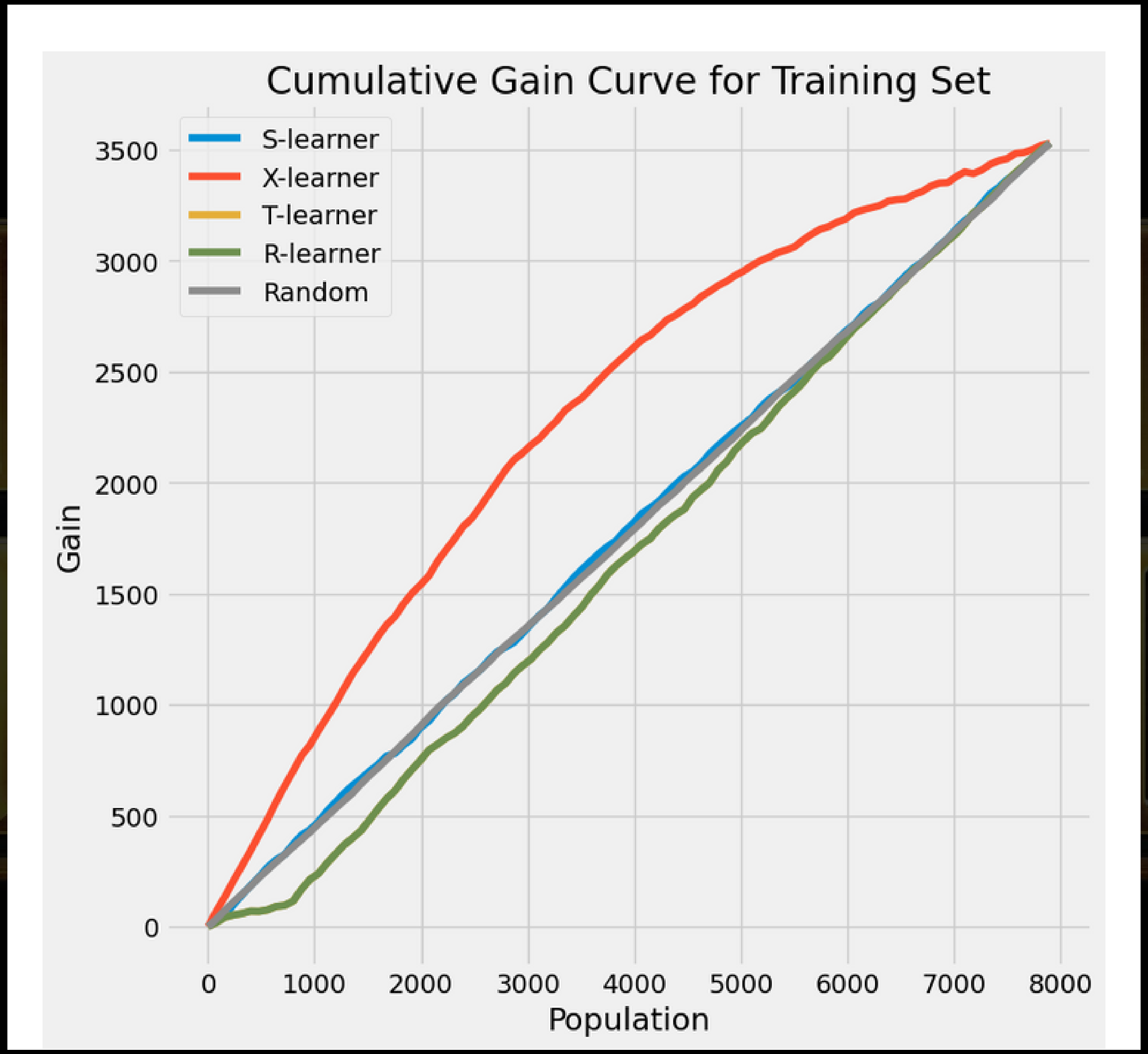
ATE result

- *S learner: 0.002*
- ***X learner: 0.77***
- *T learner: 0.12*
- *R learner: 0.12*

ATE on Train/Test Set

- *Train: 0.76*
- *Test: 0.79*

Cumulative gain curve



Confidence Interval

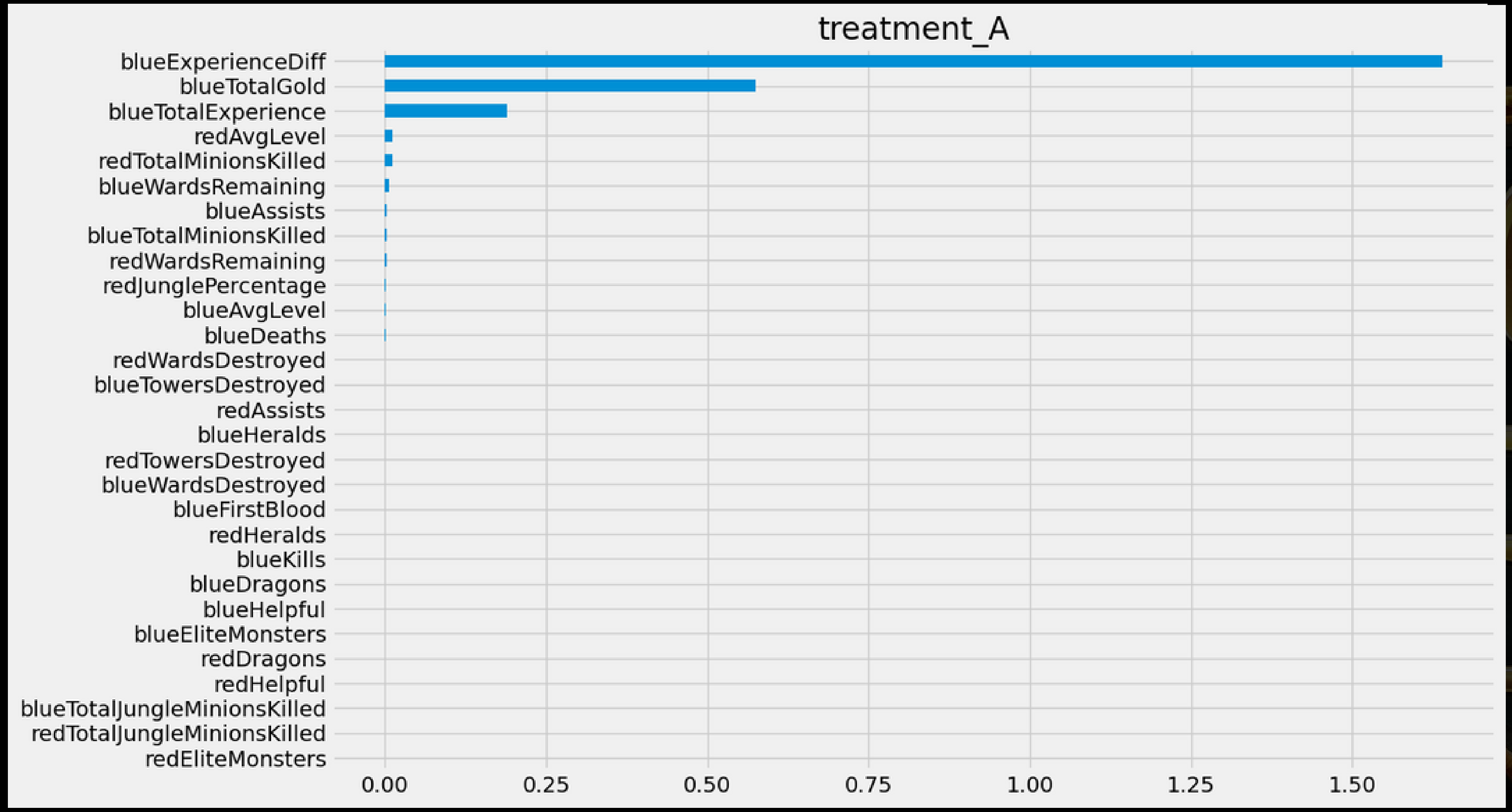
```
                     Learner Average Treatment Effect        Lower Bound CI        Upper Bound CI
                   S-Learner  [0.0005060728744939271]                   NaN                   NaN
X-Learner (Propensity Score)     [0.7935901981257701]  [0.7486730133604523]  [0.8385073828910888]
                   X-Learner     [0.7935901981257701]  [0.7486730133604523]  [0.8385073828910888]
                   T-Learner    [0.13006072874493926] [0.08449630017732056] [0.17562515731255796]
                   R-Learner    [0.13006072874493926] [0.08449630017732056] [0.17562515731255796]
```

Insights

- when choosing a model, we should consider both the **ATE** and the model's proficiency to **rank individuals for targeting**.
- X-learners perform well when there are heterogeneous treatment effects. Since the goal is to understand the overall impact of a treatment, a model with a higher ATE might be more informative.
- narrow CI -> precision

Feature Importance

High importance don't just correlate with success; they interact with high gold difference, reinforcing its effect on winning.

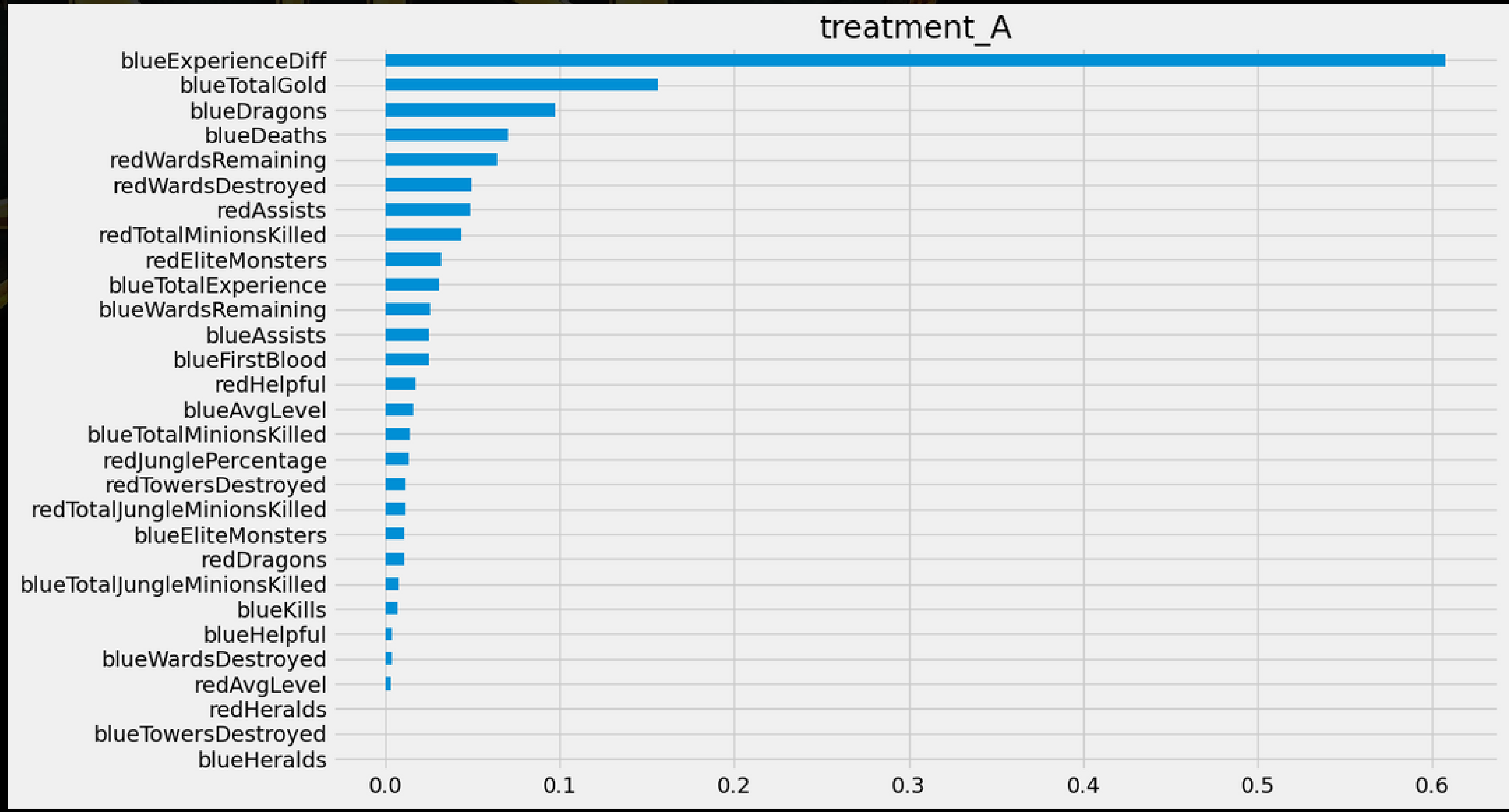


Top 3 features with Logreg:

- ***ExperienceDifference***
- ***TotalGold***
- *TotalExperience*

Top 3 features with GBM(2nd best):

- ***ExperienceDifference***
- ***TotalGold***
- *Dragons*



A Simple DAG

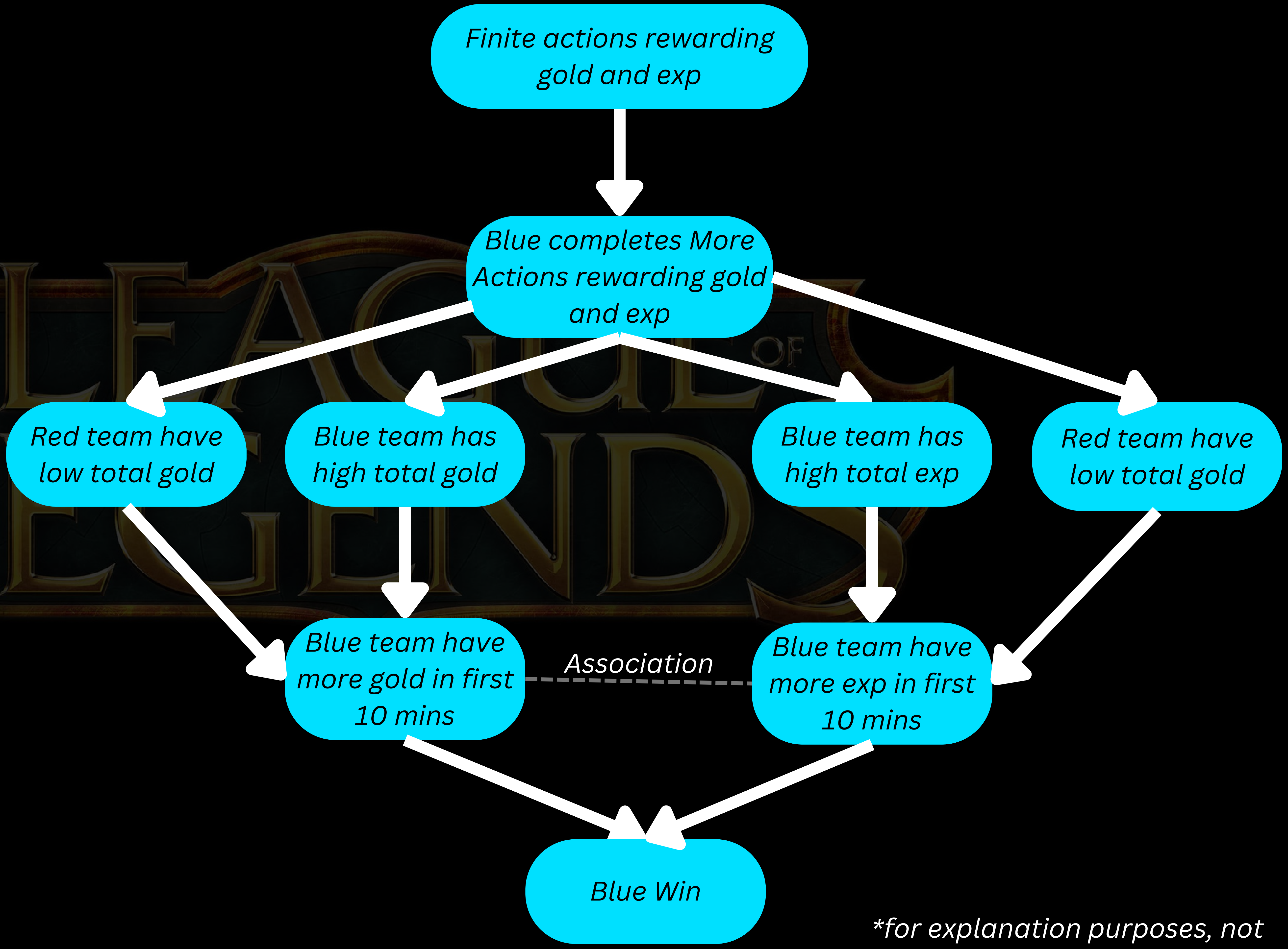
Constraints

 # actions Blue can complete
+
 # actions Red can complete
is finite

Insights

High interaction between gold and exp measures makes sense, as a lot of actions reward both. This shows:

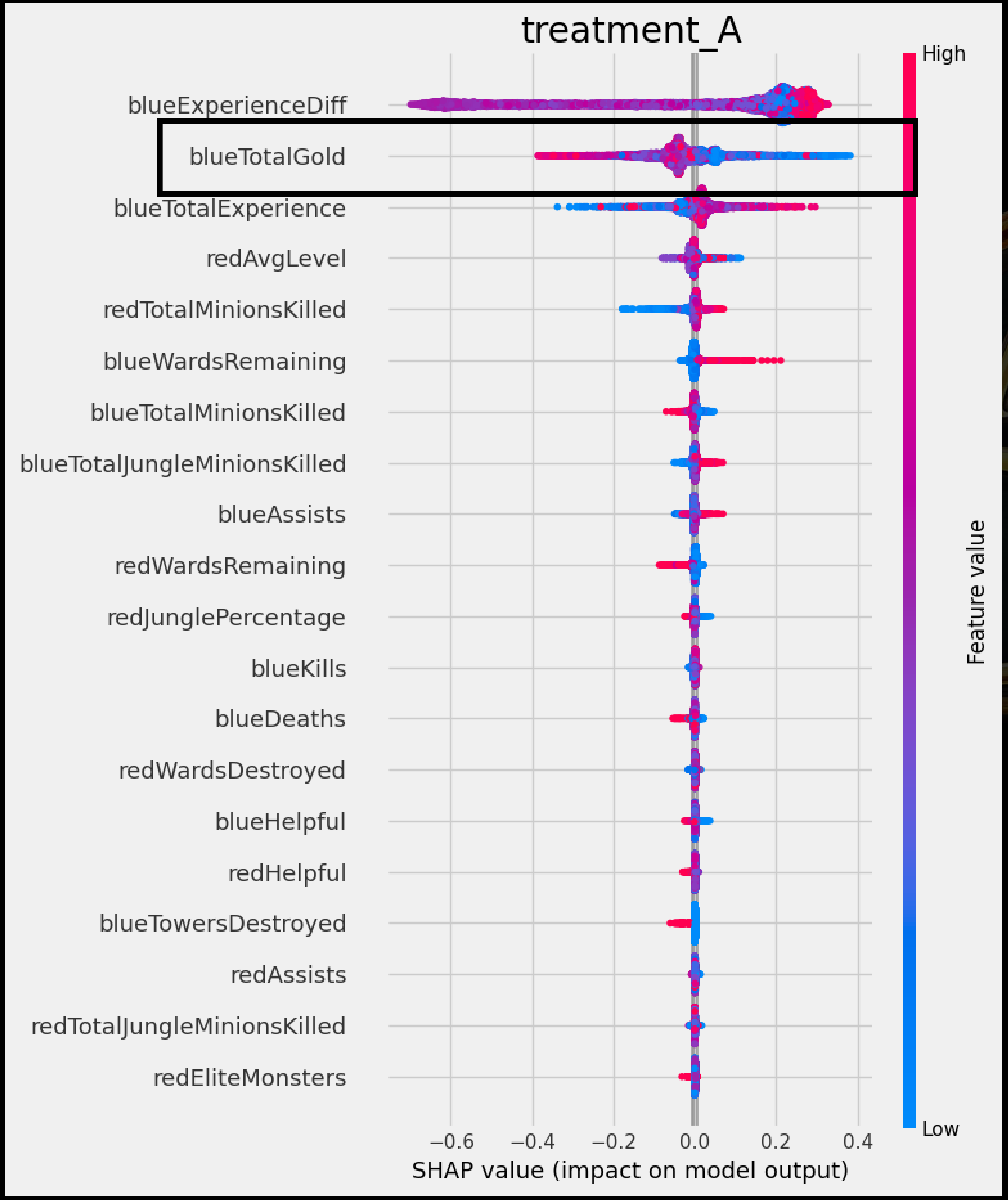
1. Association does not equal Causation
2. Completing *More* of those actions than enemy does lead to higher chance of winning



*for explanation purposes, not traditional DAG

Shapley Values

For teams with high gold difference, High total gold can have a negative impact on winning?



Main Actions Rewarding Gold in first 10 mins

Engage in fights: 1 kill = 300 gold

Farming minions: ~18 gold

1 kill = 16.7 minions

while seems counterintuitive, teams with a substantial gold (high difference + high total) could implies that this team choose to engage more fights than farming minions since killing enemies rewards way more gold.

This agressive strategy apperantly does not end well in diamond rank plays, as skilled players tend to play defensive in early game.

Insights Summary

1

Defensive Playstyle

+

2

MAINTAIN COMPETITIVE GOLD AND EXPERIENCE LEVEL

=

3

Win!

BUSINESS IMPLICATION

1. **For professional teams:**

   Insights for teams to strategize (team members selection, game strategies, ...).

2. **For Esports companies owning the professional teams:**

   Insights for the companies to be able to evaluate their team performance, calculating chances of winning and to evaluate to recruiting members.

3. **For sponsoring companies:**

   With the model providing information about team's performance, to have a competitive sponsoring offer, or updating their sponsoring offers, since the industry is about popularity and performance in matches.

NEXT STEPS

1. *Gather data and finetuning model for whole game/ tournament*
2. *Pipeline and Production*
3. *Potential Expansion to other Esport game*

LIMITATIONS

1

Many heterogeneous treatment effect estimation methods are theoretically valid only when all potential confounders are observed. These methods attempt to approximate the gold standard of RCT. We refer to them as causal estimation methods under unfoundedness. It’s important to note that any attempt to use these methods without observing potential confounders can reduce, but not eliminate, bias relative to raw correlations.

2

Unless all potential confounding factors are included in an analysis (which is unlikely to be achievable with most real-world data-sets), adding control variables to a model in many circumstances can make estimated effects of the variable(s) of interest to the researcher on the dependent variable less accurate. (Richard York (2018) Control variables and causal inference: a question of balance, International Journal of Social Research Methodology, 21:6, 675-684, DOI: 10.1080/13645579.2018.1468730)

Thank You