MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 16, No. 1(2024), pp 36-42 DOI : 10.18860/mat.v16i1.25747
Received : January 18th 2024; **Accepted** : February 4th 2024 ; **Avalaible Online** : March 28th 2024

IoT and Machine Learning in Smart Kitchen Monitoring for Enhanced Worker Health

M. Aldiki Febriantono

Abstract— **The significance of occupational health in culinary settings, particularly kitchens, is paramount due to the inherent health risks associated with these environments. This study addresses the necessity of maintaining optimal environmental conditions, such as temperature, humidity, and air quality, in kitchens to safeguard worker health. To achieve this, the study advocates for the implementation of sophisticated ventilation and air conditioning systems. The core focus of the research is the integration of Internet of Things (IoT) technology and advanced machine learning algorithms for the real-time monitoring and assessment of kitchen environments. Specifically, the study fine-tunes and evaluates several classification algorithms, including Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), aiming to accurately predict and manage kitchen conditions. The comparative analysis reveals that the DT algorithm outperforms others, demonstrating exceptional accuracy (97.41%), precision (95.35%), and proficiency in identifying relevant scenarios (88.57%). In contrast, the KNN algorithm registers the lowest accuracy (75.12%), while the SVM algorithm, despite being the least precise (86.55%), shows a relatively higher capability in recognizing pertinent cases (86.55%) compared to KNN (72.33%). This study underscores the potential of integrating IoT and machine learning in enhancing occupational health standards in kitchen settings.**

Index Terms— **Air quality, Condition monitoring, decision tree, Humidity, internet of things, Occupational health, Temperature.**

I. Introduction

IN In the domain of industrial manufacturing, the ergonomics of the workplace, encompassing the design of human-machine interfaces and the configuration of workstations, is a prevalent concern. This is addressed by Occupational Health and Safety (OHS) regulations that are instituted to ameliorate worker conditions within these environments [1]. It is imperative that the design of such workplaces is congruent with human physiological and psychological characteristics, as well as the nature of the tasks executed, to diminish occupational hazards, bolster worker well-being and safety, and curtail the probability of human error. Specifically, in culinary settings, the conditions under which kitchen staff operate particularly during prolonged cooking tasks have a substantial impact on their health. The microclimate of a kitchen, defined by variables such as ambient temperature, relative humidity, and air quality, warrants close scrutiny [2]. Thermal conditions falling outside a comfortable range can precipitate health issues ranging from hypothermia at one extremity to dehydration and heatstroke at the other. Fluctuations in humidity levels may foster microbial growth, adversely affecting dermatological and respiratory health [3]. Moreover, air quality that fails to meet established standards can pose significant health risks, often attributable to diminished oxygen saturation and exposure to noxious gases, such as LPG.

In response to these challenges, proactive monitoring and regulation of air turnover through the implementation of ventilation and air conditioning systems are vital to modulate the kitchen's thermal environment. The deployment of the Internet of Things (IoT) is instrumental in facilitating real-time surveillance of these conditions through an array of sensors that monitor temperature, humidity, and air quality indices [4]. The integration of such IoT systems is essential in ensuring effective air exchange in high-activity culinary spaces, necessitating the inclusion of smoke and steam extraction systems as part of the cooking infrastructure. The establishment of optimal ambient conditions is not only conducive to maintaining an adequate environmental quality but is also essential for safeguarding the health of kitchen personnel. To this end, an automated monitoring system has been engineered to regulate ventilation and air conditioning, thus laying the groundwork for a 'smart kitchen' paradigm that underscores occupational health [5].

Algorithms are integral to the operational accuracy and precision of monitoring systems. In this realm, the decision tree algorithm emerges as a robust tool for predicting real-time conditions. Its principal advantage lies in deconstructing intricate decision-making processes into simpler, more interpretable components, thereby facilitating an enhanced comprehension of

Manuscript received September 10, 2023. This work was supported in part by Indonesian Muslim University.
M. Aldiki Febriantono Bina Nusantara University, Jakarta and 11480, Indonesia (email: m.aldiki@binus.ac.id)

problem-solving strategies among decision-makers. Furthermore, decision trees are instrumental in data exploration, unearthing hidden correlations between input variables and their respective target outcomes [6].

The amalgamation of data examination and modeling makes decision trees particularly effective in the initial stages of modeling, extending through to the culmination of the final model. This versatility has garnered widespread utilization in process monitoring system research. An exemplar of this application is found in the work of Rajeswari et al., where the C5.0: Advanced Decision Tree (ADT) algorithm was employed to analyze soil characteristics for optimizing crop selection and sowing times. The outcomes of this research have been applied in the creation of Designing Smart Information System (DSIS) applications. Another noteworthy study involves the design of a traffic monitoring system, integrating the C5.0 decision tree algorithm with time-series analysis. Utilizing the KDD Cup 99 dataset for simulation and comparative testing against conventional traffic monitoring methodologies, the system demonstrated a notable efficacy in monitoring unanticipated attacks, achieving an accuracy rate of 96%. This research exemplifies the applicability and effectiveness of decision tree algorithms in diverse monitoring scenarios [7].

Recent advancements in research have leveraged the integration of Internet of Things (IoT) technology for environmental and health monitoring applications. Apriandy et al. explored the deployment of an IoT-based wireless system for monitoring and predicting water quality in public swimming pools, employing the ID3 (Iterative Dichotomiser 3) algorithm. The system exhibited impeccable performance with a 100% accuracy rate in its predictive capabilities. In a separate study, Binu et al. developed a smart healthcare monitoring system focused on children, integrating Apache Ranger and the C4.5 decision algorithm for secure data transmission and child behavior analysis. The empirical results underscored the C4.5 algorithm's enhanced accuracy over the ID3 algorithm in this context [8].

Furthermore, Bambang et al. utilized the C4.5 algorithm for air quality classification at sensor nodes, employing entropy values and information gain for decision tree construction and rule set formulation. This approach yielded an 85.71% accuracy rate, 81.82% precision, 60.00% sensitivity, and 92.31% specificity in air quality data classification. Additionally, Basuki et al. applied the C4.5 algorithm within a decision tree framework for real-time gas condition monitoring in a power diagnostic application. This was integrated into a SCADA system, demonstrating a prediction accuracy of 95.54% using training data [9].

The present research aims to enhance occupational health in kitchen environments by developing a smart kitchen monitoring system that integrates IoT with decision tree algorithms. This system is designed to continuously monitor key environmental parameters such as temperature, humidity, and air quality. It employs automated actuators triggered by sensor data, with decision tree algorithms (ID3, C4.5, and C5.0) being evaluated to determine the most effective model for predicting kitchen conditions based on the human health index. The goal is to establish the most efficient classification model for implementation in the smart kitchen monitoring system, contributing to improved workplace health and safety.

II. THEORETICAL FOUNDATION

A. Machine Learning

In the domain of artificial intelligence, machine learning exemplifies an innovative paradigm that utilizes algorithmic strategies and data analysis to emulate the iterative learning process of humans, thereby augmenting the decision-making acumen of computational models [10]. Classification, a pivotal machine learning technique, entails the training of algorithms to systematically categorize data into distinct classes based on learned patterns from feature sets. This methodology is imperative in a spectrum of applications, encompassing spam detection, image classification, and medical diagnostic processes. The operational framework of classification involves the initial training phase on a pre-labeled dataset followed by the predictive phase, where the trained model extrapolates the learned patterns to classify new datasets. Among the plethora of classification algorithms, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees (DT) are noteworthy. SVM is proficient in delineating high-dimensional datasets through optimal hyperplane separation [11] [12] [13]. In contrast, KNN predicates its classification on the proximity and voting majority of the closest data points, whereas DT is renowned for its transparent decision-making process, facilitated by an interpretable tree-like structure that sequentially segments the data based on feature entropy and information gain, thereby solidifying their roles as indispensable instruments for predictive modeling in various scientific and industrial sectors. The comparison of each algorithm is shown in Table I.

TABLE I. COMPARISON OF MACHINE LEARNING ALGORITHMS

	SVM	KNN	DT
Type of data	Numerical, continuous	Numerical, categorical, binary	Numerical and categorical
Speed	Slow with large datasets	Slow with large datasets (search based)	Fast, even with large datasets
Technic	Hyperplane separation	Majority voting of k-nearest	Feature value splits
Formula	w * x + b = 0	No explicit formula	Entropy & Information Gain

B. Decision Tree

The Decision Tree (DT) algorithm, a cornerstone in the repertoire of machine learning techniques, leverages the concept of information entropy to facilitate the decision-making process. Information entropy, as delineated by the equation [14].

$$H(S) = -\sum_{i=1}^{n} p(s_i)\log_2(p(s_i)) \quad (1)$$

serves as a metric for the disorder or randomness inherent in a dataset. The primary aim in employing the DT algorithm is the systematic reduction of this entropy, which is achieved by partitioning the dataset based on specific attributes.

$$\mathrm{Gain}(S, A_i) = H(S) - \sum_{\alpha \in \mathbf{A}_i}^{n} \frac{|S_i|}{|S_i|} H(S_\alpha) \quad (2)$$

In this context, the significance of information gain is paramount. It measures the anticipated decrease in entropy and consequent increase in purity following a dataset split based on a chosen attribute. Entropy, denoted as H, symbolizes the essential information required for accurately classifying an element within the dataset. The term in the entropy formula represents the optimal code length in bits needed to encode a label, based on its probability. Through the strategic selection of attributes that offer the maximum information gain at each decision node, the Decision Tree (DT) algorithm effectively constructs a model. This model minimizes the number of necessary inquiries to reach a definitive classification, thereby optimizing the efficiency of the predictive model.

C. Support Vector Machine

The Support Vector Machine (SVM) algorithm is a prominent and flexible tool in machine learning, highly regarded for classification and equally capable in regression scenarios. It specializes in processing numerical and continuous data types, effectively identifying a hyperplane in an N-dimensional space (where N indicates the number of features) to distinctly classify data points into various groups. The crux of SVM's methodology lies in maximizing the margin, which is the distance between the nearest data points of each class and the hyperplane, an approach that significantly boosts classification accuracy. The hyperplane is mathematically represented by the equation $w \times x + b = 0$, where w is the weight vector, and b is the bias. In cases where the data is not linearly separable, SVM utilizes the kernel trick to project the data into a higher-dimensional space, enabling effective separation [15]. This technique supports various kernel functions like polynomial, radial basis function (RBF), and sigmoid, thus adeptly handling intricate datasets. However, SVM's performance may diminish with very large datasets due to its computational intensity in these scenarios. Despite this, its ability to prevent overfitting, especially in high-dimensional spaces, makes SVM an invaluable asset in dealing with complex, small- to medium-sized datasets in diverse machine learning applications.

D. K-Nearest Neighbors

K-Nearest Neighbors (KNN) algorithm, a staple in machine learning, is employed for both classification and regression tasks, renowned for its straightforward yet effective approach, particularly in situations with non-linear decision boundaries. KNN functions as an instance-based or lazy learning algorithm and is adept at handling various types of data, including numerical, categorical, and binary [16]. It determines the classification of a new data point by identifying the 'K' closest data points (nearest neighbors) within the training set, using distance metrics like Euclidean, Manhattan, or Hamming distance. The final classification is derived from the majority vote among these 'K' nearest neighbors, while in regression, it is based on their average. One key characteristic of KNN is that it doesn't rely on an explicit formula for its operation. However, a significant drawback of this method is its reduced efficiency with large datasets, as it becomes computationally intensive due to the necessity of performing distance calculations for each new prediction. The choice of 'K' is critical; a smaller 'K' makes the algorithm sensitive to noise, whereas a larger 'K' tends to over-smooth the decision boundary. Despite these challenges, KNN's ability to adapt to complex and diverse data patterns without assuming any specific distribution makes it a valuable, versatile tool in the machine learning toolkit [17].

E. Industrial Kitchen Design

Effective ventilation systems are crucial in commercial kitchens for maintaining air balance, a process that necessitates the equilibrium between the expulsion of smoke, oil, and steam and the influx of fresh, clean air. High-traffic kitchen environments mandate controlled ventilation and air conditioning to ensure optimal air quality [18]. The kitchen exhaust system emerges as a pivotal component during cooking activities, as it mitigates the risk of atmospheric pollution from smoke, oil, and steam, factors that can adversely impact employee health. The implementation of efficient airflow systems in commercial kitchens is essential, considering the human health index. This entails the systematic extraction of exhaust air coupled with the introduction of clean air to establish stable airflow patterns within the kitchen space. Figure 1 illustrates the design of an industrial kitchen equipped with a monitoring system, highlighting the critical aspects of ventilation and air management [19].

F. Exhaust Systems

The integration of exhaust fans with Ventilation Air Conditioning systems forms a co-dependent exhaust mechanism in commercial kitchens. These systems are designed to function synergistically, ensuring optimal performance based on predefined parameters. Their primary effectiveness lies in the removal of heat, moisture, and gases emanating from cooking activities. It is imperative for the exhaust system to be strategically positioned above the cooking equipment, facilitating the efficient capture and treatment of air pollutants generated during the cooking process [20]. The exhaust fan plays a crucial role in expelling waste air, thereby aiding in the regulation of temperature, humidity, and gas levels within the kitchen environment. A critical aspect of this system is interlocking, which ensures a seamless connection between ventilation hoods and gas appliances. This setup entails the automatic activation of the air system when the gas supply is turned on, adhering to industrial regulations that mandate active ventilation concurrent with the operation of cooking equipment. Consequently, all exhaust gases are required to be channeled into the atmosphere through designated exhaust vents, simultaneously allowing the influx of clean air, a process crucial for maintaining a healthy and safe kitchen environment [21].

III. RESEARCH METHODOLOGY

A. *System Design*

The kitchen's infrastructure incorporates a meticulously designed ventilation and air conditioning system, specifically implemented to monitor various environmental parameters. In this study, the human health index is quantified based on three critical parameters: temperature, humidity, and air quality. The quantification of these parameters in relation to health values is systematically tabulated in Table II, providing a comprehensive overview of their impact on the overall health index in kitchen environments.. The level of health value is shown in Table II.

TABLE II. LEVEL OF HEALTH VALUE

No.	Parameter	Unit	Standard rate
1	Temperature	ºC	18 - 30
2	Humidity	%Rh	40 - 60
3	Carbon Dioxide	ppm	1000
4	Carbon Monoxide	ppm	9.00

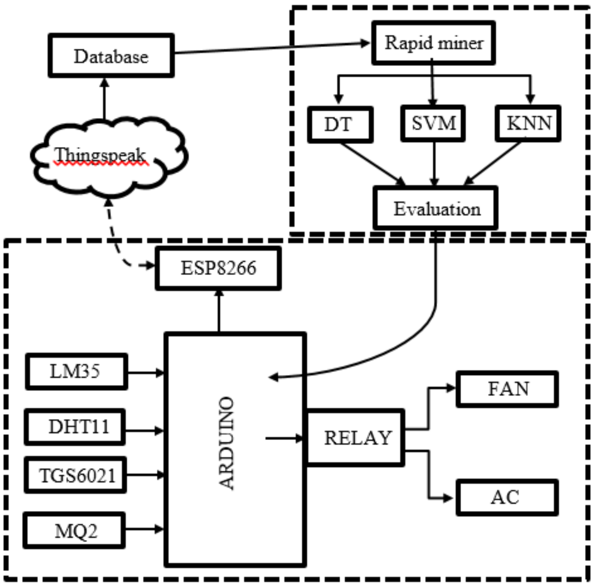

Fig. 1. System Architecture Design

Fig. 2 illustrates the system architecture designed to develop a monitoring system. This system is realized as a prototype set-box that combines an array of sensors with a microcontroller to control actuators effectively. The architecture's foundation is structured around three principal segments: the IoT module (comprising Arduino UNO and ESP8266), an array of sensors (LM35 for temperature, DHT11 for humidity, TGS2601 and MQ2 for air quality monitoring), and actuators (including a FAN and an AC unit). At the heart of the operation, the Arduino functions as the main control unit, interfacing with the ESP8266 for seamless wireless communication, thereby facilitating the relay of sensor data to the Thingspeak database [22]. The data collated in the Thingspeak database is then formatted into a CSV file, prepped for preprocessing and data labeling, which will delineate the datasets for training and testing [23]. Utilizing RapidMiner, the datasets are then subjected to training and testing phases employing machine learning algorithms such as Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Following the evaluation phase, the algorithm that exhibits the highest accuracy will be integrated into the microcontroller programming, laying the groundwork for a sophisticated smart kitchen monitoring system. The classification model's performance metrics, namely accuracy, precision, and recall, are determined using a confusion matrix.. The calculation formula is shown in the following (1)(2)(3) [24].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \tag{7}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \times 100\% \tag{9}$$

B. *Hardware Design*

The kitchen's infrastructure incorporates a meticulously designed ventilation and air conditioning system, specifically implemented to monitor various environmental parameters. In this study, the human health index is quantified based on three critical parameters: temperature, humidity, and air quality. The quantification of these parameters in relation to health values is systematically tabulated in Table II, providing a comprehensive overview of their impact on the overall health index in kitchen environments.

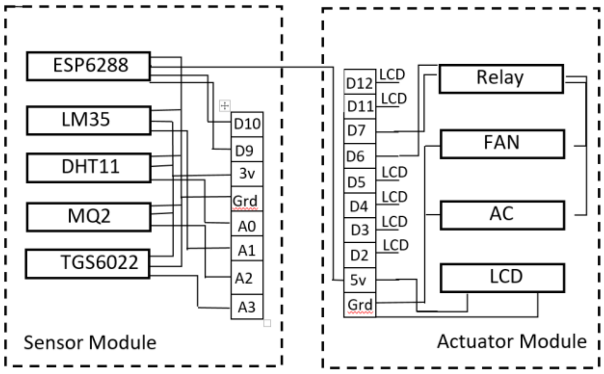

Fig. 2. Pin of Microcontroller i/o Connector

Fig. 3 illustrates the Microcontroller I/O Connector circuit deployed within a smart kitchen monitoring system. The sensor module encompasses an ESP8266 module, which is supplied with a +5V DC input voltage. Concurrently, the LM35, DHT11, MQ2, and TGS2601 sensors are provisioned with a +3V input voltage. The output from the LM35 is directed to the analog pin A1, the DHT11 to A0, the MQ2 to A2, and the TGS2601 to A3. Subsequently, the analog signals are interfaced with the microcontroller's ADC (Analog-to-Digital Converter), facilitating the transformation of analog inputs into digital data. This digital data is then presented on a Liquid Crystal Display (LCD) for visualization and monitoring purposes.

C. *Testing*

The computational apparatus employed for the development of the smart kitchen monitoring system in this study is specified as follows: the processor is an Intel Core i5 with a clock speed of 1GHz; the system's memory is comprised of 4 GB of RAM, storage capabilities are provided by a 500 GB hard drive, and the operating system utilized is Windows 7 with a 64-bit architecture.

The stages starting from the prototype design process to the evaluation of the classification model carried out these steps:

*1) **Step 1**: The prototype has been constructed in adherence to the schematic depicted in Fig. 2.*
*2) **Step 2**: Sensor data values are read and subsequently archived in the Thingspeak database in CSV format.*
*3) **Step 3**: A classification table delineating three categorical states—low, normal, and high—has been established. Labeling of data is conducted with reference to values from three sensors: LM35, DHT11, and MQ2. The standard values for the parameters are derived from prior research [21]. The parameters of humidity and air quality are each classified into three probabilistic states: normal, high, and low. Temperature measurements, however, are categorized into two probabilistic states: normal and high. A 'normal' state is designated when sensor readings fall within the specified parameter range. Readings below the range are labeled as 'low,' and those exceeding the range are classified as 'high,' according to their probability. This study has identified 14 probabilistic conditions based on the data from the three sensors, leading to the creation of 14 corresponding labels. The details of this classification are encapsulated in Table I.*

TABLE I. CONDITION CLASSIFICATION DATASET

Label	Condition	Solution
1	Temp: Normal, Rh: Normal, ppm: Normal	FAN: OFF AC: OFF
2	Temp: High, Rh: Normal, ppm: Normal	FAN: OFF AC: ON
3	Temp: Normal, Rh: High, ppm: Normal	FAN: ON AC: OFF
4	Temp: Normal, Rh: Normal, ppm: High	FAN: ON AC: OFF
5	Temp: High, Rh: High, ppm: Normal	FAN: ON AC: ON
6	Temp: Normal, Rh: High, ppm: High	FAN: ON AC: OFF
7	Temp: High, Rh: Normal, ppm: High	FAN: ON AC: ON
8	Temp: High, Rh: High, ppm: High	FAN: ON AC: ON
9	Temp: Normal, Rh: Low, ppm: Normal	FAN: ON AC: OFF
10	Temp: High, Rh: Low, ppm: Normal	FAN: ON AC: ON
11	Temp: Normal, Rh: Low, ppm: High	FAN: ON AC: OFF
12	Temp: High, Rh: Low, ppm: High	FAN: ON AC: ON
13	Temp: Normal, Rh: Normal, ppm: Low	FAN: ON AC: OFF
14	Temp: High, Rh: High, ppm: Low	FAN: OFF AC: ON

*4) **Step 4**: The dataset was acquired over a duration of one hour, from 10:00 to 11:00 WIB. Subsequent testing of the sensor array involved exposure to varied environmental conditions to evaluate its responsiveness to changes in temperature, humidity, and gas concentrations. These conditions included heating, placement within a humid environment, and exposure to smoke. Data were recorded at 15-second intervals and stored in the cloud-based Thingspeak database. Following data acquisition, preprocessing and labeling were conducted in accordance with the predefined condition classification. Illustrative examples of data corresponding to specific conditions are provided in Figs. 4, 5, and 6.*

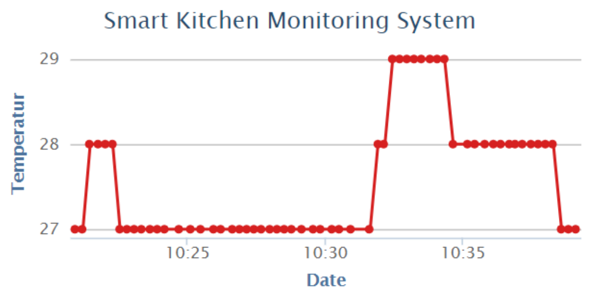

Fig. 3. Sample Data of Temperature Condition

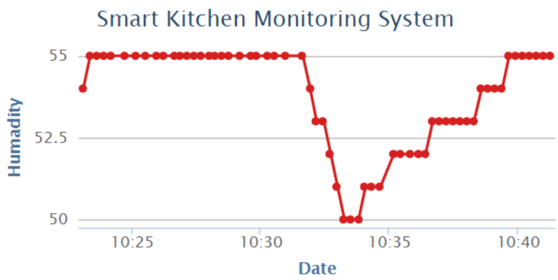

Fig. 4. Sample Data of Humidity Condition

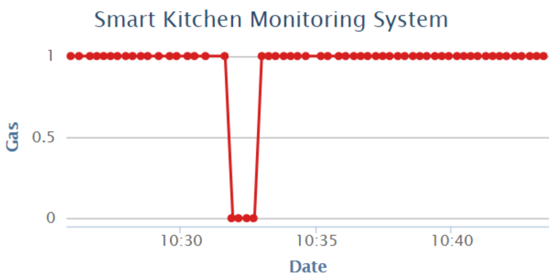

Fig. 5. Sample Data of Gas Condition

Fig. 4 delineates the temperature profile under normal and heated conditions, indicating that the baseline temperature in a normal state is 27°C, which escalates to above 29°C when subjected to heat variations. Fig. 5 illustrates the variations in humidity levels, highlighting the correlation between temperature and humidity; a normal temperature results in 55% relative humidity (Rh), which decreases to below 50% Rh upon increasing the temperature. Fig. 6 presents the air quality metrics within the room, where a normal condition is quantified as 1, while the presence of smoke is denoted by a value of 0, as detected by the sensor.

*5) **Step 5**: The dataset was subjected to evaluation utilizing machine learning algorithms including Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). The* *assessment of these models was conducted using RapidMiner, with an emphasis on calculating critical metrics like accuracy, precision, and recall.*

*6) **Step 6**: The rule sets derived from the decision tree algorithm will be integrated into the system. The model demonstrating the highest performance will be selected for implementation into the microcontroller program of the smart kitchen monitoring system.*

IV. RESULTS AND DISCUSSION

A. *Result*

An evaluation of various machine learning models, namely Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), was conducted using real-time datasets collected at 15-second intervals and stored in the Thingspeak database. The dataset for this study included 309 preprocessed and labeled instances. It was subsequently divided into 80% for training purposes and 20% for testing. The evaluation focused on calculating the metrics of accuracy, precision, and recall. The results of these performance metrics, derived using the 10-fold cross-validation method on the DT, SVM, and KNN algorithms, are detailed in Table III.

TABLE III. ALGORITHM PERFORMANCE

Algorithm	Accuracy	Precision	Recall
DT	97.41%	95.35%	88.57%
SVM	89.35%	86.55%	86.55%
KNN	75.12%	77.08%	72.33%

B. *Analysis*

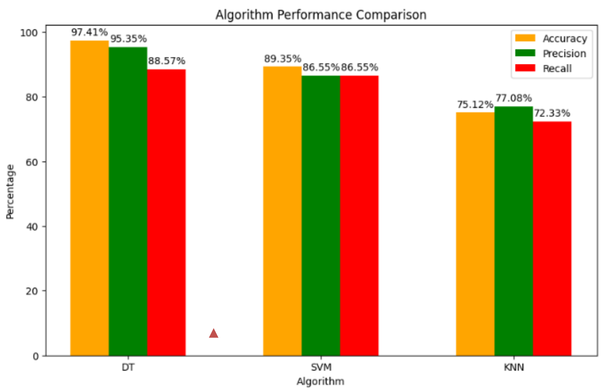

Fig. 6. Graph of Test Result

Based on Fig. 7, the Decision Tree (DT) algorithm exhibits the highest prediction accuracy among the evaluated algorithms, with performance metrics derived from Confusion Matrix test parameters showing an accuracy of 97.41%, precision of 95.35%, and recall of 88.57%. The chart reveals that the DT algorithm outperforms the others in terms of accuracy. This superior performance could be attributed to the characteristics of the dataset and the algorithm's efficacy in handling continuous data, which aligns well with the Decision Tree's capabilities in managing various data splits effectively.

```
○ Graph View ◉ Text View ○ Annotations

Tree

B > 58
|   A > 35: 5.0 {Label=1, 1.0=0, 3.0=0, 5.0=15, 2.0=0, 10.0=0, 13.0=0}
|   A ≤ 35: 3.0 {Label=0, 1.0=0, 3.0=31, 5.0=0, 2.0=0, 10.0=0, 13.0=0}
B ≤ 58
|   A > 30.500
|   |   B > 39.500: 2.0 {Label=0, 1.0=0, 3.0=0, 5.0=0, 2.0=5, 10.0=0, 13.0=0}
|   |   B ≤ 39.500: 10.0 {Label=0, 1.0=0, 3.0=0, 5.0=0, 2.0=0, 10.0=24, 13.0=0}
|   A ≤ 30.500
|   |   C > 0.500: 1.0 {Label=0, 1.0=229, 3.0=0, 5.0=0, 2.0=0, 10.0=0, 13.0=0}
|   |   C ≤ 0.500: 13.0 {Label=0, 1.0=0, 3.0=0, 5.0=0, 2.0=0, 10.0=0, 13.0=4}
```

Fig. 7. Result of Decision Tree Algorithm Rule DT

Fig. 8 depicts the processed data results within a Classification Tree, which generated a structure with 5 nodes and 6 leaves, extending over 5 levels in depth. The root node diverges into two internal nodes based on humidity: one with humidity > 58 and another with humidity ≤ 58. The derived conditions are enumerated henceforth: Condition 1 dictates that if humidity > 58 and temperature > 35, then both FAN and AC should be turned OFF. Condition 2 states that if humidity > 58 and temperature ≤ 35, then FAN should be ON and AC OFF. Proceeding from the node with humidity ≤ 58, it further divides into two nodes distinguished by temperature: one with temperature > 30 and another with temperature ≤ 30. Condition 3: if temperature > 30 and humidity is within > 39 and ≤ 58, then FAN should be OFF and AC ON. Condition 4: if temperature > 30 and humidity ≤ 39, then both FAN and AC should be ON. Condition 5: if temperature ≤ 30, gas > 0.5, and humidity ≤ 58, then both FAN and AC should be OFF. Condition 6: if temperature ≤ 30, gas ≤ 0.5, and humidity ≤ 58, then FAN should be ON and AC OFF.

V. CONCLUSION

Based on research on smart kitchen monitoring systems using the internet of things and decision tree algorithms on human health index based on Confusion Matrix, the following conclusions can be drawn:

- The algorithmic comparison within machine learning frameworks, encompassing Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), indicates a pronounced predictive efficacy in DT, evidenced by accuracy, precision, and recall metrics of 97.41%, 95.35%, and 88.57%, respectively.
- For smart kitchen monitoring systems, the DT algorithm is deduced to proficiently classify data related to human health index parameters, thereby facilitating the real-time monitoring of relative humidity (RH), temperature (°C), and gas concentration (ppm) via IoT integration.
- The Classification Tree method has successfully pinpointed the most influential parameters for the monitoring system, notably temperature, humidity, and gas levels. As a result, the decision rules extrapolated from the DT algorithm are slated for integration into the microcontroller's programming for enhanced smart kitchen surveillance.

- Accruing comprehensive training data is paramount to engender a precise real-time adaptation to kitchen environmental conditions. The magnitude of training data directly informs the solution's accuracy. Prospective studies might incorporate fuzzy logic techniques to modulate FAN speed, aiming to augment the system's monitoring efficiency and efficacy.

REFERENCES

[1] N. Umeokafor, K. Evangelinos, and A. Windapo, “Strategies for improving complex construction health and safety regulatory environments,” *Int. J. Constr. Manag.*, vol. 22, no. 7, pp. 1333–1344, 2022, doi: 10.1080/15623599.2019.1707853.

[2] C. Carlsten, S. Salvi, G. W. K. Wong, and K. F. Chung, “Personal strategies to minimise effects of air pollution on respiratory health: advice for providers, patients and the public,” *Eur. Respir. J.*, vol. 55, no. 6, Jun. 2020, doi: 10.1183/13993003.02056-2019.

[3] G. Guarnieri, B. Olivieri, G. Senna, and A. Vianello, “Relative Humidity and Its Impact on the Immune System and Infections,” *Int. J. Mol. Sci. 2023, Vol. 24, Page 9456*, vol. 24, no. 11, p. 9456, May 2023, doi: 10.3390/IJMS24119456.

[4] D. Singh, M. Dahiya, R. Kumar, and C. Nanda, “Sensors and systems for air quality assessment monitoring and management: A review,” *J. Environ. Manage.*, vol. 289, p. 112510, Jul. 2021, doi: 10.1016/J.JENVMAN.2021.112510.

[5] S. Liu *et al.*, “Improving indoor air quality and thermal comfort in residential kitchens with a new ventilation system,” *Build. Environ.*, vol. 180, p. 107016, Aug. 2020, doi: 10.1016/J.BUILDENV.2020.107016.

[6] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, “Selecting critical features for data classification based on machine learning methods,” *J. Big Data*, vol. 7, no. 1, pp. 1–26, Dec. 2020, doi: 10.1186/S40537-020-00327-4/FIGURES/13.

[7] S. Rajeswari and K. Suthendran, “C5.0: Advanced Decision Tree (ADT) classification model for agricultural data analysis on cloud,” *Comput. Electron. Agric.*, vol. 156, pp. 530–539, Jan. 2019, doi: 10.1016/J.COMPAG.2018.12.013.

[8] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, “Efficient Water Quality Prediction Using Supervised Machine Learning,” *Water 2019, Vol. 11, Page 2210*, vol. 11, no. 11, p. 2210, Oct. 2019, doi: 10.3390/W11112210.

[9] H. Sattar *et al.*, “An IoT-Based Intelligent Wound Monitoring System,” *IEEE Access*, vol. 7, pp. 144500–144515, 2019, doi: 10.1109/ACCESS.2019.2940622.

[10] C. Buizza *et al.*, “Data Learning: Integrating Data Assimilation and Machine Learning,” *J. Comput. Sci.*, vol. 58, p. 101525, Feb. 2022, doi: 10.1016/J.JOCS.2021.101525.

[11] I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, “The effect of gamma value on support vector machine performance with different kernels,” *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5497–5506, Oct. 2020, doi: 10.11591/IJECE.V10I5.PP5497-5506.

[12] Z. Yang *et al.*, “Improving The Performance of K-Nearest Neighbor Algorithm by Reducing The Attributes of Dataset Using Gain Ratio,” *J. Phys. Conf. Ser.*, vol. 1566, no. 1, p. 012090, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012090.

[13] B. Taha Jijo and A. Mohsin Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[14] S. Kaparthi and D. Bumblauskas, “Designing predictive maintenance systems using decision tree-based machine learning techniques,” *Int. J. Qual. Reliab. Manag.*, vol. 37, no. 4, pp. 659–686, Mar. 2020, doi: 10.1108/IJQRM-04-2019-0131/FULL/XML.

[15] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/J.NEUCOM.2019.10.118.

[16] S. Mohsen, A. Elkaseer, and S. G. Scholz, “Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm,” *Smart Innov. Syst. Technol.*, vol. 262 SIST, pp. 304–313, 2022, doi: 10.1007/978-981-16-6128-0_29/COVER.

[17] Z. Pan, Y. Wang, and Y. Pan, “A new locally adaptive k-nearest neighbor algorithm based on discrimination class,” *Knowledge-Based Syst.*, vol. 204, p. 106185, Sep. 2020, doi: 10.1016/J.KNOSYS.2020.106185.

[18] K. E. Paleologos, M. Y. E. Selim, and A. M. O. Mohamed, “Indoor air quality: pollutants, health effects, and regulations,” *Pollut. Assess. Sustain. Pract. Appl. Sci. Eng.*, pp. 405–489, Jan. 2021, doi: 10.1016/B978-0-12-809582-9.00008-6.

[19] J. Saini, M. Dutta, and G. Marques, “Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review,” *Int. J. Environ. Res. Public Heal. 2020, Vol. 17, Page 4942*, vol. 17, no. 14, p. 4942, Jul. 2020, doi: 10.3390/IJERPH17144942.

[20] J. González-Martín, N. J. R. Kraakman, C. Pérez, R. Lebrero, and R. Muñoz, “A state–of–the-art review on indoor air pollution and strategies for indoor air pollution control,” *Chemosphere*, vol. 262, p. 128376, Jan. 2021, doi: 10.1016/J.CHEMOSPHERE.2020.128376.

[21] S. Vardoulakis *et al.*, “Indoor Exposure to Selected Air Pollutants in the Home Environment: A Systematic Review,” *Int. J. Environ. Res. Public Heal. 2020, Vol. 17, Page 8972*, vol. 17, no. 23, p. 8972, Dec. 2020, doi: 10.3390/IJERPH17238972.

[22] A. P. Selvam, S. Najah, and S. Al-Humairi, “The Impact of IoT and Sensor Integration on Real-Time Weather Monitoring Systems: A Systematic Review,” Nov. 2023, doi: 10.21203/RS.3.RS-3579172/V1.

[23] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/J.GLTP.2022.04.020.

[24] M. Aldiki Febriantono, S. Hadi Pramono, G. Naghdy, and R. Scholar, “Classification of multiclass imbalanced data using cost-sensitive decision tree C5.0,” *IAES Int. J. Artif. Intell. (IJ-AI*, vol. 9, no. 1, pp. 65–72, 2020, doi: 10.11591/ijai.v9.i1.pp65-72.

M. Aldiki Febriantono The individual completed their undergraduate studies with a B.Eng. in Electrical Engineering from Brawijaya University, Indonesia, in 2014, followed by a postgraduate degree, M.T. in Electric Engineering, from the same university in 2020. Currently, they hold a position as a Faculty Member in the Department of School of Computer Science at Bina Nusantara University. Their research interests are centered around the Internet of Things, Artificial Intelligence, and Machine Learning. For further correspondence or inquiries, they can be reached at the email address m.aldiki@binus.ac.id.