

International Journal of Data Mining, Modelling and Management

ISSN online: 1759-1171 - ISSN print: 1759-1163

<https://www.inderscience.com/ijdmmm>

Ensemble learning models for predicting the gaming addiction behaviours of adolescents

Nongyao Nai-arun, Warachanan Choothong

DOI: [10.1504/IJDMMM.2025.10065942](https://doi.org/10.1504/IJDMMM.2025.10065942)

Article History:

Received:	14 December 2023
Last revised:	28 February 2024
Accepted:	17 May 2024
Published online:	25 February 2025

Ensemble learning models for predicting the gaming addiction behaviours of adolescents

Nongyao Nai-arun

Faculty of Science and Technology,
Phranakhon Si Ayutthaya Rajabhat University,
Phranakhon Si Ayutthaya, Thailand
Email: nnongyao@aru.ac.th

Warachanan Choothong*

Faculty of Science and Technology,
Nakhon Sawan Rajabhat University,
Nakhon Sawan, Thailand
Email: warachanan@nsru.ac.th

*Corresponding author

Abstract: This paper proposes: 1) to create a prediction model for the game addiction of adolescents using six data mining algorithms; 2) to optimise the models by adjusting the parameters; 3) to create an ensemble model. Bagging and boosting algorithms were investigated for improving the models. Data were collected from eight Northern Rajabhat Universities in Thailand. The results found that bagging with neural network had shown the highest performance with an accuracy of 99.35%, followed by the boosting with neural network (99.02%), the model with the best-optimised parameters of the neural network algorithm achieved by adjusting the learning rate. The best model was used to develop a web application for predicting the gaming addiction behaviours of adolescents, which would contribute to solve the problem.

Keywords: classification; ensemble learning; bagging; boosting; neural network; random forest; optimisation; gaming addiction behaviours.

Reference to this paper should be made as follows: Nai-arun, N. and Choothong, W. (2025) 'Ensemble learning models for predicting the gaming addiction behaviours of adolescents', *Int. J. Data Mining, Modelling and Management*, Vol. 17, No. 1, pp.103–125.

Biographical notes: Nongyao Nai-arun received her BEd degree in Computer Education from Nakhon Ratchasima Teacher College, Nakhon Ratchasima, Thailand in 1992, and MSc degree in Technology of Information System Management from Mahidol University, Bangkok, Thailand in 2000. She has a PhD in Computer Science from Naresuan University, Phitsanulok in 2014. She is currently an Assistant Professor in the Department of Information Technology, Faculty of Science and Technology, Phranakhon Si Ayutthaya Rajabhat University. Her research interests include data science, data mining, forecasting, artificial intelligence and machine learning.

Warachanan Choothong received her BSc degree in Computer Science from Nakhon Sawan Rajabhat Institute, Nakhon Sawan, Thailand in 2003, and MSc degree in Information Technology from Naresuan University, Phitsanulok,

Thailand in 2007. She is currently a Lecturer in the Department of Computer and Information Technology, Faculty of Science and Technology, Nakhon Sawan Rajabhat University. Her research interests include data mining, information technology and artificial intelligence.

1 Introduction

In 2018, the World Health Organisation (WHO) made the decision to officially incorporate symptoms of gaming disorder, also known as game addiction, in the 11th Edition of the diagnostic manual (The International Classification of Diseases: ICD). Gaming disorder has been classified in the Mental Disorders category of substance abuse and behavioural addiction. The symptoms associated with gaming addiction closely resemble those of alcohol, substance abuse, or gambling. The individuals exhibiting such behaviours are regarded as abnormal patients. The diagnosis of game addiction will involve the evaluation of abnormal gaming behaviours that have a negative impact on daily life, such as excessively long gaming sessions that continue for at least 12 months; the loss of contact with family, friends, and society; and the display of uncontrollable behaviours that have various negative effects, including affecting work and life responsibilities (World Health Organisation, 2020). In the USA, the US Psychiatric Association has proposed adding abnormal gaming behaviour to the diagnostic and Statistical Manual of Mental Disorders: DSM-5 (WebMD Editorial Contributors, 2021; Wang et al., 2019). According to a report from the US Psychological Association, about 160 million US adults play video games, with players ranging in age from under 18 to over 50 (Sherer, 2023; Richter, 2022). The market research firm Newzoo estimated that the global video game industry would be worth over \$196.8 billion in US dollars in 2025, with nearly 3.2 billion gamers worldwide in 2022 (Newzoo International B.V., 2022).

Based on the relevant research study conducted by Lee et al. (2018), an analysis of the behavioural characteristics of smartphone addiction was carried out using data mining. Ulkhaq et al. (2018) employed the game addiction scale (GAS) questionnaire as a tool to examine game addiction among players in Indonesia. Duan et al. (2021) developed a screening tool for smartphone addiction using a decision tree model during the COVID-19 pandemic. Kim and Lee (2020) presented a method for analysing and categorising the diagnosis and prediction of smartphone usage problems by comparing the important variables affecting the study using data mining algorithms, such as the Decision Tree, Random Forest, and the Support Vector methods. Nusawat and Kwangsawad (2020) developed a model to analyse the level of risk of Nomophobia, which considers smartphone usage behaviours by using data mining techniques, such as decision tree and Naive Bayes. Hafeez et al. (2017) developed a diagnostic algorithm to identify the attributes of psycho-physiological game addiction using statistical parameters. Cuzzocrea et al. (2023) used neural network algorithms to separately analyse the behaviours of human players from those of game bots. Finally, Sun et al. (2022) diagnosed internet addiction using EEG signals and neural networks. From a review of related research, it was found that there had been a few studies in Thailand, which had focused on the development of models and web applications for predicting adolescent gaming behaviour.

In 2019, Thailand had a population of approximately 27.5 million gamers. According to a global consumer survey conducted in October, gamers in Thailand were between the ages of 18 and 64. Thailand's gaming industry has been growing rapidly and consistently, with a value of 22,000 million baht in 2019, representing a growth of 13% from 2018 (Singsakda, 2020). The survey found that gaming addiction had not just been limited to children with behavioural issues, but it had also been prevalent among adults in Asia. Notably, the concern is still the abnormal gaming behaviours among children and students (Department of Mental Health, 2020). The child and adolescent mental health Institute, Rajanagarindra Institute, Department of Mental Health, and the Ministry of Public Health of Thailand, in collaboration with Asst. Prof. Chanvit Pornnoppadol of the Department of Child and Adolescent Psychiatry at the Faculty of Medicine Siriraj Hospital, developed a test to identify those children and students who may have problems with computer gaming addiction or excessive gaming. The term 'game' in the test refers to video games, Play Station, Gameboy, computer games, mobile phone games, and all types of online games (Pornnoppadol et al., 2014).

Research conducted by the Association of Radio and Media for Children and Youth in Thailand, which focused on Thai children's gaming behaviours, social responsibilities, and policy proposals, indicated that gaming affects their behaviours and their health. The National Health Commission reported that one of three of the sample group, who had played games in the past 12 months, reported playing every day for an average of five hours per day. Some of them spent more than 5,000 baht per month on gaming. Additionally, more than 40% of children who played games acknowledged that someone had told them they were addicted to gaming and nearly 20% had experienced intense emotions, including quarrelling with family members, becoming more abusive (National Health Commission Office (NHCO), n.d.). Sinsomboonthong (2018) conducted a comparative analysis of the effectiveness of predicting video game addiction among children and adolescents in Bangkok by using seven classification methods. Moreover, Inplang and Thongkam (2020) conducted a study on the sentiment analysis of player unknown's battlegrounds mobile games using text mining. The objective of this research was to explore classification techniques from five data mining techniques. Panich et al. (2021) presented an article on the topic of using the neural network technique to predict grades in the web database subject. In order to enhance the model's efficiency, they adjusted various parameters of the neural network model, including the training cycle, the learning rate, and the momentum.

According to a study of related research, various algorithms were utilised in experiments, including decision tree, random forest, support vector methods, K-nearest neighbours, Naive Bayes, neural network, and ensemble learning. Furthermore, some studies experimented with parameter tuning, such as adjusting other parameters of the neural network model, including the following:

- 1 the learning rate, which determines how much to change the weights at each step
- 2 the training cycle, which is the number of rounds in learning
- 3 the momentum, which involves increasing the proportion of the previous weights update in order to prevent extreme values in any part (not the actual maximum)
- 4 the continuity of finding the best value, resulting in a more efficient model.

Thus, the research team chose the aforementioned algorithms for experimentation since they had demonstrated accurate results.

Based on the aforementioned reasons, the researchers had the idea of studying the gaming behaviours of students by using the game addiction screening test (GAST) (Pornnoppadol et al., 2014) as a testing tool to collect data and study the gaming behaviours of students. The data was then used to create predictive models. The researchers also improved the performance of the models by adjusting the various parameter values of each model in order to obtain new knowledge and the most efficient model. This model could be used in the development of an application that predicts the gaming behaviour of students in Thailand. The results of this research will provide insights into the causes of gaming behaviours, which can be used to improve regulations, such as setting gaming time limits, controlling game characteristics, affecting brain and behaviour development, and developing an awareness of the negative effects of gaming and playing games appropriately. Additionally, this would help to create monitoring and support systems that are appropriate for those who have gaming addiction problems. All of these efforts would promote the growth of quality Thai children and youth, who could assist in developing Thailand in the future.

2 Materials and methods

2.1 The game addiction screening test

The GAST (Pornnoppadol et al., 2014) has a test structure that consists of 16 questions, which measures problems related to playing games in three areas:

- 1 pre-occupation with games
- 2 loss of control over playing games
- 3 functional impairment.

Respondents were able to choose from four answer options for each question as follows.

- 1 'never,' indicated that the respondent was 100% confident that he or she had never exhibited the described behaviours
- 2 'rarely,' indicated that the respondent was more than 50%, but less than 100%, confident that he or she had never exhibited the behaviour
- 3 'sometimes,' indicated that the respondent was more than 50%, but less than 100%, confident that he or she had exhibited the behaviour
- 4 'always,' indicated that the respondent was 100% confident that he or she had exhibited the described behaviours.

2.2 Method

The research methodology according to the principles of data mining was divided into three steps: data pre-processing, modelling, and model evaluation. The researchers designed the research framework process, as shown in Figure 1.

Figure 1 The research framework process

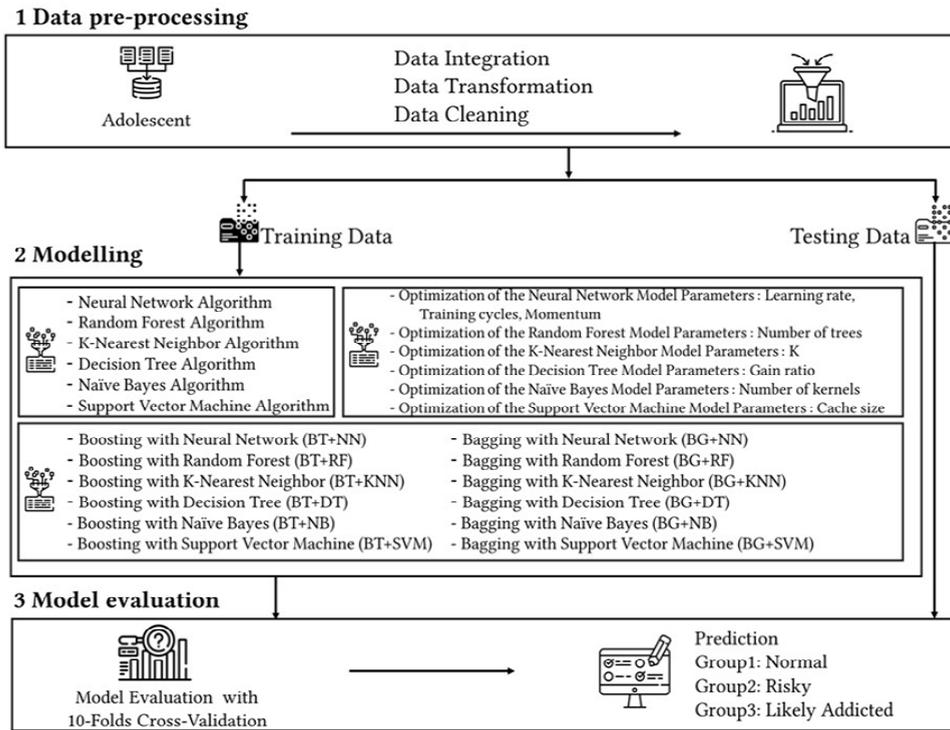


Figure 1 shows the process of the research framework, which consisted of the three following steps: data pre-processing, modelling, and model evaluation.

2.2.1 Data pre-processing

This process was applied in three steps:

- 1 Data integration: The researchers collected data from eight Rajabhat universities in Northern Thailand, consisting of Kamphaeng Phet Rajabhat University, Chiang Rai Rajabhat University, Chiang Mai Rajabhat University, Nakhon Sawan Rajabhat University, Phibun Songkhram Rajabhat University, Phetchabun Rajabhat University, Lampang Rajabhat University, and Uttaradit Rajabhat University. Using stratified sampling with the sample size calculated using equation (1) (Public Organisation), n.d.) as follows.

$$n = \frac{Nk^2\sigma_x^2}{k^2\sigma_x^2 + NE^2} \tag{1}$$

where $k = 1.96$, $\sigma_x^2 = 1$, $E = 0.05$, and the sample size was $n = 3,418$ people.

- 2 Data transformation: This step involved altering the general status questionnaire responses of the respondents into variables that were suitable for data mining, which resulted in 11 variables. Additionally, the adolescent’s game addiction test responses were transformed by converting the answer options into the following scores:

- Not at all = 0 points
- Rarely = 1 point
- Sometimes = 2 points
- Frequently = 3 points.

Once all the items had been scored, the scores were summed within each column for rarely, sometimes, and frequently. The total scores were then calculated and divided into three groups: normal, risky, and likely addicted. This was based on the criteria outlined in the gaming disorder test manual by the child and adolescent mental Health Rajanakarindra Institute at the Department of Mental Health at the Ministry of Public Health in Thailand. Pornnoppadol et al. (2014) as shown in Table 1.

Table 1 The groupings of the game addiction test

<i>Total scores from the game addiction test for females</i>	<i>Total scores from the game addiction test for males</i>	<i>The severity of the problem</i>	<i>Groups</i>
Scores less than 16	Scores less than 24	No problem with playing games	Normal
Scores between 16–22	Scores between 24–32	Starting to have problems with playing games	Risky
Scores greater than or equal to 23	Scores greater than or equal to 33	Having significant problems with playing games	Likely addicted

- 3 Data cleaning: In this case, the dataset of 3,418 individuals was cleaned to remove missing values, incomplete data, and data with missing feature values, which, following the exclusion of 348 individuals, resulted in a clean dataset of 3,070 individuals. The dataset was divided into two sets: the training dataset, which comprised 70% of the total data with 2,149 individuals (Class 1: 724, Class 2: 710, and Class 3: 715) for model development, and the testing dataset, which comprised 30% of the total data with 921 individuals (Class 1: 306, Class 2: 310, and Class 3: 305) for model evaluation. The clean dataset was composed of 10 input variables and 1 output variable, as shown in Table 2.

Table 2 Input and output variables

<i>No</i>	<i>Variables</i>	<i>Description</i>	<i>Values</i>	
1	Sex	Gender	M	Male
			F	Female
2	Faculty	Faculties or schools	Fc01	Faculty of Education
			Fc02	Faculty of Management Sciences
			Fc03	Faculty of Science and Technology
			Fc04	Faculty of Agriculture
			Fc05	Faculty of Food and Agricultural Technology
			Fc06	Faculty of Agricultural Technology
			Fc07	Faculty of Industrial Technology

Table 2 Input and output variables (continued)

<i>No</i>	<i>Variables</i>	<i>Description</i>	<i>Values</i>	
2	Faculty	Faculties or schools	Fc08	Faculty of Humanities and Social Sciences
			Fc09	Faculty of Agricultural Technology and Industrial Technology
			Fc010	School of Law
			Fc011	School of Social Sciences
			Fc012	School of Accounting
			Fc013	School of Public administration
			Fc014	School of Health Sciences
			Fc015	School of Computer and Information Technology
3	Age	Age (year)	Numeric data	
4	Income	Monthly family income (baht)	Numeric data	
5	Computer	Personal computer or laptop ownership	Y	Yes
			N	No
6	Place	The most frequently played game location	PH	House
			PD	Dormitory
			PS	Game shop
7	Type	The type of game played	VG	A video game
			PS	A play station game
			GB	The game boy
			CG	Computer game
			MG	An application game on mobile phone
			OG	An online game
8	Often	The frequency of playing games in a week	OFT01	1–2 days/week
			OFT02	3–4 days/week
			OFT03	5–6 days/week
			OFT04	Everyday
9	Weekday	Duration of online gaming from Monday to Friday (hour)	Numeric data	
10	Weekend	The duration of online gaming on Saturdays and Sundays (hour)	Numeric data	
11	Class		Group 1	Normal
			Group 2	Risky
			Group 3	Likely addicted

2.2.2 Modelling

This research created the models using RapidMiner Studio 9 software for data mining (RapidMiner Studio, 2019) as follows.

2.2.2.1 Creating 6 models using the data mining algorithm

- 1 Neural Network model: Neural network model was used to classify the data using feed-forward network architecture (Aggarwal, 2018; Beale et al., 2010). The model consisted of three layers of nodes: the input layer, the hidden layer, and the output layer.

Firstly, the input layer consisted of 36 nodes that received data from 11 variables, which were divided into two types: numeric and nominal. The numeric variables include age, income, weekday, and weekend. The values of these variables with a variety of scales were transformed using max-min normalisation to a scale ranging from 0 to 1. The nominal variables included sex, faculty, computer, place, type, and often. The values of these variables were converted into numeric values using nominal to numeric conversion.

Secondly, the hidden layer consisted of 19 nodes, which was computed by adding the number of input nodes (36) to the output nodes (3) and dividing by 2 (as shown in equation (2)). The values of the nodes in the hidden layer were computed by using the sigmoid activation function as shown in equation (3) and Figure 2.

Lastly, the output layer consisted of 3 nodes that represented groups or classes: the normal group, the risky group, and the potential game-addiction group. The model used these nodes to make predictions. As shown in Figure 2.

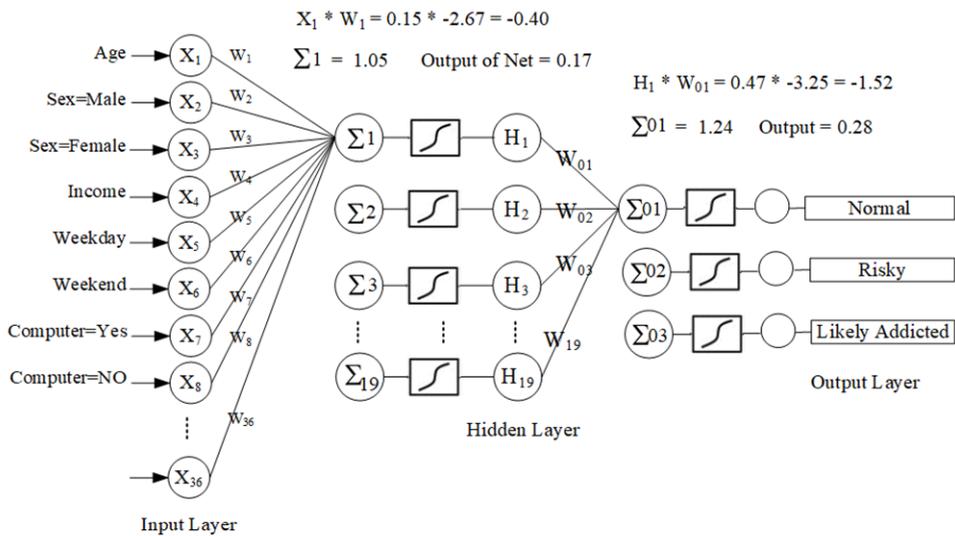
$$net_i = \sum_{j=1}^n w_{ij}x_j \quad (2)$$

where net_i is the sum of node i in the hidden layer, w is a vector of the weight value, x is a vector of the f input variable, i is the 19 nodes of hidden layer, n is the 36 nodes of the input layer and j is potential values from 1 to 36 (Nai-arun and Moungrai, 2020).

$$f(net_i) = \frac{1}{1 + e^{-net_i}} \quad (3)$$

where $f(net_i)$ is the Activation Function, e is 2.71835, net_i is the result from equation (2), i is the 19 hidden layers

Figure 2 The neural network model



2 The random forest model, developed by Breiman (2001), Ali et al. (2012), is a modelling approach that is capable of improving the accuracy of classification tasks. The algorithm begins by constructing an ensemble of trees, with each tree casting a vote for the class (as depicted in Figure 3). This presents a method for building the Random Forest model using data (D) and input variables (V). In this article, the real dataset consisted of 2,149 individuals and 10 input variables. Moreover, n denoted the number of sample groups, in which d_i and v_i represented the number of data and variables in group i, with i ranging from 1 to n. Each sample group underwent the following steps:

Data D_i , in which d_i did not exceed D, was randomly selected from D

Variables V_i , in which v_i did not exceed V, were randomly chosen from V

Trees grew and prediction classes were assigned. After iterating steps (1) to (3) for n times, these trees collectively formed a forest.

Classification was then performed by selecting the class with the highest aggregate score from all trees in the forest. After this process, all data had to be returned to the dataset before selecting a new sample group, as illustrated in Figure 3.

3 K-nearest neighbours: This algorithm uses the Euclidean distance measurement method (as in equation (4)) (Han et al., 2011; Tan et al., 2020; Ye, 2003). The model creation process consisted of five steps:

determining the value of K

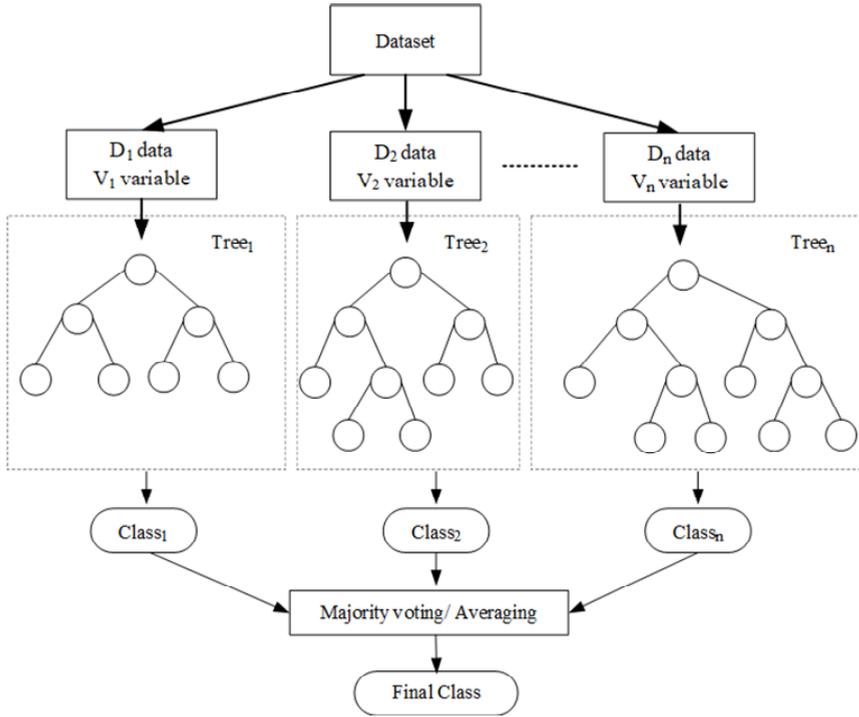
calculating the distance between the data being evaluated and the sample group

ranking the distances

selecting the nearest K neighbour data and counting the majority vote results, and

assigning a class to the data being evaluated.

Figure 3 The random forest model



The researchers created the model by setting the K values from 1 to 10, and then comparing the accuracy values to find the most suitable K value. The experimental results showed that when K had been set to 8, the highest accuracy was at 92.14%.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{4}$$

where x_i and y_i represent the parameters at the i^{th} instance of data x and y , respectively, while n represents the number of features in each instance.

- 4 The Naïve Bayes model: The principle of Bayes’ theorem was used to consider the probability of the data and to predict the target group value by considering the maximum probability value among all the possible groups. The calculation of the probability value of the numeric variables was calculated from the means and standard deviations (as shown in equation (5) below). The calculation of the probability value of the nominal variables was calculated from the frequency (as shown in equation (6) below) of all 3 possible groups: the normal group, the risky group, and the likely game-addicted group. Then the new test data was considered by selecting the highest probability value calculated and by determining the maximum value of $P(A|B)$ among all possible groups (Tan et al., 2020; Ye, 2003).

$$P(x | y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

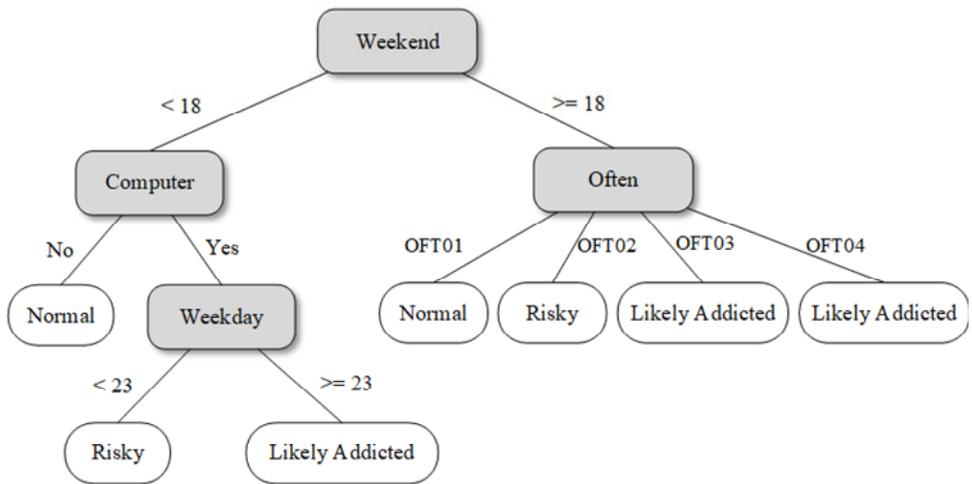
$$P(x|y) = \frac{n(x \cap y)}{n(y)} \tag{6}$$

- 5 The decision tree model: The process of building a model with the decision tree algorithm involved a recursive procedure, which allowed for the backward computation or repetition of previous steps. To create a decision tree, the root node must be considered as the first step before proceeding to consider the leaves and branches that break off. The calculation of the appropriate data to become the root node is based on the highest information gain obtained from the calculation of entropy (as shown in equations (7) and (8)) in order to classify the data into the same group as much as possible. After obtaining the root node, the decision tree was then created until the leaf nodes had been reached, which resulted in a complete decision tree (Jones, 2008), as shown in the example in Figure 4.

$$E(S) = \sum_{i=1}^n -P(V_i) \log_2 P(V_i) \tag{7}$$

$$Gain(S, A) = E(S) - \sum_{v \in values(A)} \frac{|S_v|}{S} E(S_v) \tag{8}$$

Figure 4 The decision tree model



- 6 The support vector machine model (SVM): This algorithm was designed to solve binary classification problems. The principle of SVM is to create a straight-line equation that can divide data into two groups and that can attempt to create a middle boundary line with the maximum distance between the boundary edges of both groups. SVM uses a mapping function to move data from the input space to the feature space and creates Kernel function that measures similarities on the feature space. This research uses the LibSVM algorithm (Chang and Lin, n.d.), a Multi-class support vector machine method that is used to solve problems with more than two groups by employing the principle of One-against-One, which is based on the binary classification method. When classifying three groups, the model was created $3(3-1)/2$

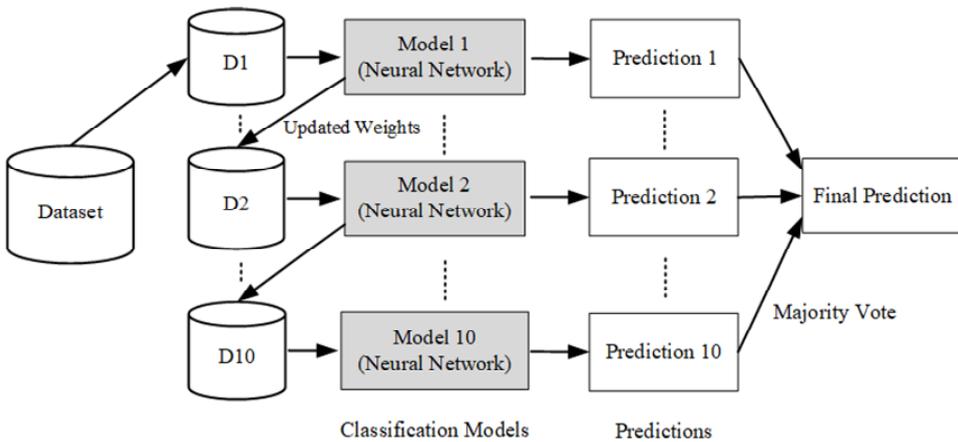
times by pairing each group as follows. (1, 2), (1, 3), (2, 3). Then the model was voted upon, and the pair with the highest value was selected.

2.2.2.2 *Creating ensemble models with data mining algorithms*

The ensemble learning: bagging and boosting algorithms were applied in this paper for improving the models (Panthong, 2023), as follows.

- 1 **Boosting:** The boosting models were created by combining the following six algorithms: boosting with neural network (BT + NN), boosting with random Forest (BT + RF), boosting with K-nearest neighbours (BT + KNN), boosting with decision tree (BT + DT), boosting with Naïve Bayes (BT + NB), and boosting with support vector machine (BT + SVM). The model creation process was composed in the following manner. Firstly, 10 sets of data (D1 ... D10) were randomly selected from a total of 2,149 individuals and 10 variables for data randomisation. Secondly, the dataset obtained from Step 1 was used to create a model with neural network algorithm. Thirdly, the probability of selecting each data instance in the dataset was adjusted in order to learn to create a model in the next round. The data that had been misclassified by the previous model had a higher chance of being selected for randomisation in the next round, while the data that had been correctly classified by the previous model had a lower chance of being selected for randomisation. This was accomplished by adjusting the updated weight or the probability of selecting the data instance for randomisation. The model created in the next round would use the data that had been misclassified by the previous models. Lastly, the final prediction was made by counting the votes and predicting the group (Auwattamongkol, 2019), as shown in Figure 5.

Figure 5 Boosting with neural network model

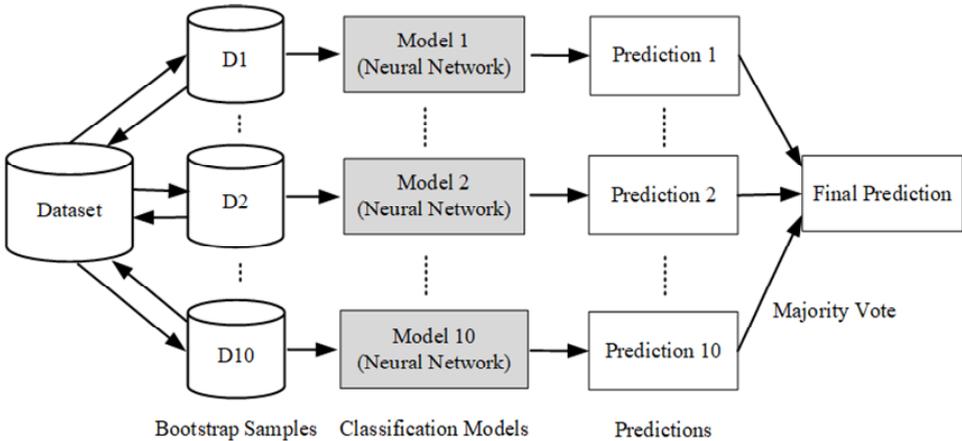


- 2 **Bagging:** A bagging model was created that combined the six algorithms: bagging with neural network (BG + NN), bagging with random forest (BG + RF), bagging with K-nearest neighbours (BG + KNN), bagging with decision tree (BG + DT), bagging with Naïve Bayes (BG + NB), and bagging with support vector machine (BG + SVM). In the process, there were 3 steps:

taking all 2,149 individuals and 10 variables and randomly sampling them into 10 sets (D1 ... D10) using the replacement method, which meant that once a set had been randomly sampled, the data in that set was returned to the original dataset creating 10 different neural network models (Model1 ... Model10) from the datasets that had been obtained in Step 1, which produced 10 different predictions (Prediction1 ... Prediction10)

taking the majority vote and predicting which group the data belonged to, and then repeating Steps 2 and 3 with the remaining basic algorithms (Breiman, 1996; Dietterich, 2000), as shown in Figure 6.

Figure 6 Bagging with neural network model



2.3 Evaluation

Evaluation of the model performance using the 10-fold cross-validation method (Han et al., 2011; Tan et al., 2020) involved dividing the dataset into K disjoint subsets, by using 10-1 subsets to create the model and by reserving 1 subset as the test data. This was repeated 10 times, with each iteration using a different 10-1 subset as the training data and a different subset as the testing data. In order to obtain accuracy, precision, and recall values, the performance of the model was evaluated by calculating the average of the evaluation results from the 10 iterations. To evaluate the performance of those models, the classification accuracy of each model is calculated. After creating the model, it will be tested using 10-fold cross-validation (as shown in equations (9)).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{9}$$

where *TP* is the number of true positives, *FP* is the number of false positives, *TN* is the number of true negative, and *FN* is the number of false negatives.

3 Results

This study consisted of the following:

- 1 the results of comparing the performance of the prediction models for gaming addiction behaviours using data mining algorithms
- 2 the results of comparing the performance of models with the adjusted parameter values for each model
- 3 the results of comparing the performance of the ensemble models using all the models combined with the data mining algorithms to predict gaming addiction behaviours, which were tested using 10-fold cross-validation and the confusion matrix tool to determine the accuracy, precision, and recall values.

The results were as follows.

3.1 *The performance comparison of models using the data mining algorithms*

The results of the performance comparison of the six data mining algorithms, are shown in Table 3.

Table 3 The performance comparison results of the six basic models

<i>Models</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>
Neural network	98.70	98.69	99.03
Random forest	98.65	98.65	98.65
Decision tree	98.56	98.55	98.55
Support vector machine	98.32	98.40	98.31
Naïve Bayes	96.18	96.20	96.18
K-nearest neighbours	92.04	92.41	92.06

Table 3 shows the results of comparing the accuracy, precision, and recall of six basic models for predicting game addiction behaviours. It was found that the model with the highest predictive performance was the neural network model with an accuracy of 98.70%, and the second-best model was the random forest model with an accuracy of 98.65%. The model having the lowest predictive performance was the K-nearest neighbours with an accuracy of 92.04%.

3.2 *Comparison of model performance improvement by adjusting each model's parameter*

Improving model performance by adjusting the parameters of each model was conducted to find the appropriate parameters that yielded higher performance. The parameters, which had been adjusted for all eight models, were:

- 1 the training cycles parameter for the neural network model
- 2 the momentum parameter for the neural network model
- 3 the learning rate parameter for the neural network model

- 4 the number of tree parameters for the random forest model
- 5 the K parameter for K-nearest neighbours model
- 6 the grain ratio parameter for the decision tree
- 7 the number of kernels parameter for the Naïve Bayes model
- 8 the cache size parameter for the SVM.

The experimental results are shown in Figure 7.

In Figure 7, the parameter values that had improved the performance of each model in predicting gaming addiction behaviours were compared. By displaying the accuracy results from highest to lowest, it was found that the neural network model with a training cycle parameter of 80 had achieved the highest accuracy of 99.02%. This was followed by the neural network model with a parameter for the momentum parameter at 0.90, which achieved an accuracy of 98.98%. The neural network model with a learning rate parameter of 0.70 had achieved the highest accuracy of 98.84%. The random forest model with a parameter for the number of trees set at 41, which achieved an accuracy of 98.79%. The decision tree model with a gain ratio parameter set at 3 achieved an accuracy of 98.76%, while the SVM with the parameter of the cache size set at 31 had achieved an accuracy of 98.52%. In addition, the Naïve Bayes model with the parameter of the number of kernels set at 70 had achieved an accuracy of 96.56%, and the K-nearest neighbours model with a K parameter set at 8 had achieved the lowest accuracy at 92.14%.

Figure 7 A comparison of the model performance results after adjusting the parameters of the eight models (see online version for colours)

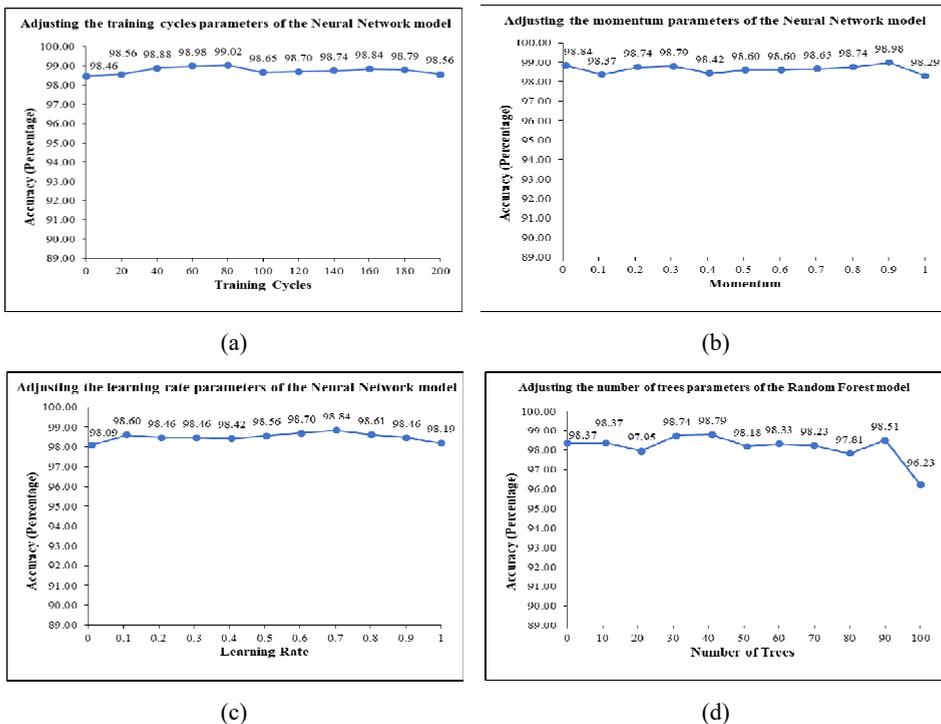
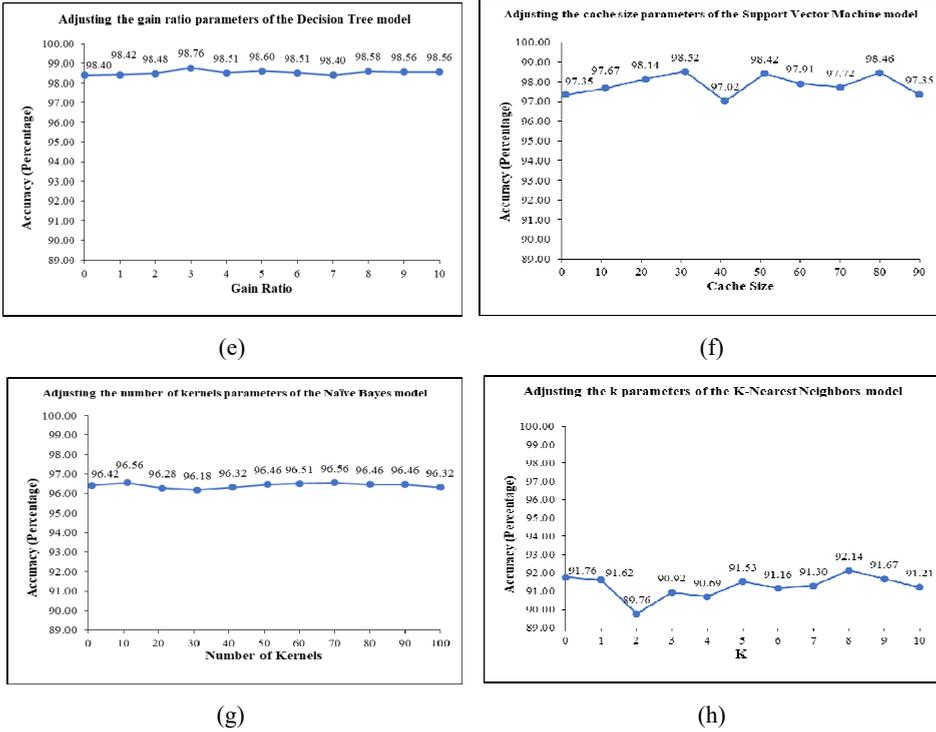


Figure 7 A comparison of the model performance results after adjusting the parameters of the eight models (continued) (see online version for colours)



3.3 A comparison of the performance of the ensemble models with the data mining algorithms

The performance comparison of the twelve ensemble models combined with data mining algorithms included boosting combined with six data mining algorithms (BT + NN, BT + RF, BT + KNN, BT + DT, BT + NB, BT + SVM) and Bagging combined with six data mining algorithms (BG + NN, BG + RF, BG + KNN, BG + DT, BG + NB, BG + SVM). The results of the comparison are shown in Table 4.

Table 4 shows the results of comparing the classification accuracy of twelve models. The top three accuracies are BG+NN, boosting with neural network (BT+NN), and bagging random forest (BG + RF) models which are 99.35%, 99.02%, and 98.98% respectively. The results also suggest that bagging and boosting algorithms improve the accuracy of neural networks and random forest. The model with the lowest prediction performance was boosting with K-nearest neighbours (BT + KNN) model, which had an accuracy of 92.09%.

Table 4 A comparison of the results of bagging and boosting performance combined with the six machine learning algorithms

<i>Models</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>
Bagging with neural network (BG+NN)	99.35	99.35	99.30
Boosting with neural network (BT+NN)	99.02	99.02	99.02
Bagging with random forest (BG+RF)	98.98	98.98	98.97
Boosting with random forest (BT+RF)	98.93	98.92	98.92
Boosting with support vector machine (BT+SVM)	98.79	98.83	98.78
Bagging with decision tree (BG+DT)	98.74	98.74	98.74
Boosting with decision tree (BT+DT)	98.60	98.61	98.59
Bagging with support vector machine (BG+SVM)	98.60	98.66	98.59
Boosting with Naïve Bayes (BT+NB)	98.56	98.54	98.54
Bagging with Naïve Bayes (BG+NB)	98.51	98.51	98.51
Bagging with K-nearest neighbours (BG+KNN)	92.69	93.14	92.72
Boosting with K-nearest neighbours (BT+KNN)	92.09	92.46	92.11

3.4 Evaluating model performance

Based on the experimental results, the top three performing models had been the BG + NN model, the boosting with neural network model, and the BG + RF model. These models were evaluated against the testing dataset, which consisted of 30% of the total data (921 individuals) and had been designated specifically for model evaluation. The performance of the models was assessed by using the values in the confusion matrix table to evaluate the prediction of gaming behaviours among students (predicted) and then comparing them with the actual values. The results are depicted in Figure 9.

Figure 9 The performance measurement results of the models (a) bagging with neural network model (b) boosting with neural network model (c) bagging with random forest model (see online version for colours)

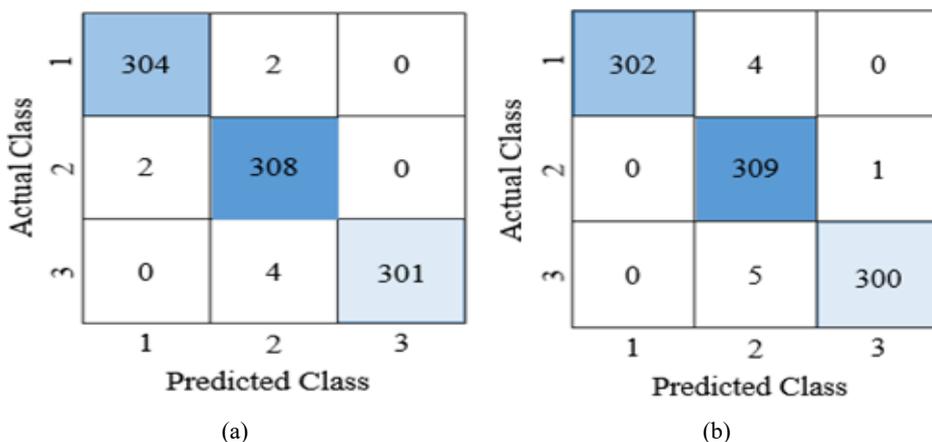


Figure 9 The performance measurement results of the models (a) bagging with neural network model (b) boosting with neural network model (c) bagging with random forest model (continued) (see online version for colours)

Actual Class	1	305	1	0
	2	4	304	2
	3	0	4	301
		1	2	3
		Predicted Class		

(c)

In Figure 9, the values of TP, FP, TN, and FN for each model’s classified data are shown. The research found that when the confusion matrix values of bagging was combined with the neural network model, it had resulted in the highest true positive value for all three groups. In other words, the data had been true, and the prediction had been correct. Based on the experimental results, it is recommended to further develop the aforementioned models into a web application that can be used to predict gaming behaviours.

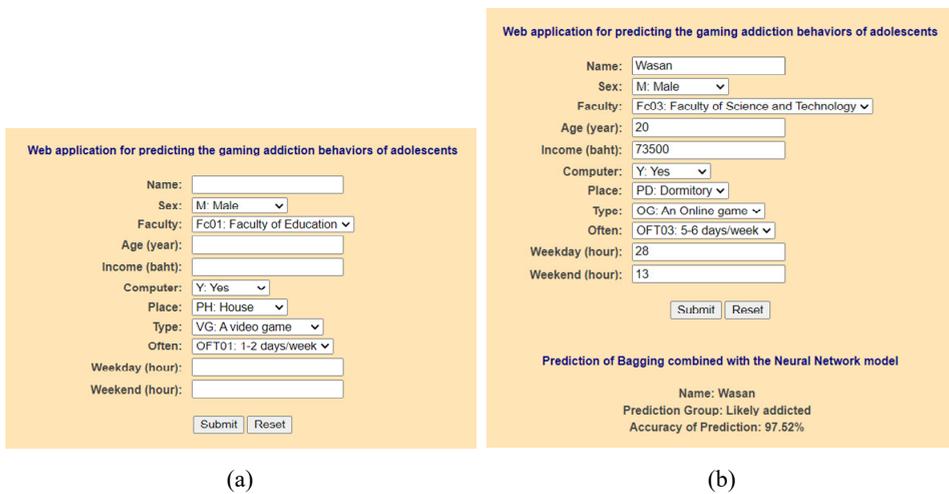
3.5 Web application

After selecting the best model, BG + NN model, we developed a web application with PHP programming and MySQL database for predicting new data entered as shown in Figure 10(a), consisting of name, sex, faculty, age(year), income(baht), computer, place, type, often, weekday(hour) and weekend(hour). In this stage, the program will evaluate the results. After that, the gaming addiction behaviour group and accuracy percentage were predicted, as shown in Figure 10(b).

Gaming addiction is a prevalent issue among adolescents, which directly affects their daily lives and potentially leads to mental health problems. While gaming can serve as a primary source of entertainment and contribute to happiness during leisure time, it can also become problematic if its usage is not appropriate. Gaming behaviour can be categorised into general gaming and gaming addiction. The distinction between these two types lies in the following: Firstly, general gamers are those individuals, who can control themselves, can appropriately allocate gaming time without disrupting their daily activities, and can ensure that gaming does not interfere with other tasks. Secondly, Gaming Addicts are those individuals, who excessively engage in gaming to the extent that gaming activities disrupt other daily activities. They may experience intrusive thoughts about gaming even when engaged in other activities, may have difficulty diverting attention away from gaming, may constantly crave gaming, and may become

irritable or restless when they are unable to play. They often prioritise gaming over essential activities, such as eating or socialising. In fact, some may even skip school or neglect sleep due to gaming. Gaming addiction, which is considered a mental health issue that requires treatment, typically arises from the brain’s pleasure response to gaming activities, leading individuals to repeatedly seek that pleasure. This addiction can significantly impact adolescents, especially considering that their brains are still developing, particularly in those areas that are related to impulse control and reasoning. Therefore, it is crucial to cultivate self-discipline and to enhance brain potential, particularly during adolescence (Arunakun, 2018).

Figure 10 Web application for predicting newly entered data (a) consisting of name, gender, faculty, age (years), income (baht), computer, location, type, frequent, weekday (hours) and weekend (hours) (b) result predict gaming behaviour groups and accuracy percentage (see online version for colours)



The findings from this study provided insights into understanding the causes of gaming behaviour. Based on Figure 10, the web application is used to predict gaming behaviour and indicates which group an individual belongs to. If they fall into the risky or likely addicted group, avenues for prevention and mitigation of the adverse effects of gaming addiction are suggested. Preventive measures for adolescent gaming addiction include fostering self-control, such as setting rules for gaming time and responsibilities before starting to play. This is a crucial aspect that should not be overlooked, since addressing the issues only after they have occurred can be challenging. Preventing and addressing gaming addiction typically involve three approaches:

- 1 Limiting gaming time: allowing 1–2 hours of gaming per day and avoiding gaming during school hours or late at night
- 2 Controlling content: ensuring that the game content is not excessively violent in order to prevent adolescents from being exposed to harmful content
- 3 Behaviour management: ensuring that gaming behaviours do not lead to inappropriate actions, such as skipping school, refusing to sleep, engaging in aggressive physical or verbal behaviour, or contemplating suicide.

4 Discussion

The objective of this research was to create models using data mining algorithms and then to compare the predictive performance of game addiction behaviours with the parameter tuning of the six models. The research found that the model created by the Bagging algorithm combined with the neural network had shown the highest predictive performance for adolescent game addiction behaviours. Following that, the boosting algorithm was combined with the neural network and the performance improvement of each model by adjusting the parameters, respectively. The reason for this was that the Bagging algorithm is a method, in which each model in the algorithm is created from a dataset, which had been randomly selected from the training dataset specified in the replacement format. As a result, multiple models were created from different datasets that were all subsets of the same original training dataset specified by majority vote, in which each model has equal weight in voting. Meanwhile, the boosting algorithm was created by randomly selecting data from the training dataset specified in the replacement format, as in bagging. However, the data was randomly sampled, and the models were created one at a time in each iteration. After each iteration, there was an adjustment of the probability of sampling each model in the training dataset to create a model in the next iteration, by adjusting the weights or probabilities of sampling the data. The model, created in the next iteration, used the data misclassified by the model that had been created in the previous iteration. The decision on the type of data used the majority vote, in which each model had a different weight value in voting. The model that had been created in the previous round had a higher weight in the vote than the model created in the following round. Therefore, bagging and boosting algorithms are algorithms that help to reduce the prediction error of the data by using multiple classification models to make decisions. Bagging aims to reduce the prediction error of the data by making the model highly fitting and complex. Boosting not only helps to reduce the prediction error in cases, in which the model is overfitting, but it also helps to reduce the prediction errors in cases, in which the model has a low specificity. Both algorithms help to improve the accuracy of these predictions. The neural network algorithm is a supervised learning method that classifies data types as a learning process from the training data to get the weight value of each connecting line. This minimises the prediction errors and enhances model performance by adjusting the learning rate parameter to obtain the most suitable and optimal value.

The research results, the best model can be recommended for use in the following areas:

- 1 in developing a web application to predict the gaming behaviours of adolescents
- 2 as a guide to
 - a finding the causes of gaming addiction
 - b improving regulations
 - c discovering the proper limitation of time for playing games
 - d exploring the characteristics of the games
- 3 in controlling the use of games that affect brain development and the behaviours of adolescent. In addition, the aims were to raise awareness of the negative consequences of game addiction and to promote appropriate gaming behaviours

among the students of the Rajabhat Universities in Thailand, so that Thai students of high quality can be developed for the country's future advancement.

The findings from this study can be beneficial for guiding the selection of classification algorithms for future use. However, there are limitations and considerations regarding future work. The dataset used in this research was relatively small, necessitating further experiments to validate these conclusions. For future studies, adjustments to prototype models may involve incorporating various methodologies. Additionally, in order to develop more efficient models, it may be beneficial to explore diverse datasets from different regions and any other factors that may be relevant to predicting gaming behaviours.

References

- Aggarwal, C.C. (2018) *Neural Networks and Deep Learning*, Springer, Cham.
- Ali, J., Khan, R., Ahmad, N. and Maqsood, I. (2012) 'Random forests and decision trees', *International Journal of Computer Science Issue*, Vol. 9, No. 5, pp.272–278.
- Arunakun, J. (2018) *Children Addicted to Games, Behavioral Problems That Need to be Addressed* [online] <https://www.rama.mahidol.ac.th/ramachannel/article> (accessed 8 February 2024).
- Auwattanamongkol, S. (2019) *Data Mining*, 2nd ed., Publisher of the National Institute of Development Administration, Bangkok.
- Beale, M.H., Hagan, M.T. and Demuth, H.B. (2010) *Neural Network Toolbox 7 User's Guide*, Math Works, Natick.
- Breiman, L. (1996) 'Bagging predictors', *Machine Learning*, Vol. 24, pp.123–140, <https://doi.org/10.1007/BF00058655>
- Breiman, L. (2001) 'Random forests', *Machine Learning*, Vol. 45, pp.5–32 [online] <https://doi.org/10.1023/A:1010933404324>
- Chang, C-C. and Lin, C-J. (n.d.) *LIBSVM -- A Library for Support Vector Machines* [online] <https://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed 15 Nov 2019).
- Cuzzocrea, A., Martinelli, F. and Mercaldo, F. (2023) 'A deep-learning approach to game bot identification via behavioural features analysis in complex massively-cooperative environments', *International Journal of Data Mining, Modelling and Management*, Vol. 15, No. 1, pp.1–29, <https://dx.doi.org/10.1504/IJDM.2023.129963>.
- Department of Mental Health (2020) *Gaming disorder, A Changing Perspective* [online] <https://dmh.go.th/news/view.asp?id=2271> (accessed 28 October 2022).
- Dietterich, T.G. (2000) 'An experimental comparison of three methods for constructing ensembles of decision trees- bagging, boosting and randomization', *Machine Learning*, Vol. 40, pp.139–157 <https://doi.org/10.1023/A:1007607513941>.
- Digital Government Development Agency (Public Organization). (n.d.) *Sampling and Estimation Techniques*, Digital Government Development Agency (Public Organization), Bangkok.
- Duan, L., He, J., Li, M., Dai, J., Zhou, Y., Lai, F. and Zhu, G. (2021) 'Based on a decision tree model for exploring the risk factors of smartphone addiction among children and adolescents in China during the COVID-19 pandemic', *Front. Psychiatry*, Vol. 12, pp.1–13, <https://doi.org/10.3389/fpsy.2021.652356>.
- Hafeez, M., Idrees, M.D. and Kim, J-Y. (2017) 'Development of a diagnostic algorithm to identify psycho-physiological game addiction attributes using statistical parameters', *IEEE Access*, Vol. 5, pp.22443–22452 [online] <https://doi.org/10.1109/ACCESS.2017.2753287>.
- Han, J, Kamber, M. and Pei, J. (2011) *Data Mining Concepts and Techniques*, 3rd ed., Morgan Kaufmann, Burlington.

- Inplang, W. and Thongkam, J. (2020) 'Opinion analysis on playerunknown's battlegrounds (PUBG) mobile using text mining', *Journal of Science and Technology Mahasarakham University*, Vol. 39, No. 5, pp.523–531.
- Jones, M.T. (2008) *Artificial Intelligence*, Infinity Science Press, Massachusetts.
- Kim, Y.J. and Lee, D.S. (2020) 'Predictive analysis of problematic smartphone use by machine learning technique', *KSCI*, Vol. 25, No. 2, pp.213–219 [online] <http://koreascience.or.kr/article/JAKO202008072030873.page> (accessed 8 March 2023).
- Lee, M., Han, M. and Pak, J. (2018) 'Analysis of behavioral characteristics of smartphone addiction using data mining', *MDPI*, Vol. 8, No. 7, pp.1–10 [online] <https://www.mdpi.com/2076-3417/8/7/1191> (accessed 8 March 2023).
- Nai-arun, N. and Mounngmai, R. (2020) 'Diagnostic prediction models for cardiovascular disease risk using data mining techniques', *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, Vol. 14, No. 2, pp.113–121 [online] <https://doi.org/10.37936/ecti-cit.2020142.199897>.
- National Health Commission Office (NHCO) (n.d.) *Research Results Show That Children and Youth are Addicted to Games, an Average of 5 Hours/Day, Spending 5,000 Baht/Month* [online] <https://www.nationalhealth.or.th/th/node/3006> (accessed 7 April 2023).
- Newzoo International B.V. (2022) *Newzoo Global Games Market Report 2022 Free Version* [online] <https://newzoo.com/resources/trend-reports/> (accessed 13 April 2023).
- Nusawat, P. and Kwangsawad, A. (2020) 'Modeling and analysis in the risk level of nomophobia from smartphone usage behavior using data mining techniques', *International Journal on Applied Biomedical Engineering*, Vol. 13, No. 1, pp.58–67.
- Panich, W., Khemapatapan, C. and Mrazek, P. (2021) 'Using neural network technique to predict grades in web database subject', in *Proceeding of NCIT 2021 the 13th National Conference on Information Technology*, pp.59–64.
- Panthong, R. (2023) 'Hybrid SMOTE and bootstrap sampling for imbalanced classification in elderly health condition dataset', *International Journal of Computers and Their Applications*, Vol. 30, No. 3, pp.274–285.
- Pornnoppadol, C., Sornpaisarn, B., Khamklieng, K. and Pattana-amorn, S. (2014) 'The development of game addiction screening test (GAST)', *Journal of the Psychiatric Association of Thailand*, Vol. 59, No. 1, pp.3–14.
- RapidMiner Studio (2019) *Solutions* [online] <https://rapidminer.com/> (accessed 12 May 2019).
- Richter, F. (2022) *Video Games Are More Than Child's Play* [online] <https://www.statista.com/chart/9821/us-gamer-demography/> (accessed 13 April 2023).
- Sherer, J. (2023) *Internet Gaming* [online] <https://www.psychiatry.org/patients-families/internet-gaming> (accessed 13 April 2023).
- Singsakda, R. (2020) *Trend of Online Games*, Department of Trade Negotiations, Nonthaburi.
- Sinsomboonthong, S. (2018) 'An efficiency comparison in prediction of child and adolescence game addition in Bangkok', *Science and Technology Journal Thammasat University*, Vol. 26, No. 3, pp.405–416.
- Sun, S., Yang, J., Chen, Y-H., Miao, J. and Sawan, M. (2022) 'EEG signals based internet addiction diagnosis using convolutional neural networks', *Applied Sciences*, Vol. 12, No. 13, pp.1–10 [online] <https://www.mdpi.com/2076-3417/12/136297>.
- Tan, P-N., Steinbach, M., Anuj, K. and Kumar, V. (2020) *Introduction to Data Ming*, 2nd ed., Pearson, London.
- Ulkhaq, M.M., Rozaq, R., Ramadhani, R., Heldianti, R., Fajri, A. and Akshinta, P.Y. (2018) 'Validity and reliability assessment of the game addiction scale: an empirical finding from Indonesia', in *ICIBE' 18: Proceedings of the 4th International Conference on Industrial and Business Engineering, Association for Computing Machinery*, New York, United States, pp.120–124.

- Wang, Q., Ren, H., Long, J., Liu, Y. and Liu, T. (2019) 'Research progress and debates on gaming disorder', *General Psychiatry*, Vol. 32, No. 3, pp.1–6 [online] <http://dx.doi.org/10.1136/gpsych-2019-100071>.
- WebMD Editorial Contributors (2021) *Is Video Game Addiction Real?* [online] <https://www.webmd.com/mental-health/addiction/video-game-addiction> (accessed 13 April 2023).
- World Health Organization (2020) *Addictive Behaviours Gaming Disorder* [online] <https://www.who.int/news-room/questions-and-answers/> (accessed 13 April 2023).
- Ye, N. (2003) *The Handbook of Data Mining*, Lawrence Erlbaum Associates, New Jersey.