**Project Title:** Game Behavior Detector
**Section:** 01

**Report on**
**Game Behavior Project**

**Submitted to:**

Meghana Chauhan
Mentor

**Submitted by:**

| Student Name | Internship ID |
|---|---|
| Mohammed Danish | INT2026-2551 |

**Date:** Feb 19, 2026

# Table of Contents

# Abstract

The exponential growth of online gaming platforms over the last decade has transformed digital entertainment into a dominant global industry. While gaming offers cognitive stimulation, stress relief, and social interaction, excessive and uncontrolled gaming behavior can lead to digital addiction, negatively affecting mental health, academic performance, professional productivity, and social relationships. Identifying early signs of gaming addiction through data-driven analysis has become increasingly important in the field of digital mental health.

This internship project focuses on the development of a Machine Learning-based classification system to detect and categorize gaming engagement levels. The dataset used in this study, titled *online_gaming_behavior_dataset*, was sourced from Kaggle and consists of 40,000 records and 13 behavioral and demographic attributes. These attributes capture various aspects of user gaming activity, including playtime duration, session frequency, achievements unlocked, and in-game purchases.

A Random Forest Classifier was implemented to classify users into three engagement categories: Low, Medium, and High. The model underwent preprocessing steps such as categorical encoding, feature selection, and dataset splitting before training. After evaluation, the model achieved an overall accuracy of 91%, demonstrating strong predictive capability and generalization performance.

This project belongs to the Healthcare domain, specifically under Digital Mental Health Analytics, and demonstrates how Artificial Intelligence can be applied to detect behavioral addiction patterns effectively. The findings highlight the importance of session frequency and duration as key indicators of high engagement levels.

# Introduction

The rapid expansion of the gaming industry, driven by advancements in internet connectivity, mobile devices, and immersive gameplay technologies, has significantly increased user engagement worldwide. Online gaming is no longer limited to casual entertainment but has evolved into competitive esports, multiplayer communities, and virtual economies.

Despite its benefits, excessive gaming has been linked to behavioral addiction. The World Health Organization (WHO) has recognized Gaming Disorder as a condition characterized by impaired control over gaming, increasing priority given to gaming over other activities, and continuation of gaming despite negative consequences.

In this context, data analytics and Machine Learning techniques offer an effective approach to detect patterns associated with excessive gaming behavior. By analyzing structured gaming behavior data, predictive models can identify high-risk individuals and enable early intervention.

The primary aim of this project is to build a supervised Machine Learning model capable of classifying users based on their engagement level using behavioral indicators. This project was carried out as part of an internship at Cepialabs, where practical implementation of AI models in real-world problem scenarios was emphasized.

The project falls under the domain of Healthcare Analytics, particularly Digital Addiction Monitoring and Behavioral Data Analysis.

# Problem Statement

With increasing accessibility to online gaming platforms, monitoring excessive engagement has become challenging. Gaming addiction develops gradually and is often unnoticed until it begins affecting mental health, productivity, and social functioning.

The main problem addressed in this project is:

To develop a Machine Learning-based classification model that can accurately predict gaming engagement levels (Low, Medium, High) using behavioral and demographic features.

The system must:

• Handle a large dataset efficiently

• Perform accurate classification

• Identify key contributing behavioral factors

• Generalize well on unseen data

The solution should be scalable and adaptable for real-world digital monitoring systems.

# Objectives

The objectives of this internship project are:

1. To explore and analyze a large-scale gaming behavior dataset.
2. To preprocess and prepare the dataset for Machine Learning modeling.
3. To implement a supervised classification algorithm.
4. To evaluate model performance using standard evaluation metrics.
5. To identify key features influencing high engagement levels.
6. To understand the practical application of Machine Learning in healthcare analytics.

# Dataset Description

The dataset used in this project was obtained from Kaggle and is titled *online_gaming_behavior_dataset*. It consists of 40,000 records and 13 columns, representing various demographic and behavioral features of gamers.

**Dataset Characteristics:**

• Number of Rows: 40,000
• Number of Columns: 13
• Target Variable: EngagementLevel

**Feature Categories:**

Demographic Features:
- Age
- Gender
- Location

Behavioral Features:
- PlayTimeHours
- SessionsPerWeek
- AvgSessionDurationMinutes
- PlayerLevel
- AchievementsUnlocked
- InGamePurchases
- GameGenre
- GameDifficulty

The target variable, EngagementLevel, is categorized into three classes:

• Low
• Medium
• High

These classes represent the intensity of user engagement in gaming activities.
Before model training, the PlayerID column was removed as it did not contribute to predictive modeling.

# Methodology

## 5.1 Data Preprocessing

Data preprocessing is a critical step in Machine Learning model development. The following preprocessing steps were performed:

• Removal of irrelevant identifier column (PlayerID)
• Encoding of categorical variables using one-hot encoding
• Encoding of target variable using LabelEncoder
• Splitting of dataset into training and testing sets (80% training, 20% testing)
• Setting random_state = 42 to ensure reproducibility

These steps ensured that the dataset was structured and suitable for training the classification model.

---

## 5.2 Model Selection – Random Forest Classifier

The Random Forest algorithm was selected due to its advantages:

• Handles both numerical and categorical data effectively
• Reduces overfitting through ensemble learning
• Provides feature importance analysis
• Performs well on large datasets
The model was configured with:
• n_estimators = 300
• random_state = 42
• class_weight = 'balanced'

The class_weight parameter ensures fair learning across all engagement classes.

# Model Implementation

The implementation was carried out in Python using the following libraries:

• pandas – Data manipulation

• numpy – Numerical operations

• scikit-learn – Machine Learning modeling

After preprocessing, the dataset was split into X (features) and y (target).
The Random Forest model was trained using the training dataset and evaluated on the testing dataset.

Evaluation metrics used:

• Accuracy Score

• Classification Report

• Confusion Matrix

• Feature Importance

# Result Analysis

The trained model achieved:

• Training Accuracy: Approximately 91%

• Testing Accuracy: Approximately 91%

The close similarity between training and testing accuracy indicates minimal overfitting and good generalization capability.

### Feature Importance Findings

The most influential features identified were:

1. SessionsPerWeek
2. AvgSessionDurationMinutes
3. PlayTimeHours
4. PlayerLevel
5. AchievementsUnlocked

This indicates that frequency and duration of gaming sessions are the strongest predictors of high engagement levels.

The results demonstrate that behavioral metrics provide strong predictive signals for detecting gaming addiction levels.

```
print("Train Accuracy:", model.score(X_train, y_train))
print("Test Accuracy:", model.score(X_test, y_test))
```

[11] ✓ 2.5s Python

```
Train Accuracy: 1.0
Test Accuracy: 0.9084551017859374
```

```
print(confusion_matrix(y_test, y_pred))
```

[12] ✓ 0.0s Python

```
[[1775   69  191]
 [  63 1834  196]
 [ 101  113 3665]]
```

# Feature Importance

```python
importance_df = pd.DataFrame({
    'Feature': X.columns,
    'Importance': model.feature_importances_
}).sort_values(by='Importance', ascending=False)

print(importance_df.head(10))
```

[10] ✓ 0.1s Python

```
                     Feature  Importance
3             SessionsPerWeek    0.391834
4   AvgSessionDurationMinutes    0.323760
1               PlayTimeHours    0.058404
5                 PlayerLevel    0.057008
6        AchievementsUnlocked    0.051632
0                         Age    0.044077
7                 Gender_Male    0.008648
16     GameDifficulty_Medium    0.007581
10               Location_USA    0.007499
8             Location_Europe    0.007050
```

# Model Testing

Generate + Code + Markdown

```python
new_player = {
    'Age': 18,
    'Gender': 'Male',
    'Location': 'USA',
    'GameGenre': 'Action',
    'PlayTimeHours': 40,
    'InGamePurchases': 10,
    'GameDifficulty': 'Hard',
    'SessionsPerWeek': 4,
    'AvgSessionDurationMinutes': 10,
    'PlayerLevel': 95,
    'Achievements': 80
}
new_df = pd.DataFrame([new_player])
new_df = pd.get_dummies(new_df)
new_df = new_df.reindex(columns=X.columns, fill_value=0)

prediction = model.predict(new_df)
probability = model.predict_proba(new_df)

print("Predicted:", le.inverse_transform(prediction))
print("Probabilities:", probability)
```

[13] ✓ 0.0s Python

```
Predicted: ['Low']
Probabilities: [[0.08       0.74333333 0.17666667]]
```

# Conlcusion

This mini project successfully implemented a Random Forest-based classification model to detect gaming engagement levels using behavioral data. With an achieved accuracy of 91%, the model demonstrates strong predictive performance and reliability.

The project highlights the importance of session frequency and duration as primary indicators of excessive gaming behavior. It also demonstrates the practical application of Artificial Intelligence in Healthcare Analytics.

The experience gained during this internship provided valuable exposure to real-world data preprocessing, model training, evaluation, and interpretation.