LAPORAN AKHIR
SKEMA PENELITIAN DASAR (PENELITIAN DASAR FUNDAMENTAL DAN PENELITIAN DASAR KERJA SAMA DALAM NEGERI)
Peneliti hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

A. JUDUL
Tuliskan judul usulan penelitian maksimal 20 kata

Classification of Occupant Position Based Ventilation in The Under-Actuated Zone Using K-Nearest Neighbor's

B. RINGKASAN
Isian ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, metode, dan luaran yang ditargetkan

Post-pandemic, the education sector is transitioning to a new normal, with universities in Jakarta, such as Trilogy University, resuming on-campus activities. Library facilities, integral to learning, require efficient air circulation management. Currently, the Trilogy University library employs a traditional HVAC system, manually controlled via wall-mounted buttons and not accounting for occupancy levels. This study aims to develop a system to detect room occupancy using the K-Nearest Neighbors (KNN) algorithm. The system utilizes data from Wi-Fi network sensors placed on room vents, capturing signal strength (RSSI), frequency, and limit values from each Access Point (AP). A dataset of 90 samples was used to train the KNN model, achieving an accuracy of 94% and an error rate of 0.05. Further validation with 15 real-time samples yielded 100% accuracy. The findings indicate that the system can accurately detect room occupancy and effectively handle real-time data, demonstrating its potential for optimizing HVAC control in library environments.

C. KATA KUNCI
Isian 5 kata kunci yang dipisahkan dengan tanda titik koma (;)

Internet of Things, Occupant Position, K- Nearest Neighbor, HVAC System, Sensor

D. PENDAHULUAN
Pendahuluan penelitian tidak lebih dari 1000 kata yang memuat, latar belakang, rumusan permasalahan yang akan diteliti, pendekatan pemecahan masalah, state-of-the-art dan kebaruan, peta jalan (road map) penelitian setidaknya 5 tahun. Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan.

Indonesia is one of the countries trying to recover from the COVID-19 pandemic. Several sectors began to actively move the wheels of life in the new normal. One of them is lecture activities held at Trilogy University, Jakarta. Students can carry out academic activities both in face-to-face learning with lecturers, tutoring, and discussions where previously online learning was carried out [1]. Trilogy University provides facilities in the form of a library that supports the needs of lecture activities. This library provides various types of books, writing guidance rooms, discussion rooms, and BI corners [2]. Currently, the Trilogy University library uses a Heating Ventilation and Air Conditioning (HVAC) system as an air conditioner. The use of the HVAC system in the Trilogy University library is still traditional. Users can turn off, turn on, and control the room temperature using the buttons on the library wall [3]. This makes the library's flow of hot and cold temperatures not optimal. Sometimes a room temperature that is cold enough is good for storing books but not for regulating the temperature of the human body. It also has an impact on the energy expended that is considered less than optimal. Other factors can also affect the comfort of library visitors in their activities indoors [4].

For example, the research of Krishna Chaitanya Jagadeesh Simma, et al [5] discussing real-time occupancy estimation using WiFi Networks to Optimize HVAC Operation, explained about the performance of energy on infrastructure in the US is quite high but has not been able to adjust occupants. Although the topic seems to overlap some with our research

concept, we differentiate ourselves by focusing on (1) measuring variations in ventilation performance (2) data management processes (3) grouping classifications in ventilation systems based on occupant detection data. In addition, the research carried out only focused on the dimensions of infrastructure in buildings.

As another example, research conducted by Wooyoung Jung [6] presents human-in-the-loop HVAC operations: A quantitative review of occupancy, comfort, and energy-efficiency dimensions on classification using the context of sensing, control, and evaluation. Differences in our research concept (1) data retrieval (2) data managers in our research are direct or real-time (3) data grouping in our system using K-Nearest Neighbors. in addition, this study focused only on the results of calculations on occupants and performance on HVAC in buildings.

The research conducted by Donya Sheikh Khan [7] with the title Application of an occupant voting system for continuous occupant feedback on thermal and indoor air quality – Case studies in office spaces, explain the air quality in office spaces. This topic carries data collection using voting on room occupants regarding comfort in HVAC which aims to evaluate discomfort. The difference between this research and the research we conducted are: (1) data collection using variations in ventilation performance not based on voting, (2) the evaluation we carried out was in the form of a system that was able to detect based on the number of room occupants. This study only focuses on the comfort of office space and energy savings without conducting an HVAC performance evaluation.

Based on some of the problems above, this study requires a system that can detect the presence and number of people in a room, especially a library. Where the Internet of Things-based application on ventilation can regulate the needs of air circulation by calculating the distance between humans and the nearest ventilation and can be classified on data which will affect room temperature and energy emitted on HVAC [8]. The method used in the occupant presence detection system with ventilation is the K-Nearest Neighbors method, where the classification of occupant movement data in the underactuated zone is carried out using dBm signals. Residents are required to use the access point given when entering the library, visitors will give a signal to be received at the vent, and the signal will be managed into data on the number of visitors in each vent.

E. METODE

Isian metode atau cara untuk mencapai tujuan yang telah ditetapkan tidak lebih dari 1000 kata. Pada bagian metode wajib dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus memuat sekurang-kurangnya prosedur penelitian, hasil yang diharapkan, indikator capaian yang ditargetkan, serta anggota tim/mitra yang bertanggung jawab pada setiap tahapan penelitian. Metode penelitian harus sejalan dengen Rencana Anggaran Biaya (RAB).

The methodology of the study consists of several stages. This stage will classify the ventilation of occupant positions in libraries in less driven zones using the method of nearest neighbors K. There are four stages in the study. The first stage begins with data acquisition as data retrieval and collection. The second stage is classification is the grouping of data where this study uses the K-nearest neighbor's method. The third stage is Error data, dan the last stage is Data validation, which is to randomly enter the original data that has been calculated in the data classification of the previous process.

A. Data Acquisition

The process of data acquisition is the collection of data in the form of information in the form of values obtained on the device [9]. The system in this study was carried out with the Internet of Things using a Wifi network sensor (Wireless Adapter) to capture the value on each Access Point (AP), the value captured in the form of signal strength. The Wireless Adapter will read the occupant's presence in the University Library Trilogy.

In this data collection, a simulation is carried out directly on the location. The application

used as a signal receiver is Network Cell Info lite, this is because this application can be an adapter in capturing signals. For the simulation of Access Point, 3 mobile phones are placed as signalers which are placed under the library ventilation. This simulation application is placed in a random position to read the value. The unit received is in the form of Signal Strength, namely dBm (Disable-milliwatt). There are eleven objects that become values: ID, Position, RSSI-AP1, FQ-AP1, Limit-AP1, RSSI-AP2, FQ-AP2, Limit-AP2, RSSI-AP3, FQ-AP3, and Limit-AP3. There are several provisions in data retrieval When in the process of grouping values, namely The smaller the size of the dBM Signal Strength (Disable-milli watt), the stronger the signal obtained at the Access Point, and the greater the value received at the Access Point the signal received on the Wifi Adapter is judged to be worse, In the research conducted by Eko Prayetno et al, [10] explained about the signal strength given by mobile phones in the form of the RSSI category in table 2 as follows:

TABLE I. SIGNAL STRENGTH DESCRIPTION

Categories	Signal Strength (dBm)
Very good	< -70 dBm
Well	-70 dBm to -85 dBm
Enough	> -86 dBm to 100 dBm
Bad	>-100 dBm
Fair Poor	-110 dBm

Data collection was carried out continuously for one week at the uptime of the Trilogy University library. Data will always be recorded in the event of a change or movement in the value of the access point. One ventilation nearby access point will retrieve 30 records so that 90 records have been collected.

B. Classification

The classification of K- Nearest Neighbor is one of the well-known non-parametric classification methods and includes the easiest algorithm in problem-solving which is capable of grouping objects in a certain class based on similarities and differences [11]. KNN uses the concept of finding the closest neighbors and classification based on similarity. The initial approach is to give new points to the training data to be used as a classification of test data. The initial approach is to give new points to the training data to be used as a classification of test data. The value lookup is determined by the equation formula:

$$K = \sqrt{n}$$

The formula above is to determine the value of K, where n is the amount of data used in this study [12]. The data obtained in the previous process will be carried out by spit data sets for the sharing of training data and testing data. in this study using 80% data calculation for training data and 20% for testing data. to calculate the closest distance between the training and test data in this study using the Ecludian Distance method [13]. The equation formula can be seen as follows:

$$d(x,y) = \sqrt{\sum_{i=1k=\sqrt{n}}^{n} (x_i - y_i)^2}$$

Where:
x = data used as test variable (RRSI and Limit of each vent).
y = test attribute data (Vents tested)
d = distance between x and y
i = any data
n = amount of data access point
xi = data on the center to i
yi = data on each data to i

The process of the K-Nearest Neighbor method to determine the classification of occupant movement data in the underactuated zone is carried out using the dBm signal. After determining the parameter K, the distance of the Encludian Distance is calculated using the equation formula against the existing training data. The next step is to sort the results of the calculation of distances starting from the smallest. After that, process the search by retrieving the nearest neighboring object from the training and test data. The series of processes will produce classification results to determine the detection of residents in the library of Trilogy University, Jakarta.

C. Performance Measurement

In this process, classification data testing is carried out by calculating the error value in the previous process. This error test uses Mean Absolute Error (MAE). This method is considered capable of calculate absolute errors based on predictions [14]. The formula Mean Absolute Error is explained as follows:

$$MAE = \frac{1}{n}\sum_{i=0}^{n} | f_i - y_i$$

Where:
fi = forecasting result value
yi = actual value ventilation
n = amount of data access point

On the formula above, MAE will intuitively calculate on average the error that the weight is the same on classification data [15]. So in this study, the KNN classification data will be tested for forecasting errors using the overall data.

D. Validation

After getting the final result of the classification members and error testing on the data [16]. This process is part of the classification calculation using a confusion matrix and re-retrieval of data for direct testing. where this calculation is considered quite easy in data implementation [17]. To measure the performance of the K-Nearest Neighbor algorithm using several methods used are Accuracy, Precision, Recall, and F1- Score. As for data testing, 15 data are created that are taken directly to ensure the performance of the algorithm runs well.

F. HASIL PENELITIAN

Tuliskan secara ringkas hasil pelaksanaan penelitian yang telah dicapai sesuai tahun pelaksanaan penelitian. Penyajian meliputi data, hasil analisis, dan capaian luaran (wajib dan atau tambahan). Seluruh hasil atau capaian yang dilaporkan harus berkaitan dengan tahapan pelaksanaan penelitian sebagaimana direncanakan pada proposal. Penyajian data dapat berupa gambar, tabel, grafik, dan sejenisnya, serta analisis didukung dengan sumber pustaka primer yang relevan dan terkini.

This discussion is in the form of the results of research that has been carried out using the previous K-Nearest Neighbor method. The system is built using the Jupyter Notebook open web application for Python programming. There are several libraries that compile this

system, namely:

- Numpy = to work with numeric arrays.
- Panda = provisioning the data structure to manage.
- Scikit Learn = create data processing and classification for K- Nearest Neighbor.
- Seaborn = creates graphs and statistics of data.
- Matplotlib = visualization in 2D.
- Plotly = creation of data to make it more interesting.

In the discussion of data, there are 90 data collected as follows Table II.

TABLE II. DATA USED IN CLASSIFICATION

ID	Ventilation	RSSI -AP1	FQ -AP1	Limit-AP1	RSSI-AP2	FQ-AP2	Limit-AP2	RSSI-AP3	FQ-AP3	Limit-AP3
1	1	-46	2412	65	-68	2437	58	-71	2417	19
2	1	-47	2412	72	-56	2437	86	-66	2417	52
3	1	-55	2412	78	-62	2437	78	-70	2417	65
53	2	-71	2412	60	-48	2437	78	-78	2417	40
54	2	-78	2412	51	-51	2437	73	-62	2417	76
55	2	-68	2412	56	-50	2437	66	-77	2417	37
56	2	-71	2412	60	-57	2437	70	-68	2417	39
87	3	-70	2412	50	-62	2437	63	-44	2417	78
88	3	-80	2412	29	-70	2437	50	-48	2417	78
89	3	-72	2412	49	-60	2437	70	-56	2417	72
90	3	-78	2412	37	-65	2437	61	-50	2417	72

This data consists of 30 data obtained from access points to 3 vents in the library of trilogy university, Jakarta. The data is divided into 2, namely training and testing data with a ratio of 80% - 20%. The K- nearest neighbor's value (K_Values) is 4, where the Access point data is worth an odd number, so for this assessment, it uses an even number.

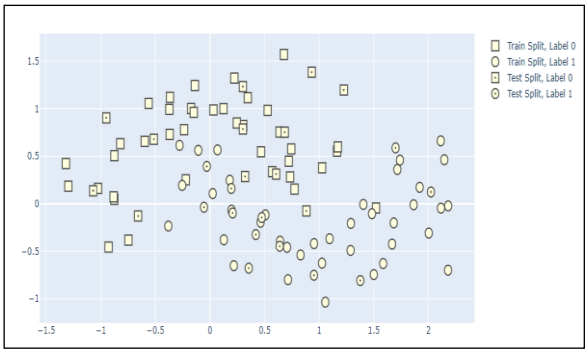

Fig. 1. Data Dissemenation

The picture above is a data dissemination containing RSSI and Limit data, the shape of a square without dots and a circle without a dot is the training data in this study totaling 72 data. Data testing has the form of a square with points in it and a circle of points with a total of 18 data. Data on the data dissemination process read in this classification system there is a positive X-axis with a negative y-axis labeled 1 having the highest number of circle training data accompanied by 7 testing data circle points labeled 1, and 1 testing data point square in it. The positive X axis and the positive Y axis have square training data as well as round and testing points that vary, there are 7 dot squares and 3-point circles in this section.

In the negative x and positive axes of the y axis, there are quite a lot of square training data, but there are 3 testing data in the form of a point square and 1 point circle testing data, and for the negative x and negative axes, the y axis only has a small amount of data,

namely 1-point square shape testing data and 1 point circle data in this section.

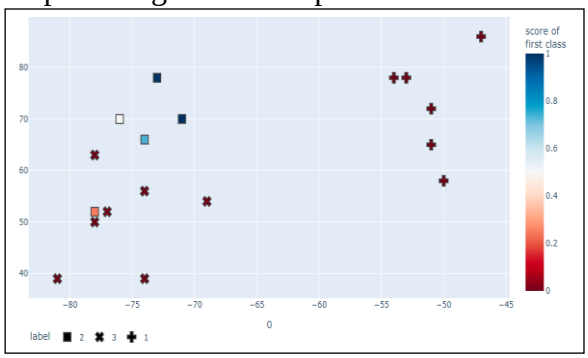

Fig. 2. Mapping Testing Data

From the 18-test data, in the testing data for ventilation 1 with the shape (+) there is 1 data that is more than 80 on the x-axis, and -50 to -45 on the y-axis. 3 data are on the 70 to 80 on the x-axis and -55 to -50 on the y-axis, while on the same y-axis but different from the x-axis by 70 to 6o there is 1 data. For other data, there is 1 data located at -50 on the y-axis and on the x-axis between 50 and 60. In total, there are 6 tests for ventilation 1. This ventilation test data 1 is grouped so that the value can be ascertained to be accurate.

For the 2nd vent, there were 5 data tested with the shape of the square. There are 2 data located between 70 to 80 on the x-axis and the location of the y-axis with values of -75 to -70. The other 1 data is on the 70 x-axes and the y-axis between the values of -80 to -75. 1 data between 70 to 60 on the x-axis and -75 to -70 on the y-axis and 1 other data is located between 60 - 50 on the x-axis and -80 to -75 on the y axis meaning that there is 1 data that has different values and is read classification in other vents.

In vent 3 with the shape (x) with a number of data as many as 7, it has 1 data that is between 70 and 60 on the x-axis with the y-axis which is at -80 to -75. There are 3 data located on the x-axis with values between 60 and 50 but different y-axis locations, 1 on the y-axis with values between -80 to -75, 1 data at -75 to -70, and 1 other data at -70 and -75. There is 1 data at -80 to -75 on the y-axis and the location of the x-axis at 50. 2 data that are below 40 on the x-axis but the y-axis are in the more -80 and others are in the -75 to -70 data on the y-axis. Conclusion for ventilation 3 these data are grouped and do not interfere with other classifications but with far distances. The results of this test data can be seen as follows:

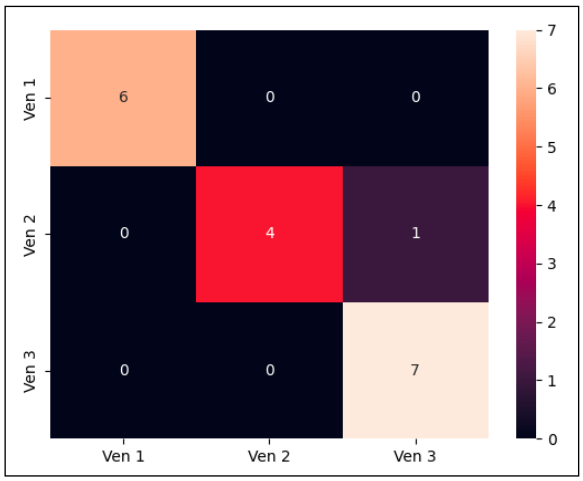

Fig. 3. Confusion Matrix

	precision	recall	f1-score	support
1	1.00	1.00	1.00	6
2	1.00	0.80	0.89	5
3	0.88	1.00	0.93	7
accuracy			0.94	18
macro avg	0.96	0.93	0.94	18
weighted avg	0.95	0.94	0.94	18

Fig. 4. Data Testing Result

Figures 3 and 4, in figures 3, and 4 explain the performance in testing data testing. in vent 1, a predicted True positive value or so-called Precision of 1.00 and a true Positive value of 1.00, and a harmonicity value between precision and recall of 1.00. However, for ventilation 2 the system is able to read a correct prediction value of 1.00 and a real value of 0.80 then affect the value of F-1, which is 0.89 can It is also seen that the system reads 1 data that is detected on the vent that it should not be. While in Ventilation 3, the prediction value is 0.88, the actual value is 1.00 and the f1 value is 0.93.

After the confusion matrix calculation process using precision, recall, and f1-score calculations for all access points produce a value of 0.94444 or 94% accuracy for error values using MAE, it is worth 0.055% of the data testing performed. To test the validation of this system, a test is made on real data consisting of 15 data, namely 5 data for each access point. In this validation test data using RSS-1, Limit-1, RSSI-2, Limit-2, RSSI-3 and Limit-3, the explanation is as follows:

TABLE III. DATA VALIDATION VENTILATION 1

No.	Data	Actual Value	Classification Value
1	-42,69,-65,78,-81,39	Vent 1	Vent 1
2	-46,65,-68,65,-75,26	Vent 1	Vent 1
3	-47,72,-56,85,-66,52	Vent 1	Vent 1
4	-52,78,-67,65,-77,50	Vent 1	Vent 1
5	-53,72,-60,78,-72,72	Vent 1	Vent 1

In this test taken from real data tested by process, the results of this test produce the same value data as the classification data, namely ventilation 1, then in this process, it is considered 100% successful.

TABLE IV. DATA VALIDATION VENTILATION 2

No.	Data	Actual Value	Classification Value
1	-70,68,-44,72,-65,39	Vent 2	Vent 2
2	-75,52,-56,78,-78,26	Vent 2	Vent 2
3	-75,52,-56,70,-62,65	Vent 2	Vent 2
4	-76,48,-54,66,-58,61	Vent 2	Vent 2
5	-79,39,-56,62,-61,65	Vent 2	Vent 2

This test is also taken from real data tested by the process, the results of this test produce the same value data as the classification data, namely ventilation 2, then in this process, it is considered 100% successful.

TABLE V. DATA VALIDATION VENTILATION 3

No.	Data	Actual Value	Classification Value
1	-82,26,-70,65,-54,78	Vent 3	Vent 3
2	-71,26,-70,58,-62,65	Vent 3	Vent 3
3	-70,50,-62,63,-44,79	Vent 3	Vent 3
4	-73,47,-61,56,-39,78	Vent 3	Vent 3
5	-69,52,-69,50,-45,72	Vent 3	Vent 3

Likewise, with the actual data on vent 3, this test produced the same results as the classification, namely ventilation 3. The results of this data validation tested 15 data and produced the same classification as actual values. 15/15 of data declared successful or 100% accurate.

G. DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

[1] I. Jun, R. Geverola, R. C. Mutya, L. M. Bañados-Siason, and A. Bonotan, "Challenges and struggles of public senior high school science teachers during the new normal," Journal of Research, Policy & Practice of Teachers & Teacher Education, vol. 12, no. 1, pp. 49–68, 2022, doi: 10.37134/jrpptte.vol12.1.4.2022.

[2] M. Selvia Laurin and D. Wahyu Wibowo, "Prototype Akses Gedung Perpustakaan Dilengkapi Sistem Peringatan Dini Kebakaran Berbasis Intenet Of Things," vol. 9, no. 2, 2022.

[3] S. Dhanalakshmi, M. Poongothai, and K. Sharma, "IoT Based Indoor Air Quality and Smart Energy Management for HVAC System," in Procedia Computer Science, 2020, vol. 171, pp. 1800–1809. doi: 10.1016/j.procs.2020.04.193.

[4] Y. Bae et al., "Sensor impacts on building and HVAC controls: A critical review for building energy performance," Advances in Applied Energy, vol. 4. Elsevier Ltd, Nov. 19, 2021. doi: 10.1016/j.adapen.2021.100068.

[5] K. C. J. Simma, A. Mammoli, and S. M. Bogus, "Real-time occupancy estimation using WiFi network to optimize HVAC operation," in Procedia Computer Science, 2019, vol. 155, pp. 495–502. doi: 10.1016/j.procs.2019.08.069.

[6] W. Jung and F. Jazizadeh, "Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions," Applied Energy, vol. 239. Elsevier Ltd, pp. 1471–1508, Apr. 01, 2019. doi: 10.1016/j.apenergy.2019.01.070.

[7] D. Sheikh Khan, J. Kolarik, and P. Weitzmann, "Application of an occupant voting system for continuous occupant feedback on thermal and indoor air quality – Case studies in office spaces," Energy Build, vol. 251, Nov. 2021, doi: 10.1016/j.enbuild.2021.111363.

[8] L. Xiong and Y. Yao, "Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm," Build Environ, vol. 202, Sep. 2021, doi: 10.1016/j.buildenv.2021.108026.

[9] F. Suryatini and F. Ilman Fauzandi, "Sistem Akuisis Data Suhu Dan Kelembaban Tanah Pada Irigasi Tetes Otomatis Berbasis Internet Of Things," 2018.

[10] E. Prayetno, D. Nusyirwan, S. Nugraha, A. Faturrahman, and D. P. Winata, "Vessel Monitoring System Using GSM Communications Network," E3S Web of Conferences, vol. 324, p. 05003, 2021, doi: 10.1051/e3sconf/202132405003.

[11] M. J. Habbibie, Y. Yaddarabullah, and A. Syahputra, "Plant recommendation using Triple Exponential Smoothing and K-Nearest Neighbor based on Internet of Things," IOP Conf Ser Mater Sci Eng, vol. 1098, no. 6, p. 062010, Mar. 2021, doi: 10.1088/1757-899x/1098/6/062010.

[12] Z. Arifien, F. A. Bachtiar, and N. Yudistira, "Pengenalan Aktivitas Manusia Menggunakan Sensor Akselerometer Dan Giroskop Pada Smartphone Dengan Metode K-Nearest Neighbor," vol. 9, no. 1, pp. 173–180, 2022, doi: 10.25126/jtiik.202295593.

[13] M. Nishom, S. Febbi Handayani, and P. Harapan Bersama, “Pillar Algorithm in K-Means Method for Identification Health Human Resources Availability Profile in Central Java,” 2021.
[14] S. G. K. Patro et al., “A Hybrid Action-Related K-Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems,” IEEE Access, vol. 8, pp. 90978–90991, 2020, doi: 10.1109/ACCESS.2020.2994056.
[15] R. Andrian, D. Maharani, M. A. Muhammad, and A. Junaidi, “Butterfly identification using gray level co-occurrence matrix (Glcm) extraction feature and k-nearest neighbor (knn) classification,” Register: Jurnal Ilmiah Teknologi Sistem Informasi, vol. 6, no. 1, pp. 11–21, 2020, doi: 10.26594/register.v6i1.1602.
[16] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, “Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking,” IEEE Access, vol. 8, pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.
[17] Isman, Andani Ahmad, and Abdul Latief, “Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal,” Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 3, pp. 557–564, Jun. 2021, doi: 10.29207/resti.v5i3.3006.