Integration of K-Means Clustering and Naïve Bayes Classification Algorithms for Smart AC Monitoring and Control in WSAN

by Marti Widya Sari

Integration of K-Means Clustering and Naïve Bayes Classification Algorithms for Smart AC Monitoring and Control in WSAN

Ryan Putranda Kristianto[1]
Informatics Department
Universitas Amikom Yogyakarta
Jl. Ring Road Utara Condong Catur
Sleman, Yogyakarta, Indonesia
ryanputrandak@amikom.ac.id

Banu Santoso[2*)]
Computer Engineering Department
Universitas Amikom Yogyakarta
Jl. Ring Road Utara Condong Catur
Sleman, Yogyakarta, Indonesia
banu@amikom.ac.id
*(*Corresponding author)*

Marti Widya Sari[3]
Informatics Department
Universitas PGRI Yogyakarta
Jl. PGRI I No. 117, Yogyakarta 55182, Indonesia
widya@upy.ac.id

***Abstract*— There is still a lot of excessive use of lamps, televisions, air conditioners (AC) and other electronic goods, resulting in a surge in electricity bills charged by electricity users due to neglect and waste of electrical energy. Based on these problems a system that is needed not only to monitor but also to control electrical equipment remotely so that electricity consumption can be controlled. Wireless Sensor and Actuator Network (WSAN) technology can monitor the physical condition of the environment which is widely applied in intelligent environments. WSAN is placed at certain regional points that will be observed the physical condition of the environment, each WSAN can use several sensors and actuators which will later be sent to the server via a wireless connection. In this research, we will test by making Smart AC (Air Conditioner) where at every point where there is AC will be installed WSAN. Data from several sensors generated from WSAN will be sent to the server to be observed and processed using intelligent computing and machine learning (K-Means and Naïve Bayes) so that the AC can turn on and off according to the physical conditions in the place. We evaluate our model too by using Confusion Matrix and obtain the score for accuracy 90%, precision 83%, recall 100% and error rate 10%, So our model can be called good model.**

Keywords*— *WSAN, Zigbee, K-Means, Naïve Bayes, Smart AC

I. Introduction

Electricity Tariff Increase Policy (TTL) conducted by the government[1]greatly affects the habits of electricity users with both household and industrial tariffs. Despite these problems, there are still many habits of electricity users[2]excessive use of lamps, televisions, air conditioners, and other electronic goods, resulting in a surge in electricity bills charged by electricity users due to negligence and waste of electrical energy. Based on these problems a system that is needed not only to monitor but also to control electrical equipment remotely so that electricity consumption can be controlled.

Wireless Sensor and Actuator Network (WSAN) is a technology that can monitor the physical condition of the environment that is often used and applied to Smart Home and intelligent environments[3] [4]. WSAN works by placing at certain points of the area that will be observed the physical condition of the environment, each WSAN can be fitted with several sensors and actuators. The use of sensors such as temperature, light, humidity and motion (PIR) sensors while actuators such as IR remotes and relays. Data from the sensor and actuator readings will be sent to the server and vice versa via a wireless connection with the Zigbee protocol.

In the previous study, WSAN was only used as monitoring data from several sensors and controlling actuators at each point where the WSAN node was located while at the same time there were data packet losses sent by WSAN to the server through several inter-room obstacles [5]. The author then tests by making Smart AC (Air Conditioner) where at every point where there is AC will be installed WSAN. Data from sensor readings generated from WSAN will be sent to the server to be observed and processed using intelligent computing and machine learning so that the AC can be turned on and off according to the physical conditions in the place.

The structure of this paper is chapter 1 covering the background of this research, chapter 2 relevant research, chapter 3 research methods, chapter 4 discussion and presentation of interpretation of results and finally chapter 5 conclusions.

II. Relevant Research

A. WSAN

Wireless Sensor and Actuator Network (WSAN) is a technology that can monitor the physical condition of the surrounding environment by using two important types of components on the node including sensors and actuators[6]. WSAN itself consists of hardware including Arduino Microcontroller, XBee radio module as a wireless transmitter and receiver, along with several sensors and actuators.

WSAN itself is installed at several points (nodes) that you want to observe the physical condition of the environment. Some sensors that will be placed on the WSAN include light sensors, motion sensors (PIR), temperature sensors and humidity sensors. After WSAN has monitored the physical condition of the environment for a while, the monitoring data will be sent to the server. The data that is already on this server will be further processed.

B. Dataset Model

The dataset in data mining is divided into three namely the Supervised dataset, the Semi-supervised dataset and the Unsupervised dataset[7]. The Supervised dataset is a dataset that has a target class attribute where the attribute will be the determinant of the key attribute used for data training. The Semi-Unsupervised dataset is a dataset that has some target class attributes and some does not have target class attributes, there are several algorithms that in the training phase require a dataset with some target class attributes and some do not. The unsupervised dataset is a dataset that does not have a target class that is a key attribute, the dataset is usually raw data generated from several sensors.

The target class itself is generated when grouping data, this target class will be a key attribute for some Supervised algorithms to look for dataset patterns in future research. The target class contains the categories for each row of data in the dataset[8].

Here are some examples of algorithms that are classified using the Supervised, Semi-supervised and Unsupervised datasets:

1. Supervised Learning: Decision Tree, SVM for classification, Linear Regression and so on[9].
2. Semi-supervised Learning: Clustering using K-Medoids, Ensemble methods and so on[10][11].
3. Unsupervised Learning: K-Means, Apriori and so on[12][13].

C. K-Means Algorithm

K-Means algorithm is a data mining algorithm which is a model of the clustering algorithm, where the algorithm is often used by researchers to group a dataset so that the dataset has target classes or key attributes[14]. The stages process is often used as data preprocessing for classification problems, where the dataset used in the classification is the Supervised dataset, so it can be said if you want to classify a dataset, but the dataset is an Unsupervised dataset, the first step that must be done is to group the data first by converting it into a Supervised dataset. K-Means is the easiest and fastest to understand clustering algorithm, using Similarity-based Instance in grouping and starting the first time by determining Centroid or Clustering center point at the beginning of random calculations[15].

D. Naïve Bayes Algorithm

Naïve Bayes algorithm is a Data Mining algorithm that is used to solve classification problems. The Naïve Bayes algorithm applies the Bayes' Theorem by assuming independence of its predictors[16], in the calculation, Naïve Bayes algorithm classifies each data into existing categories according to the existing target class attribute[17].

Because the Naïve Bayes algorithm calculation uses the target class attribute, it can be said that the algorithm is a Supervised Learning model[9]. The Naïve Bayes algorithm will calculate the probability of all class attributes in the dataset to determine which case the data belongs to which target class.

E. Confusion Matrix

Confusion Matrix is the method for measurement classification model performance, by using the Confusion Matrix, the author will obtain the score of Accuracy, Precision, Recall and Error Rate, which represent how is it good our model [20].

III. RESEARCH METHOD

The research method in this paper explains how the combination of K-Means and Naïve Bayes Algorithms in monitoring and controlling Smart AC. Smart AC that was developed from this research will have adaptive properties to the temperature and humidity of the environment. The use of WSAN sensors greatly influences the control of the Smart AC itself.

As shown in Figure 1, the data generated by WSAN sensors are in the form of an Unsupervised dataset, the authors use the classification algorithm, Naïve Bayes, to determine when Smart AC should turn on and turn off based on the current temperature and humidity (Real-Time).

However, because the dataset to be used by the classification is the Unsupervised dataset, one of the steps that must be carried out before carrying out the classification is to do a grouping of datasets or Clustering. This is also one of the steps in Preprocessing Data that will be used to classify using the Naïve Bayes Algorithm.

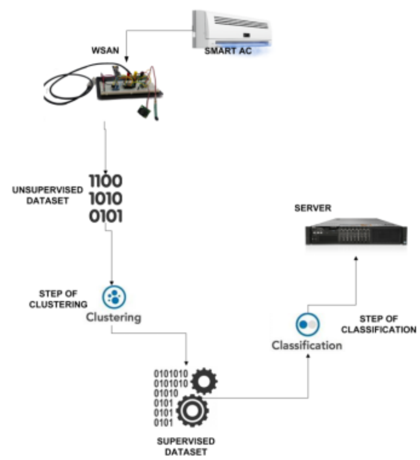

Fig1. Smart AC Infrastructure

After the classification stage is performed using the Naïve Bayes algorithm, the Supervised dataset is then trained to form a Naïve Bayes classification model. If later there is a new incoming data record from WSAN where the data record does not have the target class attribute, then using the Naïve Bayes model that has been built previously will determine which data belongs to which category, whether On, Dim-Off, Off or Dim-On. The data is then sent by the server to the IR Remote actuator on the Smart AC, so the IR Remote actuator that receives the command from the server will determine whether the Smart AC must be On or Off.

We consider why the choice of the Clustering algorithm is K-Means and the Classification algorithm is Naïve Bayes. K-Means algorithm is the most popular clustering algorithm model that is often used as a grouping of data categories[18], it is easy to implement in its application is the choice of researchers to choose the algorithm. K-Means algorithm is also very sensitive to data changes, the dataset generated by WSAN is a type of real Integer because it requires high accuracy in clustering dataset categories, therefore K-Means is the right and suitable Clustering algorithm chosen in this case.

As for classification, researchers consider choosing the Naïve Bayes algorithm. The Naïve Bayes algorithm is a Classification algorithm that takes into account probabilities[19]where the classification algorithm of this model is also easy to implement and also has no computational time, so the process of calculating this algorithm when applied to the WSAN module will be very fast and support changes to data by the WSAN sensor in Real-Time. This rapid change is needed, considering that Real-Time on Smart AC can act immediately according to environmental conditions right away, where the classification algorithm of this model is also easy to implement and also has no computational time, so the process of calculating this algorithm when applied to the WSAN module will be very fast and support changes to data by the WSAN sensor in Real-Time. This rapid change is needed, considering that Real-Time on Smart AC can act immediately according to environmental conditions right away.

The stages of the K-Means algorithm are as follow:

K-Means Algorithm

Determine the number of clusters to be used
While Euclidean Distance next iteration value is the same as the previous iteration
Do

1. Determine Centroid.
   Generate random number as many 4 cluster: Off, Dim-Off, Dim-On and On
   randBetween(1, 4) (1)
2. Perform calculations for each line with Euclidean Distance :

$$C1(X_n, X_n\_C_1) = \sqrt{(X_1 + X_1\_C_1)^2 + (X_2 + X_2\_C_1)^2} \quad (2)$$

$$C2(X_n, X_n\_C_2) = \sqrt{(X_1 + X_1\_C_2)^2 + (X_2 + X_2\_C_2)^2} \quad (3)$$

$$C3(X_n, X_n\_C_3) = \sqrt{(X_1 + X_1\_C_3)^2 + (X_2 + X_2\_C_3)^2} \quad (4)$$

$$C4(X_n, X_n\_C_4) = \sqrt{(X_1 + X_1\_C_4)^2 + (X_2 + X_2\_C_4)^2} \quad (5)$$

3. Label the results of Euclidean Distance calculations labeling is obtained from which cluster that has the smallest number
4. Update Centroid if minimum value is found.
5. Update calculations based on the new Centroid.

End While

The explanation of notation formula in K-Means Algorithm on above :

C1, C2, … , C4 = Cluster 1 – 4

X_n = dataset of clustering on n period

$X_n\_C_1$ = dataset of clustering on row of centroid 1 (same as with $X_n\_C_2$, $X_n\_C_3$, $X_n\_C_4$)

X_1 = first attribute of dataset of clustering (temperature attribute)

X_2 = second attribute of dataset of clustering (humid attribute)

$X_1\_C_1$ = first attribute of dataset of clustering on row of centroid 1

$X_2\_C_1$ = second attribute of dataset of clustering on row of centroid 1

$X_1\_C_2$ = first attribute of dataset of clustering on row of centroid 2, and so on.

The steps for calculating the K-Means algorithm above are repeated continuously until the Centroid value does not change where the results of the Euclidean Distance calculation are the same as the previous iteration.

Next steps are Naïve Bayes algorithm:

Naïve Bayes Algorithm

1. Perform the calculation of Probability P (Ci) on the target class attribute.
2. Perform calculation of the combination of the target class attribute with other attributes P (X | Ci).
3. Carry out the accumulation of P (X | Ci) combination calculations for each target class attribute category.
4. Do P(Ci) * P (X | Ci) multiplication for each target class attribute category.
5. Compare the largest value based on the results of the multiplication of P (Ci) * P (X | Ci) of each of the target class attribute categories.

Last step is evaluation of the model using Confusion Matrix. The author will takes random the dataset supervised as many 30 rows to testing it.

In the Confusion Matrix method, there are 4 attributes that must be determined as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP is condition which the classification value is positive and the actual state is positive, TN is condition which the classification value is negative and the actual state is negative, FP is condition which the classification value is positive and the actual state is negative, and FN is condition which the classification value is negative and the actual state is positive. Simplification of explanation can see on the sixth table.

		Classified	
		False	True
Actual	False	True Negative (TN)	False Positive (FP)
	True	False Negative (FN)	True Positive (TP)

Table 6. Confusion Matrix

Next step, the author analyzes how many TP, TN, FP, and FN that occurred in the classification model. The result from the amount that is used to calculated accuracy, precision, recall and error rate.

Accuracy is how good the model which represent with percentage, while the precision is the fraction of relevant instances among the retrieved instances which represent with a percentage, while the recall is the fraction of the total amount of relevant instances that were retrieved which represent with a percentage and error rate is how many

ID	TimeRTC	DateRTC	PIRSensor	HumuditySensor	TemperatureSensor	LightSensor
1	22:16:53	14/11/2019	0	0	nan	0
2	22:17:03	14/11/2019	0	0	0	16
3	22:17:13	14/11/2019	0	nan	0	20
4	22:17:23	14/11/2019	0	0	nan	22
5	22:17:33	14/11/2019	0	0	0	21
6	22:17:43	14/11/2019	0	0	0	22
7	22:17:53	14/11/2019	0	0	nan	20
8	22:18:03	14/11/2019	0	0	0	196
ID	TimeRTC	DateRTC	PIRSensor	HumuditySensor	TemperatureSensor	LightSensor
1	22:35:33	14/11/2019	0	70	28	97
2	22:35:43	14/11/2019	0	70	28	99
3	22:35:53	14/11/2019	0	69	28	98
4	22:36:03	14/11/2019	0	69	28	98

Table1. Result of generate sensor WSAN Dataset

ID	HumuditySensor	TemperatureSensor
1	25	90
2	25	90
3	25	90
4	25	89
5	25	89
6	25	89
7	25	86
8	25	86
9	25	86
0	26	92
10	26	92
11	25	90
12	25	90

Table 2. .Preprocessing Unsupervised dataset

ID	HumuditySensor	TemperatureSensor	Cluster
1	25	90	1
2	25	90	1
3	25	90	1
4	25	89	1
5	25	89	1
6	25	89	1
7	25	86	1
8	25	86	1
9	25	86	1
0	26	92	1
10	26	92	1
11	25	90	1
12	25	90	1

Table 3. .Result of K-Means Algorithm clustering

ID	HumuditySensor	TemperatureSensor	Cluster	State
1	25	90	1	On
2	25	90	1	On
3	25	90	1	On
4	25	89	1	On
5	25	89	1	On
6	25	89	1	On

Table 4. The results of labeling the target class manually based on the results of clustering

numbers of error in our model which represent with a percentage. The formula to calculated the accuracy, precision, recall and error rate as below :

Accuracy Calculation :

$$\frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (6)$$

Precision Calculation :

$$\frac{TP}{(TP + FP)} \quad (7)$$

Recall Calculation :

$$\frac{TP}{(TP + FN)} \quad (8)$$

Error Rate Calculation :

$$\frac{(FP + FN)}{(TP + TN + FP + FN)} \quad (9)$$

or

$$1 - Accuracy \quad (10)$$

IV. Results and Discussion

K-Means algorithm in this study is used for the data preprocessing stage where the generated sensor data is an Unsupervised dataset. With the help of the K-Means algorithm, the data will have a target class label and be ready to be processed using the Naïve Bayes algorithm.

The stages of the simulation combination of the K-Means and Naïve Bayes algorithms are performed using MS Excel, following an example of the Unsupervised dataset generated by the WSAN sensor as shown in Table 1.

The dataset generated by the WSAN sensors has many unused label attributes and missing data, the first step to do is preprocessing data by removing the unused label attributes and then transforming the value from the "nan" character to the number 0. The next step is to randomly sampling data by paying attention to the variance of values in the attributes used. Sampling data were taken as many as 477 data records like Table 2.

After the data preprocessing stage is carried out, the next step is to group the unsupervised dataset using the K-Means clustering algorithm, from the steps of the K-Means algorithm pseudocode described in the previous section, we will get a dataset like a Table 3.

From table 3 we get the results of clustering in the form of 1, 2, 3 and 4. In this case, the clusters are counted as many as 4 clusters, because in Smart AC itself looks for status On, Dim-Off, Off or Dim-On. Based on K-Means Clustering Algorithm calculation results are then analyzed and taken each cluster several records to determine the target class attribute label, following the results of labeling as Table 4.

After this stage (Table 4) is carried out, a Supervised dataset model that has the target class attribute label and is ready for classification using the Naïve Bayes Algorithm. Example: WSAN sensor at a certain time generates a dataset record with Moist Sensor data valued at 34 and Temperature Sensor valued at 44, from the new record data the target class attribute label or status, will be determined, whether On, Dim-Off, Off or Dim-On using the model learning from what has been generated by the K-Means algorithm (Figure 5), so that researchers or users do not need to cluster the new data record first generated by the WSAN sensor.

After using the Naïve Bayes algorithm classification as shown on Pseudocode in above (section Research Method), with predetermined Temperature and Humidity input parameters, it will produce a probability of the Smart AC status in the on the state.

The author does evaluate the model by using the Confusion Matrix. First, the author must be testing dataset supervised (which is obtained from the K-Means algorithm process) using the Naive Bayes Algorithm. The author takes random the dataset as many 30 rows to testing it. The result is shown on table 5.

Id	temperature Sensor	humiditySensor	StateOfActual	StateOfClassification_NaiveBayes
1	25	90	Off	Off
2	25	90	Off	Off
3	30	71	Off	Off
4	30	71	Off	Off
5	35	40	On	On
6	35	40	On	On
7	45	32	On	On
8	45	32	On	On
9	27	64	Dim-Off	Dim-Off
10	27	63	Dim-Off	Dim-Off
11	34	57	Dim-Off	Dim-Off
12	34	56	Dim-Off	Dim-Off
13	30	45	Dim-On	Dim-On
14	30	45	Dim-On	Dim-On
15	35	52	Dim-On	Dim-On
16	35	52	Dim-On	Dim-On
17	36	37	On	On
18	40	37	On	On
19	32	49	Dim-On	Dim-On
20	33	52	Dim-On	Dim-On
21	33	48	Dim-On	Dim-On
22	34	52	Dim-On	Dim-On
23	34	52	Dim-On	Dim-On
24	30	69	Dim-Off	Dim-On
25	32	63	Dim-Off	Dim-Off

26	33	59	Dim-Off	Dim-On
27	27	81	Off	Dim-Off
28	28	85	Off	Off
29	29	76	Off	Off
30	29	61	Dim-Off	Dim-Off

Table 5. Evaluation of the Naïve Bayes algorithm using the Confusion Matrix

The column of StateOfActual is a column which contains the actual state, is obtained from previously proceeding (clustering proceed), while the column of StateOfClassification_NaiveBayes is a column which contains the classification state which calculated using the Naive Bayes algorithm.

If we analyze the fifth table, so we obtain a score for 4 attributes as : TP = 15, TN = 12, FP = 3 and FN = 0. Using the equation of 6, 7, 8, 9 and 10, the author obtain the accuracy 90%, precision 83%, recall 100% and error rate 10%, so we can call this model is good.

V. CONCLUSION

The dataset which generated by sensors of WSAN will result as unsupervised dataset, even though finally author have to classify it so can be known the state of Smart AC when it on, dim-on, off and dim-off, it cannot because to do classify, the data must be supervised model, so the first step is doing cluster of dataset to supervised dataset, then after that, the dataset already have attribute key so can be done classify. To measurement how good the builded model, the author use Confusion Matrix. Based on the research that the author has done, it can be concluded that:

1. The K-Means algorithm can be used as data preprocessing for the Unsupervised Naïve Bayes grouping dataset.
2. Naïve Bayes algorithm can classify the WSAN sensor result dataset records to determine when the Smart AC status is On, Dim-Off, Off or Dim-On.
3. Using the Confusion Matrix, obtain the accuracy 90%, precision 83%, recall 100% and error rate 10%, so we can concluded that, our model is good.

REFERENCE

[1] B. Santoso, I. W. Mustika, and S. S. Kusumawardani, "Pemodelan Monitoring Pemakaian Dan Penghematan Energi Listrik Dengan Teknologi Jaringan Sensor Nirkabel," p. 8, 2014.

[2] W. Fungsirirut and W. Benjapolakul, "A study on energy reduction based on responsive behavior of users," in *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Phuket, 2017, pp. 580–583.

[3] M. W. Sari, P. W. Ciptadi, and R. H. Hardyanto, "Study of Smart Campus Development Using Internet of Things Technology," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 190, p. 012032, Apr. 2017.

[4] Z. Zhao, S. Kuendig, J. Carrera, B. Carron, T. Braun, and J. Rolim, "Indoor Location for Smart Environments with Wireless Sensor and Actuator Networks," in *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, Singapore, 2017, pp. 535–538.

[5] B. Santoso, "Implementasi Wireless Sensor And Actuator Network Berbasis Protokol Zigbee Untuk Pemantauan Dan Pengendalian Smart Home," *J. Din. Inform.*, vol. 7, no. 1, pp. 13–31, 2018.

[6] J. Zhu, H. Yu, Z. Lin, N. Liu, H. Sun, and M. Liu, "Efficient Actuator Failure Avoidance Mobile Charging for Wireless Sensor and Actuator Networks," *IEEE Access*, vol. 7, pp. 104197–104209, 2019.

[7] L. Nandanwar and K. Mamulkar, "Supervised, Semi-Supervised and Unsupervised WSD Approaches: An Overview," vol. 4, no. 2, p. 5, 2013.

[8] S. M. Gowreeswari and D. R. M. Chezian, "Clustering and Classification in Sentimental Data Analysis," p. 5, 2017.

[9] V. Khadse, P. N. Mahalle, and S. V. Biraris, "An Empirical Comparison of Supervised Machine Learning Algorithms for Internet of Things Data," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1–6.

[10] J. Tian, Y. Ren, and X. Cheng, "Stratified Feature Sampling for Semi-Supervised Ensemble Clustering," *IEEE Access*, pp. 1–1, 2019.

[11] F. Alalyan, N. Zamzami, M. Amayri, and N. Bouguila, "An Improved K-medoids Algorithm Based on Binary Sequences Similarity Measures," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, Paris, France, 2019, pp. 1723–1728.

[12] R. Jothi, "A Comparative Study of Unsupervised Learning Algorithms for Software Fault Prediction," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018, pp. 741–745.

[13] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, "Crime pattern detection, analysis & prediction," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, 2017, pp. 225–230.

[14] J. Qi, Y. Yu, L. Wang, and J. Liu, "K*-Means: An Effective and Efficient K-Means Clustering Algorithm," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, Atlanta, GA, USA, 2016, pp. 242–249.

[15] K. Rajeswari, O. Acharya, M. Sharma, M. Kopnar, and K. Karandikar, "Improvement in k-Means Clustering Algorithm Using Data Clustering," in *2015 International Conference on Computing Communication Control and Automation*, Pune, India, 2015, pp. 367–369.

[16] A. Arivarasan, "Classification based Performance analysis using Naïve-Bayes J48 and Random forest algorithms," p. 5, 2017.

[17] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Parapat, Indonesia, 2018, pp. 1–5.

[18] X. Tang *et al.*, "Feature Selection Algorithm Based on K-means Clustering," in *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Honolulu, HI, 2017, pp. 1522–1527.

[19] K. Netti and Y. Radhika, "A novel method for minimizing loss of accuracy in Naive Bayes classifier," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, 2015, pp. 1–4.

[20] Ting K.M. Confusion Matrix. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA, 2017

Integration of K-Means Clustering and Naïve Bayes Classification Algorithms for Smart AC Monitoring and Control in WSAN

ORIGINALITY REPORT

18%	16%	8%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	edas.info Internet Source	10%
2	ieeexplore.ieee.org Internet Source	1%
3	Windha Mega P.D, Haryoko. "Optimization Of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek's Services", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2019 Publication	1%
4	brandcora.com Internet Source	1%
5	Nancy Shaar, Mohammad Alshraideh, Lara Shboul, Iyad AlDajani. "Decision support system (DSS) for traffic prediction and building a dynamic internet community using Netnography technology in the city of	1%

Amman", Journal of Experimental & Theoretical Artificial Intelligence, 2023
Publication

6	Stefan Rauter, Franz Tschuchnigg. "CPT Data Interpretation Employing Different Machine Learning Techniques", Geosciences, 2021 Publication	1 %
7	export.arxiv.org Internet Source	1 %
8	repository.unpkediri.ac.id Internet Source	1 %
9	www.techscience.com Internet Source	<1 %
10	jdi.upy.ac.id Internet Source	<1 %
11	www.slideshare.net Internet Source	<1 %
12	Nadia Ardhianie, Rachmadita Andreswari, Muhammad Azani Hs. "SENTIMENT ANALYSIS OF 'INDONESIAN NO DATING CAMPAIGNS' ON TWITTER USING NAÏVE BAYES ALGORITHM", 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), 2019 Publication	<1 %
13	www.ijfcc.org Internet Source	

<1%

14	curve.carleton.ca Internet Source	<1%
15	www.diva-portal.org Internet Source	<1%
16	www.intechopen.com Internet Source	<1%
17	www.mdpi.com Internet Source	<1%
18	René Játiva E., Oliver Caisaluisa, Katty Beltrán, Martín Gavilánez. "Radio Frequency Pattern Matching - Smart Subscriber Location in 5G mmWave Networks", 2023 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), 2023 Publication	<1%
19	Lincoln Best, Ernest Foo, Hui Tian. "Chapter 7 Utilising K-Means Clustering and Naive Bayes for IoT Anomaly Detection: A Hybrid Approach", Springer Science and Business Media LLC, 2022 Publication	<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography On