Modeling and Analyzing Football Team Performance Using Rating System

Page Count: 15

Table of Contents

Introduction

Football has always been an important part of my life. I'm passionate about the sport, as are many of my friends. I believe the beauty of football lies not only in the game itself but also in its social aspect, which includes heated debates and discussions among fans. Every football fan has a favorite team, leading to endless arguments about which club is the greatest of all time. Among the many legendary clubs in football history, two names consistently emerge at the top of the discussion: FC Barcelona and Real Madrid. These two clubs have dominated Spanish and international football for decades and share an incredible rivalry, with their matches called *El Clásico (The Classic* in Spanish). This historic rivalry stretches back almost a century and is driven by several factors. These factors include deep political and historical tensions, with Barcelona representing Catalan nationalism and Real Madrid symbolizing Spanish centralism, which intensified during Franco's rule. Furthermore, key events, such as Luis Figo's controversial move from being Barcelona's captain to a Madrid player, further fueled the animosity. Additionally, the modern era's comparisons between Lionel Messi and Cristiano Ronaldo have even more elevated El Clásico (Mitton, 2024).

Given the deep history and ongoing debates, one question continues to divide football fans: "*Which club is truly the greatest of all time*"?

First, clear criteria for determining the best football club in history must be defined to establish a fair comparison. Football performance is dynamic and influenced by transfers, injuries, tactical approaches, and other various factors.

Rather than assessing overall historical success, I will focus on determining which club has reached the highest peak performance level over time. To achieve this, I will establish a mathematical model capable of quantifying each team's performance based on objective and reproducible metrics. This approach allows to move beyond subjective fan biases and offer a structured, data-driven evaluation of the historical debate surrounding Barcelona and Real Madrid.

Aim

The primary goal of this investigation is to develop a football power rating system for Spain's top football division, LaLiga, covering the seasons from 1995/96 to 2022/23 to determine who is historically the best between Barcelona and Real Madrid. This will be accomplished by comparing various rating methods to identify the most effective and reliable system for evaluating team performance over time.

Scope Parameters and Data Selection

Scope and delimitations

The IA focuses on Spain’s top football division, LaLiga, from the 1995/96 season to the 2022/23 season due to the dataset limitation and excludes lower divisions and other competitions, including domestic and international cups. The dataset does not include friendly matches or playoff games that are not part of the league structure.

Sample

The data used in this investigation comes from a Kaggle dataset by Kishan Kumar called La Liga Complete Dataset, which includes columns: Season, Home Team, Away Team, FTHG (Full-Time Home Goals), FTAG (Full-Time Away Goals), FTR (Full-Time Result), HTHG (Half-Time Home Goals), HTAG (Half-Time Away Goals), and HTR (Half-Time Result). However, my investigation only needs the Home Team, Away Team, and FTR (*see Appendix II for a sample of the dataset*). The author consolidated other individual datasets for each season into one to enhance convenience and usability (Kumar, 2023). The original data was sourced from Football-Data.co.uk, a reputable source for historical football statistics frequently cited in sports analytics research.

To check the accuracy and completeness of the dataset, I performed a cross-check by comparing approximately 50 match results from the dataset with the official LaLiga website, confirming that all results matched. Next, I calculated the total number of matches played between the 1995/96 and 2022/23 seasons. Since each team plays against every other team twice per season (home and away), the number of matches per season is determined by:

$$\big(T(T-1)\big)$$

where T is the number of teams in that season.

LaLiga underwent a structural change for the 1998/99 season, reducing the number of teams from 22 to 20. Therefore, the dataset consists of:

- 2 seasons with 22 teams
- 26 seasons with 20 teams

The total number of matches over this period is calculated as:

$$2\big(22(22-1)\big)+26\big(20(20-1)\big)=10804$$

Finally, I confirmed that the dataset contains 10804 rows, ensuring that the dataset is complete.

For consistency and accuracy, I did not approximate any values in the dataset or during the processing of match outcomes. However, for the sake of clarity in presenting results, some final numerical answers are rounded to two decimal places, unless otherwise stated.

Exploration

In this exploration, I will evaluate different methods to quantify football team performance over time, specifically focusing on the La Liga seasons from 1995/96 to 2022/23. Three rating systems: Win Accumulation, Win-Loss Differential, and Elo Ratings, each starting at 1000 points to avoid negative ratings, will be compared to assess which provides the most reliable evaluation of team success. All calculations for the rating systems were performed using Python and applied chronologically to match results, ensuring that each team's rating evolved over time in a realistic and data-driven manner.

Method A: Win Accumulation

Method A is a rating system based purely on the accumulation of wins. It operates on a binary principle where each match contributes either a point for a win or nothing for a loss and a draw. The primary mathematical representation of the system is a cumulative sum:

$$R_i = \sum_{t=1}^{n} S_t$$

Where:

- R_i is the rating for team i,
- S_t is the outcome of the match, which is a fixed value (1 for a win, 0 for a draw, and 0 for a loss)
- n is the total number of matches played by the team.

This model assumes that each match is independent, and the only factor influencing the rating is the binary outcome of a win or not a win, making it an objective measure of success.

Figure 1 represents the cumulative win count of various teams over time using Method A. Each line corresponds to a different team, and the vertical axis indicates the total number of wins accumulated, starting at 1000. Meanwhile, the horizontal axis represents time, spanning from 1995 to 2023.

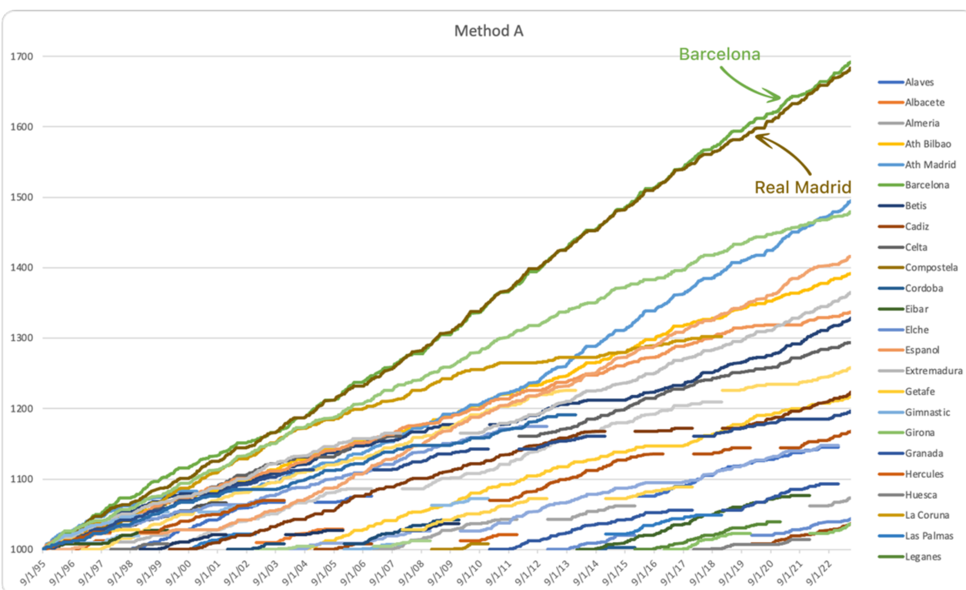


Figure 1: Team Win Trajectories in LaLiga Using Win Accumulation (Method A)

Method A provides an objective, transparent rating by tallying cumulative wins. A steeper slope directly reflects a higher win rate over time. This simplicity clearly identifies the most victorious team (with the highest total wins). This is evident as Barcelona and Real Madrid dominate the top of the graph with steep, uninterrupted trajectories, reflecting their consistently high win rates across the 28-season period. However, their lines are extremely close together, making it difficult to determine which team performed better overall. Furthermore, it fails as a comprehensive rating system because it disregards contextual factors. Ultimately, it serves as a basic win tracker rather than a rating system.

Method B: Win-Loss Differential

Method B is an extension of Method A but improves the simple win count by introducing negative points for losses. It also operates on a ternary principle where each match contributes either a point for a win, but minus a point for a loss, and nothing for a draw. This makes it more refined than Method A, as it differentiates between teams that consistently win and those that suffer many losses. The primary mathematical representation of the system is a cumulative sum:

$$R_i = \sum_{t=1}^{n} S_t$$

Where:

- R_i is the rating for team i,
- S_t is the outcome of the match, which is a fixed value (1 for a win, 0 for a draw, and -1 for a loss)
- n is the total number of matches played by the team.

Method B provides a more dynamic rating than Method A, reflecting not only a team's success but also its failures. Teams with more losses will have a lower rating, even if they have some wins, and teams with an equal number of wins and losses will have a neutral rating.

Figure 2 illustrates the cumulative win-loss count of various teams over time using Method B. Each line represents a different team. The vertical axis shows the net total of wins minus losses, starting at 1000. Meanwhile, the horizontal axis represents time, spanning from 1995 to 2023.

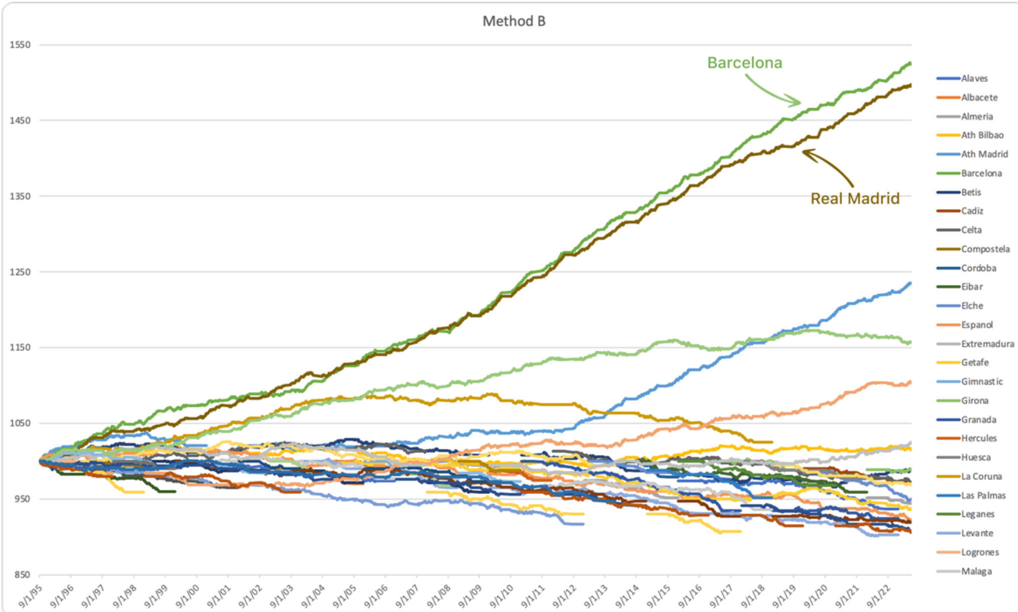


Figure 2: LaLiga Team Rating Evolution via Win-Loss Differential (Method B)

Method B effectively differentiates between strong and weak teams over time, providing a fairer representation by considering both wins and losses. The slope for each team represents its long-term performance trend, where a steeper positive slope indicates consistent success, a downward slope reflects frequent losses, and a horizontal slope shows an equal number of wins and losses. However, Method B overlooks opponent strength, meaning teams that accumulate wins against weaker opponents can seem as strong as those winning against top-tier teams, which reduces rating accuracy. Furthermore, the system only allows for +1, -1, or 0 per match, leading to rating compression, where mid-tier teams become indistinguishable, slowing down rating updates. Barcelona and Real Madrid both show steep, rising curves, confirming their dominance, but their trajectories remain very close, limiting meaningful comparisons between the two.

Method C: Elo Rating System

The Elo rating system is widely used in competitive scenarios like chess or online gaming to measure players' or teams' relative skill levels. It incorporates fairness by adjusting rating points based on the relative strengths of opponents. A strong team beating a weak opponent earns fewer points than a weak team defeating a strong opponent. This fairness is based on the probability of the player winning. The system consists of two equations. The first is used to calculate the probability of winning, and the other is used to update the rating.

Probability equation

Arpad Elo acknowledges that the performance of a player or a team fluctuates due to external factors where stronger players generally perform better but do not always win. To represent this variability, the system takes the assumption that the performance of a team or a player follows a normal distribution (Elo 1976: Section 1.31). A histogram of performance deviations typically forms a bell curve, a normal distribution. Although widely accepted, the normal distribution is not a perfect model for all scenarios, as large deviations occur more often than expected (Edgeworth 1902, Elo 1976: Section 1.34).

Since individual performances are assumed to follow a normal distribution, we write:

$$X_A \sim N(R_A, \sigma^2)$$

$$X_B \sim N(R_B, \sigma^2)$$

Where R_A and R_B represent the rating values (expected skill level) of Players A and B, respectively.

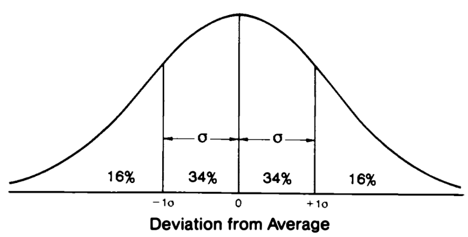


Figure 3: Normal Distribution of Player/Team Performance (Elo, 1978)

However, what determines the outcome of the match is not their absolute performance levels, but the difference between them. Since the difference of two independent normal distributions is itself normally distributed, the performance difference $D = X_A - X_B$ also follows a normal distribution.

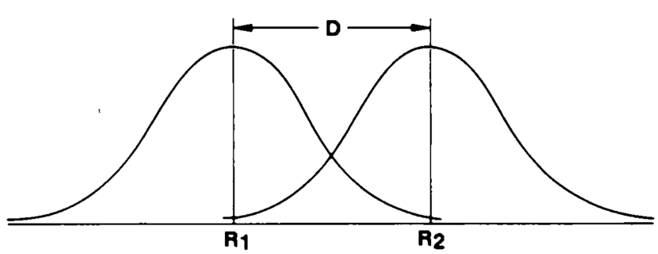


Figure 4: Performance Difference Between Two Teams Based on Elo Distributions

Furthermore, since X_A and X_B are independent, the rule extent to $\mu_A - \mu_B$, and their variances add up (Penn State University, n.d.):

$$Var(X_A - X_B) = Var(X_A) + Var(X_B)$$

Since both players' performance variances are σ^2, we get:

$$Var(D) = \sigma^2 + \sigma^2 = 2\sigma^2$$

Thus, the performance difference D follows a normal distribution:

$$D \sim N(R_A - R_B, 2\sigma^2)$$

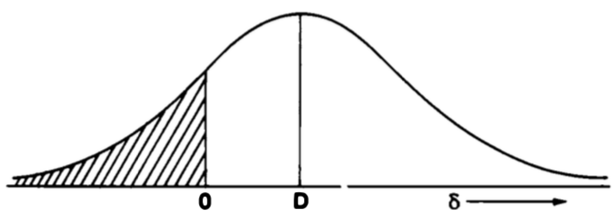


Figure 5: Shaded region representing the probability that Player A (with higher Elo) will outperform Player B, where $D > 0$.

The graph represents the normal distribution of performance differences δ around D. The shaded region represents the probability that a lower-rated team will outperform a higher-rated one.

This tells us that the probability of Player A winning the game is given by:

$$P_A = P(D > 0) = P(X_A > X_B)$$

meaning that we are looking for the probability that Player A's performance exceeds Player B's. Calculating the exact value of this probability requires the cumulative distribution function (CDF) of the normal distribution. However, before the widespread availability of computing tools, evaluating the CDF was computationally challenging, as it does not have a simple closed-form expression. To address this, Arpad Elo approximated the normal distribution using a logistic function, which is easier to compute and closely resembles the bell curve, particularly for typical rating differences in competitive play. Furthermore, the logistic function even better fits large deviations (Elo 1978, Section 8.3).

While mathematicians conventionally use base e in a logistic function, Elo used base $\sqrt{10} \approx 3.162$ because it naturally aligns with the win/loss odds of a 200-point rating difference:

$$\frac{P_{win}}{1 - P_{win}} = \frac{0.76}{0.24} \approx 3.17$$

This choice ensured that the logistic function closely matched the expected probabilities from the normal CDF.

Furthermore, a scaling factor of 200 was selected so that a 200-point rating difference corresponds to a probability of victory of approximately 0.76. Which, again, aligns with the normal distribution having a standard deviation of 200.

Thus, the formula was:

$$P_A = \frac{1}{1 + \sqrt{10}^{(R_A - R_B)/200}}$$

However, using base $\sqrt{10}$ presented the same issue as base e in that it was difficult to compute manually. To simplify calculations while preserving the same probability model, Elo reformulated the exponent using base 10 instead of $\sqrt{10}$. Since:

$$\sqrt{10} = 10^{1/2}$$

a rating difference of 200 points in the original logistic model can be rewritten using base 10 as:

$$P_A = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

This reformulation maintains the same mathematical relationship but is easier to compute manually, particularly when using logarithms or probability tables (*see Appendix I*).

This formula is a very good approximation of the normal CDF.

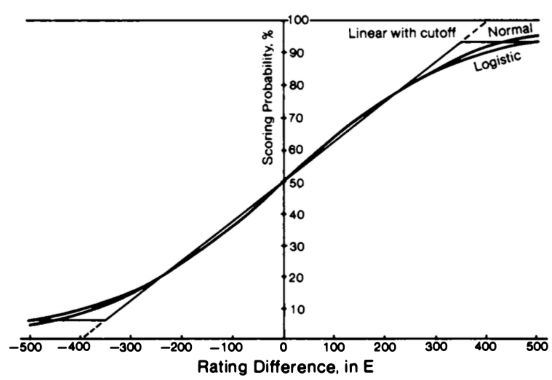


Figure 7: Comparison between logistic function, normal CDF, and linear model (Elo 1978, Section 8.73)

This graph compares probability functions for predicting match outcomes based on rating differences. The normal CDF and the logistic function are very close but deviate in higher rating differences. The linear model is a simpler alternative with a fixed cutoff.

Rating Update Equation

Now that we have determined the probability of a player winning based on their rating difference, we can use this probability to update player ratings after a match. The Elo system adjusts ratings by comparing the actual game outcome with the expected probability, ensuring that ratings reflect player performance over time.

To update the rating, the Elo system uses the formula:

$$R_{new} = R_{old} + K(S - P)$$

Where:

R_{new} is the new rating

R_{old} is the old rating

S is the outcome of the match (1 for a win, 0.5 for a draw, and 0 for a loss).

P is the probability formula found in the previous section.

The constant K in the equation determines the maximum number of points a team can gain or lose after each match. A higher K value gives more importance to recent match results, while a lower K value gives more importance to older ones (Lodder, 2012). Therefore, selecting the best K is crucial for calculating accurate Elo ratings.

The Elo system calculates new ratings by comparing the match outcomes with the predicted results based on the teams' rating differences. It is precisely this comparison that allows K to be optimized (Opisthokonta.net, 2016). The Mean Squared Error (MSE) is then used, a standard statistical measure quantifying the difference between observed and corresponding predicted values. The smaller this difference is, the more accurate are the predictions. Thus, it is needed to tune K so that the MSE is as small as possible. The MSE across n matches is given by (Frost, 2021):

$$\mathrm{MSE} = \frac{\sum(y_i - \hat{y})^2}{n}$$

Where y_i is the match observed value: 1 for a win, 0.5 for a draw, and 0 for a loss, and $\hat{y}$ is the corresponding predicted value of winning.

Using data from the Spanish LaLiga covering the 1995/96 to 2022/23 seasons (a total of 10,804 matches), the K values were tested ranging from 1 to 50 in increments of 0.1 using a custom Python script (*see Appendix IV for the complete code*). Plotting these values against their corresponding MSE revealed that $K = 9.4$ minimizes the MSE to 0.16657.

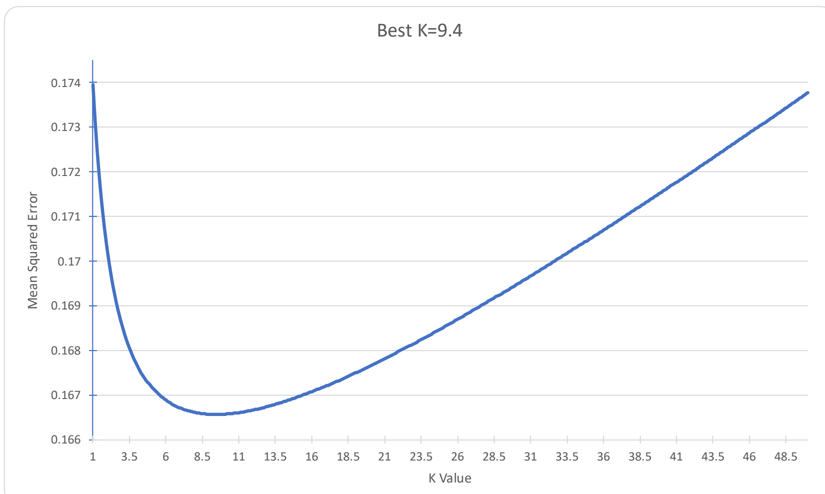


Figure 8: Mean Squared Error vs K-Value for LaLiga Predictions

Having identified $K = 9.4$ as the value that minimizes the MSE, I applied this optimal K-factor to the entire LaLiga dataset using a custom Python script (*see Appendix V for the complete code*). Figure 9 shows the resulting Elo ratings over time for each team, starting at an initial rating of 1000. The vertical axis displays the Elo rating, while the horizontal axis represents the progression of the years. Each line corresponds to a different club progression.

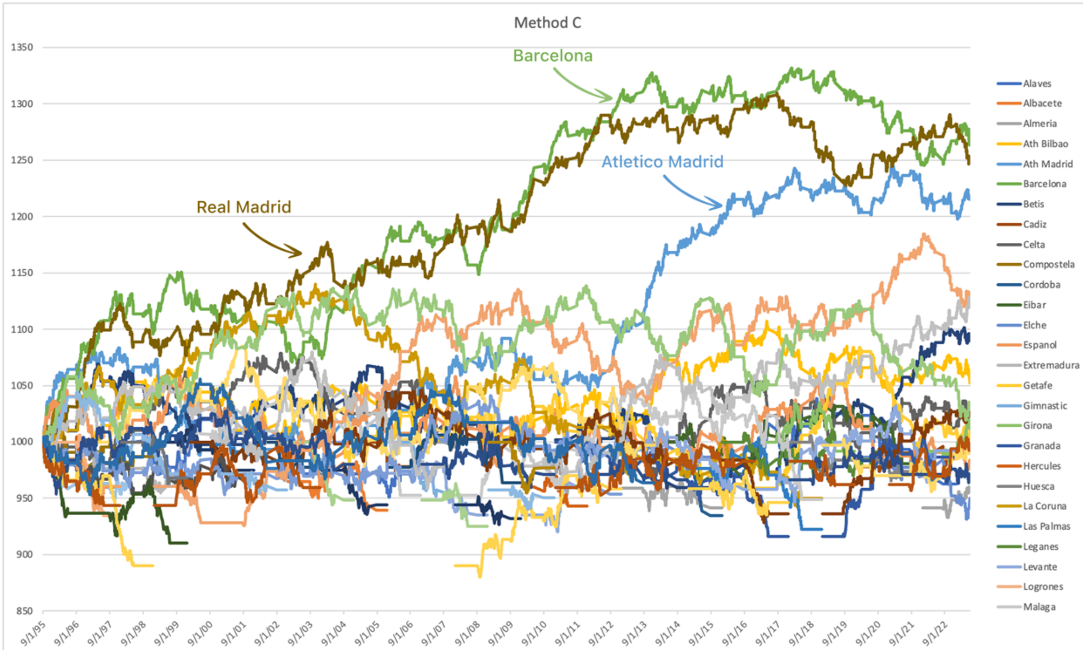


Figure 9: Method C

In the first few years of the dataset, the graph shows that the system takes time to "learn", but as the matches are played, the ratings gradually stabilize. A higher K value could be set initially and gradually reduced to minimize this learning time. The graph also shows that most teams remain clustered within the 950–1050 range, reflecting a relatively balanced competitive level, while a few elite teams gradually separate, and weaker teams go below this range. However, the weaker teams appear to deviate less from this cluster than the elite team, probably because they can get relegated to the "Segunda División", the second Spanish football division, and therefore cannot lose as many matches as the elite teams can win. The graph clearly highlights the dominance of Barcelona and Real Madrid, which have consistently been at the top since the mid-2000s. Barcelona peaked twice in the mid 1300 points range, corresponding to the Guardiola era (2008–12), often considered as the greatest teams in football history, and the Luis Enrique period (2014–17), when they won multiple titles (White, 2023) (*see Appendix II*). Meanwhile, Real Madrid maintained a high Elo rating, peaking in the low 1300 points range between 2016–18, aligning with their historic three consecutive UEFA Champions League titles (*see Appendix II*). A notable event in this dataset is Atlético Madrid's surge from 2012 onward, climbing from the 950–1050 cluster to a peak in 2020 (*see Appendix II*). This shift coincides with their actual coach Diego Simeone's historical arrival in December 2011, under whom Atlético won LaLiga in 2013/14 and 2020/21, disrupting the long-standing Barcelona–Real Madrid duopoly (Transfermarkt, 2025). Thus, the graph illustrates how external factors such as managerial changes can drive significant performance shifts

Evaluation

Strengths

This system has several strengths. Unlike traditional points-based leagues that reset each season, the Elo system preserves a team's historical performance while still reflecting current trends. Another strength is the consideration given to the difference in level between two teams facing each other, which makes the rating progression and the result of each match more indicative. Finally, it also captures changes in team performance, as seen with Atlético Madrid's surge in the 2010s and the gradual decline of former strong teams like Valencia.

Limitations

Despite its advantages, the Elo system has notable limitations in football. First, the system does not account for promotion and relegation, meaning newly promoted teams start at 1000 points, the same as established LaLiga clubs, which overestimates their actual strength. A more refined approach could assign them a lower initial rating, for instance, 950, as it aligns with the lower end of the typical team cluster, allowing for a more realistic progression. Additionally, the home-field advantage is not accounted for in this model. Other adaptations, such as FiveThirtyEight's NBA Elo system, incorporate a home-court advantage multiplier set at 100 points. This adjustment reflects how home teams generally perform better due to familiar conditions, crowd support, and travel fatigue for opponents. FiveThirtyEight found that constant 100-point boost was a simple yet effective way to improve prediction accuracy (Fischer-Baum, 2015). A similar adjustment in football Elo models could enhance accuracy. The model also does not directly factor in external influences such as injuries, transfers, tactical changes, or managerial shifts, though their effects are indirectly reflected in match results, for example, the visible rating drops of Real Madrid after Cristiano Ronaldo's 2018 departure and Barcelona after Lionel Messi left in 2021. Finally, the system excludes non-LaLiga competitions, meaning performances in continental tournaments like the UEFA Champions League and domestic cups like the Copa del Rey are not considered. Some might argue that Real Madrid's dominance in the Champions League by winning five titles between 2014 and 2022 should be reflected in their Elo rating, as it showcases their true strength beyond domestic matches. However, incorporating these competitions would require integrating all European leagues, which is impossible at my scale.

Conclusion

After evaluating the three rating methods: win accumulation, win-loss differential, and the Elo rating system, I found that while the simpler models provided a basic understanding of team performance, they fell short in capturing the complexities of football matches. The Elo rating system, by adjusting ratings based on the strength of opponents, offered the most nuanced and reliable reflection of long-term team performance.

The win accumulation method overlooked losses, while the win-loss differential still didn't account for the quality of opposition, leading to less accurate comparisons. In contrast, the Elo system's ability to adapt based on match outcomes and the relative strength of teams resulted in a more dynamic and comprehensive evaluation. Despite its limitations, such as not considering other factors such as home-field advantage or non-LaLiga competitions, the Elo method was the most effective in answering the question of which club, Barcelona or Real Madrid, has historically performed better in LaLiga.

In the context of LaLiga, Barcelona's Elo ratings consistently showed higher peaks compared to Real Madrid, particularly during the Guardiola and Luis Enrique eras. Despite Real Madrid's consistent strength and recent dominance in European competitions (which my Elo system doesn't incorporate), Barcelona's performance in LaLiga has reached greater peaks. Therefore, my final conclusion is that Barcelona surpasses Real Madrid, primarily due to its higher peaks and consistent superiority in terms of ratings within the context of Spanish LaLiga.

Extension

Building on the LaLiga analysis, one possible extension is to broaden the application of the Elo rating system beyond LaLiga. While this IA evaluates team performance within a single league, the Elo model can be adapted to include data from continental cups, such as the UEFA Champions League, to assess a global rating of clubs. Additionally, incorporating other European leagues, such as the English Premier League, Italian Serie A, and German Bundesliga, would allow for a broader comparison for Barcelona and Real Madrid, or perhaps even reveal higher peaks achieved by other teams.

Moreover, while Elo is widely used in football ratings, other systems offer different advantages. Developed by Mark Glickman, the Glicko system introduces a ratings deviation metric (RD), which quantifies the uncertainty in team ratings based on match frequency and recency. It also incorporates score margins, improving predictive accuracy (Glickman, 2016). Furthermore, developed by Alec Stephenson in 2012, the Stephenson model is an extension of the Glicko system and ranks teams using eigenvector centrality, meaning it considers indirect victories (if A beats B and B beats C, A is inferred to be stronger than C), and it applies temporal decay, discarding older matches (Glickman & Hennessy, n.d.).

In the context of machine learning, the Elo rating system is widely used to dynamically rank AI models through pairwise comparisons, enabling continuous performance updates. For instance, Yan et al. (2022) have applied it in dueling bandit frameworks to identify top-performing models while reducing computational costs. It has also been used to rank large language models (LLMs) like GPT-4 and Claude based on human preferences. However, Elo struggles with rating stability when models perform similarly, necessitating hundreds of comparisons for accurate results (Boubdir et al., 2023).

Bibliography

Boubdir, M., Kim, E., Ermis, B., Hooker, S., & Fadaee, M. (2023). *Elo Uncovered: Robustness and Best Practices in Language Model Evaluation Cohere for AI* (pp. 339–352). https://aclanthology.org/2023.gem-1.28.pdf

Elo, A. E. (1978). *The Rating of Chessplayers, Past and Present*.

Fischer-Baum, R. (2015, May 21). *How We Calculate NBA Elo Ratings*. FiveThirtyEight. https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/#

Football-Data.co.uk. (n.d.). *Spain Football Results and Betting Odds*. Www.football-Data.co.uk. Retrieved January 10, 2025, from https://www.football-data.co.uk/spainm.php

Frost, J. (2021, November 12). *Mean Squared Error (MSE)*. Statistics by Jim. https://statisticsbyjim.com/regression/mean-squared-error-mse/

Glickman, M. (2016). The Glicko system. In *Glicko.net*. Harvad University. http://www.glicko.net/research/acjpaper.pdf

Glickman, M., & Hennessy, J. (n.d.). *A COMPARISON OF RATING SYSTEMS FOR COMPETITIVE WOMEN'S BEACH VOLLEYBALL*. Retrieved February 18, 2025, from http://www.glicko.net/research/volleyball-FINAL.pdf

Kumar, K. (2023). *La Liga Complete Dataset*. Kaggle.com. https://www.kaggle.com/datasets/kishan305/la-liga-results-19952020/data

Lodder, P. (2012). *The Use of the K-factor in Estimating Individual Ability Advanced Study in Individual Differences K-factor*. https://www.paultwin.com/wp-content/uploads/Lodder-P.-2012-The-Use-of-the-K-factor-in-Estimating-Individual-Ability.pdf

Mitton, A. (2024, April 5). *Explaining how the Barcelona and Real Madrid rivalry began*. GiveMeSport. https://www.givemesport.com/how-barcelona-real-madrid-rivalry-el-clasico-began-explained/

Newcastle University. (n.d.). *Numeracy, Maths and Statistics - Academic Skills Kit*. Www.ncl.ac.uk. Retrieved February 13, 2025, from https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/distribution-functions/cumulative-distribution-function.html

Opisthokonta.net. (2016, August 12). *Tuning the Elo ratings: The K-factor and home field advantage | opisthokonta.net*. Opisthokonta. https://opisthokonta.net/?p=1387

Penn State University. (n.d.). *7.1 - Difference of Two Independent Normal Variables*. Psu.edu. Retrieved February 13, 2025, from https://online.stat.psu.edu/stat500/book/export/html/572

Transfermarkt. (2025). *Atlético de Madrid - Club achievements*. Transfermarkt.com. https://www.transfermarkt.com/atletico-de-madrid/erfolge/verein/13

White, M. (2023, November 30). *Ranked! The 10 best Barcelona sides ever*. Fourfourtwo.com; FourFourTwo. https://www.fourfourtwo.com/features/ranked-the-10-best-barcelona-sides-ever

Yan, X., Du, Y., Ru, B., Wang, J., Zhang, H., & Chen, X. (2022). *Learning to Identify Top Elo Ratings: A Dueling Bandits Approach*. ArXiv.org. https://arxiv.org/abs/2201.04480

Appendices

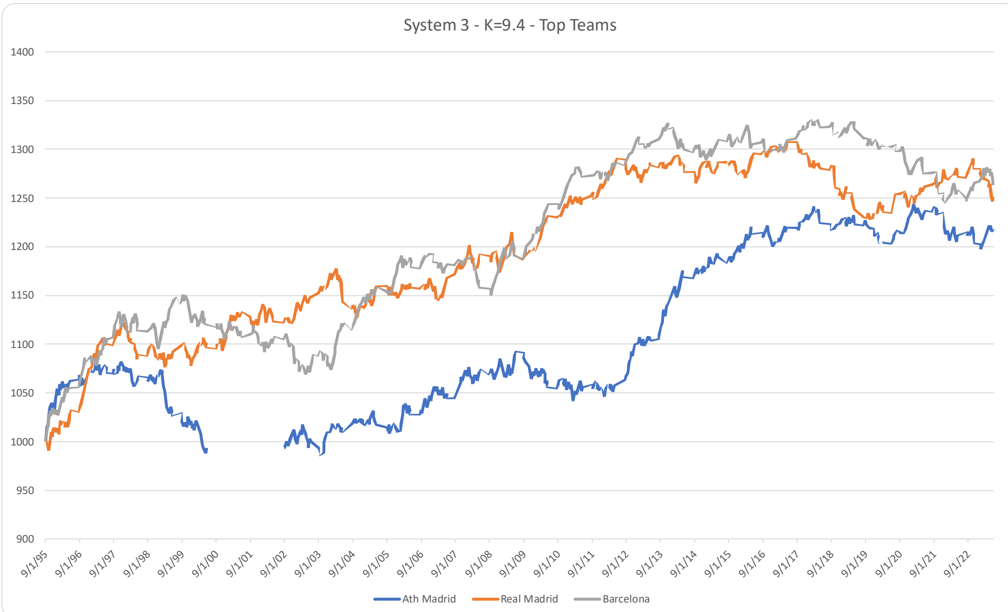
Appendix I: Logistic Probabilities to Base $\sqrt{10}$ Table (Elo 1978, Section 8.46)

Logistic Probabilities to Base $\sqrt{10}$

D	P		D	P		D	P	
Rtg. Diff	H	L	Rtg. Dif.	H	L	Rtg. Dif.	H	L
0-3	.50	.50	120-127	.67	.33	282-294	.84	.16
4-10	.51	.49	128-135	.68	.32	295-308	.85	.15
11-17	.52	.48	136-143	.69	.31	309-323	.86	.14
18-24	.53	.47	144-151	.70	.30	324-338	.87	.13
25-31	.54	.46	152-159	.71	.29	339-354	.88	.12
32-38	.55	.45	160-168	.72	.28	355-372	.89	.11
39-45	.56	.44	169-177	.73	.27	373-391	.90	.10
46-52	.57	.43	178-186	.74	.26	392-412	.91	.09
53-59	.58	.42	187-195	.75	.25	413-436	.92	.08
60-66	.59	.41	196-205	.76	.24	437-463	.93	.07
67-74	.60	.40	206-214	.77	.23	464-494	.94	.06
75-81	.61	.39	215-224	.78	.22	495-530	.95	.05
82-88	.62	.38	225-235	.79	.21	531-576	.96	.04
89-96	.63	.37	236-246	.80	.20	577-636	.97	.03
97-103	.64	.36	247-257	.81	.19	637-726	.98	.02
104-111	.65	.35	258-269	.82	.18	727-920	.99	.01
112-119	.66	.34	270-281	.83	.17	Over 920	1.00	00

H is the probability of scoring for the higher rated player, and L for the lower.

Appendix II: Elo Rating Progression of Top LaLiga Teams (1995–2023)



Appendix III: The First 50 And Last 50 Rows of The Dataset

Match Number	Date	Team A	Team B	Result A	Result B
1	1995-09-02	Albacete	Sp Gijon	L	W
2	1995-09-02	La Coruna	Valencia	W	L
3	1995-09-03	Ath Bilbao	Santander	W	L
4	1995-09-03	Ath Madrid	Sociedad	W	L
5	1995-09-03	Barcelona	Valladolid	W	L
6	1995-09-03	Betis	Merida	D	D
7	1995-09-03	Celta	Compostela	L	W

8	1995-09-03	Espanol	Salamanca	W	L
9	1995-09-03	Oviedo	Zaragoza	L	W
10	1995-09-03	Real Madrid	Vallecano	W	L
11	1995-09-03	Sevilla	Tenerife	L	W
12	1995-09-09	Albacete	Sevilla	W	L
13	1995-09-09	Ath Bilbao	Real Madrid	W	L
14	1995-09-09	Barcelona	Merida	D	D
15	1995-09-09	Betis	Zaragoza	W	L
16	1995-09-10	Ath Madrid	Santander	W	L
17	1995-09-10	Celta	Salamanca	W	L
18	1995-09-10	Compostela	La Coruna	W	L
19	1995-09-10	Espanol	Tenerife	W	L
20	1995-09-10	Oviedo	Vallecano	W	L
21	1995-09-10	Sociedad	Sp Gijon	W	L
22	1995-09-10	Valencia	Valladolid	W	L
23	1995-09-16	Ath Bilbao	Ath Madrid	L	W
24	1995-09-17	Albacete	Espanol	L	W
25	1995-09-17	Barcelona	Zaragoza	W	L
26	1995-09-17	Betis	Vallecano	W	L
27	1995-09-17	Celta	Tenerife	D	D
28	1995-09-17	Compostela	Valladolid	D	D
29	1995-09-17	La Coruna	Salamanca	W	L
30	1995-09-17	Merida	Valencia	L	W
31	1995-09-17	Oviedo	Real Madrid	W	L
32	1995-09-17	Santander	Sp Gijon	L	W
33	1995-09-17	Sevilla	Sociedad	W	L

34	1995-09-23	Barcelona	Vallecano	W	L
35	1995-09-23	Betis	Real Madrid	D	D
36	1995-09-23	Santander	Sevilla	D	D
37	1995-09-24	Albacete	Celta	W	L
38	1995-09-24	Ath Bilbao	Oviedo	D	D
39	1995-09-24	Ath Madrid	Sp Gijon	W	L
40	1995-09-24	Compostela	Merida	W	L
41	1995-09-24	Espanol	Sociedad	W	L
42	1995-09-24	La Coruna	Tenerife	D	D
43	1995-09-24	Salamanca	Valladolid	D	D
44	1995-09-24	Valencia	Zaragoza	D	D
45	1995-09-30	Ath Bilbao	Sp Gijon	W	L
46	1995-09-30	Barcelona	Real Madrid	D	D
47	1995-10-01	Albacete	La Coruna	L	W
48	1995-10-01	Ath Madrid	Sevilla	D	D
49	1995-10-01	Betis	Oviedo	W	L
50	1995-10-01	Celta	Sociedad	D	D
...	...	...	...	...	...
10754	2023-05-04	Valladolid	Vallecano	L	W
10755	2023-05-12	Cadiz	Mallorca	L	W
10756	2023-05-13	Almeria	Osasuna	L	W
10757	2023-05-13	Ath Bilbao	Villareal	L	W
10758	2023-05-13	Getafe	Real Madrid	L	W
10759	2023-05-13	Girona	Sociedad	D	D
10760	2023-05-14	Ath Madrid	Elche	L	W
10761	2023-05-14	Barcelona	Espanol	W	L

10762	2023-05-14	Celta	Valencia	L	W
10763	2023-05-14	Sevilla	Valladolid	W	L
10764	2023-05-15	Betis	Vallecano	W	L
10765	2023-05-19	Cadiz	Valladolid	W	L
10766	2023-05-20	Almeria	Mallorca	W	L
10767	2023-05-20	Ath Bilbao	Celta	W	L
10768	2023-05-20	Barcelona	Sociedad	L	W
10769	2023-05-20	Elche	Getafe	D	D
10770	2023-05-20	Girona	Villareal	L	W
10771	2023-05-21	Ath Madrid	Osasuna	W	L
10772	2023-05-21	Betis	Sevilla	D	D
10773	2023-05-21	Espanol	Vallecano	W	L
10774	2023-05-21	Real Madrid	Valencia	L	W
10775	2023-05-23	Almeria	Sociedad	L	W
10776	2023-05-23	Barcelona	Valladolid	L	W
10777	2023-05-23	Celta	Girona	D	D
10778	2023-05-24	Ath Madrid	Espanol	D	D
10779	2023-05-24	Betis	Getafe	L	W
10780	2023-05-24	Cadiz	Villareal	L	W
10781	2023-05-24	Elche	Sevilla	D	D
10782	2023-05-24	Real Madrid	Vallecano	W	L
10783	2023-05-25	Ath Bilbao	Osasuna	L	W
10784	2023-05-25	Mallorca	Valencia	W	L
10785	2023-05-27	Real Madrid	Sevilla	W	L
10786	2023-05-28	Almeria	Valladolid	D	D
10787	2023-05-28	Ath Bilbao	Elche	L	W
10788	2023-05-28	Ath Madrid	Sociedad	W	L

10789	2023-05-28	Barcelona	Mallorca	W	L
10790	2023-05-28	Betis	Girona	W	L
10791	2023-05-28	Cadiz	Celta	W	L
10792	2023-05-28	Espanol	Valencia	D	D
10793	2023-05-28	Getafe	Osasuna	W	L
10794	2023-05-28	Vallecano	Villareal	W	L
10795	2023-06-04	Almeria	Espanol	D	D
10796	2023-06-04	Ath Bilbao	Real Madrid	D	D
10797	2023-06-04	Ath Madrid	Villareal	D	D
10798	2023-06-04	Barcelona	Celta	L	W
10799	2023-06-04	Betis	Valencia	D	D
10800	2023-06-04	Cadiz	Elche	D	D
10801	2023-06-04	Getafe	Valladolid	D	D
10802	2023-06-04	Girona	Osasuna	L	W
10803	2023-06-04	Mallorca	Vallecano	W	L
10804	2023-06-04	Sevilla	Sociedad	L	W

Appendix IV: Custom Python Script for Optimizing the Elo *K*-Factor Using MSE

```
import pandas as pd
import numpy as np


def calc_elo(df, k):
    ratings = {}
    errors = []

    df = df.sort_values('Date').copy()

    for _, row in df.iterrows():
        elo_a = ratings.get(row['Team A'], 1000)
        elo_b = ratings.get(row['Team B'], 1000)

        exp = (elo_a - elo_b) / 400
        prob_a = 1 / (1 + 10 ** (-exp))

        if row['Result A'] == 'W':
            result = 1
        elif row['Result A'] == 'D':
            result = 0.5
        else:
            result = 0

        errors.append((result - prob_a) ** 2)
```

```
        delta = k * (result - prob_a)
        ratings[row['Team A']] = elo_a + delta
        ratings[row['Team B']] = elo_b - delta

    return np.mean(errors)

def find_k(df, start=1, end=50, step=1):
    k_range = np.arange(start, end + step, step)
    results = {}

    for k in k_range:
        results[k] = calc_elo(df, k)

        if len(results) % 10 == 0:
            print(f"Tested {len(results)} K values")
            best_k = min(results, key=results.get)
            print(f"Current best: K={best_k}, MSE={results[best_k]:.4f}")

    best_k = min(results, key=results.get)
    return best_k, results

if __name__ == '__main__':
    df = pd.read_csv('matches.csv')
    best_k, results = find_k(df)

    print(f"\nBest K: {best_k}")
```

```
    print(f"Best MSE: {results[best_k]:.4f}")

    pd.DataFrame({
        'K': list(results.keys()),
        'MSE': list(results.values())
    }).to_csv('mse_results.csv', index=False)
```

Appendix V: Custom Python Implementation of the Elo Rating System for LaLiga Match Results

```
import pandas as pd

def get_match_result(team_score, opp_score):
    if team_score > opp_score:
        return 1, 0
    elif team_score < opp_score:
        return 0, 1
    return 0.5, 0.5

def calc_win_prob(rating_a, rating_b):
    return 1 / (1 + 10 ** ((rating_b - rating_a) / 400))

def process_matches(file_path, k_factor=9.4, initial_elo=1000):
    df = pd.read_csv(file_path)
    df['Date'] = pd.to_datetime(df['Date'])
    df.columns = df.columns.str.replace(' ', '_')
```

```
    teams = set(df['Team']).union(df['Opponent'])
    ratings = {team: initial_elo for team in teams}

    elo_timeline = [{
        'Timeline': pd.Timestamp('1995-09-01'),
        **{team: initial_elo for team in teams}
    }]

    elo_before = []
    win_probs = []
    elo_changes = []

    for match in df.sort_values('Date').itertuples(index=False):
        team_a, team_b = match.Team, match.Opponent
        rating_a = ratings[team_a]
        rating_b = ratings[team_b]

        win_prob = calc_win_prob(rating_a, rating_b)
        result_a, result_b = get_match_result(match.Team_Score,
match.Opponent_Score)

        elo_change = k_factor * (result_a - win_prob)

        ratings[team_a] += elo_change
        ratings[team_b] -= elo_change
```

```
        elo_before.append(rating_a)

        win_probs.append(win_prob)

        elo_changes.append(elo_change)


        if not elo_timeline or elo_timeline[-1]['Timeline'] !=
match.Date:

            elo_timeline.append({

                'Timeline': match.Date,

                **ratings

            })


    df['Elo_Before'] = elo_before

    df['Win_Probability'] = win_probs

    df['Elo_Change'] = elo_changes


    timeline_df = pd.DataFrame(elo_timeline)


    df.to_csv('Updated_Match_Data.csv', index=False)

    timeline_df.to_csv('Elo_Timeline.csv', index=False)


if __name__ == '__main__':

    process_matches("Math IA - Clean Full Data.csv")
```