





## Prediction of Violence Against Women Using Ensemble Learning Models: A Comparative Study of LightGBM, XGBoost, and Others

Harco Leslie Hendric Spits Warnars<sup>1\*</sup>, Aswan Supriyadi Sunge<sup>2</sup>, Suzanna<sup>3</sup>, Beni Bevlyadi<sup>4</sup>, Maybin K. Mueyba<sup>5</sup>

<sup>1</sup> Computer Science Department, BINUS Graduate Program, Bina Nusantara University, Jakarta 11480, Indonesia

<sup>2</sup> Informatics Engineering Department, Pelita Bangsa University, Bekasi 17530, Indonesia

<sup>3</sup> Informatics Information Systems Department, Binus Online Learning, Bina Nusantara University, Jakarta 11480, Indonesia

<sup>4</sup> Management Department, Bentara Persada Batam College of Economics, Batam 29425, Indonesia

<sup>5</sup> School of Science, Engineering & Environment (SEE), University of Salford, Manchester M5 4WT, United Kingdom

Corresponding Author Email: [spits.hendric@binus.ac.id](mailto:spits.hendric@binus.ac.id)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijssse.150409>

### ABSTRACT

**Received:** 22 January 2025

**Revised:** 25 March 2025

**Accepted:** 9 April 2025

**Available online:** 30 April 2025

#### **Keywords:**

*prediction, domestic violence, woman, machine learning, ensemble learning*

Violence against women and the possibility of its occurrence among children is a very serious issue that negatively impacts the physical, psychological, and emotional aspects of the victims and those around them. Various efforts have been made to reduce violence against women and children; however, in reality, such violence still occurs significantly in many countries due to emotions and turmoil within human relationships. It is necessary to propose prediction methods so that violence can be reduced through early observation and intervention against violence experienced by women. machine learning, as one of the Artificial Intelligence algorithms, offers a solution to identify and predict the risk of violence. This study aims to explore the use of several Ensemble Learning models, such as LightGBM, XGBoost, CatBoost, and AutoEnsemble, which are expected to improve prediction accuracy and stability. This study uses a dataset consisting of 348 samples with 5 selected features that represent indicators relevant to the risk of violence against women. The test results show that XGBoost and CatBoost achieved the highest accuracy, approximately 73%, with a precision of 76%, recall of 65%, and F1-Score of 70%. TabNet demonstrated similar performance with an accuracy of 73%, but with a higher recall of 70%. Meanwhile, LightGBM showed slightly lower performance with 68% accuracy and an F1-Score of 64%. AutoEnsemble produced stable results with 73% accuracy, 76% precision, 65% recall, and 70% F1-Score. However, the practical limitation of this study lies in the relatively small dataset size, which may affect the model's generalization ability when applied to larger or more diverse features. The findings of this study indicate that Ensemble Learning models can provide accurate and effective results in predicting violence against women. It is hoped that this research can contribute to more proactive and accurate efforts to prevent violence against women in the future.

## 1. INTRODUCTION

Violence against women is one of the biggest problems in the world that has a serious impact that not only occurs to each individual, but can also affect society at large and have a negative impact on a community [1]. Violence against women can occur in various forms ranging from physical violence, sexual violence, emotional violence, and even more so economic violence that has an impact on economic conditions that often occur in everyday life [2]. Several efforts to prevent violence against women have been made to reduce the consequences and impacts, but what has been done so far is still far from ideal expectations. Global research data shows that almost a third of women in the world experience physical violence or sexual violence including from their partners, whether bound or not bound by marriage [3]. Violence does not only occur in physical form, but can include sexual harassment, bullying, and gender discrimination [4]. In

addition, other causes are very complex and closely related to power inequality, abuse of power and wealth, and cultural norms that support male dominance in family life [5].

Furthermore, early prediction of violence against women is important in overcoming and preventing the impact and consequences of violence which are increasingly evident considering that its impact is not only felt by women, but also experienced by families and communities. Efforts to reduce violence against women can involve a holistic approach that includes education, legal implementation policies, and moral support for female victims [6]. One of the biggest challenges in overcoming violence against women is the inability to detect early signs of violence or patterns from when the approach is made by women or when dating activities are carried out before further relationships towards marriage.

However, on the other hand, there are also many incidents of violence against women that are not reported by families or authorities. This is certainly due to the living conditions and

fear experienced by victims and families and of course makes the impact of violence worse [7]. For this problem, we can apply machine learning (ML) (or Artificial Intelligence algorithms) to run predictions of possible violent activities against women. The use of this prediction algorithm makes it possible to analyze data such as reports of violence against women, the behavior of victims or perpetrators of violence, and socio-economic factors that influence the occurrence of violent activities against women. The use of predictive algorithms can help to accelerate the response to potential violence and provide a greater opportunity to help prevent violence against women before they occur.

Several previous studies have shown that ML can be used to predict violence in various contexts, such as violence-prone environments [8], patients with schizophrenia spectrum disorders who show aggressive behavior [9], violent behavior in schizophrenia patients [10], violent behavior towards health workers [11], and violence against women in Africa [12].

However, although ML models are very useful and can provide very good results, they have weaknesses, one of which is overfitting or bias [13-16]. Therefore, an Ensemble Learning (EL) method is needed, one of the functions of which is to combine the results of several models to produce more accurate, more stable predictions and compare which is better than a single model [17, 18].

Overall, previous studies have shown that EL improves the ability of predictive analysis in the context of violence against women by integrating the strengths of several models to produce more stable and accurate predictions. In addition, previous studies have shown that EL also helps in building a better and more reliable prediction system.

Although ML models have shown promising results in predicting violence against women, most previous studies have primarily focused on improving predictive accuracy without paying sufficient attention to the interpretability aspect of the models. Earlier research, such as that conducted by Chen and Guestrin [19] and Ke et al. [20], demonstrated that ensemble models like XGBoost and LightGBM can achieve high classification performance in social issue prediction tasks. However, these models often operate as "black boxes," making it difficult for stakeholders such as policymakers and social workers to understand the reasoning behind the predictions. Yet, interpretability is crucial in violence prediction tasks, as interventions based on non-transparent predictions risk misjudgment or ethical concerns [21, 22]. Meanwhile, studies using TabNet [23] offer a slightly more transparent decision-making process through an attention mechanism; however, comprehensive comparisons between the trade-offs of interpretability and predictive performance remain very limited. Furthermore, there is still a lack of systematic research comparing aspects of interpretability, recall sensitivity, and real-world applicability among various EL models such as XGBoost, TabNet, LightGBM, CatBoost, and AutoEnsemble in the context of violence prediction. Addressing this research gap is essential for developing ML systems that are not only accurate in prediction but also capable of providing understandable and actionable insights [24].

From the background discussed earlier, the main contributions of this study are:

1. Designing a prediction model for violence against women.
2. Implementation of 5 EL models such: LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble in detecting potential violence against women.

3. The algorithms will be evaluated using the 4-evaluation metrics: accuracy, precision, recall and F1-Score.

## 2. SIMILAR PREVIOUS RESEARCH

Previous studies have shown that machine learning (ML) can make a significant contribution to addressing the increasing problem of violence against women through more effective predictive modeling based on datasets. Some commonly used techniques such as Decision Trees (DT), Random Forests (RF), and Support Vector Machine (SVM) can identify patterns associated with risk factors for violence against women [25-28]. Furthermore, the use of ML in preventing violence against women requires attention to the accuracy and truth of data as valid research results. In addition, the ML process requires a valid dataset with no data defects such as errors and noise to ultimately help produces more accurate and accountable predictions.

Therefore, it is important to maintain ethics in producing and using data to ensure accuracy and authenticity in the ML modeling process [29]. Furthermore, one of the methods in ML, namely Ensemble Learning (EL), is an approach that combines several ML models to improve accuracy and reduce prediction errors [30, 31]. This EL technique uses a combination of several ML models to work together to produce a more accurate ML model than a single model [32, 33].

EL models have been widely used, but some that are widely used, especially in research, are the LightGBM and XGBoost models, which are boosting algorithms that combine weak ML models into stronger ML models [34, 35]. Several previous studies have used the LightGBM model in various fields, such as stock market prediction [36], traffic congestion prediction [37], and plasma density prediction [38]. On the other hand, XGBoost has also been applied in water production prediction [39], construction delay risk [40], steel pipe bond strength [41], car insurance fraud [42], and lower limb joint angle prediction based on electromyography [43].

Furthermore, the CatBoost model, which is known for its ability to handle large data, has been proven to improve model performance in managing complex and unbalanced data [44, 45]. In addition, CatBoost is also used in previous studies such as in order book price prediction [46], company failure prediction [47], early warning for mountain slope stability [48], and in Alzheimer's disease prediction [49]. In addition, the TabNet ML model is used to handle more complex tabular data with an efficient approach in managing data dependencies [50, 51], and has been applied in breast cancer prediction [52], pathological sound prediction [53], herbal medicine infrared analysis [54], bearing fault detection [55], and DDoS detection [56].

Finally, AutoEnsemble, which is known for its ability to handle imbalanced data and focus on minority classification as the best result [57], has been used in image label prediction [58] and learning rate scheduling [59]. Thus, EL can be used to build a more accurate and reliable predictive system, as well as more efficient in forecasting or detection and especially expected to predict the prevention of violence against women.

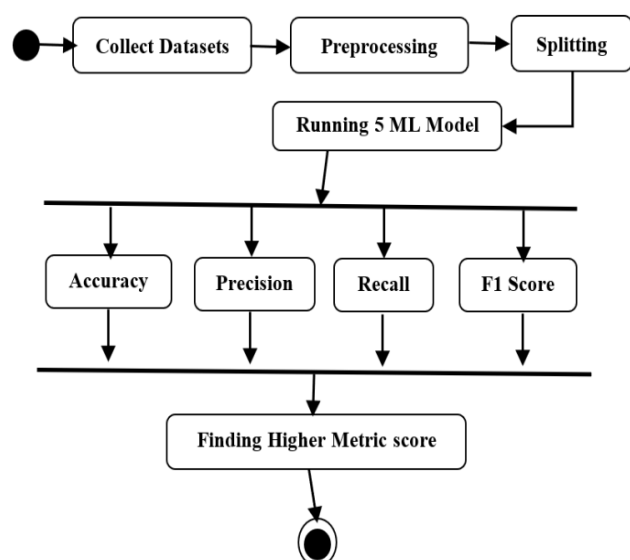
However, EL has its advantages and disadvantages, which is important for choosing the most appropriate model. XGBoost is known for its high accuracy and efficiency in handling large datasets with structured data, but it often requires careful hyperparameter tuning to avoid overfitting.

On the other hand, LightGBM offers faster training speed and lower memory usage, making it suitable for large-scale applications, although it may struggle with small datasets or highly imbalanced data. CatBoost shows strong performance on categorical and imbalanced datasets with minimal need for extensive preprocessing, but its training time can be relatively longer compared to LightGBM. TabNet provides advantages in interpretability through its careful feature selection mechanism, which is critical in high-stakes decision-making such as violence prevention; however, it generally demands larger training datasets and longer computation time. AutoEnsemble, which automatically combines the strengths of multiple models, offers strong performance, especially in minority class prediction, but can be computationally intensive and more complex to implement. Understanding this is necessary to develop machine learning systems that not only maximize predictive performance but also meet the practical needs of real-world applications in the sensitive area of violence prevention.

### 3. PROPOSED MODEL

The research that has been conducted and will be explained in this paper uses a quantitative approach in the form of an experimental study based on a literature review and theoretical framework that has been collected based on previous similar research. In this study, the performance of five different ML models, namely LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble, are compared in the context of improving the prediction of violence against women.

The evaluation of these ML models is carried out by measuring 4 white box testing metrics, namely accuracy, precision, recall, and F1-Score in order to find out which model provides the best performance. Furthermore, the main objective of this study is to provide deeper insights and meaningful contributions to the development of prediction of violence against women, with an emphasis on comparing the performance of these models.



**Figure 1.** Framework model

As seen in the activity diagram in Figure 1, this study was conducted through 6 stages, starting with data collection as the

first stage, after which the preprocessing process was carried out as the second stage in order to prepare the data used. Furthermore, the splitting stage as the third stage needs to be carried out to make the data used valid and free from noise. The fourth stage is carried out by running 5 ML models, namely LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble, and the fifth stage, the five ML models will be evaluated using 4 white box testing metrics such as accuracy, precision, recall and F1 Score. The last step, namely the sixth step, is carried out by comparing the results of the evaluation of the 4 metrics that were run previously and in order to find a higher metric score.

#### 3.1 Collect dataset

The process of collecting datasets related to cases of violence against women is taken from Kaggle data sharing on domestic violence against women [60]. This Kaggle dataset, which is secondary and public, is always updated regularly by the community, which contains 348 data with 5 features and a target class that indicates whether violence has occurred (0 = No or 1 = Yes). As seen in Table 1, the data collected are 5 features such as victim age, education, income, occupation, and marital status. The main purpose of sharing this Kaggle dataset is to estimate the likelihood of violence based on various factors, which in turn can help create more effective prevention measures and increase efforts to protect women as early as possible as a preventive measure.

**Table 1.** Dataset features and description

No.	Features	Description	Values
1	Age	Age at marriage	0 or 1
2	Education	Last education at marriage	0 or 1
3	Employment	Work when getting married	0 or 1
4	Income	Income at the time of marriage	0 or 1
5	Marital Status	Legally married or not	0 or 1

#### 3.2 Preprocessing

The data processing process begins with the pre-processing stage, which aims to prepare the data to be ready for testing and reduce the potential for errors that occur, as well as to support further data analysis. This pre-processing stage is very important to ensure the data before testing, and the steps taken can vary depending on the purpose and model used. The most commonly applied techniques include checking and cleaning data, such as irrelevant symbols, duplicates, typos, and unclear or incorrect data. The purpose of this stage is to ensure that the data is not empty or null and is free from errors. Testing is done using the syntax `data.isnull().sum()` to calculate the number of missing data, for more details it is then shown in Table 2 which illustrates that each variable has a value of 0 which means that no noise or null data was found.

**Table 2.** Test results inspecting and cleaning data

No	Features	Values
1	Age	0
2	Education	0
3	Employment	0
4	Income	0
5	Marital Status	0

### 3.3 Splitting

At this stage, the data is separated into two main parts, namely training data and testing data, which are used to run and comprehensively evaluate several proposed ML models, as well as to improve model performance. The data is divided with a proportion of 80% for training and 20% for testing. This division is intended to achieve the right balance between the two, providing sufficient data for the training process while leaving enough data to effectively evaluate the model results. The choice of the 80/20 split is based on the need to maintain a proportional distribution of samples between the two classes—violence and non-violence—in both the training and testing data. By keeping the class distribution balanced, the model can learn the characteristics of each class well during training and be fairly evaluated during testing. In addition, considering the limited size of the dataset, allocating 80% for training helps maximize the learning potential without sacrificing the robustness of the evaluation process.

### 3.4 Running 5 machine learning models

#### 3.4.1 LightGBM model

This model generally performs classification with a gradient boosting approach, which builds a prediction model based on successive decision trees. Here is the general formula in LightGBM classification:

##### a. Cumulative Prediction Model

LightGBM constructs the predictive model iteratively by adding a new prediction function at each iteration:

$$F_t(x) = F_{t-1}(x) + f_t(x) \quad (1)$$

where,

- $F_t(x)$  is the cumulative prediction model at iteration  $t$ .
- $F_{t-1}(x)$  is the model from the previous iteration.
- $f_t(x)$  is the new decision tree model added at iteration  $t$ .

##### b. Loss Function for Binary Classification

For binary classification tasks, LightGBM typically uses the binary cross-entropy loss function:

$$L = (y, p) = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2)$$

where,

- $y \in \{0,1\}$  represents the actual label.
- $p = \sigma(F_t(x)) = \frac{1}{1+e^{-F_t(x)}}$  the represents the predicted probability using the sigmoid function.

##### c. Second-Order Taylor Expansion Approximation

To accelerate optimization, LightGBM applies a second-order Taylor expansion approximation to the loss function:

$$L = (y, F_{t-1}(x) + f_t(x)) \approx L(y, F_{t-1}(x)) + g_t f_t(x) + \frac{1}{2} h_t f_t^2(x) \quad (3)$$

where,

- $g_t = \frac{\partial L}{\partial F(x)}$  is the first derivative (gradient).
- $h_t = \frac{\partial^2 L}{\partial F(x)^2}$  is the second derivative (hessian).

##### d. Gain Function for Split Optimization

When constructing decision trees, LightGBM determines

the optimal split based on the maximum gain, calculated as:

$$Gain = \frac{1}{2} \left( \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \quad (4)$$

where,

- $G_L, H_L$  = sum of gradients and Hessians for the left child.
- $G_R, H_R$  = sum of gradients and Hessians for the right child.
- $\lambda$  is the L2 regularization term.
- $\gamma$  is the penalty for creating a new split to control model complexity.

#### 3.4.2 XGBoost model

It is one of the very powerful ML models, especially in classification and regression tasks and an efficient implementation of the gradient boosting method, which focuses on optimization and speed. Here is the formula related to XGBoost:

##### a. XGBoost Model Prediction

The prediction of the XGBoost model ( $\hat{y}$ ) as shown in Eq. (5) is the result of aggregating a number of decision trees that contribute to the final prediction:

$$\hat{y} = \sum_{k=1}^K f_k(x_i) \quad (5)$$

where,

- $\hat{y}_i$  is the model prediction for the  $i$ -th data.
- $K$  is the number of decision trees in the model.
- $f_k(x_i)$  is the output of the  $k$ -th decision tree for the  $i$ -th data.

##### b. Loss Function

The loss function as shown in Eq. (6) is generally used to optimize the model. In the case of regression, for example, the one that is often used is the Mean Squared Error (MSE):

$$L = (\hat{y}_i, y_i) = (y_i, \hat{y}_i)^2 \quad (6)$$

where,

- $y_i$  is the actual value.
- $\hat{y}_i$  is the predicted value.

##### c. Objective Function

It is a combined loss of the loss function and decision tree regularization. The complete objective function as shown in Eq. (7) is:

$$Obj(\Theta) = \sum_{i=1}^N L(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (7)$$

where,

- $\Omega(f_k)$  is the complexity penalty of the  $k$ -th decision tree.
- Typically, this penalty as shown in Eq. (8) is:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

where,

- $T$  is the number of leaves in the decision tree.
- $w_j$  is the weight of the  $j$ -th leaf.

##### d. Gradient Boosting

XGBoost works by using the gradient boosting method, which iteratively adds new decision trees to reduce prediction errors.

##### e. Optimization and Split Algorithm

In XGBoost, we use an algorithm to decide how each

decision tree will “split” the data. This is done by finding the best features and threshold values that divide the data into the two most homogeneous parts.

### 3.4.3 CatBoost model

In general, this model cannot be explained with a single mathematical formula because it involves various complex concepts. However, the core of this model is the Gradient Boosting approach, which builds the model gradually by adding new decision trees at each iteration. The goal of each added tree is to correct the errors produced by the previous model, so that overall it can minimize the loss function. The formula used to achieve this goal is as follows shown in Eq. (9):

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (9)$$

where,

- $F_m(x)$  is the model in iteration  $m$ .
- $h_m$  is a tree in iteration  $m$ .
- $\gamma_m$  is the weighting coefficient determined by minimizing the loss function.
- $F_{m-1}(x)$  is the model in the previous iteration.

### 3.4.4 TabNet model

TabNet does not have a single formula like classical machine learning models such as linear regression or decision trees, but its architecture can be described through the following key components such as:

- a. Decision Tree and Attention Mechanism  
Uses an attention mechanism to select relevant features at each decision. Unlike traditional decision trees, TabNet can dynamically select relevant features for each decision step.
- b. Encoder  
The input data is first processed through an encoder that transforms it into a hidden representation. The encoder uses a series of decision steps consisting of Attentive Transformer and Gated Linear Units (GLUs).
- c. Attention Mechanism  
The model applies attention to a subset of features at each decision step. This attention mechanism helps the model to focus on the most important parts of the data as shown in Eq. (10).

$$a_t = \text{Softmax}(W_t h_{t-1} + b_t) \quad (10)$$

where,

- $h_{t-1}$  is the hidden representation at step  $t-1$ .
- $W_t$  is the weight matrix for decision step  $t$ .
- $b_t$  the bias for decision step  $t$ .
- Softmax ensures that the output  $a_t$  is a probability distribution (with sum 1).
- d. Sparsity Constraint  
A sparsity regularization term as shown in Eq. (11) is added during training to encourage the model to make decisions based on only a small subset of features, making it easier to interpret.

$$\mathcal{L}_{spat} = \lambda \sum_{t=1}^T \|a_t\|_1 \quad (11)$$

where,

- $\|a_t\|_1$  is the L1 norm of the attention mask at step  $t$ , which ensures sparsity (selecting only certain features).

- $\lambda$  is a constant that controls the strength of the regularization sparsity.

### e. Output Layer

The final layer combines the results of the attention mechanism and produces model predictions. Overall, TabNet works in a flexible and dynamic way in selecting relevant features and processing data, making it effective for various types of tabular datasets.

### 3.4.5 AutoEnsemble model

While there is no single “formula” for AutoEnsemble as the process can vary depending on the implementation and framework used, the general approach can be described as follows:

#### a. Model Selection

Selecting a range of machine learning models, such as Decision Trees, SVM, NN, K-NN, and others that combine the results of the attention mechanism and generate model predictions.

#### b. Model Training

Training each model individually using the training data, for example with a classification model with Logistic Regression as shown in Eq. (12):

$$P(y = c | x) = \frac{1}{1 + e^{-(w^t x + b)}} \quad (12)$$

where,

- $e$  is the input features.
- $x$  is the model weight vector.
- $b$  is the bias
- $P(y = c | x)$  is the predicted probability for class  $c$ .

#### c. Model Evaluation

Evaluating the performance of each model (e.g. accuracy, AUC, RMSE, and others) as shown in Eq. (13) on the validation data to identify which models perform best and which ones contribute less.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total number of Samples}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

where,

- TP is True Positive.
- TN is True Negative.
- FP is False Positive.
- FN is False Negative.

#### d. Determining Model Weights

Based on model performance, each model is given a different weight, for example a model with higher accuracy can have a greater influence or the weights can be learned automatically or determined by the meta model.

#### e. Final Prediction

Combining predictions from various models, there are several methods to combine them, namely Voting, Averaging and Stacking. For example, in a classification model with soft voting as shown in Eq. (14):

$$\hat{y}_i^{\text{ensemble}} = \arg \max (\sum_{j=1}^k w_j P(y_i = c_j | M_j)) \quad (14)$$

where,

- $w_j$  is the weight of model  $M_j$ .
- $P(y_i = c_j | M_j)$  is the probability of model  $M_j$  for class  $c_j$ .

- The prediction  $\hat{y}_i^{ensemble}$  is the class with the highest probability.

Overall, this model combines multiple models to improve overall performance by leveraging the strengths of different models and using methods such as averaging, weighted averaging, voting, or stacking to make the final prediction.

At the end of the day, the results of the evaluation using the LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble models identified the best performing models, as well as analyses that revealed influential factors and the potential for fruitful collaboration to develop more precise and efficient predictive models in supporting the prevention of violence against women.

### 3.5 Measuring 4 white box testing metrics

Table 3 shows the evaluation results of using 4 white box metrics such as accuracy, precision, recall, and F1-Score in predicting violence against women.

- Accuracy, which describes the proportion of correct predictions to the total predictions, shows that CatBoost, XGBoost, TabNet, and AutoEnsemble achieved a higher accuracy value of 73%, while LightGBM was slightly lower with an accuracy of 68%.
- Precision, which measures how accurate the model is in identifying positive cases, shows that apart from LightGBM, all models recorded a higher precision value of 76%, while LightGBM was at 70%.
- In terms of recall, which shows the model's ability to detect true positive cases, TabNet recorded the highest value with 70%, while the other models, namely XGBoost, CatBoost, and AutoEnsemble, had a recall of 65%, and LightGBM recorded a lower recall value of 60%.
- F1-Score, which is a measure of the balance between precision and recall, shows similar results for TabNet, XGBoost, CatBoost, and AutoEnsemble, at higher value at 70%, while LightGBM achieves an F1-Score of 64%.

**Table 3.** Ensemble learning model testing results

Model	Accuracy	Precision	Recall	F1-Score
LightGBM	68%	70%	60%	64%
XGBoost	73%	76%	65%	70%
CatBoost	73%	76%	65%	70%
TabNet	73%	76%	70%	70%
AutoEnsemble	73%	76%	65%	70%

### 3.6 Finding higher metric score

Overall, despite slight variations among these models, and as shown in Table 3 and the explanation then

- TabNet shows an advantage in term of recall metric, while the performance of the other models is almost similar in terms of accuracy, precision, and F1-Score.
- Meanwhile, LightGBM shows as the lowest model because all the metrics such as accuracy, precision, recall and F1-Score have low metric scores.

And then tested with a t-test model which is a statistical method used to determine whether there is a significant difference between the averages of two independent groups. This aims to broadcast whether there is a statistically significant difference between:

- Group 1: Individuals who did not experience violence (Violence = 0),
- Group 2: Individuals who experienced violence (Violence

= 1).

In this analysis, an independent two-sample t-test was conducted to compare the means of the numerical features, namely Age, which is an important variable in building a violence prediction model, because it shows a significant difference between groups.

From the interpretation of the results, the T-statistic of -3.138 shows that the difference in average age between the two groups is quite large compared to the variation in the data. The P-value of 0.00185 is much smaller than the significance level of  $\alpha = 0.05$ . Because the  $p$ -value  $< 0.05$ , the statistical decision is: Reject the null hypothesis ( $H_0$ ), which states that the average age of the two groups is the same.

This means, there is a statistically significant difference in age between individuals who experience violence and those who do not. The conclusion shows that age is significantly related to the likelihood of experiencing violence. The average age between the two groups is not the same, which means that age is an important factor to consider.

## 4. CONCLUSION AND FUTURE WORK

This study has several limitations that need to be considered, particularly regarding the generalization of the models tested on a specific dataset, which may not be applicable to other datasets or domains with different characteristics. Although various models show good performance, new data variations or class imbalance in the dataset may affect the results, especially in terms of recall and F1-Score. Additionally, this study is limited to five models, namely LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble, while many other models could potentially be more effective in the context of predicting violence against women.

The main limitations of this study also include the relatively small number of features and the limited sample size, with only 348 records, which may restrict the model's ability to generalize when applied to larger and more diverse populations. These limitations could hinder the model's capability to capture more complex patterns that may appear in broader and heterogeneous datasets.

The use of a single evaluation metric may also provide an incomplete picture of model performance, as other factors such as prediction speed and model interpretability are crucial in real-world applications. More complex models such as TabNet, although superior in recall, require more training data and are harder to interpret compared to other models like CatBoost. Therefore, the trade-off between accuracy, recall, interpretability, and model complexity must be carefully considered.

The conclusions obtained show that Ensemble Learning (EL) models such as LightGBM, XGBoost, CatBoost, TabNet, and AutoEnsemble overall demonstrate very similar performance in predicting violence against women, with an accuracy of around 73%, precision of 76%, and recall of 65%. Although the performance across models is relatively similar, TabNet is slightly superior in terms of recall, indicating its better ability to capture positive cases or critical events. While this difference is not significant, it could be an important consideration when choosing a model depending on specific needs, such as the balance between accuracy and recall or training time. Additionally, XGBoost and CatBoost excel in handling large and complex datasets, while AutoEnsemble is effective in combining the strengths of multiple models to

improve predictive accuracy.

LightGBM, despite showing slightly lower accuracy and recall, remains a good option for situations that require faster training times. Overall, all models tested are quite effective in predicting violence against women, and selecting the most appropriate model depends largely on the application goals and the design needs, based on the trade-offs between prediction speed and accuracy.

In the future, the development of this research is expected to focus on exploring and integrating more complex features, including socio-cultural attributes, economic conditions, education levels, social environment, and history of violence, which are believed to enhance the model's predictive accuracy and capability. Furthermore, future work should explore additional Ensemble Learning models such as AdaBoost, Bagging, Random Forest, or other recent boosting and bagging techniques to enrich the comparative analysis of the strengths and weaknesses of each model. The application of interpretability methods such as LIME, SHAP, or other explainable AI techniques is also strongly recommended to provide deeper insights into feature contributions, enabling models to be more adaptive to complex and diverse real-world situations.

## ACKNOWLEDGMENT

This work is supported by Bina Nusantara University as a part of Bina Nusantara University's BINUS International Research - Applied entitled "Community-Based Mobile Application for the Empowerment and Protection of Women for the Research Area in Cianjur City and its Surroundings" with contract number: 097/VRRTT/VII/2024 and contract date 02 Juli 2024." Data and code availability can be accessed at <https://github.com/spitswarnars/Woman-Ensemble-Learning>.

The author contributed to the writing as follows, where H.L.H.S. Warnars give his conceptualization, review, submit and edit the paper. He did the investigation and check the implementation. A.S. Sunge review and edit the paper. He was also did the implementation, collect the data and writing the draft paper. Suzanna reviewed the papers and added some ideas and correction. B. Bevyadi reviewed the paper, added some ideas and correction. M.K. Mueyba reviewed the paper and give his supervision upon the topic paper.

## REFERENCES

- [1] Stöckl, H., Sorenson, S.B. (2024). Violence against women as a global public health issue. *Annual Review of Public Health*, 45(1): 277-294. <https://doi.org/10.1146/annurev-publhealth-060722-025138>
- [2] Violence Against Women. (2024). <https://www.who.int/news-room/fact-sheets/detail/violence-against-women>, accessed on December 16, 2024.
- [3] Yakovleva, N., Vazquez-Brust, D.A., Arthur-Holmes, F., Busia, K.A. (2022). Gender equality in artisanal and small-scale mining in Ghana: Assessing progress towards SDG 5 using salience and institutional analysis and design. *Environmental Science & Policy*, 136: 92-102. <https://doi.org/10.1016/j.envsci.2022.06.003>
- [4] Causes and Consequences. (2024). Violence against women. <https://medicamondiale.org/en/violence-against-women/causes-and-consequences>, accessed on December 16, 2024.
- [5] Dahal, P., Joshi, S.K., Swahnberg, K.A. (2022). qualitative study on gender inequality and gender-based violence in Nepal. *BMC Public Health*, 22: 2005. <https://doi.org/10.1186/s12889-022-14389-x>
- [6] Lwamba, E., Shisler, S., Ridlehoover, W., Kupfer, M., Tshabalala, N., Nduku, P., Langer, L., Grant, S., Sonnenfeld, A., Anda, D., Eysers, J., Snilstveit, B. (2022). Strengthening women's empowerment and gender equality in fragile contexts towards peaceful and inclusive societies: A systematic review and meta-analysis. *Campbell Systematic Reviews*, 18(1): e1214. <https://doi.org/10.1002/cl2.1214>
- [7] Garfias Royo, M., Parikh, P., Walker, J., Belur, J. (2023). The response to violence against women and fear of violence and the coping strategies of women in Corregidora, Mexico. *Cities*, 132: 104113. <https://doi.org/10.1016/j.cities.2022.104113>
- [8] Parmigiani, G., Barchielli, B., Casale, S., Mancini, T., Ferracuti, S. (2022). The impact of machine learning in predicting risk of violence: A systematic review. *Frontiers in Psychiatry*, 13. <https://doi.org/10.3389/fpsyt.2022.1015914>
- [9] Parsaei, M., Arvin, A., Taebi, M., Seyedmirzaei, H., Cattarinussi, G., Sambataro, F., Pignoni, A., Brambilla, P., Delvecchio, G. (2024). machine learning for prediction of violent behaviors in schizophrenia spectrum disorders: A systematic review. *Frontiers in Psychiatry*, 15: 1384828. <https://doi.org/10.3389/fpsyt.2024.1384828>
- [10] Verrey, J., Ariel, B., Harinam, V., Dillon, L. (2023). Using machine learning to forecast domestic homicide via police data and super learning. *Scientific Reports*, 13(1): 22932. <https://doi.org/10.1038/s41598-023-50274-2>
- [11] Dobbins, N.J., Chipkin, J., Byrne, T., Ghabra, O., Siar, J., Sauder, M., Huijon, R.M., Black, T.M. (2024). Deep learning models can predict violence and threats against healthcare providers using clinical notes. *NPJ Mental Health Research*, 3(1): 61. <https://doi.org/10.1038/s44184-024-00105-7>
- [12] University of Texas at Dallas. (2022). Machine learning technique helps predict state violence in Africa. UT Dallas News Center. Retrieved from <https://news.utdallas.edu/social-sciences/machine-learning-predict-violence-2022/>.
- [13] Belkin, M., Hsu, D., Ma, S., Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32): 15849-15854. <https://doi.org/10.1073/pnas.1903070116>
- [14] Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168: 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- [15] Kolluri, J., Kotte, V.K., Phridviraj, M.S., Razia, S. (2020). Reducing overfitting problem in machine learning using novel L1/4 regularization method. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), Tirunelveli, India, pp. 934-938. <https://doi.org/10.1109/ICOEI48184.2020.9142992>

- [16] Cunningham, P., Delany, S.J. (2021). Underestimation bias and underfitting in machine learning. In: Heintz, F., Milano, M., O'Sullivan, B. (eds) *Trustworthy AI - Integrating Learning, Optimization and Reasoning (TAILOR 2020)*. Springer, Cham. [https://doi.org/10.1007/978-3-030-73959-1\\_2](https://doi.org/10.1007/978-3-030-73959-1_2)
- [17] Dietterich, T.G. (2000). Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science*, vol 1857. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)
- [18] Mienye, I.D., Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, 10: 99129-99149. <https://doi.org/10.1109/ACCESS.2022.3207287>
- [19] Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [20] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30: 3149-3157.
- [21] Arik, S.Ö., Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8): 6679-6687. <https://doi.org/10.1609/aaai.v35i8.16826>
- [22] Doshi-Velez, F., Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://doi.org/10.48550/arXiv.1702.08608>
- [23] Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 1135-1144. <https://doi.org/10.1145/2939672.2939778>
- [24] Rahman, R., Khan, M.N.A., Sara, S.S., Rahman, M.A., Khan, Z.I. (2023). A comparative study of machine learning algorithms for predicting domestic violence vulnerability in Liberian women. *BMC Women's Health*, 23(1): 542. <https://doi.org/10.1186/s12905-023-02701-9>
- [25] Rodríguez-Rodríguez, I., Rodríguez, J.V., Pardo-Quiles, D.J., Heras-González, P., Chatzigiannakis, I. (2020). Modeling and forecasting gender-based violence through machine learning techniques. *Applied Sciences*, 10(22): 8244. <https://doi.org/10.3390/app1022824>
- [26] Coll, C.V.N., Santos, T.M., Devries, K., Knaul, F., Bustreo, F., Gatuguta, A., Houvessou, G.M., Barros, A.J.D. (2021). Identifying the women most vulnerable to intimate partner violence: A decision tree analysis from 48 low and middle-income countries. *EClinicalMedicine*, 42: 101214. <https://doi.org/10.1016/j.eclinm.2021.101214>
- [27] Etzler, S., Schönbrodt, F.D., Pargent, F., Eher, R., Rettenberger, M. (2024). Machine learning and risk assessment: Random Forest does not outperform logistic regression in the prediction of sexual recidivism. *Assessment*, 31(2): 460-481. <https://doi.org/10.1177/10731911231164624>
- [28] Ye, L., Wang, L., Ferdinando, H., Seppänen, T., Alasaarela, E. (2020). A video-based DT-SVM school violence detecting algorithm. *Sensors*, 20(7): 2018. <https://doi.org/10.3390/s20072018>
- [29] Menger, V., Spruit, M., van Est, R., Nap, E., Scheepers, F. (2019). Machine learning approach to inpatient violence risk assessment using routinely collected clinical notes in electronic health records. *JAMA Network Open*, 2(7): e196709. <https://doi.org/10.1001/jamanetworkopen.2019.6709>
- [30] Corbett-Davies, S., Gaebler, J.D., Nilforoshan, H., Shroff, R., Goel, S. (2023). The measure and mismeasure of fairness. *Journal of Machine Learning Research*, 24(312): 1-117.
- [31] Khan, A.A., Chaudhari, O., Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 244: 122778. <https://doi.org/10.1016/j.eswa.2023.122778>
- [32] Mahajan, P., Uddin, S., Hajati, F., Moni, M.A. (2023). Ensemble learning for disease prediction: A review. *Healthcare*, 11(12): 1808. <https://doi.org/10.3390/healthcare11121808>
- [33] Dostmohammadi, M., Zamani Pedram, M., Hoseinzadeh, S., Astiaso Garcia, D. (2024). A GA-stacking ensemble approach for forecasting energy consumption in a smart household: A comparative study of ensemble methods. *Journal of Environmental Management*, 364: 121264. <https://doi.org/10.1016/j.jenvman.2024.121264>
- [34] Ahn, J.M., Kim, J., Kim, K. (2023). Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins*, 15(10): 608. <https://doi.org/10.3390/toxins15100608>
- [35] Ferrouhi, E.M., Bouabdallaoui, I. (2024). A comparative study of ensemble learning algorithms for high-frequency trading. *Scientific African*, 24: e02161. <https://doi.org/10.1016/j.sciaf.2024.e02161>
- [36] Zheng, X., Cai, J., Zhang, G. (2022). Stock trend prediction based on ARIMA-LightGBM hybrid model. In *2022 3rd Information Communication Technologies Conference (ICTC)*, Nanjing, China, pp. 227-231. <https://doi.org/10.1109/ICTC55111.2022.9778304>
- [37] Li, F., Nie, W., Lam, K.Y., Wang, L. (2024). Network traffic prediction based on PSO-LightGBM-TM. *Computer Networks*, 254: 110810. <https://doi.org/10.1016/j.comnet.2024.110810>
- [38] Wang, Y., Hu, W., Xiao, B., Yuan, Q., Zhang, R. (2025). Fault detection of line-averaged plasma density on EAST using LightGBM. *Fusion Engineering and Design*, 211: 114772. <https://doi.org/10.1016/j.fusengdes.2024.114772>
- [39] Qi, Z., Feng, Y., Wang, S., Li, C. (2025). Enhancing hydropower generation predictions: A comprehensive study of XGBoost and Support Vector Regression models with advanced optimization techniques. *Ain Shams Engineering Journal*, 16(1): 103206. <https://doi.org/10.1016/j.asej.2024.103206>
- [40] Alsulamy, S. (2024). Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM. *Expert Systems with Applications*, 268: 126268. <https://doi.org/10.1016/j.eswa.2024.126268>
- [41] Sheng, H., Ren, Z., Wang, D., Li, Q., Li, P. (2024). Estimation and interpretation of interfacial bond in concrete-filled steel tube by using optimized XGBoost



- and SHAP. Structures, 70: 107669. <https://doi.org/10.1016/j.istruc.2024.107669>
- [42] Ding, N., Ruan, X., Wang, H., Liu, Y. (2025). Automobile insurance fraud detection based on PSO-XGBoost model and interpretable machine learning method. Insurance: Mathematics and Economics, 120: 51-60. <https://doi.org/10.1016/j.insmatheco.2024.11.006>
- [43] Lu, Z., Chen, S., Yang, J., Liu, C., Zhao, H. (2025). Prediction of lower limb joint angles from surface electromyography using XGBoost. Expert Systems with Applications, 264: 125930. <https://doi.org/10.1016/j.eswa.2024.125930>
- [44] Hancock, J.T., Khoshgoftaar, T.M. (2020). CatBoost for big data: An interdisciplinary review. Journal of Big Data, 7: 94. <https://doi.org/10.1186/s40537-020-00369-8>
- [45] He, Y., Yang, B., Chu, C. (2024). GA-CatBoost-weight algorithm for predicting casualties in terrorist attacks: Addressing data imbalance and enhancing performance. Mathematics, 12(6): 818. <https://doi.org/10.3390/math12060818>
- [46] Bileki, G.A., Barboza, F., Silva, L.H.C., Bonato, V. (2022). Order book mid-price movement inference by CatBoost classifier from convolutional feature maps. Applied Soft Computing, 116: 108274. <https://doi.org/10.1016/j.asoc.2021.108274>
- [47] Ben Jabeur, S., Gharib, C., Meftah-Wali, S., Ben Arfi, W. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. Technological Forecasting and Social Change, 166: 120658. <https://doi.org/10.1016/j.techfore.2021.120658>
- [48] Cai, Y., Yuan, Y., Zhou, A. (2024). Predictive slope stability early warning model based on CatBoost. Scientific Reports, 14: 25727. <https://doi.org/10.1038/s41598-024-77058-6>
- [49] Shukla, R., Singh, T.R. (2024). AlzGenPred - CatBoost-based gene classifier for predicting Alzheimer's disease using high-throughput sequencing data. Scientific Reports, 14: 30294. <https://doi.org/10.1038/s41598-024-82208-x>
- [50] Zhang, B., Jin, X., Liang, W., Chen, X., Li, Z., Panoutsos, G., Liu, Z., Tang, Z. (2024). TabNet: Locally interpretable estimation and prediction for advanced proton exchange membrane fuel cell health management. Electronics, 13(7): 1358. <https://doi.org/10.3390/electronics13071358>
- [51] Zhen, Y., Zhu, X. (2024). An ensemble learning approach based on TabNet and machine learning models for cheating detection in educational tests. Educational and Psychological Measurement, 84(4): 780-809. <https://doi.org/10.1177/00131644231191298>
- [52] Shahriarirad, R., Meshkati Yazd, S.M., Fathian, R., Fallahi, M., Ghadiani, Z., Nafissi, N. (2024). Prediction of sentinel lymph node metastasis in breast cancer patients based on preoperative features: A deep machine learning approach. Scientific Reports, 14: 1351. <https://doi.org/10.1038/s41598-024-51244-y>
- [53] Alobaidi, M.H., Chebana, F., Meguid, M. A. (2018). Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Applied Energy, 212: 997-1012. <https://doi.org/10.1016/j.apenergy.2017.12.054>
- [54] Wang, Y., Jin, C., Ma, L., Liu, X. (2024). A robust TabNet-based multi-classification algorithm for infrared spectral data of Chinese herbal medicine with high-dimensional small samples. Journal of Pharmaceutical and Biomedical Analysis, 242: 116031. <https://doi.org/10.1016/j.jpba.2024.116031>
- [55] Khawaja, A.U., Shaf, A., Al Thobiani, F., Ali, T., Irfan, M., Pirzada, A.R., Shahkeel, U. (2024). Optimizing bearing fault detection: CNN-LSTM with attentive TabNet for electric motor systems. CMES-Computer Modeling in Engineering and Sciences, 141(3): 2399-2420. <https://doi.org/10.32604/cmes.2024.054257>
- [56] Setitra, M.A., Fan, M. (2024). Detection of DDoS attacks in SDN-based VANET using optimized TabNet. Computer Standards & Interfaces, 90: 103845. <https://doi.org/10.1016/j.csi.2024.103845>
- [57] Haghish, E.F., Nes, R.B., Obaidi, M., Qin, P., Stănicke, L.I., Bekkhus, M., Laeng, B., Czajkowski, N. (2024). Unveiling adolescent suicidality: Holistic analysis of protective and risk factors using multiple machine learning algorithms. Journal of youth and adolescence, 53(3): 507-525. <https://doi.org/10.1007/s10964-023-01892-6>
- [58] Ostrowski, E., Shafique, M. (2023). ISLE: A Framework for Image Level Semantic Segmentation Ensemble. arXiv preprint arXiv:2303.07898v1. <https://doi.org/10.48550/arXiv.2303.07898>
- [59] Yang, J., Wang, F. (2020). Auto-Ensemble: An adaptive learning rate scheduling based deep learning model ensembling. IEEE Access, 8: 217499-217509. <https://doi.org/10.1109/ACCESS.2020.3041525>
- [60] Kaggle. Domestic Violence Against Women. <https://www.kaggle.com/datasets/fahmidachowdhury/domestic-violence-against-women>