NUANSA INFORMATIKA ***p-ISSN :1858-3911* , *e-ISSN : 2614-5405***
Volume 20 Nomor 1, Januari 2026 https://journal.fkom.uniku.ac.id/ilkom

Interactive Dashboard-Based Mail Classification Using Deep Learning Models

Ruth Diana Purnamasari*[1], Nisa Hanum Harani[2]
[1,2]Universitas Logistik dan Bisnis Internasional, Indonesia
E-mail: *[1]ruthdianaps@gmail.com, [2]nisa@ulbi.ac.id

Abstract

The management of incoming mail archives at a large national logistics company in Indonesia generates a large volume of unstructured textual data, making manual classification inefficient and error-prone. This study evaluates the performance of deep learning models for administrative mail archives classification using data collected between 2023 and 2025. Three models are examined, namely Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Convolutional Neural Network (CNN). Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics. Experimental results indicate that CNN achieves the highest accuracy of 85.82%, outperforming LSTM and Bi-LSTM models. This superior performance is attributed to CNN's ability to capture local textual patterns through convolution operations, which are well-suited to the structured and repetitive language characteristics of official correspondence. To support practical interpretation, an interactive dashboard is implemented as a visualization tool for model evaluation results, classification outcomes, and clustering analysis. These findings demonstrate that deep learning-based approaches integrated with visual analytics can significantly improve the efficiency and accuracy of unstructured mail archive management.

Keywords*—mail archive classification, deep learning, CNN, model evaluation, interactive dashboard, clustering*

Submission: January 09, 2026 Accepted: January 18, 2026 Published: January 20, 2026
DOI Article: ***https://doi.org/10.25134/ilkom.v20i1.546***

1. INTRODUCTION

Mail archive management is an essential component in supporting the smooth operation of an organization, particularly in institutions that handle a high volume of incoming mail. At a large national logistics company in Indonesia, incoming mail processed at branch offices includes various types and subject matters originating from government institutions as well as external parties. All incoming mail is recorded in a mail log as archival documentation, after which it is distributed and followed up through established administrative mechanisms[1]. Based on internal archival records obtained from the incoming mail agenda of the company during the period 2023-2025, a total of 3,453 mail entries were recorded before data cleaning and validation. The data are in the form of unstructured text, characterized by information such as the subject of the letter, the type of incoming mail, and the sender's name. The high volume and variability of this raw data make the processes of matching, grouping, and classifying mail archives complex when performed manually. Similar conditions are also found in various types of administrative data in other institutions. The use of conventional methods in mail archive management may lead to delays, classification inconsistencies, and difficulties in the retrieval of archived documents [2].

Along with the advancement of information technology, the application of text mining and artificial intelligence has begun to be integrated into digital archiving systems to improve the effectiveness and accuracy of mail management. Several previous studies have shown that document classification algorithms are able to automate the process of mail grouping and increase productivity compared to manual

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

methods[3],[4]. However, many existing studies emphasize automation results without providing clear, interpretable, and user-oriented model evaluation mechanisms that can be easily understood by non-technical users, such as archive officers or administrative staff in institutional environments [3].

In the field of text document classification, Deep Learning approaches have been widely applied due to their ability to better understand linguistic context and the relationships between words compared to traditional methods. Models such as *Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM),* and *Convolutional Neural Network (CNN)* have been proven effective in various document and email classification tasks, particularly for texts with diverse and complex sentence structures[5],[6],[7]. However, most previous studies still focus primarily on reporting performance metrics, such as accuracy and F1-score, without sufficient analytical discussion or integration of these results into visual systems that can be directly utilized in institutional operations [8],[9],[10]. In addition to the classification process, analysis of patterns and structures within mail archive data is also an important aspect, considering the continuously increasing number of documents. Clustering methods can be used to group mail based on content similarity, thereby assisting in understanding topic distribution, characteristics, and the most frequently occurring types of mail[11],[12]. Nevertheless, clustering results are often presented only in numerical or tabular form, making them difficult to interpret without appropriate visual support[13],[14].

To ensure that the data preprocessing procedure is conducted in a systematic and structured manner, this study adopts the CRISP-DM (*Cross-Industry Standard Process for Data Mining)* methodology. This methodology provides a framework consisting of six main stages, namely *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,* and *Deployment* [15],[16]. The use of a standardized methodology in the data classification process is a fundamental principle for generating valuable information patterns for institutions[17].

Nevertheless, previous studies remain limited to the development of mail archive classification systems without comprehensively integrating model performance evaluation, analytical discussion, and clustering analysis within an interactive dashboard. Based on these conditions, a system is required that is not only capable of performing automatic mail classification but also provides informative, interpretable, and user-oriented evaluation results. Several studies on digital archiving systems indicate that the utilization of dashboards can assist users in understanding data conditions and system effectiveness more optimally[3],[8].

Therefore, this study is directed toward the implementation of mail document archive classification using deep learning methods, complemented by an interactive evaluation dashboard and clustering analysis. This research does not aim to develop new algorithms, but rather to apply and critically evaluate multiple deep learning models in a case study using data from a large national logistics company in Indonesia. The main contribution of this study lies in the comparative evaluation of multiple deep learning models combined with visual analytics to enhance interpretability and practical usability in administrative mail classification. Through this approach, it is expected that mail document archive management can be carried out effectively, transparently, and systematically to support the utilization of information technology within institutional environments.

2. METHODOLOGY

This study is designed to provide a systematic overview of stages involved in the data processing, modeling, and evaluation processes. The methodology is structured to ensure that each process is conducted in a measurable, objective, and scientifically accountable manner. In general, study includes stages of mail

p-ISSN :1858-3911 , e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

archive data collection, text data preprocessing, application of deep learning models for document classification, model performance evaluation using quantitative metrics, as well as result evaluation and clustering analysis presented through an interactive dashboard.

2.1. Research Approach and Type

This study adopts an applied *research* approach using computational experimental methods. This approach is selected because the research focuses on the application of deep learning models for classifying mail archive text documents and evaluating model performance based on measurable quantitative metrics. The evaluation results are then presented in the form of an interactive dashboard to facilitate analysis and interpretation.

The type of research employed is quantitative research, as all research outcomes are measured numerically using model performance metrics, namely *accuracy, precision, recall*, and *F1-score.* Through this approach, the effectiveness of each deep learning model can be measured objectively and compared systematically. In addition, this study also utilizes clustering analysis as a supporting method to understand patterns and structures within the mail archive data used.

2.2. Data Sources and Types

The source type of data used in this study include:

2.2.1 Data Source

The data used in this study originate from the incoming mail agenda of a large national logistics company in Indonesia for the period 2023 to 2025. The dataset constitutes secondary data obtained from the internal administrative recording system and is used as part of the organization's operational activities.

2.2.2 Data Type and Research Variables

The type of data used in this study consists of unstructured text, derived from descriptive columns in the mail archive. Based on their roles in the modeling process, the research variables are categorized as follows:

1. **Independent Variables**
   The independent variables consist of the *Subject, From/To,* and *Mail Type* columns. These three attributes are combined to form a richer textual representation and are used as input for the deep learning models.
2. **Dependent Variable**
   The dependent variable in this study is the Mail Type Category, which serves as the classification target for the model, such as *Surat Dinas*, *SPK*, *PKS*, and *KD*.

2.3. Dataset Description

The dataset used in this study consists of 3,453 raw entries of incoming mail archive data obtained from the internal administrative system of a large national logistics company in Indonesia during the period 2023-2025. Data selection and validation were conducted by removing duplicate records, incomplete entries, and irrelevant data that did not meet the research criteria. As a result of this process, 2,817 clean and validated data entries were retained and deemed suitable for use in the modeling stage.

The validated dataset was subsequently divided into training and testing sets using an 80:20 ratio, where 80% of the data were used for model training and 20% were reserved for performance evaluation. This data splitting strategy was applied to maintain a balance and objective evaluation of the proposed models.

The distribution of incoming mail archive data across categories indicates an imbalanced class condition. Several categories, such as *Surat Dinas*, *SPK* and *PKS*, dominate the dataset, while other categories contain fewer instances. The class distribution of the validated dataset is illustrated in figure 4, providing an overview of the proportion of each mail category used in this study.

2.4. Methodology Flowchart

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

The research methodology is designed in a structured manner to ensure that each stage is carried out systematically and can be scientifically justified. Conceptually, the research flow adapts the CRISP-DM [15].

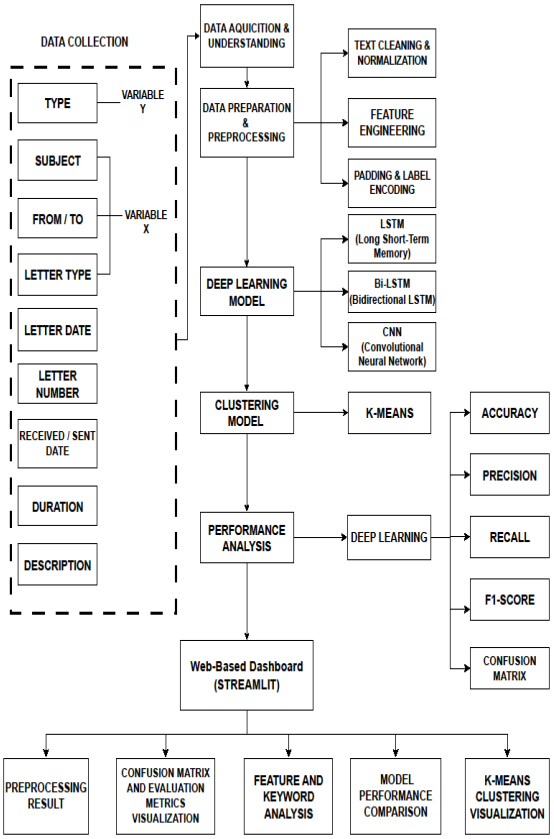


Figure 1 Research Stage

The methodology flow diagram is presented in Figure 1, illustrating the main stages of the study from data collection to visualization.

2.2.3 Description of Research Stages

Based on the flow diagram shown in Figure 1, the research stages include:

1. **Data Collection**
   Incoming mail archive data are collated from multiple internal files of a large national logistics company in Indonesia for the period 2023-2025 and combined into a unified dataset.
2. **Data Selection and Validation**
   The data are examined to remove empty, duplicate, or irrelevant entries that do not align with the research objectives.
3. **Text Data Preprocessing**
   This stage includes text cleaning, normalization text (*lowercasing*), text attribute merging, tokenization, padding, and category label encoding. For illustration purposes, an example of preprocessing transformation includes converting the original text "Surat Permohonan kerja Sama PT ABC" into the cleaned and normalized from "surat permohonan kerja sama pt abc".
4. **Dataset Splitting**
   The dataset is divided into training and testing sets using an 80:20 ratio to maintain a balance between training and evaluation processes.
5. **Deep Learning Model Training**
   L*STM, Bi-LSTM,* and *CNN* models are trained using the training dataset to learn textual patterns within the mail archive data. All models were trained using the same experimental settings to ensure a fair comparison. The training process was conducted with a batch size of 32 and a maximum of 50 epochs. Early Stopping was applied by monitoring validation loss to prevent overfitting during training.
6. **Model Evaluation**
   Model evaluation is conducted using quantitative performance metrics and a confusion matrix to analyze prediction errors.
7. **Result Visualization through Dashboard**
   The results of the evaluation and analysis are displayed through a web-based interactive dashboard using *Streamlit.*
8. **Clustering Analysis**
   The *K-Means* algorithm is used as an exploratory method to analyze patterns and distribution in letter archive data. The clustering results are not used as a classification mechanism, but serve as complementary analysis to provide additional insight into the structure

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

and thematic distribution of the classified mail archive data.

9. **Interpretation of Result and Discussion**
   All results were analyzed to draw research conclusions.

2.5. Data Preprocessing

The preprocessing stage is conducted to ensure that text data are ready to be processed by deep learning models. This process consists of several steps, including text cleaning using regular expression functions (*re.sub*), text normalization by converting all characters to lowercase, merging textual attributes, tokenization and padding processes, and label encoding to convert categorical labels into numerical representations.

2.6. *Deep Learning Models Used*

This study applies three algorithms in the text document classification process, namely *Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM),* and *Convolutional Neural Network (CNN).*

2.6.1. Long Short-Term Memory (LSTM)

LSTM utilizes a memory gate mechanism to retain important information within text sequence data. The basic equations of *LSTM* are formulated as follows:

Forget Gate (1)

$$f_t = \sigma\ (w_f \cdot \left[h_{\{-1\}}, x_t\right] + b_f$$

Input Gate (2)

$$i_t = \sigma(w_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i$$

$$\tilde{C}_t = tanh\left(w_c \cdot \left[h_{\{t-1\}}, x_t\right] + b_c\right)$$

Update Cell State (3)

$$C_t = f_t \cdot C_{\{t-1\}} + i_t \cdot \tilde{C}_t$$

Output Gate (4)

$$o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right)$$
$$h_t = o_t \cdot tanh\left(C_t\right)$$

Table 1 Algorithm LSTM Classifications

Algorithm 1 : *Long Short-Term Memory (LSTM)* Classifications

Output : hidden state h_t

1 *Forget Gate :* Computes previous information that should be retained or discarded
$$f_t = \sigma\ (w_f \cdot \left[h_{\{-1\}}, x_t\right] + b_f$$
2 *Input Gate :* Determines new information to be stored
$$i_t = \sigma(w_i \cdot \left[h_{\{t-1\}}, x_t\right] + b_i$$
3 *Candidate Cell State :* Forms candidate values for new memory
$$\tilde{C}_t = tanh\left(w_c \cdot \left[h_{\{t-1\}}, x_t\right] + b_c\right)$$
4 *Update Cell State :* Updates the cell memory
$$C_t = f_t \cdot C_{\{t-1\}} + i_t \cdot \tilde{C}_t$$
5 *Output Gate :* Determines the output of the cell
$$o_t = \sigma\left(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o\right)$$
6 *Hidden State Output :* Produces the final hidden state
$$h_t = o_t \cdot tanh\left(C_t\right)$$

2.6.2. Bidirectional LSTM (Bi-LSTM)

Bi-LSTM processes data in two directions, forward and backward, enabling the model to capture broader contextual information within text data. The model is formulated as follows :

$$h_t = [\overrightarrow{h_t} \oplus h_t^{\leftarrow}] \quad (1)$$

Where $\overrightarrow{h_t}$ represents the forward *LSTM* and $h_t^{\leftarrow}$ represents the backward *LSTM*.

Table 2 Algorithm Bi-LSTM Classifications

Algorithm 2 : *Bidirectional LSTM Memory (Bi-LSTM)* Classifications

Output : bidirectional context representation h_t

1 *Forward LSTM :* Processes the sequence from the beginning to the end
$$\overrightarrow{h_t} = LSTM_{\{forward\}}(x_t)$$
2 *Backward LSTM :* Processes the sequence from the end to the beginning
$$h_t^{\leftarrow} = LSTM_{\{backward\}}(x_t)$$
3 *Concatenation output :* Combines bidirectional representations
$$h_t = [\overrightarrow{h_t} \oplus h_t^{\leftarrow}]$$

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

2.6.3. Convolutional Neural Network (CNN)

CNN is used to extract local features from text through convolution operations and the ReLU activation function, as shown in the following equation :

$$c_i = f\left(W \cdot x_{\{i:i+h-1\}} + b\right) \quad (1)$$

Where $x_{\{i:i+h-1\}}$ denotes the word vector, W represents the convolution filter, and f is the activation function.

Table 3 Algorithm CNN Classifications

Algorithm 3 : *Convolutional Neural Network (CNN)* Classifications

Output : predicted class C_{map}

1 W *:* convolution filter weights
2 $x_{\{i:i+h-1\}}$ *:* input vector from position – i to $i+h-1$
3 b : bias
4 $f(\cdot)$ *:* activation function (*ReLU)*
5 h *:* filter size (window size)

2.7. Model Training and Evaluation Process

The model training process was carried out by applying *Early Stopping* to prevent *overfitting.* Model performance was evaluated using *accuracy, precision, recall,* and *F1-score* metrics*,* which are standard evaluation metrics for text classification problems.

2.8. Dashboard Visualization and Clustering Analysis

The model evaluation results are visualized in the form of a Streamlit-based dashboard. In addition to classification, this study also applies the *K-Means* Clustering algorithm to identify patterns in the distribution of mail archive data that cannot be observed manually. The main tools used in this study include Python, TensorFlow, and Visual Studio Code on a Windows operating system environment.

3. RESULTS

3.1. Initial Data Exploration and Dataset Statistics

The dataset used in this study consists of 2,817 incoming letters that have undergone data validation and cleaning. All data is free of empty values, duplications, and text anomalies that could potentially interfere with the deep learning model training process. The dataset has 18 target categories representing various types of official letters such as *Surat Dinas, SPK dan PKS, Laporan Nota, Kesepahaman,* and *Risalah*. This initial exploration confirms that the dataset is suitable for training and evaluating deep learning models for multi-class mail classification. This is shown in Figure 2.

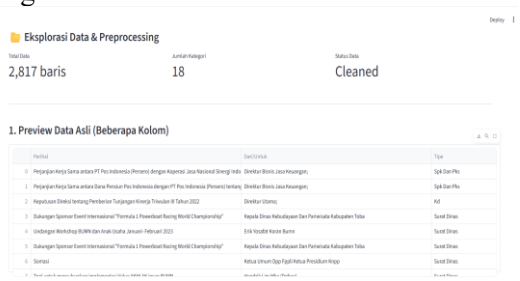


Figure 2 Data Exploration & Preprocessing

3.2. Preprocessed Data Results

Preprocessing is performed to convert raw text into a structured format that is ready to be processed by the model. This process includes text normalization, removal of non-alphanumeric symbols, and merging of several text attributes into one main column. The results can be seen in Figure 3.



Figure 3 Preview Preprocessed Data Results

3.3. Data Distribution by Letter Category

The distribution of data by letter category shows significant differences in the number of samples across classes as illustrated in Figure 4. The *Surat Dinas* category dominates the dataset, followed by *SPK* and *PKS,* while several other categories contain relatively fewer samples. This class

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

imbalance reflects real administrative conditions and presents an additional challenge for the classification task, particularly for models that are sensitive to skewed data distributions.

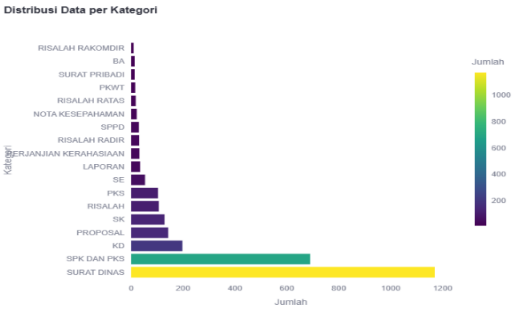


Figure 4 Data Distribution by Category

3.4. Deep Learning Model Evaluation Results

Model performance evaluation was conducted using 20% of the test data from the total dataset. The evaluation results show that the CNN model obtained the highest accuracy, precision, recall, and F1-score values compared to the LSTM and Bi-LSTM models, as shown in Table 4 and Figure 5. This superior performance indicates that CNN is more effective in capturing local and discriminative textual patterns through convolutional operations, which are common in structured and repetitive administrative mail documents. In contrast, sequence-based models such as LSTM and Bi-LSTM primarily focus on long-term dependencies, which are less dominant in formal correspondence texts.

Table 4 Accuracy Results Table

Model	Acc.	Prec.	Rec.	F1
LSTM	0.80	0.78	0.80	0.79
Bi-LSTM	0.78	0.74	0.78	0.75
CNN	0.85	0.84	0.85	0.83

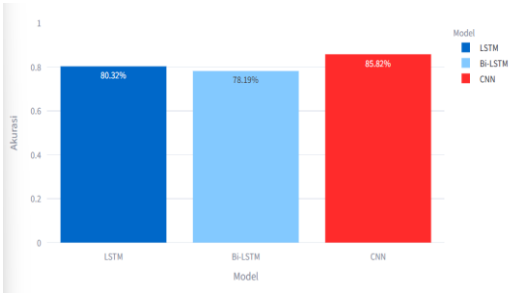


Figure 5 Accuracy Comparison (Training Results)

3.5. *Confusion Matrix Analysis*

The confusion matrix visualizations in Figures 6-8 illustrate the distribution of correct predictions and classification errors for each model. Overall, the CNN model demonstrates a higher concentration of correct predictions along the main diagonal, indicating stronger classification consistency across categories. Most misclassifications occur between semantically similar letter types, such as *SPK-PKS* and *Surat Dinas-Proposal,* which share overlapping vocabulary and structural patterns. This observation suggests that classification errors are primarily influenced by content similarity rather than model instability.

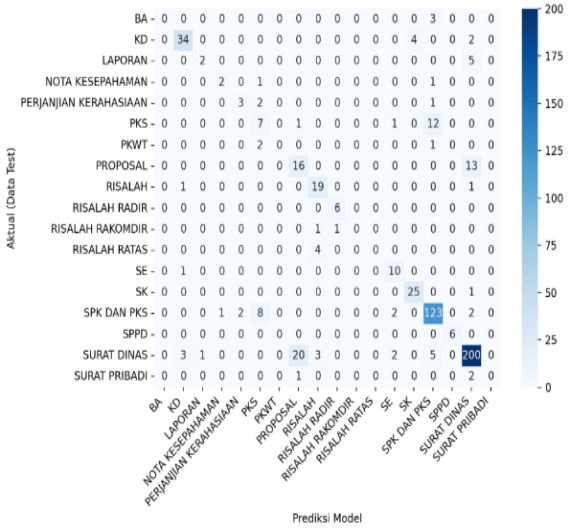


Figure 6 Confusion Matrix Model LSTM

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

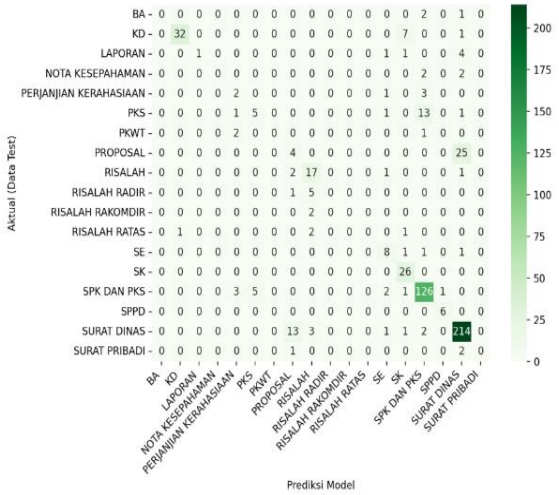


Figure 7 Confusion Matrix Model Bi-LSTM

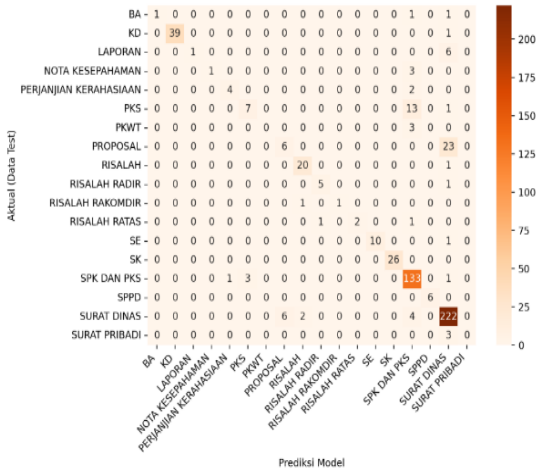


Figure 8 Confusion Matrix Model CNN

3.6. Feature Correlation Analysis Results

A feature correlation analysis using the *TF-IDF* and *Chi-Square* methods produced a list of dominant keywords that distinguish each category of letter, as shown in Figure 9 and Figure 10.



	Kategori	Kata Kunci
0	BA	kn, rekonsiliasi, pumk, 547, pelimpahan, bu, vision, evi, acara, berita
1	KD	direktur, kurir, perubahan, utama, tentang, dirut, tarif, direksi, keputusan, kd
2	LAPORAN	realisasi, mjm, prognosis, 2022, iv, triwulan, bulan, pengawasan, komisaris, laporan
3	NOTA KESEPAHAMAN	dit, corp, batang, pemanfaatan, persikindo, nota, kesepahaman, potensi, pihak, masing
4	PERJANJIAN KERAHASIAA	sembilan, src, diclosure, 097, hand, data, informasi, kerahasiaan, atau, pertukaran
5	PKS	egm, sidik, outlet, dengan, spbd, bersertifikat, kembar, kerjasama, 0725, evp
6	PKWT	sinaga, henry, lily, selanno, tenaga, sebagai, profesional, iwan, irwansyah, pkwt
7	PROPOSAL	golf, ulbi, pasundan, hpn, himpunan, mahasiswa, panitia, pengantar, sponsorship, prop
8	RISALAH	des, oktober, agustus, rabu, 17, 2024, direksi, selasa, risalah, radir
9	RISALAH RADIR	maret, pola, mei, direksi, agt, 11, risalah, radir, selasa, 2025

Figure 9 List of Keywords by Category

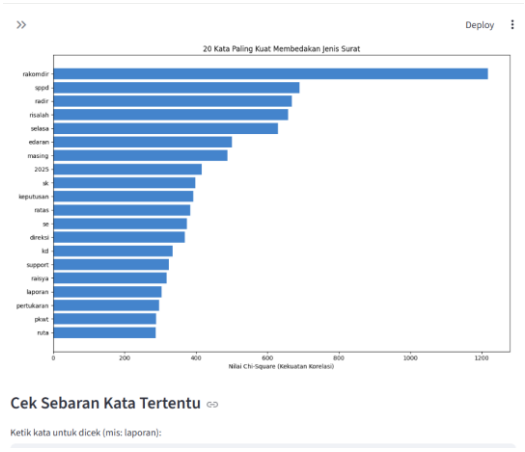


Figure 10 Bar Chart of the Strongest Words Distinguishing Types of Letters

3.7. K-Means Clustering Results

The clustering process using the *K-Means* algorithm groups mail archive data into several clusters based on textual similarity, as shown in Figures 11-17. This clustering analysis is not intended to replace the classification process, but rather to provide complementary insight into the inherent structure and thematic distribution of the dataset. By examining cluster composition and dominant keywords, this analysis helps explain why certain letter categories exhibit overlapping features, which may contribute to classification challenges observed in the confusion matrix.



Figure 11 Determining the Number of Clusters (Elbow Method)

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

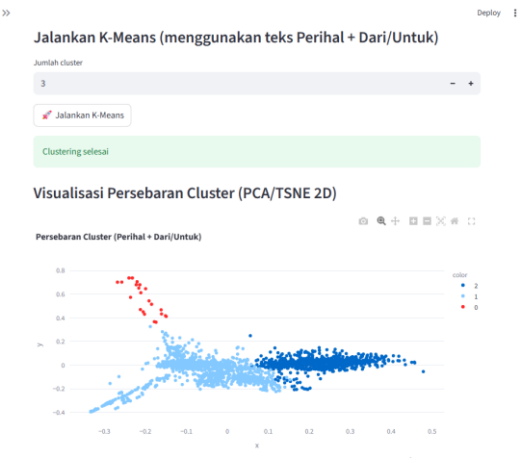


Figure 12 Cluster Distribution Visualization (PCA/TSNE 2D)



Figure 13 Keywords per Cluster (Centroid/Top terms)

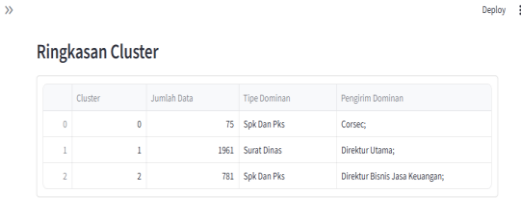


Figure 14 Cluster Summary

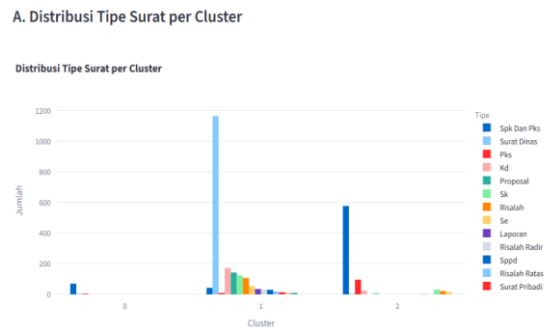


Figure 15 Letter Type Distribution per Cluster

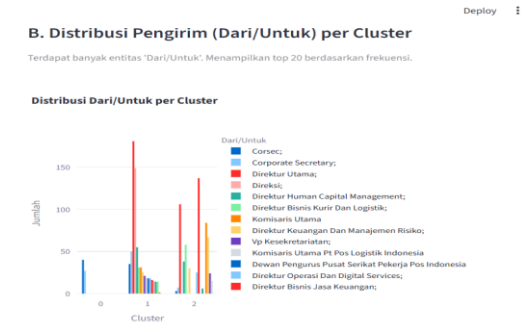


Figure 16 Distribution of Senders (From/To) per Cluster



Figure 17 Cluster Membership

3.8. Classification Trial (Manual Input)

The developed system provides a classification trial feature through manual input and new data uploads. The results of the system testing are shown in Figure 18 and Figure 19.

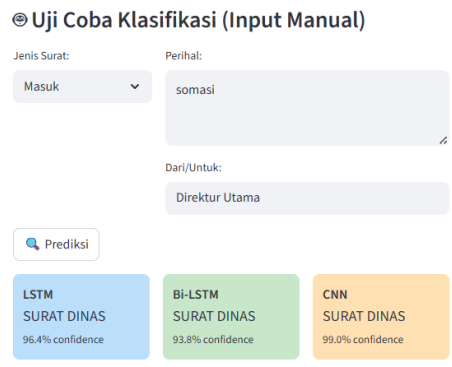


Figure 18 Classification Trial (Manual Input)

p-ISSN :1858-3911 **,** ***e-ISSN : 2614-5405***
https://journal.fkom.uniku.ac.id/ilkom

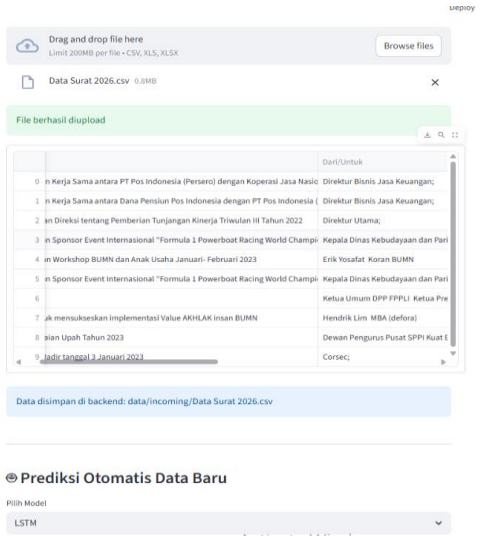


Figure 19 Upload New Data

4. DISCUSSION

4.1. Discussion of Deep Learning Model Evaluation

Based on the evaluation results, the *Convolutional Neural Network* (*CNN*) model demonstrates the most optimal performance with an accuracy of 85.82%, as well as higher precision, recall, and F1-score values compared to *LSTM* and *Bi-LSTM*. These results indicate that *CNN* is more effective in extracting local features and word patterns from administrative texts that have relatively consistent structures.

This superior performance can be attributed to the convolutional mechanism employed by CNN, which applies fixed-size filters (kernel size = 5) to capture local contextual patterns within text sequences. Such patterns are particularly prominent in administrative mail documents, where standardized phrases and recurring expressions are commonly used. By focusing on local feature extraction, CNN is able to identify discriminative textual cues efficiently without relying on long sequential dependencies, which are more emphasized in LSTM-based models

These findings are consistent with previous studies that reported the superiority of CNN-based models for document classification tasks involving structured and repetitive textual patterns.

4.2. Discussion of Confusion Matrix

Confusion matrix analysis provides a more detailed picture of the classification performance of each model in each letter category. The *CNN* model shows a higher number of correct predictions in almost all categories, with relatively fewer classification errors compared to *LSTM* and *Bi-LSTM*.

The results show that *CNN* has better class separation capabilities, especially in letter categories with similar language contexts. The error patterns that still appear generally occur in categories with overlapping text characteristics, indicating the natural challenges in administrative document classification.

4.3. Discussion of Feature Correlation Analysis

The results of feature analysis using *TF-IDF* and *Chi-Square* show that each letter category has dominant keywords that are distinctive and relevant to its administrative context. Words with high *Chi-Square* values are proven to play an important role in distinguishing one type of letter from another.

This finding reinforces the classification evaluation results, as categories with clear and consistent keywords tend to have more accurate prediction rates. In addition, this analysis provides interpretability to the model, so that the classification process is not only numerical, but also semantically understandable.

4.4. Discussion of K-Means Clustering

The clustering results using the *K-Means* algorithm show that the letter archive data naturally forms several main groups based on the similarity of the text content. The *PCA/TSNE* visualization reinforces the Elbow Method results, which indicate that **k = 3** is the optimal number of clusters for representing the thematic structure of the dataset.

Using more than three clusters tends to increase grouping complexity without revealing additional meaningful semantic

p-ISSN :1858-3911 , e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

patterns, indicating that the archive data exhibits a relatively stable and well-defined structure. In this study, clustering is not intended to replace the classification process, but rather to serve as a complementary analytical tool to help explain similarities between letter categories and overlapping textual features observed in the confusion matrix analysis.

4.5. Discussion of System Implementation

The results of system testing through manual input and new data upload show that the developed system is applicable and can be used on real data. The automatic preprocessing process ensures consistency between the training data and the new data being tested.

The *confidence* value displayed in the prediction results represents the model's level of confidence in a particular input and cannot be equated with the overall model evaluation accuracy value. This finding confirms that the system is designed not only for experimental purposes but also to support the practical process of classifying mail archives in the work environment.

5. CONCLUSION

This study demonstrates that deep learning approaches are capable of effectively classifying unstructured and imbalanced administrative mail archives. Among the evaluated models, namely LSTM, Bi-LSTM, and CNN, the CNN model achieved the best overall performance with an accuracy of 85.82%, precision of 84.04%, recall of 85.82%, and an F1-score of 83.39. This result indicates that CNN is more effective in capturing local textual patterns and recurring phrase structures commonly found in formal administrative documents.

In addition to classification performance, this study highlights the importance of interpretability and usability through the integration of feature correlation analysis, K-Means clustering, and an interactive dashboard. Feature correlation analysis provides semantic insight into dominant keywords for each letter category, while clustering analysis reveals the natural structure and similarity patterns within the dataset. The developed dashboard further supports practical usage by enabling users to evaluate model performance and perform classification on new data in an intuitive manner.

Despite these contributions, this study has several limitations. The deep learning architectures employed are relatively conventional, and class imbalance remains a challenge that affects minority categories. These limitations indicate opportunities for further methodological improvements and extensions in future work.

6. RECOMMENDATIONS

Future research is recommended to explore more advanced language models, such as transformer-based architectures including IndoBERT, to further improve classification performance and contextual understanding of administrative texts. Addressing class imbalance through resampling or cost-sensitive learning techniques is also suggested to enhance model robustness, particularly for underrepresented letter categories.

Moreover, the interactive dashboard developed in this study can be extended through usability evaluation and deployment in real operational environments. Testing the system with larger and continuously updated datasets would provide valuable insight into its scalability and effectiveness in real-world archival management scenarios.

REFERENCES

[1] F. Nikmah, D. J. Pribadi, E. A. Sukma, E. Suwarni, and I. Azmi, “Decision Support System For Handling Incoming And Outgoing Mail: To Facilitate Archives Retrieval,” *Int. J. Acad. Res. Reflect.*, vol. 10, no. 3, pp. 53–61, 2022, [Online]. Available: www.idpublications.org

[2] J. F. Indey and S. Supangat, “Implementasi Algoritma Naïve Bayes Dalam Sistem Pengarsipan Surat Berbasis Ai Di Gpi Papua Klasis Mimika Papua Tengah,” *J. Ilm. Inform.*, vol. 12, no. 2, pp. 102–113, 2024, doi:

p-ISSN :1858-3911* , *e-ISSN : 2614-5405
https://journal.fkom.uniku.ac.id/ilkom

10.33884/jif.v12i02.9087.

[3] F. R. Gerung, G. D. P. Maramis, and E. R. S. Moningkey, "Penerapan Algoritma Naive Bayes pada Arsip Surat," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 9, no. 2, pp. 676–690, 2025, doi: 10.33395/remik.v9i2.14786.

[4] R. A. Krisdiawan, "Implementasi Model Pengembangan Sistem GDLC dan Algoritma Linear Congruential Generator Pada Game Puzzle," *Nuansa Inform.*, vol. 12, no. 2, pp. 1–9, 2018, doi: 10.25134/nuansa.v12i2.1634.

[5] G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email Classification Research Trends: Review and Open Issues," *IEEE Access*, vol. 5, pp. 9044–9064, 2017, doi: 10.1109/ACCESS.2017.2702187.

[6] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751. doi: 10.3115/v1/d14-1181.

[7] R. A. Pakpahan and D. A. Utami, "Pengelolaan Arsip Surat Masuk Di Dinas Energi Dan Sumber Daya Mineral Provinsi Jawa Timur Melalui Aplikasi Tata Naskah Dinas Elektronik (TNDE)," *J. Inov. Negara dan Terap.*, vol. 4, no. 1, pp. 211–222, 2025, [Online]. Available: https://journal.unesa.ac.id/index.php/innovant/article/download/38747/12867/125823#:~:text=Uraian diatas dapat disimpulkan bahwa,elektronik dengan proses penciptaan dan

[8] D. Z. Robert, A. B. Setiawan, and D. W. Widodo, "Implementasi metode support vector machine untuk klasifikasi otomatis pengaduan publik di Kabupaten Trenggalek," *Inotek J. Inov. Teknol.*, vol. 9, p. 1617, 2025.

[9] S. Pandey, A. Taralekar, R. Yadav, S. Deshmukh, and P. S. Suryavanshi, "E-mail Spam Detection and Classification using SVM," *Int. J. Comput. Sci. Inf. Technol.*, vol. 10, no. 1, pp. 6–8, 2020, [Online]. Available: http://www.ijcsit.com

[10] I. Kasim and Sahibu, "Klasifikasi Surat Digital Menggunakan Algoritma Naïve Bayes," *J. IT Media Inf. IT STMIK Handayani*, vol. 13, no. 2, pp. 57–62, 2020.

[11] C. Magnolia, A. Nurhopipah, and B. A. Kusuma, "Penanganan Imbalanced Dataset untuk Klasifikasi Komentar Program Kampus Merdeka Pada Aplikasi Twitter," *Edu Komputika J.*, vol. 9, no. 2, pp. 105–113, 2022, doi: 10.15294/edukomputika.v9i2.61854.

[12] W. N. Ibrahem Al-Obaydy, H. A. Hashim, Y. AbdulKhaleq Najm, and A. A. Jalal, "Document classification using term frequency-inverse document frequency and K-means clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, pp. 1517–1524, 2022, doi: 10.11591/ijeecs.v27.i3.pp1517-1524.

[13] A. Liani, "Analisis Perbandingan Kernel Algoritma Support Vector Machine dalam Mengklasifikasikan Skripsi Teknik Informatika berdasarkan Abstrak," *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 240–249, 2020, doi: 10.33633/joins.v5i2.3715.

[14] J. and others MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[15] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, 2021, doi: 10.1109/TKDE.2019.2962680.

[16] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *Nuansa Inform.*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.

[17] M. A. Senubekti and L. A. Puspita Dewi, "Prinsip Klasifikasi Dan Data Mining Dengan Algoritma C4.5," *Nuansa Inform.*, vol. 16, no. 2, pp. 87–93, 2022, doi: 10.25134/nuansa.v16i2.5834.