Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 1 Iss. 1 July 2023, pp: 47-54
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Implementation of Naïve Bayes Classifier for Classifying Alzheimer's Disease Using the K-Means Clustering Data Sharing Technique

Wildani Putri[1*], Delvi Hastari[2], Kunni Umatal Faizah[3], Siti Rohimah[4], Devy Safira[5], Ella Silvana Ginting[6]

[1,2,3,4,5]Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[6]Faculty of Technology Management and Business, University Tun Hussein Onn Malaysia

E-Mail: [1]12050320361@students.uin-suska.ac.id, [2]12050320385@students.uin-suska.ac.id,
[3]12050324542@students.uin-suska.ac.id, [4]12050323845@students.uin-suska.ac.id,
[5]devysafira@gmail.com, [6]ellaginting86@gmail.com

Received March 22th 2023; Revised Jul 5th 2023; Accepted Jul 30th 2023
Corresponding Author: Wildani Putri

Abstract

Alzheimer's disease is a neurodegenerative disease that is very universal and characterized by memory loss and cognitive function decline which ultimately leads to dementia. In 2015, it is estimated that around million people worldwide will suffer from Alzheimer's disease or dementia. Globally, the number of Alzheimer's diseases will increase from 26.6 million in 2006 to 106.8 million cases in 2050. Due to the large number of people with Alzheimer's disease, it is necessary to classify symptoms that lead to indicators of Alzheimer's disease, so that data mining methods are used for data processing. Alzheimer's data taken from Kaggle amounted to 373 records, through the stages of data preprocessing, data sharing using the Hold-Out method and clustering with AK-Means algorithm. The data is processed using data mining techniques using NBC algorithms. Validation testing the accuracy value obtained the result that the NBC algorithm with K-Means Clustering data sharing has relatively better accuracy than the hold-Out method of 91.89%.

Keywords**:** Alzheimer, Classification, Data Sharing Technique, K-Means Clustering, Naïve Bayes Classifier

1. INTRODUCTION

As the most common neurodegenerative disease, etiology and clinical heterogeneity, Alzheimer's disease (AD) is rumored to be on the rise in various regions [1]. Alzheimer's disease is one of the cases ofemensial disease, which is estimated to affect more than 50 million people worldwide. Of these cases, >50% are due to Alzheimer's Disease [2]. About80% of Alzheimer's disease risk is inherited by enowitz, but the biological mechanismsin 2015 still underlie heritability and are poorly understood. An analysis was conducted to elucidate individual genetic variants ("ge-this phenom notypes") associated with changes in risk for disease phenotypes [3]. In 2020, 5.8 million Americans aged 65 and older were living with Alzheimer's, and this number is expected to reach 13.8 million by 2050 [4].

Alzheimer's disease is biologically determined by the presence of plaques containing amyloid B and neurofibrillary tangles containing tau. Alzheimer's is a genetic and sporadic neurodegenerative disease that causes amnestic cognitive impairment in its prototypical presentation and non-amnestic bognitive disorder in less common variants. Alzheimer's is a common cause of acquired cognitive impairment in middle age and late life, but its clinical impact is modified by other neurodegenerative and cerebrovascular conditions [5]. Alzheimer's disease is associated with widespread disruptions in intrinsic local specialization and global integration in the brain's functional systems. This widespread disorder upsets the balance between local specialization and global integration in the brain and consequently impairs cognitive abilities, affecting the patient's ability to perform daily tasks [6]. Alzheimer's disease is pathologically characterized by p-amyloid deposition (Ap) and accumulation of fibrillar phosphorylated tau (pTau), followed by neurodegeneration and loss of synapses in the brain [7].

Data mining is used for data mining that combines Alzheimer's Disease data with prevention and control [8]. Data mining is a method of exploring unique types of data along with the use of developed methods [9]. These methods include the use of econometric models or machine learning to help cluster data in Alzheimer's disease [10]. In prepocessing data, clustering methods are used, one of which is the K-Means algorithm and

the Hold Out method. K-Means is a grouping method used to support decisions [11]. K-Means Clustering analysis allocated various categories based on the best similarity [12]

There have been many studies related to the use of the Naïve Bayes Classifier (NBC) algorithm using hold out such as research conducted by N Pandiangan in 2019 regarding the "Implementation of Decision Tree and Naïve Bayes Classification Method for Predicting Study Period" al naïve Bayes algorithm gets the highest accuracy value found at a ratio of 80:20 with an accuracy value of 66% [13]. Then, in Ajib Susanto's research in 2021 regarding "Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method" succeeded in implementing NBC and K-Means with 74.5% accuracy and 25.5 errors [14]. In addition, research conducted by Shamima Akter in 2022 on "Convergence of Blockchain, k-medoids and homomorphic encryption for privacy preserving biomedical data classification" resulted in a better accuracy value of K-Medoids of 97.80% [15]. Based on the results of the above exposure, maka can be classified for patients who suffer from Alzheimer's disease. Next, a classification will be carried out for someone whohas Alzheimer's disease. So it can be concluded that this study raised the title Implementation of Naïve Bayes Classifier for Alzheimer's Disease Classification Using K-Means Clustering Data Sharing Technique.

2. MATERIALS AND METHODS

2.1. Stages of Research

This research is divided into several stages, namely, the planning stage, data collection, data processing / pre-processing stages and data sharing, implementation of algorithmas and the results of analysis algorithms and documentation. In Figure 1. Displayed research methodology of the research to be made.

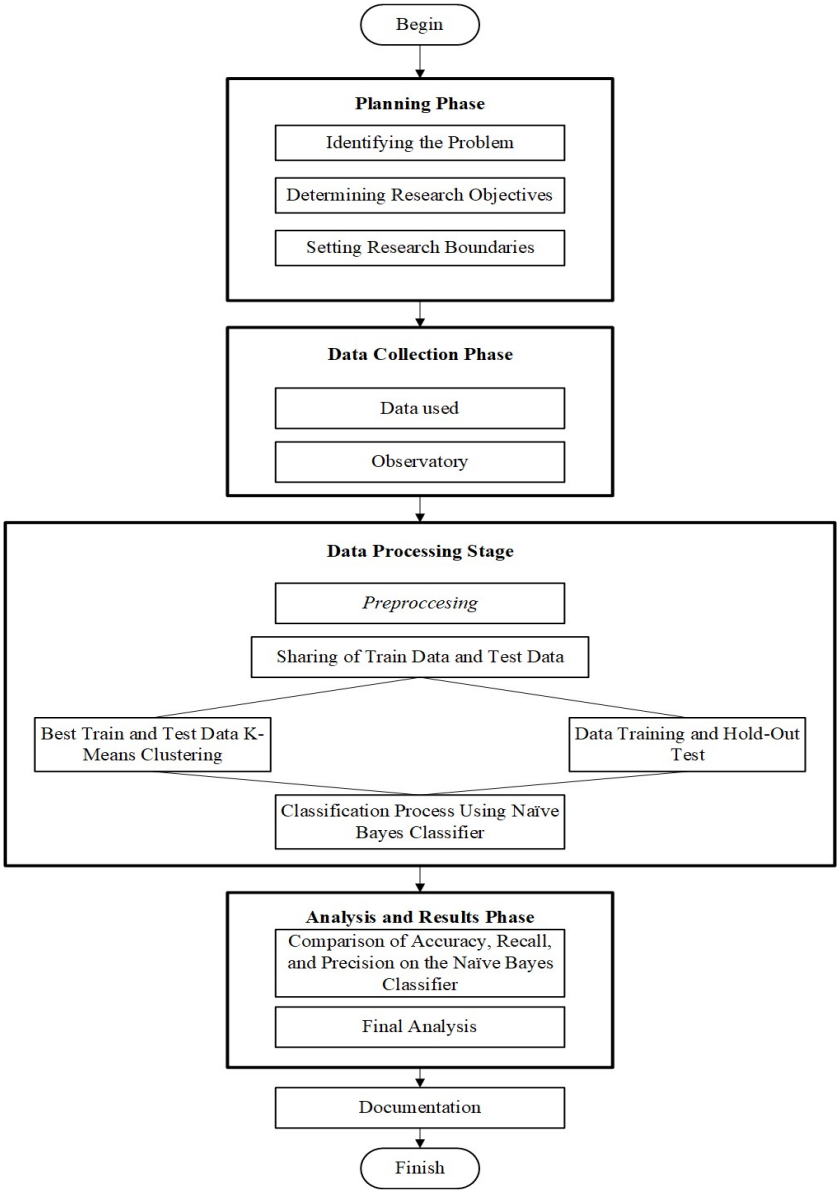

Figure 1. Research Methodology

The dataset used in this study comes from the Kaggle web which can be downloaded for research on data mining with a total of 373 data. The data is then preprocessed so that it becomes 371 data. After that, the data sharing method was carried out using the 80: 20 hold out method and the clustering method, namely the application of the K-Means algorithm so that 260 training data records and 111 testing data records were obtained. Next, the classification process is carried out using the NBC algorithm. The next stage, the analysis process and results are carried out to determine the prediction results from the classification that has been made.

2.2. Alzheimer 's

Alzheimer's disease is a very universal neurodegenerative disease characterized by memory loss and impaired cognitive function that ultimately leads to dementia [16]. Alzheimer's disease is a disease that is usually felt by the elderly aged 60 years and over, but at this time Alzheimer's can be felt by someone who has a fairly adolescent age. Alzheimer's sufferers generally have difficulty in remembering something, difficulty in dialogue so often feel isolated and affect the psychological state of the sufferer [17]. The characteristics of this disease are dementia which often begins with a decrease in the ability to remember. Reduced memory or the ability to identify something which is then aggravated due to disorders of the brain that are increased in nature so that sufferers experience the inability to remember or recognize something [18].

2.3. Classification

Classification is one of the most well-known machine learning methods for predicting new sample classes, using modeling from training data. Universally, classification is defined as a learning method that maps or classifies data instances into predefined class labels in a given data set [19]. Classification is the computational task of classifying data objects among the many classes available [20]. The purpose of classification is to predict target variables (categories) by forming a classification model based on training data, and then using that model to predict the value of the test data class [21].

2.4. K-Means

The K-Means Clustering algorithm is a very simple unsupervised learning algorithm. This algorithm provides a very convenient way to classify a particular dataset into a certain number of clusters, ie. Data sets such as x1, x2, x3, ..., xn are grouped into K-clusters. The main idea of this algorithm is to define one K-center for each cluster, the cluster center must be randomly selected [22]. Euclidean distance can be found using Equation 1 [23].

$$d_{ij} = \sqrt{\left(z_{i1} - z_{j1}\right)^2} + \left(z_{i1} - z_{j2}\right)^2 + \cdots + \left(z_{im} - z_{jm}\right)^2 \quad (1)$$

Information:

d_{ij} : The distance between the i-th curve and the j-th curve

$z_{i1}, z_{j1}, z_{i2}, z_{j2}, \ldots, z_{im}, z_{jm}$: Score the corresponding major components in matrix z

2.3 Naïve Bayes Classifier

This method is commonly used to calculate the probability of a probability based on a known attribute or trait as a means of ascertaining the most accurate class that refers to the highest probability value [24]. The advantage of this theorem is that we can create it very easily and it will work for large data sets and the disadvantage of this theorem is that it will assume all variables are dependent [25]. Not only that, the use of the Naïve Bayes method only requires a small amount of training data to ensure the estimation of the parameters needed in the classification process [26], for the NBC algorithm formula can be seen in Equation 2 [27].

$$P(B) = \frac{P(A)P(A)}{P(B)} \quad (2)$$

Information:

$P(B)$: Conditional probability from A to B

$P(A)$: Conditional probability from A to B

$P(A)$: Probability of occurrence A

$P(B)$: Event probability (B)

2.4 Hold-Out

The holdout method is a method of sharing and evaluating models where data is divided into training data and testing data randomly that are mutually independent (not overlapping) [28]. In the process of

randomizing data to divide data into training and testing data in other methods, it is possible to be overrepresented in one or more classifications, that is, the classification is dominant compared to other classifications [29]. This results in the resulting training and testing data being unrepresentative. However, the holdout method is able to guarantee that each classification can be represented in training and testing data so that the data created is proportional [30].

3. RESULTS AND ANALYSIS

3.1. Data Collection

The data used for this study is data on people with Alzheimer's disease sourced from kaggle. The data totaled 373 records and explained that the attributes behind Alzheimer's disease include gender, age, length of education, socioeconomic status, MMSE (Mini Mental State Examination), CDR (Clinical Dementia Rating), eTIV (Estimated Total Intracranial Volume), Nwbv (Normalize Whole Brain Volume) and ABF (Atlas Scaling Factor). Alzheimer's disease is the most popular neurodegenerative disease with characteristics where the sufferer will lose memory and decline in cognitive function causing dementia. Therefore, this study tries to classify data attributes to determine the accuracy value in recognizing Alzheimer's disease.

Table 1. Data Collection

No	Possible Alzheimer's	Gender	Age	Duration of Education	Socioeconomic Status	MMSE	CDR	eTIV	nWBV	ASF
1	Nondemented	M	87,000	14,000	2,000	27,000	0,000	1987,000	0,696	0,883
2	Nondemented	M	88,000	14,000	2,000	30,000	0,000	2004,000	0,681	0,876
3	Demented	M	75,000	12,000		23,000	0,500	1678,000	0,736	1,046
4	Demented	M	76,000	12,000		28,000	0,500	1738,000	0,713	1,010
….	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..
372	Nondemented	F	63,000	13,000	2,000	30,000	0,000	1327,000	0.796	1.323
373	Nondemented	F	63,000	13,000	2,000	30,000	0,000	1333,000	0.801	1.317

3.2. Preprocessing Data

The next stage is data cleaning, Data cleaning is cleaning irrelevant data and data that contains noise, such as blank, invalid, or misspelled data. Data cleaning affects data mining because of the reduced volume and complexity of data. The initial data consisted of 373 data after cleaning produced 371 data. The results of data cleaning can be seen in Table 2.

Table 2. Cleaning Result Data

No	Possible Alzheimer's	Gender	Age	MMSE	CDR	eTIV	nWBV	ASF
1	Nondemented	M	87,000	27,000	0,000	1987,000	0,696	0,883
2	Nondemented	M	88,000	30,000	0,000	2004,000	0,681	0,876
3	Demented	M	75,000	23,000	0,500	1678,000	0,736	1,046
4	Demented	M	76,000	28,000	0,500	1738,000	0,713	1,010
…..	……	…..	……	……	….	……	…..	……
370	Nondemented	F	63,000	30,000	0,000	1327,000	0,796	1,323
371	Nondemented	F	65,000	30,000	0,000	1333,000	0,801	1,317

3.3. Data Sharing

3.3.1. Hold Out

In the experiment with the hold out method, a ratio of 80:20 was used with the number of training data 297 and testing data 74. The results of this data will then be classified using the NBC algorithm to see the accuracy performance using this method. Previously in the study (Akbar, Firman and Rahmaddeni: 2022), data sharing was carried out with the 70:30 hold-out method on Alzheimer's data. The use of the 80:20 Hold-Out method can be used as an update in this study.

3.3.2. K-Means Clustering

The cleaning results are divided using the clustering method, namely with the K-Means algorithm. In its implementation, experiments were carried out with k = 2 to k = 5. The best number of clusters will be used for data sharing in a ratio of 70:30. This was done to get updates from research (Akbar, Firman and Rahmaddeni: 2022), where the research used the hold out method with a ratio of 70:30 resulting in an accuracy of 83.04% [31]. The best cluster between k=2 sd. k=5 can be seen from the performance value as shown in Figure 2.

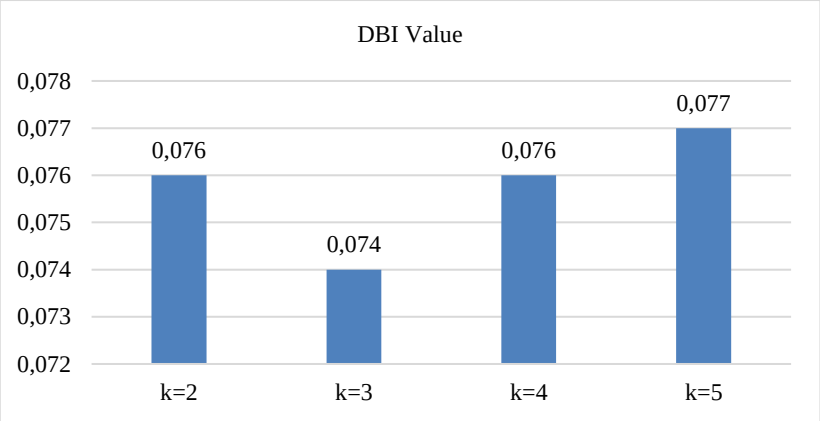

Figure 2. K-Means Cluster Validity Value

In Figure 2. It can be seen that the best cluster is found at k = 3 with the smallest DBI value of 0.074. The dataset at k=3 will be divided into data in a ratio of 70:30. The results of the distribution obtained 260 training data and 111 testing data.

3.4. Implementation of NBC Algorithm with Hold Out

In the NBC implementation experiment with the hold out method, a ratio of 80:20 was used with the number of training data 297 and testing data 74. The accuracy of the hold-out experiment is illustrated by the confusion matrix which can be seen in Table 3.

Table 3. Confusion Matrix

	True Converted	True Demented	True Nondemented	Class Precision
Pred. Converted	5	16	0	23,81%
Pred. Demented	12	99	2	87,61%
Pred. Nondemented	13	0	150	92,02%
Class Recall	16,67%	86.09%	98,68%	

In Table 4. explain the results of Alzheimer's data classification with hold out as outlined in accuracy, precision, and recall performance. The calculation results have the status if Converted as many as 5 data, If converted but Demented 16 Data, If Converted then Nondemented 0 data, If Demented but Converted 99 data, If Demented but Nondemented 2 Data, If Nondemented but Converted 13 data, If Nondemented But Demented 0 data, and If Nondemented 150 data. The accuracy results obtained are quite good at 85.52%.

3.5. Implementation of NBC Algorithm with K-Means

The application of the NBC method is carried out with rapidminer tools with datasets that have been shared before. This dataset has 3 classes including "Nondemented", "Demented", and "Convented". The results of the calculation of classification performance using 111 test data, which have the status if Converted as many as 4 data, if Converted But Demented as many as 2 data, if Demented 34 data, if Demented but Converted 3 data, if Nondemented 64 data, and if Nondemented but Converted 4 data. The results at the test stage will be explained in the form of Precision, Recall and Accuracy values which will be visualized in the form of a Confussion Matrix as outlined in Table 4.

Table 4. Confussion Matrix

	True Converted	True Demented	True Nondemented	Class Precision
Pred. Converted	4	2	0	66.67%
Pred. Demented	3	34	0	91.89%
Pred. Nondemented	4	0	64	94.12%
Class Recall	36.36%	94.44%	100.00%	

Table 3 presents a number of training and testing data which are the results of predictions per class, Precision, and Recall and Accuracy. The use of Confussion matrix can provide information related to the results of true positive values, true negative values, false positive values and false negative values. The Precision value can see how precise the prediction is to identify data according to the original class, as well as the recall value can see the ability of the classifier to redetect information according to the original class, while accuracy is a

parameter to see the performance of the entire model formed. The performance results from the application of the NBC algorithm with the K-means clustering data sharing method are relatively better with an accuracy of 91.89%.

The results of Confussion Matrix testing from both data sharing methods, namely Hold-Out and K-means from three classes of possible Alzheimer's suffer, can be analyzed in the form of average results of precision, recall and accuracy which can be seen in Figure 3.

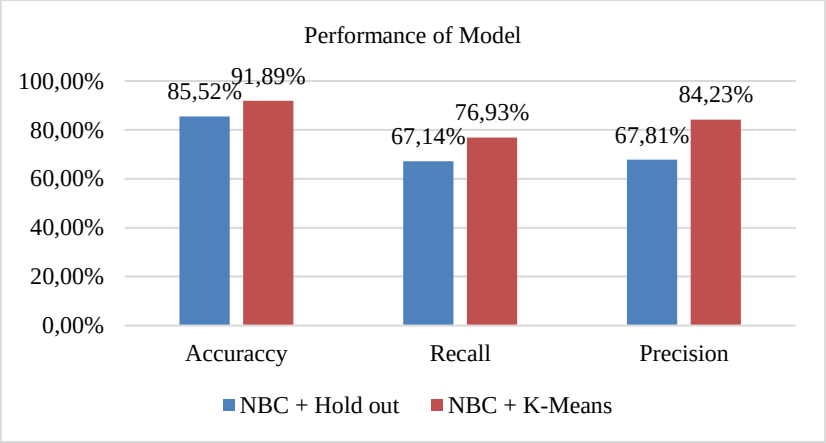

Figure 3. Performance of Test Results

Based on Figure 3. The highest accuracy value was obtained from classification using the NBC algorithm with data sharing techniques using k-means clustering with an accuracy performance of 91.89%, Recalll 76.93%, Precision 84.23%, while the use of the Hold-Out data sharing method on NBC only got an accuracy of 85.52%, recall 67.14%, and 67.81% precision.

4. CONCLUSIONS

This research was conducted to develop a data mining model to classify Alzheimer's disease with NBC algorithms with hold-out data sharing methods and K-means clustering. Alzheimer's data taken from Kaggle amounted to 371 records after cleaning, through the data sharing stage, namely the 80:20 hold-out method and K-Means Clustering. Application of K-Means clustering data sharing method with k=2 to k=5 experiments. The best number of cluster is k = 3 with a DBI value of 0.074. Testing the accuracy value of K-Means Clustering data sharing explained that the NBC algorithm has a relatively higher accuracy of 91.89% compared to the Hold-Out method which only gets 85.52% accuracy.

REFERENCES

[1] J. Lu *et al.*, "The heterogeneity of asymmetric tau distribution is associated with an early age at onset and poor prognosis in Alzheimer's disease," *60. Jahrestagung der Dtsch. Gesellschaft für Nukl.*, vol. 61, no. April, 2022, doi: 10.1055/s-0042-1746117.

[2] T. Bachmann, M. L. Schroeter, K. Chen, E. M. Reiman, and C. M. Weise, "Longitudinal changes in surface based brain morphometry measures in amnestic mild cognitive impairment and Alzheimer's Disease," *NeuroImage Clin.*, vol. 38, no. March, p. 103371, 2023, doi: 10.1016/j.nicl.2023.103371.

[3] Y. Katsumata *et al.*, "Multiple gene variants linked to Alzheimer's-type clinical dementia via GWAS are also associated with non-Alzheimer's neuropathologic entities," *Neurobiol. Dis.*, vol. 174, no. August, 2022, doi: 10.1016/j.nbd.2022.105880.

[4] D. Wang *et al.*, "Deep neural network heatmaps capture Alzheimer's disease patterns reported in a large meta-analysis of neuroimaging studies," *Neuroimage*, vol. 269, no. January, 2023, doi: 10.1016/j.neuroimage.2023.119929.

[5] Y. Wang *et al.*, "A blood-based composite panel that screens Alzheimer's disease," *Biomark. Res.*, vol. 11, no. 1, pp. 1–8, 2023, doi: 10.1186/s40364-023-00485-6.

[6] P. Chen *et al.*, "Articles Altered global signal topography in Alzheimer ' s disease," vol. 89, pp. 1–13, 2023.

[7] K. Kasuga *et al.*, "The clinical application of optimized AT(N) classification in Alzheimer's clinical syndrome (ACS) and non-ACS conditions," vol. 127, pp. 23–32, 2023.

[8] Q. Xu, L. Ning, T. Yuan, and H. Wu, "Application of data mining combined with power data in assessment and prevention of regional atmospheric pollution," *Energy Reports*, vol. 9, pp. 3397–3405, 2023, doi: 10.1016/j.egyr.2023.02.016.

[9] K. Aulakh, R. K. Roul, and M. Kaushal, "E-learning enhancement through educational data mining with

Covid-19 outbreak period in backdrop: A review," *Int. J. Educ. Dev.*, vol. 101, no. April, 2023, doi: 10.1016/j.ijedudev.2023.102814.

[10] N. Dominic, G. N. Elwirehardja, and B. Pardamean, "Data Mining for the Global Multiplex Weekly Average Income Analysis Weekly," vol. 00, no. 2022, pp. 1–8, 2023.

[11] M. Sinan, J. Leng, K. Shah, and T. Abdeljawad, "Advances in numerical simulation with a clustering method based on K–means algorithm and Adams Bashforth scheme for fractional order laser chaotic system," *Alexandria Eng. J.*, vol. 75, pp. 165–179, 2023, doi: 10.1016/j.aej.2023.05.080.

[12] R. Chen, S. Wang, Z. Zhu, J. yu, and C. Dang, "Credit ratings of Chinese online loan platforms based on factor scores and K-means clustering algorithm," *J. Manag. Sci. Eng.*, vol. 8, pp. 287–304, 2023, doi: 10.1016/j.jmse.2022.12.003.

[13] N. Pandiangan, M. L. C. Buono, and S. H. D. Loppies, "Implementation of Decision Tree and Naïve Bayes Classification Method for Predicting Study Period," *J. Phys. Conf. Ser.*, vol. 1569, no. 2, 2020, doi: 10.1088/1742-6596/1569/2/022022.

[14] A. Susanto, M. Atho'il Maula, I. Utomo, W. Mulyono, and K. Sarker, "Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method," *J. Appl. Intell. Syst.*, vol. 6, no. 1, pp. 40–45, 2021.

[15] S. Akter, F. Reza, and M. Ahmed, "Convergence of Blockchain, k-medoids and homomorphic encryption for privacy preserving biomedical data classification," *Internet Things Cyber-Physical Syst.*, vol. 2, no. May, pp. 99–110, 2022, doi: 10.1016/j.iotcps.2022.05.006.

[16] K. M. Stouffer *et al.*, "Early amygdala and ERC atrophy linked to 3D reconstruction of rostral neurofibrillary tau tangle pathology in Alzheimer's disease," *NeuroImage Clin.*, vol. 38, no. July 2022, p. 103374, 2023, doi: 10.1016/j.nicl.2023.103374.

[17] M. Mustakim *et al.*, "Journal of Biological Sciences," vol. 9, no. 1, pp. 122–129, 2022, doi: 10.24843/metamorfosa.2021.v09.i01.p12.

[18] F. Az-zahra *et al.*, "In Silico Study of Betel Leaves Compound (Piper betle L .) as Acetylcholinesterase (AChE) Enzyme Inhibitor in Alzheimer Disease," vol. 2, no. 2, pp. 44–58, 2022.

[19] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0219-y.

[20] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Comput. Methods Programs Biomed.*, vol. 197, 2020, doi: 10.1016/j.cmpb.2020.105725.

[21] J. L. Xu, S. Hugelier, H. Zhu, and A. A. Gowen, "Deep learning for classification of time series spectral images using combined multi-temporal and spectral features," *Anal. Chim. Acta*, vol. 1143, pp. 9–20, 2021, doi: 10.1016/j.aca.2020.11.018.

[22] J. Vijay and J. Subhashini, "An efficient brain tumor detection methodology using K-means clustering algoriftnn," in *2013 International conference on communication and signal processing*, 2013, pp. 653–657.

[23] G. Niu, Y. Ji, Z. Zhang, W. Wang, J. Chen, and P. Yu, "Clustering analysis of typical scenarios of island power supply system by using cohesive hierarchical clustering based K-Means clustering method," *Energy Reports*, vol. 7, pp. 250–256, 2021, doi: 10.1016/j.egyr.2021.08.049.

[24] A. D. W. Sumari, A. M. Nugraheni, and Y. Yunhasnawa, "A Novel Approach for Recognition and Identification of Low-Level Flight Military Aircraft using Naive Bayes Classifier and Information Fusion," *Int. J. Artif. Intell. Res.*, vol. 6, no. 2, 2022, doi: 10.29099/ijair.v6i1.248.

[25] H. Yoshikawa, "Can naive Bayes classifier predict infection in a close contact of COVID-19? A comparative test for predictability of the predictive model and healthcare workers in Japan: Infection Prediction in a Close Contact of COVID-19," *J. Infect. Chemother.*, vol. 28, no. 6, pp. 774–779, 2022, doi: 10.1016/j.jiac.2022.02.017.

[26] M. Vishwakarma and N. Kesswani, "A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection," *Decis. Anal. J.*, vol. 7, no. April, 2023, doi: 10.1016/j.dajour.2023.100233.

[27] A. Tariq *et al.*, "Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data," *Heliyon*, vol. 9, no. 2, p. e13212, 2023, doi: 10.1016/j.heliyon.2023.e13212.

[28] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions," *Energy Convers. Manag.*, vol. 198, no. May, p. 111793, 2019, doi: 10.1016/j.enconman.2019.111793.

[29] A. Gupta, R. Kumar, H. Singh Arora, and B. Raman, "MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis," *IEEE Access*, vol. 8, no. Ml, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.

[30] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.
[31] T. Online and F. Akbar, "Comparison of Machine Learning Algorithms to Predict Alzheimer's Disease," vol. 8, no. 2, pp. 236–245, 2022.