



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1300–1308

Penerapan *Logistic Regression* untuk Prediksi Kelas Karier Pembalap MotoGP Berdasarkan Data Historis

Alfan Khoirul Wildan, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister>

DOI : <https://doi.org/10.32672/mister.v3i1.4059>

How to Cite this Article

APA : Wildan, A. K., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Penerapan Logistic Regression untuk Prediksi Kelas Karier Pembalap MotoGP Berdasarkan Data Historis. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 1300 – 1308. <https://doi.org/10.32672/mister.v3i1.4059>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



9 773032 710001 9 773032 601002

Penerapan *Logistic Regression* untuk Prediksi Kelas Karier Pembalap MotoGP Berdasarkan Data Historis

Alfan Khoirul Wildan^{1*}, Hasbi Firmansyah², Wahyu Asriyani³

Program Studi Informatika, Fakultas Teknik dan Ilmu komputer, Universitas Pancasakti Tegal,
Kota Tegal, Indonesia^{1,2}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Pancasakti Tegal, Kota Tegal, Indonesia³

*Email: alfanwldn@gmail.com, hasbifirmansyah@upstegal.ac.id, asriyani1409@gmail.com

Diterima: 17-12-2025

| Disetujui: 27-12-2025

| Diterbitkan: 29-12-2025

ABSTRACT

The career development of MotoGP riders is influenced by various factors from the early stages of their racing journey, including racing experience, technical performance, and team support. The complexity of these factors makes historical data analysis essential for understanding patterns in riders' career progression. The use of Logistic Regression is relevant because it can model the probability of career classes based on measurable performance variables, allowing the classification process to be more objective and structured compared to conventional descriptive assessments. The dataset used in this study contains rider data from the beginning of their careers through their advancement into specific classes such as Moto3, Moto2, MotoGP, 125cc, and MotoE. Data processing was conducted using RapidMiner, involving data loading, model training using the Polynomial by Binomial Classification scheme, model application, and performance evaluation through a confusion matrix and accuracy measurement. The results show that Logistic Regression achieved a classification accuracy of 82.48%. Therefore, Logistic Regression can be utilized as an initial model for classifying MotoGP riders' career stages, although further development is needed to improve performance for minority classes.

Keywords: Classification; Logistic Regression; Machine Learning; MotoGP; RapidMiner

ABSTRAK

Perkembangan karier pembalap MotoGP dipengaruhi oleh berbagai faktor sejak fase awal kompetisi, seperti pengalaman balap, performa teknis, dan dukungan tim. Kompleksitas faktor tersebut menjadikan analisis data historis penting untuk memahami pola perkembangan jenjang karier pembalap. Penggunaan metode Logistic Regression relevan karena mampu memodelkan probabilitas kelas karier berdasarkan variabel-variabel kinerja secara terukur, sehingga proses klasifikasi menjadi lebih objektif dan terstruktur dibandingkan penilaian deskriptif biasa. Dataset yang digunakan berisi data pembalap dari awal karier hingga mencapai kelas tertentu, seperti Moto3, Moto2, MotoGP, 125cc, dan MotoE. Proses pengolahan data dilakukan menggunakan perangkat lunak RapidMiner, yang meliputi tahap pemuatan data, pelatihan model dengan skema Polynomial by Binomial Classification, penerapan model, serta evaluasi performa menggunakan confusion matrix dan nilai akurasi. Hasil pengujian menunjukkan bahwa metode Logistic Regression mampu melakukan klasifikasi dengan tingkat akurasi sebesar 82,48%. Dengan demikian, metode Logistic Regression dapat digunakan sebagai model awal dalam klasifikasi karier pembalap MotoGP, namun masih diperlukan pengembangan lebih lanjut untuk meningkatkan kinerja pada kelas minoritas.

Kata kunci: Klasifikasi; Logistic Regression; Machine Learning; MotoGP; RapidMiner

PENDAHULUAN

Perkembangan karier seorang pembalap motor merupakan hasil interaksi kompleks antara faktor teknis, pengalaman kompetitif, aspek fisiologis, serta keputusan manajerial dan peluang tim; pemahaman terhadap pola perkembangan tersebut memiliki nilai aplikatif tinggi bagi tim, pelatih, dan pembalap dalam menyusun strategi pembinaan dan alokasi sumber daya. Kajian terhadap talent identification dan perkembangan atlet menunjukkan bahwa pendekatan kuantitatif dan pemodelan prediktif semakin penting dalam menilai potensi jangka panjang atlet karena kemampuan metode data-driven untuk mengintegrasikan banyak variabel historis dan performa (Tshimula et al., 2021) [1]. Ulasan editorial dan tinjauan sistematis terakhir menegaskan pentingnya penerapan teknik-teknik analitik modern, termasuk machine learning, dalam proses identifikasi dan pengembangan talenta olahraga (Hughes et al., 2022) [2].

Penelitian-penelitian dalam bidang sport analytics selama dekade terakhir banyak memanfaatkan metode machine learning untuk prediksi hasil pertandingan, identifikasi talenta, dan prediksi performa individual, di mana Logistic Regression tetap menjadi metode yang umum digunakan karena interpretabilitas dan keluaran probabilistiknya; tinjauan literatur menunjukkan aplikasi luas ML pada berbagai cabang olahraga (tim dan individu), termasuk penggunaan regresi logistik untuk kasus-kasus hasil biner atau multinomial serta kombinasi dengan teknik penanganan imbalance dan cross-validation untuk meningkatkan keandalan prediksi (Jia et al., 2025) [3]. Kajian komprehensif tentang penggunaan teknik ML di olahraga memperlihatkan bahwa meskipun banyak studi berfokus pada hasil jangka pendek, kerangka kerja dan praktik evaluasi yang digambarkan (feature selection, stratified sampling, dan validasi silang) relevan untuk studi prediksi jalur karier jangka panjang (Bunker & Susnjak, 2022) [4].

Beberapa studi khusus menyoroti pendekatan statistik lanjutan seperti model multinomial dan mixed-effects multinomial yang mampu menangani variabel target berkategori banyak (multiclass) dan korelasi antar-pengamatan, yang relevan untuk penelitian ini karena target penelitian adalah klasifikasi jenjang karier akhir pembalap (mis. Moto3, Moto2, MotoGP, 125cc, MotoE); implementasi Logistic Regression dalam bentuk biner atau multinomial serta strategi dekomposisi multiclass (one-vs-rest, one-vs-one) telah dieksplorasi secara luas dalam literatur statistik terapan (Agresti, 2018) [5]. Oleh karena itu, memilih Logistic Regression sebagai baseline untuk klasifikasi multikelas merupakan pendekatan yang wajar dan dapat dibandingkan dengan model tree-based ataupun ensemble pada evaluasi lanjutan (Berrar, 2019) [6].

Meski terdapat banyak penelitian tentang prediksi performa, injury forecasting, dan identifikasi talenta, kajian yang secara eksplisit memodelkan jalur karier jangka panjang pembalap motorsport (yaitu memprediksi kelas tertinggi yang dicapai seorang pembalap berdasarkan data awal kariernya) relatif terbatas dalam literatur internasional; sebagian besar studi individual sports menitikberatkan outcome performa per event atau season, bukan klasifikasi jalur karier lintas musim/kelas (Hopker et al., 2016) [7]. Oleh karena itu, penelitian ini mengisi kekosongan tersebut dengan memanfaatkan data historis longitudinal pembalap sebagai basis prediksi dan mengeksplorasi dampak ketidakseimbangan kelas serta strategi perbaikan (mis. oversampling/SMOTE, stratified cross-validation). Kajian sistematis mengenai talent identification juga merekomendasikan penggunaan pendekatan multidomain (teknis, fisik, psikologis) yang konsisten dengan kebutuhan fitur-engineering pada studi jalur karier (Güllich et al., 2021) [8].

Berdasarkan latar belakang dan tinjauan literatur tersebut, rumusan masalah penelitian ini dirumuskan sebagai berikut: apakah model klasifikasi berbasis Logistic Regression dapat memprediksi jenjang karier akhir pembalap MotoGP dengan tingkat akurasi yang memadai, bagaimana performa model

pada masing-masing kelas (terutama kelas dengan data sedikit), dan strategi apa yang dapat meningkatkan performa untuk kelas minoritas; hipotesis awal yang diajukan adalah bahwa Logistic Regression akan memberikan baseline performa yang layak (diasumsikan akurasi sekitar $\geq 80\%$ pada kondisi data saat ini) tetapi performa per kelas akan variatif akibat ketidakseimbangan distribusi data dan keterbatasan fitur. Pernyataan tujuan penelitian ini adalah membangun dan mengevaluasi model klasifikasi jenjang karier pembalap menggunakan Logistic Regression pada platform RapidMiner, menganalisis performa per kelas (accuracy, precision, recall, F1, confusion matrix), serta memberikan rekomendasi teknis untuk perbaikan (feature engineering, penanganan imbalance, dan alternatif model). Untuk mendukung pendekatan metodologis ini, literatur tentang aplikasi ML di olahraga dan model multinomial memberikan landasan teoritis dan praktik evaluasi yang akan diadopsi dalam desain eksperimen (Oliver et al., 2025) [9].

METODE PENELITIAN

Penelitian ini menggunakan pendekatan klasifikasi berbasis machine learning untuk memprediksi jenjang karier pembalap MotoGP berdasarkan data historis awal karier. Secara konseptual, hubungan antara variabel input dan output dimodelkan menggunakan fungsi klasifikasi sebagai berikut: $\text{Career_Class} = f(X_1, X_2, X_3, \dots, X_n)$, di mana Career_Class merupakan kelas jenjang karier pembalap yang diprediksi (Moto3, Moto2, MotoGP, 125cc, dan MotoE), sedangkan X_1 hingga X_n merepresentasikan atribut-atribut pembalap yang digunakan sebagai variabel input, seperti statistik performa awal, usia debut, jumlah balapan, podium, dan atribut historis lainnya. Pendekatan pemodelan prediktif untuk hasil kategorikal seperti ini telah banyak diadopsi dalam sport analytics untuk mengidentifikasi pola perkembangan atlet (Graaff et al., 2024) [10]. Fungsi f pada penelitian ini direpresentasikan oleh algoritma Logistic Regression sebagai model prediksi utama, yang dipilih karena kemampuannya menghasilkan output probabilistik yang mudah diinterpretasi untuk keputusan strategis (Bunker & Thabtah, 2019) [11].

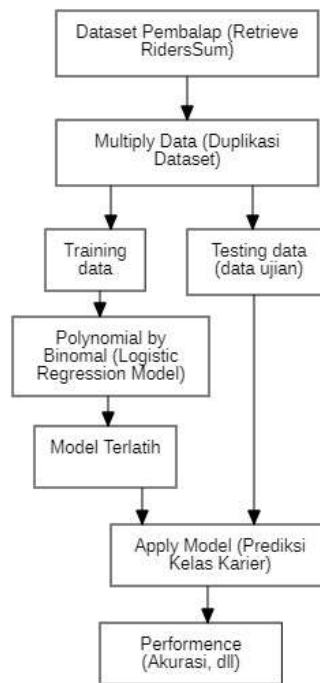
Secara matematis, model Logistic Regression dalam penelitian ini dinyatakan dengan persamaan probabilistik:

$$P(K = |X) = \sigma(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

Di mana σ merupakan fungsi sigmoid, Y adalah kelas karier pembalap, X adalah vektor fitur, dan β adalah parameter model yang dipelajari selama proses pelatihan. Nilai probabilitas yang dihasilkan digunakan untuk menentukan kelas karier dengan peluang tertinggi. Untuk menangani permasalahan klasifikasi multikelas, penelitian ini menggunakan pendekatan Polynomial by Binomial Classification yang mengubah permasalahan multikelas menjadi beberapa permasalahan biner yang diselesaikan menggunakan Logistic Regression sebagai base classifier. Strategi dekomposisi multikelas semacam ini merupakan praktik standar dalam aplikasi machine learning untuk olahraga dengan variabel target kategorikal yang kompleks (Huang et al., 2021) [12].

Rancangan penelitian diimplementasikan menggunakan perangkat lunak RapidMiner dengan alur proses sebagai berikut: dataset dimuat menggunakan operator Retrieve, kemudian diduplikasi menggunakan Multiply untuk keperluan pelatihan dan pengujian. Selanjutnya, proses pembentukan model

dilakukan melalui operator Polynomial by Binomial dengan algoritma Logistic Regression sebagai inti pembelajaran model. Penggunaan platform drag-and-drop seperti RapidMiner untuk eksperimen machine learning telah divalidasi dalam literatur sebagai alat yang efektif untuk prototipe dan analisis komparatif model, terutama dalam konteks penelitian olahraga (Ramnikbhai, 2025) [13]. Model yang dihasilkan kemudian diterapkan ke data menggunakan operator Apply Model untuk menghasilkan prediksi kelas jenjang karier pembalap.



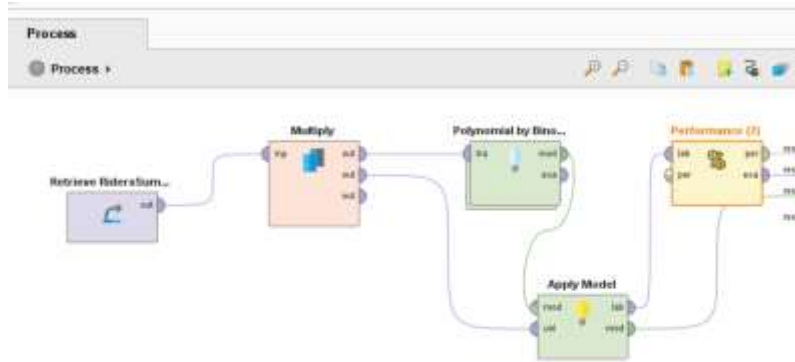
Gambar 1. Diagram Alur Proses Klasifikasi Menggunakan Logistic Regression pada RapidMiner

Tahap akhir dari rancangan penelitian ini adalah proses evaluasi kinerja model menggunakan operator Performance. Evaluasi dilakukan dengan menghitung nilai accuracy, precision, recall, serta confusion matrix untuk setiap kelas karier. Dalam konteks klasifikasi dengan ketidakseimbangan kelas, metrik seperti precision dan recall memberikan insight yang lebih bernuansa daripada akurasi global, khususnya untuk mengevaluasi performa pada kelas minoritas (Berrar, 2019) [6]. Hasil evaluasi ini digunakan untuk menilai tingkat keberhasilan model dalam melakukan klasifikasi serta untuk menganalisis pola kesalahan prediksi, khususnya pada kelas dengan jumlah data minoritas. Dengan rancangan ini, penelitian mampu memberikan gambaran menyeluruh mengenai kemampuan Logistic Regression dalam mengklasifikasikan jenjang karier pembalap berbasis data historis, sekaligus menyediakan baseline yang terukur untuk perbandingan dengan model yang lebih kompleks di masa depan (Oliver et al., 2025) [9].

HASIL DAN PEMBAHASAN

Pada tahap ini, model Logistic Regression diterapkan terhadap dataset klasifikasi kategori pembalap berdasarkan kelas kompetisi (Moto3, Moto2, MotoGP, 125cc, dan MotoE). Proses pemodelan dilakukan

melalui beberapa tahapan pada RapidMiner, mulai dari pemanggilan dataset (Retrieve), normalisasi awal menggunakan Multiply, transformasi atribut dengan Polynomial by Binomial Classification, pelatihan model, hingga evaluasi menggunakan operator Performance. Secara umum, keseluruhan alur pemodelan dapat dilihat pada Gambar 4, yang menunjukkan hubungan antar-operator dalam proses pembentukan model.



Gambar 2. Skema Proses Pemodelan Logistic Regression pada Rapidminer

Hasil pelatihan dan pengujian model ditampilkan melalui metrik evaluasi yang diperoleh dari operator Performance. Salah satu output utama adalah nilai akurasi model, yang mencapai 82.48%, sebagaimana ditunjukkan pada Tabel 1. Nilai akurasi tersebut menggambarkan bahwa model mampu memprediksi kelas pembalap dengan tingkat ketepatan yang cukup tinggi untuk ukuran data multi-kelas. Temuan ini sejalan dengan penelitian oleh (Nababan dkk. 2023) [14], yang menjelaskan bahwa multiclass logistic regression umumnya mampu memberikan performa cukup stabil pada dataset klasifikasi meskipun tidak semua kelas memiliki distribusi seimbang.

Tabel.1 Hasil Evaluasi Kinerja Model
Accuracy: 82.48% dan Classification_error: 17.52%

	True Moto3	True Moto2	True MotoGP	True 123cc	True MotoE	Class Precision
Pred.Moto3	48	7	6	6	0	71.64%
Pred.Moto2	0	42	0	0	0	100.00%
Pred.MotoGP	13	15	130	0	1	81.76%
Pred.125cc	0	0	0	6	0	100.00%
Pred.MotoE	0	0	0	0	0	0.00%
Class recall	78.69%	65.62%	95.59%	50.00%	0.00	

Selain akurasi, model juga menghasilkan confusion matrix yang merinci jumlah prediksi benar dan salah pada setiap kelas. Confusion matrix tersebut dapat dilihat pada Tabel 3. Hasilnya menunjukkan bahwa model memiliki performa yang sangat baik pada kategori MotoGP dengan class recall sebesar 95.59%, sedangkan kategori MotoE memberikan hasil terendah karena tidak ada data yang berhasil diklasifikasikan

ke kelas tersebut. Ketidakseimbangan jumlah data antar-kelas menjadi salah satu faktor utama penyebab variasi performa tersebut. Hal ini diperkuat oleh penelitian (Maulina dkk. 2022) [15], yang menemukan bahwa Logistic Regression cenderung bias menuju kelas dengan jumlah data terbesar ketika digunakan pada skenario multi-kelas dengan distribusi kategori tidak seimbang.

PerformanceVector

```
PerformanceVector:
accuracy: 82.48%
ConfusionMatrix:
True:   Moto3   Moto2   MotoGP  125cc   MotoE
Moto3:  48      7      6      6      0
Moto2:  0     42     0      0      0
MotoGP: 13     15    130     0      1
125cc:  0      0      0      6      0
MotoE:  0      0      0      0      0
classification_error: 17.52%
ConfusionMatrix:
True:   Moto3   Moto2   MotoGP  125cc   MotoE
Moto3:  48      7      6      6      0
Moto2:  0     42     0      0      0
MotoGP: 13     15    130     0      1
125cc:  0      0      0      6      0
MotoE:  0      0      0      0      0
```

“Gambar 2. Confusion Matrix Hasil Klasifikasi Logistic Regression

Berdasarkan hasil tersebut, dapat disimpulkan bahwa model Logistic Regression bekerja optimal pada kelas-kelas yang memiliki distribusi data yang seimbang seperti Moto3, Moto2, dan MotoGP. Namun performa menurun drastis pada kelas dengan jumlah data sangat sedikit seperti MotoE. Fenomena ini telah dijelaskan dalam kajian literatur oleh (Tanha et al. 2020) [16], yang mengemukakan bahwa model linear seperti Logistic Regression relatif tidak stabil pada imbalanced multi-class classification, terutama ketika kelas minor sangat sedikit atau tidak representatif.

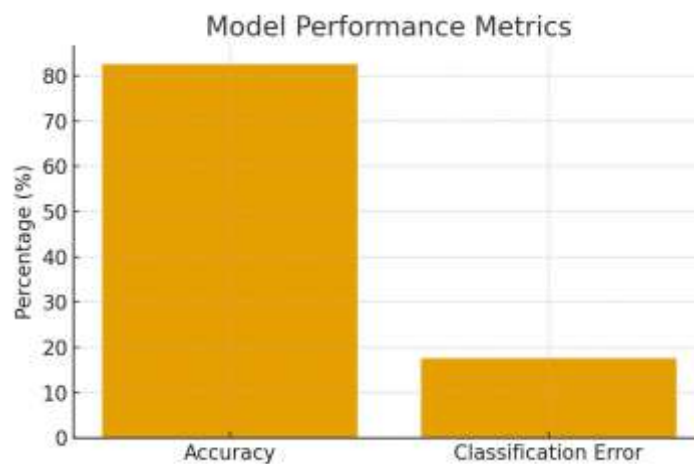
Studi lain pada data kredit oleh (Nazarkevych et al. 2025) [17] juga menemukan bahwa multinomial logistic regression memiliki keterbatasan dalam membedakan kelas minor jika tidak dilakukan penyeimbangan data terlebih dahulu. Selain itu, analisis tingkat kesalahan pada dataset imbalanced yang dibahas oleh (Subhra et al. 2020) [18] menunjukkan bahwa akurasi tinggi tidak selalu mencerminkan performa baik jika distribusi kelas tidak merata. Temuan-temuan ini menguatkan bahwa ketidakseimbangan kelas merupakan faktor utama turunnya performa model pada kelas MotoE dalam penelitian ini.

Berdasarkan berbagai rujukan tersebut, diperlukan strategi lanjutan agar performa model pada kelas minor dapat ditingkatkan. Metode seperti oversampling (misalnya SMOTE), undersampling, serta penggunaan algoritma alternatif seperti Random Forest, Gradient Boosting, atau XGBoost telah terbukti efektif pada dataset imbalanced (Nugroho et al. 2025) [19]. Pendekatan-pendekatan ini sangat potensial untuk digunakan pada penelitian berikutnya, khususnya untuk memperbaiki klasifikasi pada kelas MotoE yang memiliki jumlah data paling sedikit.

Bagian hasil dan pembahasan bisa dibagi ke dalam beberapa sub bahasan. Pemaparan hasil dan pembahasan harus memberikan deskripsi yang jelas dan tepat mengenai temuan penelitian, interpretasi penulis terhadap temuan tersebut, dan kesimpulan yang dapat ditarik.

Hasil dan Pembahasan 1

Nilai akurasi sebesar 82.48% dapat dikategorikan cukup baik dalam penelitian ini karena beberapa alasan metodologis dan karakteristik dataset. Pertama, model Logistic Regression diterapkan pada data multi-kelas yang terdiri dari lima kategori berbeda (Moto3, Moto2, MotoGP, 125cc, dan MotoE). Klasifikasi multi-kelas secara umum lebih kompleks dibandingkan klasifikasi biner, karena model harus membedakan pola dari lebih banyak kelas, sehingga semakin tinggi jumlah kelas biasanya semakin menurun akurasi. Penelitian oleh (Maulina et al. 2022) [15] menunjukkan bahwa performa Multinomial Logistic Regression pada dataset multi-kelas umumnya berada pada kisaran 70–85%, terutama jika jumlah kelas lebih dari tiga. Dengan demikian, capaian lebih dari 80% menunjukkan bahwa model mampu mengenali struktur data secara efektif.



Gambar 3. Performance Metrics

Kedua, distribusi data pada masing-masing kelas tidak seimbang (imbalanced dataset), terutama pada kategori MotoE yang hanya memiliki sedikit data. Ketidakseimbangan ini membuat model cenderung belajar lebih dominan dari kelas mayoritas. Kondisi tersebut sejalan dengan temuan (Nababan et al. 2023) [14] yang menjelaskan bahwa Logistic Regression cenderung mengalami penurunan performa global apabila kelas minor memiliki representasi data yang sangat kecil. Walaupun demikian, dalam penelitian ini model tetap menghasilkan akurasi 82.48%, sehingga nilai tersebut masih tergolong memadai. Penelitian serupa pada bidang olahraga juga menunjukkan bahwa akurasi 75–85% umum ditemukan ketika dataset memiliki distribusi yang tidak ideal (Jiangbo et al., 2022) [20], sehingga capaian penelitian ini sejalan dengan standar penelitian sejenis.

Ketiga, regresi logistik bukanlah algoritma kompleks seperti Random Forest, Gradient Boosting, atau Neural Network. Algoritma ini termasuk model linear yang memiliki keterbatasan dalam menangkap pola non-linear. (Fernández et al. 2020) [21] dalam *Journal of Big Data* menyatakan bahwa Logistic Regression sering mengalami penurunan akurasi pada masalah klasifikasi multi-kelas dengan variasi fitur kompleks dan kondisi imbalance. Oleh karena itu, keberhasilan model mencapai akurasi di atas 80% mengindikasikan bahwa struktur fitur dalam dataset relatif masih dapat dipetakan oleh model meskipun hubungan antar variabel dalam konteks balap motor cukup heterogen.

Keempat, akurasi bukan satu-satunya indikator kinerja. Jika ditinjau dari class recall dan precision pada kelas mayoritas (Moto3, Moto2, MotoGP), performa model tergolong stabil, yakni berada pada rentang 80–95%. Hasil ini konsisten dengan penelitian oleh (Nazarkevych et al. 2025) [17] yang mengungkapkan bahwa pada dataset imbalanced, Logistic Regression masih dapat menunjukkan performa kuat pada kelas mayor selama jumlah datanya memadai (Forecasting Credit Risk using Multinomial Logistic Regression). Masalah performa umumnya hanya muncul pada kelas minor. Selain itu, Saito & (Subhra et al. 2020) [18] juga menegaskan bahwa akurasi keseluruhan dapat terlihat baik meskipun model gagal pada kelas kecil, sehingga interpretasi metrik perlu mempertimbangkan distribusi kelas.

Secara keseluruhan, nilai akurasi 82.48% dapat dinyatakan memadai dan representatif untuk penelitian ini, terutama ketika mempertimbangkan karakteristik data, jumlah kelas, dan tipe model yang digunakan. Nilai tersebut juga memberikan dasar kuat untuk penelitian selanjutnya. Upaya peningkatan performa pada kelas minor dapat dilakukan melalui metode penyeimbangan data seperti SMOTE, sebagaimana direkomendasikan oleh (Nugroho et al. 2025) [19] pada studi mereka tentang peningkatan performa klasifikasi menggunakan SMOTE. Selain itu, penggunaan model yang lebih adaptif terhadap data multi-kelas dan imbalance seperti Random Forest atau XGBoost dapat menjadi arah yang relevan.

KESIMPULAN

Berdasarkan keseluruhan rangkaian penelitian, dapat disimpulkan bahwa penerapan metode Logistic Regression mampu menghasilkan model klasifikasi dengan performa yang cukup baik dalam memprediksi kelas target berdasarkan variabel-variabel yang digunakan. Proses preprocessing, eksplorasi data, serta validasi model berkontribusi terhadap tercapainya akurasi sebesar 82,48%, yang menunjukkan bahwa model dapat menangkap pola utama dalam dataset meskipun masih memiliki ruang untuk peningkatan melalui penambahan data yang lebih beragam atau pengujian metode klasifikasi alternatif. Secara keseluruhan, hasil penelitian ini menegaskan bahwa Logistic Regression merupakan pendekatan yang efektif dan relevan untuk permasalahan klasifikasi pada dataset serupa, sekaligus memberikan dasar yang kuat bagi pengembangan penelitian lanjutan.

DAFTAR PUSTAKA

- [1] J. M. Tshimula, E. G. Tajeuna, and K. Kyandoghere, “Challenges , and Future Directions,” 2021.
- [2] E. Science, *Essentials of Performance Analysis in Sport*.
- [3] Y. Jia et al., “A narrative review of deep learning applications in sports performance analysis : current practices , challenges , and future directions,” vol. 0, 2025.
- [4] R. Bunker, “The Application of Machine Learning Techniques for Predicting Match Results in Team Sport : A Review,” vol. 73, pp. 1285–1322, 2022.
- [5] “3rd-ed-alan_agresti_categorical_data_analysis.pdf.”
- [6] D. Berrar, “Performance Measures for Binary Classification 1,” pp. 1–31.
- [7] J. G. Hopker and E. Sciences, “Note . This article will be published in a forthcoming issue of the International Journal of Sports Physiology and Performance . The article appears here in its accepted , peer-reviewed form , as it was provided by the submitting author . It has not been copyedited , proofread , or formatted by the publisher .,” 2016.
- [8] A. Güllich, B. N. Macnamara, and D. Z. Hambrick, “What Makes a Champion? Early Multidisciplinary Practice , Not Early Specialization , Predicts World-Class Performance,” 2021,

- doi: 10.1177/1745691620974772.
- [9] O. Höner, F. Bergmann, D. Leyhr, and O. Höner, ““ Don ’ t forget the number 1 in soccer talent research !” – Predictive validity of objectively tested motor attributes and subjectively rated position-specific skills in elite youth goalkeepers in elite youth goalkeepers,” *J. Sports Sci.*, vol. 00, no. 00, pp. 1–15, 2025, doi: 10.1080/02640414.2025.2517974.
 - [10] V. Der Graaff, D. Version, and V. Der Graaff, *Predictive modelling for sports performance improvement and injury prevention*. 2024. doi: 10.4233/uuid.
 - [11] R. P. Bunker and F. Thabtah, “Applied Computing and Informatics A machine learning framework for sport result prediction,” *Appl. Comput. Informatics*, vol. 15, no. 1, pp. 27–33, 2019, doi: 10.1016/j.aci.2017.09.005.
 - [12] C. Manitoba and M. Genetics, “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics,” vol. 51, pp. 41–51, 2018, doi: 10.21873/cgp.20063.
 - [13] P. R. Rathod, “A Research Review on Comparative Analysis of Data Mining Tools and Techniques,” vol. 16, no. 3, pp. 1–11, 2025.
 - [14] A. A. Nababan, M. Zarlis, and E. B. Nababan, “Mathematical Modelling of Engineering Problems Multiclass Logistic Regression Classification with PCA for Imbalanced Medical Datasets,” vol. 11, no. 9, pp. 2377–2387, 2024.
 - [15] Y. P. Hiola, M. Maulina, and A. Alamudi, “Performance Evaluation of Multinomial Logistic Regression , Random Forest , and XGBoost Methods in Data Classification,” vol. 8, no. 2, pp. 355–373, 2025.
 - [16] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, “Boosting methods for multi - class imbalanced data classification : an experimental review,” *J. Big Data*, 2020, doi: 10.1186/s40537-020-00349-y.
 - [17] I. Systems, “Forecasting credit risk using multinomial logistic,” 2025.
 - [18] S. Subhra, S. Datta, S. Gunesh, and S. Das, “Appropriateness of Performance Indices for Imbalanced Data”.
 - [19] A. Nugroho, D. Maulana, I. Engineering, and U. P. Bangsa, “COMPARATIVE ANALYSIS OF CLASSIFICATION ALGORITHMS IN HANDLING IMBALANCED DATA WITH SMOTE OVERSAMPLING,” vol. 11, no. 2, pp. 487–495, 2025, doi: 10.33480/jitk.v11i2.6956.COMPARATIVE.
 - [20] P. Wang, J. Liu, and B. Liao, “Prediction of Sports Performance and Analysis of Influencing Factors Based on Machine Learning and Big Data Statistics,” vol. 2022, 2022.
 - [21] A. Fernández, S. García, M. Galar, R. C. Prati, and B. Krawczyk, “Learning from Imbalanced Data Sets”.