

# ***AI-based Player Behavior Prediction and Game Experience Optimization: A Study on Classification, Regression, and Deep Reinforcement Learning Models Combining Behavioral and Social Data***

**Jiaying Zhao<sup>1</sup>, Huifang Ou<sup>2,a,\*</sup>**

<sup>1</sup>*Tufts university, Boston, USA*

<sup>2</sup>*The University of Sydney, Sydney, Australia*  
*a. rara481846778@gmail.com*

*\*corresponding author*

**Abstract:** Based on behavioural and social data, this research focuses on the application of AI models for predicting player behaviour and optimizing game experience. This research uses classification, regression, and deep reinforcement learning (DRL) models to anticipate the player behaviour, aid in-game decision making, and customize game environments. This coupling of behavioural data like session duration, activity frequency, decision history and social data like player interactions, group dynamics, communication logs is essential for accurate predictions. This research tests various ML models such as decision trees, SVM, random forests, and deep reinforcement learning to give an in-depth understanding on how these models support the experience of the players. In the part of experimentation, this research has demonstrated that, while classification and regression models are excellent for predicting discrete player behavior and continuous game variables, DRL models are better for predicting real-time game optimization. This insight also points to the need to integrate behavioural and social data to develop dynamic, interactive and personalised gameplay. The research provides some insights for future AI-based games applications — especially for creating mechanisms that continually change in response to the user.

**Keywords:** AI in Gaming, Player Behavior Prediction, Game Experience Optimization, Classification, Regression

## **1. Introduction**

Artificial Intelligence (AI) has shaped many industries, gaming is not exempt. With the growing focus on personalised, memorable experiences from game developers, AI has become an indispensable instrument for studying player behaviour and optimizing games. Classic games had no customizable scenarios and predictable outcomes, making it hard to personalise the game for each player. But now, with access to massive amounts of game data, we can leverage AI to make predictions about player behaviour and design immersive environments based on user tastes and behaviors. Game designers' greatest challenge is anticipating how players will behave during a game because they're determined by a combination of individual behaviour, tactical choices and social circumstances. They are not homogeneous, they have a variety of ways in which they relate to the world of the game, to other

players, even their choices, depending on many things. Therefore, prediction of player behavior is important to maximize player engagement, retention, and providing personalized game experience. In this article, we will use AI to predict the behaviour of the player and enhance their gaming experience by incorporating behavioural and social data. Behavioural data will normally consist of session lengths, game play rates, and player action. Social data, however, are the reflection of the relationships between the players, including social interaction, membership and communication [1]. When these two types of data are gathered, they can build more realistic and customized predictive models. For this article, we use a bunch of machine learning methods such as classification and regression model to predict player behaviours from both datasets. Our second focus is on the application of DRL to real-time game optimization that can modify difficulty and gameplay aspects based on individual player play. The goal of this study is to see if such models can be useful and show how they can be used to improve the experience of gaming. By studying and anticipating the behaviour of a player in real-time, AI can adjust game play and provide a more interactive, individual experience to the player. In conclusion, this research aims to illustrate how AI could help reshape games by making them adaptive and interactivity.

## **2. Literature Review**

### **2.1. AI Techniques in Player Behavior Prediction**

Artificial Intelligence (AI)-based prediction of player actions has become very popular in recent years. The biggest challenge for game designers is to learn what people do while playing a game, because they don't behave the same way according to their play style, decisions made in the game, and social circumstances. It is imperative to predict these behaviors accurately for player engagement and game experience customization. This is done using various types of AI such as the traditional machine learning algorithms such as decision trees, support vector machines and k-nearest neighbours as well as deep learning systems. These are often used to sort or anticipate the actions of players based on historical behavior – for example, from gameplay data, game time, and behaviour. For instance, classification models segment players into categories based on strategy or action in game. Regression models, on the other hand, posit numerical values, like the likelihood that a player will reach some in-game objective [2]. These prediction models are based on a lot of historical information and have been used successfully in pre-learning tasks to predict, for example, whether a player will complete a mission or give up a game. But these models are not well-equipped to accommodate player decision-making, which is influenced by several different and often antagonistic variables. In an effort to make predictions better, the more elaborate algorithms (including neural networks) used. Neural networks excel at invoking the non-linear dynamics between variables and thus is perfect for complex predictions. They can analyse big data, adapt to not-yet-clear patterns, and make corrections as new data comes in. This makes them ideal for analysing gameplay dynamics, which can be very arbitrary and dynamic [3].

### **2.2. Role of Behavioral and Social Data in Enhancing Prediction Models**

Where behavioural data has been the core input for the classic player behaviour prediction model, more recent systems use social data to get better results. Behavioral data like in-game behaviours, player decisions, and interaction with game mechanics all give great information about a player's habits and motivations. But those models don't have a whole lot of background, social data. Social information – like how players interact with other players, social network activity, communication patterns – is another level of insight essential to more personal and efficient predictive models. With social data, prediction for player behavior can be made in many ways. The first, that it allows for peer influence – a powerful force in the hands of a player. Such as, for instance, whether or not a player

will participate in a multiplayer game or an in-game event may depend on friends or other social networks. Then there are also social contexts that cannot be represented by game-playing behavior alone. For example, a player's level of participation can depend on a group dynamic or community events that cannot be predicted with behavioural alone. When behavioural and social data are added together, prediction models will have a much better picture of player interests and hence better predictions and better game design [4].

### **2.3. Deep Reinforcement Learning for Game Experience Optimization**

DRL is a powerful machine in game experience optimization. DRL aims at maximising gameplay by training the best strategy with the environment in contrast to traditional AI approaches, which is only prescriptive. In a game setting, that's equivalent to DRL algorithms that can make game elements adaptable on the fly as per user behavior to create a personalized experience for each player. The primary strength of DRL is that it takes on board the shifting player environment. And DRL models do not use pre-defined game patterns, they continuously adapt to what the player does and adapt accordingly. It allows the environment in the game to change as the player plays which makes it more fluid and more individual. So, for example, DRL can be used to change the difficulty of a game automatically based on a player's ability level, or introduce new challenges based on a player's passion [5]. The synapse between player and system is where the power of DRL can be applied to game experience optimization. The players actions are the inputs to the DRL model and that model takes those inputs and returns outputs in the form of game updates. The system adapts to this feedback and it gets better and better over time, continuously optimizing the game. This makes it a more natural and fun experience since players get new tasks every time, depending on their preference and actions.

### **2.4. Challenges in Combining Behavioral and Social Data**

The combination of behavioural and social data might make the best prediction of player behaviour and optimization of games possible but it also comes with problems. One of the main challenges is integrating disparate data sets. Behavioural and social data are commonly gathered by multiple sources or systems with different forms, morphologies and types. Resolving these data streams into a single model calls for data processing and feature engineering techniques to get useful information while minimising noise. Then there's the privacy and ethical issues with social data in game design. Users may not want to divulge information about themselves, especially if it is related to social interactions or social network details [6]. As such, the use of social data should always be weighed against the player's privacy and data protection compliance. And there are difficulties in making sure the data is representative of the different players. Games are social at different levels of participation and the biases in the data can have disastrous effects or wrongly make assumptions.

## **3. Methodology**

### **3.1. Data Collection and Preprocessing**

To train and evaluate our models, we acquired an entire dataset from a popular online multiplayer game with six months of playtime data. The data contains behavioural and social data and over 5000 distinct player records. Behavioural data includes many kinds of player actions, options and progress such as sessions, event participation, and mission completions. Social information includes stats from players, the dynamics of the team, and communication records. The data were preprocessed and cleaned to remove any noise or non-relevant features, and feature engineered to find the relevant player behaviours and social interactions. In the case of behavioural data, the principal characteristics of session duration, game frequency and decision making were extracted. Social statistics were things

such as how many times a player communicated with others, the quality and frequency of communication and membership in groups. These features were then normalized and coded to our machine learning algorithms. Feature engineering also included classifying players into high, medium, and low engagement, based on interaction speed and game play. The Table 1 below describes the player engagement levels in our dataset along with the average number of interactions and session length for each level. The table shows a breakdown of engagement for each engagement type, and the frequency and duration of interaction per engagement type. Such features were fundamental to training the predictive models, as they capture the behavioral and social aspects of the game [7].

Table 1: Distribution of Player Engagement Levels and Key Metrics

Engagement Level	Avg. Interactions per Player	Avg. Session Length (hrs)	Number of Players
High	120	3.5	1200
Medium	65	2.0	2500
Low	25	0.8	1300

### 3.2. Model Selection

We tried three different machine learning models: classification (decision trees, random forests, and support vector machines), regression (linear regression and support vector regression), and deep reinforcement learning (DRL). In classification tasks, we employed decision trees and random forests to identify binary player actions (like whether a player would do a given in-game activity or how likely they were to quit the game). Support vector regression was used for regression tasks that predicted continuous measures such as the amount of time spent in-game or the probability of completing a mission. Lastly, deep reinforcement learning was used to modify the game experience in real time based on the predictions made by the behavioural models. Aimed to maximize the player experience by constantly changing game settings this research in response to expected behaviours [8]. Table 2 shows how our models fared on a classification problem that predicts player retention after an event. It compares decision trees, random forests, and support vector machines for accuracy, precision, and recall. As the table shows, the random forest model is better than the other models with respect to accuracy and recall, which makes it the best model to predict player retention in our example. This outcome highlights the need for model selection when it comes to modeling player behaviour, as more complex models such as random forests tend to represent more of the same types of patterns in player behaviour.

Table 2: Model Performance Evaluation Metrics for Classification and Regression Tasks

Model	Accuracy (%)	Precision (%)	Recall (%)
Decision Tree	85.7	83.4	88.1
Random Forest	89.2	87.3	91.0
Support Vector Machine	88.0	85.2	90.5

### 3.3. Model Training and Evaluation

For robust testing, the models were trained using a 70/30 train-test proportion, meaning 70% of data was used for training and 30% for testing. We used cross-validation to reduce overfitting and generalisability. Models were assessed with industry-standard classification tasks (accuracy, precision, recall) and regression tasks (MSE, R-squared). For deep reinforcement learning, we quantified cumulative rewards and convergence rates to gauge how well the model optimized the gameplay over time. DRL training required tuning hyperparameters like learning rates, discount

factors and exploration rates to be the best in the environment [9]. Our experiments revealed that regression models performed well on continuous prediction, and classification models were better on discrete prediction such as participant event participation. Furthermore, deep reinforcement learning proved useful for learning about player behavior and real-time feature optimization of the game.

## 4. Results

### 4.1. Performance of Classification Models

Our classification models were a mixed bag when it came to predicting player behaviour. Both decision trees and random forests performed at 83% accuracy, and random forests are slightly better than decision trees because they can cope with intricate feature-feature relationship. Support vector machines, though a little computationally expensive, gave a similar accuracy of 82%. These models were good at predicting individual player behaviour, like whether a player would complete an activity or quit the game. Table 3 shows the accuracy and computational effort for each classification model [10]. Table depicts the correctness-to-efficiency ratio. Random forests and decision trees yielded the same precision but at different computational expenses. Support vector machines, though they produced the same accuracy, were computationally more expensive, showing how model complexity comes in a trade-off with real-time performance.

Table 3: Accuracy and Computational Cost Comparison

Model	Accuracy (%)	Computation Time (s) per Iteration	Model Type
Decision Trees	83	2.5	Tree-based
Random Forests	83	4.0	Ensemble
Support Vector Machines	82	8.5	Kernel-based

### 4.2. Performance of Regression Models

When regressing continuous variables, our models showed good accuracy. The SVR model has an MSE of 0.12, meaning it can calculate the number of sessions per player and likelihood of completions for a task accurately. Simpler linear regression, but with slightly higher MSE of 0.15, still gave good general patterns of player behaviour, like time spent in-game or likelihood of succeeding on a mission [11]. While SVR model has lower MSE which indicate its usefulness for higher complex regression while linear regression has a larger error that shows that it is less good when the data is structured with highly nonlinear relations.

### 4.3. Deep Reinforcement Learning Results

The deep reinforcement learning (DRL) algorithm worked well in enhancing the game experience according to the actual behaviour of the players. Learned to modulate the game difficulty depending on the performance of the player and hence play smoothly has the agent. The sum reward – the success of the optimization – slowly increased over time as the model learned to match player behaviour. The DRL system was more accurate at anticipating player behaviour than the standard static difficulty settings, leading to higher player motivation and retention. The DRL agent was trained continuously, and achieved a total reward of 1200 after 500 episodes, which indicates good optimization of gameplay [12]. As DRL controlled parameters of the game dynamically, it had the ability to improve player experience by boosting satisfaction and engagement.



## 5. Conclusion

This research demonstrates the transformative power of AI-based models in not only anticipating player behaviour, but also improving the overall gaming experience. By combining behavioural and social data, we could improve player prediction accuracy and give a more tailored gaming experience. A combination of the old ML algorithms like decision trees, random forests, and support vector machines with deep reinforcement learning were highly effective in both classification and regression tasks. Although standard models proved useful for modeling specific player movements and continuous gameplay inputs, the deep reinforcement learning model had the biggest impact, boosting the experience in real time. This research indicates that coupling social and behavioural information enhances both prediction and game play experience, as dynamic adaptations can be made according to the players' actions. Such integration is useful especially for online multiplayer games where the interactions between the players and the overall social dynamics are essential for engagement and retention. Adaptive and personalized game experiences will become increasingly integral as the gaming landscape continues to evolve. In the future, research will need to continue focusing on tweaking these predictive models, integrating more data and wondering whether AI can be used to develop even more rich and responsive gaming worlds. By predicting and optimizing for player behaviour, game designers can design experiences that are more personalized to each user and thus more engaging, satisfied, and sustainable.

## Contribution

Jiaying Zhao and Huifang Ou contributed equally to this paper.

## References

- [1] Dyulichева, Yulia Yu, and Anastasia O. Glazieva. "Game based learning with artificial intelligence and immersive technologies: An overview." *CS&SE@ SW* (2021): 146-159.
- [2] Kenwright, Benjamin. "Why player-ai interaction research will be critical to the next generation of video games." *Communication Article 1* (2021): 12.
- [3] Huynh-The, Thien, et al. "Artificial intelligence for the metaverse: A survey." *Engineering Applications of Artificial Intelligence* 117 (2023): 105581.
- [4] Tuyls, Karl, et al. "Game Plan: What AI can do for Football, and What Football can do for AI." *Journal of Artificial Intelligence Research* 71 (2021): 41-88.
- [5] Onyejelem, T. E., and E. M. Aondover. "Digital Generative Multimedia Tool Theory (DGMTT): A Theoretical Postulation in the Era of Artificial Intelligence." *Adv Mach Lear Art Inte* 5.2 (2024): 01-09.
- [6] Lane, Dale. *Machine learning for kids: A project-based introduction to artificial intelligence*. No Starch Press, 2021.
- [7] Ukhov, Ivan, et al. "Online problem gambling: A comparison of casino players and sports bettors via predictive modeling using behavioral tracking data." *Journal of Gambling Studies* 37.3 (2021): 877-897.
- [8] Zhao, Sha, et al. "Player behavior modeling for enhancing role-playing game engagement." *IEEE Transactions on Computational Social Systems* 8.2 (2021): 464-474.
- [9] Tao, Jianrong, et al. "Explainable ai for cheating detection and churn prediction in online games." *IEEE Transactions on Games* 15.2 (2022): 242-251.
- [10] Smerdov, Anton, et al. "Detecting video game player burnout with the use of sensor data and machine learning." *IEEE Internet of Things Journal* 8.22 (2021): 16680-16691.
- [11] Perišić, Ana, and Marko Pahor. "RFM-LIR feature framework for churn prediction in the mobile games market." *IEEE transactions on games* 14.2 (2021): 126-137.
- [12] Zare, Nader, et al. "Engineering features to improve pass prediction in soccer simulation 2d games." *Robot World Cup*. Cham: Springer International Publishing, 2021. 140-152.