Available online at http://www.iorajournal.org/index.php/ijgor/index

International Journal of Global Operations Research

e-ISSN: 2722-1016
p-ISSN: 2723-1739

Vol. 6, No. 4, pp. 306-312, 2025

Decision Support System for Letter Follow-Up Using K-Means Clustering, Principal Component Analysis, and Analytical Hierarchy Process

Ilham Radan Alif[1*], Eneng Tita Tosida[2], Adriana Sari Aryani[3]

[1]*Department of Computer Science, Faculty of Mathematics and Natural Sciences, Pakuan University, Bogor, Indonesia*

[2]*Computer Science Departement, Post Graduate School, Universitas Pakuan, Indonesia*

[3]*Department of Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, Indonesia*

[1]*Corresponding author email: ilham.065120070@unpak.ac.id*

Abstract

Several institutions still face challenges in managing incoming letters due to the absence of a structured classification system. The follow-up process is often carried out based on the order of arrival without considering the urgency or importance of the content, leading to document accumulation and difficulties in determining priorities. To address these issues, this study aims to develop a Decision Support System (DSS) that assists in classifying letters and provides strategic recommendations for follow-up prioritization. The K-Means method is used to cluster letter data based on attribute similarities, supported by Principal Component Analysis (PCA) for dimensionality reduction. Furthermore, the Analytical Hierarchy Process (AHP) is applied to generate strategic recommendations for each cluster of letters. The research data were obtained from the institution's budget management letter registry, which includes attributes such as institution name, department, description, program, activity, sub-activity, and sub-detail. The results indicate that the K-Means method is less optimal for clustering complex letter data, with a silhouette score of 0.208 and a Davies–Bouldin Index (DBI) of 1.793. However, the AHP method achieved a consistency ratio (CR) below 0.1, demonstrating the reliability of the generated recommendations. Overall, the developed system effectively provides accurate and efficient recommendations for letter follow-up prioritization, thereby improving decision-making processes within the institution.

Keywords: Decision Support System, Letter, Archive, K-Means, Clustering, Principal Component Analysis, Analytical Hierarchy Process.

1. Introduction

The The XYZ Institution plays an essential role in supporting local government functions, particularly in financial administration. One of its main responsibilities involves validating and issuing budget management letters for institutions requiring operational funds. However, the existing process of handling incoming letters lacks systematic classification, as all letters are processed sequentially without considering their urgency or importance. This unstructured workflow often leads to document accumulation and difficulties in prioritizing follow-up actions, indicating the need for a more efficient and effective management system.

A Decision Support System (DSS) offers a potential solution for determining letter follow-up priorities within the institution. The Analytical Hierarchy Process (AHP) is a well-established multi-criteria decision-making technique (Ramdani et al., 2019) that can be combined with K-Means clustering for data grouping and Principal Component Analysis (PCA) for cluster validation. This integration enables the system to generate strategic recommendations for prioritizing letter clusters that require immediate attention.

Several studies have demonstrated the effectiveness of DSS and clustering methods in improving administrative efficiency. For instance, Purba et al. (2023) utilized K-Means and ELECTRE for library book procurement prioritization, while Ananda et al. (2022) and Nikmah et al. (2022) applied similar techniques for document management optimization. Nasution and Safii (2024) also showed that K-Means effectively categorized official correspondence, improving distribution efficiency and understanding of document patterns.

Building upon these findings, the present study develops a Decision Support System integrating K-Means, PCA, and AHP to enhance the prioritization of letter management processes at the XYZ Institution. The proposed system aims to support a more structured, efficient, and data-driven approach to decision-making in administrative correspondence handling.

2. Literature Review

2.1. Theoritical Foundation

A Decision Support System (DSS) is designed to assist the decision-making process by providing relevant information, guidance, and predictions to help users make better decisions (Astari et al., 2021). In data grouping, the K-Means method is commonly used because it can classify data into several clusters based on similarity characteristics, where objects with the shortest distance are grouped within the same cluster (Aprilia et al., 2022).

To enhance clustering performance, the Principal Component Analysis (PCA) method can be applied to reduce data dimensionality without losing essential information, resulting in more optimal clustering outcomes (Dewi and Pakereng, 2023). In addition, the Analytical Hierarchy Process (AHP) serves as a weighting approach that supports decision-making systematically through pairwise comparisons among criteria (Hendri et al., 2023).

2.2. Previous Studies

Several studies have previously implemented decision support methods combining clustering and analytical techniques. Purba et al. (2023) developed a decision support system for determining book procurement priorities in libraries using the K-Means and ELECTRE methods. The K-Means algorithm was employed to cluster frequently borrowed books into three groups, which were then analyzed using ELECTRE to determine procurement priorities. This approach effectively identified book clusters with high borrowing frequency to support better procurement decisions. Dewi and Pakereng (2023) applied Principal Component Analysis (PCA) to the K-Means algorithm to cluster educational attainment levels among residents of Semarang Regency. Using PCA for dimensionality reduction improved clustering efficiency and interpretability, resulting in two principal components with a cumulative variance of 70% and four distinct educational clusters. The study demonstrated that PCA can enhance the performance of K-Means in handling high-dimensional data.

Rosai et al. (2024) implemented a web-based Decision Support System (DSS) using the Analytical Hierarchy Process (AHP) to determine the issuance of warning letters for students at Institut Shanti Bhuana. The AHP method was used to calculate the weight of assessment criteria, improving objectivity and efficiency in decision-making related to disciplinary actions. In contrast to the aforementioned studies, this research integrates K-Means, PCA, and AHP within a unified Decision Support System for follow-up prioritization of correspondence documents. The combination of clustering and multi-criteria decision-making techniques aims to improve prioritization accuracy and ensure optimal handling of institutional correspondence data.

3. Materials and Methods

3.1. Materials

The materials required in this study include:

1) Books, journals, and theses used as reference materials.
2) The budget management letter registration data from August to September 2023, managed by Institution XYZ.
3) The 2019 undergraduate thesis guideline book from the Computer Science Study Program, Faculty of Mathematics and Natural Sciences, Pakuan University.
4) Tools and software used in the study, including Python programming language, Streamlit, and relevant Python libraries such as scikit-learn, pandas, and numpy.

3.2. Methods

SDLC, or System Development Life Cycle, is a series of stages in creating and modifying a system, and it has become a foundation for various system development efforts (Yudi Sobari et al., 2023). In this study, additional stages such as initial analysis, data collection, and data processing were included, which are subsequently supported by the SDLC. The workflow of this study is illustrated in Figure 1.

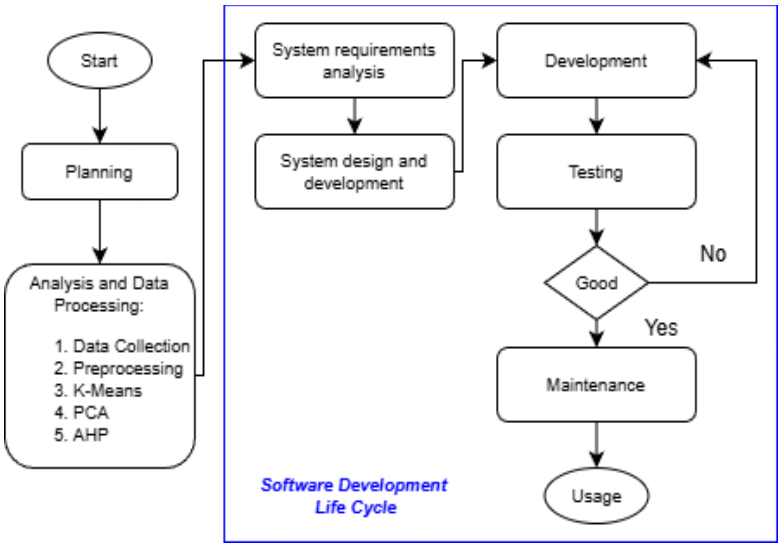

Figure 1: SDLC Method

This study follows a structured research process based on the System Development Life Cycle (SDLC), which includes planning, data analysis, system design, development, testing, and maintenance.

(a) Planning: Identifying research objectives, requirements, and potential solutions for clustering budget letters using K-Means supported by PCA, and decision-making with AHP.

(b) Data Analysis: Data was collected from Institution XYZ in CSV format and supplemented by interviews and literature review. Preprocessing was performed using text mining techniques, including tokenization, stop word removal, and lowercasing. K-Means clustering was then applied to group letters based on similarity, and PCA was used to reduce dimensionality and improve interpretability. Finally, AHP was used to determine the priority of clusters for follow-up actions based on specific criteria.

(c) System Development and Design: The system was designed using activity diagrams, use case diagrams, class diagrams, and interface prototypes. Python and Streamlit were used to implement K-Means, PCA, and AHP within the system.

(d) Testing and Maintenance: Functional and structural testing were conducted to ensure system performance. Clustering results were validated using Silhouette Score and Davies-Bouldin Index (DBI), while AHP evaluations were assessed using Consistency Index (CI) and Consistency Ratio (CR). Maintenance ensures system stability and allows for future improvements or feature additions.

This approach ensures that the decision support system is efficient, accurate, and capable of providing structured recommendations for letter follow-up priorities.

3.2.1. Equation

3.2.1.1. K-Means

The following are the steps of the clustering process using the K-Means method according to (Oktavia et al., 2020):

1) Determine the number of clusters (k) in the dataset.
2) Determine the centroid values. The initial centroid values are chosen randomly or can be set using the maximum value for high clusters and the minimum value for low clusters.
3) For each record, calculate the distance to the nearest centroid. The distance used is the Euclidean distance, calculated using the following formula:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Explanation:
De = Euclidean distance
i = Number of objects
(x, y) = Coordinates of the object
(s, t) = Coordinates of the centroid

4) Group the objects based on the nearest centroid distance to create a new centroid. The new centroid is calculated by summing the values according to the distances from the previous iteration and then dividing by the total number of objects in each cluster, using the following formula:

$$C(x,y) = \frac{\Sigma xy}{n} \quad (2)$$

5) Repeat steps 2, 3, and 4 and iterate until the centroids reach their optimal values.

3.2.1.2. AHP

The following is the AHP calculation to obtain consistent scale values (Aurachman, 2019):

1) Normalize the data in the pairwise matrix for each criterion by dividing each element in column i and column j by the sum of column i. Alternatively, it can be calculated using the following formula:

$$a_{ij} = \frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}} \tag{3}$$

2) Calculate the row averages (weights) for each criterion in the pairwise matrix.
3) Compute the maximum eigenvalue using the following formula:

$$\lambda\text{maks} = (A1 \times Y1 + B1 \times Y2 + C1 \times Y3 \ldots . n) \tag{4}$$

Explanation:
A = Sum of each column (before normalization)
Y = Row average (weight) for each criterion

4) Calculate the Consistency Index (CI) to determine the consistency of the responses, which affects the validity of the results. The formula is:

$$CI = \frac{\lambda maks - n}{n-1} \tag{5}$$

Explanation:
CI = Consistency Index
$\lambda maks$ = Maximum eigenvalue
n = Order of the matrix

5) To check if a *CI* is valid, compute the Consistency Ratio (*CR*). The matrix is considered consistent if $CR \leq 0.1$. The formula is:

$$CR = \frac{CI}{RI} \tag{6}$$

Explanation:
CR = Consistency Ratio
CI = Consistency Index
RI = Random Index

The table for random index values can be found in Saaty (2013), see Table 1.

Table 1: Random Index Value

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

6) Prioritization of alternatives can be calculated by computing the eigenvalue for each criterion and its alternatives using the formula:

$$\Sigma\ eigen\ value = (A1 \times Y1) + (B1 \times Y2) + \ldots n \tag{7}$$

Explanation:
A = Criterion weight
Y = Alternative weight under the criterion

4. Results and Discussion

4.1. K-Means and PCA Results

The clustering process in this system produced 4 types of clusters from a total of 325 data points. The number of data points and the analysis results for each cluster can be seen in Table 2.

Table 2: Clustering Result Analysis

Cluster	Amount of Data	Explanation
Cluster 1	85	Cluster 1 is dominated by the Regional Government Support Programs at the district/city level, with main activities including planning, evaluating the performance of regional apparatus, and providing office service support. The letters in this cluster have high urgency, as they focus on administrative activities that support government functions.
Cluster 2	96	Cluster 2 is dominated by the Wastewater System Development Management Program, with main activities including infrastructure development and water resource management. The letters in this cluster have medium urgency, as they relate to the development of public facilities and community services.
Cluster 3	77	Cluster 3 is dominated by the Regional Government Support Program, with main activities including infrastructure rehabilitation and maintenance, as well as the provision of communication services. The letters in this cluster have low urgency, focusing on the maintenance of facilities that support public services.
Cluster 4	67	Cluster 4 is dominated by the Regional Government Support Program, with main activities including support services, honorarium payments, and transportation services. The letters in this cluster have medium urgency, as most focus on operational support and routine administrative tasks.

The four clusters exhibit distinct characteristics based on program, activities, and letter urgency. A scatter plot of the two main PCA components visualizes these differences. Cluster 1 (top-left) contains high-urgency administrative support letters for regional government programs. Cluster 2 (bottom-center) relates to wastewater system management with medium urgency for infrastructure development. Cluster 3 (top-right) also involves regional government support programs, focusing on infrastructure rehabilitation with low urgency. Cluster 4 (center) has medium urgency, emphasizing operational support and routine administrative tasks. This visualization is shown in Figure .

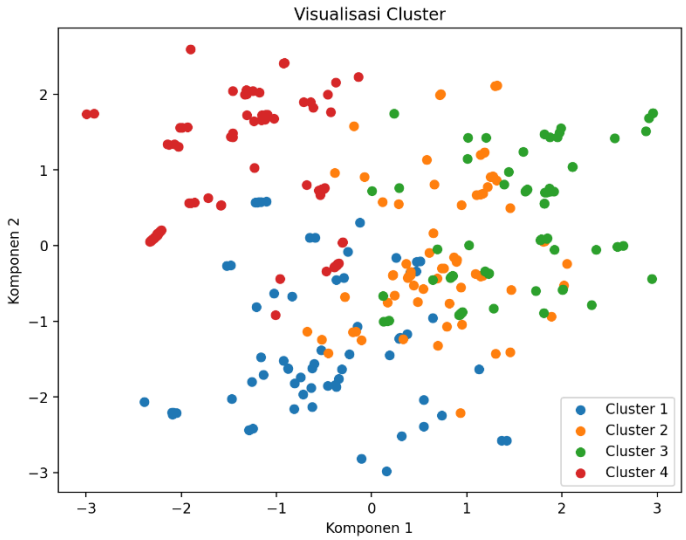

Figure 2: Scatter Plot Visualization Using PCA

Lower values on component 1 indicate higher letter urgency, while higher values on component 2 signify clusters dominated by regional government support program letters. Overall, the scatter plot illustrates the distribution of letters by activity, program, and urgency, aiding the analysis of letter patterns at XYZ Institution.

4.2. AHP Results

The AHP calculations in the system provide a list of recommended strategies for letter clusters that require immediate follow-up, as shown in Figure 3.

	Alternatif	Skor Akhir	Ranking
0	Cluster 1	0.4431	1
1	Cluster 2	0.2946	2
2	Cluster 3	0.1495	3
3	Cluster 4	0.1128	4

Prioritas Alternatif berdasarkan AHP adalah:

Figure 3: AHP Results

4.3. Discussion

This study developed a decision support system to predict letter data clusters and determine their follow-up priorities. The system successfully clustered 325 letter records that had undergone preprocessing and label encoding, producing encoder, scaler, and K-Means models for clustering and predicting new data. Additionally, the system provides cluster priority recommendations using the AHP method.

Research in this journal shows how the integration of analytical methods such as K-Means++ clustering, Principal Component Analysis (PCA), and decision-making approaches can be used as a basis for developing Decision Support Systems (DSS), including for letter follow-up cases. Through a complex data collection and processing process, this research builds a Business Intelligence model capable of presenting information in real-time to support strategic decision-making processes. The use of K-Means++ is an important component in grouping diverse data, as was done in grouping 200 vocational schools based on 36 attributes, thereby producing groups with similar characteristics that can be used for prioritizing handling. In the context of letter follow-up DSS, this approach can be applied to group letters based on urgency level, information category, or specific content patterns, thereby simplifying the sorting and automatic handling process (Tosida et al., 2020)

The K-Means clustering results show that each cluster has distinct characteristics, although some similarities and overlaps exist among certain data points. The PCA scatter plot visualization (Figure 2) illustrates that the four clusters can be clearly distinguished, despite some overlapping data. This indicates that K-Means may not be fully suitable for complex letter datasets. Furthermore, the system cannot yet perform automatic cluster analysis, requiring manual analysis with support from NLP or third-party AI models. According to the AHP results, clusters with high urgency and those related to programs supporting local government affairs are prioritized for follow-up.

Although Institution XYZ already has procedures for managing letter follow-ups, this system is designed to improve efficiency, prevent archival backlogs, and enhance overall letter management. The criteria and alternatives in the AHP process are tailored to Institution XYZ's needs and can be adjusted over time to maintain optimal effectiveness.

5. Conclusion

This study developed a Decision Support System (DSS) for determining follow-up priorities of official letters using K-Means Clustering, Principal Component Analysis (PCA), and the Analytical Hierarchy Process (AHP). The system processed 325 records of budget management letters from an institutional dataset and aimed to assist in classifying and prioritizing letters based on their urgency and content.

The clustering results showed that although K-Means could identify general groupings of letter characteristics, it was less effective for handling complex textual data. Each cluster represented different types of programs and urgency levels, indicating that additional semantic or NLP-based approaches could improve classification accuracy. The AHP method successfully generated consistent and reliable recommendations, with a consistency ratio (CR) below 0.1. These results indicate that AHP can effectively determine the priority of follow-up actions, especially for letters related to critical governmental support programs.

Overall, the integration of K-Means and AHP within the proposed DSS contributes to a more systematic and efficient decision-making process for letter management, ensuring that urgent and high-impact correspondence is handled promptly and appropriately.

Acknowledgments

The author would like to express sincere gratitude to Dr. Eneng Tita Tosida, S.TP., M.Si., M.Kom. and Adriana Sari Aryani, S.Kom., M.Cs, who provided guidance, advice, and support throughout the research process. The author also thanks XYZ Institution for providing access to data and facilities that made this research possible.

References

Ananda, W., Santi, I. H., & Kirom, S. (2022). Implementation of the K-Means Clustering Algorithm in Grouping SKCK (Police Clearance Certificate) Archives. *Journal of Informatics Engineering Students*, *6*(2), 861–867.

Aprilia, R., Afsari, K., Rahma, R., Nasution, N., Ouri, S., & Putri, D. (2022). Cluster Analysis with the K-Means Cluster Method on Letter Data Types at BPPRD North Sumatra. *Journal of Community Service*, *6*(2).

Astari, R. Y., Ginting, B. S., & Sihombing, A. (2021). Decision Support System for Determining Road Improvement Priority Using the Analytical Hierarchy Process (AHP) Method at the Public Works and Spatial Planning Office of Langkat Regency. *Kaputama Information System Journal (JSIK)*, *5*(1), 52–62.

Aurachman, R. (2019). Data Acquisition Process in AHP (Analytical Hierarchy Process) Using the Closed Loop Control System Principle. *JISI: Journal of Industrial System Integration*, *6*. https://doi.org/10.24853/jisi.6.1.55-64

Dewi, S., & Pakereng, M. A. I. (2023). Implementation of Principal Component Analysis on K-Means for Clustering the Education Level of Semarang Regency Residents. *JIPI (Scientific Journal of Informatics Research and Learning)*, *8*(4), 1186–1195. https://doi.org/10.29100/jipi.v8i4.4101

Hendri, R., Hartanto, M. B., & Agustin, A. (2023). Design and Build of a Web-Based Decision Support System for Employee Data Validation at Polda using the AHP Method. *Journal of Technology and Informatics (JEDA)*, *4*(1), 1–9.

Nasution, M. A., & Safii, M. (2024). K-Means Algorithm in Clustering Outgoing Mail at the Pematang Siantar City Religious Affairs Office. *Jayakarta Informatics Management Journal*, *4*(1), 61–71. https://doi.org/10.52362/jmijayakarta.v4i1.1304

Nikmah, F., Pribadi, D. J., Sukma, E. A., Suwarni, E., & Azmi, I. (2022). Decision Support System For Handling Incoming And Outgoing Mail: To Facilitate Archives Retrieval. *International Journal of Academic Research and Reflection*, *10*(3), 53–61.

Oktavia, R., Hardinata, J. T., & Irawan. (2020). Application of the K-means Algorithm Method in Grouping Life Expectancy at Birth by Province. *1*(4), 154–161.

Purba, A. T., Sugara, H., Simarmata, H. M. P., Saragih, D. Y., & Damanik, E. (2023). Decision Support System for Determining Library Book Procurement Priority Using K-Means and Electre Methods. *TEKINKOM*, *6*(1), 196–203.

Ramdani, H. M., Santoso, E., & Rahayudi, B. (2019). Recommendation System for Selecting Incoming Mail Priority Using the AHP-SAW Method (Case Study: DJBC KANWIL JATIM I). *Journal of Information Technology and Computer Science Development*, *3*(4), 3341–3349.

Rosai, F. R., Petrus, S., Manggara, A. D., & Noviyanti, P. (2024). Decision Support System (SPK) for Determining Warning Letter (SP) Issuance Using the AHP Method Case Study at Shanti Bhuana Institute. *Informatics Journal*, *20*(2).

Saaty, T. L. (2013). *Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process*. RWS Publication.

Tosida, E. T., Andria, F., Wahyudin, I., Widianto, R., Ganda, M., & Lathif, R. R. (2019). A hybrid data mining model for Indonesian telematics SMEs empowerment. *IOP Conference Series: Materials Science and Engineering*, *567*(1). https://doi.org/10.1088/1757-899X/567/1/012001

Tosida, E. T., Maryana, S., Thaheer, H., & Hardiani. (2017). Implementation of Self Organizing Map (SOM) as decision support: Indonesian telematics services MSMEs empowerment. *IOP Conference Series: Materials Science and Engineering*, *166*(1). https://doi.org/10.1088/1757-899X/166/1/012017

Tosida, E. T., Wahyudin, I., Andria, F., Djatna, T., Ningsih, W. K., & Lestari, D. D. (2020). Business Intelligence of Indonesian Telematics Human Resource: Optimization of Customer and Internal Balanced Scorecards. *Journal of Southwest Jiaotong University*, *55*(2). https://doi.org/10.35741/issn.0258-2724.55.2.7

Yudi Sobari, M., Purwantoro, & Susilo Yuda Irawan, A. (2023). Decision Support System for Tourism Recommendation in Karawang Regency Using the Profile Matching Method. *Journal of Informatics Engineering Students*, *7*(4).