TECHNOVATAR Jurnal Teknologi, Industri dan Informasi
Vol. 4, No. 4, Oktober 2026, pp. 50 – 60
Open Access:
https://journal.awatarapublisher.com/index.php/technovatar
ISSN: 3031-8874, **DOI**: 10.61434/technovatar.v4i4.388

PREDIKSI HASIL PERTANDINGAN DOTA 2 DENGAN MACHINE LEARNING: STUDI KASUS SATU PEMAIN DAN KONTRAS MODEL PRAPERTANDINGAN DENGAN PASCAPERTANDINGAN

Ahmad Ngiliyun[1]
[1] Program Studi Informatika, Fakultas Manajemen dan Teknik, Universitas Insan Mahardika, Kota Cirebon, Indonesia

e-mail: ahmadngiliyun@gmail.com[1]

INFO ARTIKEL
Sejarah Artikel:
Diterima: 28, September 2026
Direvisi: 01, Oktober 2026
Disetujui: 03, Oktober 2026

KEYWORDS
match outcome prediction
Dota 2
machine learning
data leakage
out-of-sample validation

ABSTRACT

Match outcome prediction in multiplayer online battle arena games is often reported with high accuracy, yet many reported figures rely on information that exists only after a match, and small samples inflate performance. This case study asks what can be predicted for a single player and how much apparent accuracy is optimism. We used 739 ranked All Pick Dota 2 matches (375 wins) of one player, retrieved through the OpenDota API, in three scenarios: pre-match win or loss classification (A), post-match explanatory classification (B), and pre-match regression of gold per minute (GPM) and KDA (C). Validation combined repeated cross-validation on 541 earlier matches for model selection, a single out-of-sample confirmation on 198 later matches, and walk-forward validation. Pre-match models reached AUC of 0.471 to 0.520 in development, 0.572 (95% CI 0.490 to 0.646) in confirmation, and 0.519 to 0.529 in walk-forward validation, close to chance; an AUC of 0.615 from an earlier exploration of 199 matches proved optimistic. The post-match model reached AUC of 0.921 but explains rather than predicts outcomes. Hero choice explained 26 to 30% of GPM variance but little of KDA. Chronological validation is needed to avoid leakage. Results describe one player and should not be generalized to Dota 2 as a whole.

This work is licensed under a Creative Commons Attribution 4.0 International License.

KATA KUNCI
prediksi hasil pertandingan
Dota 2
machine learning
kebocoran data
validasi luar-sampel

ABSTRAK

Prediksi hasil pertandingan pada permainan *multiplayer online battle arena* sering dilaporkan dengan akurasi tinggi, padahal banyak angka yang dilaporkan memakai informasi yang baru ada setelah pertandingan dan sampel kecil menggelembungkan kinerja. Studi kasus ini menanyakan apa

CORRESPONDING AUTHOR
Ahmad Ngiliyun
Universitas Insan Mahardika
Kota Cirebon
ahmadngiliyun@gmail.com

yang dapat diprediksi pada satu pemain dan seberapa besar akurasi yang tampak merupakan optimisme. Data berupa 739 pertandingan Dota 2 *Ranked All Pick* (375 menang) milik satu pemain dari OpenDota API pada tiga skenario: klasifikasi menang atau kalah prapertandingan (A), klasifikasi penjelas pascapertandingan (B), dan regresi gold per minute (GPM) serta KDA prapertandingan (C). Validasi menggabungkan validasi silang berulang pada 541 pertandingan awal untuk pemilihan model, satu kali konfirmasi luar-sampel pada 198 pertandingan yang lebih baru, dan validasi *walk-forward*. Model prapertandingan mencapai AUC 0,471 sampai 0,520 pada pengembangan, 0,572 (CI 95% 0,490 sampai 0,646) pada konfirmasi, dan 0,519 sampai 0,529 pada *walk-forward*, mendekati peluang; AUC 0,615 dari eksplorasi awal 199 pertandingan terbukti optimistis. Model pascapertandingan mencapai AUC 0,921 tetapi menjelaskan, bukan memprediksi, hasil. Pilihan *hero* menjelaskan 26 sampai 30% variasi GPM tetapi sedikit variasi KDA. Validasi kronologis diperlukan untuk menghindari kebocoran data. Hasil menggambarkan satu pemain dan tidak dapat digeneralisasi ke Dota 2 secara umum.

PENDAHULUAN

Esports telah berkembang menjadi industri hiburan dan kompetisi yang menghasilkan data pertandingan dalam jumlah besar. Dota 2 adalah permainan *multiplayer online battle arena* (MOBA) yang catatan pertandingannya mudah diakses melalui antarmuka pemrograman aplikasi terbuka seperti OpenDota (OpenDota, n.d.). Dua tim beranggotakan lima pemain, Radiant dan Dire, saling menghancurkan bangunan inti lawan dengan memilih *hero* dari lebih dari seratus pilihan, sehingga hasil pertandingan dipengaruhi oleh pemilihan *hero*, keterampilan individual, koordinasi tim, dan faktor acak. Kondisi ini menjadikan prediksi hasil pertandingan sebagai persoalan *machine learning* yang menarik sekaligus sulit.

Literatur prediksi hasil pertandingan Dota 2 dan permainan MOBA banyak memakai informasi yang berkembang selama pertandingan berlangsung. Hodge et al. (2021) memprediksi hasil pertandingan profesional secara langsung, Zhao et al. (2022) mengusulkan model prediksi kemenangan waktu nyata untuk permainan MOBA, dan Ke et al. (2022) memanfaatkan model pertarungan tim berbasis pembelajaran mendalam. Xenopoulos et al. (2022) memakai probabilitas kemenangan untuk menganalisis perbedaan *esports* profesional dan amatir, sedangkan Pedrassoli Chitayat et al. (2023) memanfaatkan parameter desain permainan agar analitik tidak bergantung pada versi *patch*. Penelitian yang lebih baru mengevaluasi ansambel pohon keputusan (Yangibaev et al., 2025), klasifikasi yang dapat dijelaskan pada siaran langsung (García-Méndez dan de Arriba-Pérez, 2025), dan *pipeline machine learning* untuk mengidentifikasi kemampuan pemain (Pengmatchaya dan Natwichai, 2024). Di Indonesia, prediksi kemenangan Dota 2 telah dikaji dengan data pertandingan (Atmaja, 2020), *Random Forest* atau XGBoost (Fandi et al., 2024), dan *Fuzzy Tsukamoto* (Tan et al., 2025). Perbandingan beberapa algoritma *machine learning* juga lazim pada penelitian terapan di Indonesia, misalnya untuk prediksi hipertensi (Permana et al., 2026), prediksi indeks prestasi kumulatif mahasiswa dengan validasi silang lima lipatan (Malik et al., 2026), serta klasifikasi sentimen yang dievaluasi dengan akurasi dan AUC (Rosmawan et al., 2024; Arlovin et al., 2026), termasuk analisis sentimen yang dikemas sebagai aplikasi untuk pemberdayaan digital masyarakat (Ngiliyun dan Nurharidhwan, 2025).

Meskipun demikian, angka akurasi yang dilaporkan sulit dibandingkan karena informasi yang dipakai tersedia pada waktu yang berbeda. Model yang memakai statistik pertandingan seperti *kills*, *deaths*, dan pendapatan emas pada dasarnya menggambarkan hasil yang sedang terjadi dan

bukan meramalkannya. Sebaliknya, prediksi sebelum pertandingan dimulai, ketika hanya pilihan *hero*, sisi, waktu bermain, dan riwayat pemain yang diketahui, sepengetahuan penulis belum banyak dievaluasi dengan ketat pada satu pemain. Data yang kecil menambah kesulitan: memilih model terbaik dari banyak kombinasi pada sampel yang sama menghasilkan estimasi kinerja yang optimistis (Cawley dan Talbot, 2010), validasi silang tidak sepenuhnya mengestimasi galat model yang sedang dipasang (Bates et al., 2024), dan kebocoran data telah dilaporkan menggelembungkan kinerja di banyak bidang (Kapoor dan Narayanan, 2023; Rosenblatt et al., 2024; Whalen et al., 2022).

Penelitian ini mengambil pendekatan studi kasus pada satu pemain, yaitu penulis sendiri, dengan 739 pertandingan *Ranked All Pick* yang diambil dari OpenDota. Pertanyaan penelitiannya ada tiga: (1) seberapa jauh hasil menang atau kalah dapat diprediksi dari informasi yang tersedia sebelum pertandingan; (2) seberapa besar kinerja meningkat, dan apa artinya, ketika informasi pascapertandingan dipakai; dan (3) seberapa besar variasi performa individual, yang diukur melalui gold per minute (GPM) dan rasio KDA, dapat dijelaskan oleh fitur prapertandingan. Temuan berlaku pada tingkat profil satu pemain dengan fitur yang terbatas dan bukan klaim bahwa hasil pertandingan Dota 2 secara umum tidak dapat diprediksi sebelum pertandingan dimulai.

Kontribusi penelitian ini ada tiga. Pertama, taksonomi fitur yang memisahkan informasi prapertandingan dari pascapertandingan sehingga kebocoran informasi dapat diaudit. Kedua, protokol validasi tiga tahap, yaitu validasi silang berulang untuk pemilihan model, uji konfirmasi luar-sampel satu kali pada data yang lebih baru, dan *walk-forward* kronologis, yang dilaporkan lengkap dengan interval kepercayaan. Ketiga, pelaporan terbuka atas optimisme sampel kecil: AUC 0,615 dari eksplorasi awal pada 199 pertandingan menyusut menjadi sekitar 0,52 pada evaluasi kronologis. Sepengetahuan penulis, kontras eksplisit antara prediksi prapertandingan dan penjelasan pascapertandingan pada data satu pemain dengan protokol semacam ini belum banyak dilaporkan, sehingga naskah ini diposisikan sebagai studi kasus metodologis dan bukan sebagai model yang siap dipakai.

METODE PENELITIAN

Desain penelitian dan sumber data

Penelitian ini merupakan studi kasus kuantitatif komputasional pada satu pemain. Data diambil dari riwayat pertandingan akun penulis melalui OpenDota API (OpenDota, n.d.), yaitu daftar pertandingan (974 pertandingan) dan rincian setiap pertandingan yang mencakup hasil, *hero* seluruh pemain, sisi, durasi, dan statistik akhir pemain. Identitas akun dan pemain lain tidak dipakai dan tidak dilaporkan. Populasi analisis dibatasi pada mode *All Pick* dengan lobi *Ranked* dan pertandingan tempat pemain tidak meninggalkan permainan (*leaver*), sehingga diperoleh 739 pertandingan (375 menang dan 364 kalah) antara 20 September 2020 dan 20 April 2026 dengan 75 *hero* berbeda. Mode lain, misalnya *Turbo*, dikeluarkan karena struktur ekonomi dan durasinya berbeda. Waktu pertandingan dicatat dalam zona waktu WIB.

Sebagai eksplorasi awal, penulis pernah mencatat 199 pertandingan secara manual. Catatan itu tidak dipakai untuk pemodelan akhir dan hanya muncul pada pembahasan tentang optimisme sampel kecil.

Fitur dan skenario

Fitur dibagi menurut saat ia dapat diketahui (Tabel 1). Fitur prapertandingan hanya memuat informasi yang sudah ada sebelum pertandingan berjalan, sedangkan fitur pascapertandingan dihasilkan oleh jalannya pertandingan. Fitur riwayat pribadi dihitung hanya dari pertandingan sebelumnya, dan *win rate* pada *hero* yang sama dihaluskan ke 0,5 dengan bobot lima pertandingan. Komposisi memuat jumlah sekutu dan lawan menurut atribut *hero* tanpa identitas pemain.

Tabel 1 Set fitur dan saat fitur tersedia

Kelompok	Fitur	Tersedia
A0	Sisi (Radiant atau Dire)	Sebelum pemilihan *hero*
A1	*Hero* yang dimainkan, sisi, bagian hari, hari dalam minggu, penanda malam	Saat pemilihan *hero*
A2	A1 ditambah jumlah sekutu dan lawan menurut atribut utama, tipe serangan, dan peran *hero*, serta atribut dan tipe serangan *hero* sendiri	Setelah pemilihan *hero* selesai
A3	A2 ditambah riwayat pribadi (*win rate* 10 dan 20 pertandingan *ranked* terakhir, *streak*, jumlah dan *win rate* pada *hero* yang sama, jeda dari pertandingan sebelumnya, jumlah pertandingan 24 jam terakhir) dan rata-rata *rank* lobi	Setelah pemilihan *hero* selesai
B	*kills*, *deaths*, *assists*, *last hits*, *denies*, GPM, XPM, *level*, *net worth*, dan durasi	Setelah pertandingan
C	Fitur A3 ditambah rata-rata GPM historis *hero* yang sama; target GPM dan log(1 + KDA)	Fitur: setelah pemilihan *hero*; target: setelah pertandingan

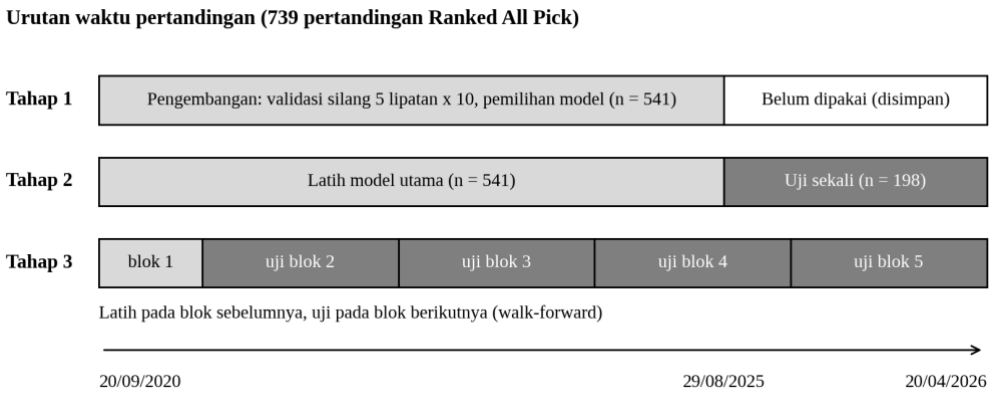
Skenario A adalah klasifikasi biner menang atau kalah dari fitur prapertandingan dengan empat set fitur berjenjang (A0 sampai A3). Skenario B adalah klasifikasi menang atau kalah dari fitur pascapertandingan yang diperlakukan sebagai model penjelas dan bukan prediktor. Skenario C adalah regresi GPM dan log(1 + KDA) dari fitur prapertandingan, dengan KDA didefinisikan sebagai (*kills* + *assists*) dibagi *deaths*, dan *deaths* minimal satu. Skenario C diuji dengan tiga kondisi fitur, yaitu semua fitur, hanya *hero*, dan tanpa *hero*, untuk memisahkan kontribusi pilihan *hero*.

Model

Tiga model diuji pada klasifikasi: regresi logistik dengan regularisasi L2 (C = 0,3, fitur distandarkan), *Random Forest* (300 pohon, daun minimal 5 sampel, akar kuadrat jumlah fitur per pemecahan), dan *Gradient Boosting* (150 pohon, kedalaman 2, *learning rate* 0,05, subsampel 0,8). Regresi memakai Ridge (alpha 10), *Random Forest*, dan *Gradient Boosting*. Seluruh model dijalankan dengan scikit-learn versi 1.8.0 (Pedregosa et al., 2011) dan *seed* 42. Pemilihan *Random Forest* dan *Gradient Boosting* di samping regresi logistik mengikuti praktik perbandingan ansambel pohon pada penelitian prediksi lain (Malik et al., 2026; Yangibaev et al., 2025). Pra-pemrosesan (imputasi median, penskalaan, pengodean satu-panas) dipasang hanya pada data latih di setiap lipatan.

Desain validasi dan evaluasi

Validasi dirancang bertingkat (Gambar 1). Tahap 1 memakai validasi silang berstrata 5 lipatan yang diulang 10 kali pada 541 pertandingan sebelum 29 Agustus 2025. Seluruh pemilihan model dilakukan hanya pada tahap ini, dan kombinasi utama adalah yang memiliki AUC tertinggi. Tahap 2 melatih kombinasi utama pada 541 pertandingan tersebut dan mengujinya satu kali pada 198 pertandingan sejak 29 Agustus 2025 tanpa pemilihan ulang. Kombinasi lain juga dilaporkan pada data konfirmasi, tetapi hanya kombinasi utama yang diperlakukan sebagai uji konfirmasi. Tahap 3 adalah *walk-forward*: data diurutkan menurut waktu dan dibagi lima blok, model dilatih pada blok sebelumnya dan diuji pada blok berikutnya.



Gambar 1 Desain validasi tiga tahap pada 739 pertandingan yang diurutkan menurut waktu

Metrik utama adalah AUC, dilengkapi akurasi, *balanced accuracy*, dan skor Brier, dibandingkan dengan *baseline* kelas mayoritas. Interval kepercayaan (CI) 95% diperoleh dengan *bootstrap*, dan AUC diuji terhadap 0,5 dengan uji Mann-Whitney satu arah. Untuk skenario B dipakai validasi silang berstrata 5 lipatan yang diulang 20 kali dan pemisahan kronologis pada titik potong yang sama. Kepentingan fitur diukur dengan kepentingan permutasi pada *Random Forest*, dan koefisien regresi logistik terstandar diestimasi dengan *bootstrap* 500 kali. Untuk skenario C dipakai R^2 dan galat absolut rata-rata (MAE) pada validasi silang 5 lipatan yang diulang 10 kali dan pada pemisahan kronologis.

Ada satu ancaman validitas yang diakui. Fitur riwayat dan rata-rata GPM historis *hero* dihitung dari pertandingan sebelumnya, sehingga pada validasi silang acak fitur pertandingan latih yang lebih baru memuat hasil pertandingan uji yang lebih lama. Kebocoran ringan ini tidak muncul pada uji luar-sampel dan *walk-forward*. Karena itu, angka validasi silang tidak dijadikan rujukan akhir, dan rujukan utama adalah evaluasi kronologis.

DISKUSI DAN PEMBAHASAN

Karakteristik data

Dari 739 pertandingan, 375 (50,7%; CI 95% Wilson 47,1% sampai 54,3%; uji binomial terhadap 50%, p = 0,713) adalah kemenangan, sehingga kelas seimbang dan tingkat menang pemain tidak berbeda dari peluang. Durasi pertandingan rata-rata 44,41 menit (SD 9,27), GPM rata-rata 502,0 (SD 127,9), dan KDA rata-rata 3,23 (SD 3,44). Pada sisi Radiant terdapat 388 pertandingan dengan tingkat menang 52,1%, dan pada Dire 351 pertandingan dengan 49,3%. Pemilihan *hero* terkonsentrasi: Lifestealer dimainkan 151 kali (20,4%) dan Chaos Knight 76 kali (10,3%).

Selisih sisi ini jauh lebih kecil daripada yang tampak pada catatan manual awal (199 pertandingan), yaitu Radiant 60,6% dan Dire 43,2% (uji permutasi p = 0,0173; setelah koreksi Benjamini-Hochberg q = 0,183). Pola waktu dan psikologis juga tidak signifikan pada catatan manual itu: menang pada permainan setelah bekerja 57,4% dibanding 50,3% pada permainan tanpa kesibukan sebelumnya (p = 0,4269), menang setelah kemenangan 49,0% dibanding setelah kekalahan 55,3% (uji Fisher p = 0,396), dan uji runs terhadap urutan hasil p = 0,417.

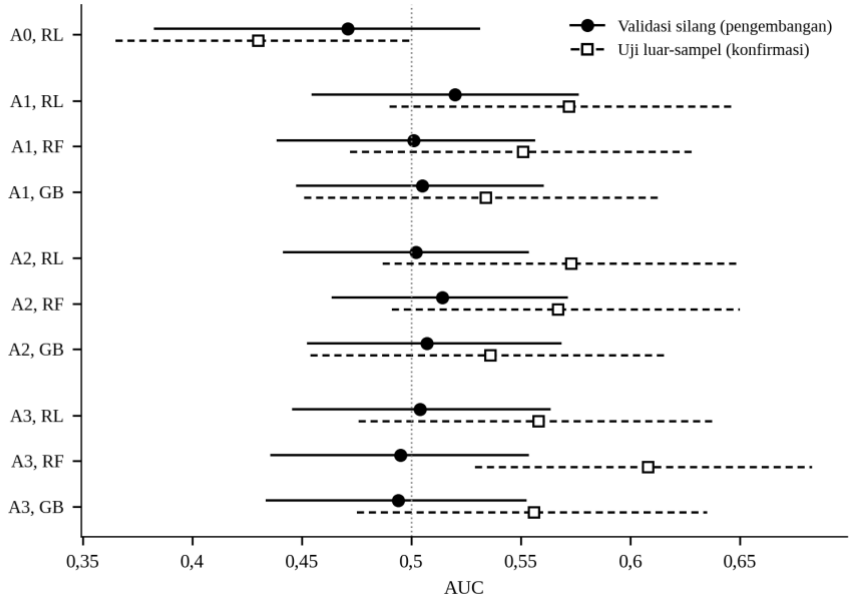
Skenario A: prediksi prapertandingan

Tabel 2 dan Gambar 2 merangkum skenario A. Pada validasi silang tahap pengembangan, seluruh kombinasi berada pada AUC 0,471 sampai 0,520 dan seluruh interval kepercayaan memuat 0,5. Kombinasi dengan AUC tertinggi adalah regresi logistik pada A1 (*hero*, sisi, waktu; AUC 0,520; CI 95% 0,455 sampai 0,576) sehingga menjadi kombinasi utama. Penambahan komposisi tim dan lawan (A2) serta riwayat pribadi dan *rank* lobi (A3) tidak meningkatkan AUC pengembangan

(0,494 sampai 0,514). Sisi saja (A0) menghasilkan AUC di bawah 0,5 (0,471), yang menunjukkan bahwa pola sisi pada data latih tidak berulang pada data uji.

Tabel 2 AUC skenario A pada validasi silang pengembangan dan uji konfirmasi luar-sampel (baris tebal: kombinasi utama)

Set	Model	AUC validasi silang (CI 95%)	AUC konfirmasi (CI 95%)	Akurasi konfirmasi	p satu arah
A0	Regresi logistik	0,471 (0,383-0,531)	0,430 (0,365-0,500)	0,429	0,975
A1	**Regresi logistik**	**0,520 (0,455-0,576)**	**0,572 (0,490-0,646)**	**0,556**	**0,040**
A1	*Random Forest*	0,501 (0,439-0,556)	0,551 (0,472-0,630)	0,525	0,107
A1	*Gradient Boosting*	0,505 (0,448-0,560)	0,534 (0,451-0,614)	0,500	0,203
A2	Regresi logistik	0,502 (0,442-0,553)	0,573 (0,487-0,649)	0,535	0,039
A2	*Random Forest*	0,514 (0,464-0,571)	0,567 (0,491-0,650)	0,540	0,052
A2	*Gradient Boosting*	0,507 (0,453-0,568)	0,536 (0,454-0,617)	0,500	0,194
A3	Regresi logistik	0,504 (0,446-0,563)	0,558 (0,476-0,639)	0,530	0,079
A3	*Random Forest*	0,495 (0,436-0,553)	0,608 (0,529-0,683)	0,530	0,004
A3	*Gradient Boosting*	0,494 (0,434-0,552)	0,556 (0,475-0,635)	0,530	0,087



Gambar 2 AUC skenario A pada validasi silang pengembangan dan uji konfirmasi (RL: regresi logistik, RF: *Random Forest*, GB: *Gradient Boosting*)

Pada uji konfirmasi 198 pertandingan, kombinasi utama menghasilkan AUC 0,572 (CI 95% 0,490 sampai 0,646), akurasi 0,556 terhadap *baseline* 0,515, dan p satu arah 0,040. Hasil ini menunjukkan sinyal yang lemah dan belum meyakinkan, karena batas bawah interval menyentuh 0,49 dan p hanya sedikit di bawah 0,05 pada satu-satunya uji yang dipilih sebelumnya. Kombinasi lain (di luar A0) memiliki AUC konfirmasi 0,534 sampai 0,608. Nilai tertinggi, 0,608 pada *Random Forest* A3 (p = 0,004), tidak diklaim sebagai temuan karena kombinasi itu hanya memiliki AUC 0,495 pada tahap pengembangan dan terlihat setelah data konfirmasi dibuka. Angka semacam itu adalah contoh pilihan yang menggelembungkan estimasi kinerja (Cawley dan Talbot, 2010).

Evaluasi *walk-forward* pada 591 pertandingan menghasilkan AUC gabungan 0,519 sampai 0,529 pada empat kombinasi A2 dan A3 (Tabel 3), dengan batas bawah CI 95% antara 0,475 dan 0,484 dan akurasi gabungan 0,497 sampai 0,530 terhadap *baseline* 0,501. AUC per blok berkisar 0,454 sampai 0,583 dan berganti-ganti di sekitar 0,5 dari satu blok ke blok berikutnya, sehingga tidak ada bukti keunggulan yang stabil. Perbedaan antara 0,572 pada konfirmasi dan sekitar 0,52

pada *walk-forward* mengingatkan bahwa 198 pertandingan menghasilkan interval selebar sekitar 0,16, yang tidak cukup sempit untuk membedakan sinyal lemah dari kebetulan.

Tabel 3 AUC *walk-forward* pada 591 pertandingan (blok 2 sampai 5)

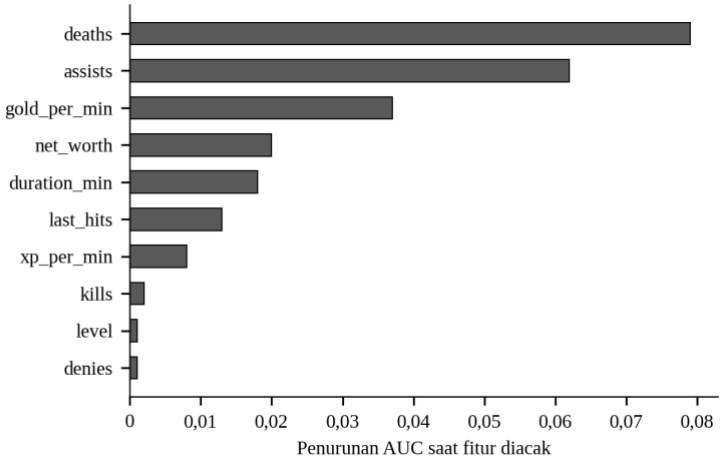
Set	Model	AUC gabungan (CI 95%)	AUC per blok (rentang)	Akurasi gabungan	Akurasi *baseline*
A2	Regresi logistik	0,529 (0,484-0,577)	0,454-0,583	0,530	0,501
A2	*Random Forest*	0,525 (0,477-0,572)	0,471-0,563	0,513	0,501
A3	Regresi logistik	0,519 (0,475-0,564)	0,461-0,576	0,523	0,501
A3	*Random Forest*	0,522 (0,476-0,567)	0,472-0,569	0,497	0,501

Skenario B: model pascapertandingan

Ketika fitur pascapertandingan dipakai (Tabel 4), AUC melonjak menjadi 0,921 (CI 95% 0,899 sampai 0,941) untuk regresi logistik dengan akurasi 0,849 terhadap *baseline* 0,507. *Random Forest* dan *Gradient Boosting* mencapai 0,905 dan 0,901. Pemisahan kronologis memberi hasil setara (AUC 0,919 untuk regresi logistik), sehingga hasil ini stabil terhadap waktu. Selisih sekitar 0,4 poin AUC dibanding skenario A tidak menunjukkan model yang lebih pintar, melainkan bahwa fitur skenario B adalah akibat dari hasil pertandingan: tim yang menang cenderung mengalami lebih sedikit *deaths*, mencatat lebih banyak *assists*, dan mengumpulkan emas lebih besar. Pola ini sejalan dengan pendekatan prediksi langsung yang memperlakukan kinerja sebagai fungsi kemajuan pertandingan (Hodge et al., 2021).

Tabel 4 Kinerja model pascapertandingan pada 739 pertandingan

Model	AUC validasi silang (CI 95%)	Akurasi validasi silang	AUC kronologis (CI 95%)	Akurasi kronologis
Regresi logistik	0,921 (0,899-0,941)	0,849	0,919 (0,878-0,957)	0,848
Random Forest	0,905 (0,882-0,927)	0,818	0,890 (0,843-0,928)	0,783
Gradient Boosting	0,901 (0,876-0,924)	0,812	0,884 (0,836-0,924)	0,773



Gambar 3 Kepentingan permutasi fitur pascapertandingan pada *Random Forest* (penurunan AUC)

Kepentingan permutasi (Gambar 3) menempatkan *deaths* (penurunan AUC 0,079), *assists* (0,062), dan GPM (0,037) di puncak, sedangkan *kills*, *level*, dan *denies* hampir tidak berkontribusi (masing-masing 0,002; 0,001; dan 0,001). Koefisien regresi logistik terstandar tidak stabil karena fitur saling berkorelasi kuat: *net worth* bernilai positif (1,582) sedangkan *last hits* negatif (-1,152). Pada 739 pertandingan, korelasi Pearson antarfitur skenario B berkisar 0,00 sampai 0,87 dalam nilai mutlak (11 dari 45 pasangan bernilai 0,70 atau lebih; tertinggi XP per menit dengan *level*, 0,87), dan faktor inflasi varians (VIF) melampaui 10 untuk *net worth* (24,1), GPM (16,8), *level* (16,2), durasi

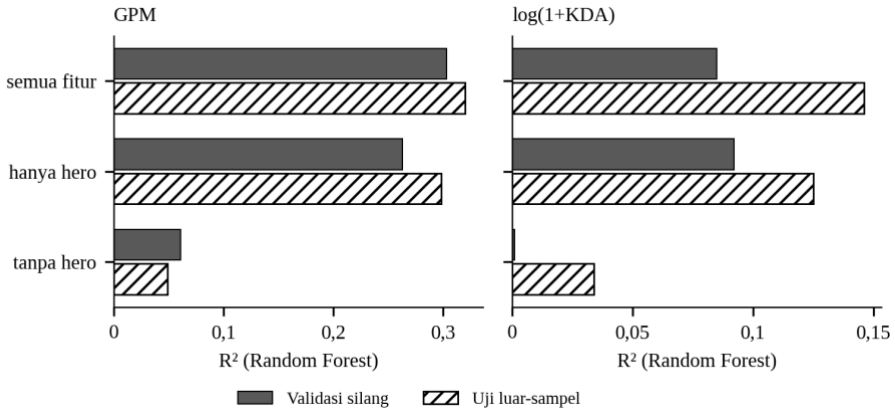
(16,1), dan XP per menit (10,8). *Net worth* berkorelasi 0,86 dengan *last hits* dan dengan GPM, sehingga koefisien keduanya saling mengimbangi. Tanda koefisien individual karena itu tidak boleh ditafsirkan sebagai pengaruh kausal, dan model B sebaiknya dibaca sebagai deskripsi hubungan statistik.

Skenario C: performa dari fitur prapertandingan

Tabel 5 dan Gambar 4 memperlihatkan bahwa pilihan *hero* menjelaskan sebagian besar variasi yang dapat dijelaskan pada skenario C. Untuk GPM, *Random Forest* dengan semua fitur prapertandingan mencapai R^2 0,303 pada validasi silang dan 0,320 pada uji luar-sampel dengan MAE sekitar 86 poin GPM dibanding 106 poin untuk *baseline* rata-rata. *Hero* saja menghasilkan R^2 0,263 dan 0,298, sedangkan tanpa *hero* R^2 turun menjadi 0,061 dan 0,049. Pilihan *hero* karena itu menjelaskan sekitar 26 sampai 30% variasi GPM, yang masuk akal karena peran *hero* menentukan pola ekonominya.

Tabel 5 R^2 dan MAE skenario C untuk GPM dan log(1 + KDA)

Target	Set fitur	Model	R^2 validasi silang	R^2 luar-sampel	MAE validasi silang	MAE luar-sampel
GPM	-	*Baseline*	-0,002	0,000	106,4	106,4
	Semua fitur	*Random Forest*	0,303	0,320	86,0	85,9
	Hanya *hero*	*Random Forest*	0,263	0,298	88,5	86,3
	Tanpa *hero*	*Random Forest*	0,061	0,049	101,8	102,6
log(1+KDA)	-	*Baseline*	-0,003	-0,003	0,410	0,392
	Semua fitur	*Random Forest*	0,085	0,146	0,393	0,369
	Hanya *hero*	*Random Forest*	0,092	0,125	0,390	0,377
	Tanpa *hero*	*Random Forest*	0,001	0,034	0,410	0,382



Gambar 4 R^2 *Random Forest* untuk GPM dan log(1 + KDA) menurut set fitur

Untuk log(1 + KDA), R^2 hanya 0,085 pada validasi silang dan 0,146 pada uji luar-sampel, dan tanpa *hero* mendekati nol (0,001 dan 0,034). KDA lebih ditentukan oleh jalannya pertandingan, seperti keberhasilan pertarungan tim dan kesalahan posisi, yang tidak diketahui sebelum pertandingan. Ridge dan *Gradient Boosting* memberi hasil yang setara (R^2 uji luar-sampel GPM dengan semua fitur masing-masing 0,363 dan 0,346). Fitur *hero_prior_gpm* membawa kebocoran ringan pada validasi silang acak, sehingga R^2 uji luar-sampel dijadikan rujukan.

Optimisme sampel kecil dan implikasi metodologis

Eksplorasi awal pada 199 pertandingan hasil pencatatan manual melaporkan AUC terbaik 0,615 (CI 95% 0,521 sampai 0,701; p permutasi 0,003). Angka itu diperoleh dengan memilih kombinasi terbaik dari banyak kombinasi model dan fitur pada sampel yang sama, sehingga memuat

bias seleksi (Cawley dan Talbot, 2010). Setelah data diperluas dan dievaluasi secara kronologis, AUC menyusut menjadi sekitar 0,52 (Tabel 2 dan Tabel 3). Selisih Radiant dan Dire yang tampak besar pada 199 pertandingan (60,6% dan 43,2%) juga mengecil menjadi 52,1% dan 49,3%. Kedua contoh sejalan dengan peringatan bahwa kebocoran data dan pemilihan model pada sampel yang sama dapat menggelembungkan kinerja (Kapoor dan Narayanan, 2023; Rosenblatt et al., 2024; Whalen et al., 2022). Secara praktis, laporan prediksi hasil pertandingan sebaiknya menyebut waktu ketersediaan setiap fitur, apakah pemilihan model terpisah dari data uji, dan interval kepercayaan, dan bukan hanya akurasi tunggal.

Keterbatasan

Penelitian ini memiliki beberapa keterbatasan. Pertama, data berasal dari satu pemain sehingga hasil tidak dapat digeneralisasi ke pemain lain, kelas peringkat lain, atau pertandingan profesional. Kedua, pilihan *hero* terkonsentrasi (Lifestealer dan Chaos Knight menyumbang 227 dari 739 pertandingan atau 30,7%), sehingga estimasi untuk *hero* yang jarang dimainkan tidak stabil. Ketiga, rentang 2020 sampai 2026 mencakup banyak *patch* yang mengubah keseimbangan permainan, sehingga hubungan antara fitur dan hasil tidak stasioner. Keempat, identitas dan peringkat sekutu dan lawan tidak dipakai padahal kekuatan tim sangat menentukan hasil, dan hanya komposisi agregat serta *rank* rata-rata lobi yang tersedia. Kelima, uji konfirmasi hanya 198 pertandingan sehingga interval kepercayaan lebar, dan pelaporan seluruh kombinasi pada data konfirmasi menimbulkan persoalan perbandingan berganda. Keenam, kebocoran ringan pada validasi silang acak untuk fitur riwayat tidak dihilangkan, melainkan diatasi dengan evaluasi kronologis.

KESIMPULAN

Penelitian ini menunjukkan tiga hal pada 739 pertandingan *Ranked All Pick* satu pemain. Pertama, prediksi hasil menang atau kalah dari informasi prapertandingan hampir tidak lebih baik daripada peluang: AUC 0,471 sampai 0,520 pada validasi silang, 0,572 (CI 95% 0,490 sampai 0,646) pada uji konfirmasi satu kali, dan 0,519 sampai 0,529 pada *walk-forward*. Kedua, model pascapertandingan mencapai AUC 0,921, tetapi hanya menjelaskan hasil yang telah terjadi, terutama melalui *deaths*, *assists*, dan GPM. Ketiga, pilihan *hero* menjelaskan sekitar 26 sampai 30% variasi GPM tetapi sedikit variasi KDA.

Dari sisi metodologi, AUC 0,615 dari eksplorasi awal terbukti optimistis, sehingga pemisahan fitur menurut waktu, pemilihan model yang terpisah dari data uji, dan validasi kronologis perlu menjadi praktik standar pada penelitian prediksi hasil pertandingan. Penelitian lanjutan dapat menambah jumlah pemain, memasukkan fitur *draft* yang lebih kaya dan kekuatan lawan, memakai fitur yang sadar terhadap *patch*, serta memanfaatkan probabilitas yang terkalibrasi untuk keperluan analitik dan bukan hanya klasifikasi.

UCAPAN TERIMA KASIH

Penulis berterima kasih kepada proyek OpenDota atas antarmuka pemrograman aplikasi terbuka yang memungkinkan pengambilan data pertandingan.

KONTRIBUSI PENULIS

Ahmad Ngiliyun: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - Original Draft, dan Writing - Review and Editing. Penulis telah membaca dan menyetujui versi akhir dari naskah ini.

PERNYATAAN KETERSEDIAAN DATA

Data yang mendukung temuan dalam penelitian ini tersedia dari para penulis atas permintaan yang wajar, dengan mempertimbangkan persetujuan etik dan peraturan institusi yang berlaku.

PERNYATAAN BEBAS KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak ada konflik kepentingan finansial maupun hubungan pribadi yang dapat mempengaruhi hasil yang dilaporkan dalam naskah ini.

PERNYATAAN ETIKA PENELITIAN DAN PUBLIKASI

Para penulis menyatakan bahwa penelitian dan penulisan naskah ini telah mematuhi standar etika penelitian dan publikasi, sesuai dengan prinsip ilmiah, serta bebas dari plagiasi.

PERNYATAAN PEMANFAATAN TEKNOLOGI ASISTIF

Penulis menyatakan bahwa Kecerdasan Buatan Generatif (*Generative Artificial Intelligence*) dan teknologi asistif lainnya tidak digunakan secara berlebihan dalam proses penelitian dan penulisan naskah ini. Ide penelitian, rumusan masalah, desain penelitian, pengumpulan data, dan penarikan kesimpulan sepenuhnya berasal dari penulis. Secara khusus, Claude (Anthropic) digunakan untuk tata tulis dan penyuntingan bahasa, pemeriksaan kosakata, pengembangan gagasan dari kerangka yang disusun penulis, pembuatan diagram dan visualisasi data, serta bantuan penyusunan skrip analisis Python yang ditinjau dan dijalankan oleh penulis. Penulis telah meninjau dan menyunting semua konten yang dihasilkan AI guna memastikan ketepatan, kelengkapan, serta kepatuhan terhadap standar etika dan ilmiah. Penulis bertanggung jawab penuh atas naskah versi akhir.

REFERENSI

Arlovin, T., Sukarsa, & Sulaeman, D. A. (2026). Analyzing public sentiment toward the establishment of Universitas Republik Indonesia using Naïve Bayes algorithm on YouTube comments. *TECHNOVATAR Jurnal Teknologi, Industri, dan Informasi*, *4*(3), 1-9. https://doi.org/10.61434/technovatar.v4i2.379

Atmaja, E. H. S. (2020). Prediksi kemenangan eSport DOTA 2 berdasarkan data pertandingan. *AVITEC*, *2*(1), 31-38. https://doi.org/10.28989/avitec.v2i1.612

Bates, S., Hastie, T., & Tibshirani, R. (2024). Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association*, *119*(546), 1434-1445. https://doi.org/10.1080/01621459.2023.2197686

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*, 2079-2107.

Fandi, R. S., Arifiyanti, A. A., & Wati, S. F. A. (2024). Evaluasi dan prediksi hasil pertandingan Dota 2 menggunakan algoritma terbaik dari Random Forest atau XGBoost. *Scientica: Jurnal Ilmiah Sains dan Teknologi*, *2*(11), 63-73.

García-Méndez, S., & de Arriba-Pérez, F. (2025). Explainable e-sports win prediction through machine learning classification in streaming. *Entertainment Computing*, *55*, 101027. https://doi.org/10.1016/j.entcom.2025.101027

Hodge, V. J., Devlin, S., Sephton, N., Block, F., Cowling, P. I., & Drachen, A. (2021). Win prediction in multiplayer esports: Live professional match prediction. *IEEE Transactions on Games*, *13*(4), 368-379. https://doi.org/10.1109/TG.2019.2948469

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, *4*(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

Ke, C. H., Deng, H., Xu, C., Li, J., Gu, X., Yadamsuren, B., Klabjan, D., Sifa, R., Drachen, A., & Demediuk, S. (2022). DOTA 2 match prediction through deep learning team fight models.

Dalam *2022 IEEE Conference on Games (CoG)* (hlm. 96-103). IEEE. https://doi.org/10.1109/CoG51982.2022.9893647

Malik, H. H., Permana, I. S., Muhammad, A. H., & Kusnawi, K. (2026). Comparison of Bayesian Markov chain Monte Carlo and machine learning algorithms for student cumulative grade point average prediction with feature engineering. *Bulletin of Network Engineer and Informatics*, *4*(1), 377-386. https://doi.org/10.59688/vfk6qw80

Ngiliyun, A., & Nurharidhwan, N. (2025). Pengembangan aplikasi DataCraft untuk analisis sentimen media sosial sebagai upaya pemberdayaan digital masyarakat. *ASMARA: Jurnal Aspirasi Masyarakat*, *2*(1), 74-81.

OpenDota. (n.d.). *OpenDota API documentation*. Diakses 30 September 2026, dari https://docs.opendota.com/

Pedrassoli Chitayat, A., Block, F., Walker, J., & Drachen, A. (2023). Beyond the meta: Leveraging game design parameters for patch-agnostic esport analytics. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, *19*(1), 116-125. https://doi.org/10.1609/aiide.v19i1.27507

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Pengmatchaya, M., & Natwichai, J. (2024). Identifying player skill of Dota 2 using machine learning pipeline. *Discover Artificial Intelligence*, *4*(1), 41. https://doi.org/10.1007/s44163-024-00139-y

Permana, I. S., Ngiliyun, A., Nurhaeni, A., Marisa, D. E., Rahayu, R., & Oktiany, T. (2026). Perbandingan algoritma machine learning dalam prediksi hipertensi berbasis data Indonesian Family Life Survey (IFLS5). *Jurnal Ilmu Kesehatan dan Keperawatan*, *4*(1), 301-312.

Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., & Scheinost, D. (2024). Data leakage inflates prediction performance in connectome-based machine learning models. *Nature Communications*, *15*(1), 1829. https://doi.org/10.1038/s41467-024-46150-w

Rosmawan, H., Setyanto, S., & Wibowo, F. W. (2024). Implementation of SVM-PSO in sentiment analysis of Google Place review users at Cafe Headquarters. *TECHNOVATAR Jurnal Teknologi, Industri, dan Informasi*, *2*(4), 97-110. https://doi.org/10.61434/technovatar.v2i4.230

Tan, M. A. A., Yulindawati, & Fahmi, M. (2025). Analisis dan prediksi hasil pertandingan Dota 2 menggunakan Fuzzy Tsukamoto. *Bulletin of Information Technology (BIT)*, *6*(4), 432-438. https://doi.org/10.47065/bit.v6i4.2382

Whalen, S., Schreiber, J., Noble, W. S., & Pollard, K. S. (2022). Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics*, *23*(3), 169-181. https://doi.org/10.1038/s41576-021-00434-9

Xenopoulos, P., Freeman, W. R., & Silva, C. (2022). Analyzing the differences between professional and amateur esports through win probability. Dalam *Proceedings of the ACM Web Conference 2022* (hlm. 3418-3427). ACM. https://doi.org/10.1145/3485447.3512277

Yangibaev, S., Mattiev, J., & Mokwena, S. (2025). DotA 2 match outcome prediction system using decision tree ensemble algorithms. *Big Data and Cognitive Computing*, *9*(12), 302. https://doi.org/10.3390/bdcc9120302

Zhao, C., Zhao, H., Ge, Y., Wu, R., & Shen, X. (2022). Winning Tracker: A new model for real-time winning prediction in MOBA games. Dalam *Proceedings of the ACM Web Conference 2022* (hlm. 3387-3395). ACM. https://doi.org/10.1145/3485447.3512274