

PREDIKSI TINGKAT KETERLIBATAN PEMAIN GAME ONLINE BERDASARKAN INTENSITAS DAN POLA AKTIVITAS BERMAIN MENGGUNAKAN PENDEKATAN HYBRID MACHINE LEARNING

R.A. Callista Salsabila, Aisy Intan Herita, Neisya Az Zahra,
Winda Kurnia Sari, Ken Ditha Tania, Fathoni

Sistem Informasi, Universitas Sriwijaya

Jl. Raya Palembang - Prabumulih No. KM. 32, Indralaya Indah, Kabupaten Ogan Ilir, Sumatera Selatan
racallistasalsabila@gmail.com

ABSTRAK

Pertumbuhan industri *online games* menghasilkan volume data aktivitas pemain yang sangat besar, sehingga analisis tingkat keterlibatan pemain (*player engagement*) secara manual menjadi kurang efektif. Permasalahan ini mendorong perlunya pendekatan analitik berbasis data untuk mengidentifikasi pola perilaku pemain secara lebih akurat. Penelitian ini bertujuan untuk memprediksi tingkat *player engagement* berdasarkan intensitas dan pola aktivitas bermain menggunakan pendekatan *hybrid machine learning* yang mengkombinasikan algoritma *Logistic Regression* dan *Artificial Neural Network* (ANN). Data yang digunakan berasal dari dataset publik Kaggle yang terdiri dari 40.034 data pemain dengan variabel target *EngagementLevel* yang diklasifikasikan ke dalam tiga kategori, yaitu *Low*, *Medium*, dan *High*. Dataset melalui tahapan *data preprocessing* dan dibagi menjadi data latih dan data uji dengan rasio 80:20. Kinerja model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix*. Hasil penelitian menunjukkan bahwa model ANN mencapai akurasi sebesar 90,76% dengan rata-rata *F1-score* 0,91, lebih tinggi dibandingkan model *Logistic Regression* yang memperoleh akurasi 82% dengan *F1-score* 0,81. Analisis *feature importance* menunjukkan bahwa variabel *PlayTimeHours* dan *SessionsPerWeek* merupakan prediktor paling signifikan dalam menentukan tingkat keterlibatan pemain.

Kata kunci : *Player Engagement, Player Behaviour, Online Games, Hybrid Machine Learning, Logistic Regression, Neural Network*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, industri online games telah mengalami pertumbuhan pesat, menarik jutaan pemain di seluruh dunia dan menjadi salah satu sektor paling berpengaruh dalam industri hiburan digital[1] [2].

Pertumbuhan ini menghasilkan volume data aktivitas pemain yang sangat besar, yang mencerminkan berbagai bentuk interaksi pemain selama bermain. Data aktivitas tersebut tidak hanya menggambarkan pola perilaku pemain, tetapi juga dapat memberikan informasi penting mengenai tingkat keterlibatan pemain (*player engagement*) dalam suatu permainan.

Player engagement merupakan salah satu faktor penting dalam keberhasilan sebuah game karena mencerminkan tingkat keterlibatan, minat, serta interaksi pemain selama bermain. Tingkat keterlibatan pemain memiliki hubungan yang erat dengan retensi pemain, kepuasan pengguna, serta keberlanjutan penggunaan game dalam jangka panjang [3].

Semakin tinggi tingkat keterlibatan pemain, semakin besar kemungkinan pemain untuk terus memainkan game tersebut. Oleh karena itu, pemahaman terhadap faktor-faktor yang memengaruhi keterlibatan pemain menjadi aspek penting bagi pengembang game dalam merancang strategi yang mampu meningkatkan pengalaman bermain serta mempertahankan partisipasi pemain dalam jangka panjang.

Keterlibatan pemain (*player engagement*) sangat berkaitan dengan perilaku pemain (*player behaviour*), yang dapat diamati melalui berbagai aktivitas dalam game seperti durasi bermain, intensitas serta pola aktivitas bermain *game* [4].

Data perilaku tersebut dapat digunakan untuk mengidentifikasi pola aktivitas pemain dan memahami dinamika interaksi pemain dalam lingkungan permainan. Namun demikian, meningkatnya jumlah pemain serta kompleksitas data aktivitas permainan menyebabkan proses analisis perilaku pemain menjadi semakin sulit apabila dilakukan secara manual [5].

Kondisi ini menimbulkan permasalahan dalam mengidentifikasi pola keterlibatan pemain secara akurat serta memanfaatkan data aktivitas permainan secara optimal dalam proses pengambilan keputusan.

Volume data aktivitas pemain yang terus meningkat menghadirkan peluang dan tantangan untuk menganalisis keterlibatan pemain (*player engagement*) dalam *game online*. Data *game* skala besar mengandung informasi berharga tentang pola perilaku pemain, termasuk intensitas bermain, frekuensi permainan, dan dinamika interaksi dalam lingkungan game [3][8].

Pola-pola ini dapat memberikan wawasan tentang bagaimana pemain berinteraksi dengan *game* dan bagaimana keterlibatan mereka berkembang dari waktu ke waktu. Namun, mengekstraksi pola yang bermakna dari data yang kompleks dan berdimensi tinggi memerlukan metode analisis yang mampu

mengidentifikasi hubungan tersembunyi dalam data [6].

Dalam beberapa tahun terakhir, teknik *machine learning* telah banyak diterapkan di berbagai bidang, termasuk analisis perilaku pengguna, sistem rekomendasi, deteksi penipuan, dan analitik prediktif [7].

Dalam konteks permainan daring, pendekatan *machine learning* dapat dimanfaatkan untuk menganalisis data perilaku pemain serta memprediksi tingkat keterlibatan pemain berdasarkan pola aktivitas bermain yang tercatat dalam dataset [[2].

Namun demikian, setiap algoritma *machine learning* memiliki karakteristik dan keterbatasan masing-masing sehingga pemilihan metode yang tepat menjadi faktor penting dalam menghasilkan model prediksi yang optimal [3].

Penelitian Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk memprediksi tingkat keterlibatan pemain (*player engagement*) pada game online berdasarkan intensitas serta pola aktivitas bermain. Penelitian ini menggunakan pendekatan *hybrid machine learning* dengan mengkombinasikan algoritma *Logistic Regression* dan *Artificial Neural Network* untuk menganalisis data perilaku pemain serta mengklasifikasikan tingkat keterlibatan pemain ke dalam tiga kategori, yaitu *Low*, *Medium*, dan *High*. Pendekatan ini diharapkan mampu memanfaatkan keunggulan masing-masing algoritma, di mana *Logistic Regression* dapat memberikan interpretasi hubungan antara fitur dan variabel target, sedangkan *Artificial Neural Network* mampu memodelkan hubungan nonlinier yang lebih kompleks dalam data [4].

Penelitian Sebagai rencana penyelesaian masalah, penelitian ini memanfaatkan dataset perilaku pemain game online yang diperoleh dari platform Kaggle. Dataset tersebut kemudian melalui beberapa tahapan pengolahan data, meliputi *data preprocessing*, *exploratory data analysis*, *feature encoding*, *feature scaling*, serta pembagian dataset menjadi data latih dan data uji. Selanjutnya dilakukan proses pemodelan menggunakan algoritma *Logistic Regression* sebagai model *baseline* dan *Artificial Neural Network* sebagai model utama dalam penelitian ini. Kinerja model kemudian dievaluasi menggunakan beberapa metrik evaluasi seperti *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk mengukur kemampuan model dalam memprediksi tingkat keterlibatan pemain secara akurat [5].

2. TINJAUAN PUSTAKA

Tinjauan pustaka ini membahas konsep-konsep dan teori yang menjadi landasan penelitian, meliputi *online games*, *player engagement*, *player behaviour*, serta metode-metode *machine learning* yang digunakan, yaitu *Logistic Regression*, *Artificial Neural Network*, dan pendekatan *hybrid machine learning*. Selain itu, disajikan pula tinjauan terhadap penelitian-

penelitian terdahulu yang relevan sebagai dasar identifikasi celah penelitian.

2.1. Online Games dan Player Engagement

Online games adalah permainan digital berbasis internet yang memungkinkan interaksi antarpemain secara *real-time*. Industri ini berkembang pesat seiring meningkatnya akses internet dan ketersediaan perangkat seluler. *Player engagement* mencerminkan tingkat keterlibatan, minat, dan interaksi pemain saat bermain yang secara langsung memengaruhi retensi dan kepuasan pengguna menemukan bahwa persepsi positif terhadap *game* (VGPP) secara signifikan memprediksi keterlibatan pemain, terutama pada pemain ahli, menggunakan metode PLS-SEM dan analisis multikelompok [1].

2.2. Player Behaviour dan Pola Aktivitas Bermain

Player behaviour mencakup durasi bermain (*PlayTimeHours*), frekuensi sesi (*SessionsPerWeek*), dan pencapaian dalam *game* (*AchievementsUnlocked*). Data perilaku ini menjadi dasar dalam membangun model prediksi keterlibatan.

Rismayanti [2] membuktikan bahwa perilaku bermain dapat diprediksi menggunakan teknik *machine learning* dengan akurasi rata-rata 84,27% menggunakan algoritma *Gaussian Naive Bayes* dan 5-fold *cross-validation* pada dataset perilaku pemain.

Rasheed et al. [3] juga menunjukkan bahwa data *game telemetry* dapat digunakan untuk mengukur keterlibatan pemain secara *real-time* secara non-intrusif, mencapai akurasi 73% dan ROC-AUC 0,83.

2.3. Logistic Regression

Logistic Regression (LR) merupakan algoritma klasifikasi yang memodelkan probabilitas kelas berdasarkan kombinasi linier fitur-fitur input. Keunggulan utamanya adalah interpretabilitas tinggi dan efisiensi komputasi menerapkan regresi logistik dua arah pada dataset jutaan pertandingan kompetitif dan berhasil mengukur hubungan kausal antara hasil pertandingan dengan *player churn*, menemukan bahwa kemenangan beruntun secara signifikan menurunkan risiko pemain berhenti bermain [4].

2.4. Artificial Neural Network

Artificial Neural Network (ANN) adalah model komputasi berlapis yang mampu memodelkan hubungan non-linier kompleks dalam data melalui fungsi aktivasi di tiap lapisan tersembunyi. Arsitektur *Multi-Layer Perceptron* (MLP) merupakan varian yang paling umum digunakan dalam klasifikasi[5]. menerapkan *Feedforward Neural Network* (FNN) untuk klasifikasi keterampilan pemain secara *real-time* dengan DDA, mencapai akurasi 99,83% pada fase validasi. Hal ini juga membuktikan keunggulan ANN dibanding LR dalam hal diskriminasi dengan AUC 0,920 berbanding 0,833 dalam konteks prediksi klinis [9].

2.5. Hybrid Machine Learning

Pendekatan *hybrid machine learning* menggabungkan dua atau lebih model untuk memanfaatkan kekuatan masing-masing dan mengatasi keterbatasan model tunggal. Melalui tinjauan literatur sistematis menyimpulkan bahwa metode hibrida secara konsisten mengungguli algoritma tunggal dalam hal kecepatan dan akurasi, terutama yang berbasis kombinasi teknik optimasi dan *machine learning* [6].

Penggabungan LR yang unggul dalam interpretabilitas dengan ANN yang kuat dalam menangkap pola non-linier merupakan salah satu bentuk pendekatan hibrida yang relevan untuk prediksi *player engagement*.

2.6. Penelitian Terdahulu

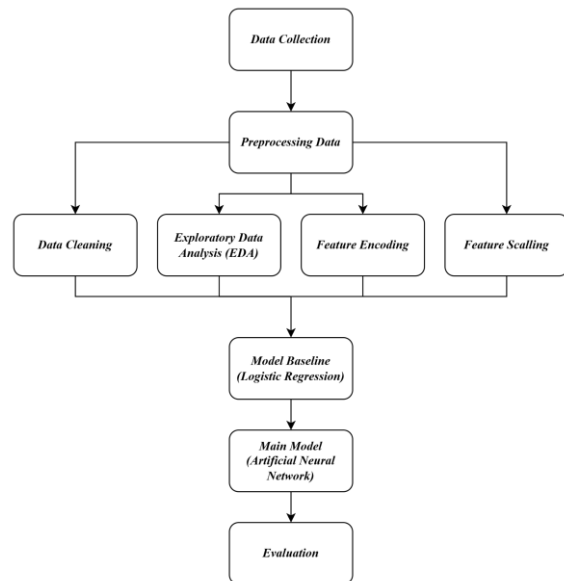
Beberapa penelitian terdahulu yang relevan adalah sebagai berikut mengembangkan sistem *Engagement-Oriented Dynamic Difficulty Adjustment* (EDDA) yang memantau *churn* secara *real-time* pada tujuh genre *game* utama dan terbukti lebih efektif dari DDA konvensional dalam mempertahankan keterlibatan pemain [7].

Selain itu juga menerapkan CNN untuk memprediksi dampak *online game* terhadap IPK mahasiswa dengan akurasi 96,69%, membuktikan efektivitas *deep learning* dalam menganalisis perilaku pengguna digital [10].

Berdasarkan tinjauan tersebut, terdapat celah penelitian yang belum banyak dieksplorasi, yaitu pendekatan *hybrid* yang secara eksplisit menggabungkan LR dan ANN untuk prediksi *player engagement* berbasis data aktivitas bermain ke dalam tiga kategori (*Low*, *Medium*, *High*). Penelitian ini berupaya mengisi celah tersebut.

3. METODE PENELITIAN

Metode penelitian ini bertujuan untuk membangun model klasifikasi yang dapat memprediksi tingkat keterlibatan pemain dalam permainan online. Alur tahapan penelitian yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1. Penelitian ini diawali dengan tahap data *preprocessing* yang meliputi *data cleaning*, *exploratory data analysis (EDA)*, *feature encoding*, dan *feature scaling* untuk mempersiapkan data sebelum pemodelan. Selanjutnya dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk proses pelatihan dan pengujian model. Tahap pemodelan menggunakan *Logistic Regression* sebagai model *baseline* serta *Artificial Neural Network* sebagai model utama dalam penelitian ini. Kinerja model kemudian dievaluasi menggunakan metrik evaluasi seperti *accuracy*, *loss*, dan *confusion matrix*.



Gambar 1. Alur diagram metode penelitian

3.1. Data Collection

Data yang digunakan dalam penelitian ini diperoleh langsung dari data publik pada platform Kaggle, dengan judul dataset “*Predict Online Gaming Behavior Dataset*” yang berisi 40.034 informasi mengenai perilaku pemain dalam bermain game online [10].

Fitur-fitur yang ada antara lain *PlayerID*, *Age*, *Gender*, *Location*, *GameGenre*, *PlayTimeHours*, *InGamePurchases*, *GameDifficulty*, *SessionsPerWeek*, *AvgSessionDurationMinutes*, *PlayerLevel*, dan *AchievementsUnlocked*. Atribut-atribut ini merepresentasikan perilaku pemain dan mengidentifikasi faktor-faktor yang mempengaruhi partisipasi dalam ekosistem bermain game [11].

Sebagai variabel target, *EngagementLevel* didefinisikan dengan tiga kategori klasifikasi yang ada, yaitu *Low*, *Medium*, dan *High*. Tujuan akhir dari pemodelan ini adalah menghasilkan prediksi akurat tentang tingkat keterlibatan pemain yang menjadi kecanduan, dengan mempertimbangkan karakteristik individu dan intensitas interaksi dalam game yang tercatat dalam dataset [12].

3.2. Data Cleaning

Tahap awal pengolahan setelah pengumpulan data ialah melakukan pembersihan data guna memastikan kualitas informasi yang kemudian diproses oleh model *machine learning* [13].

Prosesnya melibatkan identifikasi nilai yang hilang (*missing values*) dengan fungsi agregasi statistik. Dilakukan reduksi dimensi manual dengan menghapus fitur-fitur yang tidak memiliki relevansi terhadap variabel target, seperti *PlayerID* dan *Location*. *Data Cleaning* bertujuan menghilangkan *noise* sehingga model lebih fokus pada atribut signifikan yang memengaruhi data [14].

3.3. Exploratory Data Analysis (EDA)

Analisis deskriptif yang dilakukan untuk memahami distribusi frekuensi dan karakteristik statistik dari dataset [15].

Pendekatan univariat digunakan untuk meninjau sebaran variabel numerik seperti usia, durasi bermain, serta frekuensi sesi mingguan melalui teknik visualisasi histogram. Selanjutnya, analisis bivariat diimplementasikan menggunakan diagram kotak (*boxplot*) untuk mengobservasi korelasi antara variabel independen dengan tingkat keterlibatan pemain (*Engagement Level*). Tahapan ini krusial untuk mendeteksi pencilan (*outliers*) serta memahami tren perilaku pemain berdasarkan genre permainan dan tingkat kesulitan yang dipilih.

3.4. Feature Encoding

Meninjau algoritma suatu machine learning membutuhkan input dalam format numerik, maka dilakukan transformasi pada variabel kategorikal [16]. Untuk variabel target '*EngagementLevel*', digunakan teknik *Encoding* guna mengonversi tingkatan kualitatif *Low*, *Medium*, dan *High* menjadi representasi ordinal numerik. Dalam kategorikal fitur nominal seperti '*Gender*', '*GameGenre*', dan '*GameDifficulty*' diolah menggunakan *One-Hot Encoding* [17].

Pendekatan ini memastikan agar model tidak mengasumsikan adanya urutan prioritas yang salah pada variabel-variabel kategori tersebut.

3.5. Feature Scaling

Perbedaan pada rentang nilai fitur, seperti usia pemain dengan durasi bermain dalam menit, dapat memengaruhi konvergensi algoritma. Proses normalisasi data dilakukan menggunakan metode *Standardization* (*StandardScaler*) yang mentransformasikan data sehingga memiliki rata-rata nol dan standar deviasi satu [19].

Pengukuran dalam bentuk skala ini sangat penting, terutama bagi algoritma *neural network* yang berbasis gradien seperti jaringan syaraf tiruan, agar bobot setiap fitur dapat diproses secara seimbang [20].

3.6. Data Splitting

Proses evaluasi performa model secara objektif dilakukan dengan pemisahan dataset menjadi dua bagian utama, yaitu data latih (*training set*) dan data uji (*testing set*) [18].

Rasio pembagian yang ditetapkan adalah 80:20, di mana 80% yang berarti 32.027 data digunakan untuk proses pembelajaran model dan 20% yang sisanya 8.007 data dialokasikan menjadi data independen untuk pengujian. Pengaturan parameter *random state* dilakukan guna menjamin konsistensi reproduisibilitas pada setiap iterasi percobaan yang dilakukan.

3.7. Feature Importance

Analisis *feature importance* dilakukan untuk mengidentifikasi tingkat pengaruh masing-masing fitur terhadap variabel target, yaitu *EngagementLevel*. Dalam penelitian ini, nilai kepentingan fitur diperoleh dari koefisien yang dihasilkan oleh model *Logistic Regression* yang telah dilatih menggunakan data latih. Besarnya nilai koefisien menunjukkan kontribusi relatif setiap fitur dalam memengaruhi prediksi tingkat keterlibatan pemain dalam bermain game, di mana fitur dengan nilai koefisien lebih besar dianggap memiliki pengaruh yang lebih signifikan terhadap hasil klasifikasi. Analisis ini bertujuan untuk memberikan pemahaman mengenai faktor-faktor yang paling berperan dalam menentukan tingkat keterlibatan pemain, sekaligus mendukung interpretasi model *baseline* sebelum dilakukan pemodelan menggunakan metode yang lebih kompleks [14][17].

3.8. Baseline Model

Sebagai titik acuan perbandingan awal (*baseline*), penelitian ini menggunakan algoritma Regresi Logistik. Model ini dipilih karena kemampuannya dalam menangani klasifikasi multikelas dengan interpretabilitas yang tinggi [20]. Melalui model *baseline* ini, dilakukan analisis koefisien untuk mengidentifikasi tingkat kepentingan fitur (*feature importance*). Hasil dari Regresi Logistik memberikan standar performa awal sebelum diuji lebih lanjut menggunakan pendekatan pembelajaran mendalam yang lebih kompleks.

3.9. Main Model

Model utama yang dikembangkan dalam penelitian ini adalah Jaringan Syaraf Tiruan (*Artificial Neural Network*) dengan arsitektur *Multi-Layer Perceptron* (MLP). Struktur model terdiri dari satu lapisan input, dua hidden layer dengan masing-masing 64 dan 32 neuron menggunakan fungsi aktivasi ReLU, serta lapisan output dengan fungsi aktivasi Softmax untuk klasifikasi tiga kelas. Untuk mencegah terjadinya *overfitting*, diterapkan lapisan *Dropout* dengan laju 0.5 [24].

Proses optimasi dilakukan menggunakan algoritma Adam dengan fungsi kerugian *Sparse Categorical Crossentropy*. Selain itu, mekanisme *Early Stopping* diintegrasikan untuk menghentikan pelatihan saat tidak ada perbaikan pada kerugian validasi, guna memastikan model memiliki generalisasi yang optimal.

3.10. Evaluation

Kinerja model diukur menggunakan berbagai metrik evaluasi standar pembelajaran mesin. Akurasi dan kerugian (*loss*) dipantau selama proses pelatihan dan validasi untuk memastikan stabilitas model [26].

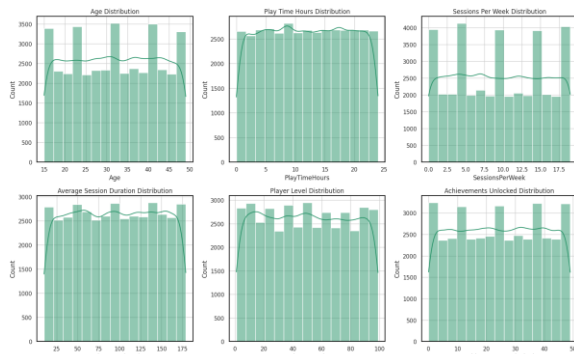
Pengujian data uji dilakukan untuk mendapatkan nilai akurasi final. Selain itu, *Confusion Matrix* digunakan untuk menganalisis distribusi prediksi benar dan salah pada setiap kelas,

memberikan gambaran mendalam mengenai presisi dan sensitivitas model dalam memprediksi tingkat keterlibatan pemain secara akurat.

4. HASIL DAN PEMBAHASAN

Penelitian ini menyajikan hasil dari serangkaian eksperimen yang telah dilakukan untuk memprediksi tingkat keterlibatan pemain (*player engagement*) menggunakan dataset perilaku permainan daring. Pembahasan mencakup analisis eksploratif data, evaluasi model *Logistic Regression* sebagai *baseline*, evaluasi model *Artificial Neural Network* (ANN), serta analisis perbandingan performa dan kepentingan fitur.

4.1. Exploratory Data Analysis (EDA)

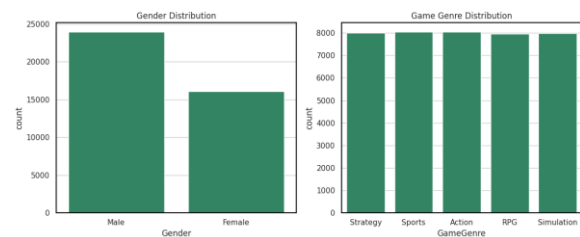


Gambar 2. Diagram distribusi numerik

Tahap awal penelitian dilakukan melalui proses pembersihan data untuk memastikan kualitas dataset sebelum dilakukan pemodelan. Pada tahap ini, atribut yang tidak memiliki relevansi prediktif terhadap variabel target, yaitu *PlayerID* dan *Location*, dihapus dari dataset karena hanya berfungsi sebagai identifikasi data dan tidak merepresentasikan perilaku pemain. Setelah proses pembersihan data, analisis deskriptif dilakukan untuk memahami karakteristik dasar dari variabel yang tersedia dalam dataset. Hasil visualisasi distribusi data pada Gambar 2 menunjukkan bahwa variabel numerik seperti *Age*, *PlayTimeHours*, dan *PlayerLevel* memiliki persebaran nilai yang cukup luas, yang mencerminkan heterogenitas profil pemain dalam dataset.

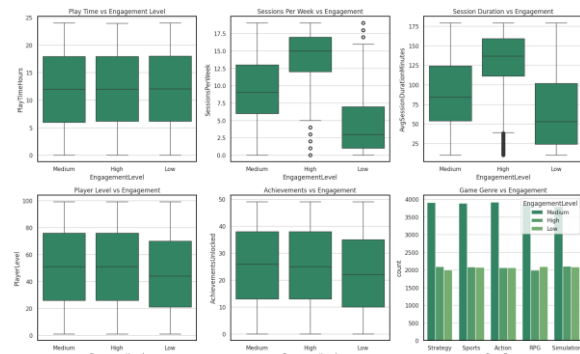
Distribusi nilai pada variabel usia menunjukkan bahwa sebagian besar pemain berada pada rentang usia dewasa muda, yang merupakan kelompok usia yang secara umum memiliki tingkat partisipasi tinggi dalam aktivitas permainan daring. Variabel *PlayTimeHours* dan *SessionsPerWeek* menunjukkan variasi yang signifikan antar pemain, yang mengindikasikan adanya perbedaan intensitas aktivitas bermain dalam populasi dataset. Sementara itu, variabel *PlayerLevel* dan *AchievementsUnlocked* menggambarkan progres pemain dalam permainan, yang dapat menjadi indikator keterlibatan jangka panjang terhadap permainan yang dimainkan. Variasi yang cukup besar pada variabel-variabel ini memberikan indikasi awal bahwa faktor perilaku bermain berpotensi menjadi

prediktor penting dalam menentukan tingkat keterikatan pemain.



Gambar 3. Diagram distribusi kategorikal

Terlihat pada Gambar 3 visualisasi lebih lanjut mengenai distribusi variabel kategorikal menunjukkan komposisi pemain berdasarkan *Gender* dan *GameGenre*. Hasil analisis menunjukkan bahwa distribusi pemain berdasarkan *gender* relatif seimbang, sehingga dataset tidak menunjukkan bias yang signifikan terhadap kelompok tertentu. Selain itu, preferensi terhadap genre permainan menunjukkan bahwa beberapa kategori permainan memiliki jumlah pemain yang lebih dominan dibandingkan genre lainnya. Variasi preferensi genre ini menunjukkan bahwa karakteristik permainan yang dipilih pemain juga dapat memengaruhi pola keterlibatan pemain dalam aktivitas bermain.

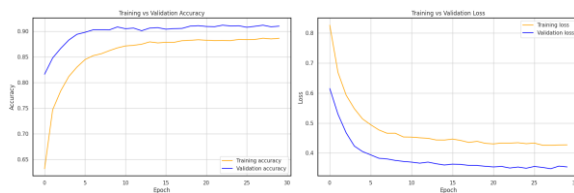


Gambar 4. Hubungan Fitur vs Engagement

Pada Gambar 4 terlihat analisis hubungan antara variabel independen dan tingkat keterikatan pemain dilakukan melalui pendekatan bivariat menggunakan visualisasi *boxplot*. Hasil analisis menunjukkan adanya hubungan positif antara durasi waktu bermain (*PlayTimeHours*), frekuensi sesi bermain (*SessionsPerWeek*), serta jumlah pencapaian (*AchievementsUnlocked*) dengan tingkat keterikatan pemain (*EngagementLevel*). Pemain yang berada dalam kategori *High Engagement* cenderung memiliki nilai rata-rata yang lebih tinggi pada variabel-variabel tersebut dibandingkan pemain pada kategori *Low* maupun *Medium*. Temuan ini menunjukkan bahwa intensitas aktivitas bermain merupakan indikator utama dalam menggambarkan tingkat keterlibatan pemain dalam ekosistem permainan daring.

4.2. Evaluasi Model ANN

Model Artificial Neural Network yang digunakan dalam penelitian ini dibangun menggunakan arsitektur Sequential dengan dua lapisan tersembunyi (*hidden layer*) dan fungsi aktivasi ReLU. Untuk meningkatkan kemampuan generalisasi model dan mencegah terjadinya *overfitting*, diterapkan mekanisme *Dropout* dengan nilai 0,5 pada lapisan jaringan. Selain itu, proses pelatihan model juga dilengkapi dengan mekanisme *Early Stopping* yang secara otomatis menghentikan proses pelatihan ketika performa pada data validasi tidak lagi mengalami peningkatan. Pendekatan ini bertujuan untuk memastikan bahwa model yang dihasilkan memiliki performa optimal tanpa mengalami penurunan kemampuan generalisasi terhadap data baru.



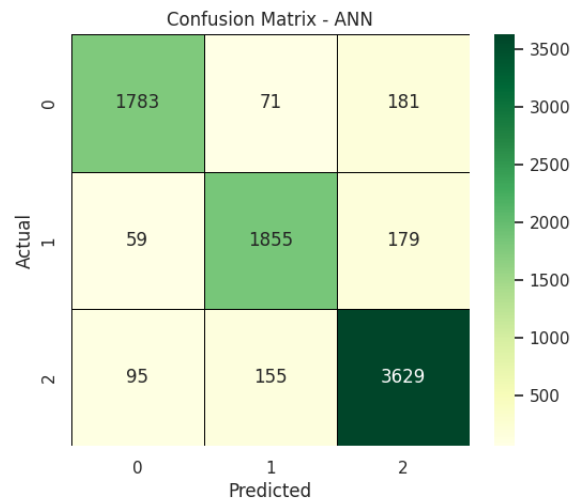
Gambar 5. Kurva pelatihan

Proses pelatihan model dipantau menggunakan kurva akurasi dan kurva kerugian (*loss curve*) yang ditampilkan pada Gambar 5. Hasil visualisasi menunjukkan bahwa nilai *training accuracy* dan *validation accuracy* mengalami peningkatan secara konsisten selama proses pelatihan. Pada saat yang sama, nilai kerugian pada data pelatihan dan validasi menunjukkan tren penurunan yang stabil hingga mencapai titik konvergensi. Pola ini menunjukkan bahwa model berhasil mempelajari pola hubungan antar fitur tanpa mengalami *overfitting* yang signifikan.

Evaluasi performa model dilakukan menggunakan data uji (*test set*) yang sebelumnya tidak digunakan selama proses pelatihan. Berdasarkan hasil pengujian, model *Artificial Neural Network* menghasilkan nilai *loss* sebesar 0,3629 dan akurasi sebesar 0,9076 yang jika dipersentasekan menjadi 90,76%. Nilai akurasi ini menunjukkan bahwa model mampu mengklasifikasikan tingkat keterikatan pemain dengan tingkat ketepatan yang tinggi. Hasil ini menunjukkan bahwa pendekatan berbasis jaringan saraf tiruan cukup efektif dalam menangkap hubungan kompleks antar variabel dalam dataset perilaku pemain game online.

Analisis lebih lanjut terhadap performa model dilakukan menggunakan *Confusion Matrix* yang ditampilkan pada Gambar 6. Matriks ini memberikan gambaran mengenai distribusi prediksi benar dan salah pada setiap kelas keterikatan pemain. Secara umum, model mampu mengklasifikasikan sebagian besar data dengan benar pada setiap kategori *EngagementLevel*. Kesalahan klasifikasi yang terjadi relatif kecil dan sebagian besar terjadi antara kategori yang memiliki

karakteristik perilaku yang serupa, seperti kategori *Medium* dan *High*.



Gambar 6. Confusion Matrix

4.3. Analisis Perbandingan Model dan Kepentingan Fitur

Sebagai pembandingan terhadap model berbasis deep learning, penelitian ini juga menerapkan algoritma Regresi Logistik sebagai model baseline. Model ini dipilih karena memiliki interpretabilitas yang tinggi dan sering digunakan sebagai metode klasifikasi awal dalam penelitian machine learning. Berdasarkan hasil evaluasi pada data uji, model Regresi Logistik menghasilkan tingkat akurasi sebesar 0,82. Meskipun nilai akurasi yang dihasilkan cukup baik, performa model ini masih berada di bawah model Artificial Neural Network.

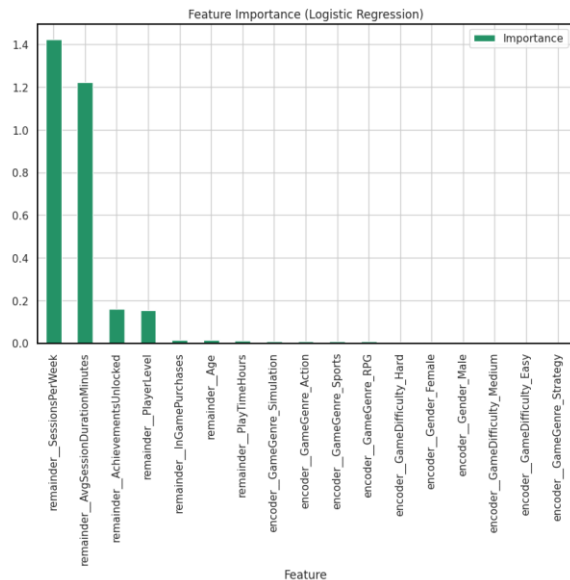
Tabel 1. Perbandingan Performa Model

Model	Accuracy	Precision Avg	Recall Avg	F1-score Avg
Logistic Regression	0.82	0.83	0.80	0.81
Neural Network	0.91	0.91	0.90	0.91

Perbedaan performa ini dapat dijelaskan melalui karakteristik masing-masing algoritma. Regresi Logistik merupakan model linear yang bekerja dengan membangun batas pemisah linear antar kelas dalam ruang fitur. Sebaliknya, Artificial Neural Network mampu memodelkan hubungan non-linear yang lebih kompleks melalui kombinasi beberapa lapisan neuron. Oleh karena itu, model ANN lebih mampu menangkap pola perilaku pemain yang kompleks dan dinamis dibandingkan model Regresi Logistik.

Selain sebagai model pembandingan, Regresi Logistik juga memberikan keuntungan dalam analisis kepentingan fitur (*feature importance*). Berdasarkan Gambar 7 yang merupakan analisis koefisien model ditemukan bahwa variabel *PlayTimeHours* dan *SessionsPerWeek* merupakan fitur yang memiliki

pengaruh paling signifikan terhadap tingkat keterikatan pemain. Kedua variabel tersebut merepresentasikan intensitas aktivitas bermain yang secara langsung mencerminkan tingkat partisipasi pemain dalam permainan. Temuan ini menunjukkan bahwa perilaku bermain aktif merupakan faktor utama yang menentukan tingkat keterlibatan pemain.



Gambar 7. Grafik visual feature importance

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa pendekatan hybrid machine learning yang mengkombinasikan *Logistic Regression* dan *Artificial Neural Network* (ANN) mampu digunakan secara efektif untuk memprediksi tingkat keterlibatan pemain (*player engagement*) pada game online berdasarkan intensitas dan pola aktivitas bermain. Hasil eksperimen menunjukkan bahwa model ANN memiliki performa terbaik dengan akurasi sebesar 90,76%, lebih tinggi dibandingkan model *Logistic Regression* sebagai baseline yang memperoleh akurasi 82%, sehingga menunjukkan bahwa model berbasis jaringan saraf lebih mampu menangkap pola hubungan nonlinier pada data perilaku pemain. Selain itu, analisis kepentingan fitur menunjukkan bahwa *PlayTimeHours* dan *SessionsPerWeek* merupakan faktor paling berpengaruh dalam menentukan tingkat keterlibatan pemain. Dengan demikian, penelitian ini membuktikan bahwa data aktivitas bermain dapat dimanfaatkan untuk memahami pola keterlibatan pemain secara lebih akurat.

Untuk penelitian selanjutnya, disarankan agar menggunakan dataset yang lebih besar dan beragam, serta mengeksplorasi algoritma lain seperti *Random Forest*, *Gradient Boosting*, atau *Deep Learning* yang lebih kompleks, serta menambahkan fitur perilaku pemain lainnya seperti interaksi sosial atau aktivitas dalam game secara real-time agar model prediksi

player engagement dapat menghasilkan performa yang lebih optimal.

DAFTAR PUSTAKA

- [1] H. Yildiz Durak, A. Haktanir, and M. Saritepeci, "Examining the Predictors of Video Game Addiction According to Expertise Levels of the Players: The Role of Time Spent on Video Gaming, Engagement, Positive Gaming Perception, Social Support and Relational Health Indices," *Int. J. Ment. Health Addict.*, vol. 22, no. 6, pp. 3683–3708, 2024, doi: 10.1007/s11469-023-01073-3.
- [2] N. Rismayanti, "Predicting Online Gaming Behaviour Using Machine Learning Techniques," *Indones. J. Data Sci.*, vol. 5, no. 2, pp. 133–143, 2024, doi: 10.56705/ijodas.v5i2.166.
- [3] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Real-Time Player Engagement Measurement Using Nonintrusive Game Telemetry," *IEEE Open J. Instrum. Meas.*, vol. 4, no. March, pp. 1–16, 2025, doi: 10.1109/OJIM.2025.3555326.
- [4] H. Kang, C. Suh, and H. K. Kim, "Match experiences affect interest: Impacts of matchmaking and performance on churn in a competitive game," *Heliyon*, vol. 10, no. 3, p. e24891, 2024, doi: 10.1016/j.heliyon.2024.e24891.
- [5] E. A. Romero-Mendez, P. C. Santana-Mancilla, M. Garcia-Ruiz, O. A. Montesinos-López, and L. E. Anido-Rifón, "The Use of Deep Learning to Improve Player Engagement in a Video Game through a Dynamic Difficulty Adjustment Based on Skills Classification," *Appl. Sci.*, vol. 13, no. 14, 2023, doi: 10.3390/app13148249.
- [6] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, *Hybrid approaches to optimization and machine learning methods: a systematic literature review*, vol. 113, no. 7. Springer US, 2024. doi: 10.1007/s10994-023-06467-x.
- [7] Q. Mi and T. Gao, "Engagement-Oriented Dynamic Difficulty Adjustment," *Appl. Sci.*, vol. 15, no. 10, 2025, doi: 10.3390/app15105610.
- [8] M. Gao, J. Wang, and J. Yang, "Research into the Relationship between Personality and Behavior in Video Games, Based on Mining Association Rules," *Mathematics*, vol. 11, no. 3, 2023, doi: 10.3390/math11030772.
- [9] S. J. Lin, C. C. Liu, D. M. T. Tsai, Y. H. Shih, C. L. Lin, and Y. C. Hsu, "Prediction Models Using Decision Tree and Logistic Regression Method for Predicting Hospital Revisits in Peritoneal Dialysis Patients," *Diagnostics*, vol. 14, no. 6, pp. 1–13, 2024, doi: 10.3390/diagnostics14060620.
- [10] M. Alshraideh, A. A. J. Abu-Zayed, M. Leiner, and I. M. Aldajani, "Beyond the Scoreboard: A Machine Learning Investigation of Online Games' Influence on Jordanian University Students' Grades," *Appl. Comput. Intell. Soft*

- Comput.*, vol. 2024, 2024, doi: 10.1155/2024/1337725.
- [11] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Descriptor: Multimodal Dataset for Player Engagement Analysis in Video Games (MultiPENG)," *IEEE Data Descr.*, vol. 2, no. February, pp. 17–25, 2025, doi: 10.1109/ieeedata.2025.3553097.
- [12] M. Kaur and M. Saini, *Role of Artificial Intelligence in the crime prediction and pattern analysis studies published over the last decade: a scientometric analysis*, vol. 57, no. 8. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10823-1.
- [13] A. Szczur, J. G. Bazan, U. Bentkowska, P. Kruczek, and S. Bazan-Socha, "Identifying and Predicting Changes in Behavioral Patterns for Temporal Data in Treatment of Neonatal Respiratory Failure," *Appl. Sci.*, vol. 15, no. 22, pp. 1–21, 2025, doi: 10.3390/app152212133.
- [14] A. Ostovar, D. D. Davari, and M. Dzikuc, "Determinants of Design with Multilayer Perceptron Neural Networks: A Comparison with Logistic Regression," *Sustain.*, vol. 17, no. 6, 2025, doi: 10.3390/su17062611.
- [15] Q. Guo, Z. He, and Z. Wang, "Monthly climate prediction using deep convolutional neural network and long short-term memory," *Sci. Rep.*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-68906-6.
- [16] W. H. Abuwatfa, N. AlSawafthah, N. Darwish, W. G. Pitt, and G. A. Husseini, "A Review on Membrane Fouling Prediction Using Artificial Neural Networks (ANNs)," *Membranes (Basel)*, vol. 13, no. 7, pp. 1–29, 2023, doi: 10.3390/membranes13070685.
- [17] D. Dey *et al.*, "The proper application of logistic regression model in complex survey data: a systematic review," *BMC Med. Res. Methodol.*, vol. 25, no. 1, 2025, doi: 10.1186/s12874-024-02454-5.
- [18] S. Cheng *et al.*, "Cartography and Neural Networks: A Scientometric Analysis Based on CiteSpace," *ISPRS Int. J. Geo-Information*, vol. 13, no. 6, 2024, doi: 10.3390/ijgi13060178.
- [19] A. Kashem, R. Karim, P. Das, S. D. Datta, and M. Alharthai, "Compressive strength prediction of sustainable concrete incorporating rice husk ash (RHA) using hybrid machine learning algorithms and parametric analyses," *Case Stud. Constr. Mater.*, vol. 20, no. March, p. e03030, 2024, doi: 10.1016/j.cscm.2024.e03030.
- [20] H. K. Maragheh, F. S. Gharehchopogh, K. Majidzadeh, and A. B. Sangar, "A Hybrid Model Based on Convolutional Neural Network and Long Short-Term Memory for Multi-label Text Classification," *Neural Process. Lett.*, vol. 56, no. 2, pp. 1–31, 2024, doi: 10.1007/s11063-024-11500-8.
- [21] C. Li and Z. Cao, "Deep learning-based AI model for predicting academic success and engagement among physical higher education students," *Sci. Rep.*, vol. 15, no. 1, pp. 1–19, 2025, doi: 10.1038/s41598-025-29000-7.