LAMPIRAN

HALAMAN
LETTER OF ACCEPTANCE (LOA)

SINERGI
Universitas Mercu Buana
http://publikasi.mercubuana.ac.id/index.php/sinergi

Dear
Zulfi Anugerahwati
Sri Lestari

Congratulation! We are pleased to inform you that your article entitled:

"Optimizing PSO for classification: comparison of Naïve Bayes and C4.5 for osteoporosis prediction"

was reviewed by the reviewer and got a positive opinion. The paper has been ACCEPTED for publication at *SINERGI* and is to be published on June 2, 2025 (Vol. 29 No. 2). Attached herewith is the revised version of the article (*if there is a mistake, please give us any comments, with a different color*) and *the Copyright Transfer Agreement*. Please send ***the final revision*** and ***the Copyright Transfer Agreement*** as soon as possible via this email.

Again, thank you for working with *SINERGI*. I believe our collaboration will help accelerate global knowledge creation and sharing one step further. *SINERGI* looks forward to your confirmation. Please do not hesitate to contact me if you have any further questions.

Sincerely,

Prof. Dr. Andi Adriansyah,

Tuesday, 21 January 2025
Editor-in-Chief *SINERGI*

HALAMAN
PLAGIARISM CHECK

SINERGI Vol. xx, No. x, February 20xx: xxx-xxx
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.xxxx.x.xxx

Optimizing PSO For Classification: Comparison Of Naïve Bayes And C4.5 For Osteoporosis Prediction

Zulfi Anugerahwati[1*], Sri Lestari[2]
[12]Master of Informatics Engineering Department, Faculty of Computer Science, Darmajaya Institute of Informatics and Business, Indonesia

Abstract
Osteoporosis is a medical disease marked by a reduction in bone density, which significantly increases the risk of fractures. Osteoporosis patients do not always exhibit symptoms and because current diagnostic techniques have limitations, early detection is frequently needed. The osteoporosis dataset consists of 1.958 records each containing 15 regular attributes and 1 special attribute as the label. The attribute represented as "1" for the presence of osteoporosis and "0" for its absence. The primary objective is to predict an individual's risk of developing osteoporosis, including age, gender, bone density, lifestyle factor, medical history, and nutritional intake of calcium and vitamin D. To achieve this, Naïve Bayes and C4.5 has been employed. PSO is employed to identify the most relevant features, thereby optimizing the efficiency and accuracy of the classification models. The initial step in data preprocessing involved handling missing values to ensure data integrity. After implementing PSO, Naïve bayes improved from 82,65% to 83,67%, while C4.5 exhibited an even greater increase, rising from 91,07% to 96,17%. PSO significantly optimizes model, with the most improvement in C4.5. PSO proves to be a valuable tool for feature selection. Age and Hormonal Change emerged as important for both models. Furthermore, Physical Activity and Calcium Intake, which despite having varying levels of influence, were consistently considered relevant. By focusing on these significant attributes, enables us more effectively monitor and recognize early signs of osteoporosis. Identifying individuals at high risk, more effective early detection and intervention, improving the potential for timely management and prevention.

Keywords:
Osteoporosis;
Decision Tree C4.5;
Naive Bayes;
Prediction;
PSO;

Article History:
Received: May 2, 2019
Revised: May 29, 2019
Accepted: June 2, 2019
Published: June 2, 2019

Corresponding Author:
Zulfi Anugerahwati
Master of Informatics Engineering Department, Darmajaya Institute of Informatics and Business, Indonesia
Email:
zulfi.2221210036@mai.darmajaya.ac.id

INTRODUCTION

Collagen, calcium, and proteins make up normal bone, which gives the bones their strength [1]. Because bone resorption occurs more quickly than bone production, bones may lose bulk and become porous, brittle, and feeble[2]. Osteoporosis is the term for bone loss[3] and medical disease marked by a reduction in bone density and loss of bone microstructure quality, significantly increasing the risk of fractures[4][5]. Over the past few decades, the prevalence of osteoporosis has grown significantly worldwide and has become one of the health problems that require serious attention[6]. Based on researcher studies[7], the prevalence of osteoporosis in the Asia-Pasific region shows that 10-30% of women over the age of 40 are affected. In contrast, in the European Union, the superiority of this medical disease in man elderly 50 years or older is 6,6%, increasing to 16,6% in guys aged eighty years older. As bones become more porous and fragile with age, osteoporosis predominantly affect the elderly and is more prevalent in women than a man[8].

Compared to men, women are likely to acquire osteoporosis. Women go through phases of pregnancy and breastfeeding, which is one of the main causes of osteoporosis. In addition, there are hormonal changes that occur throughout the postmenopausal period. A considerable loss in bone density might result from a fall in estrogen[9]. Slowly decreasing bone density is difficult for people to recognize without a professional medical evaluation is difficult to identify early because do not show typical symptoms [10]. Osteoporosis must be detected early, facilitate quicker and more efficient therapies, such as dietary modification, vitamin D and calcium supplementation, and medication use, to lower the risk of severe bone fractures and other complication[11]. However, because osteoporosis patients do not always exhibit symptoms and because current diagnostic techniques have limitations, early detection is frequently needed[12]. One of the methods for osteoporosis early detection is the Dual Energy X-ray Absorptiometry (DEXA) as it is a current technology to determine bone mineral density (BMD)[13]. However, the DEXA method is not only costly but also less accessible to remote populations. In addition, when osteoporosis patients also suffer from scoliosis, BMD dimension the usage of power DEXA becomes less accurate[14].

Data analysis techniques like clustering, classification, and prediction are developing at a faster rate than technology and data complexity, which is creating new potential for innovation and increased efficiency across a range of scientific filed[15]. Researchers and practitioners are able of make more informed decisions and more accurate predictions using the data mining techniques, which also aid in data classification and pattern recognition[16].

Classification methods like Naive Bayes and Decision Tree C4.5 can be used as the data analysis techniques [17]. As we know Particle Swarm Optimization (PSO) is employed for optimization because it robust, flexible, and efficient algorithm[18]. The use Particle Swarm Optimization (PSO) is to find the most optimal or best value of the classification process, usually indicated by an increase in accuracy when compared to a model without optimization. Particle Swarm Optimization (PSO) helps select the most relevant features so that the model becomes simpler and still effective[19].

Based on the provide explanation, it is important to investigate whether the application of PSO leads to improved evaluation metrics trough feature optimization. Several recent studies have implemented PSO as an optimization model. Dedi et al., [20] conducted PSO on the C4.5, SVM and the Naïve Bayes algorithm. Test result indicated that optimization leads to improvement in accuracy. Comparison among the Naïve Bayes and Naïve Bayes with PSO, the results showed a slight increase in accuracy, from 94,07% to 95,56%. But the precision and recall value are quite unusual with such large discrepancy[21]. The optimization of decision tree using PSO demonstrated increase in accuracy from 97,53& to 97,78%[22]. The research show that the Naïve Bayes algorithm achieved an accuracy of 93,24%, whereas the Naïve Bayes algorithm enhance with PSO reached a higher accuracy of 98,16% compared to the standard the Naïve Bayes[23]. Other study from [24], in the classification 4 methods were employed: DT, NB, SVM, and KNN. The result indicate there was increase in accuracy across all algorithms experienced an increase in accuracy. However, the most notable improvement was observed in SVM and KNN, with accuracies reaching 98,3%.

This study aims to compare the Naïve Bayes and C4.5 algorithm, with the addition of Particle Swarm Optimization (PSO) to enhance both algorithms optimization and features selection. Combining Naïve Bayes and C.45 with PSO are highly suitable for predicting osteoporosis risk due to their specific strengths in handling a complex data. The Naïve Bayes provides probabilistic prediction that account for uncertainly and variability in medical data, which is valuable for assessing various risk factors across different patient groups. C4.5 excels at handling complex dataset and determining the most relevance attribute for classification such as age, bone density, and lifestyle factors for classification. PSO further enhances these methods by optimizing model parameter, ensuring more accurate and reliable prediction. The approach is expected to yield reliable predictive results in accuracy, precision and recall, and identifying key predictors of osteoporosis risk.

METHODS AND MATERIAL

The research method is depicted in figure 1.

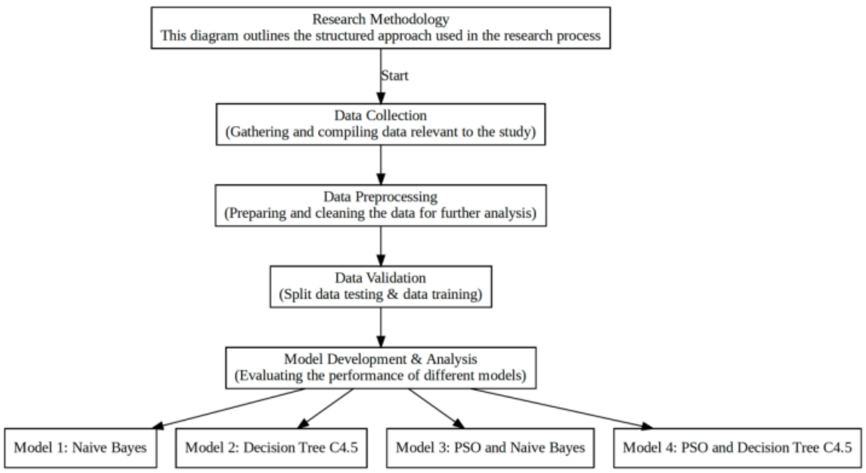

Figure 1. Researched Methodology

The research method is designed to ensure a systematic and structure approach. It begins with data collection trough data acquisition, followed by data preprocessing to prepare data analysis. The data validation step is crucial to ensure the accuracy and reliability of the data. During the model deployment process, four different models are tested to determine their effectiveness. These models include Naïve Bayes; C4.5; PSO and Naïve Bayes; PSO and C4.5.

Data Collection

The data collection was carried out using open data acquisition techniques. It is a data collection that contains searching, downloading and organizing datasets that are publicly and openly available through the Kaggle that provides datasets for analysis and predictive model development. Data collection involved searching for relevant datasets to ensure that the data used are appropriate for predicting osteoporosis. For access to the osteoporosis dataset in this research, please refer to the following link : https://www.kaggle.com/code/docxian/osteoporosis-risk-prediction/input. The osteoporosis dataset consists of 1.958 records with 15 regular attributes and 1 special attribute as a label. The dataset provides a sufficient foundation for building a predictive model, as it offers a reasonable sample size to capture pattern related to osteoporosis. The osteoporosis dataset is valid as it is complete, with no missing values, which ensures no additional data cleaning required. Furthermore, the dataset is relevant for predicting osteoporosis as it includes key risk factors such as age, gender, and other medical history, it can represent the condition effectively. With established statistical models, like Naïve Bayes and C4.5, can perform well with moderate-sized datasets, ensuring reliable predictions despite the dataset's size. The details of dataset are described in table 1.

Table 1. Osteoporosis Dataset

No.	Attribute	Description
1.	Id	Unique Identifier
2.	Age	Individual age in years
3.	Gender	Male, Female
4.	Hormonal Change	Normal, Postmenopausal
5.	Family History	Yes, No
6.	Race/Ethnicity	Caucasia, Africa-Amerika, Asia,
7.	Body Weight	Normal, Underweight
8.	Calcium Intake	Low, Adequate
9.	Vitamin D Intake	Insufficient, Sufficient
10.	Physical Activity	Sedentary, Active
11.	Smoking	Yes, No
12.	Alcohol Consumption	None, Moderate
13.	Medical Condition	Rheumatoid arthritis, None, Hyperthyroidism
14.	Medication	Corticosteroid, None
15.	Prior Factor	Yes, None
16.	Osteoporosis	0, 1

Table 1 contains lifestyle data including medical history, physical activity, smoking, and alcohol intake[25], [26]. As well as demographics information with and without osteoporosis. It is intended to support research in analyzing and predicting osteoporosis risk.

Data Pre-Processing

The initial step of preprocessing in predicting osteoporosis using Naïve Bayes and Decision Tree C4.5 is data cleaning. To ensure the accuracy and completeness of the data, the osteoporosis dataset was analyzed. During the initial stage of the analysis, it was confirmed that there were no empty, missing, or incomplete entries within the dataset. An overview of the initial analysis indicates that the dataset in a good condition for further analysis, as there are no missing or incomplete data entries. This confirms that no additional steps are necessary to address missing data. The subsequent stage involves building model using RapidMiner, specifically implementing Naive Bayes and C4.5 algorithm.

Data Validation

The data validation stage is designed to objectively assess the performance of the model and its ability to generalize to unseen data. To achieve this, the split data validation and cross-validation methods was employed. The split data validation, osteoporosis dataset comprising 1.958 records was divided into two subsets: 80% of the data (1.566 records) was allocated for model training, while the remaining 20% (392) records) war reserved for testing. This approach ensures that the model's effectiveness is evaluated on a separate test set, simulating its performance in real-world scenario. In contrast, cross-validation divided the dataset into *k* equal folds, where the model is trained on *k-1* folds and tested on the remaining fold. This process is repeated multiple times to ensure that each fold used for testing at least once, providing more comprehensive evaluation of the model's generalization ability.

Naive Bayes

As a machine learning algorithm, Naïve Bayes works according to Bayes' theorem, which relies on the conditional probability and maximum probability of an event[27]. The Naive Bayes calculation employed equation 1 as follows:

$$P(a/y) = \frac{P(y/a)P(a)}{P(y)} \quad (1)$$

The Naive Bayes calculation employed equation 1 as follows:

P(a/y) : the probability of event a given that y is true (posterior probability)
P(y/a) : the probability of event y occurring given that a is true
P(a) : the prior probability of event a
P(y) : the overall probability of event y happening

This method allows us to update our beliefs about event a based on the observation of y[28] following Bayes' theorem. Calculating probabilities in Naïve Bayes involved in five stages. The first stage entailed reading the training data that has been input into the database. The second stage involves calculating the prior probability, which represent the likelihood of class occurrence without considering specific attributes. The third stage computes the probability of each class, assuming that each attribute is independent of the other. The fourth stage involve selecting the class with the having the greatest of likelihood which indicates the likelihood of each class given the attributes. The final stage is to derive the classification result base on the probabilities[29].

C4.5

C4.5 workflow starts with building a decision tree from the given training data. This process involves selecting the most informative attributes as nodes on the tree, the variable with the having greatest gain value will be selected as the attribute that become the root of the tree[30]. Following attribute selection, smaller subsets of the training data are created based on the attribute values. Every data subset goes through this recursive procedure until all the data subsets are categorized into the same class or until a decision tree is build and specified having criteria are satisfied[31]. In a decision tree, nodes represent attributes, branches represent result, and leaves represent decisions[32].

In C4.5, the calculation begins with determining entropy using equations 2 and 3, continues through to equation 6.

$$Entropy(S) = \sum n - pi * log2\, pi \quad (2)$$

$$Entropy(S) = \sum n - pi * log_2\, pi \quad (3)$$

$$Entropy\,(S,A) = Etropy\,(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\,(Si) \quad (4)$$

$$RasioGain\,(s,j) = \frac{Gain\,(s,j)}{SplitInfo(s,j)} \quad (5)$$

$$SplitInfo(s,j) = \sum_{i=1}^{k} p(v\log_2 p(vi|s) \qquad (6)$$

Equation 4 calculates information gain, a measure used to assess how effectively attribute reduces uncertainty in dataset S. This measure quantifies the reduction uncertainly (entropy) when dataset S is partitioned based on attribute A. First, we determine the entropy of dataset S, which reflect the level of uncertainty or disorder within the dataset. Next, dataset S is divided into n subsets, Si, according to the values of attribute A [33].

For each subset Si, we calculate its relative size ISiI/ISI and multiply it by its entropy, Entropy Si, then sum these values cross all subset. Information Gain is then computed as the difference between the initial entropy of dataset S and the weighted sum of the entropies of the subset Si. Equation 5 represent the gen ratio, which evaluated how well he attribute divides the data while accounting for the number of resulting divisions. Equation 6 calculated the split information, which measures the extent to with dataset S is partitioned into smaller parts based on the values of attribute A.

Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is one of the efficacy methods influenced by the behavior of a group in universe, especially the movement and interaction of a group of particles in search of the best possible solution. In particle swarm optimization (PSO), a set of particles are considered as agents moving within the range of possible solutions. Each particle has a location and velocity that changes over time, and they move in the search space with goal of finding the best solution. The interaction among particles in Particle Swarm Optimization (PSO) is decides by means of their ability to share information about the location of optimal solution found by other particles with the population[34]. In a very short amount of time, PSO may effectively search the targeted space and identify a close to ideal solution [35].

The Particle Swarm Optimization (PSO) method begins by initializing the position (Ci) and velocities (Vi) of the particles within the swarm. Next step, it evaluates the objective function value for each particle ($f(Ci)$). The algorithm then determines the initial personal best $p(best)$ and global best ($gbest$). The velocity is updated using a specific an equation, followed by updating the position of each particle. The objective function s re-evaluated. If the new value improves upon the previous best, the personal best is updated. This process continues until the maximum number of iterations reached, at which point the algorithm stop, otherwise it returns to the updating the velocity in the particle[36].

Particle Swarm Optimization or PSO can be applied to enhance performance of Naïve Bayes and C4.5 models in a several specific ways, one of which is through feature selection. In C4.5, PSO helps identify most relevant features to be used in C4.5 algorithm, improving the tree's structure and reducing complexity. By selecting only the significant feature, the model can achieve higher accuracy and better interpretability. In Naïve Bayes model, PSO can be used to select features that the most contribute to the classification performance, enhancing the model predictive power. The feature selection process using PSO depicted in figure 2.

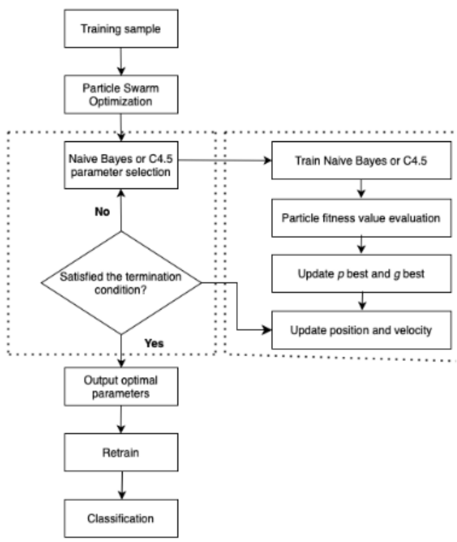

Figure 2. PSO process for enhancing Naïve Bayes and C4.5 models

Figure 2 outlines the process of using PSO to enhance model training for Naïve Bayes and C4.5. it begins with a set of training samples followed by the initialization of a swarm of particles representing potential parameter solutions. PSO select parameters based on current particle positions and trains the model with these selected values. The model's performance is then evaluated using fitness function to determine accuracy. The particles update their positions based on both global and personal best fitness values. This iterative process continues until a termination condition is met, at which point the optimal parameters are outputted. Finally, the model's is retrained with these parameters, enabling improves predictions.

Confusion Matrix

A crucial technique for assessing model performance in data processing is called confusion matrix, containing metrics used to assess the effectiveness of a model's predictions with the true values of the observed data[37]. Confusion matrix has four cells that represent the four possible outcomes of the classification process: the model correctly predicts the positive class (TP); the model incorrectly predicts the positive class when it is actually negative (FP); the model correctly predicts the negative class (TN); and the model incorrectly predict the negative class when it actually positive (FN)[38].

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (7)$$

$$Precision = TP/(FP + TP) \quad (8)$$

$$Recall = \frac{TP}{FN + TP} \quad (9)$$

Equation 7 calculates accuracy by dividing the number of corrections (TP and TN), by the total amount of observed data. Equation 8 calculates precision to measure the ratio of correct positive prediction (TP) to the sum of positive predictions generated. Equation 9 calculates sensitivity to measure how well the model detects all instances that belongs to the positive class[39].

RESULTS AND DISCUSSION

1. Split Validation

The first method applied in this study was split validation, where the performance of various model was assessed, including Naïve Bayes, C4.5, PSO and Naïve Bayes, also PSO with C4.5 to determine which approach yielded better result in predicting osteoporosis.

Naïve Bayes Algorithm

The first model was conducted using Naive Bayes, the osteoporosis dataset was taken for processing into RapidMiner, as shown in figure 3.

Figure 3. The Naive Bayes process view

Figure 3 showcases the application of the Naive Bayes algorithm using RapidMiner. It involves retrieving osteoporosis dataset, this initial step is crucial as it provides the data necessary for the subsequent analysis. The data split training set and testing set. To build the model, the Naive Bayes algorithm is implemented to the training data. Subsequently, the trained model is used to make a prediction on the testing data. Finally, performance metrics such as accuracy, precision and recall are calculated to assess the model's effectiveness. The Naive Bayes algorithm values presented on table 2. Three evaluation metrics such as accuracy, precision, and recall derived from calculation (1).

Table 2. The Naive Bayes test value

No.	Description	Naive Bayes
1.	Accuracy	82,65 %
2.	Precision	91.03 %
3.	Recall	72,45 %

The Naive Bayes model demonstrates at table 2 are solid performance. However, the recall rate is lower, suggesting the model misses some positive instance. Overall, the Naïve Bayes proves to be a reliable and efficient classifier with strength in precision, though there is room for improvement in recall[40].

C4.5 Algorithm

The second model was conducted using C4.5 algorithm, as seen in figure 4.

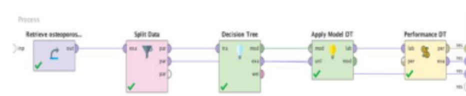

Figure 4. The C4.5 process view

Figure 4 showcases the implementation of the C4.5 algorithm using RapidMiner. It involves retrieving osteoporosis dataset, this initial step is crucial as it provides the data necessary for the subsequent analysis. The data divided into training set and testing set. To build the model, the method is utilized to the training data. Subsequently, the trained model is used to make a prediction on the testing data. Finally, performance metrics such as accuracy, precision and recall are calculated to assess the model's effectiveness. The evaluation metrics of this algorithm presented on table 3.

Table 3. The C.45 test value

No.	Description	Decision Tree
1.	Accuracy	91,07 %
2.	Precision	97,63 %
3.	Recall	84,18 %

The C4.5 algorithm are presented on Table 3 shows strong performance metric in the model deployment model. It achieved an impressing accuracy, indicating a high rate of correct classification. The model is highly reliable when predicting positive outcomes. The recall rate is a significant improvement over Naïve Bayes model, suggesting that C4.5 algorithm is effectively identified most positive instances. The C4.5 model demonstrates high accuracy, precision, and recall.

PSO and Naïve Bayes

PSO and Naïve Bayes modelling show in figure 5.

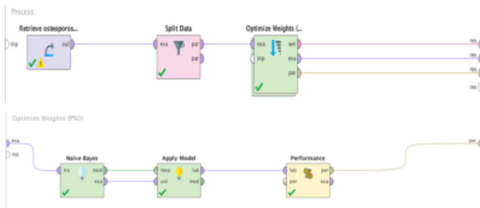

Figure 5. PSO and Naïve Bayes process view

The workflow begins with the retrieve ensembles module for feature extraction, followed by the split data module that divides the data into subsets. The optimize weight module to optimizes the model's weight using PSO. The optimized data is applied to the Naïve Bayes for classification. Finally, the performance module evaluates how well performing models by computing various metric. This process aims to optimize the Naïve Bayes model's accuracy through weight adjustment via PSO. Test results can be observed in table 4.

Table 4. PSO and Naive Bayes test value

No.	Description	Naive Bayes
1.	Accuracy	83,67 %
2.	Precision	93,42 %
3.	Recall	72,45 %

The provided data on table 4 involves that after applying PSO for weight optimization, the Naïve Bayes models show improvement in accuracy and precision, while maintaining the same recall. This suggests that while PSO optimization has enhanced the model's overall correctness and precision, making it more effective in identifying true positive cases while the recall remains unchanged.

Table 5. Attribute weights test value

No.	Description	Weight	Attribute
1.	Most Influential	0,636 – 1,000	Gender, Smoking
2.	Medium Influence	0,242 – 0,586	Age, Hormonal Changes, Physical Activity
3.	Less Influence	0,033 – 0,334	Medications, Calcium Intake
4.	No Influence	0	Prior Fractures, Medical Conditions, Alcohol Consumption, Vitamin D Intake, Body Weight, Race/Ethnicity, Family History
5.	Irrelevant	-	Id

Table 5 described the attribute weight obtained from analyzing the osteoporosis dataset using PSO and Naïve Bayes. It shows that the most influential attributes are gender and smoking.

PSO and C4.5

PSO and C4.5 modelling show in figure 6.

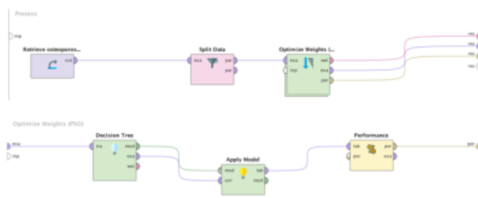

Figure 6. PSO and C4.5 process view

In this workflow, data is first collected and preprocessed. Then feature extraction is performed to identify key attributes. The Particle PSO is applied to efficacy the model's parameters. Following this, C4.5 algorithm creates a model by segmenting the data based on significant features, leading to classification or predictors. The model's effectiveness is assess using metrics such as accuracy, precision, and recall. The test value can be seen in table 6.

Table 6. PSO and C4.5 test value

No.	Description	C4.5
1.	Accuracy	96,17 %
2.	Precision	95,02 %
3.	Recall	97,45 %

Table 6 demonstrates outstanding performance in the model deployment model. By performing this model, we seek to assess how a feature selection influences the model performance and to establish the effectiveness of the C4.5 in pinpointing the most critical attribute for accurate prediction[41]. The attribute weights of the test value are shown in table 7.

Table 7. Attribute weights test value

No.	Description	Weight	Attribute
1.	Most Influential	0,939 – 1,000	Prior Fracture, Age, Hormonal Changes
2.	Medium Influence	0,455 – 0,684	Physical Activity, Calcium Intake, Smoking, Medical Conditions
3.	Less Influence	0,200 – 0,280	Family History, Body Weight, Race/Ethnicity
4.	No Influence	0	Medication, Alcohol Consumption, Vitamin D Intake, Gender
5.	Irrelevant	-	Id

Table 7 is described that prior fracture, age, hormonal change are most influential attributes. The comparison of attribute weight tables relative to Naïve Bayes and C4.5 reveals notable differences. This difference is due to their distinct methodologies. Naïve Bayes assumes feature independence, which can limit its performance when features are correlated[42]. The C4.5 do not rely on this assumption and are better at capturing complex feature interactions. Additionally, the C4.5 handled non-linier relationships more effectively, making them more adaptable to varied data pattern[43]. In addition to comparing attributes weights, the evaluation matrix results from model also compared. The results of each model test are compiled into a table containing test comparison values to facilitate analysis and evaluation of model performances. The test comparison values are displayed in table 8.

Table 8. Test comparison values on split validation method

No.	Model	Evaluation Matrix		
		Accuracy	Precision	Recall
1.	Naive Bayes	82,65%	91,03%	72,45%
2.	C4.5	91,07%	97,63%	84,13%
3.	PSO + Naive Bayes	83,67%	93,42%	72,45%
4.	PSO + C4.5	96,17%	95,02%	97,45%

In the term of effectiveness models based on table 8, shows competitive result in data classification. The C4.5 outperforms Naïve Bayes in accuracy, precision, and recall, both with and without PSO. Without PSO, C4.5 achieved an accuracy of 91,07% compared to Naïve Bayes 82,65%. With PSO, C4.5's accuracy increased to 96,17% while Naïve Bayes improved slightly to 83,67%. Precision for C4.5 was 97,63% without PSO and 95,02% with PSO, it still higher than Naïve Bayes precision, which increased from 91,03% to 93,42% with PSO. Recall for C4.5 was 81,48% without PSO, whereas Naïve Bayes remained consistent at 72,45%.

2. Cross Validation

The second method applied is cross-validation. The testing process follows the same steps as in split data validation, where the model is trained on a portion of the data and tested on a separate portion. In this experiment, four combinations were evaluated: C4.5 with cross-validation, Naïve Bayes with cross-validation, C4.5 with cross-validation and PSO, and Naïve Bayes with cross-validation and PSO. These combinations were used to access the performance of each model and technique, comparing their accuracy and generalization capabilities on the osteoporosis dataset, while ensuring that the model is not overly reliant on single training-test split, which may be sensitive to data distribution.

Naïve Bayes with cross-validation

The first model applied was Naïve Bayes with cross-validation. The data was tested starting from $k = 1$ and increasing up to $k = 10$. Throughout this process, different values of k were used to evaluate the model's performance in term of accuracy and generalization. The aim was to identify the most suitable value of k that would provide the best balance between training and testing data. The test result displayed in table 9.

Table 9. Naïve Bayes with cross-validation value

No.	Description	Naïve Bayes
1.	Accuracy	85,45%
2.	Precision	94,04%
3.	Recall	75,69 %

From table 9 shows that after testing each fold, the model achieved its optimal performance at $k = 9$, indicating that this value provided the most reliable and accurate result for the dataset is 85,45%.

C4.5 with cross-validation

The second model applied was C.45 with cross-validation. The test result displayed in table 10.

Table 10. C4.5 with cross-validation value

No.	Description	Naïve Bayes
1.	Accuracy	90,40%
2.	Precision	97,94%
3.	Recall	82,53%

Table 10 shows that when C4.5 with cross-validation model was applied, the optimal result achieved at $k = 8$, yielding an accuracy of 90,40%. This represents an improvement of 4,95% compared to the Naïve Bayes model. Additionally, the precision and recall showed by the C4.5 model were higher than those Naïve Bayes model.

PSO and Naïve Bayes with cross-validation

The third model applied was PSO and Naïve Bayes with cross-validation. The test result displayed in table 11.

Table 11. PSO and Naïve Bayes value

No.	Description	PSO and Naïve Bayes
1.	Accuracy	86,06%
2.	Precision	95,37%
3.	Recall	75,79%

Table 11 shows that when PSO was applied to the Naïve Bayes model, there was a noticeable improvement in evaluation metrics, including accuracy, precision, and recall. This optimal performance was achieved at $k = 4$ and $k = 5$, indicating that the integration of PSO enhanced the model's ability to predict osteoporosis effectively. From the test result, the attribute weights are outlined in table 12.

Table 12. Attribute weights test value

No.	Description	Weight	Attribute
1.	Most Influential	1,000	Age, Calcium Intake, Vitamin D Intake, Physical Activity, Alcohol Consumption,
2.	Less Influence	0,0092	Prior Fracture
3.	No Influence	0	Gender, Hormonal Changes, Family History, Race/Ethnicity, Body Weight, Smoking, Medical Condition, Medication,
4.	Irrelevant	-	Id

The table above presents a list of attributes along with their corresponding weights. Key attributes such as Age, Calcium intake, Vitamin D, Physical activity, and Alcohol consumption all have the highest weight of 1, indicating their strong relevance in the model or dataset. Meanwhile, Prior Fractures has a much lower weight of 0.092, showing less significance in comparison. The attributes Id, Gender, Hormonal history, Family history, Race/Ethnicity, Body weight, Smoking, Medical history, and Medications all have a weight of 0, suggesting that they were not considered relevant or influential in this analysis.

PSO and C4.5 with cross-validation

The fourth model applied was PSO and C4.5 with cross-validation. The test result displayed in table 13.

Table 13. PSO and C.45 value

No.	Description	PSO and C4.5
1.	Accuracy	91,16%
2.	Precision	99,39%
3.	Recall	82,84%

The table 13 described that when PSO and C4.5 were applied, there was a slight improvement across all metrics, though the increase was not particular significant. The most noticeable gain was in precision, which rose by 1,45% from 97,94% to 99,39%. The optimal performance was achieved at $k = 7$, indicating a modest improvement in the model's ability to correct classify positive cases, though overall effectiveness showed only minor enhancement. The attribute weights are outlined in table 14.

Table 14. Attribute weights test value

No.	Description	Weight	Attribute
1.	Most Influential	0,714-1,000	Age, Body Weight, Alcohol Consumption, Medication, Prior Fractures
2.	Less Influence	0,328-0,428	Physical Activity, Gender
3.	No Influence	0	Hormonal Changes, Family History, Race/Ethnicity, Calcium Intake, Vitamin D Intake, Smoking, Medical Condition,
4.	Irrelevant	-	Id

The table 14 categorizes attributes based on their weight and influence on the model performance. It identifies age, body weight, alcohol consumption, medication and prior fractures as the most influential, with weights ranging from 0,714 to 1,000, indicating their significant impact on predictions. Physical Activity and Gender fall into the less influential category, with weights between 0.328 and 0.428, suggesting a moderate contribution to the model's predictive power. In contrast, other attributes exhibit no influence, with weights of 0. Lastly, the attribute Id is classified as irrelevant, indicating it does not contribute to the analysis. After applying the four models, conclusions can be drawn from the evaluation result present on table 15.

Table 15. Test comparison values on cross-validation method

No.	Model	Evaluation Matrix		
		Accuracy	Precision	Recall
1.	Naive Bayes	85,45%	94,04%	75,79%
2.	C4.5	90,40%	97,94%	82,53%
3.	PSO + Naive Bayes	86,06%	95,37%	82,53%
4.	PSO + C4.5	91,16%	99,39%	82,84%

The table summarizes the performance metrics of each model, highlighting their respective strengths and weaknesses in predicting osteoporosis. When comparing between table 8 and table 15, it can be concluded that the C4.5 model, especially when optimized with PSO, exhibits superior performance in predicting osteoporosis. It achieved the highest accuracy of 96,17 in split data validation and 91,16% in cross-validation compared to Naïve Bayes. In contrast, the Naïve Bayes model improved its accuracy slightly constantly showed lower performance, with accuracy rates of 83,67% and 86,06,45% in the respective validation methods. Although incorporating PSO into Naïve Bayes model improves its accuracy slightly, it remained inferior for both the standalone C4.5 and PSO with C4.5 models. Additionally, combination of C4.5 with PSO is more reliable in improving model accuracy, precision and recall, providing better predictive performance across different validation methods compared to Naïve Bayes[44][45], there by confirming its effectiveness in osteoporosis prediction.

In the context of osteoporosis detection, Particle Swarm Optimization (PSO) proves to be a valuable tool for feature selection. By efficiently and optimizing relevant features, PSO enhances model performance in identifying predictors of osteoporosis risk. The ability of PSO to refine feature selection allows for a more accurate understanding of which attribute are most influential. For instance, Age and Hormonal Change emerged as important for both models. Age is critical factor as bone density natural decreases over time, increasing the risk fractures[46]. Hormonal changes, particularly in postmenopausal women, lead to a decline in estrogen levels, which is essential for bone health[47]. This suggests that both algorithms agree that age and hormonal changes are significant indicator in osteoporosis risk. In addition, PSO helped recognize attributes such as Physical Activity and Calcium Intake, which, despite having varying levels of influence on each model, were consistently considered relevant. Smoking has been linked to reduced bone mass and slower healing of fractures. Regularly physical activity, on the other hand, is beneficial as it helps strengthen bones and improve balance, thereby reducing the risk of fails and fractures.

One possible reason for C4.5 performance could be its ability to handle non-linier relationship and complex decision boundaries more effectively than Naïve Bayes, which assumes independence between features (the Naïve Bayes assumption). In real-world osteoporosis prediction, the relationships between risk factors (e.g., age, gender, hormonal changes) are often non-linear and interdependent, making C4.5 better suited to capture these interactions. Furthermore, PSO's role in optimizing the decision tree structure may provide further advantages by enhancing feature selection and tuning parameters to maximize predictive performance. Based on the result of the study conducted with the PSO and Naïve Bayes, PSO, utilizing he principles of Bayes' theorem, was able to reduce the initial 15 features to 7 significant features that influence osteoporosis. In contrast, when PSO was combine with the C4.5 model, its successfully selected until 10 influential features. This indicates that the approach of integrating PSO with C4.5 may be more effective 4 identifying risk of factors for osteoporosis compared to the Naïve Bayes algorithm.

CONCLUSION

The comparative of test results indicates that C4.5 algorithm is the most effecti41predicting osteoporosis, as evidenced by its superior accuracy, precision, and recall compared to the Naïve Bayes, this trend is consistent in both the split data validation and cross-validation methods, where C4.5 consistently outperformed Naïve Bayes across various metric. In additionally, the use of PSO contributes to improving the reliability and interpretability of the predictive models for osteoporosis. This research concluded that age, hormonal change, smoking, and physical activity significantly influence develop osteoporosis. These finding underscore the importance of addressing these factors to mitigates the risk of osteoporosis. This allows for preventive measures to be implemented effectively. Preventive actions include lifestyle modifications such as increasing 18ysical activity to strengthen bones, ensuring adequate intake of calcium and vitamin D, quitting smoking to improve bone health, and managing hormonal changes through medical consultation. These step help in reducing the risk of developing osteoporosis, thereby improving overall bone health and preventing fractures.

To further enhance prediction accuracy, one alternative method that could be implemented for predicting osteoporosis dataset is the use of ensemble learning techniques such as Random Forest or Gradient Boosting. These methods can effectively handle complex interaction between features, which may be present in osteoporosis risk factors. Regarding PSO, the main difficulties encountered may relate to time complexity, which could be a limitation, as PSO might require a substantial number of iterations to find an optimal solution, especially when working with more complex models. Additionally, the study could face limitations such as a small dataset size, which affects the model's ability to generalize to unseen data. With fewer data points, models are more prone to overfitting, where they perform well on the training data but fail to generalize in real-world applications.

ACKNOWLEDGMENT

The authors express gratitude to Kaggle for providing the necessary dataset and appreciate the support from the Faculty of Computer Science, Darmajaya Institute of Informatics and Business in this study.

REFERENCES

[1] V. Sromova, D. Sobola, and P. Kaspar, "A Brief Review of Bone Cell Function and Importance," Cells, vol. 12, no. 2576, pp. 1–31, Nov. 2023, doi: 10.3390/cells12212576.

[2] M. Ashrafi, F. Gholamian, and M. Doblare, "A comparison between the effect of systemic and coated drug delivery in osteoporotic bone after dental implantation," *Medical Engineering and Physics1*, vol. 107, pp. 1–12, Sep. 2022, doi: 10.1016/j.medengphy.2022.103859.

[3] Krisztina et al., "Bone Loss in Diabetes Mellitus: Diaporosis," *Int J Mol Sci*, vol. 25, no. 13, pp. 1–20, Jul. 2024, doi: 10.3390/ijms25137269.

[4] Imen et al., "Fast diagnostic of osteoporosis based on hair analysis using LIBS technique," *Med Eng Phys*, vol. 103, pp. 1–5, May 2022, doi: 10.1016/j.medengphy.2022.103798.

[5] S. H. Ahn *et al.*, "Osteoporosis and osteoporotic fracture fact sheet in Korea," *J Bone Metab*, vol. 27, no. 4, pp. 281–290, Nov. 2020, doi: 10.11005/JBM.2020.27.4.281.

[6] S. Mondal *et al.*, "A computational analysis of a novel therapeutic approach combining an advanced medicinal therapeutic device and a fracture fixation assembly for the treatment of osteoporotic fractures: Effects of physiological loading, interface conditions, and fracture fixation materials," *Med Eng Phys*, vol. 114, pp. 1–13, Apr. 2023, doi: 10.1016/j.medengphy.2023.103967.

[7] M. Chandran *et al.*, "Prevalence of osteoporosis and incidence of related fractures in developed economies in the Asia Pacific region: a systematic review," *Osteoporosis International*, vol. 34, no. 6, pp. 1037–1053, Jun. 2023, doi: 10.1007/s00198-022-06657-8.

[8] F. Borgström *et al.*, "Fragility fractures in Europe: burden, management and opportunities," *Arch Osteoporos*, vol. 15, no. 1, pp. 1–21, Dec. 2020, doi: 10.1007/s11657-020-0706-y.

[9] Y. Yang, S. Wang, and H. Cong, "Association between parity and bone mineral density in postmenopausal women," *BMC Womens Health*, vol. 22, no. 1, pp. 1–8, Dec. 2022, doi: 10.1186/s12905-022-01662-9.

[10] A. Aibar-Almazán et al., "Current Status of the Diagnosis and Management of Osteoporosis," *Int J Mol Sci*, vol. 23, no. 16, pp. 1–27, Aug. 2022, doi: 10.3390/ijms23169465.

[11] O. Gómez *et al.*, "Diagnostic, treatment, and follow-up of osteoporosis—position statement of the Latin American Federation of Endocrinology," *Arch Osteoporos*, vol. 16, no. 114, pp. 1–15, Dec. 2021, doi: 10.1007/s11657-021-00974-x.

[12] M. A. de Oliveira *et al.*, "Osteoporosis Screening: Applied Methods and Technological Trends," *Med Eng Phys*, vol. 108, pp. 1–14, Oct. 2022, doi: 10.1016/j.medengphy.2022.103887.

[13] Katherine et al., "Bone mineral density: Clinical relevance and quantitative assessment," *Journal of Nuclear Medicine*, vol. 62, no. 4, pp. 446–454, Apr. 2021, doi: 10.2967/jnumed.120.256180.

[14] N. Deshpande *et al.*, "Alternatives to DEXA for the assessment of bone density: a systematic review of the literature and future recommendations," *J Neurosurg Spine*, vol. 38, no. 4, pp. 436–445, Apr. 2023, doi: 10.3171/2022.11.SPINE22875.

[15] A. Ariani and S. Samsuryadi, "Classification Of Kidney Disease Using

Genetic Modified KNN And Artificial Bee Colony Algorithm," *SINERGI*, vol. 25, no. 2, pp. 177–184, Feb. 2021, doi: 10.22441/sinergi.2021.2.009.

[16] Yolanda et al., "Service quality dealer identification: the optimization of K-Means clustering," *Sinergi (Indonesia)*, vol. 27, no. 3, pp. 433–442, 2023, doi: 10.22441/sinergi.2023.3.014.

[17] A. A. Almazroi et al., "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," *IEEE Access*, vol. 11, pp. 61646–61659, 2023, doi: 10.1109/ACCESS.2023.3285247.

[18] Pareza *et al.*, "Improving Performance of KNN and C4.5 using Particle Swarm Optimization in Classification of Heart Diseases," *JURNAL RESTI (Rekayasa Sist. Teknol. Informasi)*, vol. 8, no. 3, pp. 333–339, 2024, doi: 10.29207/resti.v8i3.5710.

[19] N. Hayatin, G. I. Marthasari, and L. Nuraini, "Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization," *Jurnal Online Informatika*, vol. 5, no. 1, pp. 81–88, 2020, doi: 10.15575/join.v5i1.558.

[20] Dedi Saputra et al., "A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction," *International Journal of Advances in Data and Information Systems85*, vol. 1, no. 2, pp. 84–95, Nov. 2021, doi: 10.25008/ijadis.v2i2.1221.

[21] Alvina Felicia Watratan, Ema Utami, and Anggit Dwi Hartanto, "Comparison of Naive Bayes and PSO-Based Naive Bayes Algorithms for Prediction of Covid-19 Patient Recovery Data in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 4, pp. 809–816, Aug. 2023, doi: 10.29207/resti.v7i4.4893.

[22] A. Waluyo et al., "Data Mining Optimization Uses C4.5 Classification and Particle Swarm Optimization (PSO) In The Location Selection Of Student Boardinghouses," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, pp. 1–9, Jul. 2020, doi: 10.1088/1757-899X/874/1/012024.

[23] Y. K. Putra, Fathurrahman, and M. Sadali, "Comparison of Pso-Based Naive Bayes and Naive Bayes Algorithm in Determining the Feasibility of Bumdes Credit," *J Phys Conf Ser*, vol. 1539, no. 1, pp. 1–6, Jul. 2020, doi: 10.1088/1742-6596/1539/1/012030.

[24] T. S. Lestari, I. Ismaniah, and W. Priatna, "Particle Swarm Optimization for Optimizing Public Service Satisfaction Level Classification," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 1, pp. 147–155, Mar. 2024, doi: 10.23887/janapati.v13i1.69612.

[25] Varada et al., "A Decision Support System For Osteoporosis Risk Prediction Using Machine Learning And Explainable Artificial Intelligence," *Heliyon*, vol. 9, no. 12, pp. 1–19, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22456.

[26] Y. Wu et al., "Construction Of Predictive Model For Osteoporosis Related Factors Among Postmenopausal Women On The Basis Of Logistic Regression And Bayesian Network," *Prev Med Rep*, vol. 35, no. 102378, pp. 1–8, Oct. 2023, doi: 10.1016/j.pmedr.2023.102378.

[27] A. Irwanto and L. Goeirmanto, "Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia using Naïve Bayes and XGboost Classifier Algorithm," *Sinergi (Indonesia)*, vol. 27, no. 2, pp. 145–152, 2023, doi: 10.22441/sinergi.2023.2.001.

[28] Hong et al., "Improved Naive Bayes Classification Algorithm For Traffic Risk Management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s13634-021-00742-6.

[29] V. Jackins et al., "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, May 2021, doi: 10.1007/s11227-020-03481-x.

[30] F. Riandari and S. Defit, "The Application of C4.5 Algorithm for Selecting Scholarship Recipients," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 13, no. 1, pp. 11–21, Feb. 2022, doi: 10.21512/comtech.v13i1.7307.

[31] Akshansh *et al.*, "Machine learning-assisted pattern recognition algorithms for estimating ultimate tensile strength in fused deposition modelled polylactic acid specimens," *Materials Technology1*, vol. 39, no. 1, pp. 1–12, 2024, doi: 10.1080/10667857.2023.2295089.

[32] Madhusree et al., "Comparative Analysis of Machine Learning Methods to Detect Chronic Kidney Disease," in *Journal of Physics: Conference Series*, IOP

Publishing Ltd, Jun. 2021, pp. 1–12. doi: 10.1088/1742-6596/1911/1/012005.

[33] M. Sutrisno, J. K. Rambe, and A. Asruddin, "The Implementation of C4.5 Algorithm for Determining the Department of Vocational High School," *Jurnal Riset Informatika*, vol. 5, no. 2, pp. 211–218, Mar. 2023, doi: 10.34288/jri.v5i2.516.

[34] M. Mansur and M. R. Djalal, "Using Particle Swarm Optimization for Power System Stabilizer and energy storage in the SMIB system under load shedding conditions," *Sinergi (Indonesia)*, vol. 27, no. 3, pp. 423–432, 2023, doi: 10.22441/sinergi.2023.3.013.

[35] A. Khan *et al.*, "Adaptive Filtering: Issues, Challenges, and Best-Fit Solutions Using Particle Swarm Optimization Variants," *Sensors*, vol. 23, no. 18, pp. 1–28, Sep. 2023, doi: 10.3390/s23187710.

[36] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, doi: 10.1007/s11831-021-09694-4.

[37] F. Fersellia, E. Utami, and A. Yaqin, "Sentiment Analysis of Shopee Food Application User Satisfaction Using the C4.5 Decision Tree Method," *Sinkron*, vol. 8, no. 3, pp. 1554–1563, Jul. 2023, doi: 10.33395/sinkron.v8i3.12531.

[38] K. Riehl, M. Neunteufel, and M. Hemberg, "Hierarchical Confusion Matrix For Classification Performance Evaluation," *J R Stat Soc Ser C Appl Stat*, vol. 72, no. 5, pp. 1394–1412, Nov. 2023, doi: 10.1093/jrsssc/qlad057.

[39] Heni et al., "Performance Evaluation Of Feature Selections On Some ML Approaches For Diagnosing The Narcissistic Personality Disorder," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 2, pp. 1383–1391, Apr. 2024, doi: 10.11591/eei.v13i2.6717.

[40] A. Afdhaluzzikri, H. Mawengkang, and O. S. Sitompul, "Perfomance Analysis Of Naive Bayes Method With Data Weighting," *SinkrOn*, vol. 7, no. 3, pp. 817–821, Jul. 2022, doi: 10.33395/sinkron.v7i3.11516.

[41] F. Saadi et al., "An Effective Prediction Approach for the Management of Children Victims of Road Accidents," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. Inpress, no. Inpress, pp. 1–11, 2024, doi: 10.9781/ijimai.2024.02.001.

[42] A. F. A. H. Alnuaimi and T. H. K. Albaldawi, "An overview of machine learning classification techniques," *BIO Web Conf*, vol. 97, pp. 1–24, Apr. 2024, doi: 10.1051/bioconf/20249700133.

[43] A. R. Lendra and D. Firdaus, "Implementation Of C4.5 Algorithm TO Assist In The Selection Of Floor Construction Projects," *International Journal Information System and Computer Science (IJISCS)*, vol. 4, no. 3, pp. 153–161, 2020.

[44] Okfalisa et al., "Forecasting Company Financial Distress: C4.5 And Adaboost Adoption," *Engineering and Applied Science Research*, vol. 49, no. 3, pp. 300–307, 2022, doi: 10.14456/easr.2022.31.

[45] Alam et al., "Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes," *SINKRON*, vol. 8, no. 4, pp. 2641–2650, Oct. 2023, doi: 10.33395/sinkron.v8i4.12998.

[46] C. A. Inderjeeth and K. A. Inderjeeth, "Osteoporosis In Older People," *Journal of Pharmacy Practice and Research*, vol. 51, no. 3, pp. 265–274, Jun. 2021, doi: 10.1002/jppr.1743.

[47] Dave B et al., "Identification of the Risk Factors Associated with Low Bone Density in Peri- and Early Postmenopausal Women," *Dietetics*, vol. 3, no. 1, pp. 75–86, Mar. 2024, doi: 10.3390/dietetics3010007.

SINERGI_New_Journal_TemplateZulfi-1732500079679

ORIGINALITY REPORT

15%	10%	10%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Diponegoro Student Paper	2%
2	fastercapital.com Internet Source	1%
3	Submitted to Southern New Hampshire University - Continuing Education Student Paper	1%
4	iopscience.iop.org Internet Source	1%
5	Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023 Publication	1%
6	www.researchsquare.com Internet Source	1%
7	Submitted to University of Northumbria at Newcastle Student Paper	1%
8	Submitted to National College of Ireland Student Paper	<1%

9	dataaspirant.com Internet Source	<1 %
10	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
11	Submitted to Universitas Mercu Buana Student Paper	<1 %
12	ejournal.undip.ac.id Internet Source	<1 %
13	Arvind Dagur, Dhirendra Kumar Shukla, Nazarov Fayzullo Makhmadiyarovich, Akhatov Akmal Rustamovich, Jabborov Jamol Sindorovich. "Artificial Intelligence and Information Technologies", CRC Press, 2024 Publication	<1 %
14	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	<1 %
15	Rizki Rahman, Ferian Fauzi Abdulloh. "Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset", sinkron, 2023 Publication	<1 %
16	Submitted to University of Hull Student Paper	<1 %

17	www.mdpi.com Internet Source	<1 %
18	Amruta Naik, Anup A. Kale, Jyutika M. Rajwade. "Sensing the future: A review on emerging technologies for assessing and monitoring bone health", Biomaterials Advances, 2024 Publication	<1 %
19	eprints.soton.ac.uk Internet Source	<1 %
20	garuda.kemdikbud.go.id Internet Source	<1 %
21	journal.uad.ac.id Internet Source	<1 %
22	www.researchgate.net Internet Source	<1 %
23	Wahyu Supriyatin. "Palm oil extraction rate prediction based on the fruit ripeness levels using C4.5 algorithm", ILKOM Jurnal Ilmiah, 2021 Publication	<1 %
24	publikasi.mercubuana.ac.id Internet Source	<1 %
25	"Proceedings of 3rd International Conference on Smart Computing and Cyber Security",	<1 %

Springer Science and Business Media LLC, 2024
Publication

26	Submitted to University of Dammam Student Paper	<1 %
27	Herbers, Cara. "Leveraging Machine Learning Tools to Develop Objective, Interpretable, and Accessible Assessments of Postural Instability in Parkinson's Disease", University of Minnesota, 2024 Publication	<1 %
28	journals.umt.edu.pk Internet Source	<1 %
29	qu.edu.iq Internet Source	<1 %
30	www.extrica.com Internet Source	<1 %
31	Rafli Indra Gunawan, Agung Wahana, Dian Sa’Adillah Maylawati, Cepy Slamet, Cecep Nurul Alam, Nunik Destria Arianti. "Sentiment Analysis on the Pros and Cons of Cryptocurrencies using the Multinomial Naïve Bayes Algorithm", 2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED), 2023 Publication	<1 %

internationalpubls.com

32	Internet Source	<1%
33	journal2.uad.ac.id Internet Source	<1%
34	journals.plos.org Internet Source	<1%
35	repository.mercubuana.ac.id Internet Source	<1%
36	www.ijirset.com Internet Source	<1%
37	www.polgan.ac.id Internet Source	<1%
38	Submitted to Addis Ababa University Student Paper	<1%
39	Alam Alam, Divi Adiffia Freza Alana, Christina Juliane. "Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes", sinkron, 2023 Publication	<1%
40	Brivian Florentis Yustanta, Tri Mulyatno Budhi Hartanto, Ilham Khairi Siregar, Soetji Andar, Mochammad Arkansyah, Dwi Prasetyo. "Comparative analysis of algorithm C4.5 and naïve bayes in classifying adolescent domestic violence", AIP Publishing, 2024 Publication	<1%

41	Chrisdion Andrew Ramaputra, Mohammad Hamim Zajuli Al Faroby, Berlian Rahmy Lidiawaty. "Sentiment Analysis of User Reviews on Cryptocurrency Application: Evaluating the Impact of Dataset Split Scenarios Using Multinomial Naive Bayes", The Indonesian Journal of Computer Science, 2024 Publication	<1 %
42	Courage Kamusoko. "Explainable Machine Learning for Geospatial Data Analysis - A Data-Centric Approach", CRC Press, 2024 Publication	<1 %
43	David Camilo Corrales, Emmanuel Lasso, Apolinar Figueroa Casas, Agapito Ledezma, Juan Carlos Corrales. "Estimation of coffee rust infection and growth through two-level classifier ensembles based on expert knowledge", International Journal of Business Intelligence and Data Mining, 2018 Publication	<1 %
44	Submitted to University of Leeds Student Paper	<1 %
45	Yuli Astuti, Sri Ngudi Wahyuni, Dina Maulina, Fajar Muhammad Sidiq. "The Data Leakage Sentiment Analysis Using Naive Bayes Algorithm Based on Machine Learning Approach", 2022 5th International Seminar on	<1 %

Research of Information Technology and Intelligent Systems (ISRITI), 2022
Publication

46	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
47	Gao Long, Linqiao Li. "Leveraging Machine Learning Algorithms for Early Detection and Prognosis of Pediatric Autoimmune Diseases A Comprehensive Data-Driven Approach", Springer Science and Business Media LLC, 2024 Publication	<1 %
48	Lianah Lianah, Syaiful Zuhri Harahap, Irmayati Irmayati. "Implementation of the C4.5 and Naive Bayes Algorithms to Predict Student Graduation", sinkron, 2024 Publication	<1 %
49	Ramadani Ramadani, B.Herawan Hayadi, Hartono Hartono. "Comparative Analysis of Algorithms Naïve Bayes and C45 for Student Satisfaction with Administrative Services", 2023 International Conference of Computer Science and Information Technology (ICOSNIKOM), 2023 Publication	<1 %
50	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in	<1 %

Information, Communication and Computing Technology", CRC Press, 2024
Publication

51	Gita Widarma, Rice Novita, Mustakim, Nesdi Evrilyan Rozanda. "Comparing Classification Algorithms to Analyze Twitter Sentiment on Public Opinion on Fuel Oil", 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2023 Publication	<1%
52	Sheetanshu Gupta, Dhirendra Kumar, Radhika Negi, Ranjan Singh, Mohammad Javed Ansari, Shakuli Kashyap, Sudhir Mehrotra. "Genomic Intelligence: Metagenomics and Artificial Intelligence", CRC Press, 2024 Publication	<1%
53	Vibhu Krishnan Viswanathan, Ajoy Prasad Shetty, Nimish Rai, Nancy Sindhiya, Surabhi Subramanian, S Rajasekaran. "What is the role of CT-Based Hounsfield Unit assessment in the evaluation of Bone Mineral Density in patients undergoing 1- or 2-level lumbar spinal fusion for degenerative spinal pathologies? – A prospective study", The Spine Journal, 2023 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

SINERGI_New_Journal_TemplateZulfi-1732500079679

HALAMAN
<u>BUKTI KORESPONDENSI:</u>

1. BUKTI SUBMIT PAPER

DARMAJAYA

zulfi anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

[Sinergi] Submission Acknowledgement

1 pesan

Prof. Dr. Andi Adriansyah <andi@mercubuana.ac.id> 15 Juli 2024 pukul 19.48
Kepada: Zulfi Anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

Zulfi Anugerahwati:

Thank you for submitting the manuscript, "OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION TREE C4.5" to SINERGI. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Manuscript URL:
http://email-link.mercubuana.ac.id/ls/click?upn=u001.6x8wlsTvK1RwhOJmDQkr-2FOEVdheBUX6CDLcCjYNXF5h-2FR3nc2cP7OGMsbNk5alKD4Ev08gRtjXls4a0ZBzPAJzIntrUdAeDQ6dsXT93ouVtWp3MWVr614ZoSI-2Fo5jp-2BVZWsv_ve8Etq8Cfv1ilQgDCF4YkgZg-2FQwglLbLS6-2Bdb99TGtHf-2Fjf1SP-2BlEqnl3xHzH7xh3PtwJvkiaaaYQTEWkqapcppQexqAs-2FrkPsd9EWUaUD3nUolsMwdtVvEZK-2FJ1LASeszmC3IDmD-2ByEnGFMiBAZ60akQyfhVaeDsYICL7PMKf6ez3VbPY-2BxsB83jc43mcqt7kOr-2Fu-2BYNPm-2FFR6JSY1xif-2Bo-2FSDMXcicSoDW31LKfGw-3D
Username: zulfi_anugerahwati

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Prof. Dr. Andi Adriansyah
SINERGI

SINERGI
http://email-link.mercubuana.ac.id/ls/click?upn=u001.v-2FHa3-2B5Uh6vOYK4aUmmqqEH0Msw8gwTjEjUY8YpJzSmDWtLpF7g38bWvL-2B7ihkdswbllY7Gdu1bQ6tnfVXz9rQ-3D-3DLT7c_ve8Etq8Cfv1ilQgDCF4YkgZg-2FQwglLbLS6-2Bdb99TGtHf-2Fjf1SP-2BlEqnl3xHzH7xh3PtwJvkiaaaYQTEWkqapcrMrtlLo8z77z5lk-2Bvpg0b5TwdO-2FnBQJR3yZlF6tCz-2BwioD61dJaF3x5kgtLjbAlBPBm5v1Kh-2F6OZo9dhlI3wTTjfPEGsC3MXPoOl8ZAFwlqUGYj8emslkjKHAcu3pcdCpzJ9548ZDfMGWFo8UY2gNE-3D
http://email-link.mercubuana.ac.id/ls/click?upn=u001.v-2FHa3-2B5Uh6vOYK4aUmmqqG5uenjW-2FjG19vn0vaq0KAPsORwDu0yYNkT2iFByED1QYusYi5zppWVfmmZxbwtlbL2AHeXgk6FT8qm4oD9HT-2Bo-3DGUIS_ve8Etq8Cfv1ilQgDCF4YkgZg-2FQwglLbLS6-2Bdb99TGtHf-2Fjf1SP-2BlEqnl3xHzH7xh3PtwJvkiaaaYQTEWkqapcgQD5vvo7iFZryLIlUQ9cllgkMGgzl6NgjfjX2yDZzdUsK8lfLrePkhAu-2B7kXM9Z3jesVNKFnt-2FEKEtd-2B2RlzduaA9GY9c6ahNS00Vyqek9F3onvU4Out9UHd-2BWR3xuSEluuKGXlwFMTUJ8-2B-2Beb5cFw-3D

UNIVERSITAS MERCU BUANA

ISSN: 1410-2331 e-ISSN: 2460-1217

SINERGI

T. Mesin - T. Elektro - T. Industri - T. Sipil - Arsitektur

http://publikasi.mercubuana.ac.id/index.php/sinergi

ACTIVE SUBMISSIONS

ACTIVE ARCHIVE

ID	MM-DD SUBMIT	SEC	AUTHORS	TITLE	STATUS
28478	07-22	ART	Anugerahwati	OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION...	Awaiting assignment

START A NEW SUBMISSION

CLICK HERE to go to step one of the five-step submission process.

REFBACKS

2. BUKTI HASIL REVIEW PAPER

DARMAJAYA zulfi anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

[Sinergi] #28478 Editor Decision

2 pesan

Andi Andriansyah <andi@mercubuana.ac.id> 10 September 2024 pukul 11.47
Kepada: Zulfi Anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

Journal Name: SINERGI
Article Title: OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION TREE C4.5

Dear Zulfi Anugerahwati:

We have reached a decision regarding your submission to SINERGI, "OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION TREE C4.5".

Our decision is: Revisions Required

Please revise your paper according to the reviewers' comments. List down revision that you have done in a list correction table. Send the revised paper and a list correction table file in Journal System and cc to andi@mercubuana.ac.id.

We look forward to hearing from you soon.

Best regards,

Prof. Dr. Andi Adriansyah
(Scopus H-index: 10), Universitas Mercu Buana, Jakarta
Phone +62215871335
Fax +62215871335
andi@mercubuana.ac.id

Reviewer A:

The author needs to revise the comments in the attached file.

Pay attention to typos and equation errors.

Reviewer B:
• This paper addresses osteoporosis, a condition that lowers bone density and raises the risk of fracture. To avoid difficulties, early detection is essential, but the diagnostic tools available today are inadequate. In order to solve this, the osteoporosis risk prediction algorithms Naive Bayes and Decision Tree C4.5 are utilised, and Particle Swarm Optimisation (PSO) is applied to optimise feature selection and increase model efficiency and accuracy.
• The images presented in the Figure 1 are not sufficiently clear, which makes it difficult to fully interpret the data or understand the methodology. It would be helpful to improve the resolution and clarity of the images to enhance the reader's comprehension.
• Does this study's dataset provide enough information to make an accurate osteoporosis prediction? Is 1,958 records a sufficient representation of the problem to build a trustworthy prediction model?
• In what specific ways is PSO used with the Decision Tree C4.5 and Naive Bayes models? The authors should give a more thorough explanation of how PSO chooses the most pertinent traits. Illustrative materials or graphics might be included to better explain how PSO enhances model performance.
• Do the authors explain the findings enough, especially in light of how they connect to the applicability of osteoporosis detection in the real world? The reason Decision Tree C4.5 beats Naive Bayes (especially when combined with PSO) might be explored in more detail by the writers. Additional investigation into the precise

3. BUKTI EDITOR *DECISION*

DARMAJAYA

zulfi anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

[Sinergi] #28478 Editor Decision: ACCEPT SUBMISSION

3 pesan

Andi Andriansyah <andi@mercubuana.ac.id> 15 November 2024 pukul 07.40
Kepada: Zulfi Anugerahwati <zulfi.2221210036@mail.darmajaya.ac.id>

Journal Name: SINERGI
Article Title: OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION TREE C4.5

Dear Zulfi Anugerahwati:

We have reached a decision regarding your submission to SINERGI, "OSTEOPOROSIS PREDICTION USING NAIVE BAYES AND DECISION TREE C4.5".

Our decision is to: ACCEPT SUBMISSION

Your paper will be scheduled after your final paper, similarity report and payment evidence reached us !!!
Please submit your documents:
1. Final paper (in MS Word file format). Strictly adhere to the SINERGI Template: https://bit.ly/3mQKkQl. We will usually expect a minimum of 25 to 30 primarily references to recent journal articles (at least 5 (five) years)
2. Similarity report. The similarity rate should be checked using software such as iThenticate or Turnitin (the result is below 25%).
3. Please attach proof of PROOFREAD LETTERS from an official language institution or pay proofread services from the Editor Team of USD 100 (IDR 1000K).
4. Payment evidence

Please submit your final paper, similarity report and payment evidence to email: QONITA AZILLATIN qonita.azillatin@mercubuana.ac.id, cc: andi@mercubuana.ac.id within 4 (four) weeks.

We appreciate your total commitment to supporting this journal.
We look forward to hearing from you soon.

Best regards,

Prof. Dr. Andi Adriansyah
(Scopus H-index: 10), Universitas Mercu Buana, Jakarta
Phone +62215871335
Fax +62215871335
andi@mercubuana.ac.id

This journal is an OPEN ACCESS. The authors pay an open-access fee, or on their behalf to support the cost of wide open access dissemination of research results, pay the deposit to CrossRef for each published articles have a Digital Object Identifier (DOI), to manage the various costs associated with handling and editing the submitted manuscripts, and the Journal management and publication in general.

Each accepted paper will be charged (based on the first author and first institution): USD 200 (IDR 2500K). This charge is for the first 8 (eight) pages, and if any published manuscript over 8 (eight) pages will incur extra charges of USD 20 (IDR 250K) per page. Note. We only appreciate papers with a single (sole) Author. A paper with a single author will never be published in this journal.

The payment should be made by bank transfer (T/T):

Bank Account name (please be exact)/Beneficiary: MENARA BHAKTI YAYASAN
Bank Name: Bank Negara Indonesia (BNI)
Branch Office: MENTENG
City: Jakarta
Country: Indonesia