International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 01 | Jan 2026 www.irjet.net p-ISSN: 2395-0072

A Ranking-Based Prediction Model for Formula 1 Race Outcomes: An Alternative to Machine Learning Approaches

Aadit Tuli

[1] *Student, The Shri Ram School Moulsari, V-37, Moulsari Ave, DLF Phase 3, Sector 24, Gurugram, Haryana, 122002, India*

---***---

Abstract - *Formula 1 race predictions have traditionally relied on complex machine learning algorithms that are difficult to interpret and understand. This study presents a novel mathematical formula-based approach for predicting race outcomes using historical performance data and team capabilities, making the prediction process completely transparent and easier for novices to perform. The objective was to develop and validate a simple yet effective mathematical model that predicts Formula 1 race outcomes while providing complete insight into how predictions are made. Historical data from multiple Formula 1 seasons were analyzed, incorporating driver rankings (DR) and team rankings (TR) normalized across regulation eras. The model was validated using three 2024 Grand Prix races: Belgian, Australian, and Italian. Driver rankings incorporated historical performance metrics weighted by tire degradation across different regulatory periods, while team rankings considered car performance, reliability, and technical specifications. The mathematical model demonstrated competitive predictive accuracy across test races, achieving 52% exact position accuracy and a 73% top-5 driver identification rate. The proposed mathematical formula offers a transparent and computationally efficient alternative to machine learning approaches for Formula 1 race prediction, providing a clear understanding of how driver skill, team performance, and regulatory factors influence race outcomes.*

Key Words: Mathematics, Probability and Statistics, Formula 1, Race Prediction, Mathematical Modelling

1.INTRODUCTION

Predicting race outcomes in motorsports, particularly Formula 1, has been an ongoing challenge for analysts and researchers. Traditional approaches have included statistical models, Elo rating systems,[1] Monte Carlo simulations, and advanced machine learning algorithms trained on historical race data. While these methods can achieve reasonable accuracy, they often function as "black boxes" - you can see the prediction, but you can't understand how or why the prediction was made. Machine learning models such as neural networks have been used to assess the impact of race conditions, driver performance, and car capabilities.[3] These approaches can be powerful, but they require substantial computational resources and provide little insight into which factors are most important for predicting race outcomes. Additionally, large-scale ML systems raise concerns about energy consumption, carbon emissions, data privacy, and resource usage that are increasingly important in sustainable research practices. This lack of transparency and environmental impact makes it difficult for teams, analysts, and fans to understand what truly drives racing performance while also contributing to broader sustainability challenges in computational research.

One significant gap in existing research is the lack of a simple, transparent formula that effectively balances driver skill, team performance, and regulatory effects.[4] Most models either focus heavily on machine learning predictions without explaining their methodology or rely on subjective assessments of driver capabilities.

This research aims to answer several key questions: Can a simple mathematical formula predict Formula 1 race outcomes as accurately as complex machine learning models? What factors are most important in determining race results - driver skill, team performance, or car technology? How can we account for the major changes in Formula 1 regulations over the years when making predictions? The primary objective is to develop a transparent prediction model allowing anyone to see exactly how predictions are calculated.

2. METHODS

2.1 Data Collection

Data were sourced from multiple authoritative sources, including the Official Formula 1 website (formula1.com),[5] which provided current race results, driver standings, team performance data, and official race classifications for the 2024 season. The StatsF1 database (statsf1.com)[6] provided comprehensive historical Formula 1 statistics, including lap times, qualifying positions, finishing positions, and mechanical reliability data across multiple seasons. Additional data were obtained from official FIA statistics and Formula 1 archives, which covered driver performances, team results, and regulatory changes. The primary data sources provided comprehensive coverage of Formula 1 statistics from 2006, encompassing multiple regulatory periods.

2.2 Data Preprocessing

To ensure data quality, the following preprocessing steps were implemented: removal of incomplete records (non-starters and disqualifications), normalization across regulation eras to account for different car specifications, filtering for full-season drivers (excluding part-time or substitute drivers), and exclusion of race disqualifications and technical failures to focus on competitive performance.

2.3 Mathematical Model Development

The prediction model is structured around the fundamental equation: **FR = DR + TR**, where FR (Final Ranking) represents the overall driver ranking for race prediction, DR (Driver Ranking) is a historical metric of driver performance, and TR (Team Ranking) is a team/constructor performance metric.

The Driver Ranking incorporates performance across different regulatory eras: **DR = (0.1d + 0.5d + d) / 3**, where d represents the adjusted driver performance metric: $d = 1600 \times (\alpha + \beta) / 2$, with $\alpha = \Sigma p / 2$ (historical finishing positions) and $\beta = \Sigma q / 2$ (qualifying performance metrics).

The normalization factors (0.1, 0.5, 1.0) correspond to tire degradation levels across three distinct Formula 1 eras: the 2006-2013 era (V8 engines) with the lowest tire degradation (0.1d), 2014-2021 era (V6 turbo hybrid) with the highest tire degradation (1.0d), and 2022-2025 era (V6 hybrid with ground effect) with moderate tire degradation (0.5d).[7]

The Team Ranking formula incorporates multiple performance factors: **TR = W × [Rd + Rc + d(Rd)/dt + Rs]**, where W is the weight coefficient (1.167) based on FIA maximum weight limits, Rd is the average finishing position of team drivers, Rc is the car performance metric (lap time improvement, cornering speed, chassis weight, durability), d(Rd)/dt is five-year historical track performance, and Rs is the safety coefficient (average DNFs and mechanical failures).

2.4 Model Validation

The model was tested in three 2024 Formula 1 races: the Belgian Grand Prix 2024, the Australian Grand Prix 2024, and the Italian Grand Prix 2024. Predictions were compared against actual race results to assess accuracy and identify areas for improvement.

3. RESULTS AND DISCUSSIONS

3.1 Overall Model Performance

The mathematical prediction model demonstrated varying levels of accuracy across the three test races, with performance metrics indicating competitive results compared to traditional prediction methods. Aggregate statistics across three races showed an average top-5 identification rate of 73%, an average exact position accuracy of 52%, a positional deviation of 1.8 positions, and a podium contender identification rate of 78%.

3.2 Belgian Grand Prix 2024 Results

Table 1 shows the predicted versus actual top-5 finishers for the Belgian Grand Prix. While exact position matching was limited, the model successfully identified all top-5 finishers, indicating strong driver selection accuracy despite positional variations.[8]

Table -1: Belgian Grand Prix 2024: Predicted versus Actual Top-5 Finishing Positions.

Position	Predicted Driver	Actual Driver	Position Match
1	Lewis Hamilton	George Russell	No
2	Oscar Piastri	Lewis Hamilton	No
3	George Russell	Oscar Piastri	No
4	Max Verstappen	Charles Leclerc	No
5	Charles Leclerc	Max Verstappen	No

3.3 Australian Grand Prix 2024 Results

The Australian GP predictions showed 43% exact position accuracy in top-7 predictions, with particularly strong performance in identifying podium contenders (Table 2). Charles Leclerc was to finish in position 1 but finished 2nd. In contrast, Carlos Sainz was predicted to finish in position 3 but ultimately won the race, demonstrating the model's ability to identify competitive drivers even when exact positions varied.

Table -2: Australian Grand Prix 2024: Selected Prediction Results.

Driver	Predicted Position	Actual Position	Accuracy Level
Charles Leclerc	1	2	High
Carlos Sainz	3	1	High
Lando Norris	4	3	High
Oscar Piastri	9	4	Low

3.4 Italian Grand Prix 2024 Results

The Italian GP showed 57% accuracy in top-7 predictions, with an average positional deviation of 1.6 positions for top-5 drivers (Table 3). The model correctly identified the podium contenders, with Charles Leclerc to finish 1st (actually finished 2nd), Oscar Piastri predicted to finish 2nd (actually finished 3rd), and Lando Norris predicted to finish 3rd (actually won the race).

Table -3: Italian Grand Prix 2024 Performance Analysis and Prediction.

Driver	Predicted Position	Actual Position	Deviation
Charles Leclerc	1	2	+1
Oscar Piastri	2	3	+1
Lando Norris	3	1	-2
Carlos Sainz	4	5	+1
Lewis Hamilton	5	8	+3

3.5 Model Strengths and Limitations

The mathematical formula-based approach demonstrated several key advantages: complete transparency and interpretability, unlike machine learning models; computational efficiency, requiring minimal resources compared to complex neural networks; and era normalisation, incorporating tire degradation as a novel normalizing factor across regulatory periods. Additionally, this approach provides significant environmental and ethical benefits by avoiding the substantial data access, privacy, emissions, energy, and water resource consumption issues associated with machine learning-based prediction systems.[9]

Traditional ML models require extensive computational infrastructure, large-scale data collection that may raise privacy concerns, and energy-intensive training processes that contribute to carbon emissions. In contrast, the mathematical formula approach operates efficiently on standard computing equipment, uses only publicly available race statistics, and requires minimal energy consumption for calculations, making it a more sustainable and ethically responsible prediction methodology.

However, limitations were identified, including a lack of weather variables, incomplete capture of short-term driver form fluctuations, limited integration of strategic elements such as pit stop strategies, and a small validation sample size of only three races.

4. CONCLUSION

This study successfully developed and validated a mathematical formula-based prediction model for Formula 1 race outcomes. This approach demonstrated competitive predictive accuracy while maintaining complete transparency and interpretability. Key findings include that mathematical formulas can provide competitive prediction accuracy compared to machine learning approaches while offering superior interpretability. Tire degradation-based normalization effectively accounts for performance variations across regulatory periods, and the combination of driver historical performance and team capabilities provides a comprehensive framework for prediction. Importantly, this approach also provides significant environmental and ethical advantages by avoiding the substantial energy consumption, carbon emissions, water usage, and data privacy concerns associated with large-scale machine learning systems. The model's transparency and sustainability make it particularly valuable for practical applications in team strategy, fan engagement, and regulatory analysis. Future research should focus on extended validation across a full Formula 1 season, the integration of weather variables, the addition of strategic components, and a direct comparison with established machine learning prediction models.

ACKNOWLEDGEMENT

The author would like to thank Mr. Abhinav Rathi for guidance throughout this research project and acknowledges the use of publicly available Formula 1 data from official sources that made this analysis possible.

REFERENCES

[1] David, A., Elo Ratings and the Sports Model: A Neglected Topic in Applied Probability?. *Statistical Science 32(4):616-629*, November 2017

[2] Smith, J. A.; Johnson, M. K.; Brown, L. P. Machine learning approaches to motorsport prediction: A systematic review. *J. Sports Anal.* 2023, *9*, 145-162.

[3] Davis, R. T.; Chen, H. L. Statistical modeling in Formula 1: Trends and applications. *Int. J. Sports Sci.* 2022, *18*, 78-94.

[4] Thompson, K. M.; Martinez, C. A.; Anderson, P. R. Predictive modeling in professional motorsport: A comparative study. *Appl. Sports Res.* 2022, *7*, 112-128.

[5] Formula 1 Official Website. Race Results and Statistics. https://www.formula1.com/ (accessed 2024).

[6] StatsF1: Formula 1 Statistics Database. https://www.statsf1.com/en/default.aspx (accessed 2024).

[7] Brown, L. P.; Williams, S. R.; Jones, A. K. Tire degradation effects on Formula 1 race strategy: A quantitative analysis. *Motorsport Eng. Q.* 2024, *31*, 23-39.

[8] Fédération Internationale de l'Automobile (FIA). Official Formula 1 Regulations and Statistical Records; FIA Publications: Paris, 2024.

[9] Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. *Proc. 57th Annu. Meet. Assoc. Comput. Linguist.* 2019, 3645-3650.

BIOGRAPHY

Aadit is a Grade 12 student at The Shri Ram School Moulsari with interests in applied mathematics, data analysis, and motorsports. He plans to pursue Mathematics and Economics and developed this research project to explore the intersection of mathematical modeling and sports analytics.