

**LAPORAN FINAL PROJECT**  
**“Klasifikasi Engagement Level Pemain Game”**



Dosen Pengampu :  
Enda Putri Atika, M. Kom

Disusun Oleh :

Achmed Bintang Asy-Syfa M.	23.11.5818
Muhammad Rama Saetama	23.11.5854
Muhammad Naufal Yazid Akbar	23.11.5870
Muh. Hanif Nur Siddiq	23.11.5920
Erwin Eka Pratama	20.11.3720

**PROGRAM STUDI SARJANA INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA 2025**

## DAFTAR ISI

<b>BAB I</b>	<b>3</b>
<b>PENDAHULUAN</b>	<b>3</b>
1.1 Latar Belakang	3
1.2 Tujuan	3
<b>BAB II</b>	<b>4</b>
<b>METODE ANALISIS DATA</b>	<b>4</b>
2.1 EDA dan Visualisasi Data	4
1. Import Library	4
2. Load Dataset	5
3. Preprocessing Data	5
4. Split Data	6
5. Normalisasi Data	7
2.2 Analisis Korelasi	7
1. Eksplorasi dan Visualisasi Data	9
2. Visualisasi Lanjutan: Analisis Fitur terhadap Engagement Level	11
2.3 Model Regresi dan Evaluasi	13
<b>BAB III</b>	<b>16</b>
<b>HASIL DAN PEMBAHASAN</b>	<b>16</b>
3.1 Eksplorasi Data dan Visualisasi	16
3.2 Pembuatan Model dan Evaluasi	16
3.3 Pembahasan Hasil	17
<b>BAB IV</b>	<b>18</b>
<b>KESIMPULAN</b>	<b>18</b>

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Di era digital ini, industri permainan mengalami kemajuan yang pesat. Sejumlah perusahaan game memanfaatkan data pengguna guna menganalisis kebiasaan bermain, memperbaiki pengalaman pengguna, dan meningkatkan partisipasi. Karena itu, analisis data pemain menjadi krusial, terutama dalam mengidentifikasi elemen-elemen yang memengaruhi tingkat keterlibatan (Engagement Level) pemain dalam permainan.

Perkembangan teknologi digital telah memberikan pengaruh besar terhadap industri hiburan, terutama industri game. Game digital saat ini tidak hanya menjadi media hiburan semata, tetapi juga telah berkembang menjadi platform interaktif yang dapat mengumpulkan data dalam jumlah besar dari para pemainnya. Setiap tindakan yang diambil oleh pemain dalam permainan bisa dicatat dalam bentuk data, termasuk waktu bermain, lama sesi, jenis genre yang disukai, serta tingkat pencapaian dan pembelian dalam permainan.

Informasi yang tersimpan dalam data tersebut dapat dimanfaatkan untuk memahami perilaku pemain, merancang strategi kustom, dan meningkatkan pengalaman pengguna secara keseluruhan. Salah satu elemen krusial yang ingin dianalisis oleh pengembang game adalah tingkat keterlibatan pemain (engagement). Tingkat partisipasi ini berhubungan langsung dengan seberapa sering dan lama pemain memainkan permainan, serta apakah mereka berminat untuk terus menggunakan game itu.

Akan tetapi, untuk mengidentifikasi pola keterlibatan ini, diperlukan proses analisis data yang akurat. Dalam konteks ini, metode data science seperti eksplorasi data, visualisasi, analisis korelasi, dan pengembangan model prediktif seperti regresi linear dapat diterapkan untuk mengidentifikasi elemen-elemen yang memengaruhi keterlibatan pemain.

Dalam proyek ini, dilakukan analisis terhadap kumpulan data yang berisi informasi perilaku gamer untuk mengidentifikasi hubungan antara variabel dan membangun model prediktif mengenai tingkat keterlibatan pemain. Dengan demikian, diharapkan hasil analisis ini dapat memberikan pandangan bagi pengembang game dalam membuat keputusan yang lebih akurat dan berdasarkan data.

### **1.2 Tujuan**

1. Mengeksplorasi dan memvisualisasikan data pengguna game.
2. Melakukan analisis korelasi antar variabel.
3. Membangun dan mengevaluasi model regresi linear untuk memprediksi tingkat keterlibatan pemain.

## BAB II

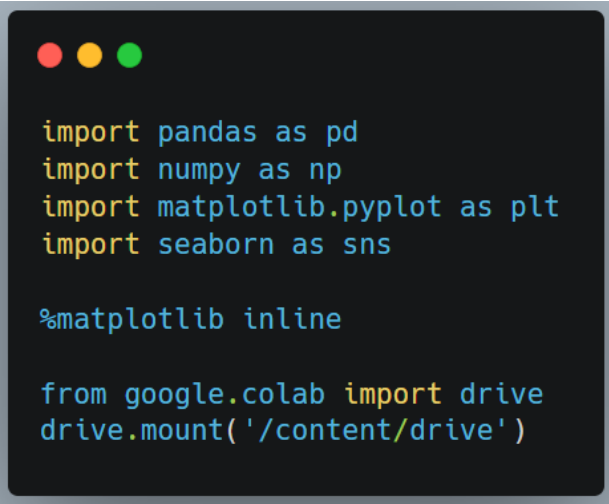
### METODE ANALISIS DATA

#### 2.1 EDA dan Visualisasi Data

##### 1. Import Library

Langkah awal dalam proses analisis data adalah mengimpor library yang diperlukan. Library merupakan kumpulan modul atau fungsi yang telah tersedia dalam Python untuk memudahkan analisis, visualisasi, dan pemodelan data

Kode berikut digunakan untuk mengimport library:



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline

from google.colab import drive
drive.mount('/content/drive')
```

Beberapa Library yang kita digunakan sebagai berikut

- Pandas adalah sebuah library yang digunakan untuk pengolahan dan analisis data. Ini membuat kita bisa melakukan operasi seperti penyaringan, pengelompokan, penggabungan, dan penataan ulang data.
- Numpy adalah sebuah library yang digunakan untuk perhitungan numerik yang menyediakan fitur-fitur berguna untuk operasi pada n-array dan matriks dalam Python.
- Matplotlib adalah sebuah library yang digunakan untuk menghasilkan visualisasi data seperti grafik, histogram, dan scatter plot.
- Seaborn adalah sebuah library yang digunakan untuk membuat visualisasi data, dan dibangun diatas matplotlib.
- Google Collab Drive  
Digunakan untuk mengakses berkas yang disimpan di Google Drive, seperti berkas dataset yang dipakai dalam proyek ini.

## 2. Load Dataset

Dataset yang digunakan dalam proyek ini bernama *online gaming behavior dataset.csv* dan disimpan di Google Drive. Berikut adalah kode yang digunakan untuk membaca file tersebut:

```
import pandas as pd
df=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/online_gaming_behavior_dataset.csv')
df.head(10)
```

Kode tersebut digunakan untuk:

- Membaca file CSV ke dalam sebuah DataFrame bernama **df**.
- Menampilkan 10 baris pertama dari dataset untuk melihat isi data.

Dan ini tampilan dataset nya :

	PlayerID	Age	Gender	Location	GameGenre	PlayTimeHours	InGamePurchases	GameDifficulty	SessionsPerWeek	AvgSessionDurationMinutes	PlayerLevel	AchievementsUnlocked	EngagementLevel
0	9000	43	Male	Other	Strategy	16.271119	0	Medium	6	108	79	25	Medium
1	9001	29	Female	USA	Strategy	5.525961	0	Medium	5	144	11	10	Medium
2	9002	22	Female	USA	Sports	8.223755	0	Easy	16	142	35	41	High
3	9003	35	Male	USA	Action	5.265351	1	Easy	9	85	57	47	Medium
4	9004	33	Male	Europe	Action	15.531945	0	Medium	2	131	95	37	Medium
5	9005	37	Male	Europe	RPG	20.561855	0	Easy	2	81	74	22	Low
6	9006	25	Male	USA	Action	9.752716	0	Hard	1	50	13	2	Low
7	9007	25	Female	Asia	RPG	4.401729	0	Medium	10	48	27	23	Medium
8	9008	38	Female	Europe	Simulation	18.152733	0	Easy	5	101	23	41	Medium
9	9009	38	Female	Other	Sports	23.942772	0	Easy	13	95	99	36	High

Penjelasan beberapa kolom penting:

- **PlayTime Hours:** total jam bermain.
- **Game Genre:** genre game yang dimainkan.
- **Game Difficulty:** tingkat kesulitan game.
- **Sessions Per Week:** berapa kali bermain per minggu.
- **Player Level:** level pemain.
- **Achievement Unlocked:** jumlah pencapaian yang telah diperoleh.
- **Engagement Level:** tingkat keterlibatan pemain (Low, Medium, High).

### 3. Preprocessing Data

Sebelum model machine learning dibangun, data perlu dipersiapkan agar dapat diproses dengan baik. Tahapan ini dikenal dengan istilah *data preprocessing*, yaitu proses untuk mengubah data mentah menjadi bentuk yang lebih mudah dipahami oleh algoritma machine learning.

#### - Encoding Data Kategorikal

Beberapa kolom dalam dataset berisi data kategorik seperti Gender, Location, Game Genre, Game Difficulty, dan Engagement Level. Karena sebagian besar algoritma machine learning tidak dapat bekerja langsung dengan data teks, maka data ini perlu diubah menjadi bentuk numerik menggunakan teknik *Label Encoding*.

Berikut kode yang digunakan:

```
label_cols = ['Gender', 'Location', 'GameGenre', 'GameDifficulty', 'EngagementLevel']
label_encoders = {}
for col in label_cols:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le
df.head()
```

Penjelasan:

- Untuk setiap kolom, **LabelEncoder** akan mempelajari (**fit**) semua kategori unik di dalamnya (misalnya, pada kolom **Gender**, ia akan menemukan 'Female' dan 'Male').
- Setelah itu, ia akan mengubah (**transform**) setiap kategori tersebut menjadi sebuah angka integer unik. Proses ini biasanya mengikuti urutan abjad, contohnya: 'Female' menjadi **0** dan 'Male' menjadi **1**.
- Kolom asli di dalam DataFrame (**df**) langsung diperbarui dengan versi numerik ini.

Setelah encoding, tampilan dataset menjadi seperti berikut:

	PlayerID	Age	Gender	Location	GameGenre	PlayTimeHours	InGamePurchases	GameDifficulty	SessionsPerWeek	AvgSessionDurationMinutes	PlayerLevel	AchievementsUnlocked	EngagementLevel
0	9000	43	1	2	4	16.271119	0	2	6	108	79	25	2
1	9001	29	0	3	4	5.525961	0	2	5	144	11	10	2
2	9002	22	0	3	3	8.223755	0	0	16	142	35	41	0
3	9003	35	1	3	0	5.265351	1	0	9	85	57	47	2
4	9004	33	1	1	0	15.531945	0	2	2	131	95	37	2

## 4. Split Data

Setelah preprocessing selesai, langkah selanjutnya adalah membagi dataset menjadi dua bagian:

- Data latih (train set): Digunakan untuk melatih model.
- Data uji (test set): Digunakan untuk mengevaluasi performa model.

Kode berikut digunakan untuk membagi data:

```
X = df.drop(columns=['EngagementLevel', 'PlayerID'])
y = df['EngagementLevel']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Penjelasan:

- X berisi semua fitur (kolom-kolom prediktor) kecuali kolom target Engagement Level dan PlayerID.
- y adalah label target yang ingin diprediksi, yaitu Engagement Level.
- train\_test\_split digunakan untuk membagi data menjadi 80% data latih dan 20% data uji dengan parameter random\_state=42 agar hasil pembagian data bersifat konsisten.

## 5. Normalisasi Data

Setelah data dibagi menjadi data latih dan data uji, langkah selanjutnya adalah melakukan normalisasi. Normalisasi bertujuan untuk menstandarkan skala dari fitur numerik agar memiliki distribusi yang seragam, biasanya dengan rata-rata 0 dan standar deviasi 1. Ini sangat penting untuk algoritma yang sensitif terhadap skala data, seperti algoritma pembelajaran berbasis jarak atau yang menggunakan regularisasi.

Kode berikut digunakan untuk normalisasi:

```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Penjelasan:

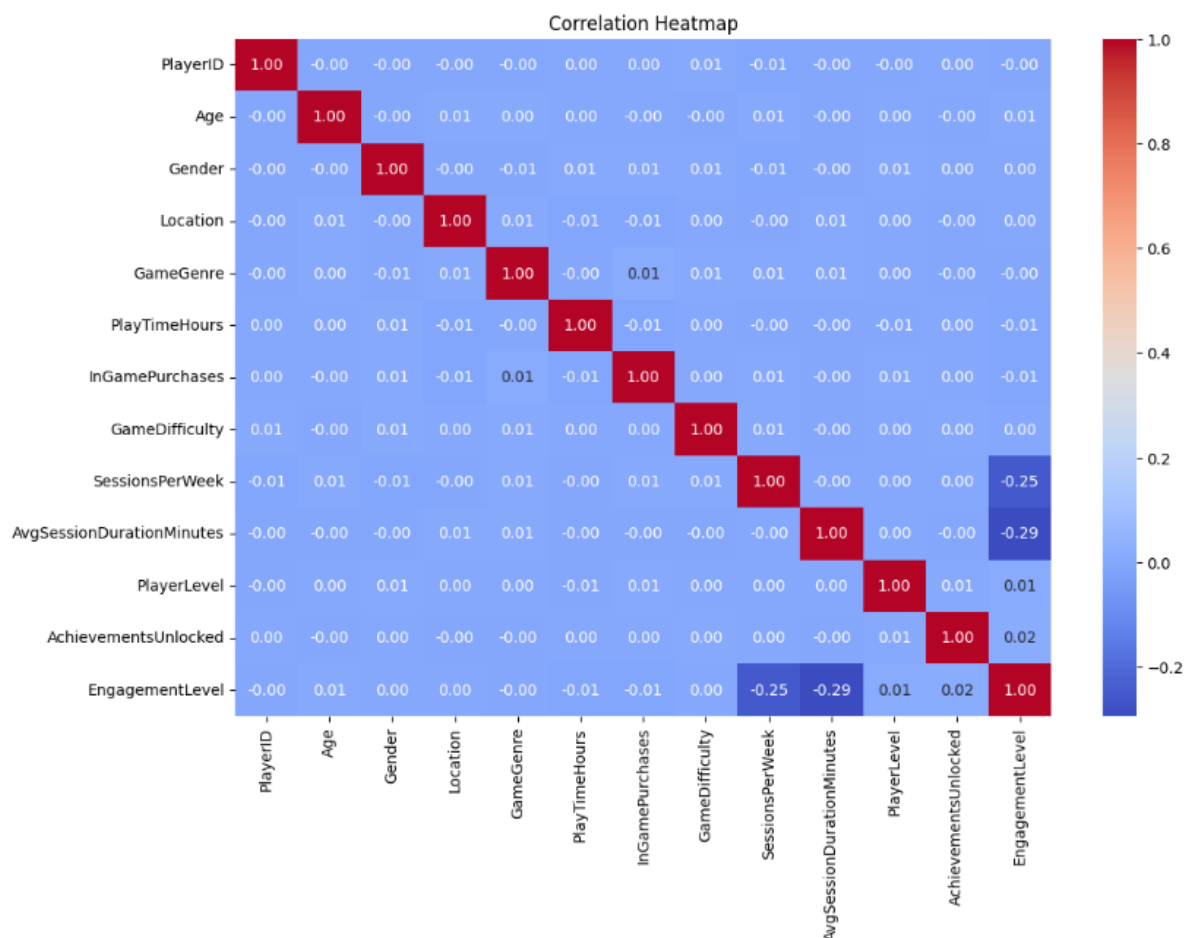
- StandardScaler dari sklearn.preprocessing digunakan untuk mengubah setiap fitur agar memiliki rata-rata = 0 dan standar deviasi = 1.
- fit\_transform() dilakukan pada data latih, sedangkan transform() pada data uji, untuk menghindari kebocoran data (*data leakage*)

## 2.2 Analisis Korelasi

Setelah preprocessing dan normalisasi selesai, dilakukan analisis korelasi untuk mengetahui hubungan antar variabel numerik dalam dataset. Korelasi diukur menggunakan metode Pearson, yang memberikan nilai antara -1 dan 1. Nilai mendekati 1 menunjukkan korelasi positif yang kuat, sementara nilai mendekati -1 menunjukkan korelasi negatif yang kuat.

Berikut adalah kode untuk visualisasi korelasi menggunakan heatmap:

```
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```





### Interpretasi Heatmap Korelasi:

- Nilai diagonal (1.00) menunjukkan korelasi variabel terhadap dirinya sendiri.
- Sebagian besar variabel memiliki korelasi rendah satu sama lain (nilai mendekati 0), yang menandakan hubungan linear yang lemah.
- Korelasi yang terlihat sedikit lebih signifikan (namun tetap kecil) adalah:
  - a. SessionsPerWeek dengan AvgSessionDurationMinutes (sekitar -0.25)
  - b. SessionsPerWeek dengan PlayerLevel (sekitar -0.29)

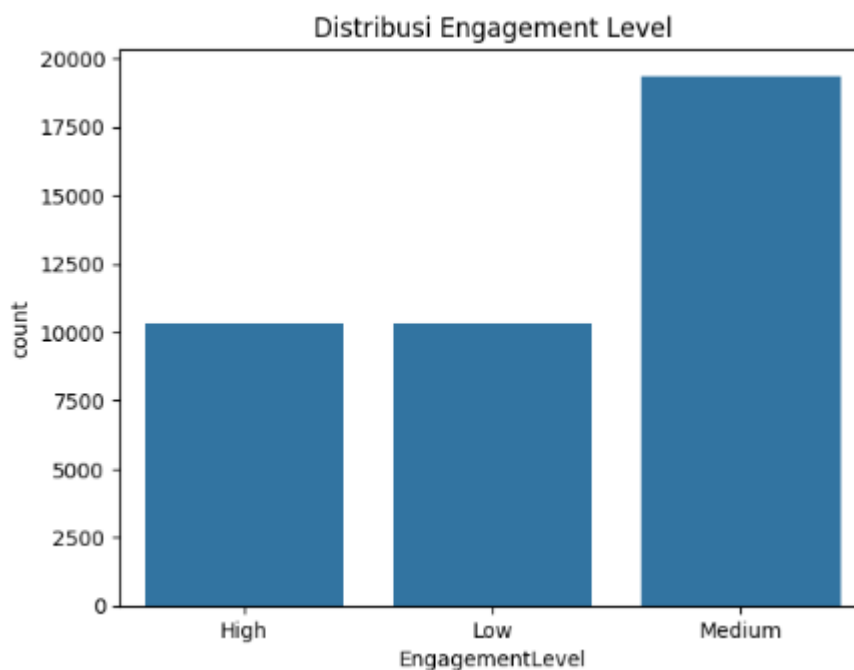
Namun, secara umum, tidak ditemukan korelasi yang sangat kuat dalam dataset ini, sehingga mungkin diperlukan teknik analisis tambahan untuk menemukan hubungan non-linear atau menggunakan algoritma yang lebih kompleks dari regresi linier biasa.

## 1. Eksplorasi dan Visualisasi Data

Setelah preprocessing dan normalisasi selesai dilakukan, langkah berikutnya adalah mengeksplorasi data target (Engagement Level) guna memahami sebaran atau distribusi dari kategori tingkat keterlibatan pemain. Proses ini dilakukan dengan membuat visualisasi berupa diagram batang dan diagram lingkaran (pie chart).

### a. Diagram Distribusi Kategori

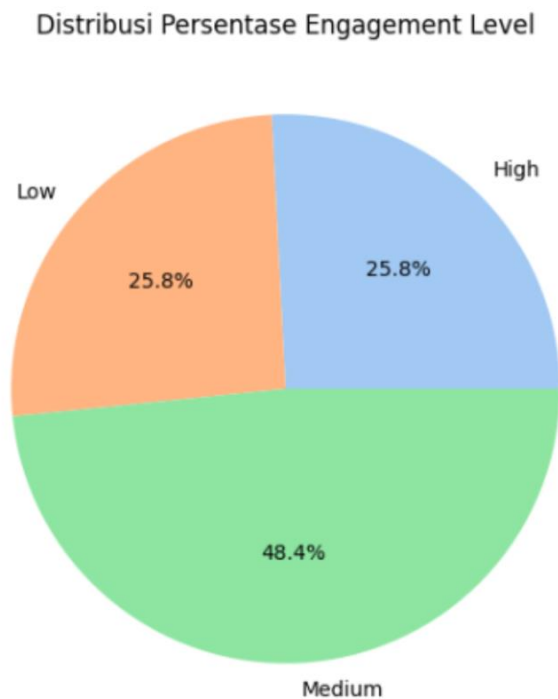
Visualisasi pertama menggunakan count plot dari Seaborn untuk menampilkan jumlah pemain pada masing-masing kategori Engagement Level (Low, Medium, High).



Penjelasan:

- Kategori Medium memiliki jumlah pemain terbanyak, mendekati 50% dari total data.
- Low dan High masing-masing memiliki jumlah pemain yang hampir sama, sekitar 25%.
- Ini menunjukkan bahwa mayoritas pemain memiliki tingkat keterlibatan sedang terhadap game

b. Diagram Lingkaran Persentase (Pie)



Visualisasi kedua dibuat untuk menampilkan distribusi persentase dari masing-masing kategori Engagement Level.

Hasil Pie Chart:

- Low: 25.8%
- Medium: 48.4%
- High: 25.8%

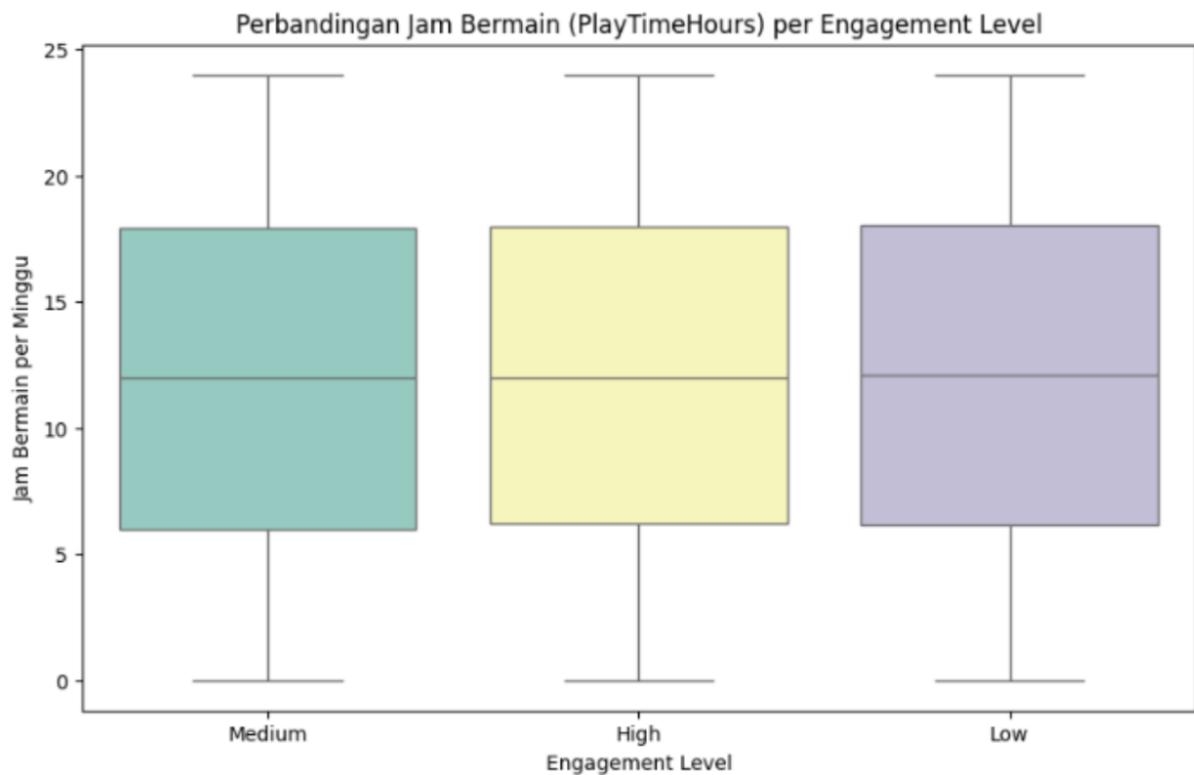
Interpretasi:

- Data memiliki distribusi yang relatif seimbang, meskipun kelas "Medium" mendominasi.
- Hal ini penting diperhatikan karena distribusi target yang tidak seimbang bisa mempengaruhi kinerja model klasifikasi.

## 2. Visualisasi Lanjutan: Analisis Fitur terhadap Engagement Level

Untuk lebih memahami karakteristik pemain pada setiap kategori **Engagement Level** (Low, Medium, dan High), dilakukan beberapa visualisasi lanjutan yang berfokus pada fitur numerik seperti PlayTimeHours, SessionsPerWeek, dan PlayerLevel.

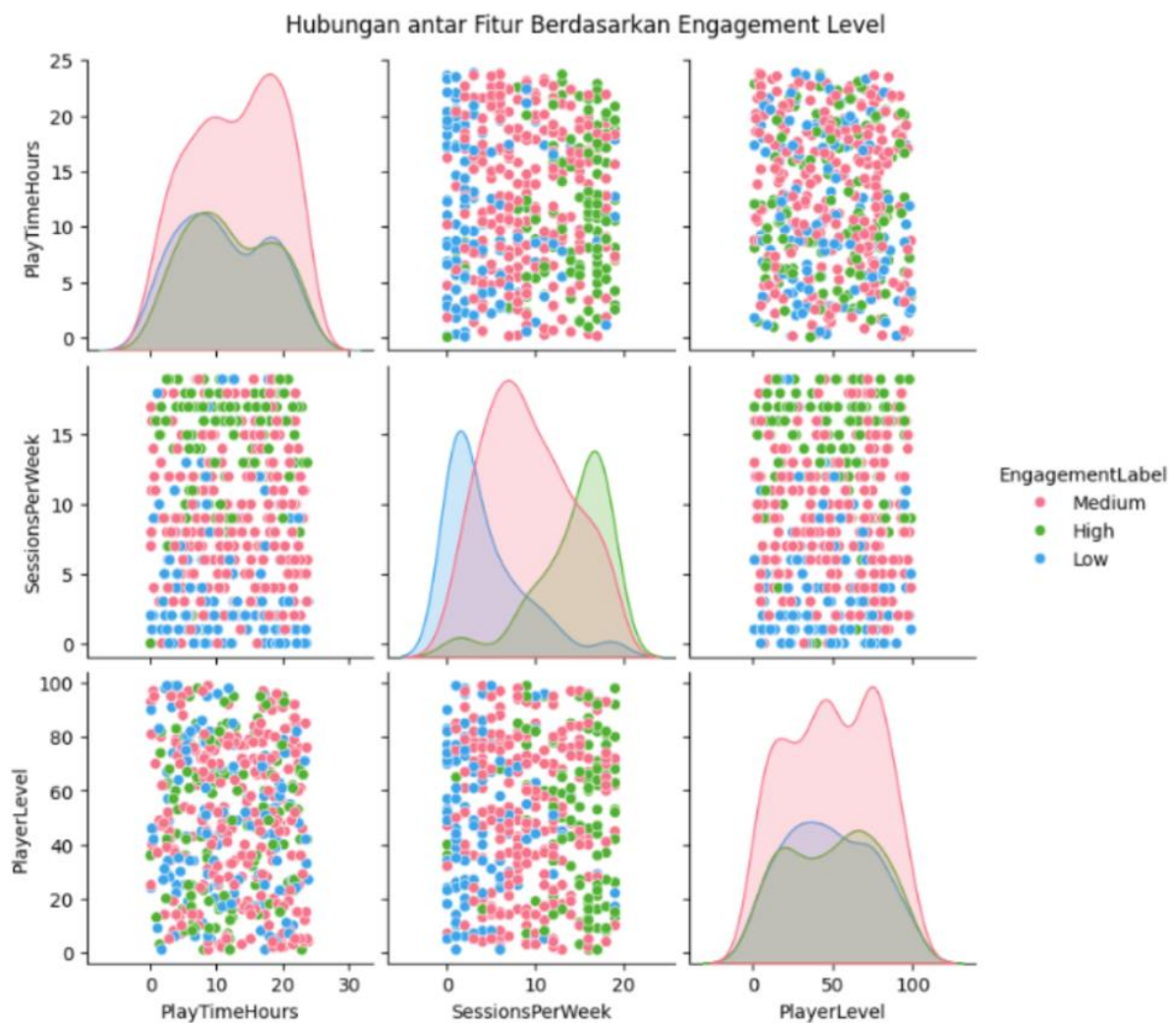
### a. Boxplot: Perbandingan Jam Bermain per Engagement Level



Penjelasan:

- Boxplot memperlihatkan distribusi jam bermain mingguan (PlayTimeHours) pada setiap level keterlibatan.
- Median jam bermain cenderung lebih tinggi pada kategori High dan Medium dibandingkan Low.
- Terdapat penyebaran nilai yang cukup lebar dan outlier di semua kategori, menunjukkan variabilitas waktu bermain yang tinggi antar pemain.

b. Pairplot: Hubungan Antar Fitur Berdasarkan Engagement Level



Penjelasan:

- Pairplot menampilkan hubungan dua variabel sekaligus untuk tiga fitur utama (PlayTimeHours, SessionsPerWeek, dan PlayerLevel) terhadap EngagementLabel.

Tampak bahwa:

- Pemain dengan EngagementLevel tinggi cenderung memiliki PlayTimeHours dan SessionsPerWeek lebih tinggi.
- PlayerLevel juga menunjukkan pola yang mirip—kategori *Medium* dan *High* cenderung berada pada level yang lebih tinggi.
- Namun, terdapat overlap cukup besar antar kelas, yang menunjukkan bahwa perbedaan antar kelas tidak sepenuhnya tegas (masih terjadi tumpang tindih).

### Kesimpulan Visualisasi:

- Kategori EngagementLevel memiliki sebaran yang dapat dibedakan berdasarkan beberapa fitur numerik, meskipun tidak terlalu ekstrem.
- Fitur PlayTimeHours, SessionsPerWeek, dan PlayerLevel memiliki potensi signifikan dalam memprediksi tingkat keterlibatan pemain.
- Visualisasi ini membantu dalam pemilihan fitur yang relevan dan mendukung proses pembuatan model klasifikasi selanjutnya.

## 2.3 Model Regresi dan Evaluasi

- A. Pemilihan dan Penyesuaian Model. Untuk melakukan klasifikasi tingkat keterlibatan pemain game (EngagementLevel), digunakan algoritma Random Forest Classifier. Algoritma ini cocok digunakan untuk masalah klasifikasi karena mampu menangani dataset dengan fitur numerik dan kategorikal, serta tahan terhadap overfitting pada dataset besar.

```
param_grid = {
    'n_estimators': [100, 200],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 5],
    'min_samples_leaf': [1, 2]
}
grid_search = GridSearchCV(RandomForestClassifier(random_state=42), param_grid,
                           cv=3, scoring='accuracy', verbose=1, n_jobs=-1)
grid_search.fit(X_train_scaled, y_train)
print('Best Parameters:', grid_search.best_params_)
best_model = grid_search.best_estimator_
```

### Penjelasan:

- GridSearchCV digunakan untuk mencari kombinasi parameter terbaik berdasarkan skor akurasi melalui validasi silang (cross-validation).

Parameter yang dicoba antara lain:

- n\_estimators: jumlah pohon dalam hutan,
- max\_depth: kedalaman maksimum pohon,
- min\_samples\_split dan min\_samples\_leaf: untuk mengontrol pembentukan cabang dalam pohon.

## B. Evaluasi Model

```
y_pred = best_model.predict(X_test_scaled)
print('Classification Report:\n', classification_report(y_test, y_pred))
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Greens')
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

```
Classification Report:
              precision    recall  f1-score   support

     0       0.92      0.87      0.90      2035
     1       0.91      0.88      0.89      2093
     2       0.90      0.95      0.93      3879

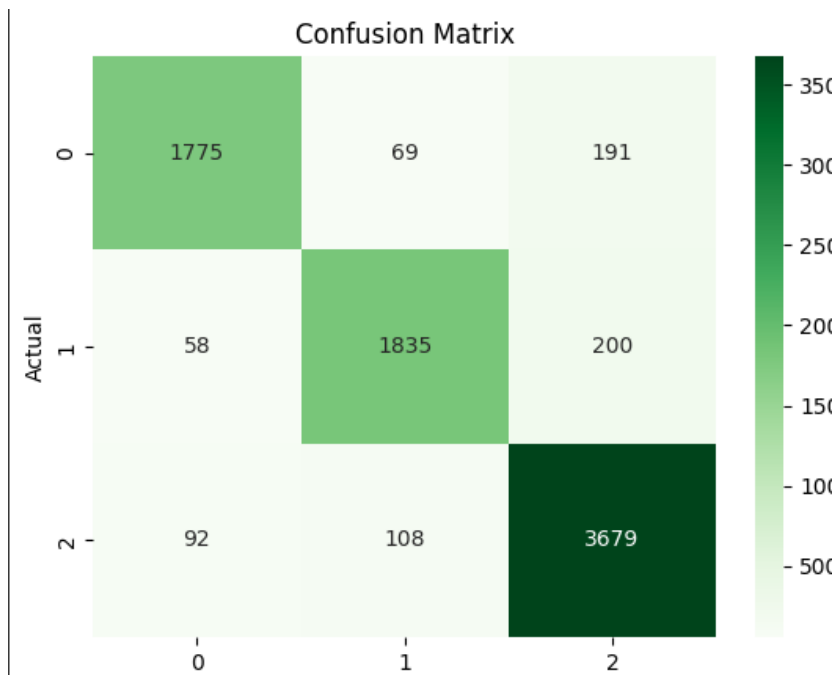
 accuracy      0.91
macro avg      0.91      0.90      0.91      8007
weighted avg   0.91      0.91      0.91      8007
```

Classification Report :

- Accuracy: 91%
- Macro average F1: 0.91 → Rata-rata dari ketiga kelas secara seimbang.
- Weighted average F1: 0.91 → Dipengaruhi jumlah data pada masing-masing kelas.

### C. Confusion Matrix

menunjukkan prediksi benar dan salah untuk setiap kelas. Dari heatmap terlihat bahwa model memprediksi mayoritas data dengan benar.



### Kesimpulan Evaluasi Model:

- Model Random Forest yang dibangun mampu melakukan klasifikasi dengan akurasi tinggi (91%).
- Semua kelas memiliki skor F1 di atas 0.89, menunjukkan performa yang stabil.
- Kelas High memiliki recall tertinggi (0.95), artinya model sangat baik mengenali pemain dengan engagement tinggi.
- Evaluasi menggunakan confusion matrix dan classification report memberikan keyakinan bahwa model cukup handal digunakan untuk klasifikasi tingkat keterlibatan pemain game.

## **BAB III**

### **HASIL DAN PEMBAHASAN**

#### **3.1 Eksplorasi Data dan Visualisasi**

Dataset yang digunakan berisi informasi terkait perilaku pemain game, seperti waktu bermain, pembelian dalam game, tingkat kesulitan game, jumlah sesi per minggu, dan tingkat pemain. Data ini kemudian dianalisis untuk memahami distribusi dan hubungan antar fitur dengan tingkat keterlibatan pemain (Engagement Level).

##### **a. Distribusi Engagement Level**

Visualisasi batang dan pie chart menunjukkan bahwa kelas "Medium" memiliki jumlah terbanyak, diikuti oleh "Low" dan "High". Distribusi ini penting untuk diketahui karena model klasifikasi perlu menghindari bias terhadap kelas mayoritas.

##### **b. Korelasi Antar Fitur**

Peta korelasi (heatmap) menunjukkan bahwa korelasi antar fitur cukup lemah, dengan beberapa fitur memiliki korelasi negatif terhadap EngagementLevel. Hal ini menunjukkan bahwa hubungan antar fitur cukup kompleks dan membutuhkan model non-linear seperti Random Forest untuk menangkap polanya.

##### **c. Visualisasi Perbandingan**

- Boxplot memperlihatkan bahwa jam bermain mingguan (PlayTimeHours) memiliki sebaran yang berbeda untuk tiap level keterlibatan. Pemain dengan engagement level "Medium" dan "High" cenderung memiliki waktu bermain yang lebih tinggi.
- Pairplot memberikan gambaran visual hubungan antara PlayTimeHours, SessionsPerWeek, dan PlayerLevel terhadap EngagementLevel. Terlihat pola yang berbeda antara kategori, mengindikasikan bahwa klasifikasi dapat dilakukan dengan cukup baik.

#### **3.2 Pembuatan Model dan Evaluasi**

##### **a. Preprocessing**

- Data kategorikal seperti Gender, Location, dan Game Genre diubah menjadi nilai numerik menggunakan Label Encoding.
- Data kemudian dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20.
- Proses normalisasi dilakukan dengan StandardScaler agar setiap fitur berada dalam skala yang sama.



#### b. Pemodelan

Model yang digunakan adalah Random Forest Classifier, yang dikombinasikan dengan GridSearchCV untuk melakukan pencarian hyperparameter terbaik.

Hasil pencarian menunjukkan bahwa kombinasi terbaik adalah:

- n\_estimators: 200
- max\_depth: None
- min\_samples\_split: 2
- min\_samples\_leaf: 1

#### c. Evaluasi Model

Hasil evaluasi dengan data uji menunjukkan bahwa model berhasil mengklasifikasikan tingkat keterlibatan pemain dengan baik:

- Akurasi: 91%
- Precision dan Recall pada ketiga kelas (Low, Medium, High) relatif seimbang.
- Confusion Matrix mengonfirmasi bahwa kesalahan klasifikasi cukup rendah, terutama pada kelas “High”.

### 3.3 Pembahasan Hasil

Model yang dibangun mampu melakukan klasifikasi dengan akurasi yang tinggi dan stabil pada semua kelas. Hal ini menunjukkan bahwa fitur-fitur yang digunakan cukup relevan untuk memprediksi tingkat keterlibatan pemain. Analisis visual juga memperkuat bahwa pola perilaku pemain berbeda tergantung pada engagement level-nya. Namun, perlu diperhatikan bahwa distribusi kelas yang tidak seimbang dapat mempengaruhi hasil model jika diterapkan ke data baru. Oleh karena itu, penggunaan teknik tambahan seperti oversampling atau SMOTE dapat dipertimbangkan di masa depan.

## **BAB IV**

### **KESIMPULAN**

Berdasarkan hasil yang telah diperoleh dari proses analisis dan klasifikasi tingkat keterlibatan pemain game, dapat disimpulkan hal-hal berikut:

1. Eksplorasi data menunjukkan bahwa pemain dengan engagement level “High” cenderung memiliki waktu bermain dan tingkat pemain yang lebih tinggi.
  2. Proses preprocessing dan encoding fitur kategorikal berhasil mengubah data menjadi bentuk yang dapat diolah oleh algoritma pembelajaran mesin.
  3. Model Random Forest Classifier memberikan hasil klasifikasi yang sangat baik dengan akurasi sebesar 91%, serta metrik evaluasi lainnya seperti precision dan recall yang konsisten.
  4. Fitur-fitur seperti PlayTimeHours, SessionsPerWeek, dan PlayerLevel terbukti memiliki pengaruh signifikan terhadap EngagementLevel.
- **Saran**
    - Untuk peningkatan performa di masa depan, dapat dicoba model lain seperti XGBoost atau LightGBM.
    - Penggunaan teknik balancing kelas dapat diterapkan jika ditemukan ketidakseimbangan lebih parah pada data baru.
    - Evaluasi dapat diperluas dengan teknik validasi tambahan seperti cross-validation dan analisis error.