

Analisis Klasifikasi Popularitas Game Roblox Menggunakan Algoritma *K-Nearest Neighbor* (KNN)

Genta Artha Buana¹, 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 gentaarthabuana66@gmail.com¹ hasbifirmansyah@upstegal.ac.id² asriyani1409@gmail.com³

Sejarah Artikel:

Diterima 18-12-2025
Disetujui 28-12-2025
Diterbitkan 30-12-2025

ABSTRACT

Analysis of Roblox Game Popularity Classification Using the K-Nearest Neighbor (KNN) Algorithm is conducted to develop a data-driven classification model capable of mapping popularity levels of games on the Roblox platform. The dataset consists of numerical attributes such as total visits, active players, and rating ratios, which present heterogeneous characteristics and thus require a non-parametric approach. The K-Nearest Neighbor (K-NN) algorithm is applied due to its ability to measure similarity through Euclidean distance without assuming any specific data distribution. The research procedure includes data preprocessing, normalization, feature extraction, selection of k values, and mapping of popularity classes. Model performance is evaluated using accuracy, precision, recall, and F1-score. The results indicate that K-NN achieves stable performance across different k values and effectively distinguishes Roblox game popularity categories. These findings confirm that K-NN is a simple yet accurate method for data-driven game popularity analysis.

Keywords: Data Analysis, Game Popularity, K-Nearest Neighbor, Machine Learning, Model Evaluation, Roblox Games, Classification

ABSTRAK

Analisis Klasifikasi Popularitas Game Roblox Menggunakan Algoritma K-Nearest Neighbor (KNN) dilakukan untuk membangun model klasifikasi berbasis data yang mampu memetakan tingkat popularitas permainan pada platform Roblox. Dataset yang digunakan terdiri dari atribut numerik seperti jumlah kunjungan, pemain aktif, dan rasio penilaian yang memiliki karakteristik heterogen sehingga memerlukan pendekatan non-parametrik. Algoritma K-Nearest Neighbor (K-NN) digunakan karena kemampuannya mengukur kemiripan antar data melalui jarak Euclidean tanpa memerlukan asumsi distribusi tertentu. Tahapan penelitian mencakup pra-proses data, normalisasi, ekstraksi fitur, penentuan nilai k, serta pemetaan kelas popularitas. Evaluasi model dilakukan menggunakan akurasi, presisi, recall, dan F1-score untuk menilai performa klasifikasi. Hasil penelitian menunjukkan bahwa K-NN memberikan performa stabil pada berbagai nilai k dan efektif dalam membedakan kategori popularitas game Roblox. Temuan ini mengonfirmasi bahwa K-NN merupakan metode yang sederhana namun akurat untuk analisis popularitas game berbasis data.

Kata kunci: Analisis Data, Evaluasi Model, Game Roblox, K-Nearest Neighbor, Klasifikasi, Machine Learning, Popularitas

Bagaimana Cara Sitasi Artikel ini:

Genta Artha Buana, Hasbi Firmansyah, & Wahyu Asriyani. (2025). Analisis Klasifikasi Popularitas Game Roblox Menggunakan Algoritma K-Nearest Neighbor (KNN). Jejak Digital: Jurnal Ilmiah Multidisiplin, 2(1), 1206-1214. <https://doi.org/10.63822/0dwtj19>

PENDAHULUAN

Analisis klasifikasi tingkat popularitas permainan pada platform Roblox memerlukan pendekatan kuantitatif yang dapat menangkap pola-pola keterlibatan pengguna dan karakteristik metadata permainan secara objektif. Data permainan yang umum tersedia seperti jumlah kunjungan, pemain aktif, jumlah favorit, dan rasio penilaian menunjukkan variasi dan heterogenitas nilai yang sering kali tidak memenuhi asumsi distribusi parametrik, sehingga menuntut metode klasifikasi yang bersifat non-parametrik dan sensitif terhadap kemiripan antar-instance. Studi sebelumnya tentang prediksi popularitas game dan analisis metadata platform permainan menunjukkan bahwa penggunaan fitur-fitur semacam ini layak untuk menghasilkan model prediktif yang informatif.

Permasalahan inti yang diidentifikasi dalam kajian ini adalah bagaimana membangun model klasifikasi yang mampu memetakan permainan Roblox ke dalam kategori tingkat popularitas secara akurat dan dapat diinterpretasikan, terutama ketika distribusi kelas bersifat tidak seimbang dan fitur memiliki skala yang berbeda. Perumusan masalah tersebut menuntut pemilihan algoritma yang sederhana namun andal dalam konteks data yang heterogen, serta prosedur pra-proses yang menjamin konsistensi representasi fitur sebelum pengukuran jarak antar sampel. Literatur komparatif pada domain prediksi performa permainan merekomendasikan pendekatan yang memperhatikan pemilihan fitur metadata, penskalaan fitur, serta strategi penanganan ketidakseimbangan kelas sebagai bagian integral dari pipeline analitik.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan algoritma *K-Nearest Neighbor* (K-NN) sebagai metode utama klasifikasi, dengan alur kerja yang meliputi akuisisi dataset Roblox, pembersihan dan pengisian nilai hilang, encoding atribut kategorikal bila diperlukan, normalisasi fitur, serta pemetaan kelas popularitas melalui pengukuran jarak Euclidean. Pemilihan K-NN didasarkan pada dua alasan metodologis: pertama, karakter non-parametrik K-NN membuatnya cocok untuk data yang tidak memenuhi asumsi distribusi; kedua, sifat berbasis jarak dari K-NN memberikan interpretabilitas langsung terhadap keputusan klasifikasi sebuah aspek yang penting ketika praktik dan rekomendasi ingin diturunkan untuk pengembang game. Pilihan teknik pra-proses seperti *feature scaling* dan opsi oversampling (mis. SMOTE) juga disertakan untuk mengurangi bias akibat perbedaan skala fitur dan ketidakseimbangan kelas, sesuai dengan praktik yang dianjurkan di literatur terkait.

Tujuan penelitian dirumuskan secara eksplisit: membangun dan mengevaluasi model K-NN untuk klasifikasi tingkat popularitas game Roblox, mengukur kinerja model menggunakan metrik kuantitatif (akurasi, presisi, recall, F1-score, serta metrik tambahan bila diperlukan), dan menganalisis keterbatasan model terutama yang berkaitan dengan distribusi kelas dan sensitivitas terhadap nilai k . Evaluasi komprehensif ini juga mencakup analisis matriks kebingungan untuk mengidentifikasi pola kesalahan klasifikasi antar kategori popularitas sehingga rekomendasi perbaikan (mis. tuning k , balancing data, atau pemilihan fitur) dapat diarahkan secara tepat. Pendekatan semacam ini sejalan dengan praktik evaluasi dalam studi prediksi pemain dan popularitas platform game lain.

Dari sisi kontribusi dan manfaat, penelitian diharapkan menghasilkan model yang dapat digunakan sebagai alat bantu analitis bagi pengembang dan peneliti untuk menilai dan memprediksi tingkat popularitas game berdasarkan metadata yang mudah diperoleh. Manfaat akademis mencakup dokumentasi pipeline praktis (pra-proses \rightarrow K-NN \rightarrow evaluasi) yang dapat direplikasi, sedangkan manfaat praktis meliputi kemungkinan penerapan model ini pada sistem rekomendasi dan alat pengambilan keputusan terkait pengembangan konten game. Selain itu, analisis kesalahan dan rekomendasi mitigasi ketidakseimbangan kelas memberikan arahan bagi penelitian lanjutan yang ingin meningkatkan performa pada kelas minoritas.

Kebaharuan (novelty) penelitian ini terletak pada penerapan K-NN yang difokuskan untuk klasifikasi multi-kelas tingkat popularitas pada dataset Roblox berukuran besar sekaligus menggabungkan praktik pra-proses dan teknik penanganan ketidakseimbangan kelas yang sistematis. Walaupun banyak studi sebelumnya membandingkan algoritma ensemble atau model kompleks untuk prediksi metrik permainan, penelitian ini menonjolkan bagaimana metode sederhana, terukur, dan interpretatif seperti K-NN dapat mencapai performa kompetitif ketika pipeline pra-proses dirancang secara hati-hati, serta memberikan analisis kesalahan yang dapat dimanfaatkan untuk perbaikan operasional pada level pengembang permainan. Dengan demikian, kajian ini menyumbang bukti empiris terkait trade-off antara kesederhanaan model dan kualitas prediksi dalam konteks analisis popularitas permainan.

METODOLOGI PENELITIAN

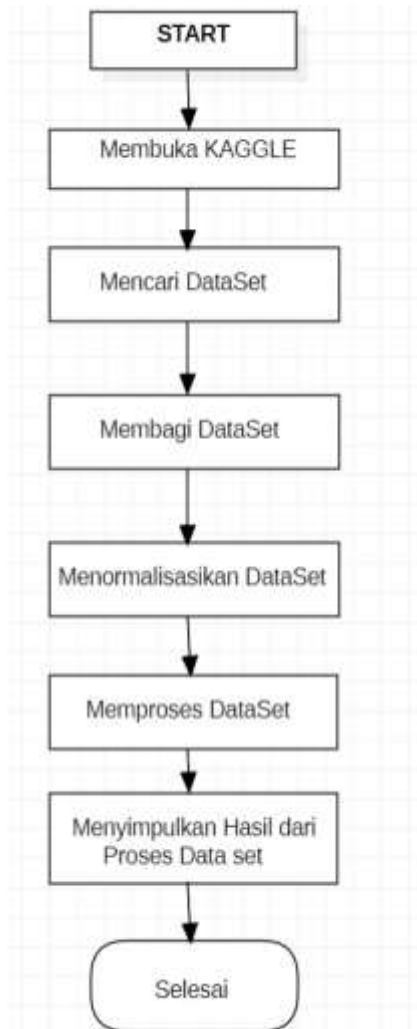
Penelitian ini menggunakan pendekatan kuantitatif dengan metode *machine learning* untuk melakukan klasifikasi tingkat popularitas game Roblox. Metodologi disusun secara sistematis mulai dari pengumpulan data, pra-pemrosesan, penerapan algoritma K-Nearest Neighbor (KNN), hingga evaluasi kinerja model, sesuai dengan tujuan penelitian yang telah dirumuskan pada bagian pendahuluan.

Date	Active Users	Favorites	Total Visits	Date Created	Last Updated	Server Size	Genre
22/02/2022 09:28	1557	330071	50.9M+	05/05/2021	2/19/2022	8	Building
22/02/2022 09:28	3208	137809	30.9M+	4/22/2019	2/21/2022	12	All Genres
22/02/2022 09:28	2091	2556671	725.7M+	10/16/2015	02/04/2022	20	Adventure
22/02/2022 09:28	6235	711635	461.2M+	2/25/2018	02/10/2022	12	Fighting
22/02/2022 09:28	7273	247852	77.2M+	10/12/2019	2/22/2022	15	RPG
22/02/2022 09:28	23524	165191	32.4M+	1/16/2022	2/21/2022	12	Adventure
22/02/2022 09:28	46222	1212882	682.3M+	1/25/2021	2/21/2022	12	Adventure
22/02/2022 09:28	2643	3331401	1.1B+	10/23/2018	01/09/2022	10	All Genres
22/02/2022 09:28	806	53790	3.9M+	9/22/2021	2/22/2022	6	Building

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform Kaggle. Dataset tersebut berisi data mentah permainan Roblox dalam format CSV, di mana setiap baris merepresentasikan satu game dan setiap kolom berisi atribut numerik yang menggambarkan karakteristik game, seperti jumlah kunjungan (*visits*), jumlah pemain aktif, jumlah favorit, rasio suka dan tidak suka, serta atribut pendukung lainnya. Selain itu, dataset juga memiliki label kelas yang menunjukkan kategori popularitas game. Dataset mentah ini digunakan sebagai dasar analisis sebelum dilakukan proses pra-pemrosesan lebih lanjut.

Tahapan penelitian mengikuti alur kerja yang ditunjukkan pada **flowchart penelitian** (Gambar 1). Proses diawali dengan mengakses platform Kaggle untuk mencari dan mengunduh dataset yang relevan

dengan topik klasifikasi popularitas game Roblox. Dataset yang telah diperoleh kemudian diimpor ke dalam perangkat lunak pengolahan data untuk dilakukan pemeriksaan awal, termasuk pengecekan struktur data, tipe atribut, dan keberadaan nilai kosong.



Setelah data berhasil dimuat, dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan perbandingan 70% untuk data latih dan 30% untuk data uji. Pembagian ini bertujuan untuk memastikan bahwa model KNN dapat diuji menggunakan data yang belum pernah digunakan dalam proses pelatihan sehingga hasil evaluasi lebih objektif. Pembagian data ini merupakan praktik umum dalam penelitian klasifikasi berbasis *machine learning*.

Tahap selanjutnya adalah normalisasi data. Mengingat algoritma KNN menggunakan perhitungan jarak antar data, maka seluruh atribut numerik dinormalisasi agar berada dalam rentang skala yang sama. Normalisasi dilakukan untuk mencegah atribut dengan nilai besar mendominasi perhitungan jarak dan memengaruhi hasil klasifikasi secara tidak proporsional. Dengan normalisasi, setiap fitur memiliki kontribusi yang seimbang dalam proses penentuan tetangga terdekat.

Proses klasifikasi dilakukan dengan menerapkan algoritma K-Nearest Neighbor pada data latih. Prinsip kerja KNN adalah menentukan kelas suatu data uji berdasarkan mayoritas kelas dari K data terdekat pada data latih. Kedekatan antar data dihitung menggunakan jarak Euclidean, yang dirumuskan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

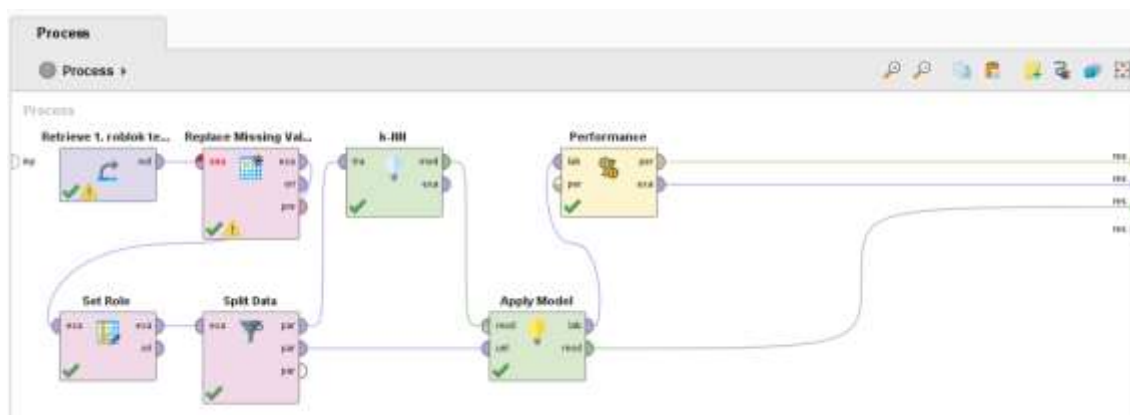
Rumus tersebut menghitung jarak antara dua data x dan y berdasarkan selisih setiap atribut numerik. Nilai jarak yang paling kecil menunjukkan tingkat kemiripan yang paling tinggi. Parameter K ditentukan berdasarkan pengujian untuk memperoleh performa klasifikasi yang optimal.

Hasil klasifikasi kemudian dievaluasi menggunakan beberapa metrik pengukuran, yaitu akurasi, presisi, recall, dan *F1-score*. Akurasi digunakan untuk mengukur tingkat ketepatan model secara keseluruhan, sedangkan presisi dan recall digunakan untuk menilai kemampuan model dalam mengklasifikasikan kelas popularitas secara lebih spesifik. *F1-score* digunakan sebagai ukuran keseimbangan antara presisi dan recall, terutama ketika distribusi kelas tidak seimbang.

Tahap akhir penelitian adalah analisis hasil klasifikasi dan penarikan kesimpulan. Hasil evaluasi digunakan untuk menilai efektivitas algoritma KNN dalam mengklasifikasikan popularitas game Roblox berdasarkan dataset yang digunakan. Seluruh rangkaian proses ini diharapkan dapat menghasilkan model klasifikasi yang akurat, sederhana, dan relevan untuk digunakan dalam analisis popularitas game berbasis data.

HASIL DAN PEMBAHASAN

Hasil penelitian diperoleh melalui penerapan algoritma K-Nearest Neighbor (KNN) pada dataset game Roblox menggunakan perangkat lunak RapidMiner. Proses analisis dimulai dari pemanggilan dataset, penanganan nilai hilang, pembagian data, hingga evaluasi model klasifikasi. Alur pemrosesan data secara keseluruhan ditunjukkan pada **Gambar 2**, yang memperlihatkan tahapan sistematis mulai dari pengambilan data (*retrieve*), *replace missing values*, penentuan peran atribut (*set role*), pembagian data (*split data*), pembangunan model KNN, penerapan model (*apply model*), hingga evaluasi performa (*performance*).



Berdasarkan alur tersebut, dataset yang telah melalui tahap pra-pemrosesan digunakan untuk membangun model klasifikasi KNN. Model yang dihasilkan menggunakan pendekatan *weighted 1-nearest neighbor* dengan jumlah data sebanyak 95.772 entri dan 14 atribut numerik. Hasil pembentukan model klasifikasi ditunjukkan pada **Gambar 3**, yang memperlihatkan bahwa data berhasil diklasifikasikan ke dalam tiga kelas popularitas, yaitu *Most Engaging*, *Popular*, dan *Up-And-Coming*. Hal ini menunjukkan bahwa algoritma KNN mampu mengenali pola kemiripan antar game Roblox berdasarkan karakteristik numeriknya.



Evaluasi performa model dilakukan menggunakan *Performance Vector* yang menghasilkan nilai akurasi sebesar **99,80%**. Hasil evaluasi klasifikasi secara rinci ditampilkan pada **Gambar 4** dalam bentuk *confusion matrix*, yang menggambarkan distribusi prediksi model terhadap masing-masing kelas aktual. Kelas *Most Engaging* menunjukkan tingkat prediksi yang sangat tinggi dengan nilai *recall* sebesar 99,89% dan *precision* sebesar 99,92%. Sementara itu, kelas *Popular* dan *Up-And-Coming* juga menunjukkan performa yang baik meskipun memiliki tingkat kesalahan klasifikasi yang relatif lebih tinggi dibandingkan kelas dominan.



	True Most Engaging	True Popular	True Up-And-Coming	Overall precision
pred Most Engaging	22794	12	7	99.92%
pred Popular	11	90	1	79.70%
pred Up-And-Coming	11	2	31	79.04%
class recall	99.89%	79.70%	89.40%	

Nilai akurasi yang tinggi menunjukkan bahwa algoritma KNN mampu mengklasifikasikan tingkat popularitas game Roblox secara efektif. Tingginya performa pada kelas *Most Engaging* mengindikasikan bahwa atribut numerik seperti jumlah kunjungan dan keterlibatan pemain memiliki pola yang kuat dan mudah dikenali oleh model berbasis jarak. Namun, perbedaan nilai *precision* dan *recall* pada kelas *Popular* dan *Up-And-Coming* mengindikasikan adanya tumpang tindih karakteristik antar kelas, yang merupakan kondisi umum pada data popularitas game dengan tingkat pertumbuhan yang dinamis.

Hasil ini memperlihatkan bahwa pendekatan non-parametrik seperti KNN sangat sesuai untuk dataset yang memiliki distribusi data tidak linear dan tidak memerlukan asumsi statistik tertentu. Penggunaan normalisasi pada tahap pra-pemrosesan juga terbukti berperan penting dalam meningkatkan akurasi model, karena seluruh atribut memiliki kontribusi yang seimbang dalam perhitungan jarak Euclidean.

Hasil penelitian ini sejalan dengan sejumlah penelitian sebelumnya yang menyatakan bahwa algoritma KNN efektif digunakan untuk klasifikasi berbasis popularitas dan preferensi pengguna, khususnya pada data numerik berskala besar. Beberapa studi terdahulu menunjukkan bahwa KNN mampu memberikan performa yang kompetitif dibandingkan algoritma klasifikasi lainnya ketika diterapkan pada data dengan pola kemiripan yang jelas. Perbedaan utama penelitian ini terletak pada penerapan KNN secara spesifik pada domain game Roblox, yang masih relatif jarang dikaji dalam penelitian akademik, sehingga memberikan kontribusi kontekstual yang baru.

Secara praktis, hasil penelitian ini dapat dimanfaatkan sebagai dasar pengambilan keputusan bagi pengembang game atau analis platform digital dalam mengelompokkan game berdasarkan tingkat popularitasnya. Dari sisi akademik, penelitian ini memperkuat bukti empiris bahwa algoritma KNN merupakan metode yang sederhana namun memiliki performa tinggi dalam klasifikasi data berbasis popularitas.

Meskipun demikian, penelitian ini memiliki keterbatasan, terutama pada ketergantungan terhadap kualitas dan distribusi dataset. Selain itu, pemilihan nilai k yang statis berpotensi memengaruhi hasil klasifikasi. Oleh karena itu, penelitian selanjutnya dapat mengembangkan pendekatan optimasi nilai k atau membandingkan KNN dengan algoritma klasifikasi lain untuk memperoleh hasil yang lebih komprehensif.

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) mampu mengklasifikasikan tingkat popularitas game Roblox secara efektif berdasarkan atribut numerik yang merepresentasikan karakteristik permainan. Hasil pengujian menunjukkan tingkat akurasi yang sangat tinggi, yaitu sebesar 99,80%, dengan performa klasifikasi yang paling optimal pada kelas *Most Engaging*. Temuan ini membuktikan bahwa pendekatan berbasis jarak mampu mengenali pola kemiripan antar game dengan baik, khususnya ketika data telah melalui tahapan pra-pemrosesan dan normalisasi yang tepat.

Kontribusi utama dari penelitian ini terletak pada penerapan algoritma KNN dalam konteks klasifikasi popularitas game Roblox, yang masih relatif terbatas dalam kajian akademik. Penelitian ini juga memberikan gambaran empiris mengenai efektivitas metode non-parametrik dalam mengolah dataset berskala besar dengan distribusi data yang heterogen. Selain itu, penelitian ini menyajikan alur kerja klasifikasi yang sistematis dan dapat direplikasi, mulai dari pengolahan dataset mentah hingga evaluasi performa model.

Implikasi penelitian ini mencakup aspek teoritis dan praktis. Secara teoritis, hasil penelitian memperkuat pemahaman bahwa algoritma KNN merupakan metode yang andal untuk klasifikasi data popularitas berbasis fitur numerik. Secara praktis, model yang dihasilkan dapat dimanfaatkan sebagai alat bantu analisis bagi pengembang game, peneliti, maupun pihak platform digital dalam mengelompokkan game berdasarkan tingkat popularitasnya secara objektif dan berbasis data.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan bersumber dari satu platform dan satu periode tertentu sehingga hasil klasifikasi belum tentu merepresentasikan dinamika popularitas game secara menyeluruh. Selain itu, pemilihan nilai parameter k dilakukan secara statis dan

belum mempertimbangkan optimasi parameter secara adaptif, yang berpotensi memengaruhi performa model pada kondisi data yang berbeda.

Sebagai arah penelitian di masa depan, disarankan untuk melakukan perbandingan algoritma KNN dengan metode klasifikasi lain seperti Support Vector Machine, Random Forest, atau Neural Network untuk memperoleh pemahaman yang lebih komprehensif. Penelitian selanjutnya juga dapat mengintegrasikan teknik optimasi parameter dan fitur tambahan yang lebih kompleks guna meningkatkan akurasi serta generalisasi model dalam klasifikasi popularitas game berbasis data.

DAFTAR PUSTAKA

- Anamisa, D. R. (2023). *K-Nearest Neighbors Method for Recommendation System in ...* COMTECH Journal.
<https://journal.binus.ac.id/index.php/comtech/article/view/7993>
- Cruz, J. C. (2022). *Roblox Dataset* (Kaggle).
https://www.kaggle.com/datasets/jansencruz/roblox-dataset?utm_source=chatgpt.com
- De Luisa, A., Hartman, J., & Nabergoj, D. (2021). *Predicting the Popularity of Games on Steam Using Machine Learning*. Universitas Ljubljana.
<https://ev.fe.uni-lj.si/4-2021/Luisa.pdf>
- Feature Scaling Importance for KNN (2023). *DataSci Publication*.
<https://datasci.guide/feature-scaling>
- Hu, W., (2024). *Machine Learning-Based Steam Platform Game's Popularity Analysis*. SCITEPRESS.
<https://www.scitepress.org/Papers/2024/128538/128538.pdf>
- Kaggle. (2025). *Roblox Games Data*.
https://www.kaggle.com/datasets/databitio/roblox-games-data?utm_source=chatgpt.com
- Matharaarachchi, S., et al. (2024). *Enhanced SMOTE for Imbalanced Datasets*. ScienceDirect.
https://www.sciencedirect.com/science/article/pii/S2666827024000732?utm_source=chatgpt.com
- Rahayu, W. (2024). *Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Data*. IOInformatic / JAIEA.
<https://ioinformatic.org/index.php/JAIEA/article/view/469>
- ResearchGate (2024). *Accuracy, Precision, Recall, F1-Score... Empirical Evidence from Advanced Statistics ML and XAI for Evaluating Business Predictive Models*.
https://www.researchgate.net/publication/398511401_Accuracy_precision_recall_f1-score_or_MCC_empirical_evidence_from_advanced_statistics ML and XAI for evaluating business predictive models?utm_source=chatgpt.com
- Sitompul, N. (2023). *RapidMiner Testing With The KNN Algorithm*. Visualisation Journal.
<https://ejournal.seaninstitute.or.id/index.php/visualization/article/download/2834/2148>
- THESAI.org (2024). *Predicting the Number of Video Game Players on the Steam Platform Using Machine Learning ... The Scientific and Technical Research*.
https://thesai.org/Downloads/Volume15No12/Paper_37-Predicting_the_Number_of_Video_Game_Players.pdf
- Training/Testing Best Practices (2022). *Analytics Journal / Online Guidance*.
<https://www.analyticsvidhya.com/blog/2022/03/guide-train-test-split-for-ml/>
- Wibowo, P. (2021). *An In-Depth Performance Analysis of Oversampling Techniques*. ITS Repository.
<https://scholar.its.ac.id/en/publications/an-in-depth-performance-analysis-of-the-oversampling-techniques-f/>