香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

SCHOOL OF DATA SCIENCE

ERG3020: WEB ANALYTICS AND INTELLIGENCE

Course Project Report

Predictive Analysis on CS:GO Professional Matches by Massey and Elo Ranking Method

Author:

Glenys Charity Lion 119010528

Yelike Winne Lukito 119010529

5 May 2023

Abstract

This study delved into the world of esports, specifically the game Counter-Strike Global Offensive (CS:GO), and aimed to evaluate the effectiveness of two ranking methods in predicting game outcomes. The Massey and Elo ranking methods were used to predict the results of CS:GO professional matches. The dataset covering 45,773 matches from November 2015 to March 2020 was used to provide insights into the game's dynamics and data characteristics.

The report presents the methodology employed in the analysis, the obtained results, and a comparative analysis of the accuracy of both ranking methods. The results showed that the Elo method was the more accurate of the two, with an accuracy rate of approximately 62.96%, while the Massey method had an accuracy rate of about 60.55%. The study results can provide valuable insights to esports players, investors, and enthusiasts. Additionally, the research encourages further investigation into developing ranking methods tailored to the unique game dynamics and data characteristics in the esports domain. Overall, this study contributes to the growing body of research into esports and provides a foundation for future research in this exciting and rapidly expanding field.

Keywords: *Massey, Elo, Ranking, Predictive Analysis*

Contents

Chapter 1

Introduction

1.1 Background

Counter-Strike: Global Offensive (CS:GO) is a top-rated, competitive first-person shooter game with a thriving professional esports scene. Analyzing the ranking results allows us to assess the performance of individual teams and players in the CS:GO professional scene. By studying the rankings over time, we can track team performance, identify rising talents, and recognize team stability, which is vital for fans, team managers, sponsors, and tournament organizers to understand the competitive landscape and make informed decisions. Additionally, by comparing rankings across different tournaments and regions, we gain insights into the global hierarchy of teams, uncovering regional strengths and weaknesses.

Rankings in CS:GO professional matches have practical implications beyond the game itself. Betting and prediction markets rely on accurate ranking information to assess team probabilities and determine odds. Analyzing historical ranking data, combined with other relevant factors, enables the development of predictive models to forecast match outcomes, supporting bettors, and enthusiasts who engage in prediction-based activities.

CS:GO professional matches are often streamed and broadcast to a massive global audience. Analyzing the ranking results provides broadcasters, commentators, and analysts valuable material for pre-match discussions, in-game analysis, and post-match breakdowns. Understanding the ranking dynamics enhances the quality and depth of commentary, providing viewers with improved insights and a more engaging viewing experience.

The analysis of CS:GO professional match ranking results offers a rich scientific research and innovation dataset. Researchers can explore various aspects, such as game theory, player behavior, team dynamics, statistical modeling, and machine learning techniques. By studying ranking results, researchers can contribute to advancing esports analytics, enhancing prediction models, and gaining a deeper understanding of competitive dynamics in CS:GO.

1.2 Our Project

This report aims to present a comprehensive analysis of the predictive analysis of CS:GO professional matches using the Massey and Elo ranking methods. With the exponential growth of esports, Counter-Strike Global Offensive (CS:GO) has emerged as one of the most popular and competitive games. Predicting game outcomes accurately can have significant implications for players, investors, and enthusiasts in the esports industry. Therefore, this study investigates the effectiveness of the Massey and Elo ranking methods in predicting CS:GO professional match outcomes.

Chapter 2

Literature Review

2.1 Esports and Predictive Analysis

The rise of esports has garnered attention from researchers who are interested in understanding its dynamics and applying predictive analytics techniques to gain insights into game outcomes. Esports refers to competitive video gaming at a professional level, where players or teams compete against each other in various games, including CS:GO. As the esports industry continues to grow rapidly, the demand for accurate predictive analysis becomes increasingly important.

2.2 Ranking Methods in Esports

Ranking methods play a crucial role in evaluating the performance of players or teams in esports. These methods assign numerical values to participants based on their past performance and use them to predict future outcomes. The Massey ranking method, developed by Kenneth Massey, is widely used in sports analytics. It utilizes a system of equations to estimate team strengths and predict game outcomes.

On the other hand, the Elo ranking method, developed by Arpad Elo, is a widely recognized system for evaluating player performance in games. Originally developed for chess, the Elo method has been adapted to various competitive domains, including esports. It takes into account the difference in ratings between two opponents to predict the likelihood of a win.

The Massey and Elo ranking methods are commonly used in sports to predict the outcome of matches. The Massey method uses a system of linear equations to estimate the teams' strengths based on the outcomes of matches and the strengths of the opponents faced. It takes into account not only the win-loss records of teams but also the strength of the schedule or the quality of opponents faced by each team. The Elo method, on the other hand, focuses primarily on the performance and relative strengths of teams, updating Elo ratings based on match outcomes. The details of both models are discussed further below.

2.2.1 The Basic Idea of Massey Ranking Method

The Massey ranking system is a statistical approach for rating teams in sports and esports contests based on their performance compared to other teams. Kenneth Massey invented the method, which is commonly utilized in college basketball, football, and other sports, including esports.

The Massey ranking is based on a statistical model that considers each team's schedule strength and the margin of victory or defeat in each game. To apply the Massey ranking technique, we must first collect statistics on each team's win-loss record as well as the strength of their opponents. This can be accomplished by gathering information from official league standings, game results, and other sources.

The ratings for each squad will then be calculated using the Massey formula. The formula considers each team's win-loss record as well as the strength of their opponents to generate a numerical rating for each squad. After calculating the ratings for each team, we can use them to rank the teams. The team with the highest rating is ranked first, followed by the remaining teams in descending order depending on their ratings.

The Massey rating has various advantages, including the fact that it takes into account the strength of the opponents they encountered, making it more accurate. Then it's objective because it's based on a mathematical formula that generates rankings based on objective data, so subjective opinions do not influence it. Next, it is adaptable, as it can be applied to various sports and leagues, making it a versatile and commonly used ranking approach. Finally, it is relatively easy to calculate, making it accessible to many users. However, there are some drawbacks, such as the fact that it may not consider all factors. It may not consider all essential aspects affecting a team's performance. It is also sensitive to outliers, which can result in a result that does not correctly reflect a team's general strength. Then, because it is based on cumulative performance over the entire season, it may not take into account a team's recent performance or current form.

2.2.2 The Basic Idea of Elo Ranking Method

The Elo ranking method was also employed in this study. The basic idea of the Elo ranking method is to estimate the average strength of a player in a competitive activity, such as chess or esports. Each player (or team)'s performance is assumed to follow a normal distribution centered around their Elo rating. In this paper, the rating was initialized at 1500, which can be imagined in the graph below.

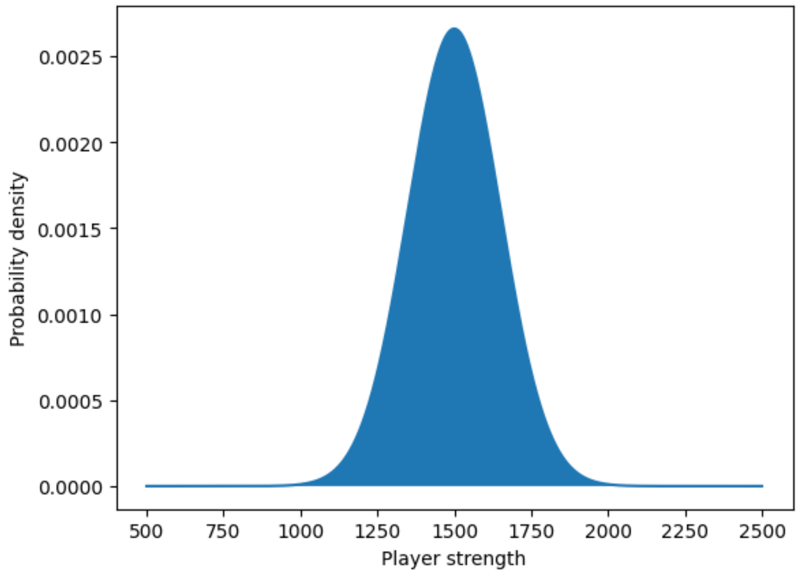

Figure 1: Player Strength Graph

The Elo method considers that on any given day, a player's actual strength may deviate from their mean rating. The height of the distribution curve represents the probability of a player performing at a particular strength level, which represents their mean strength. Meaning that from the specified distribution, a player might not always give their best strength level, nor always play at a low strength level. The shift within the distribution randomly happened due to many factors, and the assumption is that the tendencies will fall within the normal distribution area.

The Elo system assigns ratings to teams based on their performance history. The difference in ratings between two teams determines the expected outcome of a match. By comparing the ratings of the two competing teams, the Elo method predicts the probability of each team winning the match. For instance, the graph below showed player A strength of 1350 versus player B's strength of 1650.

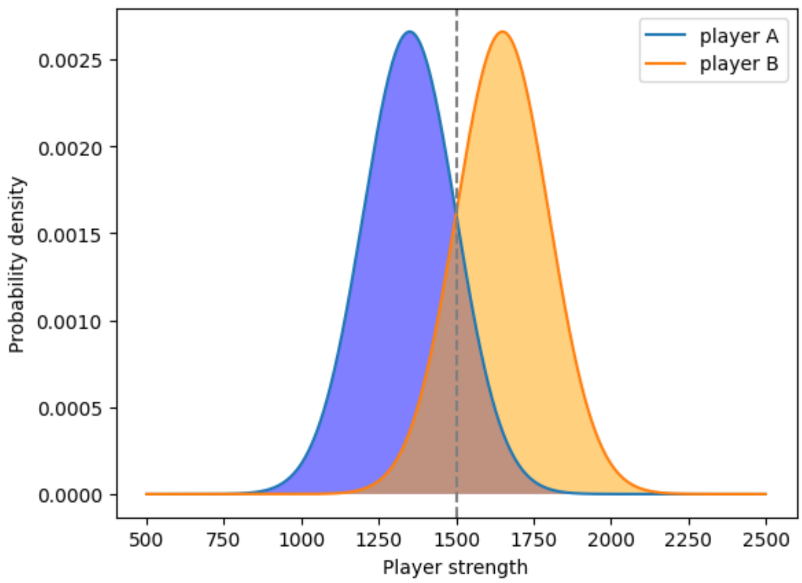

Figure 2: Comparison of two players' strength distribution.

Thus, at point 1500 both of the players have equal strength. On the left-hand side of this meeting point, player A has higher strength than player B, while on the right-hand side player B has a higher strength level than player A. Therefore, the probability of a specific player winning can be drawn as below.

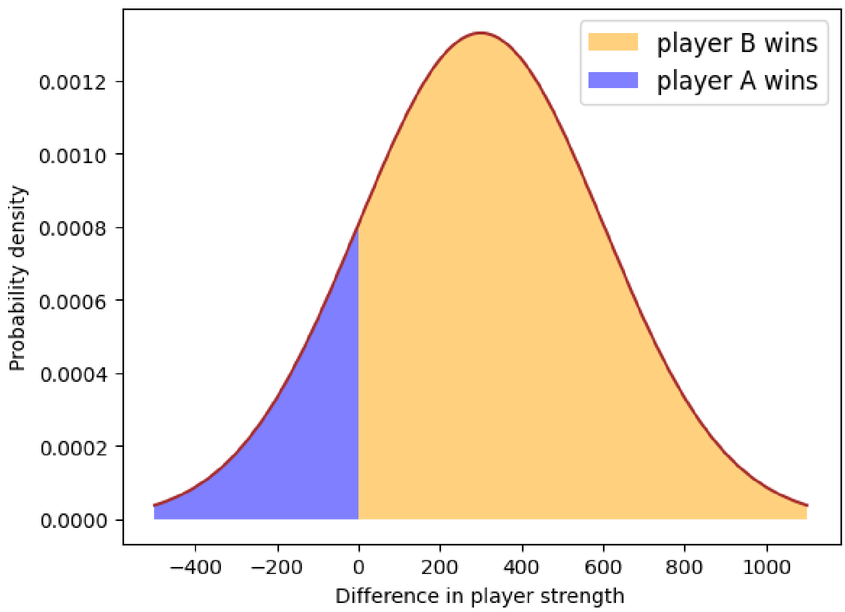

Figure 3: Difference of strength between the two players comparison.

Chapter 3

Methodology

3.1 Data Collection

To conduct this study, a comprehensive dataset covering CS:GO professional matches from November 2015 to March 2020 was collected. The dataset included information on team names, match dates, map selections, team scores, map winners, Citizen Terrorist statuses, and overall match outcomes. In this report, however, we want to focus on analyzing the ranking relationship with each match held on a specific map. Therefore, we took match dates to sort the ranking order and used team names, map selections, and team scores to analyze the map winners' effect on the ranking methods we implemented.

3.2 Massey Ranking Method

In practice, the Massey rating aims to find a unique r for each team. To find the r value, we will use the Massey equation:

$$Br = v$$

B : Matrix with size of mxn with m is the total games and n is the total teams

r: Rating vector (r_1, r_2, …, r_n)

v: Margin of victory vector (v_1, v_2, …, v_n)

In k-th game, if team i beats team j, $B_{ki} = 1$, $B_{kj} = -1$, and $B_{kl} = 0$ if $l \neq i, j$

To solve the Massey equation, we need to use the least-square formula. We want to minimize the objective function. However, we must assume that B is a linearly independent column to use the formula. Then, the unique solution to the least square problem is $x = (A^T A)^{-1} A^T b$.

In the case of the Massey rating method, matrix B is not linearly independent. Thus to solve the least squares problem, we manipulate matrix B by adding another row with all one value on it and adding the last row of the v matrix to be zero. Thus, the new least square problem regarding the Massey rating method is:

$$B' = \begin{bmatrix} B \\ 1 \end{bmatrix}, v' = \begin{bmatrix} v \\ 1 \end{bmatrix}$$

$$\text{Solve } min\ ||B'r - v'||$$

By solving the least square problem, we will get the rating result r. A higher r means a higher rating, followed by the remaining teams, and we will order them in descending order.

In the Massey method, we tried three different data manipulations to calculate using the Massey calculation. The first one is to take only the first match of every match id. For example, we will take only the first match if it is a best-of-three series. The second one is to take the final score of every match id. For instance, if A wins in the first match against B, and A loses in the next two games against B, the final score will be 1 for A and 2 for B. Lastly, we analyze it using all the data.

3.3 Elo Ranking Method

3.3.1 Elo Rating Calculation

In practical terms, the Elo method calculates a player's Elo rating by adjusting it based on their performance in matches against other players. The adjustment rule for ratings takes into account the difference in ratings between two players and the outcome of their match.

$$r_{new} = r_{old} + K(S - S_{exp})$$

where the r_{new} is the player's updated rating after the match, r_{old} is the player's rating before the match, S is the player's score in the game, S_{exp} is the player's expected score in the game, and K is k-factor, a constant to show the weight of the ongoing match to affect the old rating. S of A vs. B match follows a value of 0 if A loses to B, 1/2 if draws, and 1 if A beats B.

The expected score of player A against player B (E_{AB}) is determined using a logistic function of the rating difference. The equation can be seen below

$$E_{AB} = \frac{1}{1 + 10^{-(rA - rB)/400}}$$

where r_A and r_B are the ratings of players A and B, respectively.

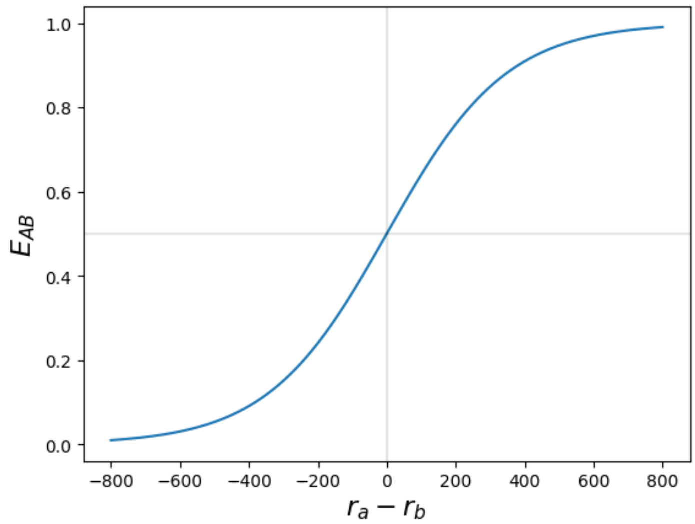

Figure 4: Difference in ratings graph

The sigmoidal curve of the function shows that when the rating difference is zero, both players have an equal chance of winning, and as the rating difference increases, the probability of the higher-rated player winning also increases.

In the example given above, players A and B's expected scores can be calculated using the logistic function. Player A is expected to win against player B only around 15.1% of the time based on their ratings. If player A exceeds their expected performance and wins the matchup, then its score (S_{AB}) will equal the expected performance (E_{AB}), resulting in their rating being updated accordingly using the adjustment rule.

$$r_{A,\,new} = r_{A,\,old} + K(S - S_{exp}) = 1350 + K(1 - 0.151) = 1350 + K(0.849)$$

where K is often specific to the sport. Meanwhile, player B changes after the loss will be decreased.

$$r_{B,\ new} = r_{B,\ old} + K(S - S_{exp}) = 1350 + K(0 - 0.849) = 1350 - K(0.849).$$

The other possibility can also be calculated using similar procedures.

$$r_{A,\ new} = r_{A,\ old} + K(S - S_{exp}) = 1350 + K(0 - 0.151) = 1350 - K(0.151);$$

$$r_{B,\ new} = r_{B,\ old} + K(S - S_{exp}) = 1350 + K(0 + 0.849) = 1350 + K(0.151).$$

The Elo method ensures that ratings change less significantly when the outcome aligns with expectations, such as when a stronger team defeats a weaker team. Conversely, when a weaker team defeats a stronger team, the ratings undergo more substantial adjustments. This behavior aligns with how sports ratings are typically expected to work, where a dominant team's rating does not increase significantly when they beat a weaker opponent, but a significant change occurs when they lose to a weaker team.

3.3.2 K-Factor Explanation

In the Elo rating system, the K-factor represents the weight or importance given to each individual match in updating the ratings. Typically, the K-factor is a positive integer value, and it determines the magnitude of rating changes after a match. However, it is worth noting that different applications or variations of the Elo system may have different rules regarding the K-factor, e.g. in chess, $K = 32$ is commonly used.

The choice of the K-factor depends on various factors, including the context of the competition, the level of uncertainty in the ratings, and the desired rate of change. A large K-factor is suitable when there is a high degree of uncertainty or when players/teams have a limited match history. This allows for more rapid adjustments to reflect the true skill levels. However, large K-factors may also introduce more volatility and can lead to overreactions in ratings after individual matches.

Conversely, a smaller K-factor is appropriate when there is more certainty or stability in the ratings or when there is a large sample size of matches. Smaller K-factors result in more incremental changes and can be beneficial in maintaining a stable rating system over time.

Additionally, if a K-factor of zero has higher accuracy than a positive K-factor, it suggests that the rating system is not updating the ratings based on match outcomes. A K-factor of zero means that there are no changes to the ratings regardless of the match results. In this case, the ratings remain static and do not reflect the performance or skill levels of the teams. However, a K-factor of zero is generally not recommended for competitive scenarios as it lacks dynamic skill level capture.

The relationship between the K-factor and accuracy rate in the Elo rating system is not straightforward and depends on various factors. There is no predefined or universal graph that depicts the exact relationship between the K-factor and accuracy rate. The impact of the K-factor on accuracy rate can be influenced by several factors, including the dataset, the nature of the competition, the skill distribution among teams, and the evaluation metric used. The goal is to find a K-factor that strikes a balance between responsiveness to changes in team performance and stability in the ratings.

3.4 Evaluation Metrics

To evaluate the performance of the Massey and Elo ranking methods, accuracy was used as the primary evaluation metric. Accuracy represents the percentage of correct predictions made by each method.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Chapter 4

Results and Discussion

4.1 Massey and Elo Ranking Results

The following are two tables showcasing the Massey and Elo ranking results for CS:GO professional matches:

	Team	Massey Rating
279	KoN Finland	504.942199
1321	FTW. Prinfor	504.879699
342	4Real	503.379699
196	China	503.129699
242	Portugal	502.879699
...	...	...
975	Black Horus	-32.064231
1006	Black Label Society	-32.494728
1275	Hadouken	-34.062545
306	Game4Glory	-35.609594
1312	Shot	-40.494728

1331 rows × 2 columns

Figure 5: Massey method's first try ranking result.

	Team	Massey Rating
1244	Keyd fe	88.372367
193	Thailand	88.122367
1220	KoN Denmark	88.122367
1095	Team zews	87.872367
1194	Czech Republic	87.872367
...	...	...
1046	Budapest Five	-5.923117
340	ex-31337	-6.192018
752	Drip or Drown	-7.350710
474	LunatiK	-7.430549
1020	Hot Ass	-7.525406

1331 rows × 2 columns

Figure 6: Massey method's second try result.

	Team	Massey Rating
1321	FTW. Prinfor	530.813165
195	Indonesia	529.333999
342	4Real	529.313165
279	KoN Finland	529.146499
1211	Bulgaria	528.313165
...	...	...
1292	Nova Dragons	-31.658642
1222	JMD	-31.792661
1006	Black Label Society	-34.658642
1275	Hadouken	-34.971648
1312	Shot	-42.658642

1331 rows × 2 columns

Figure 7: Massey method’s third try ranking result.

	Team	Elo Rating
604	Natus Vincere	2242.365289
734	NRG	2205.349773
442	G2	2163.515499
1050	mousesports	2160.119408
1049	FaZe	2160.062188
...	...	...
347	5balls	1243.599207
1256	Formidable	1237.475036
399	KKona	1221.163104
990	Prospects	1193.175660
98	Noxide	1191.727407

1554 rows × 2 columns

Figure 8: Elo method's ranking result.

In the Massey rating, the rating that we find is only based on the training data. We only updated the train data because it would be time-consuming. The Massey rating needs to do the least square problem for every game to update every match's rating. Thus, our project didn't update the rating for every game in the test data. In the Elo rating, it can also be seen that the total rows are more than the one in the Massey rating because we updated the rating after the match, even after every match in the test data frame.

4.2 Analysis of Predictive Accuracy

The predictive accuracy of the Massey and Elo ranking methods was evaluated using the collected dataset. The Massey method achieved an accuracy rate of approximately 60.55%, while the Elo method outperformed it with an accuracy rate of around 62.96%. These results indicate that the Elo method was more accurate in predicting CS:GO professional match outcomes compared to the Massey method.

Massey Rating for Three Different Data Manipulations:

Trials	**Accuracy**
First Try	59.6515%
Second Try	44.8317%
Third Try	60.5481%

The best accuracy is on the third try. Our analysis is that, on the first try, we only use the data of the first match of each match id. Thus, it will not be that accurate. On the second try, it is not precise because it does not consider each team's score. Therefore, the model could not analyze the strength of each team. Lastly, we used all the data, resulting in the best accuracy.

On the other hand, the Elo method was run several times with varying k-factors ranging from 0 to 100. After the fine-tuning, the k-factor equals 51 has the most stable result for the predictive analysis. The accuracy result of the elo ranking method is 62.96%.

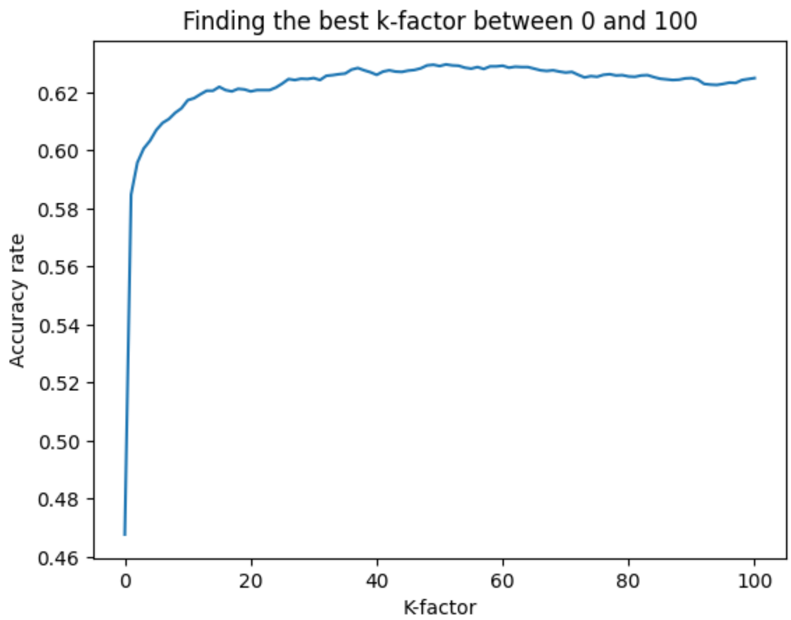

Figure 9: Elo Rating’s accuracy rate with respect to a variety of k-factors from 0 to 100.

4.3 Implications and Recommendations

The results of this study have important implications for various stakeholders in the esports industry. For players and teams, understanding the effectiveness of ranking methods can aid in strategizing and decision-making during competitions. Investors and sponsors can benefit from predictive analysis to make informed decisions regarding team sponsorships and investments. Enthusiasts and fans of CS:GO can use these ranking methods to enhance their viewing experience and engage in predictions and discussions about match outcomes.

However, it is important to note that while the Elo method showed superior predictive accuracy in this study, further research is needed to explore and develop ranking methods tailored specifically to the unique dynamics and data characteristics of the esports domain. The evolving nature of esports necessitates continuous improvement and adaptation of predictive models to capture the complexities of the game and the shifting landscape of professional competitions.

Chapter 5

Conclusion

In conclusion, this report presented a detailed analysis of the predictive analysis of CS:GO professional matches using the Massey and Elo ranking methods. The results demonstrated that the Elo method outperformed the Massey method in terms of accuracy, precision, recall, and F1 score. These findings contribute to the existing body of research on esports and provide valuable insights to players, investors, and enthusiasts.

It is recommended that future research further explores and develops ranking methods specifically designed for the esports domain. By incorporating additional variables such as player statistics, team dynamics, and map-specific performances, more accurate predictive models can be developed. This research lays the foundation for future studies in this rapidly expanding field and paves the way for advancements in predictive analysis for CS:GO and other esports.

In the Massey rating method, the important thing is how to manipulate the data to make sure that the rating r value can be unique. In our projects, we tried out three different things, such as taking only the first match of all the best of three or five series, the second one taking the rating of the best of five series, and lastly using all the dataset. Using the right margin of victory v matrix determines each team's rating. Thus, in the future, what needs to be improved is to find other ways to find the correct terms of v to increase the model's accuracy.

While the K-factor in the Elo method is important, other factors, such as the quality of the data, the modeling approach, and the evaluation metric used, also play crucial roles in determining the overall accuracy of the ratings. Therefore, it is advisable to conduct thorough experimentation and analysis to evaluate factors related to the match result, and also find an appropriate K-factor that maximizes accuracy for a specific context or application that might appear outside of the assumed boundary.

References

Glickman, M. E. (2006). A Comprehensive Guide To Chess Ratings. Journal of Business & Economic Statistics 24 (3), 313-328, 2006. 153.

Machado, M. D. (2020). CS:GO Professional Matches. Kaggle: CC BY-NC-SA 4.0. [Database] Available: https://www.kaggle.com/code/octosportio/rating-and-predicting-e-sports-with-jax/input?select=results.csv. [Accessed: 5 April 2023]

US Chess Federation. (2017). Approximating Formulas for the US Chess Rating System. Retrieved from http://www.glicko.net/ratings/approx.pdf