Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 2 (2025) 276 - 282 e-ISSN: 2580-0760

Multi-Process Data Mining with Clustering and Support Vector Machine for Corporate Recruitment

Ruri hartika Zain[1*], Randy Permana[2], Sarjon Defit[3]
[1,2,3]Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK Padang
[1]rurihartika_zain@upiyptk.ac.id, [2]randypermana@upiyptk.ac.id, [3]Sarjon_defit@upiyptk.ac.id

Abstract

Having an efficient and accurate recruitment process is very important for a company to attract candidates with professionalism, a high level of loyalty, and motivation. However, the current selection method often faces problems due to the subjectivity of assessing prospective employees and the long process of deciding on the best candidate. Therefore, this research aims to optimize the recruitment process by applying data mining techniques to improve efficiency and accuracy in candidate selection. The method used in this research utilizes a multi-process Data Mining approach, which is a combination of clustering and classification algorithms sequentially. In the initial stage, the K-Means algorithm is applied to cluster candidates based on administrative selection data, such as document completeness and reference support. Next, a classification model was built using a Support Vector Machine (SVM) to categorize the best candidates based on the results of psychological tests, medical tests, and interviews. The experimental results show that the SVM model produces high evaluation scores, with an AUC of 87%, Classification Accuracy (CA) of 90%, F1-score of 89%, Precision of 91%, and Recall of 90%. With these results, it can be concluded that this model is able to improve accuracy in the employee selection process and help companies make more measurable and data-based recruitment decisions.

Keywords: data mining; clustering algorithm; k-means; support vector machine; employee recruitment

How to Cite: R. H. Zain, Randy Permana, and Sarjon Defit, "Multi-Process Data Mining with Clustering and Support Vector Machine for Corporate Recruitment", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 2, pp. 276 - 282, Mar. 2025.
DOI: https://doi.org/10.29207/resti.v9i2.6197

1. Introduction

The process of recruiting new employees is one of the most important parts that needs proper attention. The recruitment process is necessary to select candidates who want to join the company because companies need professional, loyal, hardworking and highly committed employees to develop and improve their performance. A good employee recruitment process can ensure the fulfillment of resource needs, the acquisition of the best quality employees, which will have an impact on the increase of company productivity and the reduction of employee turnover in the company. Companies need to have a good understanding of good human resources (HR) when conducting employee recruitment [1], [2]. In order to achieve this, companies need to pay attention to the principle of sustainability in the recruitment process, where sustainability means that the recruitment process is conducted in a fair, transparent and responsible manner.

Although the company already has standards for ensuring the sustainability of the recruitment process, this process does not always get satisfactory results. The best employees based on the selection process often do not have a good enough impact on the company, so the company needs to repeat the recruitment process within a short period of time from the previous recruitment activity. However, while there has been a lot of research into sustainability, there has been less discussion about sustainable practices in the recruitment process [3].

Computers equipped with AI through hardware and software can also learn from experience gained from processing data to perform specific tasks. The quality and amount of learning produced by a machine or computer are affected by the quality and amount of data processed [4], [5]. Artificial intelligence has become a catalyst for strategic innovation in many industries, including human resource management. Human resource management is still a manual process that takes a lot of time and sometimes creates bias towards employees [6]. In recent years, advances in information technology have led to significant changes in the way

Received: 28-11-2024 | Accepted: 14-03-2025 | Published Online: 25-03-2025

companies conduct recruitment and selection. Today's business world is changing rapidly, and Human Resource Departments (HRDs) are confronted with new realities [7]. One branch of AI is machine learning (ML). ML involves the design and development of algorithms that allow computers to learn behaviour based on empirical data, such as data from sensors and databases [8]. In cases where data is fuzzy, unclear or noisy, recent research has shown that the use of artificial intelligence is more thorough and efficient than traditional rule-based approaches [9]. Development of accurate data mining-based energy load prediction models. The development of accurate data mining-based energy load prediction models usually includes the following steps feature engineering, model training, model evaluation, result visualisation, model interpretation, etc [10]. Data mining is a process that uses machine learning, statistics, mathematics and artificial intelligence to find and extract important information from very large databases [11].

Some research with data mining related to employee recruitment focuses on finding variable insights that determine the best employees, such as Yuanyuan Chen's research [12], which found a strong correlation between educational background, work experience, and salary level. Ling Sun's research [13] applied the concept of clustering to provide better insight into the types of jobs available in the online recruitment market, which can help companies in the recruitment process and job seekers in finding suitable jobs. Meanwhile, a research concept relevant to this study was conducted by Qiqi Miao [14], where this study used two main models for classification prediction, called the C5.0 decision tree model and the Bayesian network model. These models were chosen for their effectiveness in handling the complexity of recruitment data and providing insight into job matching. Utilizing algorithms like MST for clustering recruitment data enhances the efficiency of matching candidates with job openings, reducing the time needed for index updates[15]. Data mining technologies can extract relevant information from complex datasets, allowing for a better understanding of job market demands[16].

In this research, we propose a multi-process data mining concept that aims to get the best candidates to support the success of the company's employee recruitment process. The multi-process method is inspired by the stages of the traditional employee recruitment process applied by the majority of companies in Indonesia, where the cluster method will handle the administrative selection by grouping and reducing the number of candidates, and the classification model will ensure that the employees selected from the previous stage are the best employee candidates.

Clustering algorithms create discrete groups that describe certain patterns or characteristics in the data, which can be used to identify trends, predict behaviour, or group similar entities for further analysis. However, proper selection of the clustering method and accurate interpretation of the clustering results are important steps to ensure that the method is used successfully in various applications [17], [18]. Clustering is an unsupervised learning technique that separates various objects or entities into a limited number of homogeneous groups. Some features are used to distinguish between these objects [19]. The data set is divided into a number of clusters using the K-Means unsupervised clustering technique. Each cluster should be defined so that the data points within it are similar to each other but different from the data points within other clusters [20]. K-means attempts to partition the current data into one or more clusters, like other non-hierarchical clustering methods [21]. This classification is used to classify the results of psychological tests and interviews into a class of accepted or rejected. This is done using a Support Vector Machine (SVM) [22], [23]. SVM is a statistical machine learning algorithm. Support Vector Machines provide classification learning models and algorithms rather than regression models and algorithms [24].

The main purpose of using data mining techniques, the K-means clustering algorithm, for the company's recruitment process, so that it can help the HRD in making decisions at the administrative stage, called by complete or incomplete. The use of the K-means clustering algorithm makes it possible to group applicant data based on letter of application, CV, diploma, ID card (KTP), family card (KK), experience and other completeness. This research also aims to apply machine learning methods to company selection data in HRD to improve the performance of the classification algorithm by eliminating some features that have no correlation with the target label and have a significant impact on the classification results. Using K-means clustering and SVM algorithm provides high accuracy and spends very little time for the whole processing.

2. Research Methods

AI is a very hot topic and its applications are very broad, ranging from small to large. Technically, creating an AI programme, cannot be separated from ML [25], [26]. Machine learning (ML) techniques focus on automatically improving the performance of learning systems through experience, enabling computers to learn automatically without human intervention and to adapt appropriate actions [27]. This research methodology includes several main stages, called: data preparation, application of clustering algorithm, initial selection based on clustering results, classification process using Support Vector Machine (SVM), and the final stage of employee selection based on analysis results. The research flow chart is shown in Figure 1.

This research methodology begins with a data preparation stage, where applicant data is grouped based on the type of process to be applied. This data is divided into two main categories, namely Clustering, which includes administrative data and applicant

abilities, and Support Vector Machine (SVM), which consists of psychological test results, interviews, and medical examinations. After the data is prepared, the clustering process is carried out, which is the grouping of data based on similar characteristics. Clustering aims to form groups (clusters) where data in one cluster has a high level of similarity, while between clusters has significant differences. The result of this process is called the Stage 1 Selection Result. Furthermore, data that has gone through the initial selection stage is analyzed using the Support Vector Machine (SVM). This method is used to find a hyperplane that can optimally separate two classes of data. SVM has the advantage of determining the distance using a support vector so that the computation process becomes faster and more efficient. From the results of the SVM application, a final decision is obtained regarding applicants who are accepted and not accepted as company employees. This methodology is designed to improve efficiency and accuracy in the employee selection process by combining clustering techniques for initial classification and SVM for final decision-making based on psychological, interview, and health factors.

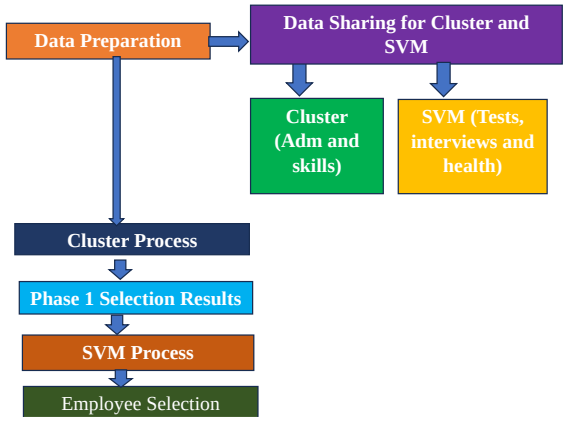

Figure 1. Multi-process data mining

2.1 Dataset Preparation

The first stage is data collection (data preparation) and the process is shown in Figure 2.

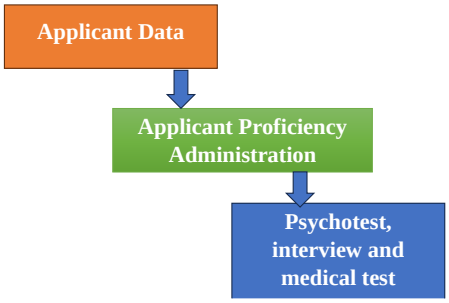

Figure 2. Data preparation

A non-hierarchical data clustering method known as the K-means algorithm divides data into one or more groups or clusters so that groups of data with similar characteristics belong to the same group and groups of data with different characteristics belong to different groups. Data collection, the data used by the author as a dataset is data on the number of applicants who apply to the company and the dataset variables used are administration, skills, psychological tests, interviews and health.

2.2 Preprocessing data

In the preprocessing stage of data analysis, there are two main processes that greatly affect the quality of the analysis results, called by data selection and data transformation. Data selection is the process of selecting data that become variables that will be used in clustering calculations using the K-Means algorithm.

This selection aims to ensure that only relevant, informative, and high-quality data is used in the clustering process. Meanwhile, data transformation is the process of changing the format or structure of data to make it more suitable for the algorithm used. In the context of K-Means, the transformation is done by normalizing and standardizing the dataset used.

2.3 Cluster Process

The clustering stage in data analysis aims to group data based on similar characteristics. The process includes data presentation, determining the number of clusters, and determining the centroid using the Sum of Squared Errors (SSE) calculation.

In the initial stage, data that has gone through the preprocessing process is prepared. The next step is to determine the number of clusters, After the number of clusters is determined, the next process is to determine the centroid, which is the center point that represents each cluster. In K-Means, the initial centroid is chosen randomly and then updated iteratively by calculating the average position of all data in a cluster.

This process continues until the centroid position stabilizes and the SSE value reaches a minimum, indicating that the clustering is optimal.

2.4 Support Vector Machine (SVM) Process

The advantages of this model are speed of processing and better performance compared to similar models. SVM works by maximising the distance between two classes to produce the best hyperplane in classifying classes. The characteristics of SVM are that the first requires a training process in storing the best support vector used in the testing process, the second model and solution are the same, the third SVM is able to separate data linearly or non-linearly, the fourth SVM does not require a dimension reduction process, the fifth SVM process is determined by the amount of data. Formula 1 is used in the calculation of SVM.

$$\text{yi}\ (w_1\ x_1 + w_2 x_2 + b) \geq 1 \quad (1)$$

The value of y is the expected output, while W is the weight of variables x1, x2, and so on. While b is a bias value called linear classifier which plays a role in improving performance.

3. Results and Discussions

The findings of this research lie in the multi-data mining process, a model is formed in the recruitment of potential employees of the company. This research refers to several studies that apply data mining to find variables that support employee selection. [12], [13], [14]. The difference in the point of view of applying data mining in this study is to use the proposed model to decide which employees will be accepted by the company, so that the company's sustainability process will be better.

The proposed model applies two data mining processes, for the first stage selection of data using clustering with the K Means algorithm and the second selection using a Support Vector Machine (SVM). The data processed by K-Means are data derived from administrative data submitted by candidates to the company. The data includes administrative data such as a letter of application, curriculum vitae, latest diploma, transcript, identity card, and letter of good conduct. Meanwhile, driving licences, certificates of employment, internship certificates and skills are used as supporting data for the selection. All input data is given a value by the company's internal team to be further processed in K-Means data mining processing. In this study, we collected about 100 data on prospective employees who submitted applications to the company as shown in Table 1.

Table 1. Dataset

No	Name	Document evaluation	Support evaluation
1	Candidate 1	75	50
2	Candidate 2	60	75
3	Candidate 3	55	50
4	Candidate 4	80	50
5	Candidate 5	40	50
6	Candidate 6	50	50
7	Candidate 7	60	50
8	Candidate 8	70	50
9	Candidate 9	30	50
10	Candidate 10	100	80
…	…	…	…
96	Candidate 96	40	50
97	Candidate 97	85	50
98	Candidate 98	85	80
99	Candidate 99	70	50
100	Candidate 100	100	50
101	Candidate 101	75	50

The clustering process is used as an initial selection of the company's employee candidates. All the scored data are then grouped using the K-means algorithm.

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(xi - ci)^2} \quad (1)$$

Based on Formula 1, it can be mentioned that x consisting of x1, x2,..., and xn is the data vector. C which consists of c1, c2,..., and cn is the centroid vector and meanwhile d(x,c) is the Distance between data x and centroid c. The centroid is randomly selected in the K-means processing based on the centroid case of the employee candidate selection data obtained from Candidate 1 and Candidate 11 as shown in Table 2.

Table 2. Initial centroid determination

Name	Document evaluation	Support evaluation	Centroid
Candidate 1	75	10	C1
Candidate 11	40	10	C2

By using the centroid formula, each data is grouped based on its proximity to a predetermined centroid point. The following is one of the procedures for finding groups of data by the nearest previously determined centroid point.

The calculation continues until the last data and then continues by entering the second centroid value. The clustering process is performed by selecting the data that are closest to the centroid. K-Means produces two groups of data, C1 and C2. Group C1 is a group of data that falls into the pass category, while C2 is a group of data that is declared to have failed the employee selection in the first stage. Table 3 is the C1 data grouped using K-means.

Table 3. Cluster C1 group passed level 1

C1	Passed Stage 1
1	Candidate 1
2	Candidate 2
3	Candidate 4
4	Candidate 8
5	Candidate 10
6	Candidate 12
7	Candidate 13
8	Candidate 15
9	Candidate 18
10	Candidate 20
11	Candidate 21
…	…
48	Candidate 93
49	Candidate 95
50	Candidate 97
51	Candidate 98
52	Candidate 99
53	Candidate 100
54	Candidate 101

The C1 group results in a data set with a total of 54 candidates. All data in C1 is input data used in the next process using SVM. The input data used in the SVM process are assessment data related to psychological, health and also interviews conducted by the company team and the target is accepted or rejected in the selection process. To strengthen this model, the data is divided into data used as training data and test data. The following is a representation of the data set used in the SVM process.

Datasets from number one to eight are used as training data, while datasets number nine to eleven are used as test data to test the classification performance of the SVM. In this research, we applied all of the SVM kernels like linear, sigmoid, polynomial, and RBF (The Radial Basis Function). We used a model with a dataset division of 70:30 and 80:20 from a total of 54 data points, which are candidate data that have entered the stage 1 pass group through the clustering process. The results of the training can be seen in Table 5.

Table 4. SVM Processing Dataset (Selection Level 2)

No	Name	Psychological	Health	Interview	Results
1	Candidate 1	70	65	75	not accepted
2	Candidate 2	100	70	89	accepted
3	Candidate 4	30	100	75	not accepted
4	Candidate 8	70	80	75	not accepted
5	Candidate 10	30	70	69	not accepted
6	Candidate 12	100	90	80	accepted
7	Candidate 13	70	75	70	not accepted
8	Candidate 15	100	100	75	accepted
9	Candidate 18	70	70	77	not accepted
10	Candidate 20	30	80	75	accepted
11	…	..	…	…	…
52	Candidate 98	30	70	90	not accepted
53	Candidate 99	30	90	76	not accepted
54	Candidate 00	100	100	82	accepted

Table 5. Model Performace Result from divison data

Data Divison	AUC	CA	F1	Precision	Recall
70:30	0.89	0.89	0.89	0.89	0.89
80:20	0.91	0.93	0.93	0.93	0.93

Using the linear SVM model, the results of two data-sharing scenarios using cross-validation achieved the result that the 80:20 data division provides better model performance where the AUC (Area Under Curve) value reaches 91% which indicates excellent discriminatory ability in distinguishing two classes, CA (Classified Accuracy) value reaches 93% stating the model works better by using data division 80:20, F1 Score has increased which states that the increasing amount of training data provides a good balance between precision and recall so that classification errors are smaller. The model can capture positive instances.

To strengthen the cluster model and the SVM, tests were carried out using orange data mining software.

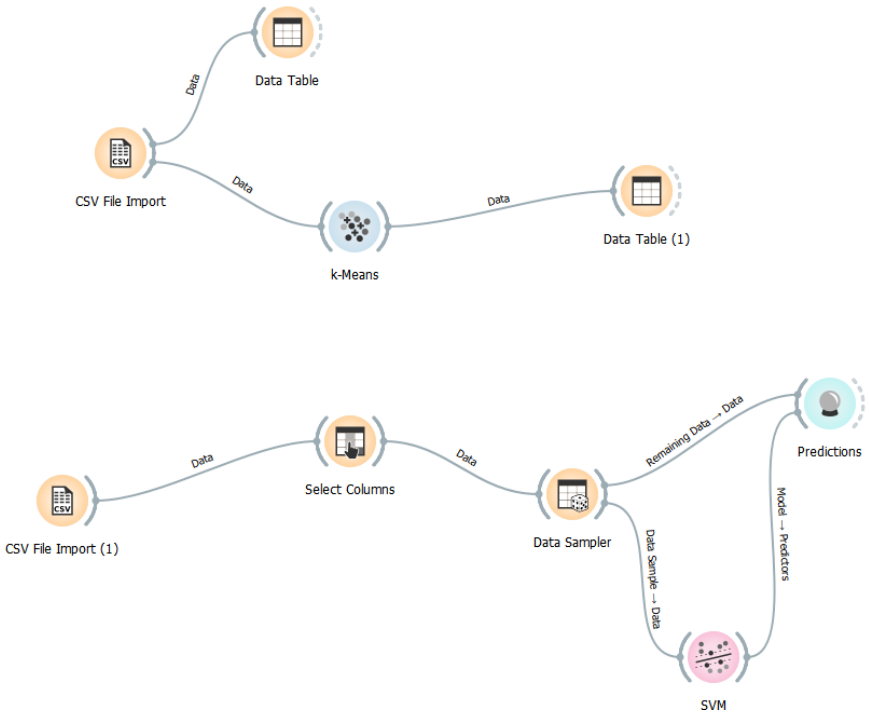

Figure 3. Cluster model and SVM for employee candidate selection

The classification results provided by this model have similar results to the process performed prior to using the software.

Based on the test results using all SVM kernels, where the basic formula of kernel function.

$$K(x, x') = \phi(x) \cdot \phi(x') \quad (2)$$

Formula 2 is a kernel function, denoted as K(x,x'), is a function used to calculate the degree of similarity between two data, called by x and x′. This function allows Support Vector Machines (SVM) to work in a more complex feature space without having to explicitly calculate a transformation to a higher dimension. This data transformation is referred to as $\phi(x)$ and $\phi(x')$, which aims to map the data from its original space to a higher dimensional feature space so that more complex relationships can be more easily separated or analyzed.

The performance results obtained can be seen in Table 6.

Table 6. Kernel-based performance evaluation

SVM Kernel	AUC	CA	F1	Precision	Recall
sigmoid	0.63	0.72	0.70	0.69	0.72
linear	0.91	0.93	0.93	0.93	0.93
polynomial	0.92	0.95	0.95	0.95	0.95
RBF	0.91	0.90	0.90	0.91	0.90

Based on the testing of the kernels of the proposed SVM model, it can be concluded that the best results of the candidate recruitment case are obtained by the SVM Polynomial kernel, where the AUC has the highest value of all kernels. In addition, the accuracy of this kernel reaches 95%, indicating that this kernel is very suitable to be applied to the case of recruitment of prospective employees.

To strengthen the model generated from this research, the SVM model with 80:20 data division and polynomial kernel was compared using other classification models. The results of the comparison can be seen in Table 7.

Table 7. Kernel-based performance evaluation with Cross Validation

Classification Model	AUC	CA	F1	Precision	Recall
SVM Polymonial	0.92	0.95	0.95	0.95	0.95
KNN	0.91	0.86	0.86	0.87	0.86
Logistic Regression	0.90	0.93	0.93	0.93	0.93

Based on the test results, the Polynomial SVM model provides the best results of the three models used in classifying the best employee candidates. SVM Polynomial has the highest AUC, Accuracy, F1 Score, Precision, and Recall metrics. While the classification model using Logistic Regression has lower performance than the SVM model. In the K-NN Classification model, the performance metric value is quite poor, making it less suitable if applied to the case of recruiting prospective employees.

4. Conclusions

The application of machine learning in the employee selection process provides convenience and a new breakthrough in the selection process of employee candidates. The combination of two machine learning processes in this paper is inspired by the multilevel selection process carried out by the majority of companies in Indonesia. The application of machine learning facilitates the file selection process by using K-Means and SVM to strengthen the classification results that have been carried out in the employee selection process in the second stage and providing corrections to the classification results so as to provide the most optimal results in the employee candidate selection process.

References

[1] Z. Zhao, "Research on Marketing Strategies of Enterprise in the Context of the Digital Economy," *Adv. Econ. Manag. Polit. Sci.*, vol. 91, no. 1, pp. 173–179, 2024, doi: 10.54254/2754-1169/91/20241055.

[2] G. Boehncke, "Talents for future – do top talents care about CSR corporate communication in recruiting? An empirical study," *Soc. Responsib. J.*, 2024, doi: 10.1108/SRJ-11-2023-0623.

[3] G. Koman, P. Boršoš, and M. Kubina, "Sustainable Human Resource Management with a Focus on Corporate Employee Recruitment," *Sustain.*, vol. 16, no. 14, 2024, doi: 10.3390/su16146059.

[4] Muhammad Haris Diponegoro, Sri Suning Kusumawardani, and Indriana Hidayah, "Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 131–138, 2021, doi: 10.22146/jnteti.v10i2.1417.

[5] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications," *Mech. Syst. Signal Process.*, vol. 147, 2021, doi: 10.1016/j.ymssp.2020.107077.

[6] L. Ghedabna, R. Ghedabna, Q. Imtiaz, A. Faheem, A. Alkhayyat, and M. S. Hosen, "" Artificial Intelligence in Human Resource Management : Revolutionizing Recruitment , Performance , and Employee Development "," vol. 10, pp. 52–68, 2024.

[7] M. A. Ahmadi, R. Ayuningtyas Fachrunisa, A. Baihaqi, F. Kurniawan, M. Ilham, and T. Abdillah, "Transforming Human Resources Recruitment: The Impact of Artificial Intelligence (AI) on Organizational Attractiveness and Applicant Intent," *Benefit J. Manaj. dan Bisnis*, vol. 9, no. 1, pp. 99–114, 2024.

[8] L. Hoki, V. Augusman, and T. Informatika, "PENERAPAN MACHINE LEARNING UNTUK MENGATEGORIKAN SAMPAH," vol. X, no. 1, pp. 1–5, 2021.

[9] K. Fukami, K. Fukagata, and K. Taira, "Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows," 2021, doi: 10.1017/jfm.2020.948.

[10] C. Zhang, J. Lu, and Y. Zhao, "Generative pre-trained transformers (GPT)-based automated data mining for building energy management: Advantages, limitations and the future," *Energy Built Environ.*, vol. 5, no. 1, pp. 143–169, 2024, doi: 10.1016/j.enbenv.2023.06.005.

[11] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.

[12] Y. Chen and R. Pan, "Research on Data Analysis and Visualization of Recruitment Positions Based on Text Mining," *Adv. Multimed.*, vol. 2022, 2022, doi: 10.1155/2022/9047202.

[13] L. Sun, X. Cao, and D. Su, "Statistical Analysis of Online Recruitment Information Based on Text Mining Technology," in *2022 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, 2022, pp. 431–436.

[14] Q. Miao, "Data mining design for development job recruitment in Wuhan," in *2022 2nd International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)*, 2022, pp. 107–112.

[15] H. Li and N. Hu, "Data Clustering Mining Method of Social Network Talent Recruitment Stream Based on MST Algorithm," in *International Conference on Advanced Hybrid Information Processing*, 2022, pp. 99–111.

[16] M. Shi, H. Zhou, and Z. Lang, "Design and Implementation of Employment Information Data Collection System Based on Data Mining Technology," in *2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)*, 2024, pp. 795–799.

[17] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol.

3, no. 1, pp. 46–56, 2024.
[18] K. Chowdhury, D. Chaudhuri, and A. K. Pal, "An entropy-based initialization method of K-means clustering on the optimal number of clusters," *Neural Comput. Appl.*, vol. 33, no. 12, pp. 6965–6982, 2021, doi: 10.1007/s00521-020-05471-9.
[19] M. Jahangoshai Rezaee, M. Eshkevari, M. Saberi, and O. Hussain, "GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game," *Knowledge-Based Syst.*, vol. 213, p. 106672, 2021, doi: 10.1016/j.knosys.2020.106672.
[20] H. A. Ulvi and M. Ikhsan, "Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia," *Sinkron*, vol. 8, no. 3, pp. 1671–1685, 2024, doi: 10.33395/sinkron.v8i3.13815.
[21] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
[22] U. Rahamathunnisa, M. K. Nallakaruppan, A. Anith, and K. S. S. Kumar, "Vegetable Disease Detection Using K-Means Clustering and Svm," *2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020*, pp. 1308–1311, 2020, doi: 10.1109/ICACCS48705.2020.9074434.
[23] S. Saikin, S. Fadli, and M. Ashari, "Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results," *JISA(Jurnal Inform. dan Sains)*, vol. 4, no. 1, pp. 22–27, 2021, doi: 10.31326/jisa.v4i1.881.
[24] N. Tilapia and U. Arduino, "JITE (Journal of Informatics and Telecommunication Engineering) Automation of Aquaponic Choy Sum and Nile Tilapia Using Arduino," vol. 4, no. January, 2021.
[25] F. Zamachsari and N. Puspitasari, "Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 203–212, 2021, doi: 10.29207/resti.v5i2.2952.
[26] R. M. Arias Velásquez, "Support vector machine and tree models for oil and Kraft degradation in power transformers," *Eng. Fail. Anal.*, vol. 127, no. April, 2021, doi: 10.1016/j.engfailanal.2021.105488.
[27] R. Sistem and F. Bencana, "JURNAL RESTI Algoritma Fungsi Perlatihan pada Machine Learning berbasis ANN," vol. 1, no. 10, pp. 254–264, 2021.

