

# Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian pada *Game Genshin Impact* Menggunakan Klasifikasi *Naive Bayes*

Hadi wiranata<sup>1</sup>, Sucipto<sup>2</sup>, Arie Nugroho<sup>3</sup>

<sup>1,2,3</sup> Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

<sup>1</sup>hadiwiranata311@gmail.com, <sup>2</sup>sucipto@unpkediri.ac.id, <sup>3</sup>arienugroho@unpkediri.ac.id

**Abstrak**-Perkembangan industri game online telah menciptakan model bisnis berbasis mikrotransaksi, dimana pemain dapat membeli item virtual menggunakan uang nyata. Salah satu game yang mengadopsi model ini adalah *Genshin Impact*, yang menarik perhatian banyak pemain melalui sistem gacha dan konten karakter yang variatif. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi keputusan pembelian item virtual dalam game *Genshin Impact* menggunakan metode klasifikasi *Naive Bayes*. Data dikumpulkan melalui kuesioner online dengan total 314 responden. Variabel yang dianalisis meliputi lama bermain, kepuasan terhadap sistem gacha, kostum, alur cerita, pengaruh sosial, serta pengalaman pembelian sebelumnya. Model dikembangkan menggunakan algoritma *Gaussian Naive Bayes* dan divalidasi menggunakan metode *Stratified K-Fold Cross-Validation*. Hasil penelitian menunjukkan bahwa model mampu mengklasifikasikan keputusan pembelian dengan akurasi sebesar 80,57%. Meskipun model menunjukkan kinerja tinggi dalam mengklasifikasikan pemain yang membeli item, akurasi terhadap pemain yang tidak membeli masih perlu ditingkatkan. Penelitian ini dapat menjadi acuan dalam memahami perilaku pembelian pemain serta pengembangan strategi pemasaran digital dalam industri game online.

**Kata Kunci:** *Genshin Impact*, Keputusan Pembelian, Game Online, *Naive Bayes*, Data Mining

**Abstract**-The growth of the online gaming industry has led to the emergence of microtransaction-based business models, where players can purchase virtual items using real money. One of the games implementing this model is *Genshin Impact*, which has attracted a wide audience through its gacha system and diverse character content. This study aims to identify the factors that influence players' decisions to purchase virtual items in *Genshin Impact* using the *Naive Bayes* classification method. Data were collected through an online questionnaire involving 314 respondents. The analyzed variables include play duration, satisfaction with the gacha system, costumes, storyline, social influence, and previous purchase experience. The model was developed using the *Gaussian Naive Bayes* algorithm and validated through the *Stratified K-Fold Cross-Validation* method. The results show that the model achieved an accuracy of 80.57% in classifying purchase decisions. While the model performed well in identifying players who made purchases, its performance in classifying non-purchasing players requires improvement. This research is expected to serve as a reference for understanding player purchasing behavior and developing digital marketing strategies in the online gaming industry.

**Keywords:** *Genshin Impact*, Purchase Decision, Online Game, *Naive Bayes*, Data Mining

## 1. PENDAHULUAN

Perkembangan teknologi kini telah mengubah cara orang berinteraksi dan berhibur, salah satunya melalui game online. Game online kini menjadi alternatif bentuk hiburan yang paling diminati di seluruh dunia, menarik perhatian seluruh pemain game dari berbagai kalangan usia, mulai dari anak-anak hingga orang dewasa[1]. Teknologi informasi telah memberikan dampak signifikan terhadap perilaku masyarakat dalam memilih sarana hiburan, dimana game online menjadi salah satu pilihan utama yang digemari lintas generasi.

Salah satu game yang telah mencuri perhatian publik secara global adalah *Genshin Impact*, sebuah game action role-playing yang dirilis oleh pengembang asal Tiongkok, yaitu miHoYo, pada tahun 2020[2]. *Genshin Impact* berhasil menarik perhatian jutaan pemain di seluruh dunia tidak hanya karena kualitas grafisnya yang memukau, tetapi juga karena gameplay yang dinamis serta beragam karakter menarik yang dapat dikumpulkan oleh pemain. Game ini tidak hanya menawarkan grafis yang menarik dan gameplay yang inovatif, tetapi juga menerapkan sistem mikrotransaksi yang memungkinkan pemain untuk membeli item virtual dengan uang nyata.

Dalam *Genshin Impact*, pemain dapat melakukan pembelian item melalui sistem *gacha*, sebuah sistem yang mengandalkan keberuntungan untuk memperoleh karakter dan senjata secara acak. Sistem ini memberikan daya tarik tersendiri bagi pemain karena memberikan sensasi kejutan dan antisipasi. Namun demikian, sistem ini juga menimbulkan pertanyaan mengenai faktor-faktor yang mempengaruhi keputusan pemain dalam melakukan pembelian item virtual. Berbagai elemen seperti kepuasan terhadap sistem *gacha*, ketertarikan terhadap desain kostum karakter, kepuasan terhadap konten game secara keseluruhan, riwayat pembelian sebelumnya, lama waktu

bermain, serta pengaruh dari komunitas sosial dan teman sebaya diyakini memainkan peran penting dalam pengambilan keputusan untuk melakukan pembelian [3][4].

Model bisnis mikrotransaksi, dimana konsumen dapat melakukan pembelian atau pembayaran dalam jumlah kecil, telah muncul sebagai solusi bagi pemain yang ingin bersaing atau menikmati fitur tertentu dalam game tanpa harus menghabiskan waktu bermain yang lebih banyak dibandingkan pemain lainnya[5].model ini telah menjadi strategi umum dalam industri game modern, karena memungkinkan pengembang mendapatkan keuntungan yang berkelanjutan sambil memberikan fleksibilitas kepada pemain dalam memilih bagaimana mereka ingin menikmati konten dalam game tersebut. Oleh karena itu, memahami faktor-faktor yang mempengaruhi minat dan keputusan pembelian menjadi hal yang sangat penting, terutama untuk mengidentifikasi pola-pola konsumsi digital yang terus berkembang [6][7].

Untuk memahami dan menganalisis faktor-faktor yang mempengaruhi keputusan untuk membeli item virtual dalam game *Genshin Impact*, diperlukan pendekatan yang sistematis dan terstruktur. Salah satu metode yang dapat digunakan adalah algoritma klasifikasi *Naive Bayes*. Algoritma ini berbasis pada prinsip teori probabilitas dan statistik, dimana setiap fitur dianggap independen terhadap fitur lainnya. *Naive Bayes* dikenal karena kemudahan dalam implementasi serta kecepatan dalam memproses data yang besar. Metode ini sangat efektif dalam mengklasifikasikan data yang kompleks dan masif, serta mampu memberikan hasil prediksi yang cukup akurat dalam berbagai studi analisis data, termasuk pada ranah perilaku pengguna [8][9].

Penelitian sebelumnya menganalisis faktor-faktor yang mempengaruhi keputusan pembelian item pada game *Guardian Tales* dengan menggunakan metode klasifikasi *Naive Bayes* ditemukan bahwa pengalaman bermain, kepuasan terhadap sistem *gacha*, serta keinginan untuk memiliki karakter tertentu menjadi elemen dominan dalam keputusan pembelian. Penerapan metode *Naive Bayes* dalam penelitian tersebut menghasilkan tingkat akurasi sebesar 80,95%, yang membuktikan efektivitas metode ini dalam mengidentifikasi faktor-faktor yang relevan terhadap perilaku konsumen dalam game online [10].

Meskipun terdapat perbedaan dari segi grafis, gaya seni, dan alur cerita antara *Guardian Tales* dan *Genshin Impact*, kedua game ini memiliki kesamaan dari aspek model bisnis, khususnya dalam hal penerapan sistem *gacha* dan mikrotransaksi. Oleh karena itu, penerapan metode yang sama dalam konteks *Genshin Impact* diperkirakan dapat memberikan hasil relevan dan informatif .

Tujuan penelitian ini adalah untuk menerapkan metode klasifikasi *Naive Bayes* dengan menggunakan Jupyter Notebook sebagai alat bantu analisis untuk mengidentifikasi dan menganalisis faktor-faktor yang mempengaruhi keputusan pembelian item virtual di dalam game *Genshin Impact* secara komprehensif [11]. Penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai pola perilaku pembelian pemain, serta menjelaskan hubungan antara berbagai faktor seperti psikologi pemain, waktu bermain, dan respons terhadap sistem *gacha*

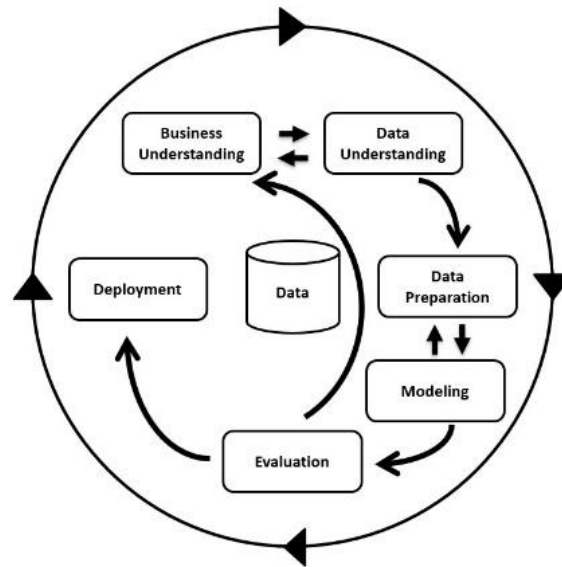
Selain itu, hasil dari penelitian ini juga dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang sistem informasi dan data mining[12], khususnya dalam konteks perilaku pengguna pada game online. Dengan demikian, penelitian ini tidak hanya untuk memahami faktor-faktor yang mempengaruhi keputusan pembelian, tetapi juga untuk mengukur tingkat akurasi metode *Naive Bayes* dalam memprediksi keputusan tersebut. Hasil dari penelitian ini dapat memberikan gambaran yang lebih jelas, terukur, dan terstruktur tentang hubungan antara berbagai faktor dan keputusan pembelian dalam konteks game *Genshin Impact*, sehingga bisa menjadi dasar untuk penelitian lanjutan maupun pengembangan strategi pemasaran digital pada industri game.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai kerangka kerja dalam pelaksanaan proses data mining. CRISP-DM dipilih karena model ini bersifat fleksibel, sistematis, dan telah terbukti digunakan secara luas dalam berbagai industri. Model ini terdiri dari enam fase utama yang saling berkaitan dan memungkinkan untuk dilakukan secara iteratif, sesuai dengan kebutuhan dan kondisi data yang dianalisis. Seperti terlihat pada gambar 1.

### 2.1 Business Understanding

Tahapan pertama bertujuan untuk memahami konteks bisnis dan tujuan utama dari dilakukannya proses datamining. Dalam penelitian ini, konteks bisnisnya adalah industri game online yang mengandalkan sistem mikrotransaksi sebagai sumber pendapatan. Tujuan spesifik penelitian adalah untuk mengidentifikasi faktor-faktor utama yang memengaruhi keputusan pemain dalam melakukan pembelian item virtual pada game *Genshin Impact*. Pemahaman ini dituangkan ke dalam rumusan masalah dan tujuan penelitian sebagai dasar seluruh tahapan selanjutnya.



**Gambar 1.** Alur CRISP-DM

## 2.2 Data Understanding

Pada tahap ini dilakukan proses pengumpulan, eksplorasi, dan pemahaman karakteristik data. Data diperoleh melalui kuesioner online yang disebarakan kepada pemain *Genshin Impact* di berbagai platform komunitas game. Data yang terkumpul mencakup atribut-atribut seperti: tingkat kepuasan terhadap sistem gacha, daya tarik karakter atau kostum, intensitas pembelian sebelumnya, durasi bermain, serta pengaruh sosial. Proses eksplorasi dilakukan untuk mendeteksi kemungkinan permasalahan seperti nilai kosong (*missing values*), data duplikat, dan ketidaksesuaian tipe data. Pemahaman data yang baik sangat penting untuk menentukan kualitas model yang akan dibangun.

## 2.3 Data Preparation

Tahap data preparation bertujuan untuk mempersiapkan data agar siap digunakan dalam proses klasifikasi. Proses ini mencakup pembersihan data (*data cleansing*) untuk menghapus nilai yang hilang atau tidak konsisten, seleksi atribut (*feature selection*) untuk memilih variabel yang relevan, serta transformasi data seperti *label encoding* atau *one-hot encoding* untuk mengubah data kategorikal menjadi numerik. Setelah itu, dilakukan normalisasi dan penyusunan dataset akhir agar sesuai dengan format yang dibutuhkan oleh algoritma *Naïve Bayes*.

## 2.4 Modeling

Pada tahap ini dilakukan pembangunan model klasifikasi menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* berfungsi untuk klasifikasi suatu variabel tertentu dengan memanfaatkan metode probabilitas dan statistic[13]. Algoritma ini bekerja berdasarkan teorema Bayes, yang mengasumsikan bahwa setiap fitur (atribut) bersifat independen terhadap fitur lainnya, meskipun dalam kenyataan hal ini tidak selalu terjadi. Kelebihan dari algoritma ini adalah kecepatan komputasi yang tinggi dan kemampuan menangani dataset berukuran besar.

Rumus dasar *Naive Bayes* adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Dalam penerapannya, untuk mengevaluasi performa model secara lebih obyektif dan menghindari bias akibat pembagian data yang tidak merata, digunakan metode validasi silang (*cross-validation*), khususnya teknik *K-Fold Cross Validation*. *Cross-validasi*, atau yang dikenal sebagai estimasi rotasi, adalah teknik untuk menilai performa model dengan mengukur sejauh mana hasil analisis statistik dapat diterapkan atau digeneralisasi ke data lain yang bersifat independen[14]. Metode ini bertujuan untuk menguji kemampuan generalisasi model dengan membagi dataset menjadi K bagian (*fold*) yang berukuran kurang lebih sama. Proses pelatihan dan pengujian dilakukan sebanyak K iterasi, di mana pada setiap iterasi satu *fold* digunakan sebagai data uji (*testing set*) dan sisanya digunakan sebagai data latih (*training set*). Nilai evaluasi (seperti akurasi, presisi, recall, dan F1-score) dari setiap iterasi kemudian dirata-ratakan untuk memperoleh estimasi performa model yang lebih stabil dan menyeluruh.

## 2.5 Evaluation

Evaluasi dilakukan untuk mengukur kinerja model klasifikasi. Evaluasi ini menggunakan metrik *Confusion Matrix*, yang menggambarkan perbandingan antara hasil prediksi dengan nilai aktual. Confusion Matrix digunakan untuk menilai seberapa akurat hasil dari suatu proses klasifikasi. Akurasi menunjukkan persentase prediksi yang benar dari total keseluruhan prediksi yang dilakukan. Untuk menghitung nilai akurasi, presisi, dan *recall* [15]. Adapun struktur confusion matrix untuk dua kelas (*binary classification*) dapat dilihat pada tabel 1.

**Tabel 1.** *Confusion Matrix*

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Dari matriks tersebut dapat dihitung metrik evaluasi sebagai berikut:

Akurasi:

Akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-score:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

## 2.6 Deployment

Tahap akhir dalam proses klasifikasi adalah implementasi model ke dalam konteks nyata atau sistem pendukung keputusan. Meskipun dalam penelitian ini tahap *deployment* belum direalisasikan secara langsung pada platform produksi, rancangan implementasinya telah disiapkan secara konseptual. Model yang dihasilkan memiliki potensi untuk diintegrasikan ke dalam sistem berbasis web atau aplikasi analitik guna mendukung proses pengambilan keputusan secara cepat dan objektif. Pendekatan ini memungkinkan pemanfaatan hasil klasifikasi sebagai alat bantu analisis non-invasif, baik dalam konteks akademik maupun pengembangan sistem cerdas. Dengan demikian, keseluruhan tahapan dalam kerangka kerja CRISP-DM telah dilaksanakan secara metodologis dan dapat dipertanggungjawabkan dari segi ilmiah.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Business Understanding

Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan faktor-faktor yang memengaruhi keputusan pembelian item dalam game Genshin Impact menggunakan algoritma Naïve Bayes. Pembelian item seperti genesis crystal, battle pass dan kostum merupakan bagian penting dari model bisnis game ini. Permasalahan utama yang diangkat adalah bagaimana memprediksi keputusan pembelian pemain berdasarkan sejumlah faktor, yaitu lama bermain, kepuasan terhadap gacha, kosmetik, alur cerita, pengaruh sosial, pengalaman pembelian sebelumnya, dan keputusan pembelian itu sendiri. Model yang dibangun diharapkan mampu memberikan prediksi akurat serta wawasan dalam penerapan data mining pada perilaku konsumen digital.

## 3.2 Data Understanding

Tahap *Data Understanding* bertujuan untuk memperoleh pemahaman yang mendalam terhadap data yang digunakan dalam proses klasifikasi. Data diperoleh melalui penyebaran kuesioner kepada pemain game Genshin Impact secara daring, dengan jumlah responden sebanyak 314 orang. Kuesioner disusun untuk mengukur variabel-

variabel yang diduga memengaruhi keputusan pembelian item dalam game.berikut adalah tabel data yang didapat dari penyebaran kuisioner dapat dilihat pada tabel 2.

**Tabel 2.** Dataset

No	Usia	Jenis kelamin	Lama Bermain	Kepuasan Gacha	Kepuasan Kosmetik	Kepuasan Alur cerita	Pengaruh sosial	Pembelian Sebelumnya	Keputusan Pembelian
1	Laki - Laki	17 - 25 Tahun	Sangat Lama	Tinggi	Sedang	Tinggi	Tinggi	Positif	Ya
2	Laki - Laki	17 - 25 Tahun	Sangat Lama	Sedang	Tinggi	Sangat Tinggi	Sangat Rendah	Negatif	Ya
3	Laki - Laki	17 - 25 Tahun	Sedang	Rendah	Tinggi	Sedang	Tinggi	Netral	Tidak
4	Laki - Laki	<17 Tahun	Sangat Lama	Sedang	Tinggi	Sangat Tinggi	Sedang	Netral	Ya
5	Laki - Laki	17 - 25 Tahun	Sangat Lama	Sedang	Sedang	Tinggi	Sangat Tinggi	Netral	Ya
...	.....	.....	.....	.....	.....	.....	.....	.....	.....
315	Laki - Laki	17 - 25 Tahun	Baru	Sangat Rendah	Rendah	Tinggi	Rendah	Netral	Tidak

Adapun atribut yang digunakan dalam perhitungan Naïve Bayes Classifier dapat dilihat pada tabel 3.

**Tabel 3.** Atribut yang Digunakan

No	Nama Atribut	Tipe Data	Keterangan
1	Lama Bermain	Kategorikal	Tingkat durasi pemain bermain: Baru, Sedang, Lama, atau Sangat Lama
2	Kepuasan terhadap Gacha	Kategorikal	Penilaian pemain terhadap sistem gacha (Sangat Tinggi – Sangat Rendah)
3	Kepuasan terhadap Kosmetik	Kategorikal	Penilaian pemain terhadap item kostum/kosmetik (Sangat Tinggi – Sangat Rendah)
4	Kepuasan terhadap Cerita	Kategorikal	Penilaian pemain terhadap konten cerita dalam game (Sangat Tinggi – Sangat Rendah)
5	Pengaruh Pembelian Sebelumnya	Kategorikal	Pengaruh pembelian item sebelumnya (Positif, Netral, Negatif)
6	Pengaruh Sosial	Kategorikal	Apakah keputusan dipengaruhi oleh teman/komunitas (Sangat Tinggi – Sangat Rendah)
7	Keputusan Pembelian	Kategorikal	Label target, yaitu keputusan akhir: Membeli atau Tidak Membeli (Ya/Tidak)

### 3.3 Data Preparation

Tahap *data preparation* merupakan langkah penting dalam proses data mining untuk memastikan bahwa data dalam kondisi optimal sebelum digunakan dalam pemodelan. Proses ini mencakup pembersihan data, seleksi fitur, transformasi data kategorikal, dan pembagian data untuk pelatihan dan pengujian. Berikut adalah rincian tahapannya:

a. **Penghapusan kolom non atribut**

Kolom-kolom yang tidak relevan untuk proses klasifikasi, seperti Usia, dan Jenis kelamin, dihapus dari dataset karena tidak termasuk dalam variabel yang dianalisis. Fokus utama adalah pada variabel-variabel yang menjadi faktor dalam pengambilan keputusan pembelian.

**Tabel 4.** *Atribut selection*

No	Atribut yang di gunakan
1	Lama bermain
2	Kepuasan terhadap gacha
3	Kepuasan terhadap kosmetik
4	Kepuasan terhadap alur cerita
5	Pengaruh sosial
6	Pengaruh pembelian sebelumnya
7	Keputusan pembelian

b. **Transformasi Data Kategorikal (*Label Encoding*)**

Metode encoding di gunakan pada fitur Lama bermain, Kepuasan Gacha, Kepuasan Kosmetik, Kepuasan alur cerita, Pengaruh sosial, Pengalaman pembelian sebelumnya dan juga keputusan pembelian yang akan di rubah menjadi numerik yang dapat di proses oleh bahasa mesin.

**Tabel 5.** *Label Encoding*

No	Fitur	Nilai Kategorikal	Nilai Kodefikasi
1	Lama Bermain	Sangat Lama, Lama, Sedang, Sebentar	3, 2, 1, 0
2	Kepuasan Gacha	Sangat Tinggi, Tinggi, Sedang, Rendah, Sangat Rendah	4, 3, 2, 1, 0
3	Kepuasan Kosmetik	Sangat Tinggi, Tinggi, Sedang, Rendah, Sangat Rendah	4, 3, 2, 1, 0
4	Kepuasan Alur Cerita	Sangat Tinggi, Tinggi, Sedang, Rendah, Sangat Rendah	4, 3, 2, 1, 0
5	Pengaruh Sosial	Sangat Tinggi, Tinggi, Sedang, Rendah, Sangat Rendah	4, 3, 2, 1, 0
6	Pengalaman Pembelian Sebelumnya	Positif, Netral, Negatif	2, 1, 0
7	Keputusan Pembelian (Label)	Ya, Tidak	1, 0

c. **Pemisahan Fitur dan Label**

Label yang digunakan sebagai target klasifikasi adalah variabel Keputusan Pembelian yang terdiri dari dua kategori "Ya" untuk pengguna yang melakukan pembelian, dan "Tidak" bagi yang tidak melakukan pembelian. Pemisahan fitur dan label ini merupakan tahapan, karena memungkinkan algoritma pembelajaran mesin untuk memahami hubungan antara fitur dan label secara sistematis.

**Tabel 6.** *Pemisahan fitur dan label*

No	Fitur	Label
1	Lama bermain	Keputusan pembelian
2	Kepuasan gacha	
3	Kepuasan kosmetik	
4	Kepuasan alur cerita	
5	Pengaruh sosial	
6	Pengalaman pembelian sebelumnya	

d. **Pembagian Data Latih dan Uji (*Train-Test Split*)**

Untuk evaluasi awal model, data dibagi menjadi dua bagian, yaitu 60% sebagai data latih dan 40% sebagai data uji. Proses ini membantu mengukur performa model terhadap data yang belum pernah dilatih sebelumnya.

**Tabel 7. Split Data**

Jenis data	Jumlah Data	Presentase
Data Latih	188	60%
Data uji	126	40%

### 3.4 Modeling

#### a. Pelatihan Model

Setelah tahap prapemrosesan data selesai, langkah selanjutnya adalah melakukan pelatihan model klasifikasi menggunakan algoritma *Gaussian Naive Bayes*. Algoritma ini dipilih karena memiliki efisiensi komputasi yang tinggi serta mampu menangani data dengan fitur kategorikal yang telah dikodekan secara numerik. Model dilatih untuk mengenali pola dari sejumlah fitur input, seperti lama bermain, kepuasan terhadap sistem gacha, kepuasan kosmetik, alur cerita, pengaruh sosial, dan pengalaman pembelian sebelumnya, guna memprediksi keputusan akhir pemain dalam melakukan pembelian item virtual. Proses pelatihan dilakukan dengan membagi data menjadi dua bagian, yaitu data latih sebesar 60% dan data uji sebesar 40%. Pemisahan ini bertujuan agar model dapat belajar dari mayoritas data (data latih), dan kemudian diuji kemampuannya terhadap data yang belum pernah dilihat sebelumnya (data uji). Model dilatih dengan mengestimasi distribusi probabilitas setiap fitur terhadap kelas target menggunakan pendekatan distribusi Gaussian, yang mengasumsikan bahwa data memiliki sebaran normal. Selama proses pelatihan, model mempelajari hubungan antar fitur dengan label target (keputusan pembelian), dengan mengoptimalkan probabilitas posterior berdasarkan Teorema Bayes. Hasil dari pelatihan ini berupa model klasifikasi yang mampu memprediksi apakah seorang pemain akan melakukan pembelian item virtual berdasarkan kombinasi nilai-nilai pada fitur yang diberikan.

#### b. Validasi data

Untuk mengevaluasi kinerja model secara objektif, digunakan metode Stratified K-Fold Cross-Validation dengan  $k=5$ . Teknik ini membagi dataset menjadi lima fold dengan distribusi kelas yang seimbang, lalu melakukan pelatihan dan pengujian sebanyak lima kali secara bergantian. Metode ini mampu mengurangi bias evaluasi, menjaga konsistensi proporsi kelas, dan mendeteksi potensi overfitting, sehingga memberikan estimasi performa model yang lebih stabil dan representatif. Rangkaian proses pelatihan dan pengujian pada masing-masing *fold* disajikan secara sistematis pada Tabel 13 berikut. Tabel ini menunjukkan urutan kombinasi data latih dan data uji yang digunakan dalam setiap iterasi, serta proses pelatihan yang dilakukan:

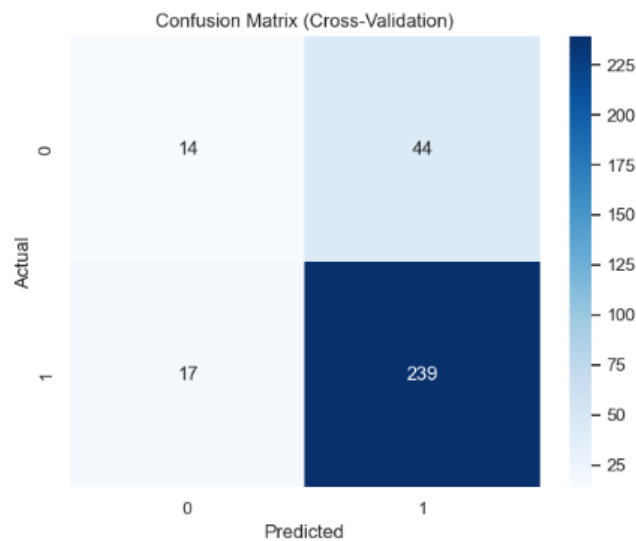
**Tabel 8. K-Fold Cross Validation**

Fold ke-	Data Latih	Data Uji	Proses yang Dilakukan
1	Fold 2, Fold 3, Fold 4, Fold 5	Fold 1	Model dilatih dengan Fold 2-5, diuji dengan Fold 1
2	Fold 1, Fold 3, Fold 4, Fold 5	Fold 2	Model dilatih dengan Fold 1,3,4,5; diuji dengan Fold 2
3	Fold 1, Fold 2, Fold 4, Fold 5	Fold 3	Model dilatih dengan Fold 1,2,4,5; diuji dengan Fold 3
4	Fold 1, Fold 2, Fold 3, Fold 5	Fold 4	Model dilatih dengan Fold 1,2,3,5; diuji dengan Fold 4
5	Fold 1, Fold 2, Fold 3, Fold 4	Fold 5	Model dilatih dengan Fold 1-4, diuji dengan Fold 5

### 3.5 Evaluation

#### a. Confusion Matrix

*Confusion matrix* digunakan untuk mengevaluasi performa klasifikasi dengan membandingkan label yang diprediksi oleh model dengan label yang sebenarnya. Ini memberikan gambaran jelas tentang seberapa banyak prediksi yang benar dan salah pada masing-masing kelas. Hasil dapat di lihat pada gambar 2.



**Gambar 2.** Confusion Matrix

Gambar 2 merupakan hasil *confusion matrix* menunjukkan bahwa model lebih akurat dalam memprediksi individu yang membeli item dibandingkan dengan yang tidak. Sebagian besar prediksi benar terjadi pada kelas mayoritas (pembeli), sedangkan kelas minoritas (tidak membeli) sering salah diklasifikasikan. Hal ini mencerminkan ketidakseimbangan kelas dalam data, yang menyebabkan model cenderung bias terhadap kelas mayoritas.

**b. Akurasi Model**

Akurasi model *Naive Bayes* mencapai 80,57%, yang berarti bahwa model ini berhasil mengklasifikasikan sekitar 80,57% dari data uji dengan benar. Nilai akurasi ini diperoleh sebagai hasil rata-rata dari penerapan teknik validasi *Stratified K-Fold Cross-Validation* dengan nilai  $k = 5$ . Metode ini memberikan evaluasi yang lebih menyeluruh terhadap performa model dengan memanfaatkan seluruh data secara bergantian sebagai data latih dan data uji. Hasil akurasi pada setiap fold disajikan pada Tabel 9.

**Tabel 9.** *K-Fold Cross-Validation*

Fold ke-	Akurasi (%)
1	82,54
2	80,95
3	79,37
4	77,78
5	82,86
Rata-rata	80,57

**c. Classification Report**

Hasil *classification report* menunjukkan bahwa model memiliki precision dan recall yang tinggi pada kelas pembeli (0.84 dan 0.93), menandakan kinerja yang baik dalam mendeteksi individu yang membeli item. Sebaliknya, performa pada kelas tidak membeli jauh lebih rendah (precision 0.45, recall 0.24), mengindikasikan bahwa model kurang efektif dalam mengidentifikasi individu yang tidak melakukan pembelian.

**Tabel 10.** *Classification Report*

Kelas	Precision	Recall	F1-score	Support
0 (Tidak Membeli)	0.45	0.24	0.31	58
1 (Membeli)	0.84	0.93	0.89	256
<b>Akurasi</b>	-	-	0.81	314
<b>Macro avg</b>	0.65	0.59	0.60	314
<b>Weighted avg</b>	0.77	0.81	0.78	314

Model menunjukkan performa tinggi pada kelas pembeli, dengan precision sebesar 0.84, recall 0.93, dan F1-score 0.89, menandakan akurasi dan konsistensi yang baik dalam mendeteksi individu yang melakukan pembelian. Sebaliknya, performa pada kelas tidak membeli jauh lebih rendah (precision 0.45, recall 0.24), mengindikasikan kelemahan model dalam mengklasifikasikan kelas minoritas. Metrik rata-rata menunjukkan akurasi keseluruhan sebesar 81%, namun masih diperlukan peningkatan pada prediksi kelas non-pembeli.

### 3.6 Deployment

Pada tahap *deployment*, model Naive Bayes yang telah dilatih digunakan untuk memprediksi keputusan pembelian item dalam game *Genshin Impact*. Model diimplementasikan menggunakan pustaka *scikit-learn* dan diterapkan pada data baru berisi fitur pemain, seperti lama bermain dan kepuasan gacha. Hasil prediksi memberikan probabilitas apakah pemain cenderung membeli item atau tidak. Evaluasi dilakukan menggunakan data uji untuk memastikan akurasi model. Model ini dapat digunakan untuk analisis perilaku pemain dan akan dipantau secara berkala, termasuk retraining jika diperlukan, guna menjaga akurasi dan relevansi dalam jangka panjang.

## 4. KESIMPULAN

Variabel seperti kepuasan terhadap sistem gacha, kepuasan terhadap kostum, alur cerita, pengaruh sosial, pengalaman pembelian sebelumnya, dan lama bermain memiliki kontribusi dalam pengambilan keputusan pembelian. Model klasifikasi yang dibangun menggunakan algoritma *Gaussian Naive Bayes* menghasilkan akurasi sebesar 80,57%, dengan performa yang tinggi dalam mengklasifikasikan pemain yang melakukan pembelian. Namun, model masih kurang optimal dalam mengidentifikasi pemain yang tidak membeli item, ditandai dengan rendahnya precision dan recall pada kelas tersebut. Hal ini menunjukkan adanya ketidakseimbangan kelas pada data yang digunakan. Secara keseluruhan, metode Naive Bayes terbukti efektif untuk mengidentifikasi pola perilaku pembelian dalam game online.

## REFERENSI

- [1] I. Permata Sari and U. Muhammadiyah Hamka, "DAMPAK GAME ONLINE MOBILE LEGENDS TERHADAP PERKEMBANGAN EMOSI SISWA KELAS V SDI," *Jurnal Cakrawala Pendas*, vol. 8, no. 4, 2022, doi: 10.31949/jcp.v8i2.2798.
- [2] Aldiansyah Pratama, M., Hasan, F. N., & Pratama, M. A. (2024). Comparison of the Naïve Bayes Method and Support Vector Machine in Sentiment Analysis of Genshin Impact Game Reviews. In MECOMARE (Vol. 13, Issue 2), 76-88. <https://doi.org/10.35335/computational.v13i2.198>
- [3] I. Ahmed, M. M. Hoque, N. Banik, A. Rahman, M. Nur-E-Alam, and M. A. Islam, "Optimizing Classification Methods for Online Buyers' Purchase Intentions in Bangladesh," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 12, no. 6, pp. 17–24, Mar. 2024, doi: 10.35940/ijrte.E7987.12060324.
- [4] Q. M. Ilyas, A. Mehmood, A. Ahmad, and M. Ahmad, "A Systematic Study on a Customer's Next-Items Recommendation Techniques," Jun. 01, 2022, *MDPI*. doi: 10.3390/su14127175.
- [5] P. Saputra, "Dampak Microtransaction pada Kepuasan Pemain di PUBG Mobile," *Equator Journal of Management and Entrepreneurship (EJME)*, vol. 11, no. 02, p. 070, Apr. 2023, doi: 10.26418/ejme.v11i02.64334.
- [6] B. Naqvi, K. Perova, A. Farooq, I. Makhdoom, S. Oyedeji, and J. Porras, "Mitigation strategies against the phishing attacks: A systematic literature review," Sep. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.cose.2023.103387.
- [7] R. R. Saxena, "Examining Reactions about COVID-19 Vaccines: A Systematic Review of Studies Utilizing Deep Learning for Sentiment Analysis," Apr. 09, 2024. doi: 10.36227/techrxiv.171262869.98334885/v1.
- [8] A. C. Ribeiro, A. Sizo, and L. P. Reis, "Investigating the reviewer assignment problem: A systematic literature review," *J Inf Sci*, 2023, doi: 10.1177/01655515231176668.
- [9] A. A. Al-Kadhimi, M. M. Singh, and M. N. A. Khalid, "A Systematic Literature Review and a Conceptual Framework Proposition for Advanced Persistent Threats (APT) Detection for Mobile Devices Using Artificial Intelligence Techniques," Jul. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app13148056.
- [10] Amin, M., & Fathurrahman, M. A. (n.d.). INFORMASI (Jurnal Informatika dan Sistem Informasi) Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Item pada Game Guardian Tales:



- Pendekatan Klasifikasi dengan Metode Naive Bayes. *INFORMASI Jurnal Informatika Dan Sistem Informasi*. Volume 15 No.2, 171-180. <https://doi.org/10.37424/informasi.v15i2.257>.
- [11] Bacht, A. S., Agustama, A. T., Wiranata, H., Daniati, E., & Ristyawan, A. (2024). Prediksi Harga Mobil Toyota Bekas Menggunakan Algoritma Regresi Linear Berganda. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8(1), 173–183. <https://doi.org/10.29407/INOTEK.V8I1.4924>.
- [12] Moh. Iqbal, I. J., Afrizal A. B .P., Aidina R., Erna D. (2024). Klasifikasi Penggunaan Video Game Dengan Menggunakan Metode Algoritma Naive Bayes *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8 (2), 910-918. <https://doi.org/10.29407/inotek.v8i2.5020>.
- [13] A. C. Khotimah and E. Utami, “COMPARISON NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON TWITTER ACCOUNT,” *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 3, 2022, doi: 10.20884/1.jutif.2022.3.3.254.
- [14] F. Tempola, M. Muhammad, and A. Khairan, “PERBANDINGAN KLASIFIKASI ANTARA KNN DAN NAIVE BAYES PADA PENENTUAN STATUS GUNUNG BERAPI DENGAN K-FOLD CROSS VALIDATION COMPARISON OF CLASSIFICATION BETWEEN KNN AND NAIVE BAYES AT THE DETERMINATION OF THE VOLCANIC STATUS WITH K-FOLD CROSS VALIDATION,” vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik20185983.
- [15] S. Sucipto, D. Dwi Prasetya, and T. Widiyaningtyas, “Educational Data Mining: Multiple Choice Question Classification in Vocational School,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 379–388, Mar. 2024, doi: 10.30812/matrik.v23i2.3499.