Designing a Smart Chemical Store with the Aid of the K-Means Algorithm

Samah Faris Kamil[1*] , Mohammed N. H. Al-Turfi[1] , Riyadh S. Almukhtar[2]

[1] College of Engineering, Al-Iraqia University, Baghdad 10041, Iraq
[2] Department of Chemical Engineering, University of Technology, Baghdad 10041, Iraq

Corresponding Author Email: samah.f.kamil@aliraqia.edu.iq

ABSTRACT

Daily dealing with chemicals, directly or indirectly, imposes a specific mechanism of knowledge that coincides with safe handling to reduce risk. This paper introduces an intelligent chemical storage system using the K-Means collection algorithm to improve the safe handling of chemicals and guarantee inventory management. The study aims to aggregate chemicals based on seasonal data variations and employ data collection, feature extraction, and collection techniques. The data set, designed at the Faculty of Chemical Engineering at the University of Technology, consists of different organic and inorganic chemicals stored in different containers. The system collects data on the weather conditions of the storage during summer and winter and captures changes that may affect chemical properties. Data preprocessing involves cleaning, minimization, and Scaler to ensure the integrity of the analysis. The optimal number of groups is determined using the Elbow Plot and Silhouette method, with the K-Means algorithm used for aggregation. The effectiveness of the system is verified by comparing the results with previous research, demonstrating its ability to enhance safety and efficiency in chemical storage management. It can be inferred from its ability to enhance and take the necessary measures promptly to maintain safety and efficiency in managing the storage of chemicals, and as a result, protect humans and the environment from the risks to which they may be exposed.

1. INTRODUCTION

Smart chemical storage has been extensively evaluated as environmentally friendly. Smart stores can enhance unwanted elements and enrich the analysis. Additionally, using a user-closed raw material storage container can improve convenience. Combining substances with carriers and spray drying principles can increase stability and enable the storage of materials, chemical reaction components, and living cells. Materials chemistry plays a role in energy storage research, synthesizing and characterizing electrode materials used in batteries and super capacitors. Figure 1 demonstrates an example of chemical material storage [1, 2].

Figure 1. Smart Storage for chemical material

Chemicals are essential materials used in various fields like chemistry, biology, medicine, materials science, and environmental science. They include elements, compounds, mixtures, and solutions, and are crucial for conducting experiments and understanding material properties. Proper handling and storage of these materials ensures the safety of researchers and protects the environment from potential risks [3, 4].

Volatile organic compounds (VOCs) are rare organic species emitted from natural and man-made sources, contributing to ozone and secondary organic air pollution. High concentrations pose health and environmental concerns in urban and industrial areas. Aromatic and oxygenated VOCs dominate regional concentrations, with industry-related sources dominating in some areas and vehicle-related citations dominating in others [5]. To minimize VOC exposure, store them in well-ventilated containers, label and handle them properly, and monitor storage areas regularly. Implementing process controls and downstream treatment technologies requires innovation and improvement [6].

Artificial intelligence (AI) is increasingly being recognized in the chemistry field due to its significant role in analyzing molecular structures and determining molecular properties. AI is also being actively explored in the development of chemical storage using machine learning techniques and algorithms, an area of research that is currently ongoing and rapidly evolving [7].

Intelligent chemical storage systems using machine learning

groups are a sophisticated method for managing chemical inventory. These systems collect data, extract features, group algorithms, label, organize groups, and improve secure storage. The data includes chemical properties, safety information, frequency of use, and expiration dates. Machine learning algorithms extract essential features for material safety [8].

Smart Chemical Storage Systems use K-Means [9, 10], hierarchical clustering, and density-based clustering to group chemicals based on data features, reducing chemical reactions and accidents. They store items safely considering temperature, humidity, and compatibility. The system updates collections based on new data and trials, prioritizing safety and compliance with MSDS and regulations. It offers an easy-to-use interface for chemical search, safety information, storage verification, and inventory tracking [11, 12].

One of the most difficult challenges faced in this research study is that maintaining clean and healthy air, whether indoors or outdoors, can be extremely difficult. This is especially true in areas where harmful gases such as strong chemicals, ammonia, and alcohol vapors are present. These pollutants are not only dangerous to the health of people in the building, but they can also create unsafe and harmful conditions in the environment. Traditional air quality monitoring systems suffer from some problems. They cannot respond quickly, do not control things well, and cannot track multiple environmental factors at the same time. In addition, it is critical to ensure that the system operates effectively during power fluctuations and that chemical cleaning is done properly to meet environmental safety standards. These are important problems that need good solutions.

The research proposes a model for collecting chemicals in a smart storage system based on artificial intelligence through machine learning. The research paper will be divided into several sections, starting with the second section, will address the topic of comparison between chemical security and chemical protection, and the third section which will address the study and analysis of previous research. The fourth section will discuss the methodology of the proposed system in detail. The discussion of the results of the proposed model will be mentioned in Section V and finally the conclusions of the paper and future work will be the subject of Section VI.

2. SAFETY AND SECURITY IN CHEMICAL MATERIALS

Chemical security and chemical safety are frequently talked about together. They deal with aspects of managing chemicals. Chemical safety is an aspect that emphasizes the prevention of accidents and injuries while working with chemicals. It entails the adoption of practices and protocols to ensure the handling, storage and disposal of chemicals ultimately safeguarding human health and the environment. Some key elements encompassed in safety are:

(1) Risk Assessments: Assessing the dangers linked to exposure, to chemicals.

(2) Safety training is essential: for educating workers on the methods of handling, storing and disposing of chemicals in a manner.

(3) Personal Protective Equipment (PPE): is crucial, in reducing the risk of coming into contact, with dangerous chemicals.

(4) Emergency Preparedness: It is crucial to have developed strategies and protocols in place to handle events such as spills, exposures, and other accidents [13, 14].

Chemical security, however, focuses on the prevention of access, to chemicals that have the potential to be used for intentions, such, as the production of chemical weapons or acts of terrorism. This encompasses:

(1) Limiting access: to chemicals is crucial to ensure that only authorized personnel can handle them.

(2) Maintaining inventory: records of chemicals are crucial, to identify any instances of access or theft [15].

(3) Vulnerability assessments involve identifying and reducing risks associated with the security of chemical facilities and stocks.

(4) Ensuring compliance: with regulations is crucial to prevent the handling of chemicals [16, 17].

Figure 2 explains the difference between chemical safety and chemical security.
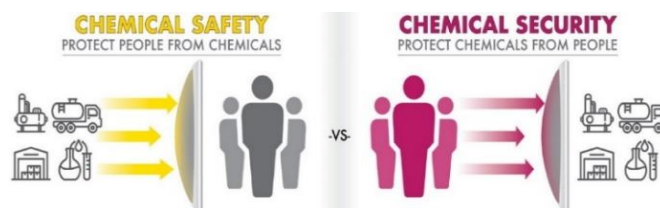
Figure 2. Differential between chemical safety and chemical security [18]

3. RELATED WORKS

This section discusses some of the most recent papers studying chemicals that have been covered by previous researchers, including in 2017 by Borboudakis et al. [19]. Proposed the artificial intelligence approached by using a machine learning algorithm using the predictive modeling tool Just Add Data v0.6 to predict the gas storage properties of metal oxide frameworks (MOF) using machine learning. They tested a machine learning-based methodology to screen MOFs for large-scale gas storage. The researchers used a dataset consisting of structural parameters of 100 metal-organic frameworks while measuring their carbon dioxide and hydrogen storage properties. The results were that the chemical properties of MOFs were stochastically predictable, with prediction accuracy increasing with sample size.

In 2019, an approach was proposed in Rahnama and Sridhar [20] that applies data tools to metal hydrides to improve hydrogen storage. The researchers' goal was to identify distinct bonds and cluster metal hydrides, regardless of the class classification of the parent material. K-Means clustering algorithm and discrete linear convolution methods were used. The data set adopted hydrogen storage materials provided by the US Department of Energy. Three optimal groups were identified, with significant changes in the behavior of the groups occurring when temperature or composition data were removed.

In 2021, Macías-Quijas et al. [21] proposed an approach that works on the concept of employing a compact and affordable electronic device that uses MOS sensors and the filter diameter determination method (FDM) to detect toxic compounds in the air. There was an aspect that the device was used in various settings, including indoor facilities, public transportation, mobile robots and wireless sensor networks. The electronic nose consists of a plastic base with six MOS sensors that change their resistance when in contact with gas. The sensors are placed in a mesh-like structure that filters out

suspended particles, allowing only gaseous elements to enter the chamber. The prototype also includes an electronic unit with a microcontroller for signal acquisition, a USB interface with a computer, and air pump control.

In 2022, Al Hasani et al. [22] proposed an IoT-based fire alarm system that is easy to install and more effective than current systems. It continuously monitors the environment for flames using several sensors. The system uses a central microcontroller, wireless network connectivity, and MQTT communication protocols to queue messages for fire department and user alarms. The prototype was tested and shown its potential. The system provides timely notifications with an average latency of 20 seconds. The prototype shows its real-world applicability.

Cardenas et al. [23] suggested a study-based machine learning strategy to classify iron-nickel smelting lime chemical composition in 2023. Kiln operational data was used to estimate lime formation online. The researchers classified using principal component analysis (PCA and XGBoost). The dataset had 146,632 observations from 6 groups. The accuracy ranged from 82.1% to 85.9%.

Ismael et al. [24] presented work in 2023 to estimate permeation flow in vacuum membrane distillation (VMD) using a hybrid machine learning model. The spotted hyena optimizer (SHO) and support vector machine (SVR) were used to create a unique hybrid model that the authors verified against other machine learning techniques. According to the study's findings, the SVR-SHO model performed better than the other models and had a high accuracy correlation coefficient of 0.94. The most important component that affected flow was found to be feed temperature.

Hossain et al. [25] aimed to develop output deficits brought on by either overstocking or understocking, a continuous chemical manufacturing system's raw material will be stored using inventory management software. Keeping enough packaging materials on hand, controlling inventory prices, and handling supply chain uncertainty are the issues. The process entails utilizing numerical computations for inventory turnover, weeks of supply, and days of supply to analyse and optimize the factors influencing inventory management. The site uses SAP for inventory management; nevertheless, connection with production planning is required to minimize mistakes and inaccuracies. The results show that effective inventory management may lessen the bullwhip impact on the supply chain.

Idama and Ekruyota [26] in 2023 developed a smart storage system for agricultural produce using IoT to address food waste due to inefficient storage. The method involved creating a storage system with a power supply, chamber, CPU, and PCI heater/fan, controlled by an Arduino microcontroller and environmental sensors. The system achieved an efficiency of 85% and a failure rate of 15%, which means a significant advance in automated agricultural storage.

4. PROPOSED SYSTEM METHODOLOGY

The objective of this scientific investigation is to provide a method that may be utilized to create an intelligent storage system that makes use of the K-Means clustering algorithm in order to achieve efficient classification of substances. The system that is being presented incorporates a number of different data processing and clustering algorithms in order to scrutinize and categorize chemicals according to the variations in seasonal data. The process is broken down into the following steps.

4.1 Data collection

Data collection in Chemical Storage provides input for machine learning models, which in turn leads to the improvement of the accuracy and efficiency of the system. This data allows models to adapt to environmental changes and detect chemical leaks early. In addition, it contributes to enhancing sustainability and the ability to predict potential problems, this makes it an essential part of improving operating efficiency and safety. Data collection enhances the personal safety of workers and contributes to improved inventory management and compliance with environmental standards.

The dataset stored for chemicals designed at the University of Technology, Faculty of Chemical Engineering, the area of storage on the ground is $(13.127 \times .82)$ m^2 as shown in Figure 3. The store contains various packages of organic and inorganic chemicals (577 substances), some of which are liquid (168 items) and solid (409 items) stored in containers of (20L, 5L, 2.5L, and 1L). Figure 4 shows a picture of the storage in which the proposed system is designed.

The initial phase involves the systematic collection of store atmosphere conditions [Temp., Humidity, VOC, Alcohol, Ammonia, and Gases] data during two critical seasonal periods, summer, and winter. This seasonal approach is designed to capture variations in data that may arise due to temperature fluctuations and other seasonal factors that can affect the properties of chemicals. These readings are demonstrated in Figures 5 and 6.

Figure 5 shows the summer readings. In summer, chemicals experience a sharp impact due to rising temperatures, leading to an increase in evaporation. This evaporation causes high vapor pressure inside the containers, which increases the risk of leakage, and this is evident through the VOC, alcohol, and ammonia sensors. In addition, they may cause unwanted chemical reactions and damage to materials, leading to illness and potential fires for workers.

Figure 6 shows the winter readings. In winter, the internal environment of the chemical store is affected, as temperatures drop and humidity increases, which causes the risk of freezing some chemicals and leads to a change in their properties and this can cause damage. To avoid these problems, these detectors were used to monitor the environmental conditions inside the storage to maintain the stability of materials, ensure the safety of workers and protect the environment from the negative effects of chemical sabotage.

4.1.1 Data preprocessing

The reason for focusing on data collected in Summer and Winter specifically is that the environment in Spring and Autumn is more friendly and does not affect the chemicals either by evaporation (due to high temps) or destroying the material properties (due to low temps), hence the aggregated data from both seasons is combined to form a comprehensive dataset. This combined dataset then undergoes a cleanup process where all null or missing values are removed to ensure the integrity of the analysis. After the cleanup, the data is condensed so that every minute of runtime is represented by a single data entry. This miniaturisation step is critical for managing data volume and focusing on the information most relevant to clustering.

Chemical Store Scheme
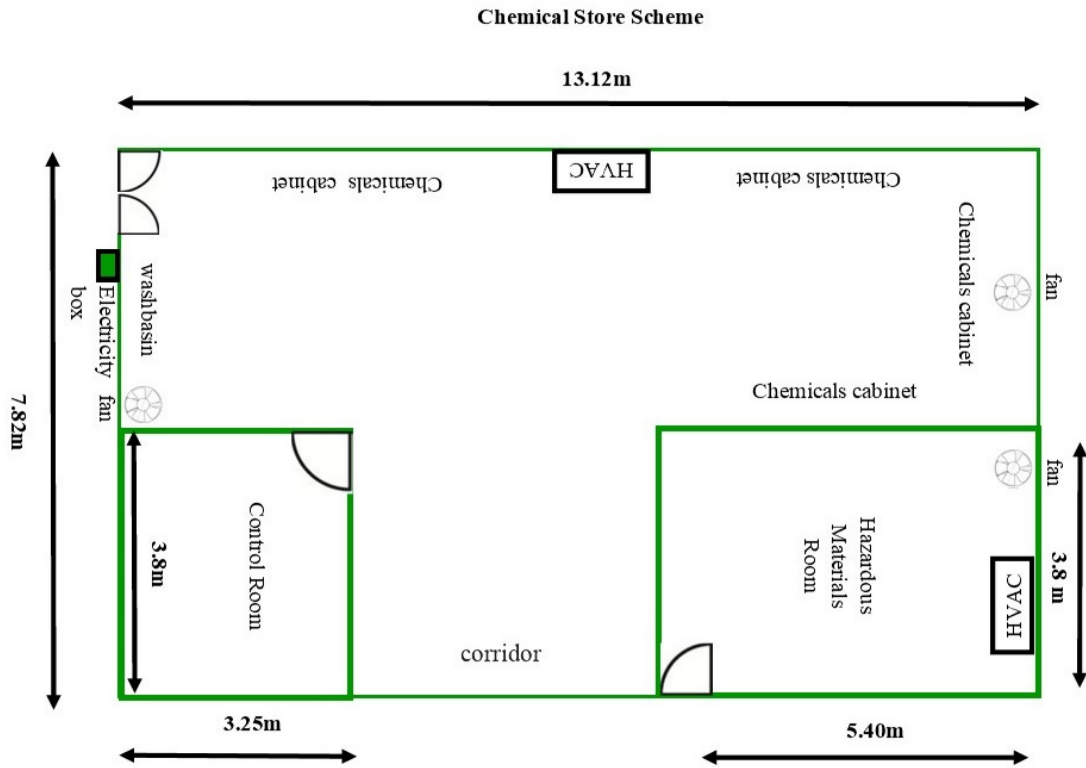


Figure 3. Chemical store that used in this paper

Figure 4. Chemical materials storage

4.1.2 Data scaler

Before the Clustering phase, a scaling technique is applied to the data to standardize feature values to a uniform range. Since the K-Means method focuses on distance computations that might be impacted by variables with different scales, proper scaling is necessary for the algorithm to operate correctly. The data scaler equation is:

$$X_{new} = \frac{X_i - X_{mean}}{Standard\ Deviation} \tag{1}$$
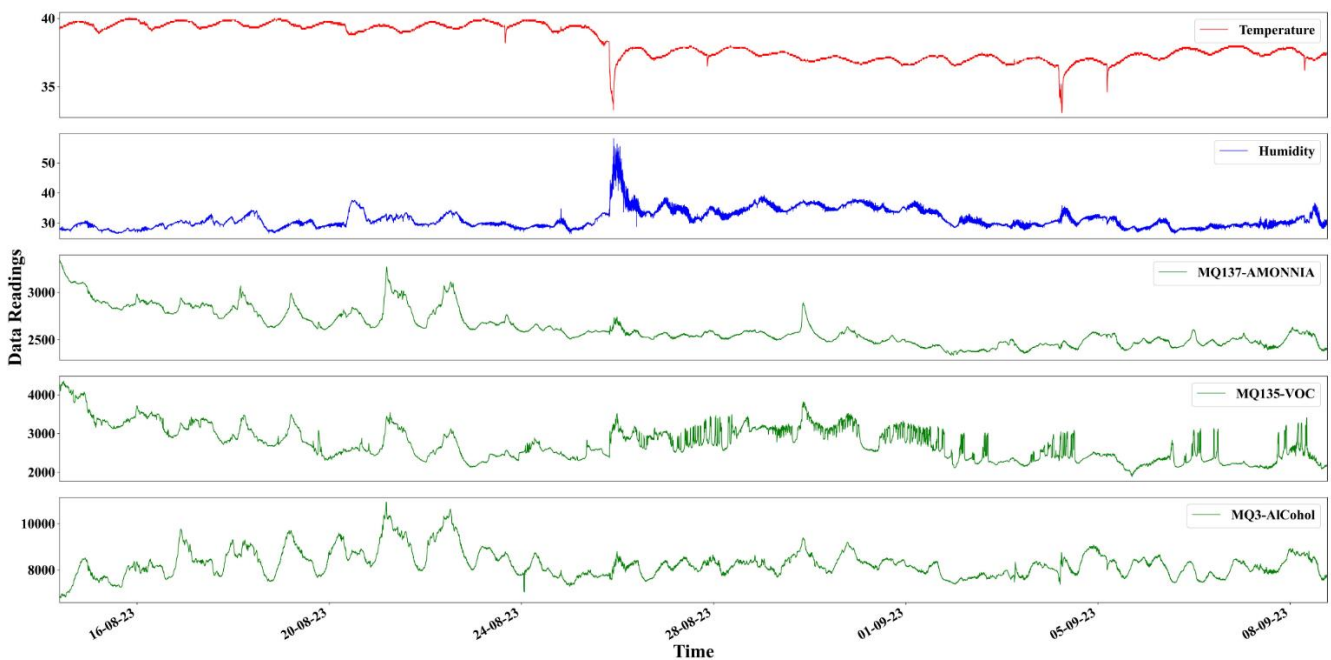
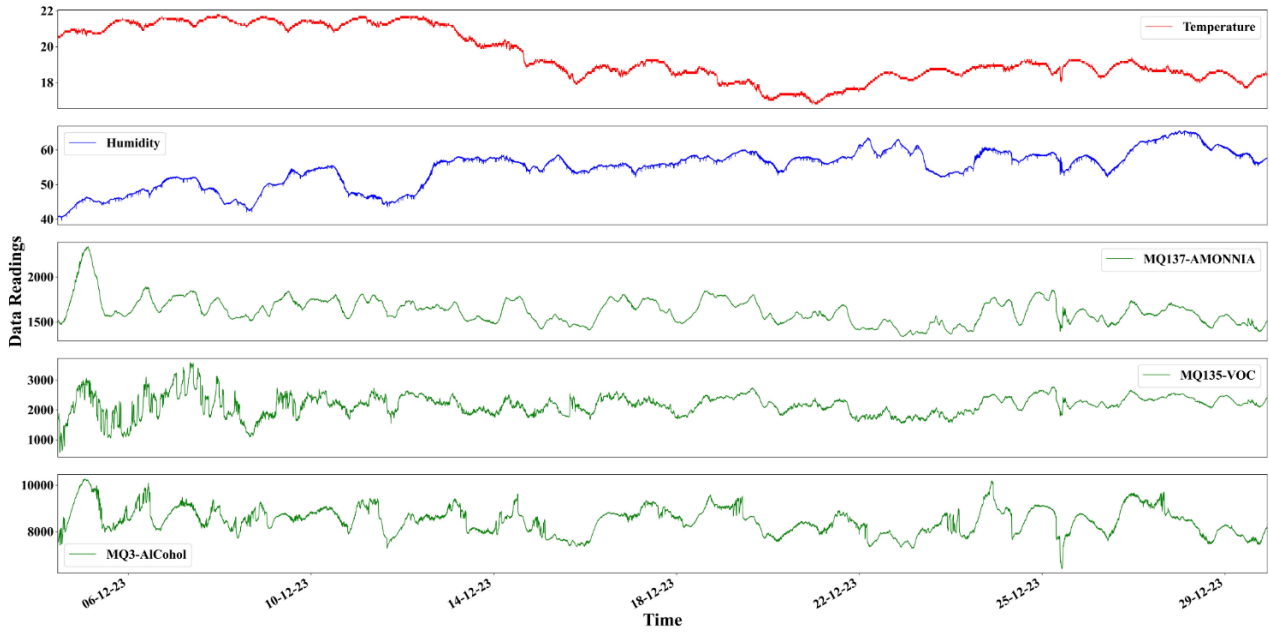
Figure 5. Summer data collection

Figure 6. Winter data collection

4.2 Clustering analysis

Proceed to the cluster analysis stage once the dataset has been scaled and cleaned. Two assessment techniques are used to determine the ideal number of groups for the K-Means algorithm:

4.2.1 Elbow plot

Plotting the Within Cluster Sum of Squares (WCSS), or sum of squares within a group, versus the number of sets results in an elbow diagram. An ideal number of combinations for the K-Means algorithm is shown by the "elbow point" on the figure, which is the point at which the rate of fall in WCSS becomes less noticeable. Determine the WCSS:

$$WCSS = \sum_{i-1}^{n_i} \sum_{x \in 1}^{0} d\left(x_{ij}, c_{kj}\right)^2 \qquad (2)$$

4.2.2 Silhouette method

The silhouette method is used in parallel to assess consistency within groups. The degree of silhouette close to +1 indicates a clear demarcation between groups. This method also suggests an optimal range of groups between 3 and 8. To select the minimum of either WCSS or BCSS.

$$minF = \frac{WCSS}{BCSS} \qquad (3)$$

For each sample, the silhouette value is calculated as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad (4)$$

where, the value $a(i)$ represents the distance from a data point to all points within its group while $b(i)$ signifies the average distance from a data point to all other points, in the nearest neighbouring cluster.

Therefore, the BCSS and WCSS will be used to determine the number of k with the aid of the elbow method.

The elbow method shows relates the total set WCSS as a function of the number of sets, which will be demonstrated later in this paper. Where mathematically, WCSS is calculated by adding the square difference between each point and the middle point of the cluster assigned to it. As the cluster population increases, WCSS tends to decrease because clusters are smaller, and the midpoints are closer to the points in the mass. It is worth noting that this method found that the best number of clusters that can be stopped at is five clusters.

4.2.3 K-Means clustering algorithm

Using the K-Means method, clustering is one of the unsupervised learning approaches in data mining. This approach takes the "K" centroids from the dataset "D" and divides the non-overlapping data points among the nearest clusters. To create the clusters, the following mathematical equation will be used.

Let $\{X_i, i=1,2,\ldots,n\}$ to be set as n patterns.

Further denote $\{x_{ij}, j=1,2,\ldots,d\}$ as the j-th feature of X_i.

J features related to each pattern define:

$$w_{ik} = \begin{cases} 1 & \text{if the ith pattern belongs to the kth cluster} \\ 0 & \text{otherwise} \end{cases}$$
$$i = 1, 2, \ldots, n \text{ and } k = 1, 2, \ldots, K$$

where, K is the number of clusters Matrix.

$$W = |w_{ik}| \text{ has the properties that } w_{ik} \in 0,1 \text{ and } \sum_{k=1}^{k} w_{ik} = 1 \forall i$$

The centeriod of the k_{th} cluster.
$c_k = (c_{k1}, c_{k2}, \ldots, c_{kd})$ is calculated as:

$$c_{jk} = \frac{\sum_{i=1}^{n} w_{ik} x_{ij}}{\sum_{i=1}^{n} w_{ik}} \forall k, j$$

where, c_{kj} determine the mean of the j_{th}.

Feature in the k_{th}.

$clusterc=(c_1, c_2,…,c_d)$ defined is the feature centroid vector.

$$c_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}; j = 1, 2,…,d \qquad (5)$$

The K-Means clustering algorithm steps are [27]:

(1) The K-Means method, also known as centroids, picks several k-points per cluster.

(2) Each data point creates a cluster with its nearest midpoints, resulting in k clusters.

(3) The centroid midway for each cluster is determined by its present members. Here we have new midpoints.

(4) When new midpoints occur, K-Means repeats steps 2 and 3 to find the closest distance between each data point and the new midpoints, connecting them to (k) the new clusters. Repeat this method until convergence occurs, which means the midpoints do not change.

The K-Mean method consists of clusters, each with its own centroid midway. The squareness of the difference between the midpoint and the data points inside the cluster is equal to the sum of the cluster's square value. Furthermore, when the sum of the square values of all clusters is added, the total falls within the range of the square value of the cluster solution. Know that as the number of clusters rises, this value continues to decline, but if you draw the result, you will notice that the sum of the square distance falls abruptly to a certain value of k, and then gradually after that [28, 29]. Figure 7 shows the ideal number of clusters.

Figure 8 shows the serial diagram of the stages that the proposed system goes through, which will be detailed in the following sections.
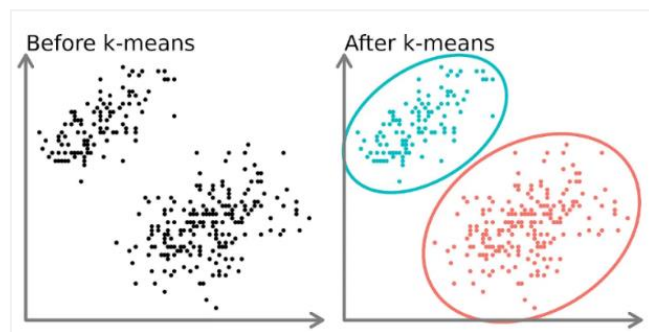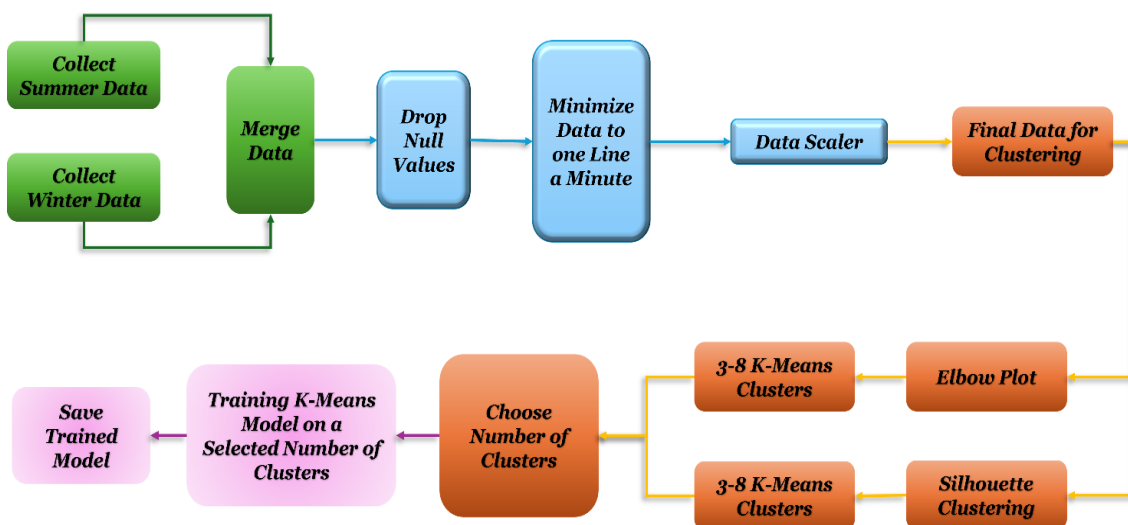

Figure 7. Example of K-Means clustering

Figure 8. Proposed system

4.2.4 Training and selection model

After analyzing the results from both the elbow and silhouette drawing methods, an appropriate number of clusters must be selected. The K-Means aggregation algorithm is then trained on the final dataset using that specific number of sets. The training process involves repeatedly assigning data points to groups and adjusting the position of the middle points of those groups until convergence is achieved. It should be noted that the percentage of convergence required to stop at it is the dangerous affinity between the chemicals in the store.

After successful training of the proposed system that achieves limitation, achievability on the ground, time-bound, measurable, and finally relevant. The K-Means model is saved for operational deployment within the intelligent storage system. The saved model serves as a basic analytical tool for classifying chemicals based on specific groups. It allows real-time classification and decision-making regarding the storage, handling, and processing of chemicals in a storage environment. Integrating this model into a smart storage system enhances the efficiency and safety of chemical management by providing a robust, data-driven approach to material classification.

5. RESULTS AND DISCUSSION

This section illustrates the results shown in the following Figures 9 and 10.

5.1 Silhouette method

The silhouette approach gives a clear visual representation of each object's classification accuracy. A measure of an object's cohesiveness (similarity to its group) and detachment (from other groups) is called the silhouette value.

A high silhouette value implies that the item is poorly matched to neighboring groups and well-matched to its cluster. The silhouette value is a number between -1 and 1. When the

majority of the items have a high silhouette value, the arrangement of grouping is suitable. In the silhouette charts you've provided, each piece corresponds to a different number of combinations (n=3, 4, 5, 6, 7, 8). The width of each piece represents the silhouette score for each sample in the group, and the dashed red line indicates the average silhouette score of the grouping.

The optimal number of groups is the one with the highest average silhouette score and where the score for most individual samples is also high, indicating strong aggregation. The silhouette value ranges between -1 and 1. If the value is close to 1, it indicates that the item is well similar to members of the group to which it belongs and distinct from other groups.

If the value is close to -1, this indicates that there is overlap between the groups, meaning that the item is not similar to the members of the group to which it belongs, but is similar to the members of other groups. If the value is close to 0, it indicates that the item can be equally similar to members of the group to which it belongs and to members of other groups.

In Figure 9, Within Sum of Squares (BCSS) decreases rapidly as the number of groups increases from 3 to 5 and then slowly from 5 to 8. This suggests that after five clusters, adding more sets does not improve the fit of the model. Therefore, it was sufficient for the number of groups in the model to be five clusters.
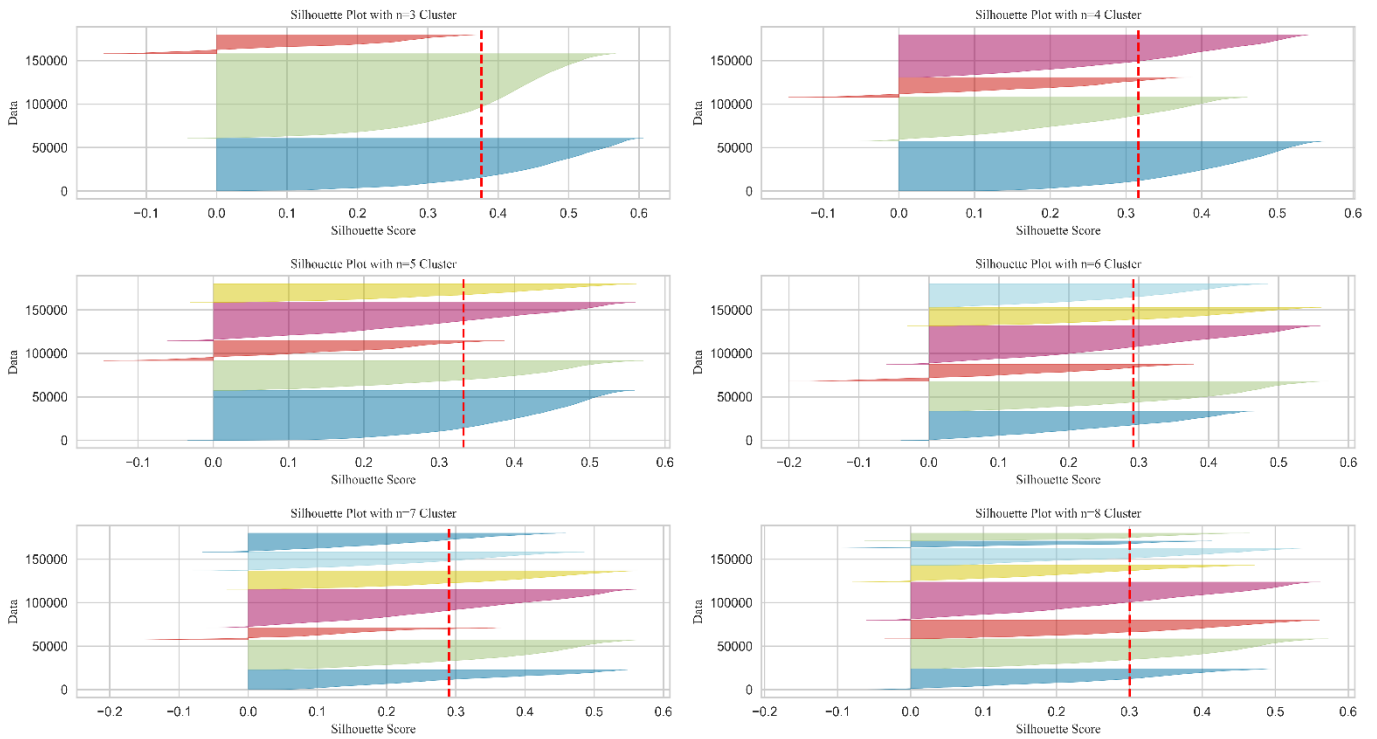

Figure 9. Silhouette method

Figure 10. Elbow method

5.2 Elbow method

As a result, these methods were applied to determine the aggregation of data from the storage, which may include different conditions during the summer and winter. Through the methods used, five optimal clusters were identified in the storage depending on the trial and application made. This means that the system recognizes five distinct periods during the day (day or night) and year (depending on the season) that are important for storage operations, which can correspond to different shifts, temperatures, humidity levels, or other factors related to storage activities. By identifying these five groups, a storage can optimize operations and use energy and other logistics to conform to these specific patterns.

It should be noted that the work initially started by recording ten readings per minute from the store environment as data input to the system, this data is checked and stored as an indication that the process is acting healthily and normally. The number of readings per minute used for the inter-processing is reduced to one reading per minute since the rate of change inside the store doesn't exceed that limited threw a minute, specially within the times of sunrise and sunset
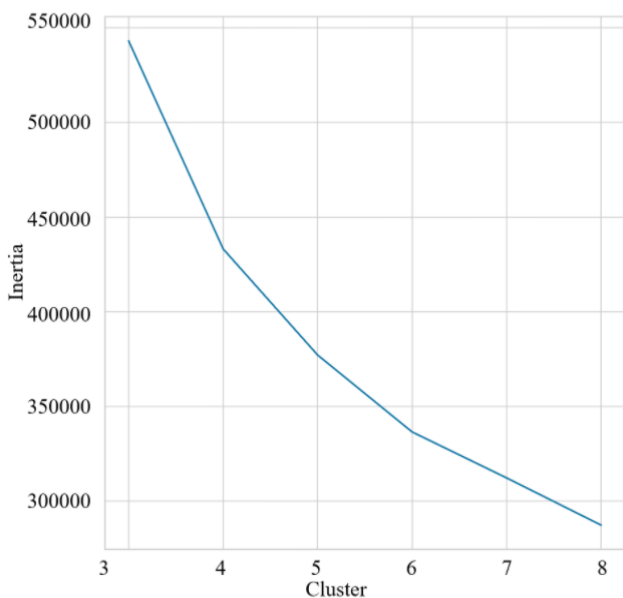
periods due to the change in the position of the sun where higher rates of change in the readings can be indicated, which occurs in less than half an hour, including the change in Heat degrees threw summer and winter. This mechanism was adopted to avoid long-time duration of model training.

Figure 11 demonstrating the notation that adjusting data readings on a reading-minute basis changed the distribution of data points so that five clusters became the most appropriate model based on Silhouette Method and Elbow Method which indicate a better fit of data with five groups to guarantee minimum intersection between the groups.

The results section indicates that after adjusting the temporal details of the data readouts, five clusters were found to be the most appropriate clustering using the K-Means algorithm. This is supported by the clear separation and lack of overlap in the first image, which indicates that each group was distinguished according to its degree of risk and was included within its own cluster. This result thus achieves a more reliable and interpretable overall solution to the data in the store.

These methods were applied to determine the aggregation of data from the storage, which may include different conditions during the summer and winter. Through these methods, determining that dividing times into 5 main times is optimal. This means that the system recognizes five distinct periods during the day that are important for storage operations, which can correspond to different shifts, temperatures, humidity levels, or other factors related to the activities of the

storage. By identifying these five groups, a storage can optimize operations and use energy and other logistics to conform to these specific patterns.

Figure 11 illustrates the cluster distributions of the reading can be tested by the images depicted. So, the X and Y are the new parts that simplify the original data. They are used in the 2D plot to show patterns and connections in the data after PCA has reduced its dimensions. You can look at the cluster distributions of reading in the chart above. These numbers clearly show that changing how we read the data based on the minute led to a different group of data points. As a result, 5 sets turned out to be the best model because the elbow and silhouette methods suggested that 5 groups fit the data better and reduced overlap between the groups.

This is supported by clear separation and no overlap in the first image, showing that each group was marked by its intensity and given its own category. So, this result provides a more reliable and easy-to-understand overall solution for the data in the store. Using these methods, it's clear that dividing time into 5 main sections is the best way to go. The system recognizes 5 important times during the day that are key for running the store. These times could match different work shifts, temperatures, and humidity levels. Levels or other features related to storage activities. Dividing these five groups helps make storage better, reduce energy use, and organize other logistics to match these different patterns.

In Table 1, it is illustrated to compare the results of previous studies with the current model.
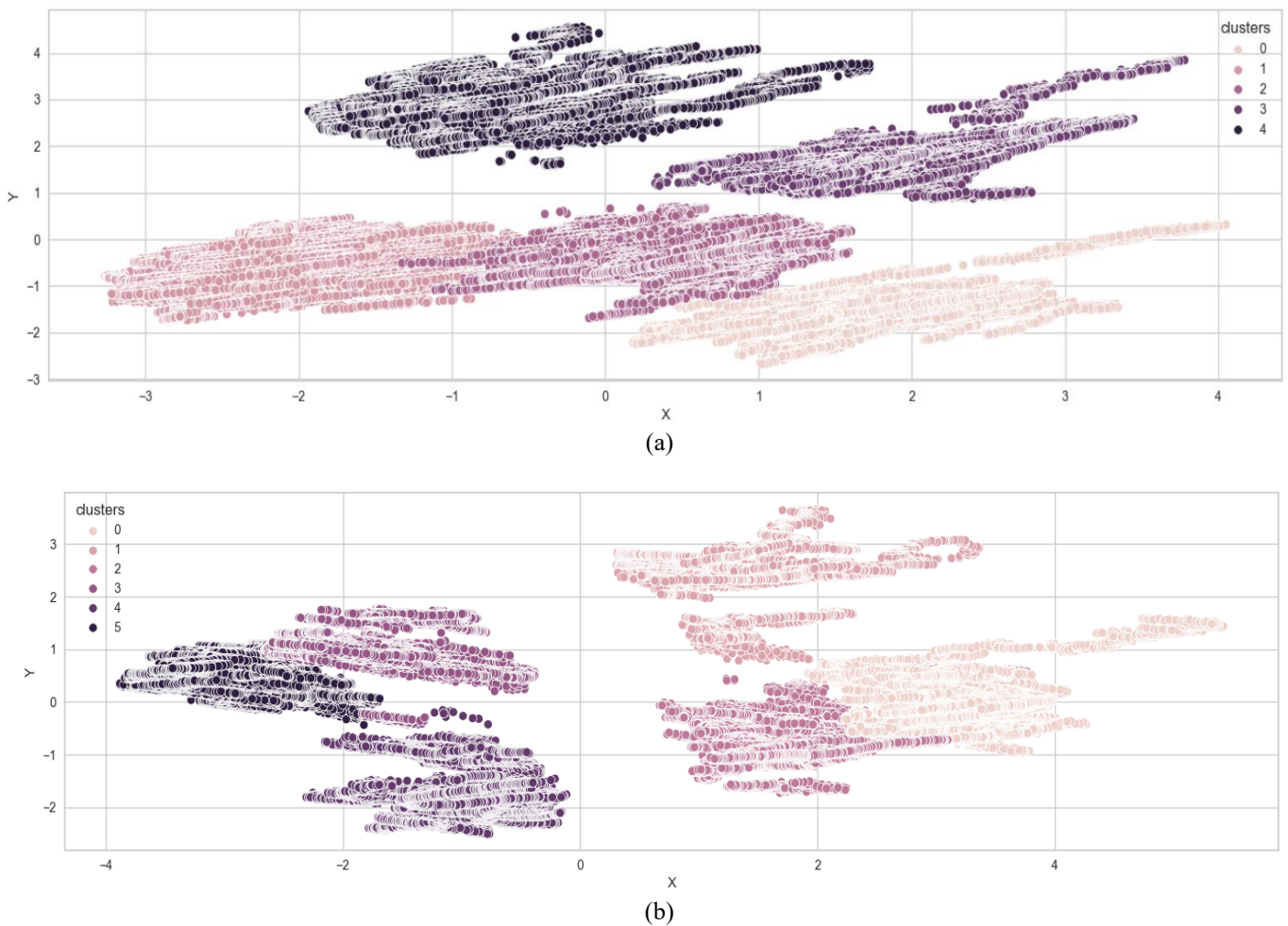

(a)

(b)

Figure 11. The clustering distributions after the readings each minute

Table 1. Comparison table between previous studies and the current model

Ref. ID	Method Used	Results
[19]	The study utilizes machine learning techniques for large-scale screening of MOFs and employs a JAD analysis pipeline for training predictive models.	Developed smart storage system had 85% efficiency in testing. Prototype had a failure rate of 15%.
[20]	Unsupervised machine learning with K-Means algorithm for clustering. Discrete linear convolution method for anomaly detection and identifying outliers. Kernel density estimations to analyze data point distribution.	Fluctuations in the supply chain can lead to significant impacts. Bullwhip effect can be reduced by good inventory management programs.
[21]	Filter Diagonalization Method (FDM) for signal spectral analysis.	SVR-SHO model predicted flux pressure accurately with R-value of 0.94. Feed temperature was the most influential parameter on flux. Permeate flux increased with rising feed temperature. SVR-SHO model outperformed ANN, SVR, and MLR models.
[22]	IoT-based fire alarm system with ESP8266 nodes and various sensors. Prototype tested sensor values, SMS alerts, and sensor node creation.	The proposed method achieved accuracy between 82.1% and 85.9%. Clusters 1 to 3 have the highest Fe and Ni concentrations. Clusters 5 and 6 demand highest electrical power.
[23]	Data-driven approach combining PCA and XGBoost for chemical composition classification. K-Means method for clustering in production planning and productivity increase.	Prototype is functional with suitable FDM approach for classification stage. Preliminary results show the effectiveness of the proposed e-nose device.
[24]	Hybrid model with spotted hyena optimizer and support vector machine. Global sensitivity analysis to interpret results.	Identified optimal clusters as 3 despite 8 material classes. Outliers mainly belong to the complex material class. A 2 B, complex hydrides, and Mg-based alloys clustered together. Removal of temperature or heat of formation alters clustering behavior.
[25]	Inventory management program optimization for continuous chemical manufacturing systems. Addressing gaps between inventory and production planning systems. Analyzing fluctuations in the supply chain to reduce the bullwhip effect. Managing uncertainties in production demand and packaging type demand. Maintaining inventory of over seventy different packaging items.	MOFs' chemical properties are predictable using machine learning methods. Accuracy of predictions increases with sample size. Machine learning guides material discovery by predicting properties of new materials.
[26]	Prototyping methodology with four main components: power source, storage chamber. System performance rating calculated using specific equations.	Prototype successfully sent SMS to warn users and fire department. Average delay of 20 seconds in sending SMS alerts.
Proposed Model	K- Means Algorithm	The algorithm achieved good clustering results, as it was able to divide chemicals under different environmental and time conditions into five clusters with the least percentage of overlap.

6. CONCLUSIONS AND FUTURE WORKS

The paper presents an intelligent chemical storage system that uses the K-Means clustering algorithm to improve chemical management safety and efficiency. The system clusters chemicals based on seasonal data variations, focusing on summer and winter, to optimize storage conditions and reduce the risk of hazardous reactions. Data preprocessing, including scaling and cleaning, ensures accurate analysis. The Elbow Plot and Silhouette method determine the optimal number of clusters, which the study identifies as five. This configuration allows the system to adapt to environmental changes and detect issues early, improving safety and sustainability in chemical storage. The intelligent system not only categorizes chemicals but also detects and mitigates potential hazards, making chemical management more robust and environmentally sustainable. The system's interpretability is enhanced by reducing overlap and ensuring clear separation between clusters. Future work suggests integrating IoT-based sensors for real-time monitoring and enhanced control over chemical storage conditions.

Using the elbow method, it was determined that five groups reduce the variance within the group without unnecessarily increasing the number of clusters. The silhouette method likely confirmed that if there are five groups, data points, on average, are appropriately close to their cluster centers and appropriately distant from other groups. This two-way approach is a powerful way to validate the aggregation option. It should be noted that the elbow was used, and after experimentation and application, it became clear that distributing the data into five clusters according to their degree of severity is the optimal distribution. To verify the result of this method, the Silhouette method was applied, and it proved that distribution into five clusters is optimal and correct after experimentation and application. Future work will be on designing a system of sensors and devices based on the Internet of Things to manage chemicals in smart storage.

REFERENCES

[1] Sadiq, A.A. (2013). Chemical sector security: Risks, vulnerabilities, and chemical industry representatives' perspectives on CFATS. Risk, Hazards & Crisis in Public Policy, 4(3): 164-178. https://doi.org/10.1002/rhc3.12032

[2] Bajjou, T., Sekhsokh, Y., Amine, I.L., Gentry-Weeks, C. (2019). Knowledge of biosafety among workers in private and public clinical and research laboratories in Morocco. Applied Biosafety, 24(1): 46-54. https://doi.org/10.1177/1535676018797140

[3] Lee, J., Mahendra, S., Alvarez, P.J. (2010). Nanomaterials in the construction industry: A review of their applications and environmental health and safety considerations. ACS Nano, 4(7): 3580-3590. https://doi.org/10.1021/nn100866w

[4] Wilujeng, N.F., Swastanto, Y., Joostensz, T.G. (2021). Counter-terrorism cooperation in the ASEAN regional forum (ARF) from the perspective of Indonesia defense diplomacy. Jurnal Pertahanan: Media Informasi tentang Kajian dan Strategi Pertahanan yang Mengedepankan Identity, Nasionalism dan Integrity, 7(2): 205-216. https://doi.org/10.33172/jp.v7i2.728

[5] Mozaffar, A., Zhang, Y.L. (2020). Atmospheric volatile organic compounds (VOCs) in China: A review. Current Pollution Reports, 6: 250-263. https://doi.org/10.1007/s40726-020-00149-1

[6] Kamil, S., Al-Turfi, M., Almukhtar, R. (2024). Advancements in chemical materials: Exploring smart storage equipment and protection systems. Journal of Applied Engineering and Technological Science (JAETS), 5(2): 1086-1101. https://doi.org/10.37385/jaets.v5i2.4096

[7] Mweene, P., Muzaza, G. (2020). Implementation of interactive learning media on chemical materials. Journal Educational Verkenning, 1(1): 8-13. https://doi.org/10.48173/jev.v1i1.24

[8] Council, N.R., Earth, D., Studies, L., Sciences, B.C., Laboratory, C.P.P., Update, A. (2011). Prudent practices in the laboratory: Handling and management of chemical hazards, updated version. Choice Reviews Online, 49(4): 2076. https://doi.org/10.5860/choice.49-2076

[9] Deng, F., Gu, W., Zeng, W., Zhang, Z., Wang, F. (2020). Hazardous chemical accident prevention based on K-Means clustering analysis of incident information. IEEE Access, 8: 180171-180183. https://doi.org/10.1109/ACCESS.2020.3028235

[10] Chiang, L. H., Braun, B., Wang, Z., Castillo, I. (2022). Towards artificial intelligence at scale in the chemical industry. AIChE Journal, 68(6): e17644. https://doi.org/10.1002/aic.17644

[11] Choudhary, N., Bharti, R., Sharma, R. (2022). Role of artificial intelligence in chemistry. Materials Today: Proceedings, 48: 1527-1533. https://doi.org/10.1016/j.matpr.2021.09.428

[12] Bagawan, K., Roshni, M., Jagadeesan, D. (2022). An overview of volatile organic compounds (VOCs). Resonance, 27(12): 2183-2211. https://doi.org/10.1007/s12045-022-1513-0

[13] Kumar, T.P., Rahul, M., Chandrajit, B. (2011). Biofiltration of volatile organic compounds (VOCs): An overview. Research Journal of Chemical Sciences, 1(8): 83-92.

[14] Yang, X., Wang, Y., Byrne, R., Schneider, G., Yang, S. (2019). Concepts of artificial intelligence for computer-assisted drug discovery. Chemical Reviews, 119(18): 10520-10594. https://doi.org/10.1021/acs.chemrev.8b00728

[15] Jain, D., Singhb, T., Singhb, S., Kaur, B.P., Pasricha, R. (2020). Biosensors: An effective toxicity biomonitoring tool. Journal of the Indian Chemical Society, 97: 1416-1425. https://doi.org/10.5860/choice.49-2076

[16] Cao, B., Adutwum, L.A., Oliynyk, A.O., Luber, E.J., Olsen, B.C., Mar, A., Buriak, J.M. (2018). How to optimize materials and devices via design of experiments and machine learning: Demonstration using organic photovoltaics. ACS Nano, 12(8): 7434-7444. https://doi.org/10.1021/acsnano.8b04726

[17] Suyoto, A.W. (2018). Implementasi customer relationship management (CRM) dengan pendekatan clustering berbasis knowledge management (KM). Bachelor's thesis, Universitas Islam Indonesia. https://dspace.uii.ac.id/handle/123456789/6538.

[18] Bashir, S., Hina, M., Iqbal, J., Rajpar, A.H., Mujtaba, M. A., Alghamdi, N.A., Ramesh, S. (2020). Fundamental concepts of hydrogels: Synthesis, properties, and their applications. Polymers, 12(11): 2702. https://doi.org/10.3390/polym12112702

[19] Borboudakis, G., Stergiannakos, T., Frysali, M., Klontzas, E., Tsamardinos, I., Froudakis, G.E. (2017). Chemically intuited, large-scale screening of MOFs by machine learning techniques. npj Computational Materials, 3(1): 40. https://doi.org/10.1038/s41524-017-0045-8

[20] Rahnama, A., Sridhar, S. (2019). Application of data science tools to determine feature correlation and cluster metal hydrides for hydrogen storage. Materialia, 7: 100366. https://doi.org/10.1016/j.mtla.2019.100366

[21] Macías-Quijas, R., Velázquez, R., De Fazio, R., Visconti, P., Giannoccaro, N.I., Lay-Ekuakille, A. (2022). Reliable e-nose for air toxicity monitoring by filter diagonalization method. International Journal of Electrical and Computer Engineering, 12(2): 1286-1298. https://doi.org/10.11591/ijece.v12i2.pp1286-1298

[22] Al Hasani, I.M.M., Kazmi, S.I.A., Shah, R.A., Hasan, R., Hussain, S. (2022). IoT based fire alerting smart system. Sir Syed University Research Journal of Engineering & Technology, 12(2): 46-50. https://doi.org/10.33317/ssurj.410

[23] Cardenas, D.A.V., Leon-Medina, J.X., Pulgarin, E.J.L., Sofrony, J.I. (2023). Data-driven classification of the chemical composition of calcine in a ferronickel furnace oven using machine learning techniques. Results in Engineering, 18: 101028. https://doi.org/10.1016/j.rineng.2023.101028

[24] Ismael, B.H., Khaleel, F., Ibrahim, S.S., Khaleel, S.R., AlOmar, M.K., Masood, A., Alsarayreh, A.A. (2023). Permeation flux prediction of Vacuum Membrane distillation using hybrid machine learning techniques. Membranes, 13(12): 900. https://doi.org/10.3390/membranes13120900

[25] Hossain, N.U.I., Sokolov, A.M., Turner, H.V., Merrill, B. (2023). Development of an inventory management program for warehouse storage of raw materials in a continuous chemical manufacturing system to prevent production deficiencies. In Proceedings of the International Conference on Industrial Engineering and

Operations Management, Manila, Philippines, pp. 85-94.

[26] Idama, O., Ekruyota, O.G. (2023). Design and development of a model smart storage system. Turkish Journal of Agricultural Engineering Research, 4(1): 125-132. https://doi.org/10.46592/turkager.1297511

[27] Chen, R., Wang, S., Zhu, Z., Yu, J., Dang, C. (2023). Credit ratings of Chinese online loan platforms based on factor scores and K-Means clustering algorithm. Journal of Management Science and Engineering, 8(3): 287-304. https://doi.org/10.1016/j.jmse.2022.12.003

[28] Sinaga, K.P., Yang, M.S. (2020). Unsupervised K-Means clustering algorithm. IEEE Access, 8: 80716-80727. https://doi.org/10.1109/ACCESS.2020.2988796

[29] Janßen, A., Wan, P. (2020). K-Means clustering of extremes. Electronic Journal of Statistics, 14: 1211-1233, 2020. https://doi.org/10.1214/20-EJS1689