Available online at www.HighTechJournal.org

HighTech and Innovation Journal

ISSN: 2723-9535

Vol. 6, No. 1, March, 2025

# Fine-Tuned Attribute Weighted Naïve Bayes with Modified Partial Instances Reduction for Gaming Disorder Classification

Anastasya Latubessy [1, 2], Retantyo Wardoyo [3], Aina Musdholifah [3*], Sri Kusrohmaniah [4]

[1] Doctoral Program Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.

[2] Department of Informatics Engineering, Faculty of Engineering, Universitas Muria Kudus, Indonesia.

[3] Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.

[4] Department of Psychology, Faculty of Psychology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.

Received 15 August 2024; Revised 14 January 2025; Accepted 02 February 2025; Published 01 March 2025

**Abstract**

Fine Tuning Attribute Weighted Naïve Bayes (FTAWNB) is a reliable modified Naïve Bayes model. Even though it is able to provide high accuracy on ordinal data, this model is sensitive to outliers. To improve the performance of FTAWNB, this research modified the Partial Instances Reduction (PIR) technique to make the FTAWNB more adaptive to outliers. Nevertheless, in contrast to the original PIR technique, which substitutes missing values for data values deemed outliers, the PIR technique suggested in this study replaces data values deemed outliers using a Naïve Bayes weighting approach. The attribute values from the outlier data are replaced with the highest probability values for the attributes in the actual class. This PIR technique is referred to as modified PIR. The FTAWNB model with modified PIR has been evaluated using the Gaming Disorder dataset. Replacing the four attributes with the least amount of information resulted in accuracy gains of 99.74%, an increase of 1.53% over the FTAWNB model. The experimental result shows that adding the modified PIR technique to the FTAWNB model can handle the outlier in the data, proving it by increasing the performance in terms of accuracy, precision, and recall without pruning the dataset used.

*Keywords:* Classification; Attribute Weighted; Fine-Tune; Naïve Bayes; Instances Reduction; Gaming Disorder.

## 1. Introduction

Uncontrolled gaming patterns can cause Gaming Disorder (GD). The World Health Organization (WHO) has defined GD as a psychiatric disorder characterized by a pattern of persistent gaming or repetitive behavior, both online and offline, which is manifested by: 1) diminished ability to control the onset, frequency, intensity, duration, termination, and context of gaming; 2) increase priority on playing games, games take precedence over daily activities and other interests in life.; and 3) continue to increase gaming patterns even though negative consequences occur [1]. Prior to Gaming Disorder (GD) being classified as a mental disorder by the World Health Organization (WHO) in 2018, Jap et al. [2] conducted research on the level of gaming addiction in Indonesia by taking samples from several schools. Among

* Corresponding author: aina_m@ugm.ac.id

http://dx.doi.org/10.28991/HIJ-2025-06-01-05

the 3,264 participants, 1,477 had continued to play games at least once a month. It is estimated that 150 of the 1,477 participants may experience addiction, with a total of 89 participants possibly falling into the severe category. This study estimates that there is a 6.1% prevalence of people experiencing gaming addiction. Furthermore, a study showed that as many as 14% of teenagers in the capital city of Jakarta were indicated to be addicted to the Internet, and the two most common activities when surfing were playing online games and playing social media [3]. Next, a potential correlation test for Internet Gaming Disorder (IGD) was carried out on 639 Indonesian medical students in Jakarta. The IGD prevalence rate was 2.03% [4].

In line with this, a study demonstrates a connection between young men's mental health and the amount of time they spend playing video games. Based on the research findings at one of the Vocational High Schools in North Sulawesi with 102 respondents, 42.2% of students fell into the high category regarding how long they played games. This number correlates with the mental health picture of teenagers at that school, which is in the poor/poor category with a percentage of 58.8% [5]. Concurrently, a study conducted on a sample of Garut City students revealed that 38.5% of them did not have an online game addiction, 15.3% had a mild addiction, and 46.2% had a severe addiction [6].

Research on the detrimental effects of excessive gaming indicates that GD should be given top priority in both physical and mental health. There is a significant gap between the healthcare needs of individuals vulnerable to GD and the resources available in the region [7]. While many health professionals recognize the profound impact of GD and express concern, they often struggle to respond effectively due to several barriers. A critical issue is the unavailability of high-quality consultation materials or clear procedural guidelines, which hampers their ability to provide targeted care [8]. Furthermore, existing research in this area is often undermined by weak sampling methods and inconsistent measurement tools, limiting the reliability and applicability of findings [9].

Classification is a way for researchers to organize, describe, and relate to their scientific disciplines, including psychology, with two main principles: validity and utility. The principle of validity ensures whether the classification scheme provides an accurate picture of understanding in accordance with science and symptoms. In comparison, the utility principle determines how functional the classification is. The purpose of classification is to enable and make it easier for experts to communicate about a disorder without having to make a long list of signs of a disorder [10].

The GD classification process can use methods in computer science to manage and synthesize psychological data. The collaboration of these two fields of science is called Psychoinformatics [11], when computer and information science changes the views and methodology of traditional psychology [12]. There has been a paradigm shift in the field of psychosocial and behavioral health from traditional experimental techniques towards the use of technology, enabling the study of people in their daily lives at the level of pertinent behavioral, psychological, and medical variables, such as communication patterns and psychophysiological data [13]. In fact, the basis of Artificial Intelligence (AI) relies on cognitive approaches in psychology. Psychiatry and clinical psychology that use machine learning techniques specifically utilize multidimensional data sets to learn statistical functions to predict individual outcomes [14].

One of the reliable classification models in machine learning is the Naive Bayes classifier. Naïve Bayes (NB) is a simple and reliable classification model for supervised classification [15]. This model is included in the ten best algorithms [16]. Probability estimates are the basis of Naive Bayes. Therefore, Naïve Bayes is suitable for computing high-dimensional text classification problems [17]. Nevertheless, assuming conditional independence between attributes is a weakness of Naïve Bayes [18]. Many studies have improved the performance of Naïve Bayes by adding structural extensions. Additionally, the process of selecting and weighting instances and attributes is another way to improve its performance.

Almost all existing NB development models only focus on reducing the unrealistic assumption of attribute conditional independence or only emphasizing getting better conditional probability estimates. However, a study contends that both are equally significant. Then, it combines the attribute weighting concept with fine-tuning to create a framework called the FTAWNB (Fine-Tuning Attribute Weighted Naïve Bayes) model [18]. A Fine Tuned NB (FTNB), Boosted NB (BNB), Correlation-based Featured Weighting Filter for NB (CFW) and Standard NB (NB) have been compared with this FTAWNB model [19–22]. Compared to other models on the dataset, FTAWNB performs exceptionally well, outperforming NB and all other cutting-edge models. The FTAWNB model has notable weaknesses, particularly during the fine-tuning phase. One of its key limitations is its heightened sensitivity to outliers, which can significantly compromise its performance and reliability [18]. This sensitivity makes the model less robust in handling outlier data, highlighting a critical area for improvement.

An instance is said to be an outlier because the instance dramatically deviates from the other instances in its class label. To reduce this sensitivity, it is recommended to use the PIR (Partial Instances Reduction) technique [23]. This technique does not remove all instances like traditional noise and outlier-filtering techniques but only removes some suspicious instances.

This research uses ordinal data, so the PIR technique used is slightly modified. The original PIR technique will replace attribute values that are considered outliers with missing values. However, this research uses a Naïve Bayes weighting approach to replace data values that are considered outliers. Data considered an outlier is at the farthest distance from the class centroid, and data on attributes has the most minor mutual information. The attribute values from the outlier data are then replaced with the highest probability values for the attributes in the actual class. Therefore, this PIR technique is called Modified PIR.

This paper is structured as follows: Section 2 reviews the literature relevant to the methods applied in this study. Section 3 introduces the FTAWNB model with the modified PIR technique. Section 4 presents the experimental and analytical results for the gaming disorder datasets. Finally, Section 5 provides the conclusions of the study.

## 2. Related Work

Every region faces challenges in diagnosing health issues, according to reports from the WHO since 2001 [24]. The lack of treatment facilities and the drawn-out, laborious diagnosis procedure were also addressed by WHO [25]. As a result, many nations employ advanced technology and knowledge advancement to tackle mental health issues. Naïve Bayes is used in a number of studies to identify mental health issues. Research on mental health issues such as depression, anxiety disorders, and other conditions will still largely rely on machine learning models for prediction even in 2024 and beyond [26–31].

Nine classification algorithms—Gradient Boosting, Multi-Layer Perceptron (MLP), AdaBoost, XGBoost, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, Decision Tree, and Gaussian Naïve Bayes (GNB)—were compared in a study on anxiety disorder. Out of the nine algorithms, MLP has the highest accuracy at 99.96%, while GNB has the lowest accuracy at 79.46% [32]. Naïve Bayes performed better in cross-validation settings in Ioannidis et al.'s research on Internet addiction, but its PR-AUC performance varied more [33]. Additionally, a study that used data samples from 100 students assessed internet addiction. 88 of the 100 data samples that were used can be correctly classified by the Naive Bayes model [34]. An investigation was carried out to compare the auto-sklearn and multinomial logistic regression machine learning algorithms to the Naive Bayes model for mood and anxiety disorders prediction. Although each of the three models did well in terms of accuracy, the auto-sklearn model performed better than the other two [35].

Depression is another mental health condition that extensively utilizes machine learning models for early detection. One study developed a hybrid model combining Support Vector Machines (SVM) and Neural Networks for early detection of depression [30], another study employed a robust tuned extreme gradient boosting model generator to identify depression [31]. The Naive Bayes model was used in a number of studies to analyze depression; the datasets from Reddit [36], the Australian Data Archive [37], and survey data [38] yielded accuracy rates of 74.35%, 94%, and 86.364%, respectively. Furthermore, a study classifies internet addiction and depression using questionnaire data. The research yielded an accuracy of 84.1% for the Naive Bayes model of Internet addiction and 88.9% for the depression model [39]. According to some of this literature, the Naive Bayes model has not yet proven to be the most accurate machine learning algorithm when compared to the algorithms employed in other research.

The FTAWNB model was also used in previous research on ordinal data using the anxiety disorder dataset. The accuracy, precision, and recall of the model were found to be good, outperforming Gaussian, Categorical, and Multinomial Naïve Bayes models. With an accuracy value of 99.22%, According to the test results, the FTAWNB performs better in terms of recall, accuracy, and precision than the other three models. The accuracy of Multinomial Naïve Bayes is 61.104%, categorical is 91.592%, and Gaussian NB is 91.132% [40].

One of the improvements to the Naïve Bayes model is the Fine Tuning Attribute Weighted Naïve Bayes (FTAWNB) model, which combines the ideas of attribute weighting and fine-tuning to improve the Naive Bayer’s performance [18]. These two factors are thought to be equally significant in enhancing Naïve Bayes' performance. It is well known that Naïve Bayes (NB) predicts its class label and estimates the probability of its class membership using Equations 1 and 2. P(c) is the prior probability of class c; $a_j$ is the value of the $j_{th}$ attribute $A_j$ of x; and $P(a_j|c)$ is the conditional probability of $A_j = a_j$ which belongs to class C and is estimable using Equations 3 and 4. In this context, the set of all potential c class labels is denoted by C; the amount of attributes is m. In this case, $a_{ij}$ is the value of the $j_{th}$ attribute of the $i_{th}$ training instance, $c_i$ is the class label of the th training instance i, $n_j$ is the total value of the $j_{th}$ attribute $A_j$, and the indicator function δ(x,y) is one if x=y and zero otherwise.

$$P(c|x)_{NB} = \frac{P(c)\prod_{j=1}^{m} P(a_j|c)}{\sum_{c\in C} P(c)\prod_{j=1}^{m} P(a_j|c)}, \tag{1}$$

$$C(x)_{NB} = arg\ max_{c\in C} P(c|x), \tag{2}$$

$$P(c) = \frac{\sum_{i=1}^{n} \delta(c_i,c) + \frac{1}{q}}{n+1}, \tag{3}$$

$$P(a_j|c) = \frac{\sum_{i=1}^{n} \delta(a_{ij},a_j)\delta(c_i,c) + \frac{1}{n_j}}{\sum_{i=1}^{n} \delta(c_i,c)+1}, \tag{4}$$

The difference between FTAWNB and NB standard is in the equation used to compute the conditional probability $P'(a_j \mid c)$ in Equations 5 and 6. In this case, attribute weights and fine-tuning will be applied to the FTAWNB.

$$P(c|x)_{FTAWNB} = \frac{P(c)\prod_{j=1}^{m} P\prime(a_j|c)}{\sum_{c\in C} P(c)\prod_{j=1}^{m} P\prime(a_j|c)} \tag{5}$$

$$C(x)_{FTAWNB} = arg\ max_{c\in C} P(c|x) \tag{6}$$

Various machine learning algorithms in the literature address the outlier problem in different approaches. Some learning algorithms, such as decision trees, include an embedded pruning phase that primarily removes some tree branches in order to address outlier [41]. Other methods use a different phase for outlier reduction in which noisy events are detected and removed according to certain criteria [42–45] or corrected by changing the suspected value. There is a risk associated with labeling or correcting an instance because it may introduce a new noisy value [46, 47].

## 3. Proposed Method

This study was approved by the Medical and Health Research Ethics Committee (MHREC), Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, with an ethics number of KE/FK/0090/EC/2024. It begins with the procedure for gathering data, where the data used is questionnaires data with a sample age range of 12 to 20 years old were randomly sent to Indonesian schools in order to gather primary data on gaming disorders. The next step involves conducting the data preprocessing phase. The characteristics of the data utilized do not permit the presence of missing values within the dataset. Therefore, during the preprocessing stage, measures are taken to ensure that the data is in an ideal condition. The dataset pertaining to gaming disorders contains 782 instances.

There are 45 attributes used in this study. Statements of symptoms of gaming disorder are found in attributes 1 through 44 taken from the Gaming Disorder Detection Questionnaire (GDDK) adopted from the Internet Addiction Diagnostic Questionnaire (KDAI) [48], and the 45 attribute is the class label of GD prediction. The following is how Table 1 displays each statement's weight: Very rarely = 1, rarely = 2, Sometimes = 3, Often = 4, Very often=5, Always = 6, It is not in accordance with=0. The total of all the respondents' input is added to determine the final score. Those who score more than 170 are classified in the GD class. There are two class label used, namely No, and GD shown in Table 2.

**Table 1. Frequency rating of classification GD questionnaire**

| Frequency rating | Weight |
|---|---|
| Very rarely | 1 |
| Rarely | 2 |
| Sometimes | 3 |
| Often | 4 |
| Very often | 5 |
| Always | 6 |
| Not in accordance | 0 |

**Table 2. Classification GD**

| GD Prediction | Score | Class |
|---|---|---|
| No | ≤107 | 0 |
| GD | >107 | 1 |

The model validation procedure employing 10-fold cross-validation comes next. Ten equal-sized portions of the data are separated out. Nine parts become training data, and one part becomes testing data for every fold, ranging from one to ten folds. FTAWNB consists of two algorithms: the classification algorithm and the training algorithm. As a result, the training process will further refine the FTAWNB model and the modified PIR technique, ultimately generating class predictions. A flowchart outlining the classification methodology employed in this study is provided in Figure 1.

Meanwhile, as depicted in Figure 2, the proposed model operates in three distinct stages: the initialization phase, fine-tuning of conditional probabilities, and the modified partial instance reduction (PIR) phase. In the original PIR technique, instances identified as outliers are treated as missing values. However, the current research introduces a novel approach by incorporating a Naïve Bayes-based weighting method. This method replaces the attribute values of outliers with the values corresponding to the highest probability within the actual class of the dataset. This modification aims to preserve the integrity of the dataset while improving classification performance.

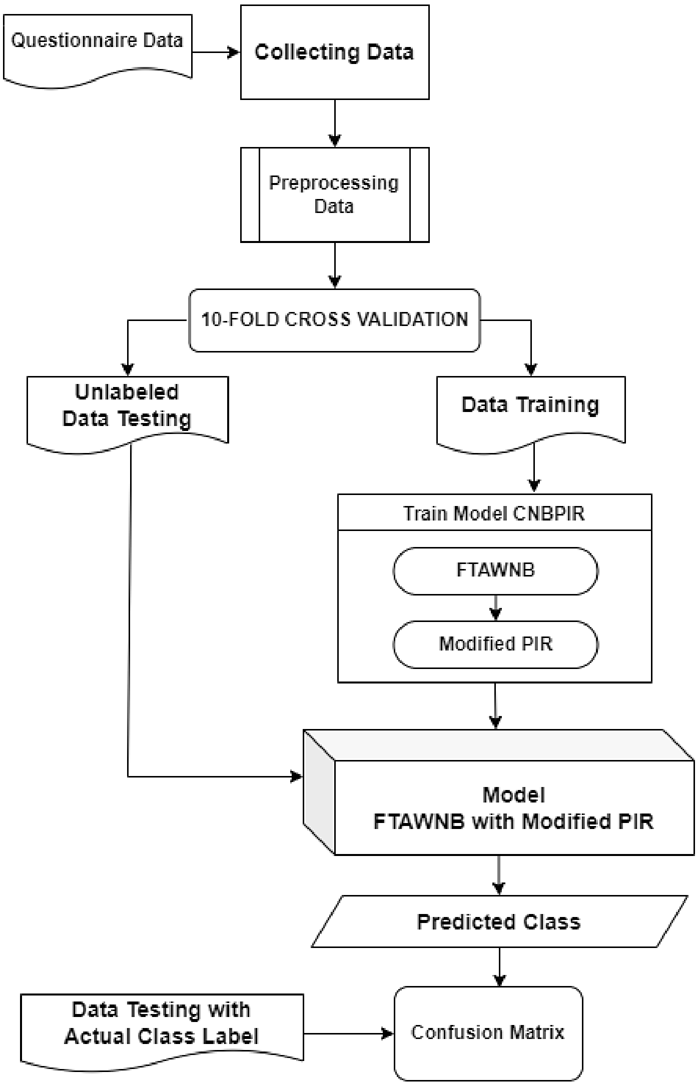


**Figure 1. The suggested model classification's workflow diagram**

Following this replacement, an accuracy check is performed. If the accuracy fails to improve or shows a decline compared to the previous iteration, the process is terminated, as illustrated in the third phase of Figure 2. Notably, the proposed model focuses solely on modifying the attribute values of data points considered outliers rather than pruning or eliminating data from the gaming disorder dataset. This approach ensures that all data is retained while addressing outlier effects, thereby maintaining the dataset's comprehensiveness and facilitating more robust classification results

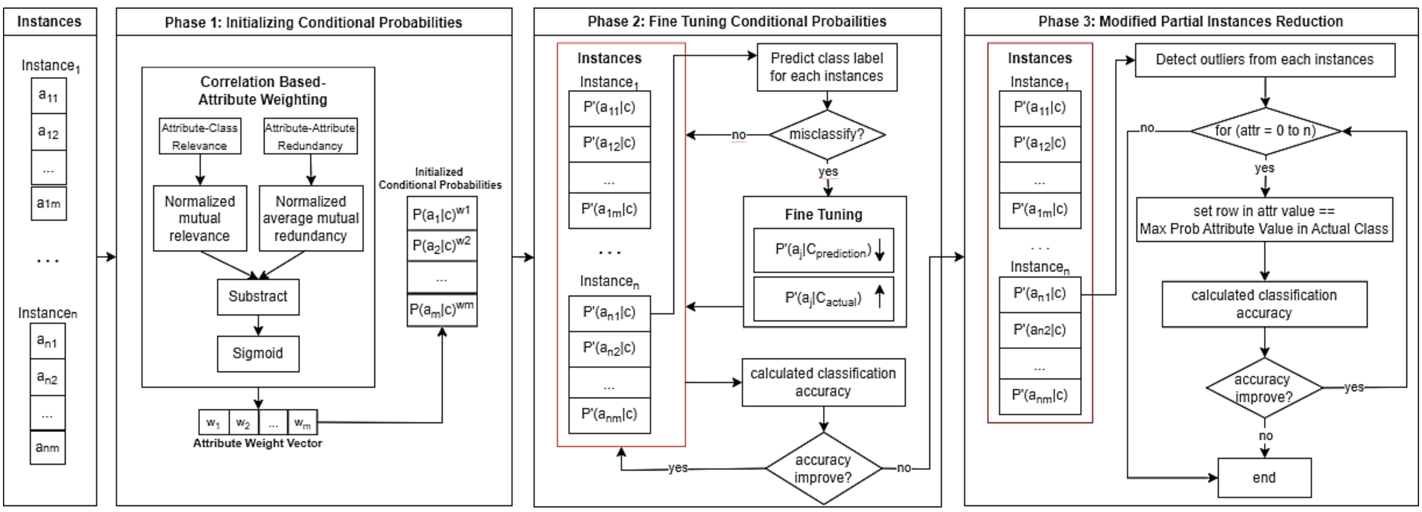


**Figure 2. Framework FTAWNB with modified PIR**

### 3.1. Initializing Conditional Probabilities Phase

The weights of the attributes are determined in the first phase by considering the redundancy of the attributes and the relevance of the classes. Initializing conditional probabilities is the term for this phase. The same information is constructed and initialized using conditional probabilities, which aim to quantify the correlation between every pair of discrete random variables. Equations 7 and 8 define the computations of attribute-class relevance and attribute-attribute inter-correlation, respectively. $I(A_j;A_k)$ denotes attribute inter-correlation, and $I(A_j;C)$ denotes attribute-class relevance.

$$I(A_j;C) = \sum_{aj}\sum_{c} P(a_j,c) log \frac{P(a_j,c)}{P(a_j)P(c)}, \tag{7}$$

$$I(A_j;A_k) = \sum_{aj}\sum_{ak} P(a_j,a_k) log \frac{P(a_j,a_k)}{P(a_j)P(a_k)}. \tag{8}$$

Normalization is carried out into $NI(A_j;C)$ and $NI(A_j;A_k)$ using Equations 9 and 10 in order to maintain $I(A_j;C)$ and $(A_j;A_k)$ in the range [0,1].

$$NI(A_j;C) = \frac{I(A_j;C)}{\frac{1}{m}\sum_{j=1}^{m} I(A_j;C)} \tag{9}$$

$$NI(A_j;A_k) = \frac{I(A_j;A_k)}{\frac{1}{m(m-1)}\sum_{j=1}^{m}\sum_{k=1\,\wedge k\neq j}^{m} I(A_j;A_k)}, \tag{10}$$

Next, in order to determine the weight of the $j_{th}$ attribute, $D_j$, the subtraction procedure is executed utilizing Equation 11. The weight of each attribute is determined by proportionally reducing the normalized mutual relevance and the normalized average mutual redundancy, as demonstrated by Equation 11. Because $D_j$, as defined by Equation 11, can be negative, $D_j$ is converted to [0, 1] by Equation 12 using the standard sigmoid logistic function. where $w_j$ represents the $j_{th}$ attribute's discriminatory weight.

$$D_j = NI(A_j;C) - \frac{1}{m-1}\sum_{k=1\,\wedge k\neq j}^{m} NI(A_j;A_k) \tag{11}$$

$$w_j = \frac{1}{1+e^{-Dj}} \tag{12}$$

### 3.2. Fine Tuning Conditional Probabilities Phase

Based on the conditional probabilities of the training instances, fine-tuning is done in the second stage. First, for each training instance $T_i(i=1,2,...,n)$, predict the class label ($C_{prediction}$) in turn. In the event that a training instance is misclassified ($C_{prediction} \neq C_{actual}$), adjust the relevant conditional probabilities. Equations 13 and 14 provide a clearer illustration of the fine-tuning formula for each misclassified training instance, where $c_{actu}$ and $c_{pred}$ represent the actual class and class prediction, respectively.

$$P'(a_j|c_{actu}) = P'(a_j|c_{actu}) + \delta(a_{j,}c_{actu}) \tag{13}$$

$$P'(a_j|c_{pred}) = P'(a_j|c_{pred}) - \delta(a_{j,}c_{pred}) \tag{14}$$

Hereafter, the learning rate is controlled by parameter $\eta \in [0,1]$. Likewise, δ(aj,cpred) must be reduced in proportion to the error, the difference between β. $P'(a_j|c_{pred})$ and $P'_{min}(a_j|c_{pred})$, and the learning rate η. Equations 15, 16, and 17 provide the formulas for varying the step sizes $\delta(a_j,c_{actu})$ and $\delta(a_j,c_{pred})$ based on this analysis.

$$\delta(a_j,c_{actu}) = \eta.(\alpha\,.\,P'_{max}(a_j|c_{actu}) - P'(a_j|c_{actu})).error \tag{15}$$

$$\delta(a_j,c_{pred}) = \eta.(\beta\,.\,P'(a_j|c_{pred}) - P'_{min}(a_j|c_{pred})).error \tag{16}$$

$$error = P(c_{pred}|T_i) - P(c_{actu}|T_i) \tag{17}$$

### 3.3. Modified Partial Instance Reduction Phase

The third phase begins by detecting the presence of outliers using the calculation of the Euclidean distance of each instances to the actual class centroid point. An instance is considered as outlier when the closest distance to the centroid of a particular class is different from the predicted class. Equation 18 is used to calculate the c center point on the n attribute. Where k= total number of instances in label c, i= instances, and $v_{i,n}$ = value of row i in attribute n with label c. In the meantime, equation 19 is used to determine the distance between each set of data and the cluster center. Where d(i,c) is the distance of data i to the center of cluster c, ($x_{ni}$) is the i data on the n attribute, and ($x_{nc}$) the c center point on the n attribute.

$$X_{nc} = \frac{1}{k} \times \sum_{i=0}^{k} v_{i,n}; \tag{18}$$

$$d(i,c) = \sqrt{[(x_{1i}-x_{1c})^2 + (x_{2i}-x_{2c})^2 + \cdots + (x_{ni}-x_{nc})^2]} \tag{19}$$

At this stage, we also look for the information gain (IG) value of each attribute and sort the IG value of each attribute from smallest to largest. Instances that are considered outliers will be replaced with attribute values starting from the attribute with the smallest IG to the largest according to the number of attributes selected. The modified PIR technique suggested in this study replaces data values that are deemed outliers using a Naïve Bayes weighting approach, as opposed to the original PIR technique, which substitutes missing values for outlier-class data values. Naive Bayes weighting is used to find the highest probability value of each attribute in the actual class.

Data on attributes with the least information gain and data that are the furthest from the class centroid are regarded as outliers. The highest probability values for the attributes in the actual class are then used to replace the attribute values from the outlier data. Therefore, this method is known as modified PIR. Algorithm 1 shows the procedures carried out in the FTWNB with modified PIR method.

**Algorithm 1. FTWANB with Modified Partial Instances Reduction algorithm**

```
Input: dataset, i
sorting IG value (dataset)
buildAndEvaluate(FTAWNB(), dataset, fold:10)
for (i = 0 to i < countFeature)
   buildcentroid()
   getouliters(centroid, dataset, igValList,i)
   buildAndEvaluate(FTAWNB(), dataset, fold:10)
      if(new accuracy < old accuracy)
        break;
      end if
    else
       return new accuracy
end for
Output: new dataset
```

## 4. Result and Discussion

### 4.1. Model Training Results

The dataset is information collected from a survey of children aged 12 to 20 based on questions related to gaming disorders. There are 45 columns total—44 columns for each questionnaire question, 1 column for the class label, and 782 rows to indicate the total number of participants. This 782 data are tested on the FTAWNB Model with modified PIR. The first experiment was carried out on the original FTAWB model; the next experiment was tested on FTAWNB with modified PIR by replacing outlier values on one, two, three, four and five attributes with the smallest Information Gain (IG) sequentially. Based on these experiment, the best accuracy was obtained when the value of the outlier data on the four attributes that had the smallest IG values were replaced with attribute values that had the greatest probability of attribute values in the actual class.

The cross-validation results are presented in Table 3. The accuracy is as follows: 98.98% for FTAWNB with modified PIR (1 attribute); 99.23% for FTAWNB with modified PIR (two attributes); 99.49% for FTAWNB with modified PIR (three attributes); 99.74% for FTAWNB with modified PIR(four attributes) and 99.62% for FTAWNB with modified PIR (five attributes). Table 3's final row demonstrates that the number of instances in the dataset used for each model test has not decreased. It was also shown that there was a progressive decrease in the total number of outliers found. Using the original FTAWNB model yielded 70 outliers, utilizing FTAWNB with modified PIR (1 attribute) yielded 61 outliers, FTAWNB with modified PIR (2 attributes) produced 59 outliers, FTAWNB with modified PIR (3 attributes) produced 57 outliers, and FTAWNB with modified PIR (4 attributes) produced 47 outliers.

**Table 3. Stratified cross-validation of gaming disorder classification**

| Parameter | FTAWNB | FTAWNB with modified PIR (1 attribute) | FTAWNB with modified PIR (2 attributes) | FTAWNB with modified PIR (3 attributes) | FTAWNB with modified PIR (4 attributes) | FTAWNB with modified PIR (5 attributes) |
|---|---|---|---|---|---|---|
| Accuracy | 98.21% | 98.98% | 99.23% | 99.49% | 99.74% | 99.62% |
| Correctly Classified Instances | 768 | 774 | 776 | 778 | 780 | 779 |
| Incorrectly Classified Instances | 14 | 8 | 6 | 4 | 2 | 3 |
| Kappa statistic | 0.9362 | 0.9633 | 0.9725 | 0.9815 | 0.9908 | 0.9862 |
| Mean absolute error | 0.0194 | 0.0145 | 0.0116 | 0.0102 | 0.0085 | 0.0081 |
| Root mean squared error | 0.1045 | 0.0869 | 0.074 | 0.0669 | 0.0597 | 0.0583 |
| Relative absolute error | 7.0215% | 5.2544% | 4.2157% | 3.7072% | 3.0822% | 2.9389% |
| Total number of Outliers | 70 | 61 | 59 | 57 | 47 | 41 |
| Total number of instances | 782 | 782 | 782 | 782 | 782 | 782 |

As indicated in Table 4, model performance is evaluated using three criteria: recall, precision, and accuracy. Meanwhile, detailed performance by class is shown in Table 5. The reliability of the suggested FTAWNBmPIR model is evaluated by comparing its performance to that of the original FTAWNB model. Furthermore, the performance of the FTAWNBmPIR is compared to other well-known outlier-handling strategies, including reliable approaches like LOF (Local Outlier Factor). The FTAWNBmPIR (four attributes) model obtained the highest accuracy value of 99.74%. Thus, it can be demonstrated that the accuracy of the FTAWNB model on the dataset of gaming disorders was increased by 1.53%. Meanwhile, the single-attribute, two-attribute, three-attribute, five-attribute modified PIR, and FTAWNB with Local Outlier Factor (LOF) techniques each also show greater accuracy than the original FTAWNB model.

**Table 4. Model performance comparison of gaming disorder classification**

| Model | Accuracy | Precision | Recall |
|---|---|---|---|
| FTAWNB | 98.21% | 98.3% | 98.2% |
| FTAWNB with modified PIR (one attribute) | 98.98% | 99.0% | 99.0% |
| FTAWNB with modified PIR (two attributes) | 99.23% | 99.3% | 99.2% |
| FTAWNB with modified PIR (three attributes) | 99.49% | 99.5% | 99.5% |
| FTAWNB with modified PIR (four attributes) | 99.74% | 99.7% | 99.7% |
| FTAWNB with modified PIR (five attributes) | 99.62% | 99.6% | 99.6% |
| FTAWNB with LOF | 98.59% | 98.6% | 98.6% |

**Table 5. Comprehensive performance indicators by class**

| Model | Class | Precision (%) | Recall (%) | TP Rate (%) | FP Rate (%) |
|---|---|---|---|---|---|
| FTAWNB | 0 | 99.4 | 98.5 | 98.5 | 0.31 |
| | 1 | 92.6 | 96.9 | 96.9 | 0.15 |
| FTAWNB with modified PIR (one attribute) | 0 | 99.7 | 99.1 | 99.1 | 0.16 |
| | 1 | 95.5 | 98.4 | 98.4 | 0.09 |
| FTAWNB with modified PIR (two attributes) | 0 | 99.8 | 99.2 | 99.2 | 0.08 |
| | 1 | 96.2 | 99.2 | 99.2 | 0.08 |
| FTAWNB with modified PIR (three attributes) | 0 | 99.8 | 99.5 | 99.5 | 0.08 |
| | 1 | 97.7 | 99.2 | 99.2 | 0.05 |
| FTAWNB with modified PIR (four attributes) | 0 | 100 | 99.7 | 99.7 | 0 |
| | 1 | 98.5 | 100 | 100 | 0.03 |
| FTAWNB with modified PIR (five attributes) | 0 | 100 | 99.5 | 99.5 | 0 |
| | 1 | 97.7 | 100 | 100 | 0.05 |
| FTAWNB with LOF | 0 | 99.4 | 98.9 | 98.9 | 0.31 |
| | 1 | 94. | 96.9 | 96.9 | 0.11 |

A comparison of the performance of the models tested on the gaming disorder dataset is shown in Figure 3. The results of the experiments that were done show that the suggested model, which is FTAWNB with a modified PIR (four attributes), achieved the highest accuracy, 99.74%. The same thing applies to precision and recall, which obtained the highest results in the FTAWNB with the modified PIR model (four attributes) of 99.70%.

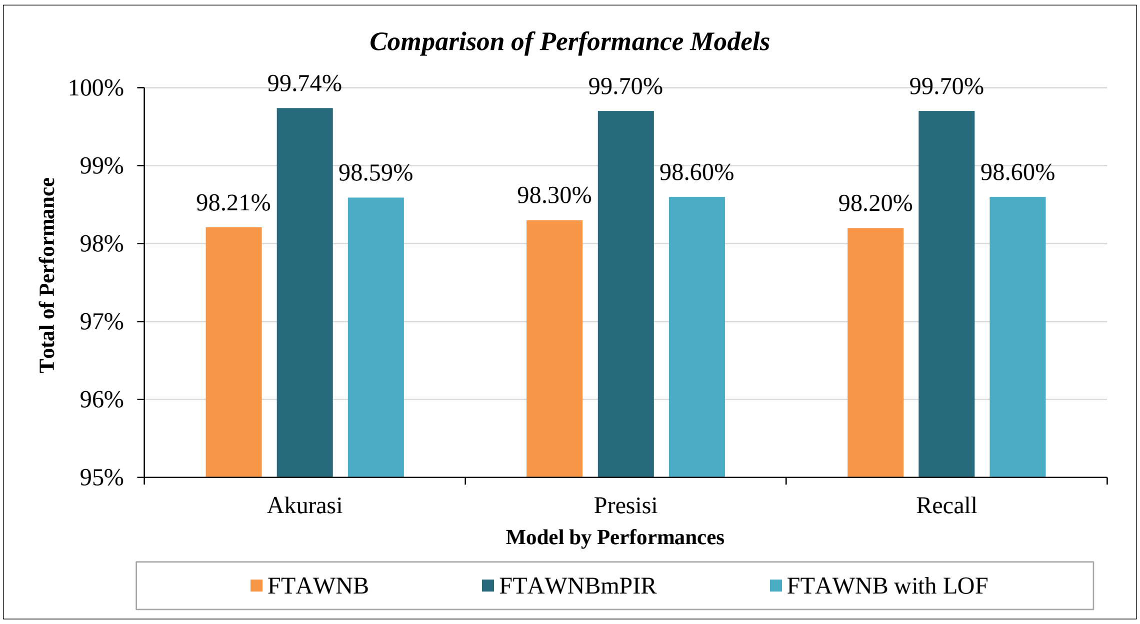


**Figure 3. Comparison of Performance Models: FTAWNB, FTAWNB with modified PIR, and FTAWNB with LOF**

As previously explained, in this study, gaming disorder was classified into two classes, namely No and GD (0.1). Table 6 presents the confusion matrix results from each test. In the FTAWNB model, there were 643 true negatives, 125 true positives, 10 false negatives, and 4 false positives. The FTAWNB model with modified PIR (one attribute) achieved 647 true negatives, 127 true positives, 6 false negatives, and 2 false positives. When modified with PIR (two attributes), the model produced 648 true negatives, 128 true positives, 5 false negative, and 1 false positives. With PIR modification using three attributes, the model recorded 650 true negatives, 128 true positives, 3 false negative, and 1 false positives. Further modification with PIR (four attributes) resulted in 651 true negatives, 129 true positives, 2 false negatives, and 0 false positives. The FTAWNB model with PIR (five attributes) reported 650 true negatives, 129 true positives, 3 false negatives, and 0 false positives. Finally, In the FTAWNB with LOF model, there were 646 true negatives, 125 true positives, 7 false negatives, and 4 false positives.

Within the original FTAWNB model, 14 cases were classified incorrectly. Based on the modified PIR stage in the third phase, 10 false positive instances at index 79, 173, 176, 389, 468, 490, 664, 678, 706, and 720 were identified as outlier data. In contrast, the four false negative cases were not recognized as data outliers. The index numbers of the four false negative cases are 172, 200, 527, and 702. However, Table 6b shows that the developed model can produce 0 false negative examples out of 4 false negative examples, which is not outlier data.

Meanwhile, the 10 false positive instances, which were outlier data, experienced a decrease in the number of misclassified data and achieved the best performance when using FTAWNB with modified PIR (four attributes). It can be seen that the number of false positive cases has decreased to two, as shown in Table 6a. These two data have respective index numbers of 664 and 678. In the meantime, three false positive instances—indexes 664, 678, and 706—are produced when FTAWNB is used in conjunction with a modified PIR (five attributes). Based on these findings, the proposed model stops replacing outlier data at the fourth attribute and displays the best accuracy results.

**Table 6. Confusion matrix (a) class 0 (b) class 1**

| Model | Class | Classified as | | |
|---|---|---|---|---|
| | | 0 | 1 | |
| FTAWNB | 0 | 643 | 10 | |
| FTAWNB with modified PIR (one attribute) | 0 | 647 | 6 | |
| FTAWNB with modified PIR (two attributes) | 0 | 648 | 5 | (a) |
| FTAWNB with modified PIR (three attributes) | 0 | 650 | 3 | |
| FTAWNB with modified PIR (four attributes) | 0 | 651 | 2 | |
| FTAWNB with modified PIR (five attributes) | 0 | 650 | 3 | |
| FTAWNB with LOF | 0 | 646 | 7 | |

| Model | Class | Classified as | | |
|---|---|---|---|---|
| | | 0 | 1 | |
| FTAWNB | 1 | 4 | 125 | |
| FTAWNB with modified PIR (one attribute) | 1 | 2 | 127 | |
| FTAWNB with modified PIR (two attributes) | 1 | 1 | 128 | **(b)** |
| FTAWNB with modified PIR (three attributes) | 1 | 1 | 128 | |
| FTAWNB with modified PIR (four attributes) | 1 | 0 | 129 | |
| FTAWNB with modified PIR (five attributes) | 1 | 0 | 129 | |
| FTAWNB with LOF | 1 | 4 | 125 | |

### 4.2. Model Testing Results

Testing of the FTAWNBmPIR model was conducted using 73 testing data, separate from the 782 training data points used during the model training phase. Table 7 presents the performance evaluation results of the FTAWNBmPIR model on the test data. The evaluation parameters include the number of correctly and incorrectly classified instances, accuracy, precision, recall.

**Table 7. FTAWNBmPIR model testing results**

| Parameter | FTAWNBmPIR |
|---|---|
| Incorrectly Classified Instances | 3 |
| Correctly Classified Instances | 70 |
| Accuracy | 95,89 % |
| Precision | 98,35 % |
| Recall | 95,89 % |

Based on the results of the model testing on the test data, the FTAWNB model was able to classify 70 data correctly. Two respondent data that were detected early on for gaming disorder in the actual class could be classified correctly. Meanwhile, 3 misclassified data were data that were in the 4th index, the 66th index, and the 71st index. The three data were data whose number of respondent answers was close to the minimum gaming limit, which was 107. The number of answers for the three misclassified data were respectively as follows: 105, 105, and 99. The FTAWNBmPIR model produces an accuracy value of 95.89%. The accuracy value obtained in this test shows that the model is able to classify well. The comparison of actual class labels and predicted classes of the FTAWNBmPIR model is shown in Table 8. The test results show that the FTAWNBmPIR model is highly influenced by data distribution, in order to be able to perform good classification.

**Table 8. Confusion matrix testing FTAWNBmPIR model**

| Model | Class | Classified as | |
|---|---|---|---|
| | | 0 | 1 |
| FTAWNBmPIR | 0 | 68 | 3 |
| | 1 | 0 | 2 |

### 4.3. Evaluation

### 4.4. Model Using Depression Dataset

This study also utilizes a depression dataset as benchmark data to evaluate the performance of the proposed FTAWNBmPIR model. The dataset, publicly available on Kaggle, comprises 2,556 instances with four commonly observed attributes representing depression symptoms and one class label attribute. Model training was conducted using both the original FTAWNB model and the enhanced FTAWNBmPIR model. The results of cross-validation are presented in Table 9.

The FTAWNBmPIR model demonstrated superior accuracy compared to the original FTAWNB model, achieving an improvement of 4.3819%. Specifically, the FTAWNBmPIR model attained an accuracy of 86.5806%, whereas the original FTAWNB model achieved an accuracy of 82.1987%. In terms of correctly classified instances, the FTAWNBmPIR model accurately predicted 2,213 cases, surpassing the 2,101 correctly classified cases by the original FTAWNB model.

**Table 9. Stratified cross-validation using depression dataset**

| Parameter | FTAWNB | FTAWNBmPIR |
|---|---|---|
| Accuracy | 82.1987% | 86.5806% |
| Correctly Classified Instances | 2101 | 2213 |
| Incorrectly Classified Instances | 455 | 343 |
| Kappa statistic | 0 | 0.3493 |
| Mean absolute error | 0.2686 | 0.2087 |
| Root mean squared error | 0.3662 | 0.3205 |
| Relative absolute error | 91.7395% | 71.2822% |

These findings demonstrate the potential applicability of the FTAWNBmPIR model in the mental health domain, particularly in addressing issues such as gaming disorder and depression. The model's effectiveness is substantiated by the observed improvements in accuracy across both datasets utilized in the study. The increase in accuracy highlights the model's ability to effectively capture the intricate patterns and relationships inherent in mental health-related data. This capability is critical for developing reliable tools that can support mental health diagnostics and interventions. Specifically, the enhanced performance of the FTAWNBmPIR model suggests its suitability for practical applications in identifying and categorizing complex mental health conditions based on symptomatology and other relevant attributes.

## 5. Conclusion

Data in the mental health domain is mostly ordinal because measuring instruments are used as questionnaires. The partial instance reduction technique needs to be modified when using ordinal data. Missing values will be substituted for outlier values in the original PIR technique, but for ordinal data, it is preferable to avoid missing values. The partial instance reduction technique must be adjusted when dealing with ordinal data. The PIR technique can be modified to find attribute values with the highest probabilities in the actual class using Naïve Bayes probabilities. These values can then be used to replace attribute values in outlier instances. This partial instance reduction modification technique can improve the performance of the FTAWNB model on the dataset used. The greatest accuracy in the gaming disorders dataset was obtained when using the FTAWNB with a modified PIR (four attributes) model, amounting to 99.74%. In the case of the depression dataset, the FTAWNBmPIR model demonstrates superior performance compared to the original FTAWNB model. This improvement is evidenced by a notable increase in accuracy of 4.3819%.

This research shows that adding the mPIR technique to the FTAWNB model can increase its performance without pruning data on the dataset used. This research also proves that the proposed model can reduce the number of outlier data in the dataset used. The results highlight the potential of the FTAWNBmPIR model for applications in the mental health field, especially in addressing conditions like gaming disorder and depression. The model's capability is supported by the notable improvements in accuracy observed across the datasets analyzed in this study. To improve the mPIR technique, future research could explore the alternative of feature selection methods and other distance measurement methods.

## 6. Declarations

### 6.1. Author Contributions

Conceptualization, A.L. and R.W.; methodology, A.L.; validation, A.L., R.W., A.M., and S.K.; formal analysis, A.L.; investigation, A.L. and R.W.; resources, A.L., R.W., A.M., and S.K.; writing—original draft preparation, A.L., R.W., A.M., and S.K.; writing—review and editing, A.L., R.W., A.M., and S.K.; visualization, A.L. All authors have read and agreed to the published version of the manuscript.

### 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6.3. Funding

Financial support provided by the Center for Higher Education Funding (BPPT) and Indonesia Endowment Funds for Education (LPDP) on decree no. 01054/J5.2.3./BPI.06/9/2022 is sincerely appreciated.

### 6.4. Acknowledgements

The authors would like to thank the Center for Higher Education Funding (BPPT) and Indonesia Endowment Funds for Education (LPDP) for funding this research.

### 6.5. Institutional Review Board Statement

This study was approved by the Medical and Health Research Ethics Committee (MHREC), Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, with an ethics number of KE/FK/0090/EC/2024. Thus, the informed consent form used is that presented by the Universitas Gadjah Mada Medical and Health Research Ethics Committee (MHREC).

### 6.6. Informed Consent Statement

Not applicable.

### 6.7. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 7. References

[1] W.H.O. (2018). ICD-11 Chapter 06 Mental, behavioural or neurodevelopmental disorders. ICD-11 International Classification of Diseases 11th Revision, World Health Organization, World Health Organization, Geneva, Switzerland.

[2] Jap, T., Tiatri, S., Jaya, E. S., & Suteja, M. S. (2013). The Development of Indonesian Online Game Addiction Questionnaire. PLoS ONE, 8(4), 4–8. doi:10.1371/journal.pone.0061098.

[3] Kurniasanti, K. S., Wiguna, T., Wiwie, M., & Winarsih, N. S. (2018). Internet addiction among adolescents in Jakarta: A challenging situation for mental health development. Journal of International Dental and Medical Research, 11(2), 711–717.

[4] Siste, K., Hanafi, E., Sen, L. T., Wahjoepramono, P. O. P., Kurniawan, A., & Yudistiro, R. (2021). Potential correlates of internet gaming disorder among Indonesian medical students: Cross-sectional study. Journal of Medical Internet Research, 23(4), e25468. doi:10.2196/25468.

[5] Erik, S., & Syenshie, W. V. (2020). Relationship Between Duration of Playing Online Games and Mental Health in Male Adolescents. Scientific Journal of Mental Health, 2(2), 69–75.

[6] Ikbal, I., Wikanengsih, W., & Septian, M. R. (2021). Profile of Online Game Addiction Level of Students in Class Xma Plus Al Mujammil Garut. FOKUS (Guidance & Counseling Study in Education), 4(1), 56. doi:10.22460/fokus.v4i1.6138.

[7] King, D. L., & Delfabbro, P. H. (2018). Internet Gaming Disorder: Theory, Assessment, Treatment, and Prevention. In Internet Gaming Disorder: Theory, Assessment, Treatment, and Prevention, 1–276. doi:10.1016/C2016-0-04107-4.

[8] King, D. L., Delfabbro, P. H., Wu, A. M. S., Doh, Y. Y., Kuss, D. J., Pallesen, S., Mentzoni, R., Carragher, N., & Sakuma, H. (2017). Treatment of Internet gaming disorder: An international systematic review and CONSORT evaluation. Clinical Psychology Review, 54(November 2016), 123–133. doi:10.1016/j.cpr.2017.04.002.

[9] Mihara, S., & Higuchi, S. (2017). Cross-sectional and longitudinal epidemiological studies of Internet gaming disorder: A systematic review of the literature. Psychiatry and Clinical Neurosciences, 71(7), 425–444. doi:10.1111/pcn.12532.

[10] Kaplan, H. I., & Sadock, B. J. (1988). Synopsis of psychiatry: Behavioral sciences clinical psychiatry. Williams & Wilkins Co, Baltimore, United States.

[11] Yarkoni, T. (2012). Psychoinformatics: New Horizons at the Interface of the Psychological and Computing Sciences. Current Directions in Psychological Science, 21(6), 391–397. doi:10.1177/0963721412457362.

[12] Pandey, D. (2020). Psychoinformatics: a theoretical approach on information science and psychology. Journal of Management and Science, 10(2), 7–10. doi:10.26524/jms.2020.2.2.

[13] Baumeister, H., & Montag, C. (2019). Digital Phenotyping and Mobile Sensing. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics, 466. doi:10.1007/978-3-030-98546-2.

[14] Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. Annual Review of Clinical Psychology, 14, 91–118. doi:10.1146/annurev-clinpsy-032816-045037.

[15] Stephens, C. R., Huerta, H. F., & Linares, A. R. (2018). When is the Naive Bayes approximation not so naive? Machine Learning, 107(2), 397–441. doi:10.1007/s10994-017-5658-0.

[16] Zhang, H., Jiang, L., & Yu, L. (2021). Attribute and instance weighted naive Bayes. Pattern Recognition, 111, 107674. doi:10.1016/j.patcog.2020.107674.

[17] Bai, Y., & Bain, M. (2022). Optimizing weighted lazy learning and Naive Bayes classification using differential evolution algorithm. Journal of Ambient Intelligence and Humanized Computing, 13(6), 3005–3024. doi:10.1007/s12652-021-03135-7.

[18] Zhang, H., & Jiang, L. (2022). Fine tuning attribute weighted naive Bayes. Neurocomputing, 488, 402–411. doi:10.1016/j.neucom.2022.03.020.

[19] Jiang, L., Zhang, L., Li, C., & Wu, J. (2019). A Correlation-Based Feature Weighting Filter for Naive Bayes. IEEE Transactions on Knowledge and Data Engineering, 31(2), 201–213. doi:10.1109/TKDE.2018.2836440.

[20] El Hindi, K. (2014). Fine tuning the Naïve Bayesian learning algorithm. AI Communications, 27(2), 133–141. doi:10.3233/AIC-130588.

[21] Elkan, C. (1997). Boosting And Naive Bayesian Learning. Department of Computer Science and Engineering University of California, San Diego La Jolla, California, United States.

[22] Langley, P., Iba, W., & Thompson, K. (1992). Analysis of Bayesian classifiers. Proceedings Tenth National Conference on Artificial Intelligence, 223–228.

[23] Jamjoom, M., & El Hindi, K. (2016). Partial instance reduction for noise elimination. Pattern Recognition Letters, 74, 30–37. doi:10.1016/j.patrec.2016.01.021.

[24] Morris, B. (2001). 54th World Health Assembly. International Journal of Health Care Quality Assurance, 14(6), 5. doi:10.1108/ijhcqa.2001.06214fab.005.

[25] Funk, M. (2016). Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. World Health Organization, World Health Organization, Geneva, Switzerland.

[26] Rahmadani, A., Setianingsih, C., Dirgantara, F. M., Ambarita, A. R., Arifin, H. I., Manalu, I. P., & Pratama, M. L. (2023). Depression, anxiety, and stress disorders detection in students during the Covid-19 pandemic using Naïve Bayes algorithm. BIO Web of Conferences, 75, 1003. doi:10.1051/bioconf/20237501003.

[27] Jin, Y., Xu, S., Shao, Z., Luo, X., Wang, Y., Yu, Y., & Wang, Y. (2024). Discovery of depression-associated factors among childhood trauma victims from a large sample size: Using machine learning and network analysis. Journal of Affective Disorders, 345(September 2023), 300–310. doi:10.1016/j.jad.2023.10.101.

[28] Zhong, Y., He, J., Luo, J., Zhao, J., Cen, Y., Song, Y., Wu, Y., Lin, C., Pan, L., & Luo, J. (2024). A machine learning algorithm-based model for predicting the risk of non-suicidal self-injury among adolescents in western China: A multicentre cross-sectional study. Journal of Affective Disorders, 345(May 2023), 369–377. doi:10.1016/j.jad.2023.10.110.

[29] Liu, Q., Zhou, B., Zhang, X., Qing, P., Zhou, X., Zhou, F., Xu, X., Zhu, S., Dai, J., Huang, Y., Wang, J., Zou, Z., Kendrick, K. M., Becker, B., & Zhao, W. (2023). Abnormal multi-layered dynamic cortico-subcortical functional connectivity in major depressive disorder and generalized anxiety disorder. Journal of Psychiatric Research, 167(May), 23–31. doi:10.1016/j.jpsychires.2023.10.004.

[30] Saha, D. K., Hossain, T., Safran, M., Alfarhood, S., Mridha, M. F., & Che, D. (2024). Ensemble of hybrid model based technique for early detecting of depression based on SVM and neural networks. Scientific Reports, 14(1), 1–18. doi:10.1038/s41598-024-77193-0.

[31] Ananthanagu, U., & Agarwal, P. (2024). DepXGBoot: Depression detection using a robust tuned extreme gradient boosting model generator. IAES International Journal of Artificial Intelligence, 13(4), 4352–4363. doi:10.11591/ijai.v13.i4.pp4352-4363.

[32] Rahman, A. A., Khalid, L. I., Siraji, M. I., Nishat, M. M., Faisal, F., & Ahmed, A. (2021). Enhancing the Performance of Machine Learning Classifiers by Hyperparameter Optimization in Detecting Anxiety Levels of Online Gamers. 24th International Conference on Computer and Information Technology, ICCIT 2021, 18–20. doi:10.1109/ICCIT54785.2021.9689911.

[33] Ioannidis, K., Chamberlain, S. R., Treder, M. S., Kiraly, F., Leppink, E. W., Redden, S. A., Stein, D. J., Lochner, C., & Grant, J. E. (2016). Problematic internet use (PIU): Associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. Journal of Psychiatric Research, 83, 94–102. doi:10.1016/j.jpsychires.2016.08.010.

[34] Nandhini, C., & Krishnaveni, K. (2016). Evaluation of internet addiction disorder among students. Indian Journal of Science and Technology, 9(19), 93864. doi:10.17485/ijst/2016/v9i19/93864.

[35] van Eeden, W. A., Luo, C., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., Wardenaar, K. J., Hoos, H., & Giltay, E. J. (2021). Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. Psychiatry Research, 299. doi:10.1016/j.psychres.2021.113823.

[36] Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and Suicide Analysis Using Machine Learning and NLP. Journal of Physics: Conference Series, 2161(1). doi:10.1088/1742-6596/2161/1/012034.

[37] Haque, U. M., Kabir, E., & Khanam, R. (2021). Detection of child depression using machine learning methods. PLoS ONE, 16(12 December 2021), 1–13. doi:10.1371/journal.pone.0261131.

[38] Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. Materials Today: Proceedings, 58, 217–226. doi:10.1016/j.matpr.2022.01.467.

[39] Rubaiyat, N., Apsara, A. I., Chaki, D., Arif, H., Israt, L., Kabir, L., & Rabiul Alam, M. G. (2019). Classification of depression, internet addiction and prediction of self-esteem among university students. 2019 22nd International Conference on Computer and Information Technology, ICCIT 2019, 18–20. doi:10.1109/ICCIT48885.2019.9038211.

[40] Latubessy, A., Wardoyo, R., Musdholifah, A., & Kusrohmaniah, S. (2024). Fine tuning attribute weighted naïve Bayes model for detecting anxiety disorder levels of online gamers. International Journal of Electrical and Computer Engineering, 14(3), 3277–3286. doi:10.11591/ijece.v14i3.pp3277-3286.

[41] Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. doi:10.1007/bf00116251.

[42] Frénay, B., & Verleysen, M. (2014). Classification in the presence of label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems, 25(5), 845–869. doi:10.1109/TNNLS.2013.2292894.

[43] Massie, S., Craw, S., & Wiratunga, N. (2007). When similar problems don't have similar solutions. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 4626 LNAI, 92–106. doi:10.1007/978-3-540-74141-1_7.

[44] Delany, S. J., Segata, N., & Mac Namee, B. (2012). Profiling instances in noise reduction. Knowledge-Based Systems, 31, 28–40. doi:10.1016/j.knosys.2012.01.015.

[45] Son, S. H., & Kim, J. Y. (2006). Data reduction for instance-based learning using entropy-based partitioning. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 3982 LNCS, 590–599. doi:10.1007/11751595_63.

[46] Yang, Y., Wu, X., & Zhu, X. (2004). Dealing with predictive-but-unpredictable attributes in noisy data sources. Knowledge Discovery in Databases: PKDD 2004: 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004. Proceedings 8, 471-483. doi:10.1007/978-3-540-30116-5_43.

[47] Brodley, C. E., & Friedl, M. A. (1999). Identifying Mislabeled Training Data. Journal of Artificial Intelligence Research, 11, 131–167. doi:10.1613/jair.606.

[48] Siste, K., Wiguna, T., Bardasono, S., Sekartini, R., Pandelaki, J., Sarasvita, R., Suwartono, C., Murtani, B. J., Damayanti, R., Christian, H., Sen, L. T., & Nasrun, M. W. (2021). Internet addiction in adolescents: Development and validation of Internet Addiction Diagnostic Questionnaire (KDAI). Psychiatry Research, 298(71), 113829. doi:10.1016/j.psychres.2021.113829.