ILKOM Jurnal Ilmiah Vol. 17, No. 1, April 2025, pp. 54-63
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Tackling Attendance Analysis: Unraveling Employee Patterns using K-means Clustering for Workforce Optimization

Arfiani Nur Khusna [a,1,*]; Wisdah Efendi [a,2]; Nur Arina Hidayati [a,3]
[a] *Universitas Ahmad Dahlan, Jl. Kapas 9, Semaki, Umbulharjo, Yogyakarta 55166, Indonesia*
[1] *arfiani.khusna@tif.uad.ac.id;* [2] *wisdah1800018130@webmail.uad.ac.id* [2]; [3] *nur.hidayati@pmat.uad.ac.id*
** Corresponding author*

Article history: Received August 12, 2024; Revised December 01, 2024; Accepted April 13, 2025; Available online April 20, 2024.

Abstract

This study aims to apply the K-Means Clustering method using employee attendance data. The background of this research problem is to improve the understanding and management of employee attendance by identifying similar attendance patterns in different groups. Employee attendance impacts their morale, sense of responsibility, discipline, cooperation with supervisors or colleagues, and their level of productivity. The K-means Clustering method divides employees into groups based on their attendance patterns, to create groups with similar attendance characteristics. This research has important benefits in decision-making related to human resource management, scheduling, and employee performance evaluation. The results of the study were measured using the Silhouette Coefficient, with a value of 0.3140272065284342, which shows a moderate level of accuracy in separating groups based on attendance patterns. Furthermore, the study also achieved a 100% truth value, signifying the success of consistent and accurate grouping. The main contribution of this research is the use of the K-Means Clustering method as an effective tool in analyzing the attendance of employees and providing valuable insights into managing employee attendance by understanding existing attendance patterns.

Keywords: Attedance; Clustering; Employee; K-Means Clustering; Silhoutte Coefficient.

Introduction

In this increasingly advanced digital age, data collection and analysis has become imperative for various organizations and industries. Performance evaluation is one of the critical aspects of effective business decision making [1]. In this context, a data clustering technique called K-Means Clustering has become popular as a tool for conducting group analysis and performance evaluation [2]. K-Means Clustering is a method of grouping data that focuses on partitioning data into groups that are close to each other based on the similarity of certain features or attributes [3]. This method has been used extensively in many fields, including finance, social sciences, computer science, and more. In the context of performance evaluation, K-Means Clustering can provide valuable insights into the classification and ranking of performance of individuals, teams, products, or business processes [4].

However, although K-Means Clustering has been widely used in performance evaluation, there are still challenges that need to be overcome. First, the use of this method requires selecting the right number of groups, which can be challenging for less experienced users. In addition, the selection of relevant features is also an important aspect in performance evaluation using K-Means Clustering. Selecting appropriate features can help identify relevant patterns in the data and produce more accurate evaluation results. Therefore, this study aims to overcome several challenges in performance evaluation by using K-Means Clustering [5].

Artificial Intelligence (AI) is a field within computer science that aims to create computer systems that can perform tasks that require human intelligence [6]. AI encompasses a wide range of approaches and methods, including machine learning, natural language processing, pattern recognition, logical reasoning, and robotics [7].

The process of analyzing employee absences is a complex and time-consuming task for organizations [8]. In facing this challenge, an efficient approach is needed to identify and understand employee attendance patterns. The use of artificial intelligence (AI), especially the K-Means Clustering method, is an attractive alternative [9]. This method allows grouping of employees based on similar attendance characteristics, enabling organizations to optimize

attendance management and identify potential problems. By implementing K-Means Clustering, organizations can improve operational efficiency and make more informed decisions in managing employee attendance [10].

Analyzing employee attendance data manually is often difficult and time consuming. Complex attendance patterns and large volumes of data make traditional analysis processes less effective [10]. In this case, the K-Means Clustering method can be an efficient solution. By implementing this algorithm, employees can be grouped according to their presence, revealing groups with similar attendance characteristics. This helps in identifying attendance trends, optimizing scheduling, and addressing attendance issues that may occur. The use of K-Means Clustering in employee absence analysis provides a smarter and more efficient approach to managing attendance in the work environment [11].

Adopting the K-Means Clustering method in analyzing employee attendance brings significant benefits to the organization [12]. By understanding existing attendance patterns, organizations can make better plans in terms of scheduling, redistribution of tasks, and developing policies regarding employee attendance. Additionally, by using clustering algorithms, organizations can identify and group employees with similar attendance patterns, enabling more effective monitoring and appropriate action. By implementing K-Means Clustering, organizations can improve operational efficiency, increase productivity, and create a more organized and effective work environment [13].

The Human resource management plays an important role in the reputation of any organization. Absenteeism at work, it affects productivity which is detrimental to the growth of the organization [14]. Due to the decrease in productivity, the financial status of an organization decreases drastically. Absenteeism is one of the factors that affect employee discipline. So, it is necessary to maintain regularity at work. Analyzing employee discipline is the overall responsibility of the organization to maintain its reputation in the industry [15]. This analysis is very useful in the appraisal of employee performance. There are many reasons for absenteeism at work. When analyzing individual discipline, it is customary to take into account all the reasons for absenteeism at work [16]. Machine learning algorithms can be used to analyze absences from job data. To analyze the nature of employee discipline failures based on various reasons for work absence, such as:

1. Survey existing literature related to absenteeism at work.
2. Study the impact on organizational productivity and economics.
3. Analyze the discipline of individual employees based on various reasons for work absence.
4. Compare different classifier models in analyzing individual disciplines to find out the best model suitable for analysis.

This analysis plays an important role in the evaluation of employee performance over time and also in finding the failure rate of discipline in any organization; Organizations can incorporate better measures for human resource management, ultimately the effect of absenteeism on productivity aspects can be reduced. It uses five classifier models in analyzing data [17]. The performance of the models is also compared to find out the better model options for analysis. Five classifier models were used to analyze the attendance dataset. Bayesnet, Naivebayes, SMO, J48 and Multi-layer Perceptron models were implemented on the dataset and resulted in Bayesnet 95.5405%, Naïve bayes 95.4054%, SMO 100%, J48 99.4595%, and Multilayer Perceptron 100% accuracy, respectively [10].

Professionals in the field of primary education have a high rate of sick absenteeism due to work-related diseases, ranking among the first positions in Brazil [18]. Among the main diseases are voice disorders, diseases of the respiratory and musculoskeletal systems, as well as mental and behavioral disorders. Mental health-related notifications have recently increased, as has the case of the state of São Paulo. The working conditions these professionals face, with long working days, large classes and lack of recognition, tend to add to the problem, with direct consequences in the personal lives of teachers [19]. However, there are still few indicators that analyze the condition of teachers at every level of primary education, from early childhood to high school, areas that show differences and specificities in work contexts. Studies that seek to produce indicators usually use aggregate values from different stages of primary education or are based on surveys in specific areas that depend on the actions of respondents, and can incur high operational costs [20]. This study aims to predict the risk of absenteeism due to morbidity of teachers working in early childhood education in urban public schools, using machine learning algorithms and predict cases of sickness absenteeism in public school teachers using machine learning using public data [21]. build a predictive model capable of estimating the risk of absenteeism of early childhood education teachers working in municipal public schools, taking into account all citizens in the state of São Paulo, using artificial intelligence and machine learning algorithms known for high performance, especially in the health field.

The study contributes to more assertive predictions in the field of public health and worker health, allowing to monitor and help prevent the absence of these workers due to morbidity. Cross-sectional study to predict the risk of work absenteeism due to morbidity (work-related illness or disease) in teachers working in early childhood education (preschool period) of the entire public network of the state city of São Paulo, using secondary, public and anonymous data. Classified in RAIS with CBO. For the preprocessing and loading data stages, we use the MS SQL Server Database Manager system. To analyze the data and build predictive models, we used R software. Five machine learning algorithms were developed: logistic regression, decision trees, random forest, XGBoost and artificial neural networks and had an accuracy of 71.52% [21].

Human resources are valuable assets for an agency [22]. The success of an institution is determined not only by the quality of its human resources, but also by the level of discipline. The discipline of an employee in an agency can be seen and measured from the level of attendance in doing a job, because the level of attendance is one of the factors that determine productivity [23]. The problem that exists today is the level of company management that has difficulty in monitoring and controlling employee attendance data. There needs to be mapping and grouping to figure out the pattern of absenteeism [21]. The mapping or pattern obtained helps the management to monitor employees, approach and take action so as to improve employee discipline. This study aims to categorize the groups of attendance and heavy delay, attendance and late as well as absenteeism and delay that are still at normal levels. Grouping into categories whether employees are included in the group of attendance and severe delay, sufficient attendance and delay and absenteeism and delay that are still at normal levels. The patterns and knowledge gained will help the management level to monitor, approach and take action so as to improve work discipline. With the cluster, it can be known the pattern of employee absenteeism based on severe, moderate and normal absenteeism groups. Descriptive modeling with the application of the kmeans clustering method [24].

This study will propose an innovative approach to selecting the optimal number of groups and selecting relevant features to increase the accuracy and usefulness of the evaluation results. It is hoped that this research can provide practical guidance for business decision makers in using K-Means Clustering for more effective and efficient performance evaluation.

Method

The research flow followed on **Figure 1** for this study was the first to collect attendance data which included information on employee attendance. This data includes attendance date, entry time, return time, and other related attributes. Furthermore, the preprocessing process is carried out to clean and prepare the data before analysis. This process includes removing duplicate data, handling missing or incomplete data, and transforming data when necessary. After the data is ready, the K-Means Clustering method is implemented. This algorithm divides employees into groups based on similar attendance patterns. This grouping is done by calculating the distance between each data point to the nearest cluster center. After the clustering process is complete, the results are evaluated and further analyzed to understand the characteristics and interpretation of each group formed. Thus, this study includes data collection attendance, data preprocessing, implementation of K-Means Clustering, and analysis of the resulting results.

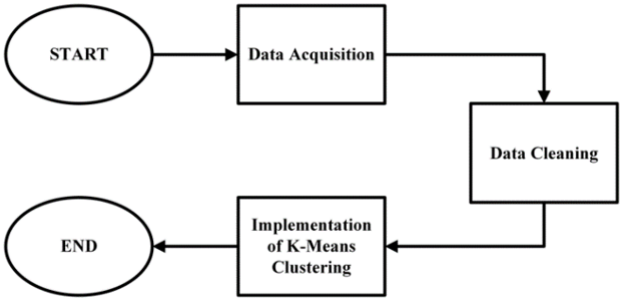

Figure 1. Research Flowchart

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Dataset

The dataset used in this study consists of attendance records, comprising 368 data samples collected over two periods: September to October 2020 and January to February 2021, spanning a total of six months. The attendance data for each educational staff member ranges from a minimum of 0 to a maximum of 26 days per month. This dataset is

utilized to predict the absenteeism of educational staff in the subsequent year. The attendance data is categorized into two types: effective and reduction. "Effective" refers to the actual days attended by the educational staff within a month, while "reduction" indicates the days the staff were absent within the same period. The total of these two categories equals 26, representing the maximum possible attendance days in a month.

B. K-Means Clustering

K-means clustering is a data analysis method used to group a set of data into k groups (clusters) based on similarity of attributes [3]. The main goal of k-means clustering is to minimize variance within each group and maximize similarity between data points within the same group [2]. The k-means clustering algorithm works with the following steps [25]:

1. Initialization, Specify the desired number of k groups. Select k random points as the starting position of the centroid, representing the center of each group.
2. Assignment, For each data point in the dataset, calculate the distance between that data point and each centroid. Assign the data point to the group with the closest centroid. This is usually done using Euclidean distances.
3. Update centroid, Once all data points are placed in a group, recalculate the centroid position for each group by taking the average of all data points in that group. This new centroid position will be the central point representing the group.
4. Repeat steps 2 and 3 until convergence is reached, i.e. there is no change in the placement of data points or the number of iterations has reached the maximum limit. At each iteration, data point placement and centroid position are updated to optimize clustering.
5. Output, Once convergence is achieved, the algorithm will generate clusters with centroid positions representing those groups.

K-means clustering has several advantages, such as being simple and easy to implement, efficient in grouping big data, and can be used for various types of data [27]. However, this algorithm also has some drawbacks, such as being sensitive to centroid initial initialization and prone to achieving optimal local solutions [28]. In practice, k-means clustering is often used in data analysis [29], consumer clustering [2], market segmentation [30], pattern recognition [31], and data compression [32].

C. Evaluation

Evaluating the effectiveness of clustering algorithms through cluster quality analysis using the silhouette score entails finding the ideal number of clusters based on the silhouette coefficient. The quality of clusters is assessed by the silhouette coefficient, which takes into account how closely packed and distinct data points are within each cluster.

Results

A. Processing Data Clustering with K-Means

This stage performs the process of grouping with the K-means algorithm from 368 data. The grouping results produce 4 data clusters. **Table 1** shows the resulting cluster data with membership in each segment. Referring to table 3 that the average attendance for each cluster.

Table 1. Data cluster

Cluster	Average Attendance
1	16.4
2	13.7
3	9.8
4	2.75

B. Experimental Results

The following is a visualization of the distribution of attendance data on the **Figure 2** with a total of 368 data using K-means Clustering using 4 data clusters, including:

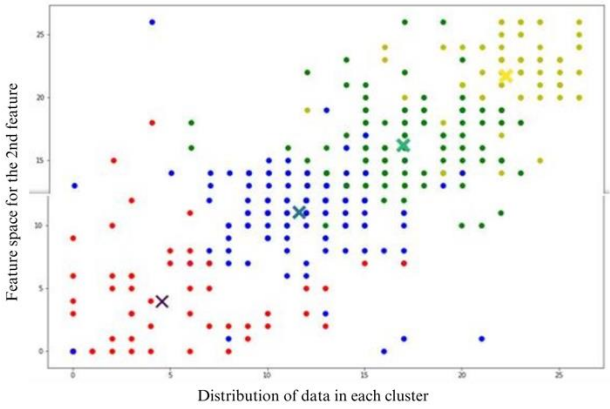

Figure 2. Data visualization

From the test results on **Figure 3** using silhouette coefficient against all education staff attendance data, it can be seen that cluster 0 has a silhouette coefficient value, which is close to 0.6 with a structured value, cluster 1 has a score of 0.5 with a structured value, cluster 2 with a score of 0.5 with an unstructured value because it is close to -0, while cluster 3 with a score of 0.5 has an unstructured value of close to -0.1. Each cluster has a score of 0.5 and 0.6, where the score in this category has a medium structure predicate.

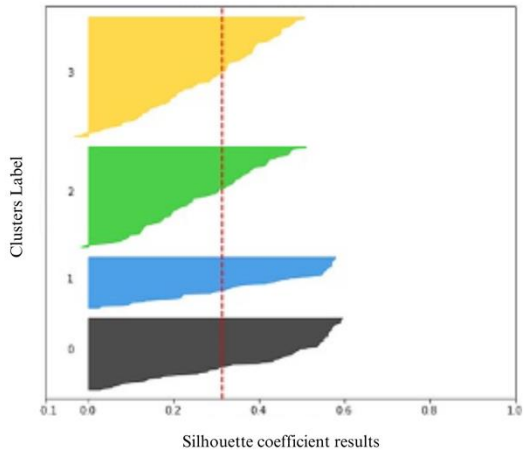

Figure 3. Silhouette coefficient results

Testing the validity of the system in this study used expert judgment testing. Where, testing is carried out by comparing the results of the system with the results of manual calculations, namely if the results of the system and manual are the same, it can be categorized as "Appropriate" and if not the same it will be categorized as "inappropriate". S is mean system and M is mean manual. The results we can see on the **Table 2**.

Table 2. Expert Judgment Testing

ID	Respondent	Cluster Results		Results	Validation
		S	M		
1	Respondent 1	C1	C1	Highly Effective	*Appropriate*
2	Respondent 2	C2	C2	effective	*Appropriate*
3	Respondent 3	C4	C4	Less Effective	*Appropriate*
4	Respondent 4	C2	C2	effective	*Appropriate*
5	Respondent 5	C3	C3	Quite Effective	*Appropriate*
6	Respondent 6	C2	C2	effective	*Appropriate*
7	Respondent 7	C3	C3	Quite Effective	*Appropriate*
8	Respondent 8	C1	C1	Highly Effective	*Appropriate*

ID	Respondent	Cluster Results		Results	Validation
		S	M		
9	Respondent 9	C3	C3	Quite Effective	*Appropriate*
10	Respondent 10	C4	C4	Less Effective	*Appropriate*

From the test results using expert judgment above, the presentation of the truth value of the results obtained will be calculated, it can be concluded that the truth value of expert judgment testing obtained a presentation value of 100% with the appropriate category.

Discussions

K-means clustering is one of the methods used in cluster analysis or data clustering. The purpose of K-means clustering is to divide a certain amount of data into interconnected groups based on similarity of features or attributes [2]. The end result of K-means clustering is the division of data into groups that optimize the distance between the data and the centroid within each group [26]. K-means clustering can be used in areas such as data analysis, pattern recognition, and customer segmentation. The results of K-means clustering for attendance data employees can focus on the interpretation of the groups formed, the characteristics of each group, and the practical implications that can be taken from the clustering results. Based on the analysis of K-means clustering on the attendance data employees, it can be seen that several groups have formed that divide employees based on their attendance patterns. These groups can provide useful insights in further understanding employee attendance in the organization. One of the groups identified is the group of employees with very consistent and regular attendance. They are fully present in every work schedule and show a high commitment to their duties and responsibilities. This can be a positive reference in the management and appreciation of employees who have a good attendance rate. In addition, there are groups of employees with varied attendance patterns and more flexibility. They may have an presence that is affected by factors such as project schedules or specific tasks that require flexibility. The management of this group can involve setting a more dynamic schedule and understanding the needs of individuals in completing their tasks. Another group that emerged was the group of employees with inconsistent attendance. Their attendance patterns tend to be unstable and can affect team productivity. In this case, further investigation is needed to understand the causes of attendance instability and take appropriate steps to improve employee discipline and accountability in carrying out their duties. By understanding employee attendance patterns in these groups, Company management can take more effective actions in schedule planning, performance management, and policy development as appropriate to ensure optimal attendance and organizational efficiency.

Research [27] presents an intelligent attendance system that extracts distinguishable individual phase characteristics to enable the recognition of various targets. The frequency distribution histogram is extracted as a fingerprint and the K-means grouping method is used for target recognition with similar features. The results of the system evaluation show the efficiency and accuracy of the system with an average accuracy of 92%. In addition, the evaluation of the system shows that the design is strong against the differences in clothes worn and the time of day, which further verifies the success of the system's performance. Educational Data Mining (EDM) is a new and growing area of research that uses data mining techniques to extract useful information about the progress behavior of employees or students [28]. Data mining techniques are used to forecast and evaluate students' academic performance based on their academic record and participation in forums. The current study seeks to consider the factors that contribute to the improvement of students' academic achievement in Pakistan. Simple and parallel grouping techniques are implemented and analyzed to show their best features. The Parallel K-Mean algorithm overcomes the problem of a simple algorithm and the result of the parallel algorithm is always the same, which improves the quality of the cluster, the number of iterations, and the elapsed time. The proposed study is more useful for sorting scientific research data, since the statistics of scientific research data are more accurate. Attendance is an important part of educational institutions, industry, and workplaces, and is usually done manually [9]. This paper proposes a system that marks student or employee attendance by recognizing their faces and creating attendance sheets automatically [29]. The accuracy rate of face recognition is affected by changes in illumination, posture, expression, and occlusion. The rules of the K-means clustering algorithm are used to research facial expressions, and the biometric features of the face units are extracted. SVM methodology is used to classify image features, and reports are generated for interpretation. The Assembly Line Worker Assignment and Balancing Problem (ALWABP) is an important research topic originating from sheltered work centers for persons with disabilities. It is important to hire temporary workers to fill labor shortages, and the movement of workers between stations on the assembly line can increase the flexibility of worker assignment. This study investigates new risk aversion ALWABP with uncertain availability of disabled workers, limited temporary workers, and mobile workers. A risk-averse, two-stage stochastic programming model was formulated to assign specific tasks to stations, while a genetic

algorithm combining the K-means clustering approach and variable environment search (GAKV) was designed. The experimental results showed the advantages of GAKV in terms of solution quality and computational time compared to sample average approximation (SAA). In addition, managerial insights are drawn.

Table 3. Competitive results

Author	Years	Amount of Data	Methods	Accuracy
[30]	2018	672	Improved entropy weight method and K-Means cluster algorithm	N/A
[10]	2020	740	Bayesnet, Naïve bayes, SMO, J48 and Multi-layer Perceptron	It is observed that SMO and multi-layer Perceptron models are the best models suitable for analysis with 100% accuracy. Bayesnet (95.5405%), Naivebayes (95.4054%), J48 (99.4595%), Multi-layerPerceptron (100%)
[31]	2021	740	K-Means Clustering	N/A
[32]	2022	26,741	K-Means Clustering	90.20%
[13]	2022	250	Fuzzy K-Means Clustering	N/A
Proposed Model	**2023**	**368**	**K-Means Clustering**	**68.60%**

Conclusion

Based on the results of the K-means Clustering method applied to the problem of grouping the attendance employee with a truth value of 100% and an accuracy rate using a Silhouette Coefficient of 0.3140272065284342, it can be concluded that the success of achieving a 100% truth value shows that grouping the attendance personnel using the K-means Clustering method is very effective. This suggests that groups formed based on attendance patterns have clear and consistent separations. The accuracy rate measured using Silhouette Coefficient of 0.3140272065284342 shows that grouping attendance employees using K-means Clustering has a moderate success rate. The Silhouette Coefficient measures the degree to which each data matches its own group compared to other groups. The higher the Silhouette Coefficient value, the better the separation between groups. Although the accuracy value is 68.60% can still be improved, the use of K-means Clustering provides valuable insight in grouping the attendance employee. Similar attendance patterns are grouped together, and separation between groups can help management to take appropriate action in attendance management and related policy development.

As a future work related to grouping the attendance employee using the K-means Clustering method, here are some things that can be considered:

1. Explore Additional Features.

   Consider incorporating additional relevant features in the analysis, such as employee demographic data, work performance data, or other data that can provide a more comprehensive understanding of attendance patterns. The use of additional features can improve the accuracy and understanding of the resulting grouping.

2. Evaluate Other Metrics

   In addition to the Silhouette Coefficient, can evaluate other clustering metrics, such as the Calinski-Harabasz Index or Davies-Bouldin Index. Using a variety of evaluation metrics can provide a more complete perspective on the quality and success of the resulting grouping.

3. External Validation

Perform external validation by involving domain experts or related management to verify the clustering results and ensure that the groups formed have meaningful interpretations and can be used in better decision making.

4. Solution Implementation and Impact Evaluation

   After getting better grouping results, future work can involve implementing solutions and evaluating their impact on employee attendance management. This evaluation may involve further monitoring and analysis to assess the effectiveness and sustainability of the resulting grouping.

Developing further research in these matters can help improve the understanding and use of clustering methods for grouping of attendance employee, as well as gain deeper and practical insights for decision making and human resource management.

References

[1] M. A. M. Ali, A. Basahr, M. R. Rabbani, and Y. Abdulla, "Transforming Business Decision Making with Internet of Things (IoT) and Machine Learning (ML)," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 2020, pp. 674–679. doi: 10.1109/DASA51403.2020.9317174

[2] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Basel)*, vol. 9, no. 8, p. 1295, 2020. doi: 10.3390/electronics9081295

[3] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796

[4] S. Bandyopadhyay, S. S. Thakur, and J. K. Mandal, "Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society," *Innov. Syst. Softw. Eng.*, vol. 17, no. 1, pp. 45–52, Mar. 2021, doi: 10.1007/s11334-020-00372-5

[5] H. H. Zhao, X. C. Luo, R. Ma, and X. Lu, "An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables," *IEEE Access*, vol. 9, pp. 48405–48412, 2021, doi: 10.1109/ACCESS.2021.3067499

[6] R. Cioffi, M. Travaglioni, G. Piscitelli, A. Petrillo, and F. De Felice, "Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions," *Sustainability (Switzerland)*, vol. 12, no. 2. MDPI, Jan. 01, 2020. doi: 10.3390/su12020492

[7] N. Haefner, J. Wincent, V. Parida, and O. Gassmann, "Artificial intelligence and innovation management: A review, framework, and research agenda☆," *Technol Forecast Soc Change*, vol. 162, p. 120392, 2021, doi: 10.1016/j.techfore.2020.120392

[8] G. N. S. Bandi, T. S. Rao, and S. S. Ali, "Data Analytics Applications for Human Resource Management," in *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402300

[9] N. Palanivel, S. Aswinkumar, and J. Balaji, "Automated Attendance Systems Using Face Recognition by K-Means Algorithms," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1–8. doi: 10.1109/ICSCAN.2019.8878790

[10] M. Raman, N. Kaliappen, and C. L. Suan, "A Study on Machine Learning Classifier Models in Analyzing Discipline of Individuals Based on Various Reasons Absenteeism from Work," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 2020, pp. 360–364. doi: 10.1109/DASA51403.2020.9317017

[11] V. Kakulapati, K. K. Chaitanya, K. V. G. Chaitanya, and P. S. Akshay, "Predictive analytics of HR - A machine learning approach," *Journal of Statistics and Management Systems*, vol. 23, pp. 959–969, 2020, [Online]. doi: 10.1080/09720510.2020.1799497

[12] Q. Xiao, X. Zhong, and C. Zhong, "Application Research of KNN Algorithm Based on Clustering in Big Data Talent Demand Information Classification," *Intern J Pattern Recognit Artif Intell*, vol. 34, no. 06, p. 2050015, 2020, doi: 10.1142/S0218001420500159

[13] S. Liu, "Teaching Management System and Algorithm Implementation Based on Big Data Fuzzy K-Means Clustering," *Math Probl Eng*, vol. 2022, p. 8970414, 2022, doi: 10.1155/2022/8970414

[14] D. Vrontis, M. Christofi, V. Pereira, S. Tarba, A. Makrides, and E. Trichina, "Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review," *The international journal of human resource management*, vol. 33, no. 6, pp. 1237–1266, 2022. doi: 10.1080/09585192.2020.1871398

[15] S. Chowdhury *et al.*, "Unlocking the value of artificial intelligence in human resource management through AI capability framework," *Human Resource Management Review*, vol. 33, no. 1, p. 100899, 2023. doi: 10.1016/j.hrmr.2022.100899

[16] P. Tambe, P. Cappelli, and V. Yakubovich, "Artificial intelligence in human resources management: Challenges and a path forward," *Calif Manage Rev*, vol. 61, no. 4, pp. 15–42, 2019. doi: 10.1177/0008125619867910

[17] H. A. H. A. Almarashda, I. Bin Baba, A. A. Ramli, A. H. Memon, and I. A. Rahman, "Human resource management and technology development in artificial intelligence adoption in the UAE energy sector," *Journal of Applied Engineering Sciences*, vol. 11, no. 2, pp. 69–76, 2021. doi: 10.2478/jaes-2021-0010

[18] L. N. Dyrbye, T. D. Shanafelt, P. O. Johnson, L. A. Johnson, D. Satele, and C. P. West, "A cross-sectional study exploring the relationship between burnout, absenteeism, and job performance among American nurses," *BMC Nurs*, vol. 18, pp. 1–8, 2019. doi: 10.1186/s12912-019-0382-7

[19] K. A. Davis, S. Allen-Milton, and S. Coats-Boynton, "Making the case for results-based accountability as an intervention for chronic absenteeism in schools to improve attendance," *Race Justice*, vol. 9, no. 1, pp. 46–59, 2019. doi: 10.1177/2153368718816

[20] M. C. Barreto Moreira Couto *et al.*, "Prevalence and work-related factors associated with lower back musculoskeletal disorders in female shellfish gatherers in Saubara, Bahia-Brazil," *Int J Environ Res Public Health*, vol. 16, no. 5, p. 857, 2019. doi: 10.3390/ijerph16050857

[21] S. A. Ali Shah, I. Uddin, F. Aziz, S. Ahmad, M. A. Al-Khasawneh, and M. Sharaf, "An enhanced deep neural network for predicting workplace absenteeism," *Complexity*, vol. 2020, no. 1, p. 5843932, 2020. doi: 10.1155/2020/5843932

[22] A. A. Davidescu, S.-A. Apostu, A. Paul, and I. Casuneanu, "Work flexibility, job satisfaction, and job performance among Romanian employees—Implications for sustainable human resource management," *Sustainability*, vol. 12, no. 15, p. 6086, 2020. doi: 10.3390/su12156086

[23] A. D. P. Chiavegatto Filho, "Prediction of absenteeism in public schools teachers with machine learning," *Rev. saúde pública (Online)*, p. 23, 2021. doi: 10.11606/s1518-8787.2021055002677

[24] W. Nengsih and M. M. Zain, "Descriptive Modeling Uses K-Means Clustering for Employee Presence Mapping," *International Journal of Information Engineering & Electronic Business*, vol. 12, no. 4, 2020. doi: 10.5815/ijieeb.2020.04.02

[25] S. Wang *et al.*, "K-means clustering with incomplete data," *IEEE Access*, vol. 7, pp. 69162–69171, 2019. doi: 10.1109/ACCESS.2019.2910287

[26] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J (Basel)*, vol. 2, no. 2, pp. 226–235, 2019. doi: 10.3390/j2020016

[27] Q. Miao, F. Xiao, H. Huang, L. Sun, and R. Wang, "Smart attendance system based on frequency distribution algorithm with passive RFID tags," *Tsinghua Sci Technol*, vol. 25, no. 2, pp. 217–226, 2019. doi: 10.26599/TST.2018.9010141

[28] R. Shang, B. Ara, I. Zada, S. Nazir, Z. Ullah, and S. U. Khan, "Analysis of simple K-mean and parallel K-mean clustering for software products and organizational performance using education sector dataset," *Sci Program*, vol. 2021, no. 1, p. 9988318, 2021. doi: 10.1155/2021/9988318

[29] M. Liu, Z. Liu, F. Chu, R. Liu, F. Zheng, and C. Chu, "Risk-averse assembly line worker assignment and balancing problem with limited temporary workers and moving workers," *Int J Prod Res*, vol. 60, no. 23, pp. 7074–7092, 2022. doi: 10.1080/00207543.2021.2002960

[30] H. Xu, C. Ma, J. Lian, K. Xu, and E. Chaima, "Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China," *J Hydrol (Amst)*, vol. 563, pp. 975–986, 2018. doi: 10.1016/j.jhydrol.2018.06.060

[31] J. G. Saren, H. Talirongan, F. J. B. Talirongan, and C. L. Malicay, "Mining student behavioral concern through referrals using k-means clustering," *Min. Student Behav. Concern Through Referrals Using K-Means Clustering*, vol. 71, no. 1, p. 19, 2021. doi: 10.47119/IJRP100711220211768

[32] F. Bowen, C. Gentle-Genitty, J. Siegler, and M. Jackson, "Revealing underlying factors of absenteeism: A machine learning approach," *Front Psychol*, vol. 13, p. 958748, 2022. doi: 10.3389/fpsyg.2022.958748