

Formula 1 World Driver Champion 2025 Prediction Using XGBoost

Auliya Viswakarma Hestia, Adityawan

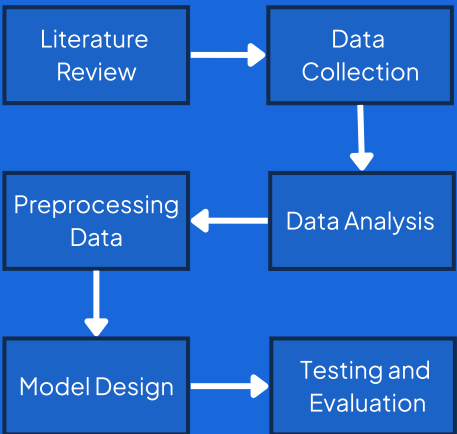
BACKGROUND

Formula 1 (F1) is the pinnacle of motorsport, a highly complex and data-driven competition where success is determined by a combination of driver skill, engineering excellence, and strategic acumen. Predicting the World Driver's Champion is a significant challenge due to the vast number of variables, including vehicle performance, team strategy, and technical regulations. Following several seasons of dominance by Max Verstappen, who secured four consecutive titles by 2024, the 2025 season presents a compelling opportunity for predictive analysis to forecast the next potential champion in this dynamic environment.

To address this challenge, this research proposes the development of a predictive model utilizing the Extreme Gradient Boosting (XGBoost) algorithm, which is renowned for its high performance and effectiveness in handling complex tabular data. The model is built using a comprehensive historical dataset from the Jolpi API, spanning from 1950 to 2025,

RESEARCH

METHODOLOGY



CONCLUSION

- This research successfully developed a robust predictive model using the XGBoost algorithm to forecast the 2025 Formula 1 World Driver Champion.
- The model, optimized with hyperparameter tuning and trained on a decade of recent historical data (2014-2024), achieved excellent performance with an R² score of 0.9441.
- Based on performance data up to the 10th round of the 2025 season, the model predicts that Oscar Piastri has the highest potential to win the championship title.

REFERENCES

[1] F. 1, "Everything you need to know about f1 – drivers, teams, cars, circuits and more," <https://www.formula1.com/en/latest/article/drivers-teams-cars-circuits-and-more-everything-you-need-to-know-about.7lQfL3Rivf1comzdgV5jwc>, 2025.

[2] S. Mitchell, "Data analytics: Managing f1's digital gold," <https://www.raceca-r-engineering.com/articles/data-analytics-managing-f1s-digital-gold/>, 2022.

RESULT AND DISCUSSION

EVALUATION METRICS ACROSS SPLIT SCENARIOS

Scenario	Years	Data Train	MAE	RMSE	R2 Score
Scenario 1	10	2014-2024	1.0610	1.3634	0.9441
Scenario 2	20	2004-2024	1.3288	1.5868	0.9243
Scenario 3	30	1994-2024	1.1596	1.5234	0.9302

RESULTS FROM MODEL SCENARIO 1

