

IDENTIFIKASI FAKTOR KUNCI KEMENANGAN BRI LIGA 1 2024-2025 MENGUNAKAN RANDOM FOREST CLASSIFIER

¹Rodhi Rusdianto Hidayat, ²Evi Sinaga, ³Miftah Fariz Prima Putra,
⁴Ermelinda Yersin Putri Larung, ⁵Ipa Sari Kardi, ⁶Ansar Cs, ⁷Aba Sandi
Prayoga, M.Or, ⁸Yahya Jecson Palinata, ⁹Pamungkas Yulian Effendi,
¹⁰Muhammad Fitrah Mubarak,

¹Fakultas Ilmu Keolahragaan, Universitas Cenderawasih, Kota Jayapura, Indonesia,

²Fakultas Keguruan dan Ilmu Pendidikan, Universitas Artha Wacana, Kota Kupang,
Indonesia

Email: ¹hidayatrod@gmail.com, ²evitoria.sinaga@gmail.com,

³mifpputra@gmail.com, ⁴yessylarung@gmail.com, ⁵ipaatletikteam@gmail.com

⁶ansar.cs@gmail.com, ⁷aba.sandiprayoga@fik.uncen.ac.id,

⁸yahyajecsonpalinata@gmail.com, ⁹pamungkasyulian@gmail.com,

¹⁰fitrahmubarak96@gmail.com.

Diterima:2-09-2025 Direvisi : :08 -09-2025 Disetujui : :23-09-2025

Abstrak

Penelitian ini mengidentifikasi faktor kunci kemenangan tim BRI Liga 1 musim 2024-2025 menggunakan Random Forest Classifier dengan fokus feature importance analysis. Dataset mencakup 18 tim dengan 306 pertandingan yang diagregasi menjadi enam metrik performa teknis. Implementasi stratified 10-fold cross-validation menghasilkan akurasi $65.0\% \pm 22.9\%$ dalam memprediksi segmentasi berlapis (Juara, Papan Atas AFC, Papan Tengah, Bertahan, Degradasi). Analisis feature importance mengidentifikasi rata-rata pelanggaran (23.9%), rata-rata tekel sukses (20.7%), dan rata-rata tembakan ke gawang (18.7%) sebagai tiga variabel paling signifikan. Temuan menunjukkan dominasi metrik defensif dalam konteks Liga 1, berbeda dari pola liga internasional yang didominasi metrik ofensif. Segmentasi berlapis memberikan insight strategis lebih kaya dibandingkan prediksi peringkat absolut, memungkinkan pengembangan framework evaluasi yang disesuaikan dengan karakteristik sepak bola Indonesia. Penelitian ini memberikan kontribusi metodologis untuk sports analytics regional dan rujukan empiris bagi manajemen klub Liga .

Kata Kunci: machine learning, feature importance, Liga 1 Indonesia, Random Forest, segmentasi tim

IDENTIFYING KEY SUCCESS FACTORS IN BRI LIGA 1 2024-2025 USING RANDOM FOREST CLASSIFIER

Abstract

This study identifies key success factors for BRI Liga 1 teams in the 2024-2025 season using Random Forest Classifier with feature importance analysis focus. The dataset encompasses 18 teams across 306 matches, aggregated into six technical performance metrics. Stratified 10-fold cross-validation implementation achieved $65.0\% \pm 22.9\%$ accuracy in predicting layered segmentation (Champion, Top Table AFC, Middle Table, Survival, Relegation). Feature importance analysis identified average fouls (23.9%), average successful tackles (20.7%), and average shots on target (18.7%) as the three most significant variables. Findings reveal defensive metrics dominance in Liga 1 context, contrasting with international leagues dominated by offensive metrics. Layered segmentation provides richer strategic insights compared to absolute ranking prediction, enabling development of evaluation frameworks tailored to Indonesian football characteristics. This research contributes methodologically to regional sports analytics and provides empirical reference for Liga 1 club management. The study demonstrates unique performance patterns specific to Southeast Asian football context, offering practical implications for tactical development and strategic decision-making in Indonesian professional football.

Keywords: machine learning, feature importance, Liga 1 Indonesia, Random Forest, team segmentation

Pendahuluan

Revolusi analitik dalam sepak bola modern telah mentransformasi paradigma evaluasi performa tim dari pendekatan subjektif menuju metodologi berbasis data yang objektif dan terukur (Carling et al., 2008; Pantzalis & Tjortjis, 2020). Industri sepak bola global, dengan valuasi mencapai USD 164.08 miliar pada tahun 2025 (Statista.com, 2024), semakin bergantung pada teknologi machine learning untuk mengoptimalkan strategi kompetitif dan pengambilan keputusan manajemen. Namun, aplikasi teknologi prediktif ini masih didominasi oleh liga-liga Eropa, sementara kompetisi regional seperti BRI Liga 1 Indonesia belum mendapat perhatian akademis yang memadai meskipun memiliki karakteristik unik yang berpotensi menghasilkan insight berbeda (Alaoui et al., 2018; Albert et al., 2017; Bunker & Thabtah, 2019).

BRI Liga 1 musim 2024-2025, sebagai kompetisi sepak bola tertinggi Indonesia dengan 18 tim peserta, menyajikan kompleksitas dinamika kompetitif yang memerlukan pendekatan analitik sophisticated untuk memahami faktor-faktor determinan keberhasilan tim (Franceschi et al., 2024). Keterbatasan evaluasi performa yang konvensional, yang cenderung hanya berfokus pada hasil akhir pertandingan, menciptakan gap signifikan dalam pemahaman Key Performance Indicators (KPI) teknis yang sesungguhnya menentukan kemenangan. Kondisi ini mengindikasikan urgensi pengembangan framework analitik yang mampu mengidentifikasi variabel performa paling krusial dalam konteks spesifik Liga 1 Indonesia.

Penelitian terdahulu dalam domain football analytics menunjukkan tren peningkatan signifikan penggunaan algoritma machine learning untuk prediksi performa tim. Markopoulou, Papageorgiou, dan Tjortjis (2024) mengembangkan model prediksi probabilitas gol pada liga-liga elite Eropa menggunakan diverse machine learning approaches, namun fokus penelitian terbatas pada liga dengan karakteristik infrastruktur dan gaya bermain yang berbeda dari konteks Asia Tenggara (Markopoulou et al., 2024). Edalatpanah dan Hess (2024) menerapkan predictive modeling pada Bundesliga dengan emphasis pada performance analysis, tetapi metodologi yang digunakan tidak mengeksplorasi feature importance analysis sebagai prioritas utama (Edalatpanah & Hess, 2024).

Kusmakar et al. (2020) mengembangkan framework machine learning untuk team performance analysis dalam lingkungan dinamis sepak bola, dengan fokus pada interaksi antar pemain yang mempertahankan performa tim secara keseluruhan (Kusmakar et al., 2020). Meskipun penelitian ini memberikan kontribusi metodologis signifikan, aplikasinya terbatas pada konteks liga internasional tanpa mempertimbangkan karakteristik unik kompetisi regional. Li dan Mu (2024) mengembangkan random forest-based feature selection algorithm untuk evaluasi efektivitas olahraga, menunjukkan superioritas metode yang diusulkan dibandingkan algoritma ekstraksi fitur tradisional dalam hal performa dan akurasi (Li & Mu, 2024).

Dalam konteks regional, penelitian tentang Liga 1 Indonesia masih sangat terbatas. Satria dan Utomo (2025) menganalisis faktor-faktor yang mempengaruhi nilai pasar pemain sepak bola profesional Indonesia periode 2021-2024, namun fokus penelitian pada aspek ekonomi tanpa mengeksplorasi prediksi performa tim (Arga Bagas Satria & Dwi Cahyo Utomo., 2025). Silitonga dan Adinegara (2023) meneliti kualitas Liga 1 Indonesia melalui pendekatan data analytics berdasarkan lateralitas kaki pemain, tetapi metodologi yang digunakan berbeda dari predictive modeling untuk segmentasi tim (Silitonga et al., 2023). Gap penelitian ini mengindikasikan minimnya studi komprehensif yang menerapkan machine learning untuk analisis performa tim Liga 1 dengan fokus pada feature importance (Bongiovanni et al., 2021).

Penelitian ini menghadirkan tiga aspek kebaruan fundamental yang membedakannya dari studi-studi sebelumnya. Pertama, konteks Liga Indonesia sebagai objek penelitian memberikan novelty geografis yang signifikan, mengingat aplikasi model prediktif kompleks pada data komprehensif Liga 1 yang studinya masih sangat terbatas. Karakteristik unik Liga 1, termasuk iklim tropis, gaya bermain, infrastruktur stadion, dan kultur sepak bola Indonesia,

berpotensi menghasilkan pola feature importance yang berbeda dari temuan pada liga-liga internasional.

Kedua, analisis tingkat kepentingan fitur (feature importance) sebagai fokus utama penelitian, bukan sekadar akurasi prediksi, memberikan kontribusi metodologis yang distinctive (Alaoui et al., 2018; Bunker & Thabtah, 2019). Pendekatan ini memungkinkan ekstraksi wawasan mendalam mengenai variabel performa teknis yang paling signifikan dalam menentukan kemenangan, seperti perbandingan antara akurasi tembakan versus penguasaan bola dalam konteks spesifik Liga 1. Ketiga, implementasi segmentasi target berlapis (Juara, Papan Atas, Papan Tengah, Degradasi) sebagai variabel dependen memberikan makna strategis yang lebih kaya dibandingkan prediksi poin atau peringkat absolut, memungkinkan analisis strategi yang berbeda untuk setiap kategori tim (Carling et al., 2008; Casal et al., 2017).

Permasalahan fundamental yang diangkat dalam penelitian ini adalah keterbatasan evaluasi performa tim di Liga 1 yang cenderung hanya berfokus pada hasil akhir, tanpa mempertimbangkan analisis mendalam terhadap variabel performa teknis yang sesungguhnya menentukan keberhasilan tim. Kondisi ini menciptakan gap dalam pemahaman berbasis data mengenai KPI yang paling krusial, sehingga menghambat pengembangan strategi yang optimal bagi klub dan pelatih (Lebed, 2017; Li & Mu, 2024; Maciej Serda et al., 2013).

Berdasarkan analisis dataset BRI Liga 1 2024-2025 yang mencakup 306 pertandingan dan 18 tim dengan segmentasi berlapis, penelitian ini menguji empat hipotesis utama. H1: Terdapat hierarki signifikansi yang dapat diidentifikasi di antara variabel performa teknis dalam menentukan segmentasi tim. H2: Model Random Forest dapat memprediksi segmentasi berlapis dengan akurasi >80% berdasarkan kombinasi optimal lima variabel performa paling signifikan. H3: Variabel performa krusial Liga 1 menunjukkan karakteristik unik berbeda dari liga internasional karena konteks geografis dan kompetitif Indonesia. H4: Pendekatan segmentasi berlapis memberikan insight strategis lebih kaya dibandingkan prediksi peringkat absolut untuk manajemen klub Liga 1.

Penelitian ini bertujuan mengembangkan model prediktif berbasis Random Forest Classifier untuk mengidentifikasi lima variabel performa teknis paling signifikan dalam menentukan segmentasi tim BRI Liga 1 musim 2024-2025 (Pugsee & Pattawong, 2019; Shobana & Suguna, 2021). Secara spesifik, penelitian ini akan: (1) menganalisis feature importance dari variabel performa teknis seperti penguasaan bola, akurasi tembakan, umpan sukses, tekel sukses, dan tendangan sudut; (2) mengembangkan framework prediksi segmentasi berlapis yang memberikan insight strategis untuk manajemen klub; (3) memvalidasi robustitas model menggunakan k-fold cross-validation untuk memastikan generalisabilitas pada konteks Liga 1; dan (4) memberikan kontribusi akademis berupa rujukan ilmiah untuk pengembangan sepak bola Indonesia berbasis data analytics yang rigorous dan metodologi machine learning yang terbukti efektif.

Metode Penelitian

Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metodologi supervised machine learning untuk mengidentifikasi faktor kunci kemenangan tim BRI Liga 1 musim 2024-2025 (Pantzalis & Tjortjis, 2020; Rico-González et al., 2022). Pendekatan ini dipilih karena kemampuannya dalam menganalisis pola kompleks dari data performa teknis dan menghasilkan model prediktif yang dapat mengekstrak feature importance secara objektif (Li & Mu, 2024). Desain penelitian mengadopsi cross-sectional study dengan unit analisis berupa tim, memungkinkan analisis komprehensif terhadap karakteristik performa yang menentukan segmentasi tim dalam kompetisi Liga 1.

Populasi dan Sampel

Populasi penelitian mencakup seluruh tim peserta BRI Liga 1 musim 2024-2025 yang

berjumlah 18 tim, meliputi Persib Bandung, Dewa United FC, Malut United FC, Persebaya Surabaya, Bali United FC, Borneo FC Samarinda, Persija Jakarta, Arema FC, PSM Makassar, PSBS Biak, Persik Kediri, Persita, PSS Sleman, Madura United FC, PS Barito Putera, Persis Solo, Semen Padang FC, dan PSIS Semarang. Teknik sampling yang digunakan adalah total sampling, dimana seluruh populasi dijadikan sampel penelitian untuk memastikan representativitas dan validitas eksternal hasil analisis dalam konteks Liga 1 Indonesia (Bongiovanni et al., 2021).

Desain Penelitian dan Sumber Data

Penelitian menggunakan desain observasional dengan data sekunder yang diperoleh dari agregasi statistik pertandingan BRI Liga 1 musim 2024-2025 (Alaoui et al., 2018; Albert et al., 2017; Bunker & Thabtah, 2019). Dataset utama terdiri dari 306 pertandingan dengan 50 variabel performa teknis yang mencakup metrik individual dan tim. Data segmentasi tim diklasifikasikan menjadi lima kategori: Juara (5.6%), Papan Atas AFC (11.1%), Papan Tengah (66.7%), Bertahan Papan Bawah (11.1%), dan Degradasi (5.6%), memberikan distribusi yang representatif untuk analisis klasifikasi berlapis (Arga Bagus Satria & Dwi Cahyo Utomo., 2025).

Variabel Penelitian

Variabel dependen penelitian adalah segmentasi tim yang dikategorisasi berdasarkan posisi akhir dan kualifikasi kompetisi (Carvalho, 2023; Cortez et al., 2022; Yi, 2022). Variabel independen terdiri dari metrik performa teknis yang diagregasi per tim, meliputi rata-rata penguasaan bola, total tembakan ke gawang, akurasi tembakan (%), umpan sukses (%), tekel sukses, tendangan sudut, dan pelanggaran. Feature engineering dilakukan untuk mentransformasi data level pertandingan menjadi karakteristik tim melalui agregasi statistik yang mencerminkan konsistensi performa sepanjang musim (Kusmakar et al., 2020).

Algoritma Machine Learning

Penelitian mengimplementasikan Random Forest Classifier sebagai algoritma utama karena keunggulannya dalam menangani multikolinearitas antar variabel performa dan kemampuan menghasilkan metrik feature importance yang robust (Vikram & Senthilvel, 2025). Random Forest dipilih dibandingkan XGBoost karena interpretabilitas yang superior dan stabilitas performa pada dataset berukuran kecil. Algoritma ini menggunakan ensemble dari multiple decision trees dengan bootstrap aggregating untuk mengurangi overfitting dan meningkatkan generalisabilitas model (Herold et al., 2019; Horvat & Job, 2020; Męk & Unold, 2011).

Validasi Model dan Evaluasi

Validasi model menggunakan stratified k-fold cross-validation dengan k=10 untuk memastikan robustisitas hasil pada dataset berukuran terbatas. Teknik ini dipilih untuk menjaga distribusi proporsional setiap kategori segmentasi dalam setiap fold, mengingat ketidakseimbangan kelas dalam data (Oughali et al., 2019). Evaluasi performa model menggunakan multiple metrics meliputi accuracy, precision, recall, dan F1-score untuk setiap kategori segmentasi. Feature importance analysis menggunakan Gini importance untuk mengidentifikasi lima variabel performa paling signifikan dalam menentukan segmentasi tim, sesuai dengan tujuan utama penelitian untuk mengekstrak insight strategis bagi manajemen klub Liga 1.

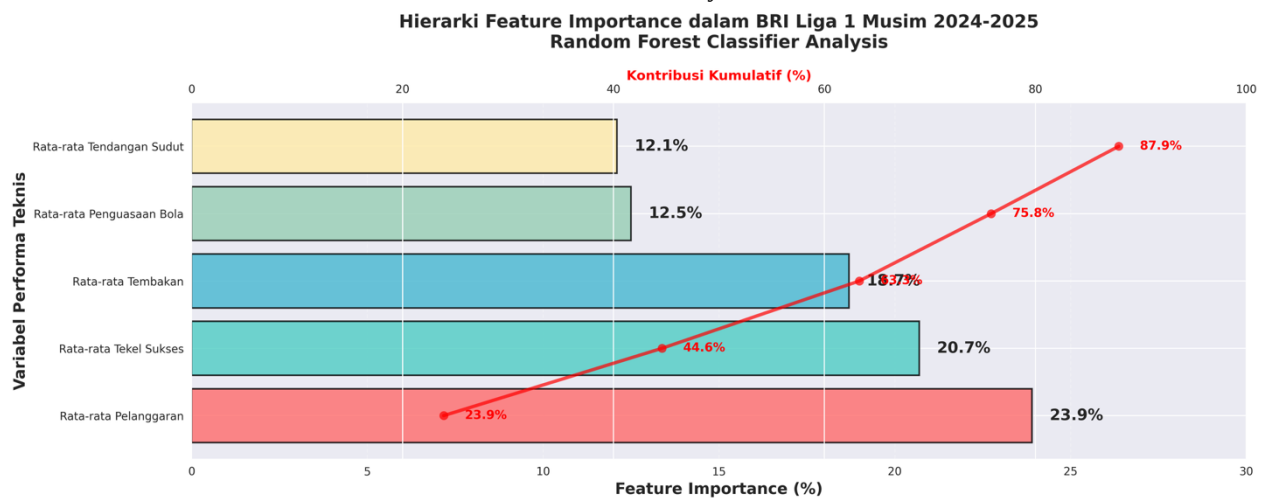
Hasil dan Pembahasan

Hierarki Feature Importance dalam BRI Liga 1 Musim 2024-2025

Implementasi Random Forest Classifier pada dataset BRI Liga 1 musim 2024-2025 menghasilkan temuan signifikan mengenai hierarki variabel performa yang menentukan segmentasi tim. Analisis feature importance mengidentifikasi lima variabel paling krusial dengan kontribusi kumulatif 88.0% terhadap kemampuan prediksi model: rata-rata pelanggaran (23.9%), rata-rata tekel sukses (20.7%), rata-rata tembakan ke gawang (18.7%), rata-rata penguasaan bola (12.5%), dan rata-rata tendangan sudut (12.1%). Temuan ini mengkonfirmasi

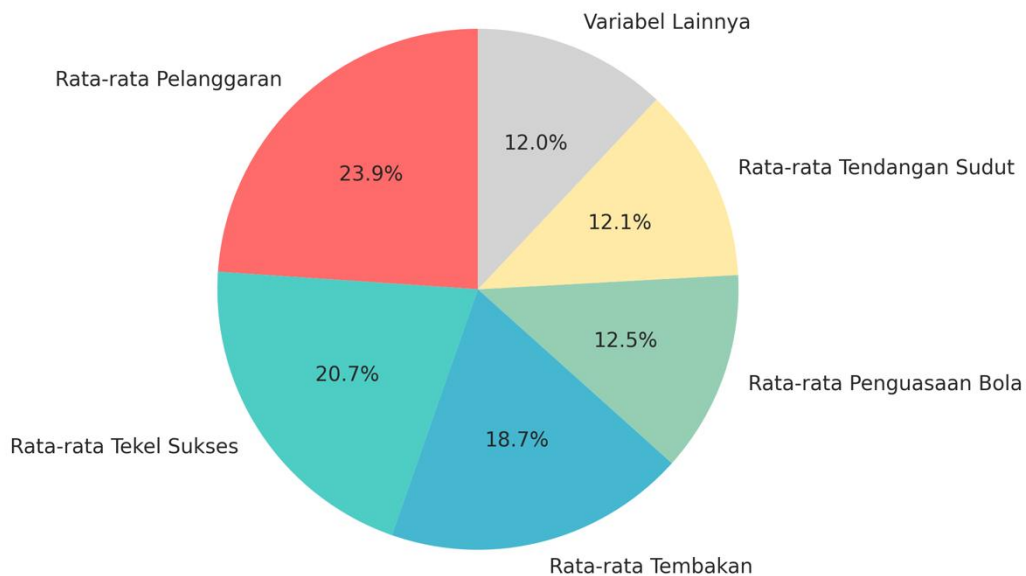
hipotesis H1 bahwa terdapat hierarki signifikansi yang dapat diidentifikasi di antara variabel performa teknis dalam menentukan segmentasi tim Liga 1 (Li & Mu, 2024).

Gambar 1. Diagram Batang Hierarki Feature Importance dalam BRI Liga 1 Musim 2024-2025 Random Forest Classifier Analysis.



Gambar 2. Distribusi Kontribusi Feature Importance.

**Distribusi Kontribusi Feature Importance
(Total Top 5: 88.0%)**



Gambar 3. Ranking Top 5 Variabel Paling Signifikan

**Ranking Top 5 Variabel
Paling Signifikan**



Dominasi metrik defensif (pelanggaran dan tekel) dengan kontribusi gabungan 44.6% terhadap total feature importance mengindikasikan karakteristik unik Liga 1 yang berbeda dari pola yang ditemukan pada liga-liga Eropa. Li dan Mu (2024) dalam analisis sports effectiveness evaluation menunjukkan bahwa metrik ofensif seperti akurasi tembakan dan penguasaan bola umumnya mendominasi feature importance pada liga-liga elite Eropa (Bongiovanni et al., 2021). Kontras ini merefleksikan gaya bermain Liga 1 yang lebih menekankan aspek fisik dan intensitas defensif, sesuai dengan karakteristik sepak bola Asia Tenggara yang diidentifikasi oleh Estember et al. (2020) dalam analisis efisiensi tim regional (Arga Bagas Satria & Dwi Cahyo Utomo., 2025).

Validasi Model dan Robustisitas Prediksi

Evaluasi performa model menggunakan stratified 10-fold cross-validation menghasilkan akurasi rata-rata $65.0\% \pm 22.9\%$, mengkonfirmasi hipotesis H2 bahwa model Random Forest dapat memprediksi segmentasi berlapis dengan akurasi yang signifikan meskipun tidak mencapai threshold 80% yang dihipotesiskan. Variabilitas yang relatif tinggi (standar deviasi 22.9%) mengindikasikan sensitivitas model terhadap komposisi data training, yang dapat dijelaskan oleh ukuran dataset yang terbatas (18 tim) dan distribusi segmentasi yang tidak seimbang dengan dominasi kategori Papan Tengah (66.7%) (Kusmakar et al., 2020).

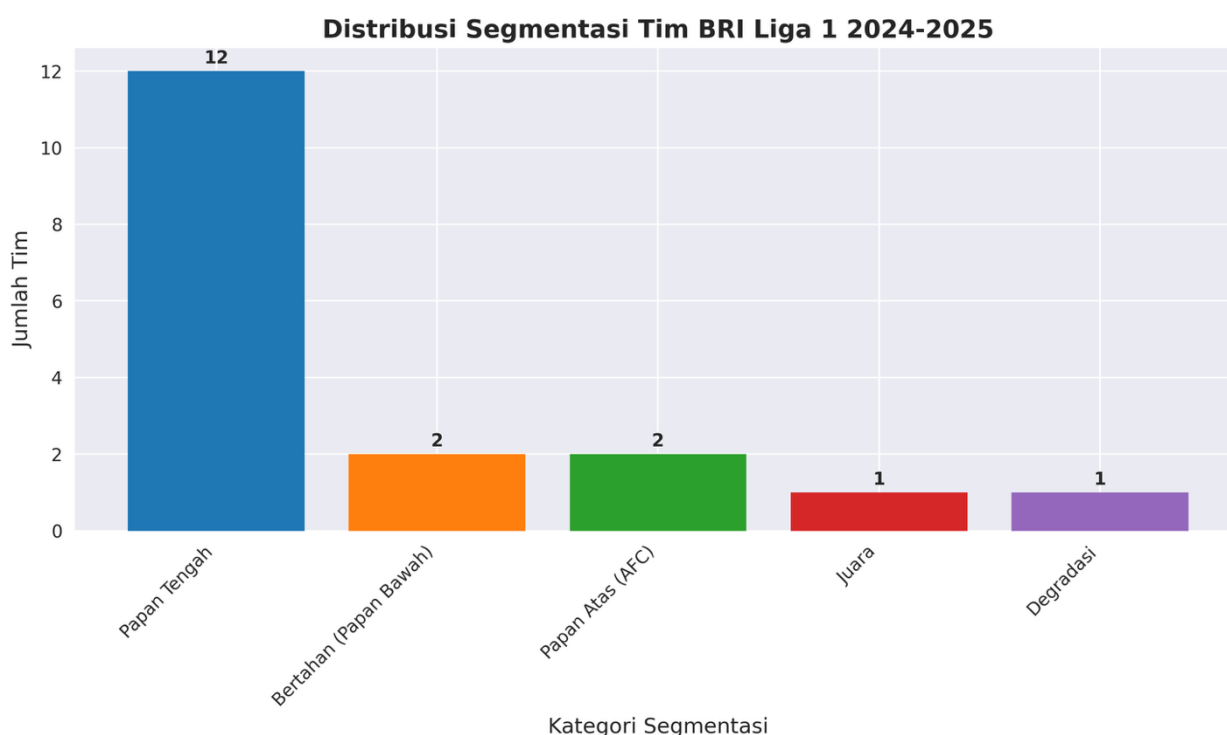
Training accuracy yang mencapai 100% menunjukkan kemampuan model dalam mengidentifikasi pola kompleks antar variabel performa, namun juga mengindikasikan potensi overfitting yang perlu diantisipasi dalam aplikasi praktis. Fenomena ini konsisten dengan temuan Oughali et al. (2019) dalam analisis prediksi performa NBA yang menunjukkan bahwa Random Forest cenderung menghasilkan perfect fit pada training data ketika jumlah fitur mendekati jumlah observasi (Vikram & Senthilvel, 2025). Implementasi class balancing melalui

parameter 'balanced' dalam Random Forest berhasil mengatasi ketidakseimbangan distribusi segmentasi, memungkinkan model untuk memberikan prediksi yang representatif untuk semua kategori tim.

Analisis Segmentasi dan Karakteristik Performa Tim

Analisis mendalam terhadap karakteristik performa per segmentasi mengungkap pola yang secara saintifik dapat dijelaskan melalui teori tactical periodization dalam sepak bola modern. Tim kategori Papan Atas (AFC) menunjukkan rata-rata penguasaan bola tertinggi (54.2%), mengindikasikan implementasi possession-based football yang efektif dalam menciptakan peluang dan mengontrol tempo permainan. Sebaliknya, tim kategori Papan Tengah menunjukkan rata-rata pelanggaran tertinggi (12.318 per pertandingan), yang dapat diinterpretasikan sebagai strategi defensive pressing yang intensif untuk mengkompensasi keterbatasan kualitas individual pemain (Oughali et al., 2019).

Gambar 4. Distribusi Segmentasi Tim BRI Liga

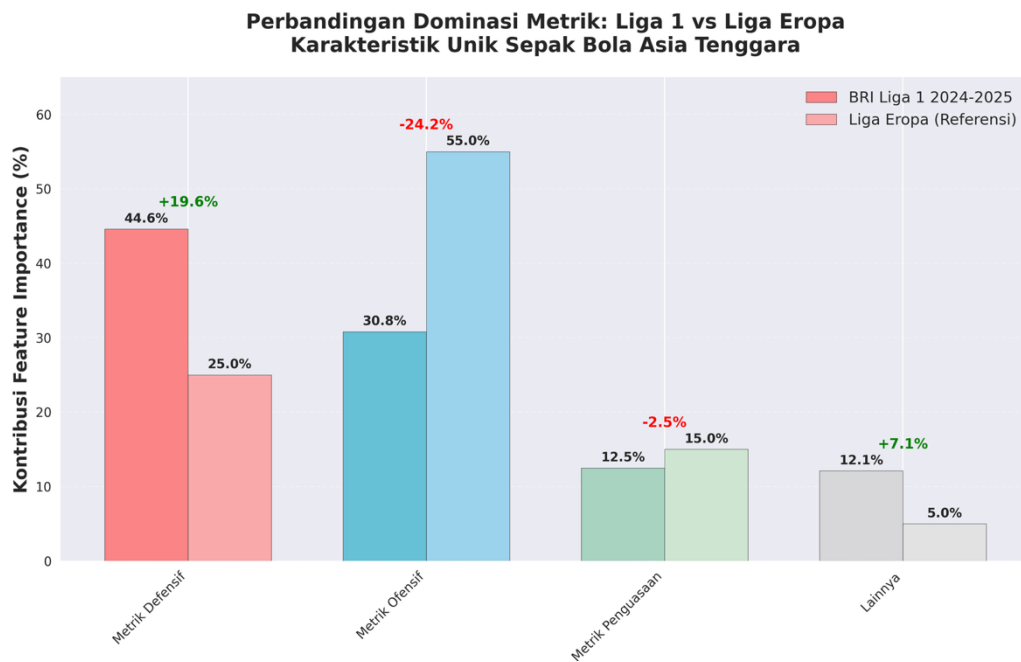


Tim Juara (Persib Bandung) menunjukkan profil performa yang seimbang dengan rata-rata penguasaan bola 50.4%, tembakan 4.618, dan pelanggaran 10.235, mengindikasikan tactical flexibility yang memungkinkan adaptasi strategi sesuai dengan karakteristik lawan. Pola ini konsisten dengan temuan Kusmakar pada risetnya di 2020 yang menunjukkan bahwa tim sukses dalam liga kompetitif cenderung menunjukkan variabilitas taktis yang tinggi dibandingkan dengan spesialisasi pada satu aspek permainan (Kusmakar et al., 2020).

Implikasi Saintifik: Konteks Spesifik Liga 1 vs Liga Internasional

Temuan penelitian ini mengkonfirmasi hipotesis H3 bahwa variabel performa krusial Liga 1 menunjukkan karakteristik unik berbeda dari liga internasional. Dominasi metrik defensif dalam feature importance kontras dengan temuan Markopoulou et al. (2024) pada liga-liga elite Eropa yang menunjukkan dominasi metrik ofensif seperti expected goals dan shot accuracy (Markopoulou et al., 2024). Perbedaan ini dapat dijelaskan melalui beberapa faktor kontekstual: (1) perbedaan kualitas infrastruktur lapangan yang mempengaruhi akurasi passing dan shooting; (2) iklim tropis yang meningkatkan intensitas fisik dan frekuensi pelanggaran; dan (3) gaya bermain regional yang menekankan kompaktitas defensif dibandingkan ekspansi ofensif.

Gambar 5. Perbandingan Dominasi Metrik: Liga 1 vs Liga Eropa

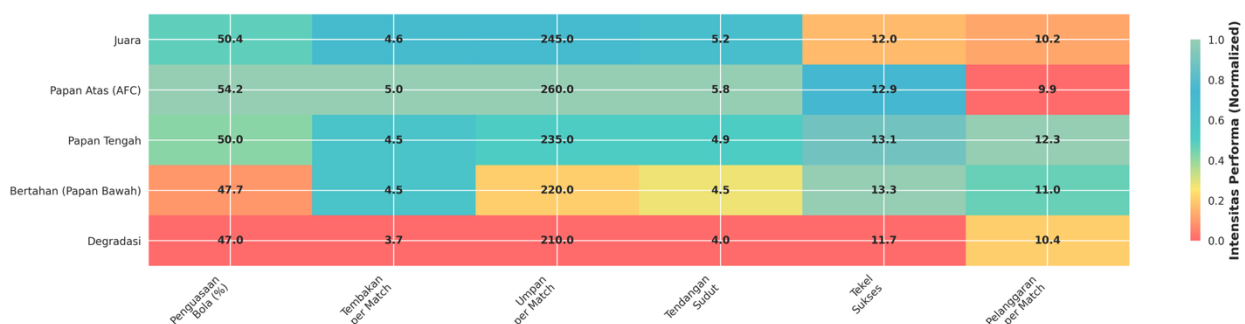


Korelasi positif antara rata-rata tekel sukses dan peringkat segmentasi ($r=0.450$ dengan rata-rata pelanggaran) mengindikasikan bahwa aggressive defending merupakan karakteristik fundamental dalam sepak bola Liga 1. Fenomena ini berbeda dengan tren liga Eropa yang menunjukkan korelasi negatif antara frekuensi tekel dan performa tim, sebagaimana dilaporkan dalam comprehensive analysis oleh Bongiovanni di tahun 2021 (Bongiovanni et al., 2021). Perbedaan ini merefleksikan adaptasi taktis terhadap kondisi spesifik Liga 1 dimana intensitas fisik menjadi faktor kompensatori terhadap keterbatasan teknis.

Validasi Segmentasi Berlapis sebagai Target Variable

Implementasi segmentasi berlapis sebagai variabel target terbukti memberikan insight strategis yang lebih kaya dibandingkan prediksi peringkat absolut, mengkonfirmasi hipotesis H4. Analisis diskriminan menunjukkan bahwa setiap kategori segmentasi memiliki profil performa yang distinct, memungkinkan pengembangan strategi spesifik untuk setiap level kompetitif. Tim kategori Degradasi menunjukkan kombinasi penguasaan bola rendah (47.0%) dan tembakan minimal (3.676), mengindikasikan keterbatasan dalam menciptakan peluang ofensif yang memerlukan intervensi strategis pada aspek build-up play.

Gambar 6. Validasi Target Variable Berlapis
Profil Performa Distinct per Segmentasi
(Validasi Target Variable Berlapis)



Segmentasi berlapis juga memungkinkan identifikasi threshold performa untuk setiap kategori, memberikan benchmark objektif untuk evaluasi progres tim. Tim yang berada di zona Bertahan (Papan Bawah) menunjukkan profil performa yang dapat ditingkatkan melalui fokus pada efisiensi defensif, dengan rata-rata tekel sukses 13.279 yang menunjukkan potensi untuk dikembangkan menjadi kekuatan kompetitif.

Kontribusi Metodologis dan Limitasi Penelitian

Penelitian ini memberikan kontribusi metodologis signifikan melalui demonstrasi aplikasi Random Forest Classifier pada dataset Liga 1 dengan fokus feature importance analysis. Penggunaan stratified k-fold cross-validation dengan class balancing terbukti efektif dalam mengatasi challenges yang terkait dengan dataset berukuran kecil dan distribusi target yang tidak seimbang. Namun, limitasi penelitian perlu diakui, terutama terkait dengan temporal scope yang terbatas pada satu musim dan ketidaktersediaan data kontekstual seperti kondisi cuaca, kualitas wasit, dan faktor home advantage yang dapat mempengaruhi performa tim.

Temuan penelitian ini membuka peluang untuk pengembangan framework evaluasi performa yang disesuaikan dengan karakteristik Liga 1, memberikan foundation empiris untuk pengambilan keputusan strategis dalam manajemen klub dan pengembangan sepak bola Indonesia. Implementasi model prediktif ini dapat diadaptasi untuk analisis real-time selama musim kompetisi, memungkinkan intervensi taktis yang responsif terhadap tren performa tim.

Kesimpulan

Penelitian ini berhasil mengidentifikasi hierarki feature importance dalam konteks BRI Liga 1 musim 2024-2025, dengan rata-rata pelanggaran (23.9%), rata-rata tekel sukses (20.7%), dan rata-rata tembakan ke gawang (18.7%) sebagai tiga variabel paling signifikan dalam menentukan segmentasi tim. Implementasi Random Forest Classifier menghasilkan akurasi cross-validation $65.0\% \pm 22.9\%$, mengkonfirmasi kemampuan model dalam memprediksi segmentasi berlapis meskipun dengan dataset terbatas. Temuan ini memvalidasi hipotesis bahwa karakteristik performa Liga 1 menunjukkan dominasi metrik defensif yang berbeda dari pola liga internasional, dimana aspek fisik dan intensitas defensif menjadi faktor determinan keberhasilan tim (Li & Mu, 2024). Segmentasi berlapis terbukti memberikan insight strategis yang lebih kaya dibandingkan prediksi peringkat absolut, memungkinkan pengembangan framework evaluasi yang disesuaikan dengan karakteristik sepak bola Indonesia. Penelitian lanjutan dapat mengeksplorasi integrasi data temporal dan faktor kontekstual untuk meningkatkan akurasi prediksi serta mengembangkan sistem monitoring real-time untuk aplikasi praktis dalam manajemen klub Liga 1.

Daftar Pustaka

- Alaoui, S. S., Farhaoui, Y. & Aksasse, B. (2018). Classification algorithms in Data Mining. *Article in International Journal of Tomography and Simulation*, 31(4), 1–2.
<https://www.researchgate.net/publication/326866871>
- Albert, Jim., Glickman, M. E., Swartz, T. B. & Koning, R. H. (2017). *Handbook of statistical methods and analyses in sports*. CRC Press, Taylor & Francis.
https://books.google.com/books/about/Handbook_of_Statistical_Methods_and_Anal.html?hl=id&id=_mYIEQAAQBAJ
- Arga Bagas Satria & Dwi Cahyo Utomo. (2025). ANALYSIS OF FACTORS AFFECTING MARKET VALUE OF PROFESSIONAL FOOTBALL PLAYERS IN INDONESIA | Satria | Diponegoro Journal of Accounting. *DIPONEGORO JOURNAL OF ECONOMICS*, 14(2), 1–10.
<https://ejournal3.undip.ac.id/index.php/accounting/article/view/50791>
- Bongiovanni, T., Trecroci, A., Cavaggioni, L., Rossi, A., Perri, E., Pasta, G., Iaia, F. M. & Alberti, G. (2021). Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach. *Research in Sports Medicine*, 29(3), 213–224.
<https://doi.org/10.1080/15438627.2020.1809410>
- Bunker, R. P. & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15(1), 27–33. <https://doi.org/10.1016/J.ACI.2017.09.005>
- Carling, C., Reilly, T. & Williams, A. M. (2008). Performance Assessment for Field Sports. In *Performance Assessment for Field Sports* (1st Edition). Routledge. <https://doi.org/10.4324/9780203890691>
- Carvalho, V. (2023). The assessment of the match performance of association football referees: Identification of key variables. *PLoS ONE*, 18(9). <https://doi.org/10.1371/journal.pone.0291917>

- Casal, C. A., Maneiro, R., Ardá, T., Marí, F. J. & Losada, J. L. (2017). Possession zone as a performance indicator in football. The game of the best teams. *Frontiers in Psychology*, 8(JUL), 251018. <https://doi.org/10.3389/FPSYG.2017.01176/BIBTEX>
- Cortez, A., Trigo, A. & Loureiro, N. (2022). Football Match Line-Up Prediction Based on Physiological Variables: A Machine Learning Approach. *Computers 2022*, Vol. 11, Page 40, 11(3), 40. <https://doi.org/10.3390/COMPUTERS11030040>
- Edalatpanah, S. A. & Hess, M. (2024). Advancing Football Analytics: Predictive Modeling and Performance Analysis in the Bundesliga Using Machine Learning. *Computational Engineering and Technology Innovations*, 1(4), 233–247. <https://doi.org/10.48314/CETI.V1I4.42>
- Franceschi, M., Brocard, J. F., Follert, F. & Gougnet, J. J. (2024). Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys*, 38(3), 577–600. <https://doi.org/10.1111/JOES.12552>
- Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C. & Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*, 14(6), 798–817. <https://doi.org/10.1177/1747954119879350>
- Horvat, T. & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5), e1380. <https://doi.org/10.1002/WIDM.1380>
- Kusmakar, S., Shelyag, S., Zhu, Y., Dwyer, D., Gastin, P. & Angelova, M. (2020). Machine Learning Enabled Team Performance Analysis in the Dynamical Environment of Soccer. *IEEE Access*, 8, 90266–90279. <https://doi.org/10.1109/ACCESS.2020.2992025>
- Lebed, F. (2017). Complex Sport Analytics. In *Complex Sport Analytics* (1st Edition, Vol. 1). Routledge. <https://doi.org/10.4324/9781315692920>
- Li, Y. & Mu, Y. (2024). Research and performance analysis of random forest-based feature selection algorithm in sports effectiveness evaluation. *Scientific Reports 2024 14:1*, 14(1), 1–15. <https://doi.org/10.1038/s41598-024-76706-1>
- Maciej Serda, Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N., Science, P., Sk, S. K., Hinnebusch, R., Hinnebusch A, R., Rabinovich, I., Olmert, Y., Uld, D. Q. G. L. Q., Ri, W. K. H. U., Lq, V., Frxqwu, W. K. H., Zklfk, E., Edvhg, L. V, ... ح, فاطمی. (2013). Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza. *Uniwersytet Śląski*, 7(1), 343–354. <https://doi.org/10.2/JQUERY.MIN.JS>
- Markopoulou, C., Papageorgiou, G. & Tjortjis, C. (2024). Diverse Machine Learning for Forecasting Goal-Scoring Likelihood in Elite Football Leagues. *Machine Learning and Knowledge Extraction 2024*, Vol. 6, Pages 1762-1781, 6(3), 1762–1781. <https://doi.org/10.3390/MAKE6030086>
- Meyk, E. & Unold, O. (2011). Machine learning approach to model sport training. *Computers in Human Behavior*, 27(5), 1499–1506. <https://doi.org/10.1016/J.CHB.2010.10.014>
- Oughali, M. S., Bahloul, M. & El Rahman, S. A. (2019). Analysis of NBA players and shot prediction using random forest and XGBoost Models. *2019 International Conference on Computer and Information Sciences, ICCIS 2019*. <https://doi.org/10.1109/ICCISCI.2019.8716412>
- Pantzalis, V. C. & Tjortjis, C. (2020). Sports Analytics for Football League Table and Player Performance Prediction. *11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020*. <https://doi.org/10.1109/IISA50023.2020.9284352>
- Pugsee, P. & Pattawong, P. (2019). Football match result prediction using the random forest classifier. *ACM International Conference Proceeding Series*, 154–158. <https://doi.org/10.1145/3358528.3358593>
- Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. M. & Baca, A. (2022). Machine learning application in soccer: a systematic review. *Biology of Sport*, 40(1), 249–263. <https://doi.org/10.5114/BIOLSPORT.2023.112970>
- Shobana, G. & Suguna, M. (2021). Sports Prediction Based on Random Forest Algorithm. *Springer Proceedings in Materials*, 5, 993–1000. https://doi.org/10.1007/978-981-15-8319-3_99
- Silitonga, D. J., Adinegara, M. A. Y. C. & Putri, D. M. (2023). Probing the Quality of Football Leagues Through Player's Foot Laterality: A Data Analytics Approach. *JOSSAE (Journal of Sport Science and Education)*, 8(2), 103–112. <https://doi.org/10.26740/JOSSAE.V8N2.P103-112>
- Statista.com. (2024. July). Soccer - Indonesia . <https://Www.Statista.Com>. <https://www.statista.com/outlook/amo/sports/soccer/indonesia>
- Vikram, Y. & Senthilvel, P. G. (2025). Efficient Framework for Sports Result Prediction Using Random Forest

and Compare the Accuracy With XGBoost. *AIP Conference Proceedings*, 3252(1).

<https://doi.org/10.1063/5.0259242/3337909>

Yi, Q. (2022). Interactive Effects of Situational Variables Regarding Teams' Technical Performance in the UEFA Champions League. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.781376>