Final Project – Attempt to Predict the Dota 2 International Tournament Outcomes

Course: EN 605.662 (Data Visualization)

Professor: J. Caban

Date: 8/11/2025

By: Mike Xie

Table of Contents

Abstract

This paper shows an attempt to build visualizations that predict who will win the Dota 2 TI 2025 tournament. Previous research on Dota 2 has shown that the draft influences the match results, and the data shows that it is common for a team's hero pool to influences the draft. If we model those two links (*a team's hero pool influence their drafting decision, and their drafting decisions influence the outcome of a match*) , we can predict a winner before the draft even starts and then estimate the likelihood of each team to win the whole tournament. I combine data visualizations, a machine learning model, and Monte Carlo simulations to estimate team win probabilities. The paper starts by defining key gameplay terms, then maps team hero pools, trains a model, simplifies it for speed, and runs Monte Carlo simulations to get tournament odds. Future work should use larger datasets and add features like current player rosters and player-specific hero pools as well as using better simulation methods.

Disclaimer

This paper presents model-based predictions for TI 2025, which had not occurred at the time of writing. The results are estimates from the models and should not be taken as guarantees of future outcomes.

The tournament's structure, format, or other details may change between writing and the actual event. Because this project was completed beforehand, some specifics in the paper may not match what ultimately happens.

Team names can also change. For example, Parivision, a direct invite to TI 2025, appears in some places as PVISION. This may be because "pari" means wager in Russian, and the name in the TI 2025 Tournament could show up as "Parivision" or "Pvision". Many sites have not updated yet, and the dataset I use still lists "Parivision" as the team name. For consistency, this paper, datasets, machine learning models and the visualizations will refer to that team as only Parivision.

I do not have any conflict of interests in the outcomes of the Dota 2 TI 2025.

Chapter 1: Introduction

This paper shows an attempt to build a visualization to help explore which team will win the Dota 2 TI 2025. Dota 2 is a multiplayer online battle arena (MOBA) developed and published by Valve. Every year a massive tournament is set up for professional teams to take part in to win, called the Interantional. This years international will happen on September 11 – 14 in Germany's Hamburg with 16 teams competing [1].

The game of Dota 2 is played in matches between two teams of five players. A game will have a drafting phase, and then the actual gameplay. The game is won if one team is able to defend their base while destroying the enemy team's base. Each player controls one of 126 heroes, which is selected during the drafting phase. Once a player chooses a hero, other player's cannot select that same hero for the same match. It is already well documented that teams that are able to draft heroes that complement each other, while also countering the enemy heroes, are more likely to win [2]. Meaning that one could create a machine learning model to predict which team would win based on the heroes each team picked.

Many players have a set hero pool and a strategy to execute before entering the drafting phase. Meaning that for some teams, they will go into the drafting phase with a draft order and team composition already in mind. And if that is true, then one could predict which team would win before the drafting phase even starts.

By combining data visualization and machine learning models, and creating a monte carlo simulation, this project will attempt to predict which team will win each match, to then predict which team will win the TI 2025 tournament overall.

To see the web app data visualization, and to mess around with the monte carlo simulation visualization, visit this link: https://data-visualization-final-project-production-8917.up.railway.app/

To see the datasets, ml models, python google colab note books and the code for the web app application, visit this github link, https://github.com/mikexie360/data-visualization-final-project/tree/main

Drafting Phase

A good draft means picking heroes that work well together and that players are comfortable on, while banning heroes that threaten the plan or that the opponent is particularly strong with.

In practice, teams pick from their own comfort heroes and synergies, and they ban counters, meta threats, and the opponent’s signature heroes.

For this paper, I do not model the draft. Instead, I estimate what each team is likely to pick from its hero pool and use pre-draft information to make predictions. Prior work shows the draft is a big factor in Dota 2 and is strongly linked to match outcomes [3], so I expand on it by predicting what heroes might be drafted by looking at the team’s hero pool.

Hero Pool

Every team has a hero pool, which is the set of heroes they consistently use. Some teams have deep, flexible pools of many heroes that they can play. Others keep a smaller set that fits a specific strategy. Because each team drafts to match its playstyle, hero pools end up being unique.

Playstyles can counter each other in Dota 2, which means hero pools can counter hero pools. A team with a narrow but well-targeted pool can beat a “stronger” team if its usual picks line up well against the opponent’s comfort heroes. In other words, team vs team matchup can flip expectations based on hero pools.

Hero pools also exist at the player level and change over a career. Some players are famous for a few signature heroes. Others are solid on many heroes. Professional teams will reflect that since teams are made up of 5 players.

If a team's hero pool shapes the draft, and the draft shapes the match, then knowing what a team has played before lets us estimate results without predicting every draft choice in real time. That idea is the basis for the modeling in this paper.

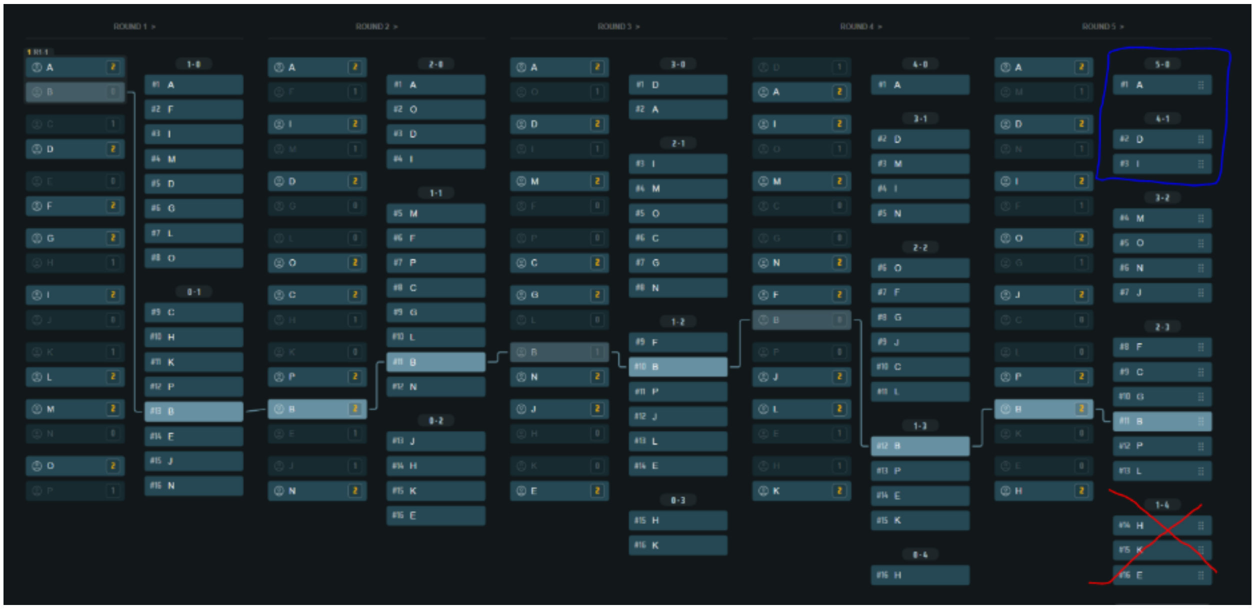
Dota 2 TI 2025

The International (TI) is Dota 2's most prestigious annual championship tournament. The International 2025 will take place in Hamburg, Germany on September 14, 2025. And this is what we will try to make predictions on. There will be 16 different teams and three stages. Some teams are directly invited and guaranteed a spot to compete, other teams have to qualify in their region to complete. The three stages are swiss-system, elimination, and double-elimination bracket [4].

Swiss Style Tournament Format, Variations and Changes

In order for the tournament organizers to make every match matter, they have decided to match teams in the swiss system against teams that are similar in strength. A normal swiss system might match the strongest teams against the weakest teams, to make sure that the strongest teams survive to the finals, and to also reward teams to do well in the swiss tournament. There have been cases where if a tournament is formatted unfairly for winning teams, some teams might try and lose one match on purpose to increase their odds of winning the overall tournament. Such as that one instance, where in 2012, several badminton teams were losing matches on purpose, to avoid stronger teams to increase their overal chance of winning the tournament [5]. Meaning that because of the structure and overall format that valve chose to eliminate teams and to match up teams, some teams may have an advantage from the seeding.

Overall, the way Valve structured the swiss format is 5 rounds over three days. All series are best of three, each round teams play others with the same record. Teams with 4 wins auto advance to the next playoffs, teams with 4 losses are eliminated, the rest ener into the elimination round.



The image above shows what the Swiss tournament style may look like. 5 rounds, with teams advancing if they receive 4 wins, and teams being eliminated when they receive 4 losses.

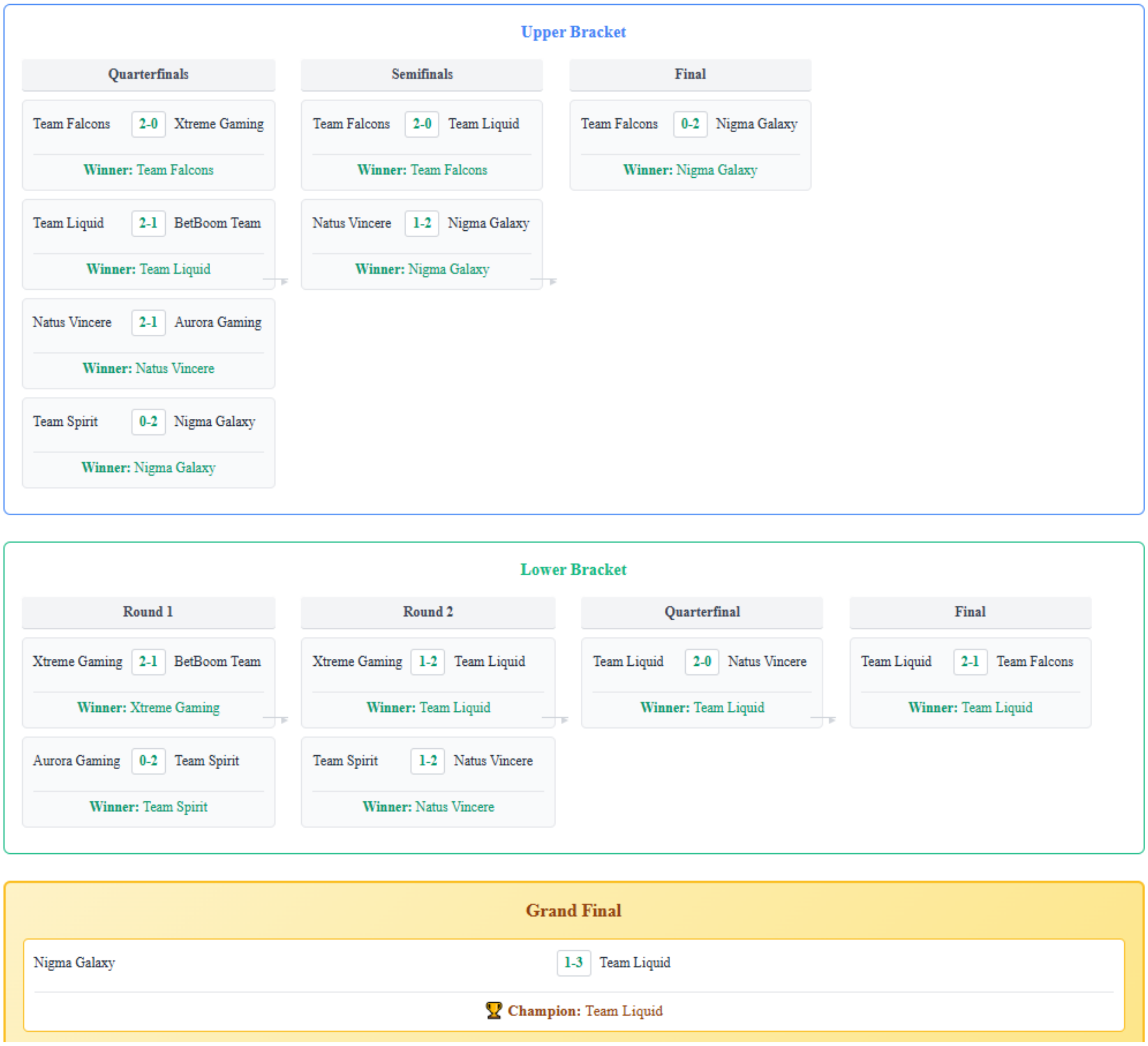
Elimination Round

After the Swiss stage, the three teams with four wins qualify directly for the playoffs, and the three teams with four losses are eliminated. The remaining five teams are seeded from weakest to strongest and play best of three series. Winners advance to the playoffs, and losers are out.

Playoffs and the Double Elimination Format

8 teams enter in the upper bracket. If a team in the upper bracket looses a match, they enter into the lower bracket. If a team in the lower bracket looses, then they are eliminated. Teams are seeded and matched up from strongest to weakest seeds at the start.

This double elimination format means teams are allowed to lose one match without being eliminated, and being eliminated requires to lose two matches. This reduces the variance of a strong team loosing to early after one bad match.



Upper Bracket

Quarterfinals

Team Falcons 2-0 Xtreme Gaming
Winner: Team Falcons

Team Liquid 2-1 BetBoom Team
Winner: Team Liquid

Natus Vincere 2-1 Aurora Gaming
Winner: Natus Vincere

Team Spirit 0-2 Nigma Galaxy
Winner: Nigma Galaxy

Semifinals

Team Falcons 2-0 Team Liquid
Winner: Team Falcons

Natus Vincere 1-2 Nigma Galaxy
Winner: Nigma Galaxy

Final

Team Falcons 0-2 Nigma Galaxy
Winner: Nigma Galaxy

Lower Bracket

Round 1

Xtreme Gaming 2-1 BetBoom Team
Winner: Xtreme Gaming

Aurora Gaming 0-2 Team Spirit
Winner: Team Spirit

Round 2

Xtreme Gaming 1-2 Team Liquid
Winner: Team Liquid

Team Spirit 1-2 Natus Vincere
Winner: Natus Vincere

Quarterfinal

Team Liquid 2-0 Natus Vincere
Winner: Team Liquid

Final

Team Liquid 2-1 Team Falcons
Winner: Team Liquid

Grand Final

Nigma Galaxy 1-3 Team Liquid
🏆 Champion: Team Liquid

The image above shows what the double elimination tournament might look like. 8 teams enter in the brackets with the strongest teams facing off the weakest teams. Teams can lose once, and the Grand Final match winner is the TI 2025 winner.

Chapter 2: Background

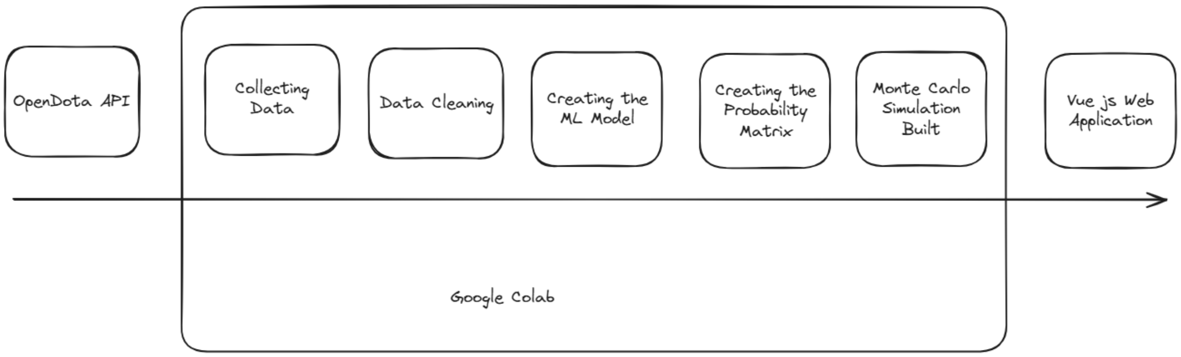
Most of the research into predicting matches results was done by using a machine learning model to learn from the drafting phase. Such as using a linear regression to predict the best heroes for the drafting phase. https://old.fruct.org/publications/fruct24/files/Por.pdf There hasn't been much done outside of the drafting phase to predict overall tournament outcomes.

There are other research in visualizing the most optimal draft phase picks and bans for video games, and giving insight. They used an interactive application to help users better understand the draft phase, and insights of how the probabilities will change based on the hero choices [6]. Some ideas of this will be borrowed, such as the interactive visualizations, but it will be trying to show how the structure of the tournament works at a high level, rather than showing how the match result probabilities change during the drafting phase.

There has been research on predicting tournament outcomes through a mathematical model using pairwise matrix to calculate the probabilities without running a monte carlo simulation [7]. However, this was only done on various single elimination and double elimination brackets, and was not done on a swiss system. I elected to use a monte carlo simulation, so that one could possibly see different tournament runs and see how the teams move up the tournaments by analyzing a specific tournament run.

Chapter 3: Approach

This image shows what steps were taken to create the visualization.



Collecting Data Through OpenDota 2 API

OpenDota API, owned by OpenAI, allows one to collect data through a REST API interace. One can collect team information, a team's overall rating, match history, and what heroes a team has played as well as what heroes that team has won with.

OpenDota API, also allows users to execute sql statements directly throught the API interface [8]. Allowing users to see the exact schema of the tables in the database, create joins to get the exact data that they need. By doing this I was able to get a list of matches associated with a team through a join clause. Which was much faster than doing it through API calls and then filtering data through python.

A thing to be worried about is the network latency when doing many API calls iteratively, such as iterating through pages to collect a large collection of data. When I had to collect data of many matches through multiple API calls, it took me around an hour to collect the data that was needed.

After collecting the datasets from OpenDota API, the data would then have to be analyzed and cleaned through python with Google Colab. Some datasets would be joined together to be flattened for the machine learning model.

Creating the Model

The machine learning model needed to learn from the hero pool of each team, and also how well that team do against another team with a different hero pool. So the dataset that trained the model required the match results, the 2 teams that took part in the match, and the hero pool of the two teams. This resulted in a very large dataset of around 750 features and over 18,000 rows.

Some features were not used, such as the team's overall rating, what happened during or in the match, such as the score, or duration. The team's overall rating was not used, as that feature would have overpowered the team's hero pool features. Meaning that the ML model would have relied too heavily on a specific feature, instead of the overall hero pool of the team. The score and duration would not have mattered as we are trying to predict the outcome of a match before the drafting phase, and including the score or duration of the match may have caused leakage.

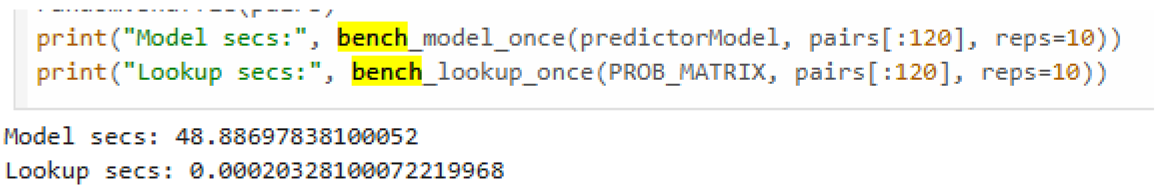
The machine learning model that was chosen was, HistGradientBoostingClassifier from sci-kit learn. As that ml model is great for large datasets. XGBoost was tried, but it was found to be overfitting the training data, and gave a suspiciously high accuracy score. The final benchmarks of accuracy was the following.

```
Train size: 15106 | Test size: 3777 | Features: 751

Accuracy: 0.7180 | ROC AUC: 0.8037 | LogLoss: 0.5325

Majority-class Accuracy baseline: 0.5234
```

Distilling the Model into a Probability Matrix with the 16 Teams

The problem with running machine learning models in production is that they are computationally expensive. It is always a good idea to distill or dumb it down to a simpler model, so that it runs faster with less compute power; however the downside is that the distilled model may have less predictive power or use less parameters. But the cost may sometimes be worth it.

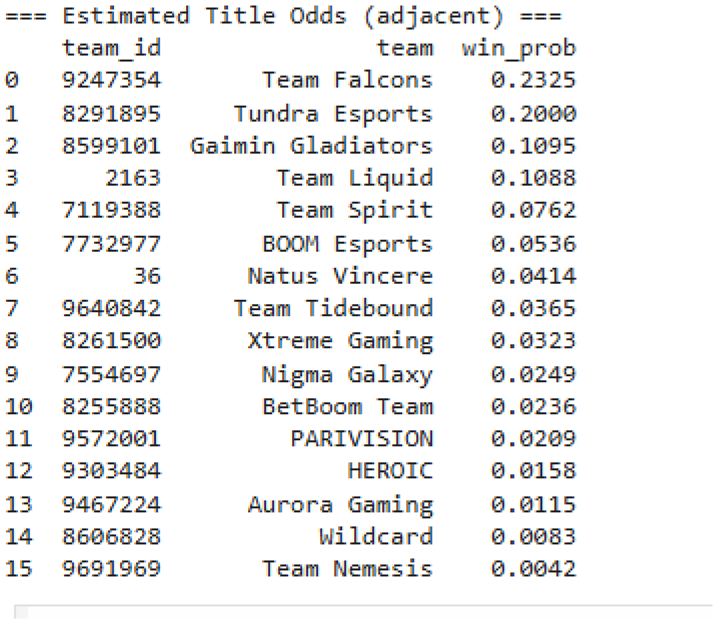
I managed to distill the model down into a simple probability matrix for the 16 teams that will participate in the TI 2025 tournament. Meaning that it is essentially a 16 by 16 look up table, meaning that the answers are just memorized and cached. This lead to an over 20,000 times improvement for my requirements.



```
print("Model secs:", bench_model_once(predictorModel, pairs[:120], reps=10))
print("Lookup secs:", bench_lookup_once(PROB_MATRIX, pairs[:120], reps=10))

Model secs: 48.88697838100052
Lookup secs: 0.00020328100072219968
```

Creating and Running a Monte Carlo Simulation to Predict Results

After using the ML model’s answer for the 16 qualified TI 2025 teams and turning it into a probability matrix, we can now run Monte Carlo simulations at incredible speeds. Creating the Monte Carlo simulation required creating the tournaments format and structure into python code that can be ran over and over again to show a statistical outcome. Giving results that showed teams that were directly qualified a higher chance of winning the tournament over the regional teams.



```
=== Estimated Title Odds (adjacent) ===
    team_id               team  win_prob
0   9247354       Team Falcons    0.2325
1   8291895     Tundra Esports    0.2000
2   8599101  Gaimin Gladiators    0.1095
3      2163        Team Liquid    0.1088
4   7119388        Team Spirit    0.0762
5   7732977       BOOM Esports    0.0536
6        36      Natus Vincere    0.0414
7   9640842     Team Tidebound    0.0365
8   8261500      Xtreme Gaming    0.0323
9   7554697       Nigma Galaxy    0.0249
10  8255888       BetBoom Team    0.0236
11  9572001         PARIVISION    0.0209
12  9303484             HEROIC    0.0158
13  9467224      Aurora Gaming    0.0115
14  8606828           Wildcard    0.0083
15  9691969       Team Nemesis    0.0042
```

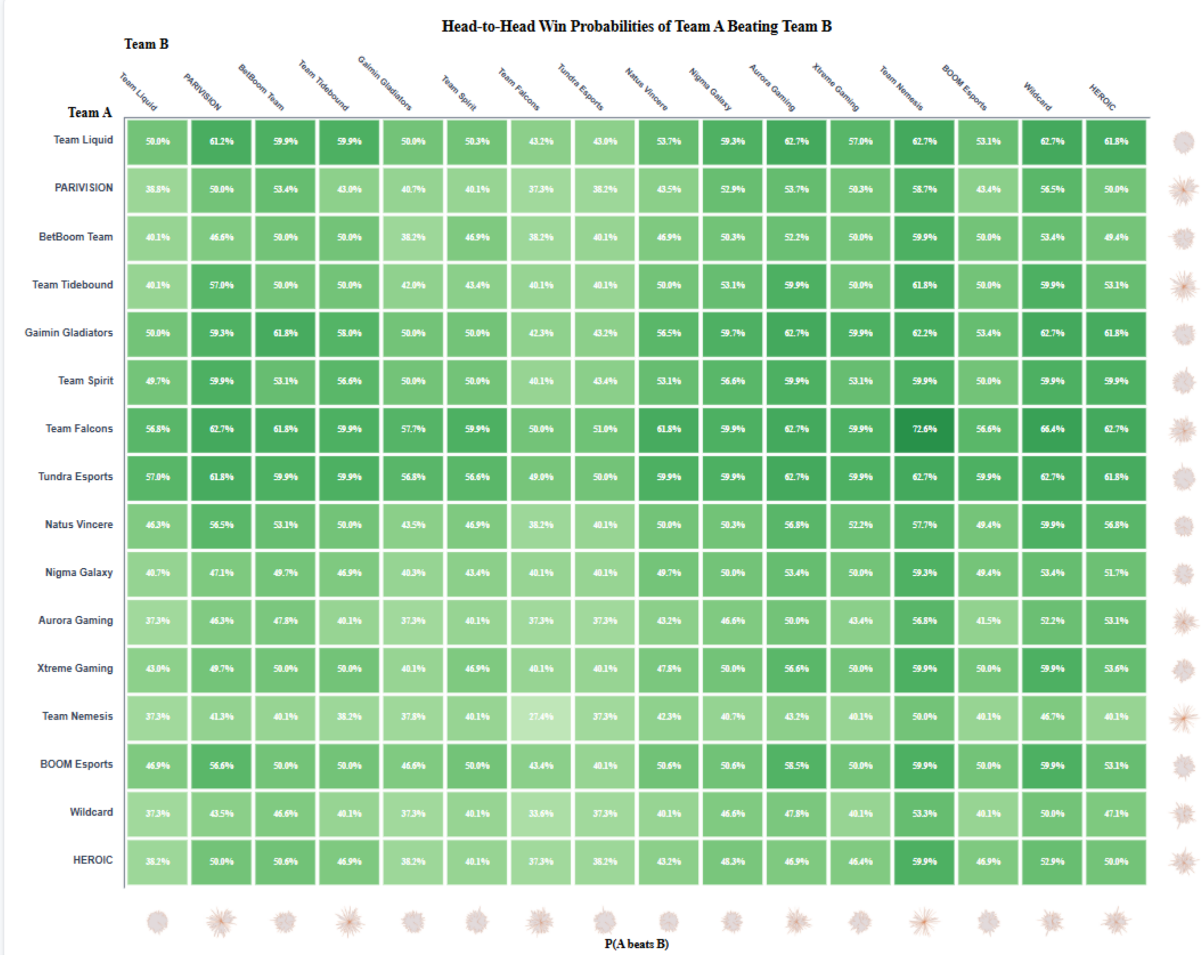
The monte carlo simulation has added parameters to change how the tournament seeds the matches, from random seeds to the user picking and choosing seeds, to having teams match up with teams in similar strength or prefer teams to match up against teams with wildly different strength.

Creating a Web Application to help Visualize the Probability Matrix and Monte Carlo Simulation Runs.

Current modern dashboards and visualization tools on the market do not do well with tournament bracket visualizations. The tournament visualization tools are limited, and not officially supported by large JavaScript libraries. And if they do support tournament brackets, they do not support the swiss system tournaments. So I had to make my own web application to do this to support the 3 different tournament formats.

There were some deliberate design choices that was made to increase the effectiveness of the visualizations. The web application was made with the vue js framework and only contains a frontend. By using vue js, the app can efficiently store state globally using the Pinia js library [9], meaning that switching between different visualizations won't break filters or destroy previously loaded data. The web app is also device agnostic and will display well on any screen resolution. Meaning that you can run the visualization on your computer screen or your phone screen, and both will run just fine using the same code base.

Some other features of the web app includes, all datasets are loaded into the client to decrease network latency. A heatmap, built using d3, shows the probability matrix that is used for the monte carlo simulation. The monte carlo simulation have parameters for the user to input to see differences in tournament format and how it can affect the teams.



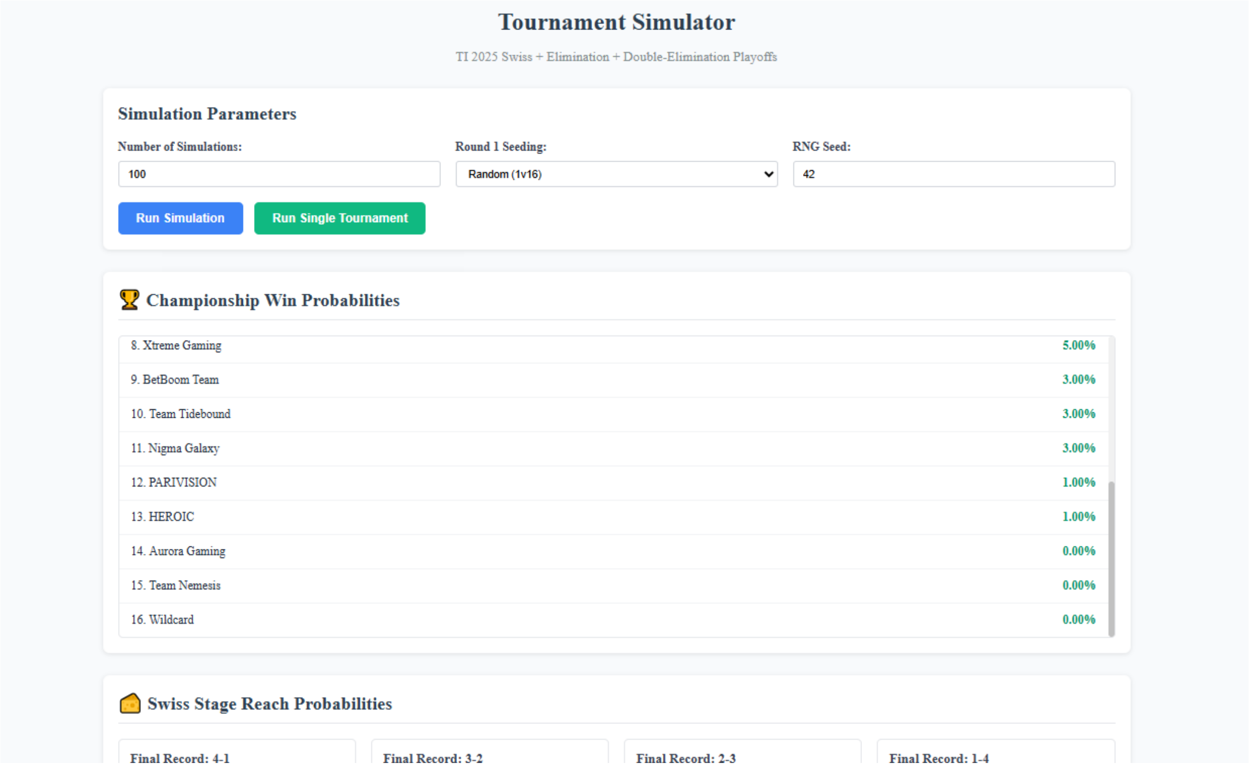
Head-to-Head Win Probabilities of Team A Beating Team B

Team B

Team A	Team Liquid	PARIVISION	BetBoom Team	Team Tidebound	Gaimin Gladiators	Team Spirit	Team Falcons	Tundra Esports	Natus Vincere	Nigma Galaxy	Aurora Gaming	Xtreme Gaming	Team Nemesis	BOOM Esports	Wildcard	HEROIC
Team Liquid	50.0%	61.2%	59.9%	59.9%	50.0%	50.3%	43.2%	43.0%	53.7%	59.3%	62.7%	57.0%	62.7%	53.1%	62.7%	61.8%
PARIVISION	38.8%	50.0%	53.4%	43.0%	40.7%	40.1%	37.3%	38.2%	43.5%	52.9%	53.7%	50.3%	58.7%	43.4%	56.5%	50.0%
BetBoom Team	40.1%	46.6%	50.0%	50.0%	38.2%	46.9%	38.2%	40.1%	46.9%	50.3%	52.2%	50.0%	59.9%	50.0%	53.4%	49.4%
Team Tidebound	40.1%	57.0%	50.0%	50.0%	42.0%	43.4%	40.1%	40.1%	50.0%	53.1%	59.9%	50.0%	61.8%	50.0%	59.9%	53.1%
Gaimin Gladiators	50.0%	59.3%	61.8%	58.0%	50.0%	50.0%	42.3%	43.2%	56.5%	59.7%	62.7%	59.9%	62.2%	53.4%	62.7%	61.8%
Team Spirit	49.7%	59.9%	53.1%	56.6%	50.0%	50.0%	40.1%	43.4%	53.1%	56.6%	59.9%	53.1%	59.9%	50.0%	59.9%	59.9%
Team Falcons	56.8%	62.7%	61.8%	59.9%	57.7%	59.9%	50.0%	51.0%	61.8%	59.9%	62.7%	59.9%	72.6%	56.6%	66.4%	62.7%
Tundra Esports	57.0%	61.8%	59.9%	59.9%	56.8%	56.6%	49.0%	50.0%	59.9%	59.9%	62.7%	59.9%	62.7%	59.9%	62.7%	61.8%
Natus Vincere	46.3%	56.5%	53.1%	50.0%	43.5%	46.9%	38.2%	40.1%	50.0%	50.3%	56.8%	52.2%	57.7%	49.4%	59.9%	56.8%
Nigma Galaxy	40.7%	47.1%	49.7%	46.9%	40.3%	43.4%	40.1%	40.1%	49.7%	50.0%	53.4%	50.0%	59.3%	49.4%	53.4%	51.7%
Aurora Gaming	37.3%	46.3%	47.8%	40.1%	37.3%	40.1%	37.3%	37.3%	43.2%	46.6%	50.0%	43.4%	56.8%	41.5%	52.2%	53.1%
Xtreme Gaming	43.0%	49.7%	50.0%	50.0%	40.1%	46.9%	40.1%	40.1%	47.8%	50.0%	56.6%	50.0%	59.9%	50.0%	59.9%	53.6%
Team Nemesis	37.3%	41.3%	40.1%	38.2%	37.8%	40.1%	27.4%	37.3%	42.3%	40.7%	43.2%	40.1%	50.0%	40.1%	46.7%	40.1%
BOOM Esports	46.9%	56.6%	50.0%	50.0%	46.6%	50.0%	43.4%	40.1%	50.6%	50.6%	58.5%	50.0%	59.9%	50.0%	59.9%	53.1%
Wildcard	37.3%	43.5%	46.6%	40.1%	37.3%	40.1%	33.6%	37.3%	40.1%	46.6%	47.8%	40.1%	53.3%	40.1%	50.0%	47.1%
HEROIC	38.2%	50.0%	50.6%	46.9%	38.2%	40.1%	37.3%	38.2%	43.2%	48.3%	46.9%	46.4%	59.9%	46.9%	52.9%	50.0%

P(A beats B)

The monte carlo simulation has a loading screen that shows it to be running the simulation as well what tournament it is currently running. Once the simulation runs finished, it then outputs the results. It shows the probability of a team visiting a bracket. For example it will show the probability of a team reaching round 5 in the 4-1 bucket, but for every team in that bracket. This allows users to estimate the odds of a specific team reaching a specific bracket and make bets on those odds.



Tournament Simulator

TI 2025 Swiss + Elimination + Double-Elimination Playoffs

Simulation Parameters

Number of Simulations: 100

Round 1 Seeding: Random (1v16)

RNG Seed: 42

🏆 Championship Win Probabilities

8. Xtreme Gaming	5.00%
9. BetBoom Team	3.00%
10. Team Tidebound	3.00%
11. Nigma Galaxy	3.00%
12. PARIVISION	1.00%
13. HEROIC	1.00%
14. Aurora Gaming	0.00%
15. Team Nemesis	0.00%
16. Wildcard	0.00%

Swiss Stage Reach Probabilities

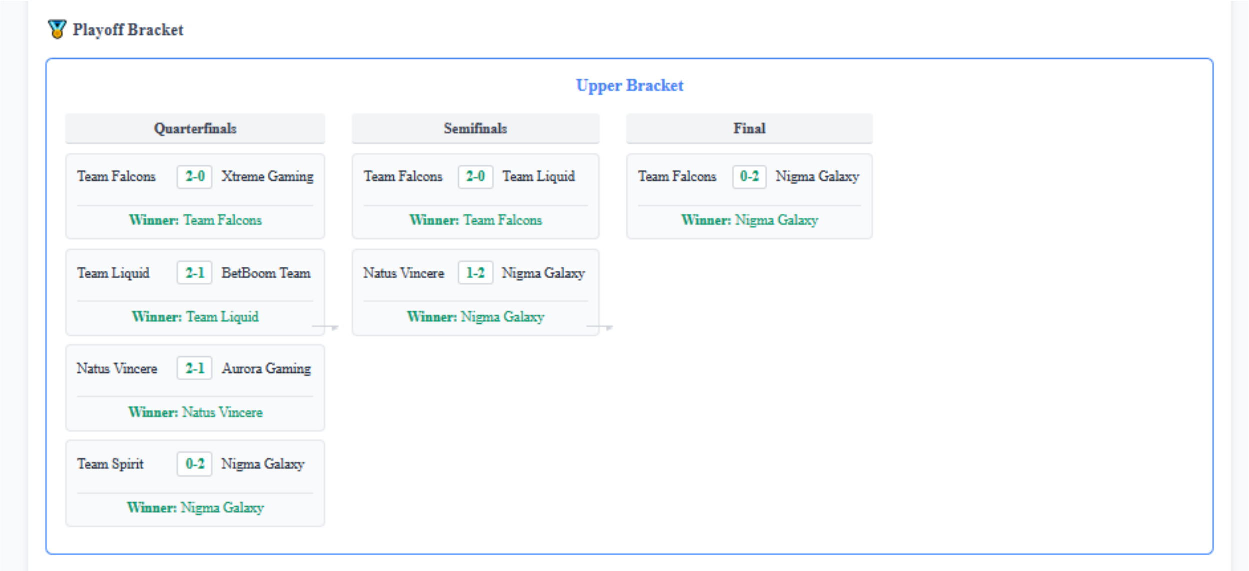
Final Record: 4-1

Final Record: 3-2

Final Record: 2-3

Final Record: 1-4

The monte carlo simulation also shows singular tournament runs and shows how a singular tournament might look like. It will show the structure of the Swiss tournament, the results of it, the elimination tournament, and the results of it, and the structure of the play offs and the result of it. This singular tournament run is meant to help users visualize how the tournament is visually structured and formatted, giving them a better understanding of how the tournament works.

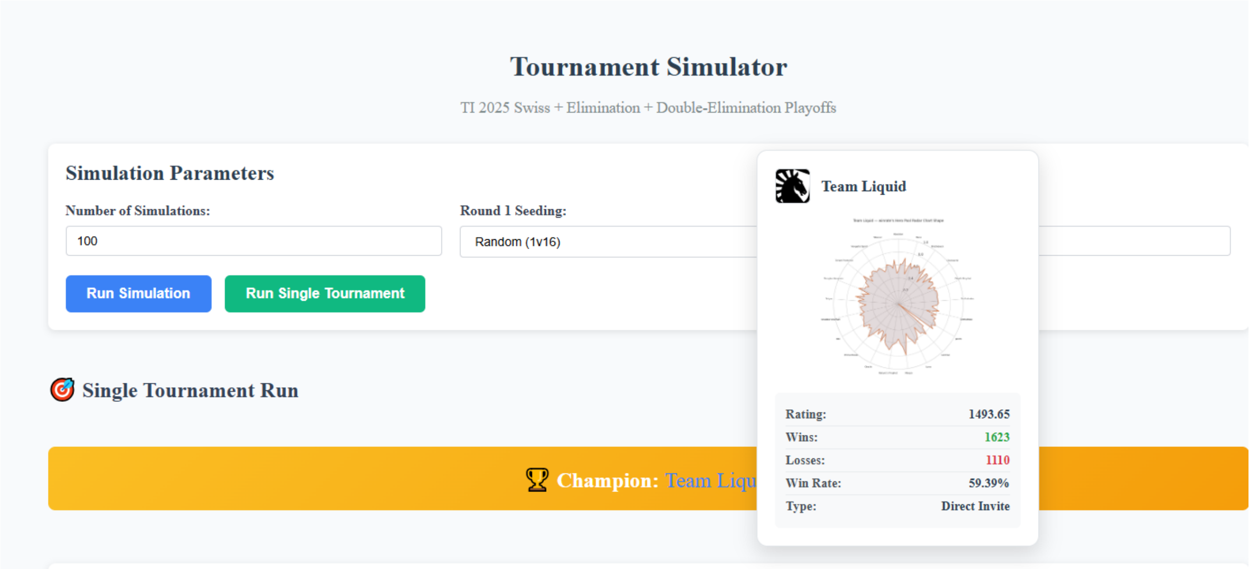


🏅 Playoff Bracket

Upper Bracket

Quarterfinals	Semifinals	Final
Team Falcons 2-0 Xtreme Gaming **Winner:** Team Falcons	Team Falcons 2-0 Team Liquid **Winner:** Team Falcons	Team Falcons 0-2 Nigma Galaxy **Winner:** Nigma Galaxy
Team Liquid 2-1 BetBoom Team **Winner:** Team Liquid	Natus Vincere 1-2 Nigma Galaxy **Winner:** Nigma Galaxy	
Natus Vincere 2-1 Aurora Gaming **Winner:** Natus Vincere		
Team Spirit 0-2 Nigma Galaxy **Winner:** Nigma Galaxy		

When hovering over a team name, the user can see details of that team as a card. This allows users to easily see more information on specific teams without having to look them up. This allow users to stay where they are currently in the web app and use the application more efficiently.



Tournament Simulator

TI 2025 Swiss + Elimination + Double-Elimination Playoffs

Simulation Parameters

Number of Simulations: 100

Round 1 Seeding: Random (1v16)

🎯 Single Tournament Run

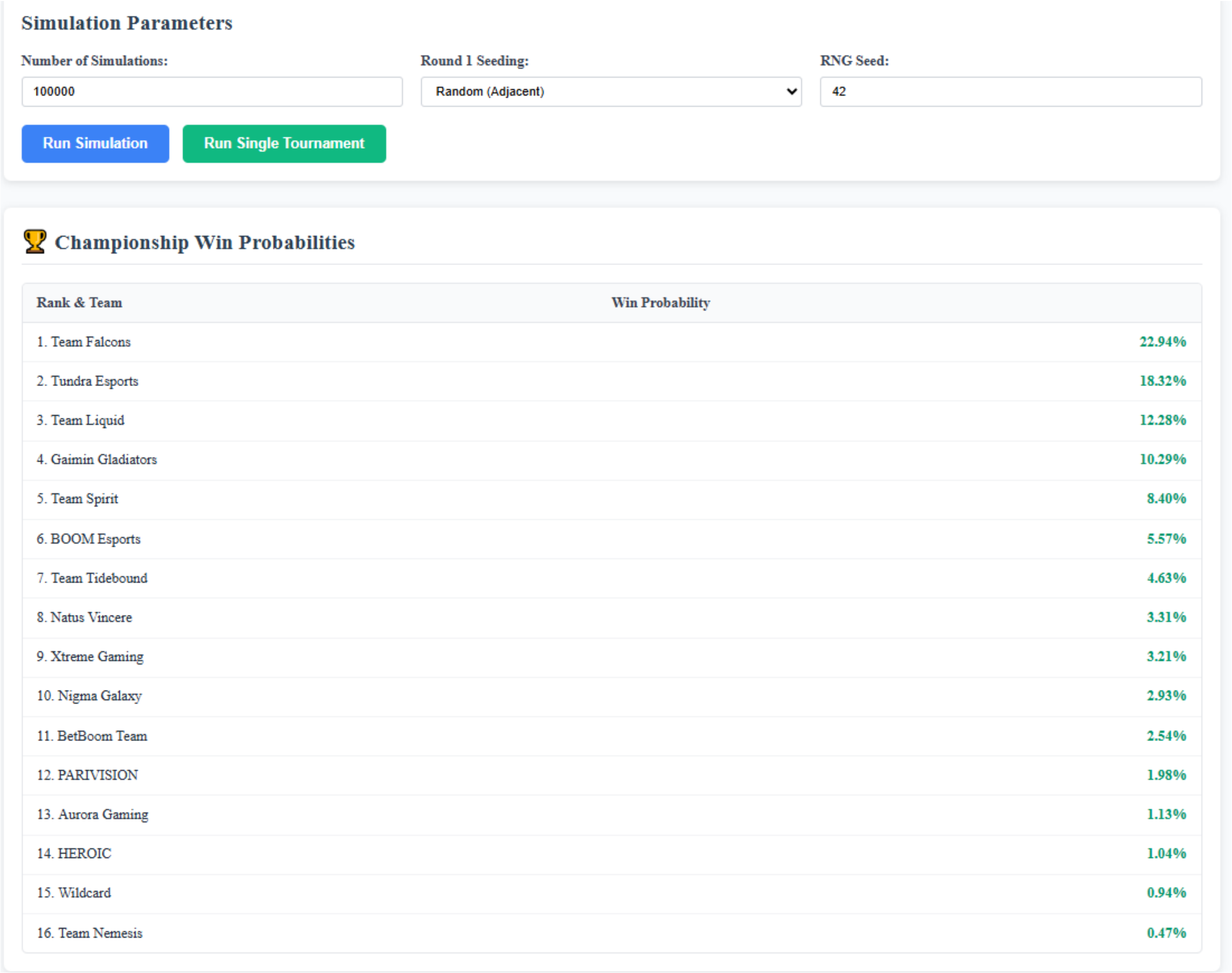
🏆 Champion: Team Liqu

Team Liquid

Rating:	1493.65
Wins:	1623
Losses:	1110
Win Rate:	59.39%
Type:	Direct Invite

Chapter 4: Results

With a monte carlo simulation of 100,000 runs, with Random (adjacent) seeding, we can see a table of probabilities of teams to win the TI 2025 Tournament.



Simulation Parameters

Number of Simulations: 100000

Round 1 Seeding: Random (Adjacent)

RNG Seed: 42

🏆 Championship Win Probabilities

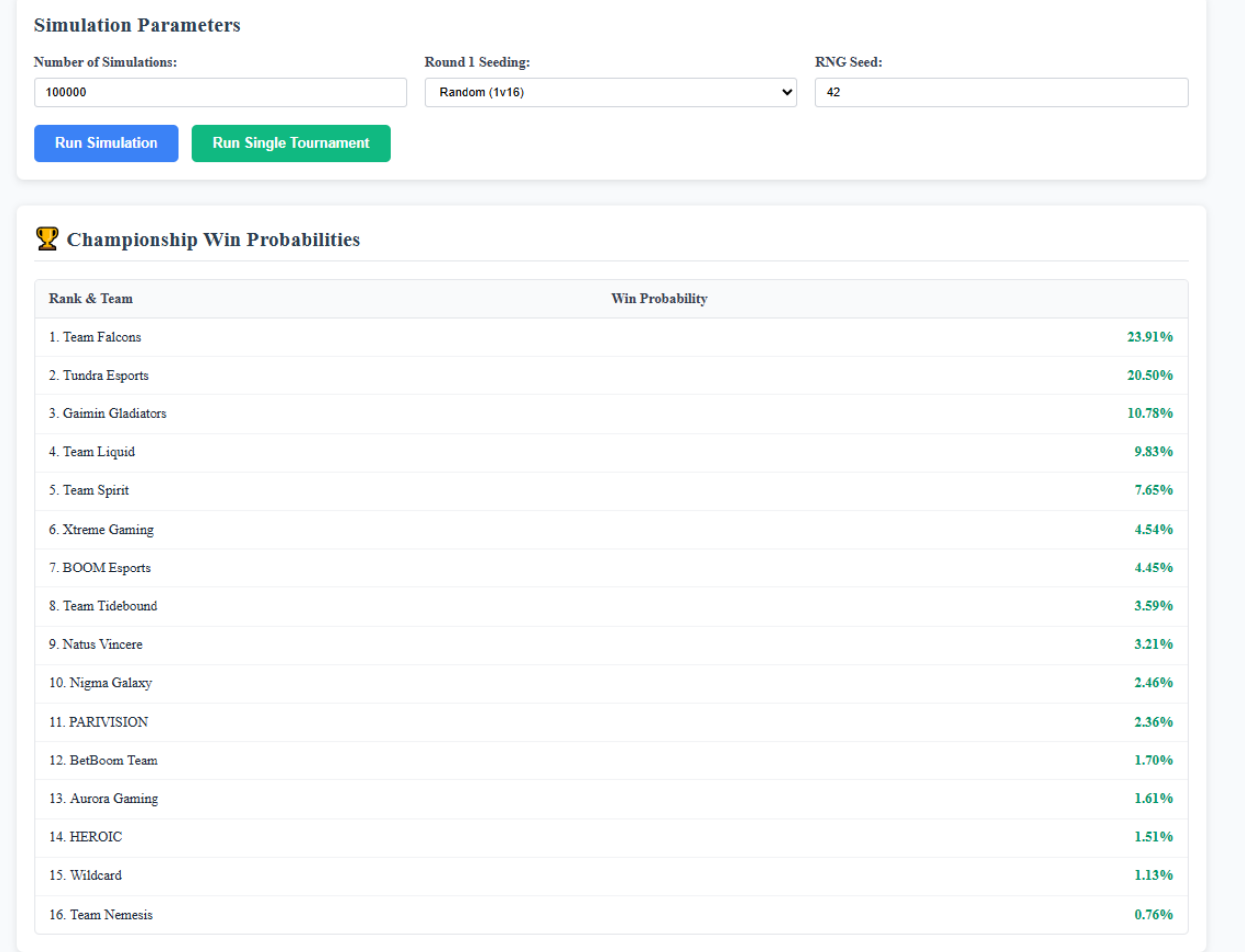
Rank & Team	Win Probability
1. Team Falcons	22.94%
2. Tundra Esports	18.32%
3. Team Liquid	12.28%
4. Gaimin Gladiators	10.29%
5. Team Spirit	8.40%
6. BOOM Esports	5.57%
7. Team Tidebound	4.63%
8. Natus Vincere	3.31%
9. Xtreme Gaming	3.21%
10. Nigma Galaxy	2.93%
11. BetBoom Team	2.54%
12. PARIVISION	1.98%
13. Aurora Gaming	1.13%
14. HEROIC	1.04%
15. Wildcard	0.94%
16. Team Nemesis	0.47%

The results are quite insightful. First we see that Team Falcons has a 22.94 % chance of winning TI 2025. If every team in the tournament had an equal chance of winning, then we would have expected around 1/16 chance of winning for a particular team. Which equals to around a 6.25 % chance of winning for a particular team. This means that Team Falcons has an over 3 times chance of winning than if it was purely random.

Another insight is that the top 5 teams that have the highest chance of winning the championship are all direct invites. Proving why they were directly invited to the tournament.

The last insight is that running different formats of how teams are to be matched up in the swiss tournament does affect the probabilities of different team winning.

You can see here with strongest to weakest team match ups in the swiss bracket lead to changing the ranking of the leaderboard for some teams.



Simulation Parameters

Number of Simulations: 100000

Round 1 Seeding: Random (1v16)

RNG Seed: 42

🏆 Championship Win Probabilities

Rank & Team	Win Probability
1. Team Falcons	23.91%
2. Tundra Esports	20.50%
3. Gaimin Gladiators	10.78%
4. Team Liquid	9.83%
5. Team Spirit	7.65%
6. Xtreme Gaming	4.54%
7. BOOM Esports	4.45%
8. Team Tidebound	3.59%
9. Natus Vincere	3.21%
10. Nigma Galaxy	2.46%
11. PARIVISION	2.36%
12. BetBoom Team	1.70%
13. Aurora Gaming	1.61%
14. HEROIC	1.51%
15. Wildcard	1.13%
16. Team Nemesis	0.76%

In the chart you can clearly see that for Random 1v16 it is slightly different to Random Adjacent. The Random Adjacent Chart leads to Gamin Gladiators to now be above Team

Liquid, and Parivision to now be above BetBoom Team. Probablities have changed, and lead to some teams being switched around in the win probability leaderboard.

Chapter 5: Conclusion

What Went Well

Overall the web application to display the visualization does manage to show the monte carlo simualtion, and allows users to hover over team names to see more details about the team information. This helps uses in visualizing the information, such as the format and structure of the tournament and the data of the teams.

The global states of the web application, allows users to switch between pages of the web application without losing state, allowing users to switch between the monte carlo simulation and other areas of the web application,

Using a distilled probability matrix from a machine learning model, by reducing the parameters so that we only use relevant teams partaking the the tournament lead to a massive increase in performance. Not only that but a probability matrix is programming language agnostic, meaning that the machine learning model created in python would have to be used in a python environment. While the probability matrix is just a look up table and can be ported to any programming language, allowing it to work in JavaScript, even though it was generated in python.

Playing around with the web application and running the monte carlo simulation, does provide insight on what teams are more likely to win the overall tournament. Not only that, but the probability matrix also shows that some teams do have a large advantage over other teams depending on the different hero pools

To see the web application, visit this link. https://data-visualization-final-project-production-8917.up.railway.app/

Possible Improvements

The model right now uses a monte carlo simulation to get the statisitical probability of teams reaching a certain bracket and winning the overall tournament. But there are already ways to find that probability without running any simulations. If we used a mathematical approach, we could see at least “two orders of magnitude faster” [7] in compute time compared to the monte calro simulation.

The model also right now only tracks the hero pool of the team. It does not track the hero pools of individual players for that team. It may be more accurate to track the individual hero pools of the actual players playing, than tracking the overall team, as teams will change their player roster over time.

The web application could also possibly show more visualizations, and different kinds of marks to show trendlines. Right now most of the visualization in the web app is a heatmap, or in a tabular format. The cards to show the team information is actually in a tabular format, where each card that represents a team is actually just a table. The monte carlo simulation and the single tournament runs are also in a tabular format, just in a different style. This doesn’t exactly show trendlines. If I would have to improve the visualization of the web application, I would attempt to show trendlines, and not just rely on tabular formats for visualizations.

References

[1] https://www.dota2.com/newsentry/569242769355177990

[2] https://dota2freaks.com/drafting

[3] https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=5489&context=gc_etds

[4] https://www.dota2.com/newsentry/536597286667419796

[5] https://www.theguardian.com/sport/2012/aug/01/london-2012-badminton-disqualified-olympics

[6] https://dl.acm.org/doi/pdf/10.1145/3637303

[7] https://journals.sagepub.com/doi/10.1177/22150218251313905

[8] https://docs.opendota.com/#tag/explorer

[9] https://pinia.vuejs.org/core-concepts/state.html