Racing Beyond the Gate: Predicting Speedway Results with Expected Points (xP)

Piotr Kaczorek[0009−0003−8536−2691], Jędrzej Słupski[0009−0000−7334−5534], Franciszek Kornobis[0009−0008−7118−0965], Michał Zaręba[0000−0003−2604−0484], Tomasz Górecki[0000−0002−9969−5257], and Tomasz Piłka[0000−0003−1206−2076]

Faculty of Mathematics and Computer Science,
Adam Mickiewicz University in Poznań, Poland
{piokac4, jedslu, frakor}@st.amu.edu.pl
{michal.zareba, tomasz.gorecki, tomasz.pilka}@amu.edu.pl

Abstract. This study presents a novel predictive framework for individual rider outcomes in the Polish speedway PGE Ekstraliga. Using 2018-2024 data, we introduce the Expected Points (xP) metric and apply feature engineering, including Elo ratings, gate- and track-specific indicators, and rivalry stats. An XGBoost model, optimized with hyperparameter tuning, achieves over 50% accuracy across four classes. Feature importance analysis shows the critical role of long-term rider attributes, dynamic ratings, and contextual variables such as riders' gate preferences. The validated xP sets a benchmark for speedway analytics and enables practical applications in tactical planning, rider evaluation, and talent identification. Future work should explore the integration of telemetry and cross-league data for improved predictions.

Keywords: Speedway racing · Predictive modeling · Machine learning · Classification Methods · Sports analytics.

1 Introduction

Speedway racing is among Poland's most popular sports, with the PGE Ekstraliga considered the world's top speedway league [16]. However, quantitative analysis of speedway performance remains limited compared to other sports.

Speedway presents unique analytical challenges due to the complex interplay of rider skill, bike setup, track conditions, and psychology. Accurate rider performance prediction could provide significant competitive advantages in team selection and tactics, and enhance the fan experience.

In Polish speedway, the primary competition format is the league match, consisting of 15 heats each. Every heat features four riders, who start simultaneously from four gates in a stationary position on 500cc motorcycles. Heat line-ups follow the official race schedule, although team managers may make tactical substitutions when their team is at least six points behind. Riders (see Figure 1) then race four laps on a 300–400 meters long oval track, reaching speeds up to

140 km/h. Each heat lasts 55–70 seconds, depending on track length[1]. Points are awarded as follows: 3 to the winner, 2 for second place, 1 for third, and none for the last rider.



Fig. 1: Speedway riders in action during FIM Speedway Grand Prix round at National Stadium in Warsaw, Poland, 14^{th} May 2022 (author: Wojciech Tarchalski)

In speedway, each heat features four riders competing for points, generating twelve permutations of rider pairs. Our model predicts the winner of each pair and extracts the win probabilities, which are later aggregated into the xP (expected points) metric for each rider in a heat. Inspired by football's xG (expected goals) [12], xP shares some similarities: it quantifies the likelihood of each scoring opportunity (in this case, a rider's heat score) and relates directly to match points. Crucially, unlike xG, calculated post-event [2], xP is calculated pre-race, offering strategic foresight.

This study makes three key contributions: a systematic analysis of factors affecting rider scoring in the PGE Ekstraliga, the development and comparison of predictive models, and the quantification of the importance of key variables in predicting race outcomes. To the best of our knowledge, this is the first comprehensive machine learning application for predicting individual heat results in speedway, providing a foundation for future research. The xP metric also has practical uses: enhancing TV broadcasts with pre-race expected points, supporting coaches in tactical decisions, and enabling objective post-match evaluation by comparing xP to actual scores.

2 Background and Literature Overview

While sports analytics have rapidly evolved, motorsports – particularly speedway – have lagged in advanced statistical applications. Recent work in motorsports

[1] `https://www.youtube.com/watch?v=BcKkrC_E6F8` (accessed on July 26, 2025)

analytics, such as that by Van Kesteren and Bergkamp [17], which provided Bayesian insights into Formula 1 race outcomes, has established foundational methodologies for assessing driver and team effects. Heilmeier et al. [6] employed neural networks for strategic decision-making in motorsport, while Więckowski et al. [20] successfully predicted Formula 1 lap times using machine learning. Peng et al. [14] further extended these predictive approaches into broader racing contexts, proposing advanced rank position forecasting methods for car racing.

In team and individual sports such as tennis and basketball, sophisticated predictive models are already well-established. Kovalchik [8] introduced statistical methods for predicting tennis outcomes [17], and similarly, the hybrid neural network model known as MambaNet [7] demonstrated significant success in forecasting NBA playoff outcomes. These studies underscore the effectiveness of machine learning-based predictive modeling across various sports, highlighting the potential for analogous advancements in motorsports.

Speedway-specific literature has primarily focused on isolated elements rather than comprehensive predictive modeling. Martin et al. [11] identified the crucial impact of starting gate selection and initial positioning, while Markowski et al. [10] emphasized reaction times and rider experience as influential factors. Additionally, Williamson [18] addressed systematic gate position biases. However, these works did not explore predictive modeling for individual heat outcomes or incorporate dynamic skill ratings.

Our research addresses these gaps by introducing the Expected Points (xP) metric for speedway. By integrating the Elo rating system [5], updated dynamically after each heat, we establish a robust, skill-based metric superior to traditional point averages. This innovation significantly enhances prediction accuracy and provides tactical foresight. Unlike prior studies, our model utilizes comprehensive pairwise rivalry features, gate- and track-specific indicators, and systematically engineered rider attributes. This holistic and dynamic approach sets our research apart, creating a benchmark framework for predictive analytics in speedway racing and aligning speedway analytics with advancements seen in other sports domains. Moreover, our study introduces rigorous feature selection and validation procedures, systematically identifying the most predictive rider attributes and contextual variables through advanced model interpretability techniques such as SHAP values. This ensures a precise understanding of rider performance factors, a dimension that has often been overlooked or insufficiently explored in previous research. Thus, our approach not only enhances predictive accuracy but also delivers valuable strategic insights for teams, coaches, and stakeholders.

3 Methodology

3.1 Data Collection

Given speedway's limited global popularity, no public datasets exist for the sport. Therefore, a dedicated dataset was constructed using match data from the Polish

PGE Ekstraliga[2] covering all matches from the 2018–2024 seasons. Collected information includes:

- Information about match (date, starting time, referee, track).
- Teams participating in a match with their line-ups.
- Single heat results (starting gates, substitutions, and rider points, representing also riders' positions).

3.2 Data Preprocessing

In each speedway heat (assuming no exclusions, falls, and engine failures), riders earn points by position: 3 (1^{st}), 2 (2^{nd}), 1 (3^{rd}), and 0 (4^{th}). A bonus point, not affecting match results but included in salaries, rewards riders finishing directly behind a teammate ahead of an opponent, designed to prevent intra-team rivalry [15]. The raw data underwent cleaning procedures detailed in Figure 2 before dataset creation.

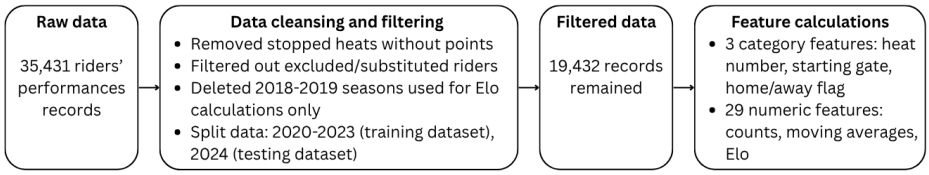


Fig. 2: The process of preparing the dataset

As shown in Figure 3, the dataset contains data on the performances of 19,432 riders in 5,004 races and 336 competitions, as well as information on 158 riders recorded in the database.

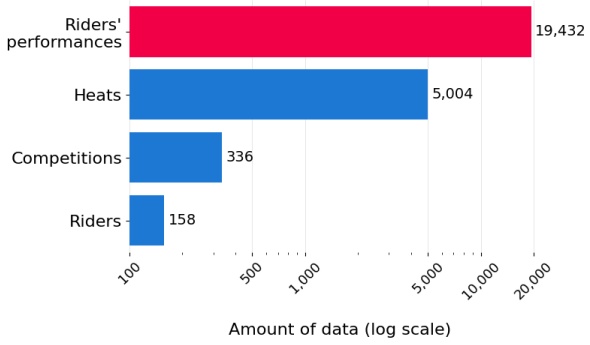


Fig. 3: Dataset summary

[2] https://www.ekstraliga.pl/

3.3 Performance Metrics

Mean Absolute Error (MAE) was used for error quantification due to its interpretability and direct proportionality to xP values [19]. Given balanced classes, standard accuracy sufficed for classification evaluation without error weighting.

Standard accuracy requires an exact match between predicted and actual points. However, due to the bonus point rule in speedway, some predictions may be considered accurate, even though they were initially counted as inaccurate. When a rider finishes directly behind a teammate and at least one opponent is behind them, a bonus point is awarded to the latter. This point does not affect the match result, only the rider's salary. The rule was introduced to limit intra-team rivalry and adds a tactical aspect to specific heats. In bonus point situations, the front rider often focuses on overtaking, while the latter adopts more defensive tactics to avoid being overtaken by rivals. Therefore, we introduced an Accuracy with bonuses metric: predictions are also considered correct if a bonus point is awarded. For n predictions, the formula for Accuracy with bonuses is:

$$\text{Accuracy}_B := \frac{\sum_{i=0}^{3} (T_i + F_{i;B})}{n} \cdot 100\%,$$

where T_i is the number of correctly classified i-point gains, $F_{i;B}$ are the total cases of falsely predicted i-point gains, in a bonus point situation.

4 Approach Dilemma

4.1 Heat Approach

Speedway riders compete in four-rider heats, making heat context crucial for point prediction. To capture this, a quadruple merge was performed on the dataset. As a result, each row contains not only the target rider's features but also those of the three competitors in the heat, thereby contextualizing target rider data concerning all opponents. This methodology, referred to as the Heat Approach, provides comprehensive modeling of the competitive environment of the heat. During development, multiple algorithms were compared to select the top-performing model for the dataset, with XGBoost [3] achieving the best results (MAE = 0.786, Accuracy_B = 47.04), as detailed in Table 1.

Table 1: Results of tested models

Model	MAE	Accuracy_B
Bayesian Ridge	0.792	46.87
LightGBM	0.787	45.78
Random Forest	0.806	44.88
SVM	0.791	46.77
XGBoost	**0.786**	**47.04**

4.2 Pair Approach

Speedway is a dangerous and unpredictable (engine failures, crashes, false starts) sport, what often leads to rider exclusions, preventing them from participating in re-runs and complicating predictions where excluded riders features influence predictions. The Pair Approach addresses this by decomposing heats into rider duels, with each record containing paired rider-opponent features and four new metrics, capturing the dynamics of the rivalry between each pair.

Mirroring the Heat Approach methodology, a double merge was applied to preserve competitive context while excluding inactive riders, appending the in-pair opponent's features to each row. This ensures each row contains complete rider-opponent data without exclusion artifacts.

Heat-level predictions aggregate duel results: predicted points equal duel wins, and xP sums duel win probabilities, as shown in Table 2.

Table 2: Conversion from duels results to heat results for one rider (R1)

Rider1	Rider2	Rider1 prob	Rider2 prob
R1	R2	0.696	0.304
R1	R3	0.425	0.575
R1	R4	0.881	0.119

Duels

⇒

Rider	xP	Duels won
R1	2.002	2

Heat

5 Variables Considered in Pair Model

5.1 Feature Engineering

After switching to the Pair Approach, the newly-trained model underwent extensive feature engineering using Shapley Additive exPlanations (SHAP) [9] to identify significant predictors of rider performance. Features were retained if either the rider's or opponent's SHAP values exceeded an absolute threshold of 0.1. For features below this threshold, the difference between rider and opponent statistics was evaluated for predictive power. If so, the differential feature was retained and the individual rider/opponent components were removed (as an example shows in Figure 4); otherwise, all were discarded.

This process resulted in 46 essential features capturing rider-opponent dynamics (36 of which were calculated for both the rider and his opponent, and the remaining 10 represented the differences between corresponding attributes of the rider and his opponent). The calculated and considered features are described in Table 3. Key findings:

– Match-specific averages (e.g., rider's points per match) showed low predictive value, while cumulative heat points within the game proved highly informative.

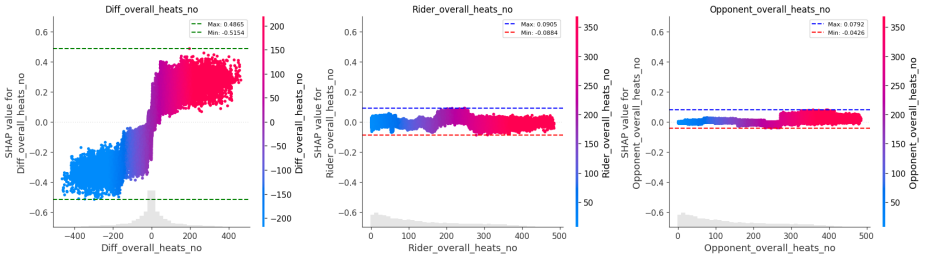


Fig. 4: Example of feature salvage: `Rider_overall_heats_no` and `Opponent_overall_heats_no` showed insufficient individual significance, but their difference demonstrated predictive power

– Gate performance statistics gained significance when calculated over multi-season periods.
– Transient factors (previous heat results, team score sums, and differentials) had minimal impact.

Table 3: Overview of feature categories employed in the xP predictive model

Features category and quantity	Description
Categorical variables (4)	Indicators of the rider's starting gate assignment and whether the heat is contested on their home track. Those two features, calculated for both riders, capture positional and venue-related contextual factors.
Rider characteristics (14)	Quantitative measures of individual form and efficacy, including cumulative point totals, season-long and career averages, and short-term moving averages (last 3 and 5 matches), computed for both the focal rider and their opponent to enable relative class assessment.
Track-related statistics (16)	Performance metrics stratified by track and gate combinations, such as point averages and win ratios, together with the number of heats elapsed since the track's most recent grading, reflecting venue-specific proficiency.
Duel statistics (3)	Head-to-head encounter data encompassing the total number of pairwise duels, individual win counts for each rider, and the resultant win–loss balance, thus quantifying direct competitive history.
Difference-based features (9)	Derived variables representing the differential between rider and opponent across key statistics (e.g., average points, total points, track/gate performance), serving to model relative competitive advantage.

5.2 Elo Rating

To enhance the input feature set for our predictive models, it has become apparent that there is a need to include an additional variable that captures the skill level of riders, beyond their average point performance. For this purpose, the Elo rating system [4] was adopted.

Each heat is treated as a set of six independent pairwise match-ups (duels) between riders. Within-team duels are ignored. Therefore, each race results in four pairwise match-ups between riders from opposing teams. The probability that rider i beats rider j is calculated using the standard Elo formula:

$$E_{i,j} = \frac{1}{1 + 10^{\frac{R_j - R_i}{400}}},$$

where R_i and R_j represent the Elo ratings of riders i and j. The constant value 400 from the original Elo system was used, as it helps to maintain the interpretability of the rating differences. After each heat, the ratings are updated according to the following formula:

$$R'_i = R_i + K \cdot (S_{i,j} - E_{i,j}),$$

where R'_i is the updated rating of the rider i, K is the sensitivity factor, $S_{i,j}$ is the actual result (1 if i wins, 0 if j wins), and $E_{i,j}$ is the expected probability of i winning.

In practice, the ratings are not updated immediately after each heat. Instead, rating changes are accumulated in a temporary buffer during the competition and applied to each rider at the end of the event. This ensures all duels within a competition are evaluated using the same pre-competition ratings. The selection of the K factor and thresholds was based on empirical evaluation. Eight different Elo rating variants were generated, each based on a different combination of K values and heat thresholds. For each variant, a predictive model was trained and evaluated using the corresponding Elo ratings as input features. The lowest mean absolute error was obtained with K set to 40 for riders with fewer than 75 heats, 30 for riders with 75 to 150 heats, and 20 for riders with more than 150 heats, thereby reducing volatility and stabilizing the position of the experienced rider. This configuration was therefore adopted due to its superior predictive performance.

6 Results

The final XGBoost model, developed using the Pair Approach and trained on data from the 2020-2023 seasons, demonstrated strong predictive performance on the 2024 test set. Hyperparameter tuning was conducted with Optuna [1], and the parameter ranges considered during optimization with best values found are listed in Table 4. Results are presented in Table 5 and visualized in Figure 5.

Table 4: XGBoost model hyperparameter search ranges with values found for best model

Hyperparameter	Range	Best values found
`learning_rate`	[0.001, 0.3]	0.289
`max_depth`	[3, 10]	5
`subsample`	[0.1, 1.0]	0.886
`colsample_by_tree`	[0.1, 1.0]	0.862
`min_child_weight`	[1, 20]	2

Table 5: Final results of XGBoost model with the Pair Approach

Model	MAE	Accuracy_B
XGBoost	**0.7387**	**50.112**

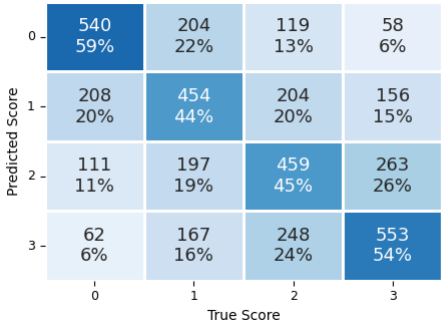


Fig. 5: Confusion Matrix of XGBoost model with the Pair Approach

The final model achieves the highest accuracy for classes corresponding to 0 and 3-point outcomes. This result aligns with expectations, as predicting intermediate positions -— 2nd and 3rd place -— proves more challenging. In many heats, the first and last positions are more distinct, and the competition for middle positions tends to be more balanced.

6.1 Feature Importance Analysis

XGBoost feature importance analysis was conducted using permutation importance calculation [13]. Although individual importance scores exhibit some stochastic variation, Elo ratings and rolling sums of points in the current match consistently emerged as the strongest predictors. Figure 6 shows these results.

These results confirm that the new Speedway Elo Rating is an excellent indicator of riders' performance level, as both `Rider_ELO` and `Opponent_ELO` exhibit the highest permutation importance, surpassing other features by a considerable margin. The high importance of the `*_comp_sum` features (cumulative points already scored in the current match) indicates that a rider's overall performance

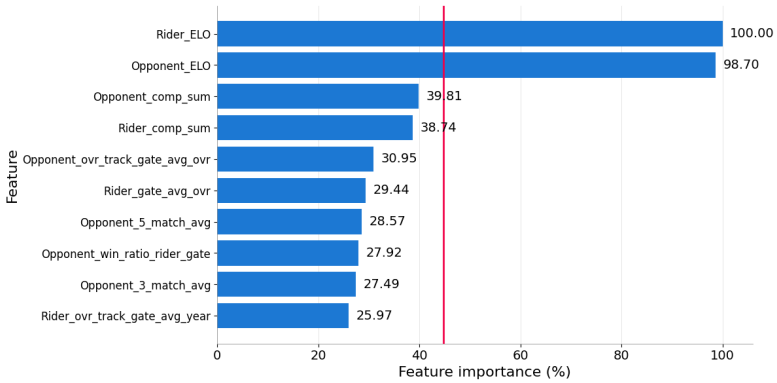


Fig. 6: The top 10 most important features (permutation importance). The red vertical line represents the mean feature importance for the top 10 features.

throughout the game, as reflected by total points accumulated so far, is a strong predictor of their expected performance in the upcoming heat.

7 Discussion

This study demonstrates that long-term historical features outperform short-term form indicators in predicting speedway rider performance. The Elo rating – computed over a fixed two-season window – emerged as the most significant predictor, despite not being used in practice, while the seasonal heat average ranked substantially lower. Moreover, feature differences, particularly the difference in seasonal averages between competitors, yielded stronger predictive signals than individual metrics.

The predictive model achieved very satisfactory overall accuracy but exhibited two systematic biases. First, debuting riders with limited historical data (e.g., 2024 entrants Cook and Jepsen Jensen) incurred higher MAE values (1.120 and 0.956 compared to the 0.7387 average). Second, prediction errors followed a U-shaped dependence on Elo rating, indicating greater volatility among mid-tier riders. Extreme scores (0 and 3 points) were forecasted more accurately than intermediate outcomes (Figures 5 and 7).

8 Conclusions

The results indicate that XGBoost successfully captures the complex relationships underlying speedway performance. Despite the inherent randomness of the data, the model achieved 50% accuracy with bonuses in a challenging four-class prediction task. Longitudinal rider profiles and contextual features, particularly those summarizing extended historical performance, proved to be the strongest determinants of predictive success.

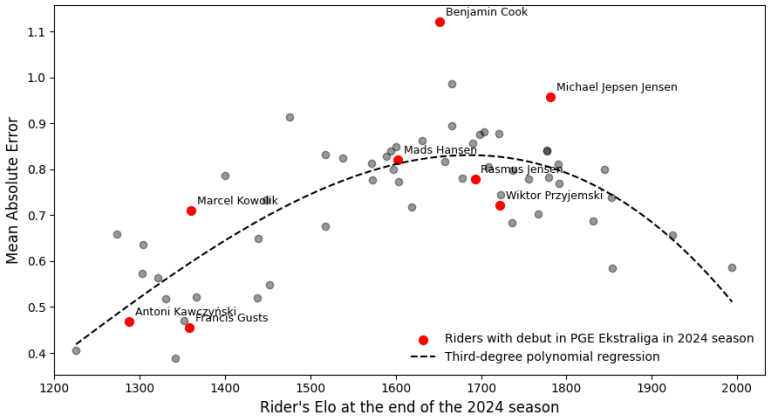


Fig. 7: Relationship between riders' Elo at the end of the 2024 season and prediction error (MAE). Each dot represents a rider with more than 20 appearances in the 2024 season. The dashed curve is a third-degree polynomial regression fitted to the data. Riders debuting in the 2024 season are shown in red and labelled.

To further improve model accuracy, future research should adapt the Elo framework to reflect speedway-specific dynamics (e.g., gate position and track characteristics) and disentangle individual contributions from team effects. Incorporating results from lower divisions could significantly improve initial skill estimates for newcomers, provided appropriate weighting schemes are developed. Finally, expanding the feature set to include telemetry data – such as reaction times, first-corner performance, acceleration, and racing lines – may further enhance predictive accuracy and support advanced decision-support tools for team managers.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 2623–2631. KDD '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3292500.3330701, https://doi.org/10.1145/3292500.3330701
2. Berrar, D., Lopes, P., Davis, J., Dubitzky, W.: Guest editorial: Special issue on machine learning for soccer. Machine Learning **108**(1), 1–7 (2019). https://doi.org/10.1007/s10994-018-5763-8
3. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 785–794 (2016). https://doi.org/10.1145/2939672.2939785
4. Elo, A.E.: The Proposed USCF Rating System, Its Development, Theory, and Applications. Chess life **22**(8), 242–247 (1967)

5. Elo, A.E.: The Rating of Chessplayers, Past and Present. Arco Publishing, New York (1978)
6. Heilmeier, A., Thomaser, A., Graf, M., Betz, J.: Virtual strategy engineer: Using artificial neural networks for making race strategy decisions in circuit motorsport. Applied Sciences **10**(21), 7805 (2020). https://doi.org/10.3390/app10217805
7. Khanmohammadi, R., Saba-Sadiya, S., Esfandiarpour, S., Alhanai, T., Ghassemi, M.M.: Mambanet: A hybrid neural network for predicting the nba playoffs. SN Comput. Sci. **5**(5) (Jun 2024). https://doi.org/10.1007/s42979-024-02977-0
8. Kovalchik, S.A.: Searching for the goat of tennis win prediction. Journal of Quantitative Analysis in Sports **12**(3), 127–138 (2016). https://doi.org/10.1515/jqas-2015-0059
9. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 4768–4777. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017). https://doi.org/10.5555/3295222.3295230
10. Markowski, M., Szczepan, S., Zatoń, M., Martin, S., Michalik, K.: The importance of reaction time to the starting signal on race results in elite motorcycle speedway racing. PLOS ONE **18**(1), e0281138 (2023). https://doi.org/10.1371/journal.pone.0281138
11. Martin, S., Starbuck, C., Coyle, M.: The relationship between heart rate and positioning in speedway racers. Journal of Sports Sciences **32**(sup2), s47–s48 (2014). https://doi.org/10.1080/02640414.2014.968386
12. Mead, J., O'Hare, A., McMenemy, P.: Expected goals in football: Improving model performance and demonstrating value. PLOS ONE **18**(4), 1–29 (04 2023). https://doi.org/10.1371/journal.pone.0282295
13. Molnar, C.: Interpretable Machine Learning. 3 edn. (2025), https://christophm.github.io/interpretable-ml-book, Accessed: 2025-06-14
14. Peng, B., Li, J., Akkas, S., Araki, T., Yoshiyuki, O., Qiu, J.: Rank position forecasting in car racing. In: Proceedings of the IEEE International Parallel and Distributed Processing Symposium (IPDPS). pp. 724–733 (2021). https://doi.org/10.1109/IPDPS49936.2021.00082
15. Polish Automobile and Motorcycle Federation: Regulations (in Polish). https://pzm.pl/zuzel/regulaminy/, Accessed: 2025-06-14
16. Speedway Media: About Poland's Infatuation With Speedway Racing. https://speedwaymedia.com/2024/04/29/about-polands-infatuation-with-speedway-racing/, Accessed: 2025-06-14
17. Van Kesteren, E.J., Bergkamp, T.: Bayesian analysis of formula one race results: disentangling driver skill and constructor advantage. Journal of Quantitative Analysis in Sports **19**(4), 273–293 (12 2023). https://doi.org/10.1515/jqas-2022-0021
18. Williamson, C.A.: Reducing starting position bias in the Speedway Grand Prix. Journal of Sports Analytics **6**(2), 99–109 (2020). https://doi.org/10.3233/JSA-200453
19. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research **30**(1), 79–82 (2005). https://doi.org/10.3354/cr030079
20. Więckowski, J., Paradowski, B., Kizielewicz, B., Shekhovtsov, A., Sałabun, W.: Decision Support System Based on MLP: Formula One (F1) Grand Prix Study Case. In: Proceedings of the 30th Neural Information Processing International Conference (ICONIP). p. 265–276 (2023). https://doi.org/10.1007/978-981-99-8079-6_21