



Strategic and Environmental Variables as Predictors of Formula 1 Race Outcomes: A Machine Learning Approach

Garvit Singhal

ABSTRACT

With Formula 1 generating US\$4 billion in annual revenue, race strategy is one of the most consequential decisions in motorsport, yet remains dependent on human judgment and proprietary simulation tools. This paper investigates whether machine learning models trained on historical race data can learn relationships between strategic and environmental variables and race outcomes. Data from the 2023, 2024, and 2025 Formula 1 seasons was collected through the OpenF1 API, covering 903 driver-race sequences across 23 circuits and 14 drivers. Features included tire compound, stint duration, tire age, track temperature, humidity, rainfall, and starting grid position. Four models were developed and compared: a Transformer encoder, xLSTM, Random Forest, and XGBoost classifier, each predicting finishing position across four classes. The Transformer achieved the highest mean accuracy at 64.8% ($\pm 1.6\%$), while sequential models consistently outperformed tabular models by approximately 5 to 8 percentage points. Mixed-effects logistic regression identified starting grid position as a statistically significant predictor across three of four classes. Predictability was strongest for podium (P1-3) and outside-the-points (P11+) finishes, while P4-5 and P6-10 outcomes showed substantially greater uncertainty. These findings demonstrate that strategic and environmental variables contain predictive signals for race outcomes and that sequential models capture temporal patterns in race strategy.

Keywords - sports analytics; sequential modeling; transformer; classification

INTRODUCTION

Formula 1 is one of the most strategy-driven sports in the world, with a cumulative global audience of 1.83 billion in 2025, a 6.8% increase compared to 2024 and the largest total audience recorded in five years (Formula 1, 2025). While the car and driver play a huge role in performance, race strategy such as the sequencing of tire compounds, pit stop timing, and reacting to changing environmental conditions often determines the final result. Teams that deploy superior strategies consistently outperform rivals with comparable machinery, and strategic errors have repeatedly cost drivers championship positions despite having a pace advantage. A prominent example is the 2024 Formula 1 season, in which Lando Norris lost points across multiple races due to a combination of strategic errors and execution mistakes, including in Canada where McLaren misjudged a safety car call and at the British Grand Prix where a questionable strategy contributed to Norris finishing behind Lewis Hamilton. Sky Sports F1 noted that these costly errors were ultimately decisive in his failure to challenge Max Verstappen for the championship (Sky Sports, 2024). Similarly, Ferrari's 2022 campaign was undermined by a combination of strategic mistakes and reliability failures, with the official Formula 1 website noting that had there been better reliability, better strategic calls, and fewer errors, the outcome of the season could have been very different (Formula 1, 2022). A single poorly timed pit stop, an incorrect tire compound selection, or a failure to respond to a safety car can undo an entire race weekend of preparation and cost valuable championship points. The variables that affect race strategy are numerous, interdependent, and continuously evolve throughout a race. The type of tire compound affects how quickly the tires wear out, while tire degradation is also influenced by track temperature, humidity, and other weather conditions. These factors determine how long a driver can stay on a set of tires before making a pit stop, which also depends on the strategies being used by rival teams. As Piccinotti et al. (2021) describe, the decision of whether to pit stop and which compound to use must be made at the end of each lap, giving teams only the duration of a single lap to evaluate all available information and commit to a course of action. Rain can completely change the race within a single lap, forcing teams to abandon their original plans and make quick decisions under pressure.

Because all these variables are connected, finding the best strategy is extremely challenging. Even experienced strategy engineers, supported by powerful real-time simulator software, still make costly mistakes. Sky Sports further noted that Ferrari cost Charles Leclerc a century of points through decision-making and reliability failures during the 2022 season, illustrating how consequential poor strategic calls can be even for the most well-resourced teams in the sport (Sky Sports, 2022). High-profile strategy errors, including missing the opportunity to perform an undercut, selecting the wrong tire compound, or misjudging how long a set of tires will last, remain common at the highest level of the sport.

A data-driven approach provides a more systematic way of making these decisions instead of relying mainly on experience or intuition. By training a mathematical model on historical Formula 1 data, it becomes possible to identify relationships between race conditions, strategic variables, and race outcomes that may not be obvious to human analysts. Prior research has demonstrated that algorithmic approaches can produce results competitive with or superior to

actual race strategies. Piccinotti et al. (2021) modeled the tire strategy problem as a sequential decision-making problem and applied a Monte Carlo Tree Search algorithm, showing that their automated planning agent was able to match or improve upon the real strategies used by teams in the majority of ten historical races. Quiroga García (2024) built a data-driven optimization model and applied it to the 2022 Hungarian Grand Prix, finding that the model was able to identify the best strategy from around lap 20 onwards, and that Ferrari had made a costly error by switching Leclerc to Hard tires too late in the race. Bonomi et al. (2023) showed that data collected from free practice sessions alone contains enough information to build a viable race strategy using a Genetic Algorithm, with results validated against real tire data provided by Pirelli. Sulsters (2018) built a race simulation model using 2016 season data and showed that simulated finishing positions closely matched real race results, demonstrating that historical data can reliably replicate race outcomes. Rather than depending on strategies developed from a relatively small number of races, a data-driven model trained across multiple seasons and circuits can recognize broader patterns and generate recommendations supported by statistical evidence rather than human judgment alone.

This paper investigates whether a machine learning model trained on historical Formula 1 race data can learn the relationship between strategic and environmental variables and race outcomes, with the goal of quantifying how much of a race result is attributable to these factors.

DATA

Data for this study was sourced from the OpenF1 API (OpenF1), a publicly available telemetry and timing data platform that provides granular race data from the official Formula 1 timing feed. The dataset spans three complete Formula 1 seasons, 2023, 2024, and 2025, covering 23 circuits and 14 drivers who competed in at least 90% of races across all three seasons. This filtering criterion was used to ensure that each driver had sufficient longitudinal data for sequence modeling. The original dataset included 20 drivers across the three seasons. Drivers were identified by name rather than driver number to ensure consistency across seasons, as some drivers were assigned different numbers in different years. Circuits that did not appear in all three seasons were also excluded, reducing the circuit count from 24 to 23. China, which returned to the calendar in 2024 after being absent in 2023, was excluded on this basis. The final dataset consists of 903 driver-race sequences, with each sequence representing one driver's complete race divided into individual stints.

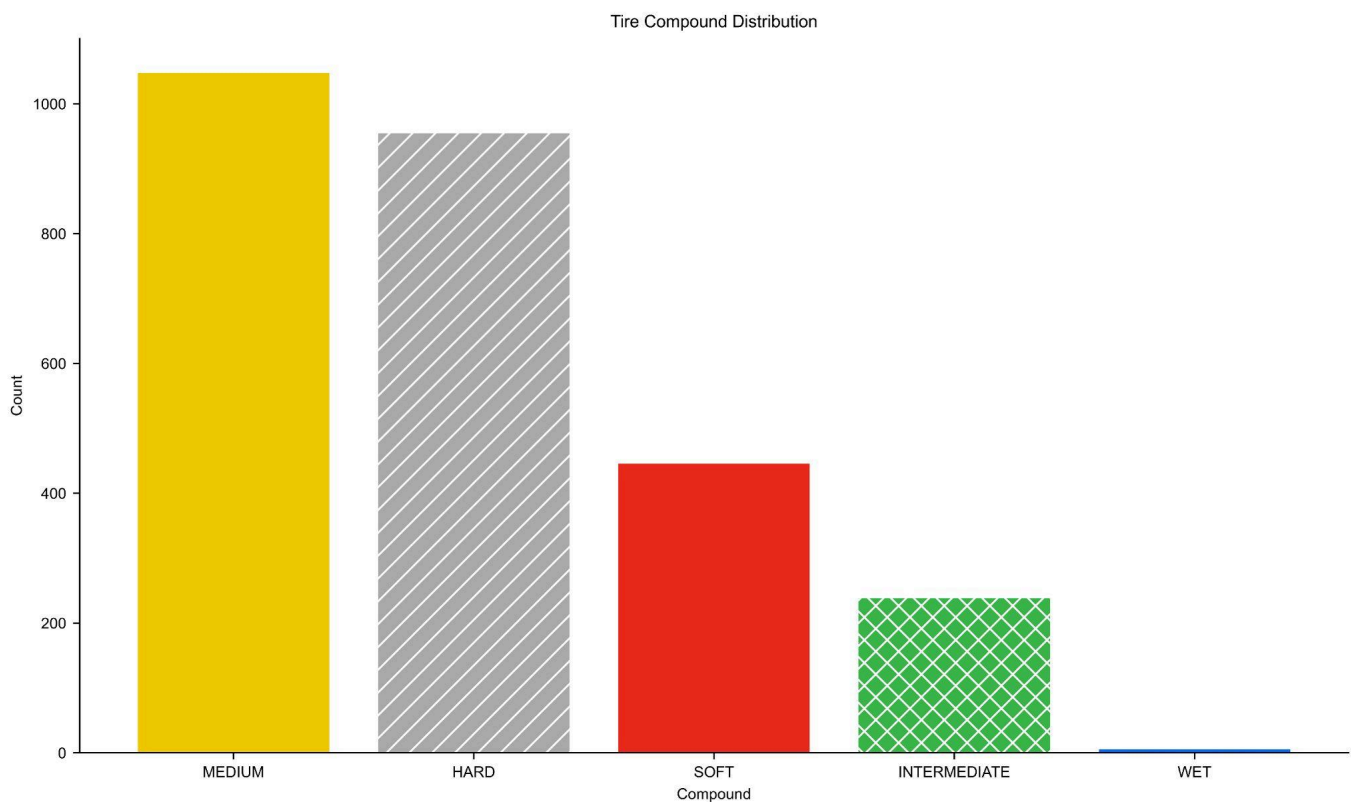
Each stint-level record contains nine features: tire compound, encoded as a numerical index, tire age at the start of the stint, stint duration in laps, track surface temperature, relative humidity, a binary rainfall indicator, starting grid position, stint number, and stint end position. Stint end position represents the driver's on-track position at the end of each stint, used as a proxy for race position at the time of pitting. Including this variable allows the model to condition its predictions on the current competitive context, rather than treating each stint in isolation. For the final stint, this value is set to zero to avoid data leakage, since the final stint's end position is the target variable.

Driver identity was encoded using a 14-dimensional one-hot vector and appended to each timestep, allowing the model to account for driver-specific effects, including differences in driving style, racecraft, and team affiliation that may influence finishing outcomes independently of tire strategy and race conditions.

The target variable is the driver's final finishing position, grouped into four classes based on strategically meaningful outcome thresholds within the FIA championship points system: podium finish (P1-3), strong points finish (P4-5), minor points finish (P6-10), and outside the points (P11+). The dataset exhibits class imbalance, with 301 of the 903 sequences (33.3%) belonging to the P11+ category. To address this imbalance, inverse-frequency class weights were applied to the training loss function, causing misclassifications of underrepresented classes to receive greater penalties during model training.

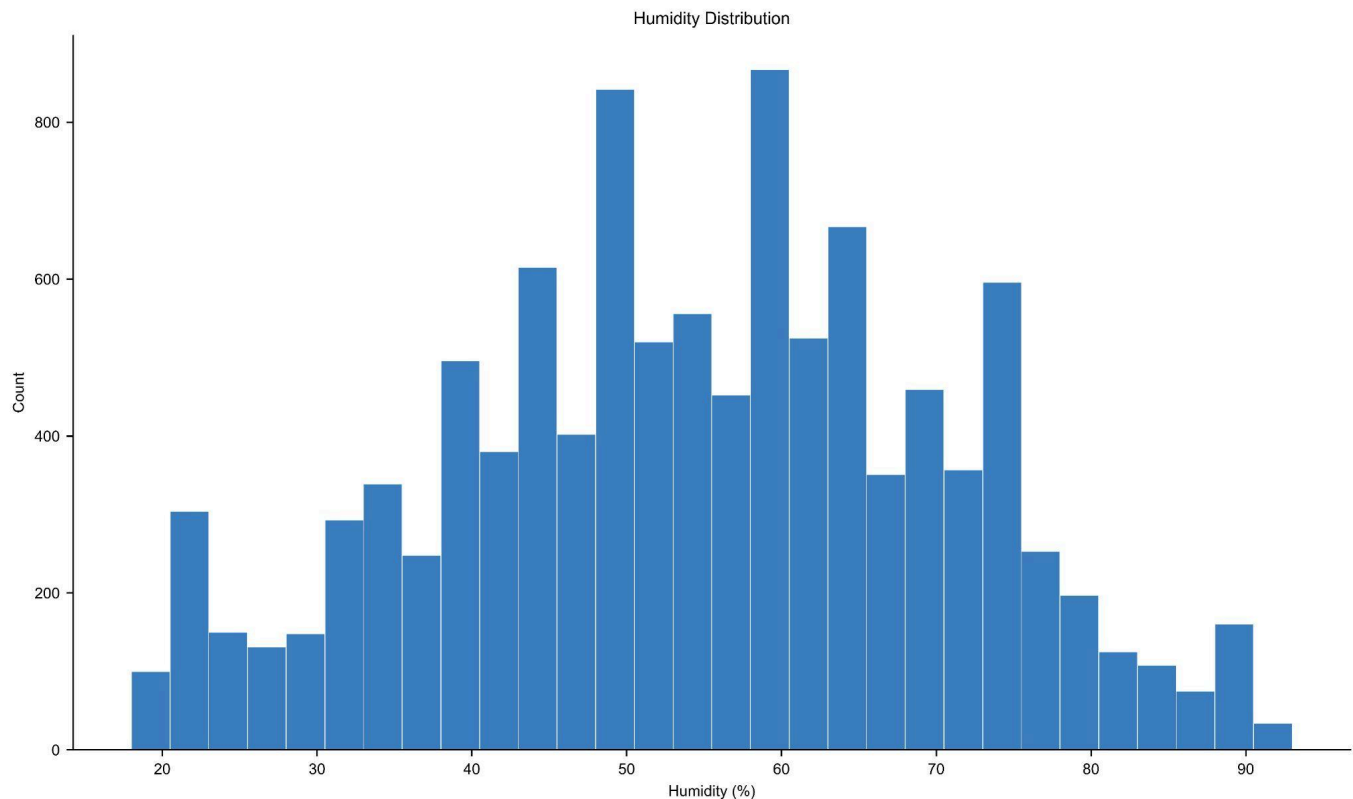
Exploratory Data Analysis

Exploratory analysis of the dataset revealed several notable patterns across drivers, circuits, and race conditions.

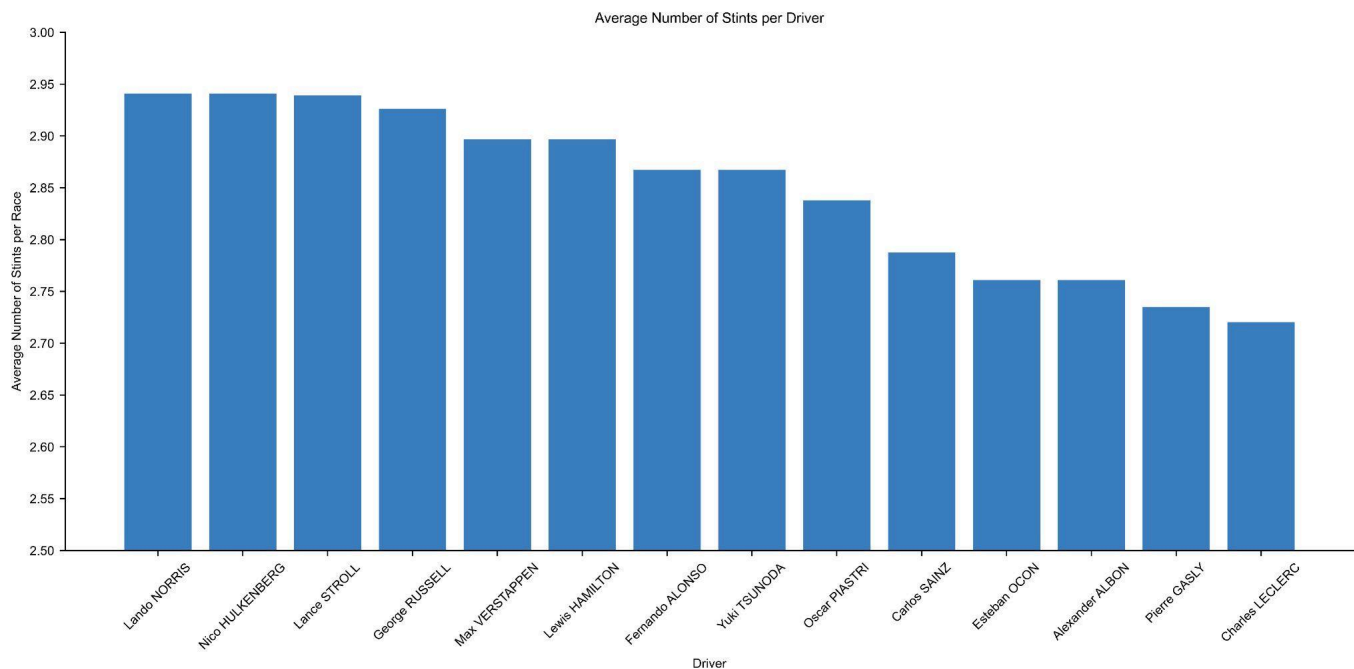


The compound distribution chart shows that Medium and Hard compounds dominate race strategies across the three seasons, accounting for approximately 1,048 and 955 stint usages respectively out of the total 2,695 stints in the dataset. Soft compound stints account for roughly 446 usages, while Intermediate tires appear approximately 239 times, reflecting the relatively infrequent occurrence of wet or mixed conditions across the 23 circuits. Wet tires are used

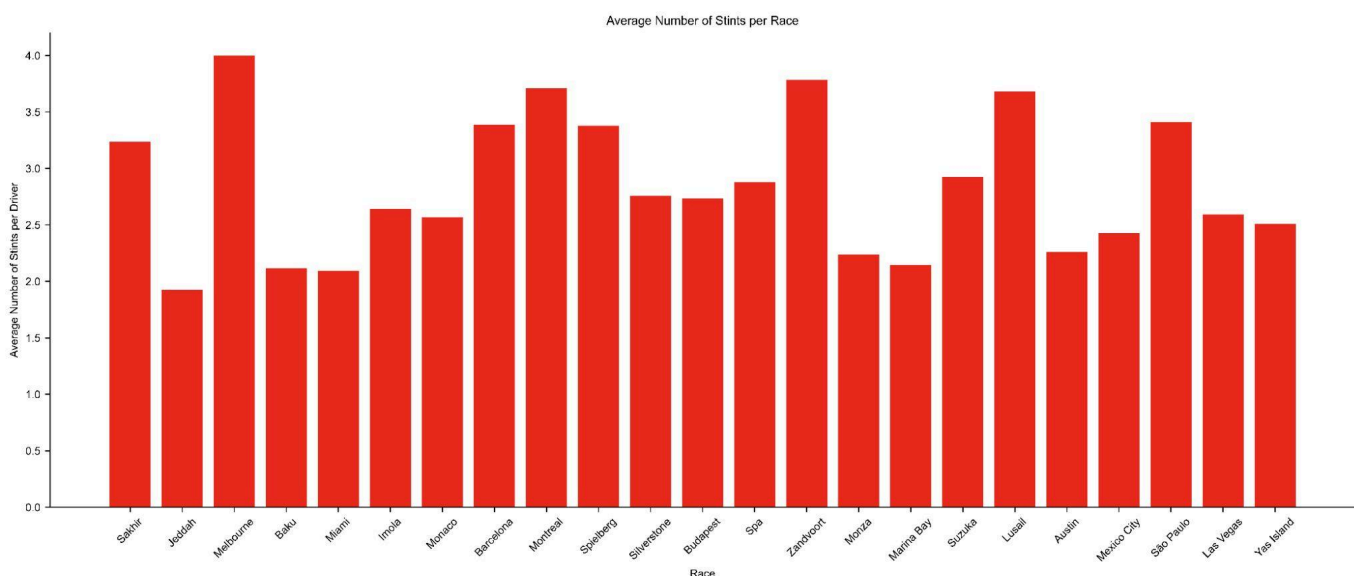
rarely, appearing in fewer than 10 stints across all three seasons. The dominance of Medium and Hard compounds reflects the FIA regulation requiring drivers to use at least two different dry compounds per race, with most teams opting for the more durable compounds to minimize pit stop frequency.



The humidity histogram reveals a bimodal distribution across all race weekends, with two concentration peaks occurring around 50% and 60% relative humidity. Values range from approximately 18% to 95%, reflecting the diverse geographic and seasonal conditions across the 23 circuits in the dataset. Circuits in Southeast Asia such as Marina Bay (Singapore) and in South America such as São Paulo contribute to the higher humidity observations, while Middle Eastern circuits such as Sakhir and Jeddah tend toward the lower end of the distribution.



The average number of stints per driver was relatively consistent across the 14-driver subset, ranging from 2.72 (Leclerc) to 2.94 (Norris, Hulkenberg, Stroll), with all drivers averaging between 2.7 and 3.0 stints per race. This narrow range reflects the FIA regulation requiring at least two different compounds per race, which constrains teams toward 2-stop strategies as the baseline.



A noticeable variation exists across circuits, with averages ranging from approximately 1.9 stints

at Jeddah to nearly 4.0 stints at Melbourne. Circuits such as Melbourne, Spa, Montreal, Lusail, Barcelona, and São Paulo exhibit the highest average number of stints, suggesting that races at these venues typically involve multiple pit stops. This trend is likely influenced by factors such as higher tire degradation and Safety Car interventions. Conversely, circuits including Jeddah, Baku, Miami, Monza, and Marina Bay display lower averages, indicating that one-stop or conservative two-stop strategies are more common.

Table 1. Dataset Summary Statistics

Category	Variable	Mean (SD)	Min	Max
Dataset Overview	Seasons covered	2023, 2024, 2025		
	Number of circuits	23		
	Number of drivers	14		
	Driver-race sequences	903		
	Total stints	2,695		
Stints per Race	Stints per driver-race	2.85 (1.08)	1	8
Stint Duration	Laps per stint	19.93 (12.26)	0	78
Track Temperature	Track temperature (°C)	35.52 (9.17)	16.6	54.6
Humidity	Relative humidity (%)	54.21 (16.18)	18.0	93.0

Table 2. Tire Compound Distribution

Tire Compound	Count
Medium	1,048
Hard	955
Soft	446
Intermediate	239
Wet	6
Total	2,695

METHODS

Four models were trained and evaluated on the dataset: a Transformer encoder, an Extended Long Short-Term Memory network (xLSTM), a Random Forest classifier, and an XGBoost classifier. All models were trained on data from the 2023 and 2024 seasons, with 596 training sequences. The 2025 season was held out as a final test of 307 sequences and was not used at any point during training or model selection.

Transformer

The primary model is a standard Transformer encoder (Vaswani et al., 2017). Each driver-race is represented as a sequence of up to 8 timesteps, with each timestep representing one stint. Sequences with fewer stints are padded with zeros. The raw 23-dimensional feature vector at each timestep is projected into a 32-dimensional embedding, with a learnable positional encoding added to preserve the order of stints. The output from the final real timestep is then passed to a linear classification head that produces four class logits. The model was trained using the AdamW optimizer (Loshchilov & Hutter, 2019) with an ordinal cross-entropy loss that penalizes larger misclassifications more heavily. For example, predicting P11+ when the true class is P1-3 receives a greater penalty than predicting P4-5. Class imbalance was addressed using inverse-frequency class weighting.

xLSTM

An Extended Long Short-Term Memory network (xLSTM; Beck et al., 2024), a recently proposed architecture that extends the standard LSTM through matrix memory cells and improved gating mechanisms, was implemented as a comparison model. It used the same 23-dimensional feature vectors, padding scheme, and aggregation approach as the Transformer, with the output from the final timestep used for classification. The same optimizer, loss function, and class weights were applied to ensure a consistent and fair comparison between the two sequential models.

Random Forest & XGBoost

Since Random Forest (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016) operate on flat feature vectors rather than sequences, each driver-race was represented as a single row. The per-stint features across all 8 stint slots were concatenated and padded with zeros, along with aggregate race-level features including number of stints, number of distinct compounds used, average tire age, average track temperature, average humidity, and average rainfall. Class imbalance was addressed using balanced class weights for Random Forest and inverse-frequency sample weights for XGBoost.

The four models represent two fundamentally different approaches to the problem. The Transformer and xLSTM treat each driver-race as a temporal sequence, allowing them to capture how stint order, tire compound progression, and on-track position change throughout a race. The Transformer's self-attention mechanism allows it to relate information across different timesteps, while the xLSTM processes the sequence step by step and maintains information

through its memory state. Both sequential models are therefore designed to capture the cumulative nature of race strategy decisions. In contrast, Random Forest and XGBoost represent each driver-race as a single flat feature vector. Although they use the same underlying information, the sequential structure must be represented through flattened stint features and race-level aggregates rather than being modeled directly. Given these architectural differences, the sequential models would be expected to outperform the tabular models if the ordering and progression of stints contain meaningful predictive information. The results are evaluated against this expectation in the following section.

Evaluation

All models were evaluated on the 2025 held-out test set of 307 sequences. Performance was assessed using overall classification accuracy, along with per-class precision, recall, and F1 score. To ensure that performance estimates were stable and not sensitive to a particular training data split, models were trained across 5 folds using three random seeds, producing 15 independently trained models for each architecture. In each of the 15 fold evaluations, a model was trained on the corresponding cross-validation subset of the 2023-2024 training data and evaluated directly on the full 2025 held-out test set of 307 sequences. Mean accuracy and standard deviation were computed across these 15 evaluations, providing a stable estimate of out-of-sample performance while accounting for sensitivity to training data composition. All models were tested for confounding variables. Model comparisons are made using each model's best configurations, ensuring that performance differences reflect architectural choices rather than suboptimal tuning. To prevent data leakage from the 2025 test season, the StandardScaler used to normalize continuous features was fitted exclusively on the 2023 and 2024 training sequences and applied to the full dataset thereafter.

RESULTS

The Transformer achieved the highest test accuracy at 64.8% ($\pm 1.6\%$), closely followed by the xLSTM at 63.2% ($\pm 2.5\%$). Both sequential models clearly outperformed the tabular models, with XGBoost achieving 58.7% ($\pm 1.6\%$) and Random Forest achieving 56.4% ($\pm 2.2\%$). The 6 to 8 percentage-point performance gap suggests that the order and timing of stints within a race provide meaningful predictive information that flat feature representations cannot fully capture. The Transformer also showed greater stability across repeated evaluations, as indicated by its lower standard deviation compared with the xLSTM.

Table 3. Model Comparison

Model	Test Accuracy (2025)	Std
Random Forest	56.4%	$\pm 2.2\%$
XGBoost	58.7%	$\pm 1.6\%$
xLSTM	63.2%	$\pm 2.5\%$
Transformer	64.8%	$\pm 1.6\%$

Table 4. Per-Class Performance Breakdown

Model	Class	Precision	Recall	F1	Support
Transformer	P1-3	0.71 ± 0.03	0.78 ± 0.03	0.75 ± 0.02	66
	P4-5	0.44 ± 0.05	0.44 ± 0.08	0.44 ± 0.06	43
	P6-10	0.58 ± 0.04	0.41 ± 0.04	0.48 ± 0.04	85
	P11+	0.71 ± 0.03	0.83 ± 0.03	0.77 ± 0.02	113
xLSTM	P1-3	0.74 ± 0.05	0.76 ± 0.09	0.74 ± 0.05	66
	P4-5	0.41 ± 0.05	0.45 ± 0.11	0.42 ± 0.07	43
	P6-10	0.55 ± 0.06	0.40 ± 0.09	0.46 ± 0.07	85
	P11+	0.71 ± 0.04	0.81 ± 0.05	0.75 ± 0.02	113
XGBoost	P1-3	0.73 ± 0.03	0.70 ± 0.04	0.71 ± 0.02	66
	P4-5	0.37 ± 0.03	0.37 ± 0.06	0.37 ± 0.04	43
	P6-10	0.45 ± 0.03	0.47 ± 0.06	0.46 ± 0.04	85
	P11+	0.71 ± 0.03	0.69 ± 0.03	0.70 ± 0.02	113
Random Forest	P1-3	0.67 ± 0.03	0.72 ± 0.03	0.70 ± 0.02	66
	P4-5	0.36 ± 0.04	0.36 ± 0.05	0.36 ± 0.03	43
	P6-10	0.43 ± 0.04	0.36 ± 0.05	0.40 ± 0.04	85
	P11+	0.66 ± 0.03	0.70 ± 0.04	0.68 ± 0.03	113

Across all four models, per-class performance follows a consistent pattern. The P1-3 (podium) and P11+ (no points) classes are predicted most reliably, while P4-5 and P6-10 are consistently the most difficult classes to classify.

For the Transformer, P1-3 achieved a mean precision of 0.712 (± 0.035) and recall of 0.783 (± 0.034), indicating that the model reliably identifies podium finishers. P11+ achieved a mean precision of 0.713 (± 0.030) and recall of 0.834 (± 0.032), reflecting strong performance in identifying drivers who finish outside the points. P4-5 was the most challenging class, with a mean recall of 0.436 (± 0.076), meaning the model frequently misclassified drivers finishing in the upper points positions as either podium finishers or lower points finishers. P6-10 showed similar difficulty, with a mean recall of 0.405 (± 0.040).

The xLSTM achieved slightly higher precision for P1-3 (0.736 ± 0.050) than the Transformer (0.712 ± 0.035). However, the overlapping variation across runs suggests that this difference should not be interpreted as a clear performance advantage. Overall, the Transformer achieved higher mean accuracy than the xLSTM (64.8% vs. 63.2%) and showed lower variation across runs ($\pm 1.6\%$ vs. $\pm 2.5\%$), indicating more consistent performance across models trained on different training subsets.

Statistical Significance of Features

A complementary mixed-effects logistic regression was used to assess the statistical significance of the input features. The analysis used a one-versus-rest binary logistic model with driver and circuit included as random intercepts. Starting grid position was statistically significant for three of the four finishing classes. For P1-3, starting position had a significant negative coefficient (coef = -0.151, $p < 0.001$), indicating that drivers starting closer to the front of the grid were more likely to finish on the podium. For P11+, starting position had a significant positive coefficient (coef = 0.216, $p < 0.001$), indicating that drivers starting further back were more likely to finish outside the points. Total race laps (coef = -0.059, $p < 0.001$) and the number of distinct compounds used (coef = 0.047, $p = 0.001$) were also statistically significant predictors of P11+ outcomes.

DISCUSSION

Model Behavior & Class Imbalance

The weaker performance on P4-5 and P6-10 reflects a genuine challenge in the data rather than a modeling failure. The boundary between these classes is strategically ambiguous, as the difference between finishing fifth and sixth is often determined by safety car timing, mechanical reliability, or in-race incidents rather than tire strategy and race conditions, which are the primary inputs to these models. Inverse-frequency class weighting partially reduced the effects of class imbalance, but P4-5 recall remained the weakest across all four models, with the Transformer achieving a mean recall of 0.436 (± 0.076). This is further supported by the mixed-effects model, which found no statistically significant features for the P6-10 class, making it the only class without a significant predictor. In contrast, starting grid position was the most statistically significant predictor across three of the four finishing classes, with a negative coefficient for P1-3 (coef = -0.151, $p < 0.001$) and a positive coefficient for P11+ (coef = 0.216, $p < 0.001$). Total race laps and the number of distinct compounds used were also significant predictors of P11+

outcomes, suggesting that longer and more strategically complex races are associated with drivers finishing outside the points. The lack of significant predictors for P6-10 further suggests that this class is difficult to distinguish from neighboring finishing classes using the available features.

Sequential vs Tabular Models

The consistent performance gap between the sequential models, the Transformer at 64.8% ($\pm 1.6\%$) and the xLSTM at 63.2% ($\pm 2.5\%$), and the tabular models, XGBoost at 58.7% ($\pm 1.6\%$) and Random Forest at 56.4% ($\pm 2.2\%$), suggests that the temporal ordering of stints within a race carries meaningful predictive information. A driver's compound sequence, stint lengths, and tire age progression across a race tell a richer story than the same features aggregated into a single flat vector, and the sequential models are better positioned to capture these patterns. The Transformer's lower standard deviation compared to the xLSTM also indicates greater stability across repeated evaluations, suggesting it generalizes more consistently to unseen race data. Taken together, the results demonstrate that strategic and environmental variables, including tire compound selection, stint sequencing, weather conditions, and starting grid position, carry enough signal to predict race finishing outcomes substantially above chance, supporting the central premise of this study.

LIMITATIONS

The most significant missing feature is engine and powertrain data. Power unit performance, including energy deployment, battery management, and engine modes, plays an important role in determining race pace and overtaking ability. However, this information is proprietary and closely guarded by each constructor, making it unavailable through public APIs such as OpenF1. As a result, the model cannot distinguish between a driver running at reduced power to preserve the power unit and one operating at full deployment. This may contribute to misclassification in the P4-5 and P6-10 ranges. One possible direction for future work would be to expand the dataset to include IndyCar, where powertrain specifications are more standardized and baseline engine data is more publicly accessible. This could allow the effects of powertrain-related features on finishing outcomes to be studied more directly.

The model also does not account for pit stop time loss or tire degradation curves. Each pit stop costs approximately 22 seconds of track time, meaning that a strategy involving an additional stop must produce enough lap time improvement from fresher tires to justify the added time loss. The model currently treats tire compound as a categorical feature without modeling how pace changes as a stint progresses. This limits its ability to evaluate strategies where tire degradation is a major factor in determining the optimal strategy. Class imbalance remains a structural limitation given the dataset size of 903 sequences. The P4-5 class in particular contains relatively few examples, making it difficult for the models to reliably distinguish strong points finishes from neighboring classes. Although class weighting can reduce the effects of imbalance, it cannot fully address this limitation without additional data.

Finally, the findings are specific to the 2023 to 2025 regulatory era. Formula 1 has undergone what has been described as the biggest regulation change in its history ahead of the 2026 season, with new rules affecting power units, aerodynamics, tires, and fuel (ESPN, 2026). The patterns learned from this period may not generalize beyond 2025 as team performance hierarchies change under the new regulations. However, the broader finding that machine learning can learn meaningful predictive relationships from publicly available race data is not dependent on a specific regulatory framework. The approach can therefore be retrained on future seasons as new data becomes available.

CONCLUSION

This paper investigated whether machine learning models trained on historical Formula 1 race data can learn the relationship between strategic and environmental variables and race outcomes, with the goal of quantifying how much of a race result can be explained by these factors rather than providing real-time strategic recommendations. Data from the 2023, 2024, and 2025 Formula 1 seasons was collected through the OpenF1 API, covering 23 circuits, 14 drivers, and 903 driver-race sequences. Four models were developed and compared: a Transformer encoder, an xLSTM, a Random Forest, and an XGBoost classifier. Each model was trained to predict a driver's finishing position class using sequence-level features including tire compound, stint duration, tire age, track temperature, humidity, rainfall, starting grid position, and on-track position at stint end.

The Transformer achieved the highest mean accuracy of 64.8% ($\pm 1.6\%$) across 15 evaluations on the unseen 2025 season, with the sequential models consistently outperforming the tabular models by approximately 5 to 8 percentage points. This performance gap suggests that the temporal ordering of stints within a race contains meaningful predictive information that is not fully captured by flat feature representations. A complementary mixed-effects logistic regression further identified starting grid position as a statistically significant predictor for three of the four finishing classes, while no features reached statistical significance for the P6-10 class. This highlights the difficulty of predicting midfield outcomes using the available strategic and environmental variables.

Overall, the findings demonstrate that sequence-based machine learning models can learn meaningful patterns from publicly available Formula 1 race data. However, the absence of proprietary engine and powertrain data, the lack of tire degradation modeling, and the introduction of new technical regulations in 2026 limit the generalizability of the results. Despite these limitations, this approach provides a foundation for future work that incorporates more detailed data sources and realistic race simulation components. As Formula 1 continues to grow in popularity and strategic complexity, machine learning has the potential to complement the expert judgment that remains central to race strategy decisions at the highest level of the sport.

REFERENCES

1. Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J. and Hochreiter, S. (2024) 'xLSTM: Extended long short-term memory', *Advances in Neural Information Processing Systems*, 37. Available at: https://proceedings.neurips.cc/paper_files/paper/2024/hash/c2ce2f2701c10a2b2f2ea0bfa43cfaa3-Abstract.html
2. Bonomi, A., Turri, E. and Iacca, G. (2023) 'Evolutionary F1 race strategy', in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation (GECCO 2023 Companion)*, Lisbon, Portugal, 15-19 July 2023. New York: ACM, pp. 1925-1932. Available at: <https://doi.org/10.1145/3583133.3596349>
3. Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5-32. Available at: <https://doi.org/10.1023/A:1010933404324>
4. Chen, T. and Guestrin, C. (2016) 'XGBoost: A scalable tree boosting system', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, 13-17 August 2016. New York: ACM, pp. 785-794. Available at: <https://doi.org/10.1145/2939672.2939785>
5. ESPN (2026) '2026 F1 rules: What's new on the cars; how will changes affect the racing?', *ESPN*, 5 March. Available at: https://www.espn.com/racing/f1/story/_/id/48090668/2026-f1-rules-whats-new-cars-how-changes-affect-racing
6. Formula 1 (2022) 'Analysis: The good, the bad and the future — What's next for Charles Leclerc after his 2022 title defeat?', *Formula1.com*, 17 December. Available at: <https://www.formula1.com/en/latest/article/analysis-the-good-the-bad-and-the-future-whats-next-for-charles-leclerc.6WhHWZoRqEkNzY9bc5OLhg>
7. Formula 1 (2025) 'Formula 1's record-breaking 2025 season in numbers', *Formula1.com*, 19 December. Available at: <https://www.formula1.com/en/latest/article/formula-1s-record-breaking-2025-season-in-numbers.iq7aR8PcyAw7ysO72vJn>
8. Loshchilov, I. and Hutter, F. (2019) 'Decoupled weight decay regularization', in *7th International Conference on Learning Representations (ICLR 2019)*, New Orleans, LA, 6-9 May 2019. Available at: <https://arxiv.org/abs/1711.05101>
9. OpenF1 (2023) *OpenF1 API: Open-source Formula 1 telemetry and timing data [Data set]*. Available at: <https://openf1.org/>

10. Piccinotti, D., Likmeta, A., Brunello, N. and Restelli, M. (2021) 'Online planning for F1 race strategy identification', in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 12624–12632.
11. Quiroga García, R. (2024) *Optimization of pit stop strategies in Formula 1 racing: A data-driven approach*. MSc thesis, Universidad de Chile. Available at: <http://repositorio.uchile.cl/handle/2250/199664>
12. Sky Sports (2022) 'Explaining Ferrari's blunders and Charles Leclerc's dwindling Formula 1 title dreams', *Sky Sports F1*, 21 August. Available at: <https://www.skysports.com/f1/news/12433/12663464/hungarian-gp-explaining-ferraris-lat-est-blunder-and-charles-leclercs-dwindling-formula-1-title-dreams>
13. Sky Sports (2024) 'Lando Norris: The factors that undid McLaren driver's F1 world title bid against Max Verstappen in 2024', *Sky Sports F1*, 30 November. Available at: <https://www.skysports.com/f1/news/12433/13260572/lando-norris-the-factors-that-undid-mclaren-drivers-f1-world-title-bid-against-max-verstappen-in-2024>
14. Sulsters, W. (2018) *Simulating Formula One race strategies*. MSc thesis, Vrije Universiteit Amsterdam. Available at: <https://vu-business-analytics.github.io/internship-office/papers/paper-sulsters.pdf>
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) 'Attention is all you need', *Advances in Neural Information Processing Systems*, 30. Available at: <https://arxiv.org/abs/1706.03762>