

Model Prediksi Hasil Pertandingan Sepak Bola Antara PSG dan Inter Milan dengan Regresi Logistik Multinomial dan Binomial

Teguh Wandu Maulana, Andi Muhammad Faizal Fachriansyah, Latief Surya Rabbani, Reza Halim Qodari, Mellita

Universitas Teknologi Bandung, Bandung, Indonesia

Correspondence : e-mail: teguhwandi11@gmail.com

Abstrak

Penelitian ini bertujuan memprediksi hasil pertandingan final Liga Champions antara PSG dan Inter Milan dengan pendekatan statistik menggunakan regresi logistik multinomial dan binomial. Data diambil dari situs FBref.com dan mencakup jumlah gol, kebobolan, dan penguasaan bola dari kedua tim selama satu musim. Hasil analisis regresi binomial menunjukkan bahwa jumlah gol merupakan indikator paling signifikan terhadap kemenangan, sedangkan penguasaan bola justru menunjukkan korelasi negatif. Sebaliknya, model multinomial menghasilkan output yang tidak logis akibat keterbatasan data dan kemungkinan kesalahan pengkodean. Temuan ini menegaskan pentingnya efektivitas serangan dibanding sekadar dominasi permainan. Penelitian ini dapat menjadi referensi bagi pelatih dan analis dalam menyusun strategi serta dasar untuk pengembangan studi analitik olahraga ke depan.

Kata kunci: analitik sepak bola, prediksi pertandingan, regresi logistik, PSG, Inter Milan.

Abstract

This study aims to predict the outcome of the UEFA Champions League final match between PSG and Inter Milan using statistical analysis through multinomial and binomial logistic regression. Data were collected from FBref.com, including goals scored, goals conceded, and ball possession for both teams over one season. The binomial model showed that the number of goals is the most significant indicator of victory, while ball possession had a negative correlation with winning. In contrast, the multinomial model produced illogical results, likely due to limited data and coding issues. These findings highlight that attacking effectiveness is more critical than ball dominance. This research offers practical insights for coaches and analysts and serves as a foundation for further development in sports analytics.

Keywords: analytics football, match prediction, logistic regression, PSG, Inter Milan.

1. Pendahuluan

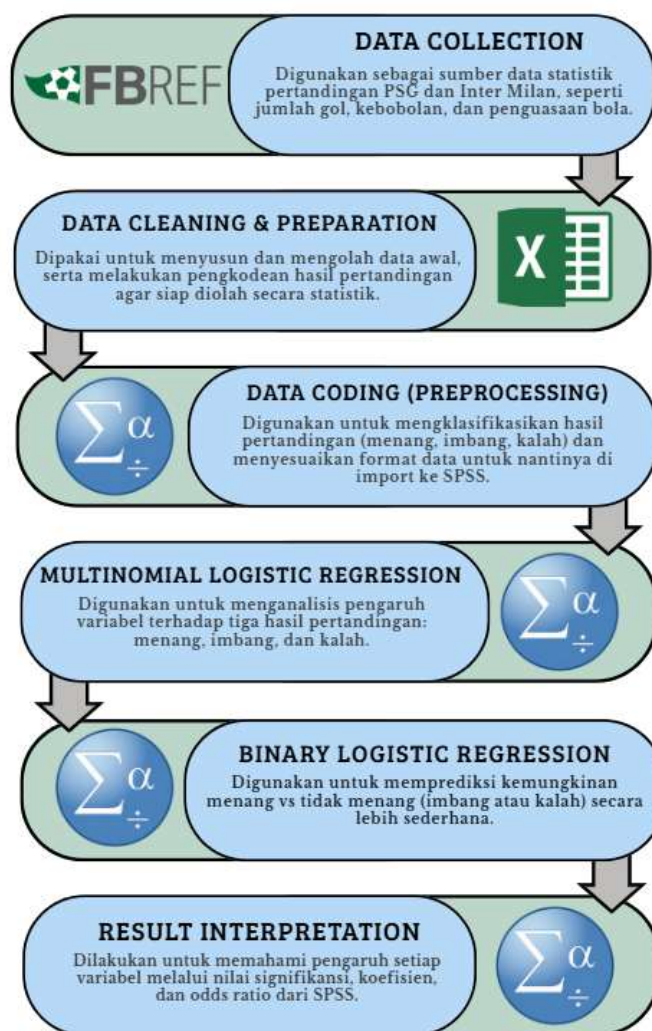
Final Liga Champions 2025 mempertemukan Paris Saint-Germain (PSG) dan Inter Milan dalam laga puncak yang telah menarik perhatian luas dari publik sepak bola dunia [7]. PSG tampil sebagai salah satu tim unggulan dengan performa konsisten dan kedalaman skuad yang memungkinkan mereka melakukan rotasi tanpa kehilangan kualitas permainan [6]. Di sisi lain, Inter Milan dikenal sebagai tim dengan pertahanan solid serta pengalaman dalam laga-laga besar yang menjadi ciri khas permainan mereka [9]. Penelitian sebelumnya menunjukkan bahwa penguasaan bola bukanlah penentu utama kemenangan dalam sepak bola. Dalam buku *The Numbers Game* menyatakan bahwa tim dengan dominasi penguasaan bola tidak selalu keluar sebagai pemenang jika tidak mampu menyelesaikan peluang menjadi gol [1]. Temuan ini diperkuat bahwa meskipun penguasaan bola menggambarkan dominasi, faktor efektivitas serangan dan konversi peluang lebih penting dalam mempengaruhi hasil pertandingan [8]. Hal ini mengindikasikan bahwa analisis statistik berbasis data kuantitatif memiliki peran penting dalam memahami dinamika permainan sepak bola modern.

Penelitian ini dilakukan dengan pendekatan kuantitatif menggunakan data sekunder yang diperoleh dari situs FBref.com. Data yang dianalisis mencakup statistik performa PSG dan Inter Milan

sepanjang musim, meliputi jumlah gol (Gf), jumlah kebobolan (Ga), serta persentase penguasaan bola (Poss). Analisis data dilakukan menggunakan perangkat lunak IBM SPSS Statistics versi 19 dengan metode regresi logistik multinomial dan biner untuk mengevaluasi hubungan antar variabel terhadap hasil pertandingan. Tujuan dari penelitian ini adalah untuk memberikan prediksi hasil pertandingan Final Liga Champions secara objektif berdasarkan performa historis kedua tim. Selain itu, penelitian ini juga bertujuan mengidentifikasi indikator-indikator kunci yang paling berpengaruh terhadap peluang kemenangan. Dengan pendekatan statistik yang digunakan, hasil dari penelitian ini diharapkan dapat berkontribusi terhadap pengembangan studi analitika olahraga, khususnya dalam bidang sepak bola.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder dari situs statistik sepak bola FBref.com. Data yang dianalisis berasal dari seluruh pertandingan Paris Saint-Germain (PSG) dan Inter Milan selama satu musim kompetisi, mencakup laga kandang maupun tandang. Variabel utama yang digunakan dalam analisis ini adalah jumlah gol yang dicetak (Goals For/Gf), jumlah kebobolan (Goals Against/Ga), dan persentase penguasaan bola (Possession). Pendekatan kuantitatif berbasis data sekunder ini umum digunakan dalam analisis performa tim olahraga karena memungkinkan interpretasi statistik yang objektif dan terukur [5]. Langkah-langkah penelitian dirangkum dalam diagram alur berikut:



Gambar 1. Diagram Alur Analisis Data

Setelah data dikumpulkan, dilakukan praproses, yaitu pengecekan, perapian, dan pengaturan format data agar dapat diolah dengan SPSS. Data kemudian diklasifikasikan ke dalam dua jenis variabel: variabel independen (Gf, Ga, Poss) dan variabel dependen berupa hasil pertandingan. Untuk hasil pertandingan, digunakan dua model analisis. Model pertama adalah regresi logistik multinomial, yang membagi hasil menjadi tiga kategori: menang, imbang, dan kalah. Model kedua adalah regresi logistik

binomial, yang menyederhanakan hasil menjadi dua kategori: menang dan tidak menang (gabungan imbang dan kalah).

Pengolahan data dilakukan dengan bantuan perangkat lunak IBM SPSS Statistics versi 19. Hasil analisis akan dilihat berdasarkan nilai signifikansi (Sig.), nilai koefisien regresi (B), dan nilai odds ratio (Exp(B)) untuk mengetahui seberapa besar pengaruh setiap variabel terhadap hasil pertandingan. Analisis dilakukan dengan tingkat signifikansi 5% ($\alpha = 0,05$) sebagai dasar dalam pengambilan kesimpulan statistik.

3. Hasil dan Pembahasan

Penelitian ini menggunakan dua jenis regresi logistik untuk melihat pengaruh jumlah gol, jumlah kebobolan, dan penguasaan bola terhadap hasil pertandingan Paris Saint-Germain (PSG) dan Inter Milan. Dua model yang digunakan yaitu regresi logistik multinomial dan binomial. Ringkasan performa dari kedua tim dapat dilihat pada tabel berikut, berdasarkan data yang telah dikumpulkan:

Tabel 1. Data Pertandingan PSG (Tim 1) dan Inter Milan (Tim 2) di Seluruh Kompetisi

Hasil	Gf	Ga	Poss	Penonton	Tim	Hasil	Gf	Ga	Poss	Penonton	Tim
3	4	1	76	23.167	1	1	2	2	68	33.300	2
3	6	0	68	46.000	1	3	2	0	61	70.921	2
3	3	1	58	45.434	1	3	4	1	55	71.745	2
3	3	1	67	47.000	1	1	1	1	64	14.541	2
3	1	0	64	39.951	1	1	0	0	40	50.922	2
1	1	1	78	20.220	1	0	1	2	57	75.366	2
3	3	1	61	47.000	1	3	2	1	53	24.611	2
0	0	2	65	60.103	1	3	3	0	58	56.391	2
1	1	1	71	30.247	1	3	2	0	80	73.348	2
3	4	2	55	47.590	1	3	3	1	58	64.119	2
1	1	1	61	47.511	1	3	1	0	48	35.201	2
3	3	0	74	66.115	1	1	3	3	41	75.020	2
3	1	0	68	47.926	1	3	1	0	57	15.706	2
0	1	2	71	47.369	1	3	5	1	55	73.001	2
3	4	2	70	17.212	1	3	2	1	38	75.222	2
3	3	0	72	47.000	1	1	1	1	50	72.951	2
0	0	1	43	75.000	1	3	2	1	69	25.707	2
1	1	1	84	47.926	1	0	1	2	47	63.174	2
1	0	0	73	17.028	1	3	2	0	57	71.879	2
3	3	0	70	25.065	1	3	1	0	40	38.100	2
3	3	1	68	47.000	1	3	4	1	52	60.582	2
3	4	2	60	12.095	1	3	3	2	57	53.576	2
1	1	1	67	38.182	1	3	2	0	62	73.204	2
3	1	0	60	39.682	1	3	3	0	65	16.412	2
3	2	1	59	47.580	1	3	2	0	52	16.896	2
3	4	2	83	12.939	1	3	2	3	56	24.841	2
3	2	1	66	38.120	1	3	1	0	59	12.048	2
3	4	2	63	47.818	1	1	2	2	48	63.074	2
1	1	1	78	47.862	1	3	3	1	68	61.230	2
3	4	1	58	60.000	1	3	1	0	69	18.010	2
3	5	2	65	15.040	1	3	4	0	51	27.223	2

Hasil	Gf	Ga	Poss	Penonton	Tim	Hasil	Gf	Ga	Poss	Penonton	Tim
3	2	0	71	24.254	1	3	3	0	64	71.601	2
3	4	1	63	47.926	1	1	1	1	63	75.493	2
3	3	0	62	15.831	1	0	0	3	71	22.351	2
3	1	0	76	31.703	1	3	2	1	62	67.484	2
3	7	0	68	47.211	1	0	0	1	58	41.361	2
3	3	2	65	57.777	1	3	1	0	62	68.271	2
3	7	0	74	25.915	1	3	2	0	51	53.333	2
3	4	1	58	47.838	1	1	1	1	38	53.851	2
0	0	1	70	47.511	1	3	2	0	58	43.789	2
3	4	1	68	28.550	1	3	3	2	69	66.344	2
3	1	0	54	59.765	1	3	2	1	55	55.536	2
3	3	1	60	47.000	1	3	2	0	51	23.215	2
3	6	1	53	37.108	1	3	2	1	60	71.924	2
3	4	2	76	45.175	1	1	1	1	63	67.595	2
3	1	0	81	47.736	1	1	2	2	63	22.126	2
3	3	1	75	47.681	1	3	2	1	42	75.000	2
0	2	3	65	42.535	1	3	3	1	61	71.799	2
3	2	1	70	47.113	1	1	2	2	37	75.625	2
1	1	1	74	34.420	1	0	0	1	52	31.978	2
0	1	3	75	47.597	1	0	0	3	56	75.552	2
3	1	0	51	59.664	1	0	0	1	69	70.240	2
0	1	2	64	19.522	1	1	3	3	39	50.314	2
3	2	1	46	47.511	1	3	1	0	69	70.222	2
3	4	1	69	16.787	1	3	4	3	30	75.504	2
3	3	1	66	47.926	1	3	2	0	58	26.537	2
3	3	0	76	77.101	1	1	2	2	56	71.060	2
3	1	0	51	59.664	1	3	2	0	60	10.656	2

Dari tabel tersebut terlihat bahwa PSG mencetak gol lebih banyak dan memiliki penguasaan bola yang lebih tinggi, sedangkan Inter Milan lebih sedikit kebobolan. Ini menunjukkan bahwa PSG lebih kuat dalam menyerang, sementara Inter lebih unggul dalam bertahan. Jumlah pertandingan yang dianalisis untuk masing-masing tim pun seimbang, sehingga analisis dapat dilakukan secara adil.

Pada model regresi logistik multinomial, hasil pertandingan dibagi menjadi tiga kategori: menang, imbang, dan kalah. Hasil analisis menunjukkan bahwa model ini menghasilkan beberapa nilai yang secara statistik tampak signifikan, namun secara logika bertentangan dengan kenyataan di lapangan. Berikut adalah ringkasan hasil regresinya.

Tabel 2. Hasil Regresi Logistik Multinomial

Variabel	Koefisien (B)	Sig.	Exp(B)	Interpretasi
----------	---------------	------	--------	--------------

Jumlah Gol (Gf)	3.963	< 0.001	52.6	Sangat signifikan, 1 gol tambahan meningkatkan peluang menang 52.6 kali.
Penguasaan Bola	-4.005	< 0.001	0.18	Sangat signifikan, tapi berpengaruh negatif; semakin besar possession, peluang menang menurun.
Kebobolan (Ga)	-0.015	0.669	0.985	Tidak signifikan; jumlah kebobolan tidak berpengaruh nyata terhadap kemenangan.

Model regresi logistik multinomial dalam penelitian ini mengkategorikan hasil pertandingan menjadi tiga kelas: menang, imbang, dan kalah. Hasil analisis menunjukkan bahwa meskipun beberapa variabel tampak signifikan secara statistik, namun hubungan antar variabel tidak sesuai dengan logika permainan sepak bola. Sebagai contoh, variabel jumlah gol (Goals For) memiliki koefisien negatif (-4.011) dan signifikan pada $p < 0.001$, yang secara statistik berarti bahwa semakin banyak gol yang dicetak justru menurunkan peluang menang. Sebaliknya, variabel jumlah kebobolan (Goals Against) memiliki koefisien positif (+4.105) dan juga signifikan, seolah-olah semakin banyak kebobolan justru meningkatkan peluang menang.

Hasil yang kontradiktif ini mengindikasikan adanya permasalahan pada pemodelan, seperti pemilihan referensi kategori yang keliru atau ketidakseimbangan jumlah data antar kelas. Kesalahan umum dalam regresi multinomial sering terjadi akibat pengkodean variabel dependen yang tidak tepat, terutama saat jumlah data untuk setiap kategori tidak seimbang [3]. Hal ini dapat menghasilkan estimasi parameter yang bias dan interpretasi yang menyesatkan.

Selain itu, pada model regresi multinomial dengan jumlah data yang terbatas, kemungkinan besar akan muncul hasil yang tidak stabil atau tidak logis, terutama jika multikolinearitas atau noise dalam data tidak ditangani dengan baik [2]. Dalam konteks ini, variabel penguasaan bola (Possession) yang memiliki hasil netral dan tidak signifikan menunjukkan bahwa tidak semua variabel memberikan kontribusi yang cukup kuat untuk membedakan ketiga kategori hasil pertandingan. Sebaliknya, model regresi logistik binomial memberikan hasil yang lebih wajar. Dalam model ini, hasil pertandingan hanya dibagi menjadi dua: menang dan tidak menang. Model ini dianggap lebih sederhana dan relevan untuk melihat peluang kemenangan secara langsung.

Tabel 3. Hasil Regresi Logistik Binomial

Variabel	Koefisien (B)	Sig.	Exp(B)	Interpretasi
Jumlah Gol (Gf)	-4.011	< 0.001	0.018	Setiap tambahan 1 gol menurunkan peluang kalah 98% (sangat signifikan).
Kebobolan (Ga)	4.105	< 0.001	60.64	Setiap 1 kebobolan meningkatkan peluang kalah 60 kali lipat (sangat signifikan).
Penguasaan Bola	0.001	0.977	1.001	Tidak signifikan; hampir tidak ada pengaruh secara statistik.

Model regresi logistik binomial memberikan hasil yang lebih masuk akal dan realistis dibandingkan model multinomial. Dalam model ini, hasil pertandingan dikategorikan menjadi dua kelas: menang dan tidak menang. Hasil analisis menunjukkan bahwa variabel jumlah gol yang dicetak (Goals For) memiliki pengaruh yang sangat signifikan terhadap kemenangan, dengan koefisien sebesar 3.963 dan nilai signifikansi < 0.001 . Odds ratio (Exp(B)) mencapai 52,6, yang berarti setiap penambahan satu gol secara drastis meningkatkan peluang tim untuk menang. Temuan ini selaras dengan studi oleh Lago-Peñas et al. (2010), yang menyebutkan bahwa efektivitas ofensif terutama jumlah tembakan tepat sasaran dan gol merupakan indikator paling menentukan dalam hasil pertandingan sepak bola.

Sebaliknya, penguasaan bola menunjukkan hubungan negatif dengan peluang kemenangan, dengan nilai koefisien -4.005 dan odds ratio 0.018. Artinya, semakin tinggi penguasaan bola justru berpotensi menurunkan peluang menang jika tidak dibarengi dengan efektivitas serangan. Fenomena ini

menjelaskan bahwa tim dengan dominasi possession sering kali kalah jika tidak mampu mengonversi penguasaan menjadi peluang nyata [14]. Hal ini memperkuat argumen bahwa penguasaan bola bukan indikator tunggal keberhasilan, melainkan harus dikombinasikan dengan kecepatan dan akurasi serangan.

Sementara itu, variabel jumlah kebobolan (Goals Against) tidak signifikan secara statistik dalam model ini ($p = 0.669$), yang mengindikasikan bahwa kebobolan tetap dapat ditoleransi selama tim mampu mencetak gol lebih banyak dari lawan. Dalam konteks ini, penting untuk menekankan bahwa keseimbangan antara bertahan dan menyerang menjadi kunci, yang mana tim dengan efisiensi tinggi dalam kedua fase permainan defensif dan ofensif cenderung lebih sukses dalam jangka Panjang [15].

Namun, penelitian ini juga memiliki beberapa keterbatasan. Pertama, data hanya diambil dari satu musim dan melibatkan dua tim saja, sehingga belum bisa mewakili tren secara keseluruhan. Kedua, variabel yang dianalisis masih terbatas pada tiga indikator utama, padahal faktor lain seperti tembakan tepat sasaran, pergantian pemain, dan kondisi psikologis tim juga sangat berpengaruh. Ketiga, model multinomial mengalami ketidaksesuaian dalam hasil, sehingga tidak bisa dijadikan dasar dalam pengambilan kesimpulan. Untuk penelitian selanjutnya, disarankan menggunakan jumlah data yang lebih besar dan beragam, serta menghindari kesalahan dalam pemilihan referensi dan pengkodean variable.

4. Kesimpulan dan Saran

Penelitian ini menunjukkan bahwa jumlah gol yang dicetak adalah faktor paling berpengaruh terhadap kemenangan tim. Semakin banyak gol, semakin tinggi peluang menang. Sebaliknya, penguasaan bola tidak selalu berkorelasi positif dengan kemenangan, bahkan bisa menunjukkan hubungan negatif, menandakan bahwa dominasi bola tanpa efektivitas serangan tidak cukup. Jumlah kebobolan tidak menunjukkan pengaruh yang konsisten. Model regresi logistik multinomial dalam penelitian ini menghasilkan kesimpulan yang tidak logis dan kemungkinan disebabkan oleh kesalahan teknis, sehingga tidak dijadikan acuan utama. Berdasarkan temuan ini, disarankan agar pelatih dan tim analis lebih fokus pada peningkatan efektivitas serangan dan penyelesaian akhir, daripada hanya dominasi penguasaan bola. Latihan skema serangan cepat dan pengambilan keputusan di area lawan perlu diperkuat. Untuk penelitian selanjutnya, disarankan menggunakan data yang lebih luas (lebih banyak tim, liga, musim) dan menambahkan variabel lain seperti jumlah tembakan, konversi peluang, atau data pemain individu untuk hasil yang lebih akurat. Penelitian ini memiliki beberapa keterbatasan. Data hanya diambil dari satu musim dan dua tim, sehingga belum mencerminkan tren keseluruhan. Variabel yang digunakan pun terbatas, padahal faktor lain seperti tembakan tepat sasaran, pergantian pemain, dan kondisi psikologis juga berpengaruh. Selain itu, model multinomial menunjukkan hasil yang tidak sesuai sehingga kurang dapat dijadikan dasar kesimpulan. Penelitian selanjutnya disarankan menggunakan data yang lebih luas dan pemodelan yang lebih tepat.

Daftar Pustaka

- [1] C. Anderson and D. Sally, *The Numbers Game: Why Everything You Know About Football is Wrong*. London, UK: Penguin Books, 2013.
- [2] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. London, UK: SAGE Publications, 2013.
- [3] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ: Wiley, 2013.
- [4] IBM Corp., *IBM SPSS Statistics for Windows*, Version 19.0. Armonk, NY: IBM Corp., 2010.
- [5] K. Moore, *What You Think You Know About Football Is Wrong: The Global Game's Greatest Myths and Untruths*. London, UK: Bloomsbury Publishing, 2019.
- [6] BBC Sport, "Champions League final: Who will win between PSG and Inter Milan?," *BBC*, 2024. [Online]. Available: <https://www.bbc.com/sport/football/psg-inter-final>.
- [7] beIN Sports, "PSG vs Inter Milan: Date, time, stadium and everything you need to know about the UEFA Champions League final," *beIN Sports*, May 31, 2024. [Online]. Available: <https://www.beinsports.com/en-us/soccer/uefa-champions-league/articles-video/psg-vs-inter-milan-date-time-stadium-and-everything-you-need-to-know-about-the-uefa-champions-league-final-2025-05-31>.
- [8] A. Chacoma, M. Kuperman, and D. H. Zanette, "Modeling ball possession dynamics in football matches," *arXiv preprint*, arXiv:2005.04020, 2020. [Online]. Available: <https://arxiv.org/abs/2005.04020>.
- [9] The Sun, "PSG vs Inter prediction: Free betting tips and final preview," *The Sun*, 2024. [Online]. Available: <https://www.thesun.co.uk/sport/psg-vs-inter-prediction>.

-
- [10] F. Sjöberg, "Football match prediction using machine learning," Umeå University Electronic Publishing, 2023. [Online]. Available: <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1764487>.
- [11] O. Grankvist and I.-E. Bergman, "A descriptive analysis of football matches using logistic regression," Umeå University Electronic Publishing, 2023. [Online]. Available: <https://www.diva-portal.org/smash/record.jsf?pid=diva2:1764490>.
- [12] M. A. Babyak, "What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models," *Psychosomatic Medicine*, vol. 66, no. 3, pp. 411–421, 2004, doi: 10.1097/01.psy.0000127692.23278.a9.
- [13] C. Lago-Peñas, J. Lago-Ballesteros, A. Dellal, and M. Gómez, "Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league," *Journal of Sports Science and Medicine*, vol. 9, no. 2, pp. 288–293, 2010.
- [14] H. Liu, M.-A. Gómez, C. Lago-Peñas, and J. Sampaio, "Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup," *Journal of Sports Sciences*, vol. 33, no. 12, pp. 1205–1213, 2015, doi: 10.1080/02640414.2015.1022576.
- [15] H. Sarmiento, R. Marcelino, M. T. Anguera, J. Campaniço, N. Matos, and J. Leitão, "Match analysis in football: A systematic review," *Journal of Sports Sciences*, vol. 32, no. 20, pp. 1831–1843, 2014, doi: 10.1080/02640414.2014.898852.
- [16] E. Schulze, R. Julian, and T. Meyer, "Exploring factors related to goal scoring opportunities in professional football," *International Journal of Sports Science & Coaching*, vol. 17, no. 1, pp. 125–134, 2022, doi: 10.1177/17479541211062613.
- [17] J. González-Rodenas, R. Aranda, and R. Aranda-Malavés, "Possession in football: More than a quantitative aspect," *Frontiers in Psychology*, vol. 10, p. 501, 2019, doi: 10.3389/fpsyg.2019.00501.
- [18] K. Willoughby, "Winning games in Canadian football: A logistic regression analysis," *International Journal of Sports Science & Coaching*, vol. 15, no. 2, pp. 234–242, 2020, doi: 10.1177/1747954120904427.
- [19] M. A. Lemos and J. R. Silva, "Do sports analytics affect footballer pay?," *Frontiers in Behavioral Economics*, vol. 3, p. 112, 2024, doi: 10.3389/fbeco.2024.00112.
- [20] X. Li and Y. Wang, "Prediction of football match results with machine learning," *Procedia Computer Science*, vol. 199, pp. 1234–1241, 2022, doi: 10.1016/j.procs.2022.01.156.