AI-Based Classification System for Management of Internal Memoranda in Tertiary Institutions

Ifreke John Udoeka[1*], Mfoniso Polycarp Asuquo[2], Anthony Okpe Okwori[3], Idongesit Ifreke Udoeka[4], Augustine Ogbaji Otobi[5], Daniel Thursday Sunday[6], Holly Zachariah Edidem[7]

[1,2,7]*Department of Computer Science, Akwa Ibom State University, Ikot Akpaden, Nigeria*
[3]*Department of Computer Science, Federal University Wukari, Taraba State, Nigeria*
[4]*Department of Microbiology, Akwa Ibom State University, Ikot Akpaden, Nigeria*
[5]*Department of Computer Science, University of Calabar, Cross River State, Nigeria*
[6]*Department of Computer Science, Akwa Ibom State Polytechnic, Ikot Osurua, Nigeria*
[*]*ifrekeudoeka@aksu.edu.ng*

DOI: https://doi.org/10.60787/aasd.vol4no1.117

Abstract

Internal memoranda (memos) serve as formal written communications for institutions to convey decisions, instructions, policies, or announcements among staff, units, or management. Inconsistent implementation, delayed communication, poor distribution, and confidentiality concerns pose challenges to the proper management of memos. This paper presents an AI-based classification system to enhance accuracy, speed, and consistency in the management of internal memoranda across various units of tertiary institutions. This study utilizes 2,846 internal memoranda obtained from the Department of Computer Science at Akwa Ibom State University. Four AI and machine learning (ML) classifiers, including RoBERTa-base, DistilBERT, Support Vector Machine (SVM), and Random Forest (RF), were used to evaluate the system. The data were split into training, validation, and testing sets using an 80:10:10 ratio through 5-fold stratified cross-validation. The performance of the classification models was evaluated using accuracy, macro F1-score, micro F1-score, and weighted F1-score. The results showed that the RoBERTa-base classifier outperformed the other AI/ML classifiers, with a mean accuracy of 98.60% ± 1.20% across the 5-fold cross-validation, indicating strong and reliable classification performance. The standard deviation of 1.20% indicates little variation in model performance across the various validation folds. Compared with DistilBERT, SVM, and RF classifiers, the model achieved 96.49% ± 0.50%, 96.14% ± 0.77%, and 95.09% ± 0.51%, respectively. The study was limited to inflow and outflow memos from one department of an institution. However, the results showed the strength of AI and ML techniques in improving the classification and management of internal memoranda. The study therefore suggests the adoption of intelligent memorandum management to improve the accuracy of institutional communication.

Keywords: Internal Memoranda; Classification; AI; Machine Learning; Natural Language Processing; Tertiary Institutions

Introduction

Effective management of institutional information is crucial to the smooth operation of tertiary institutions. Memoranda are a major form of administrative communication that disseminate official decisions, directions, announcements, and schedules among departments and units. Memoranda remain a significant mode of conveying information in learning institutions and organizations (Pagapong & Buctot, (2026). The day-to-day generation and dissemination of such documents pose a management challenge because organizations are often expected to manage a large volume of memos and reports with varying degrees of urgency, authority, and administrative importance. Internal memoranda support an organization in sharing information such as

instructions, policies, announcements, decisions, requests, and other administrative information (Olarinde *et. al.,* 2022). In tertiary institutions, these documents can focus on academic activities, departmental administration and staff issues, institutional policies, or management decisions. Each memo has a specific purpose and can be sent to a specific group of people, so it is important to determine what is contained in the memo and to categorize it for proper information retrieval and administrative coordination. There is a growing volume of institutional records, increasing the challenges for document organization. Organizations have to deal with various types of digital and paper-based documents like reports, invoices, academic records, and official communications (Surabhi *et. al.,* 2026). If classification relies primarily on manual effort, it may be cumbersome and could lead to inconsistencies, especially when documents are distributed to various departments. Manual filing systems are inefficient in returning the information that is relevant on time when documents are not filed in the right category. The need for automated document management methodologies that can enhance the classification and organization of document collections has been pointed out in previous studies as the number of documents increases (Surabhi *et. al.,* 2026; Zhang *et. al.,* 2024).

AI offers potential solutions to these problems as it allows computer systems to process information and classify with less reliance on manual effort. Its performance in automated classification has shown significant promise in various fields (Zahoor et. al., 2024). Comparatively, Onedigbo *et al.* (2017) focused on reducing mail loss in transit as well as delays, while AI-based classification enhances memorandum labelling, retrieval, and institutional efficiency. In educational settings, AI can assist in managing administrative tasks and streamlining processes for handling, organizing, and accessing information in institutions of higher learning (Bosede *et. al.,* 2024). These techniques are increasingly relevant with the increasing availability of digital textual information. As textual documents contain a lot of information, the possibility of using text mining is emphasized by Penchala et *al.* 2024). AI in the management of memos might thus offer a more systematic way to deal with institutional communications data.

Some of the computational methods that can be used for textual information analysis and categorization include machine learning, deep learning, and natural language processing, which are all AI-based models. Ahmed *et. al.,* (2025) utilized in document classification to detect patterns in the content of documents and apply them to classify documents into pre-defined categories. Akpan *et. al.* (2017) offers a statistical perspective on linking optimality criteria, relevant to evaluating AI-based classification models for managing internal memoranda. The Naive Bayes algorithm has also been proven to be effective in document classification in educational documents. Textual information from academic meetings has been classified using the deep learning architecture Long Short-Term Memory (LSTM) and reported high classification accuracy (Zhang *et. al.,* 2024; Stow *et. al.*, 2023). These studies lay the foundation for investigating the possibility of using AI techniques to structure official memoranda.

Although there have been advancements in these areas, the classification of internal memoranda in tertiary institutions still has requirements that might not be satisfied by general-purpose document classification systems. A memo may not be sufficiently specific in terms of administrative priority, urgency, and institutional hierarchy. Therefore, the content of a memorandum ought to be taken into consideration when deciding on its classification and the needs of the users for this memorandum. Solving this problem would enhance the organization of official communication, assist decision-making in administration, and decrease the stream of finding relevant information. To address this challenge, this study proposes an Artificial Intelligence (AI) based system for the classification of official memoranda in tertiary institutions.

Related Work

Classification of documents, especially memos, has become a very important phenomenon, as memos serve as the major means of message passing in diverse organizations. Attempts have been made to modernize memo processing using AI; Pagapong & Buctot (2026) examined how to automate the administration of

memoranda-related tasks by creating a web-based application using artificial intelligence (AI) for Northeastern Mindanao State University (NEMSU) in the Philippines. For the processing of memoranda and notification of recipients, the system used the Document Understanding Transformer (DONUT) algorithm. The application was not just about managing document information; it was also intended to transform the details of a memorandum into a calendar item and a notification, thus allowing for the communication of administrative activities and eliminating manual effort. The study also showed that the accuracy and event-level correctness of the DONUT model outperformed the rule-based, statistical, and transformer-based baseline models. The paper did not provide specific numerical results, but its methodology provides an example of the ways in which AI can be integrated into the overall memorandum or note-taking workflow.

All universities in and around the globe continue to hold meetings across units and generate minutes, which are most often communicated in the form of memos and reports to convey information. Stow *et. al.* (2023) developed a model for classification of reports from Senate deliberations in the University of Port Harcourt. The system was planned to classify discussion and decision-making based on various categories such as National Youth Service Corps (NYSC) matters, election matters, and malpractice matters. For this purpose, the researchers used LSTM, a deep learning model that would be appropriate for handling sequential textual information. Their results achieved a 99% accuracy of their model and were able to classify decisions made in Senate meetings. The researchers attributed this ability to the reduction of conflicting decisions in the university Senate activities. The study was focused on reports rather than the general classification of university-related text documents, but it shows that deep learning can be effective for arranging textual documents related to the university.

An automatic system for classification of university documents has been developed in recent studies. Baimakhanova *et. al.* (2023) proposed a deep Conv2D classification model used to classify scanned documents, where each document is classified into one of seven different categories. A dataset from Khoja Akhmet Yassawi International Kazakh-Turkish University in Turkistan, Kazakhstan was used. The researchers combined convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to learn the features of the documents and to capture the word relationships between documents. The results obtained revealed the classification accuracy of 97.3% and the model was proven to be robust against variations in scanned document quality and optical character recognition (OCR) accuracy. The results suggest that deep learning methods could be helpful for document classification even if the quality of the digitized documents has varied, which would be beneficial for institutions with huge document archives.

Aubaid *et. al.* (2024) investigated approaches to structuring the ever-increasing amount of electronic textual documents. The study proposed a rule-based embedding approach and contrasted it with the conventional document classification approaches. The proposed approach yielded an accuracy of 89.91%, whereas the other traditional approaches used for comparison yielded an accuracy less than 89.91 %. This study points out the possibility of using word representation techniques and rule-based classification to enhance the classification of textual documents. The results are relevant in the institutional context where documents are distinguished by the information contained within the documents' text. Real-time classification of documents enhances the confidentiality of sensitive memos. Another approach explored to enhance the management of confidential information is through the document classification system. Tan *et. al.* (2024) proposed an automatic classification method using keyword retrieval and the k-nearest neighbours (kNN) algorithm to facilitate the identification and management of sensitive enterprise documents. The following four feature selection techniques have been tested: Term Frequency - Document Entropy (TF-DE), Term Frequency - Inverse Document Frequency (TF-IDF), Term Frequency - Odds Ratio (TF-OR), and Term Frequency - Relevance Frequency (TF-RF). The TF-DE approach achieved better F1-measure (86.3% to 93.5%) when the dimension of

the feature vector was increased from 1,000 to 15,000. The results underline the importance of feature selection and representation on the classification performance, particularly when dealing with long text documents. The study further illustrates the application of automated classification in the context of other document management requirements, beyond classifying documents by subject.

Amensisa (2020) used the Enhanced Sentence-Vector Space Model (ES-VSM) to represent text information. The problem investigated in this thesis was the ability to represent the content of documents in a manner that would ease the task of categorising the document into a new unseen category. To enhance text representation, the researcher extended ES-VSM to the bigram representation model to enable the classification process to consider the semantic information and relationships between words. This was performed in combination so that an average success rate of 95% was recorded. This work shows that the choice of proper text representation techniques for designing a document classification system is crucial, as the quality of the features extracted has a significant impact on the differentiability of the documents. In the educational field, Ekireghwo *et. al.* (2025) proposed an automatic document classification system to assist in educational reform data processing based on the Naive Bayes algorithm. The researchers surveyed and collected pertinent literature and document data to form an annotated set of documents to test different classification methods. The models used in the experiments were Naive Bayes (NB), random forest (RF), support vector machine (SVM), and multilayer perceptron (MLP). The findings revealed that Naive Bayes is the best algorithm, with an average F1 score of 95% in all the educational reform document classification tasks examined. The study demonstrated how machine learning could help to classify vast amounts of educational information and aid in the management of information more efficiently. The results further encourage the adoption of using conventional machine learning algorithms to design an automatic classification solution in academic settings.

The approach adopted by Ekireghwo *et. al.* (2025) was the application of data harmonization in institutional information management. They suggested a framework based on machine learning to tackle the inconsistencies between heterogeneous data sources in tertiary institutions. The purpose of this framework was to automatically link together different sources of information seeking to better harmonize, enrich and distribute information across educational platforms. A relational database management system and a decision support system with the help of natural language processing (NLP) and large language models (LLMs) were integrated. The researchers reported a level of accuracy of 85% for the decision support system based on NLP. The study does not involve the categorisation of memoranda but rather the harmonisation of data, but it demonstrates the application of intelligent techniques of information-processing to administrative management in higher education in general.

Abdullahi *et. al.,* (2025 investigated the use of AI for compliance verification and document workflow in Higher Education Institutions. In their proposed system, they used CNN, RNN, and CNN-RNN models to classify documents. The researchers reported that a classification of 73% was achieved by CNN and 44% by RNN, while the hybrid model achieved a classification of 27% when the original dataset was used. The accuracy rates following data improvement increased to 76%, 48%, and 79%, respectively. The results of the upgraded model showed that the hybrid CNN-RNN model outperformed the other models evaluated with the highest accuracy. The study also reported an improvement in the recall and mean squared error for the hybrid model and an increase in the F1 score for the hybrid model.

Zahoor et. al. (2024) performed a detailed analysis of the model created for text classification and sentiment analysis using a popular Explainable AI tool named Local Interpretable Model-agnostic Explanations (LIME). The results verify that it is essential to evaluate machine learning models using explainable AI tools, as accuracy and other related metrics do not ensure the correctness, fairness, and reliability of the model. They

also present a comparison of the explainability and interpretability of various machine learning algorithms using LIME. Out of the four, random forest has outperformed the other algorithms. The results of category classification are given, which clearly mark the supremacy of the random forest algorithm with 95% accuracy and 96% F1-Score. The study concluded that model interpretation is indispensable, and the interpretations provided by LIME are a good source to optimize the model performance as well as establishing trust by providing evidence-based analysis results to users.

Aside from manual memos, electronic memos (e-memos) can save costs and offer high speed of processing, reliability, and security (Olabanjo *et. al.,* 2022). A novel semantic web-based memo platform was designed and developed that allows for handling memos within an organisation. The tools used include HTML, CSS, PHP, JavaScript, jQuery, AJAX, and MySQL. The developed system was tested in three stages. Testing was done in the Department of Computer Science of the institution, then in the Faculty of Science, and finally in the entire Lagos State University. The designed system was hosted on a cloud server, and a total of 43 test cases were considered, and an average of four minutes was recorded for a complete memo cycle. The e memo allows users to generate, sign, and treat memo electronically.

Beyond memo text classification, AI-based systems can still be employed for other purposes within tertiary institutions. Ahmed & Sorour (2024) used NLP methods, feature engineering using the Count Vectorizer method to convert questions into feature vectors, and a classification module based on the Logistic Regression (LR) algorithm to categorize tertiary institutions' exam questions into categories like knowledge, comprehension, application, analysis, synthesis, and evaluation. Experimental results indicate that their system achieved an average accuracy of 98.5%. Showing the capabilities of AI in evaluating the quality of university exam papers beyond manual process that can be tedious, lengthy, and in some cases, inconsistent.

Surabhi *et. al.* (2026) presented a Machine Learning-driven framework for automated document classification and file organization using Natural Language Processing techniques. They did a comparison of several literature surveys of Document Classification Systems. Afterward, they found many limitations, including that many people have tried to make these systems better at sorting documents using Machine Learning models. Not many of these systems can actually put the documents in the right category after they have been sorted. The study introduced NLP, an AI-based algorithm for solving this problem.

Materials and Methods

This study adopts an incremental development method where memo generation, classification, and notification were separately developed and tested and then combined into a single system; this enabled each component to be tested separately without waiting for all testing to be completed.

Proposed System Architecture

This system utilized PHP/CodeIgniter, which serves as an interface for users, manages user authentication, role-based access control (RBAC), memo creation, sending, routing, and email notifications through PHP Mailer. The AI Classification Microservice receives memo text through a REST API endpoint, tokenizes the text, and processes it using a fine-tuned AL/ML classifier to predict the appropriate administrative category. The predicted class is returned to the PHP application and associated with the relevant administrative category. The MySQL Memo Database stores memo records and related information, supports communication between the application and the database, and enables the retrieval of memos by authorized users. Figure 1 shows the architecture of the proposed system.

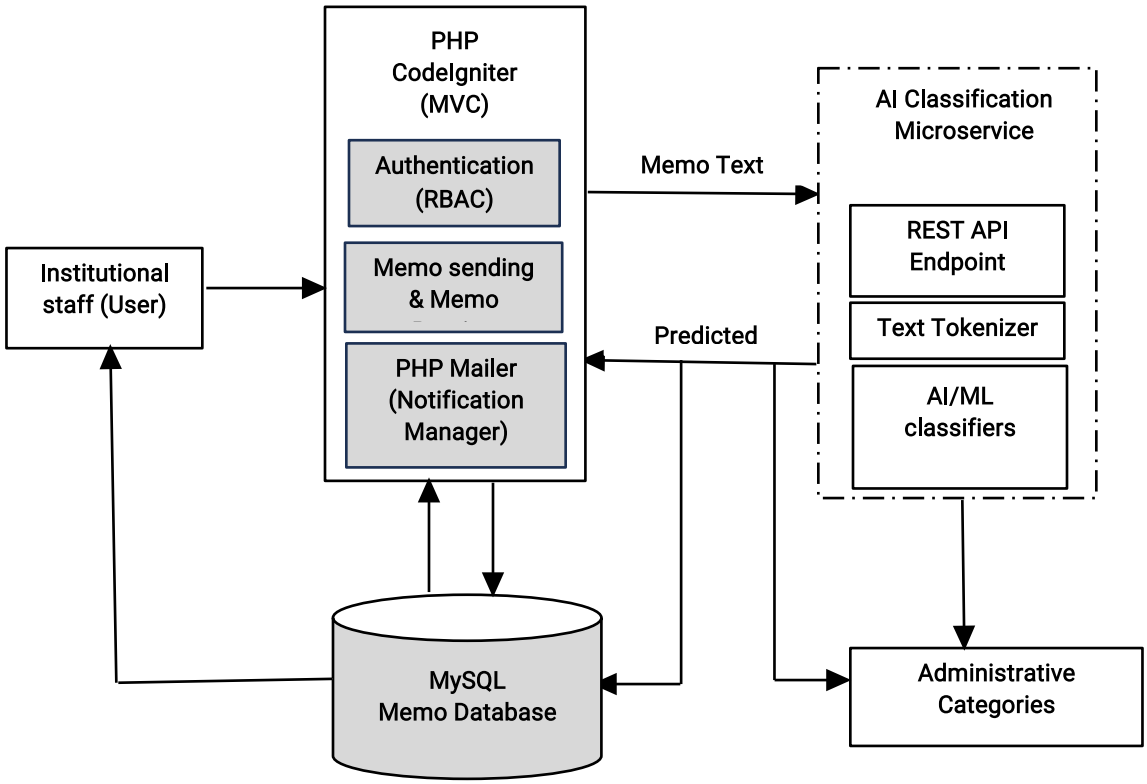


Figure 1: Proposed System Architecture

Data Collection

The dataset contains 2,846 institutional internal memoranda retrieved from the archives of the Department of Computer Science in Akwa Ibom State University. The documents represent institutional correspondence between academic departments and administrative units within the university management system. Instead of relying on manual evaluations, the memoranda were systematically organised into six statutory administrative categories depending on the originating unit, recipient, and directives contained in the document. These administrative categories include Academic Affairs, Examination and Records, Staff Administration, Governance and Discipline, Finance and Student Services, and Operations. Each datum is formatted as a structured tuple consisting of metadata details, including sender, recipient, and directive. The corpus was systematically partitioned via a stratification technique into train/validate/test splits at ratios of 80:10:10. This guarantees equal proportions of records belonging to all six administrative divisions for the purpose of benchmark evaluation during experimentation.

Data Preprocessing

Prior to training the model, the collection of institutional documents was pre-processed to prepare a clean representation for modelling. First, nonfunctional institutional noise, such as logos, graphics, watermark stamps, and internal referral numbers, were removed. Textual content was then normalized by converting text to lowercase, expanding abbreviations such as HOD (Head of Department) and DVC (Deputy Vice-Chancellor), and removing extra whitespace. Most importantly, to preserve crucial routing information embedded in documents for semantic analysis, each document was parsed to structure its syntactic composition in x = "FROM: [origin] | TO: [destination] | SUBJECT: [directive]" format. This preprocessing enables self-attention models to identify relationships among sender-receiver entities and directives. Empirical length profiling

indicated relatively short sequences, with a mean μ = 18.4 words and a maximum length of 33.0 words. Accordingly, the maximum token sequence length cap L_max was set to 64 tokens. This length accommodates 100% of tokenized records and reduces transformer self-attention complexity by more than 25% compared with the 256/512-token lengths conventionally used. Target categorical labels were assigned unique discrete encodings $y \in \{0,1,2,3,4,5\}$, which correspond, respectively, to six statutory administrative divisions.

Modelling Architectures and Fine-Tuning

In selecting an optimal machine learning algorithm for this study, experiments were conducted using classical machine learning algorithms and transformer-based architectures to establish benchmarks for institutional document routing purposes. Traditional models were represented by a linear support vector machine (Linear SVM) and a Random Forest classifier algorithm using sub-linear term frequency-inverse document frequency (TF-IDF) unigram and bigram representations, whereby calibrated probability estimates were made via Platt Scaling of Linear SVM model output probabilities.

The deep learning model chosen for semantic interpretation of institutional documents includes the BERT family of pre-trained models, which leverage the bidirectional encoding representation capability to understand the semantics within a memo. Although BERT-family models such as BERT-base provide high-level semantic representation for the texts fed into the system, the massive number of parameters used in such models makes them computationally expensive. Hence, DistilBERT, which approximates nearly 97% of the contextualized semantic representation capabilities compared to BERT using 40% fewer parameters, was chosen for lightweight institution-wide application deployment purposes. Furthermore, RoBERTa-base models were fine-tuned using byte-level Byte-Pair Encoding along with a dynamically masked language modelling technique for deeper contextual representation. Both Transformer models were fine-tuned on structured memo documents consisting of the sender, recipient, and subject lines. Tokenizers convert the text into numerical tokens before clipping/padding to a maximum length of 64 tokens. After padding the tokenized texts to create embeddings for input processing, masks were created to distinguish between actual tokens and padding during the attention process. A classification head consisting of Dense layers connecting to a softmax activation layer was added on top of the transformer's pooled representations to produce probability predictions on six administrative classes consisting of Academic Affairs, Examinations and Records, Staff Administration, Governance and Discipline, Finance, and Student Services & Operations. The fine-tuning procedure used the Adam optimization function alongside the Early Stopping callback function during the training process. During the fine-tuning stage, models were able to capture unique vocabulary and hierarchy specific to the respective operational domains in the institution. For instance, academic terms such as accreditation, curriculum, and matriculation are recognized as belonging to Academic Affairs, whereas timetables, transcripts, and result vetting belong to the Examinations and Records category. Similarly, annual appraisal (A&PC), cadre conversion, and promotion are recognized as belonging to Staff Administration, whereas disciplinary panels, misconduct summons, and Council directives belong to Governance and Discipline. Payment vouchers, petty cash, and procurement fall under the Finance Department, and hostels accommodation, SIWES placements, and student unions' activities belong to the Students' Affairs & Operations department.

Evaluation Metrics

We evaluated the model's performance on 285 memos sampled from an independent held-out testing dataset using multi-faceted evaluation metric approach to provide insight on both accuracy and reliability of the model on each administrative division. The overall performance evaluation used Accuracy and the three types of F1-scores: Macro-average, Micro-average, and Weighted Average. Out of these metrics, the Macro-average F1-Score is considered as the main performance benchmark metric because it calculates the average of all the F1

scores of all the six classes. Thus, this ensures that each administrative department is given equal importance without neglecting any of them. The Class-specific metrics include precision and recall metrics at each class level. Such metrics become important especially when considering situations in which error classification may cause wrong routing of organizational mail correspondence. Moreover, the ROC-AUC curve was plotted for each classifier model at one-vs-rest basis. A multi-class confusion matrix was used as a visualization tool to understand correct vs. incorrect classification for each administrative division. Lastly, we measured inference time, Disk usage, and Training time on a generic CPU to assess practicality and scalability for production use in a memo management system.

Results

The RoBERTa-based fine-tuned classifier, which was chosen as the classification system for the deployed system, obtained 98.60% accuracy across the four metrics. The primary reference metric was thus chosen to be Macro F1 as the distribution of support was not equal between the classes of urgency. The classification report is shown in Table 1.

Table 1. Classification Report of DistiBERT

	Precision	Recall	F1-Score	Support
Academic affairs	0.9796	1.0000	0.9897	48
Examinations and Records	1.0000	0.9792	0.9895	48
Staff administration	0.9792	1.0000	0.9895	47
Governance and Discipline	1.0000	0.9787	0.9892	47
Finance	1.0000	1.0000	1.0000	47
Student Services and Operations	0.9583	0.9583	0.9583	48
accuracy	------	------	0.9860	285
macro avg	0.9862	0.9860	0.9860	285
weighted avg	0.9861	0.9860	0.9860	285

The model achieved a flawless routing of the financial class across Precision, Recall and F1-score with zero false positives and negatives. Both Academic affairs and Staff administration classes attained a 100% recall meaning that all academic policies or personnel memos were correctly classified and routed.

Comparative Evaluation Across Three Models

In order to compare the accuracy performance of RoBERTa to alternative methods using same input training dataset, this research evaluated three additional architectures, comprising of Support Vector Machine (SVM) trained with TF-IDF feature representation, Random Forest model trained with TF-IDF, and Fine-tuned DistilBERT-base transformer. Accuracy and F1-Score performance of the four classifiers is provided in Table 2.

Table 2. Comparative Performance Across Three Models

Model	5-Fold CV Accuracy	Holdout Accuracy	Macro-F1	Micro-F1	Weighted -F1
Linear SVM+TF-IDF	96.52 ± 0.77%	96.14%	96.17%	9.14%	96.16%
Random Forest+TF-IDF	95.64% ± 0.51%	95.09%	95.08%	95.09%	95.07%
DistilBERT (Fine-Tuned)	96.80% ± 0.65%	96.49%	96.53%	96.49%	96.51%

The results reveal that both fine-tuned transformer networks exceed classical machine learning algorithms, with RoBERTa Base obtaining the best performance with 98.60% Holdout Accuracy and 98.60% Macro-F1. Specifically, RoBERTa Base exceeded the best classical algorithm, which is Calibrated Linear SVM by 2.46% and outperformed the Random Forest algorithm by 3.51% where it predicted 281 out of 285 Holdout memoranda. DistilBERT base achieved 96.49% holdout accuracy and 96.53% Macro-F1, which exceeds the Linear SVM model by 0.35% and the Random Forest model by 1.40%. It should be noted that DistilBERT base has about half of the number of parameters compared to the RoBERTa Base Model. The results revealed close internal convergence among Macro-F1, Micro-F1, and Weighted F1 metrics, with no deviation of more than 0.05% in relation to the overall accuracy score. It suggests that all models maintained consistent performance across all six statutory administrative divisions without inflation of the aggregate scores due to overprediction in dominant departments.

Detailed observation of Class-Level Performance highlights consistent superior performance of contextual transformer algorithms to classical machine learning classifiers. For instance, RoBERTa Base had a perfect prediction performance with 100.0% Precision, Recall and F1 score in classifying Finances memoranda, while obtaining F1 scores higher than 98.9% for Academic Affairs (98.97%), Examination and Records (98.95%), Staff Administration (98.95%) and Governance & Discipline (98.92%). In addition, even in case of the highly diverse institution division known as Student Services & Operations division, RoBERTa Base maintained 95.83% Balanced Precision and Recall predictions. The tuned Linear SVM produced an effective model for low latency prediction with 100% Precision in classifying academic directives memoranda. It however exhibited minimal boundary confusion between normal request by students and petty cash voucher transactions in financial class.

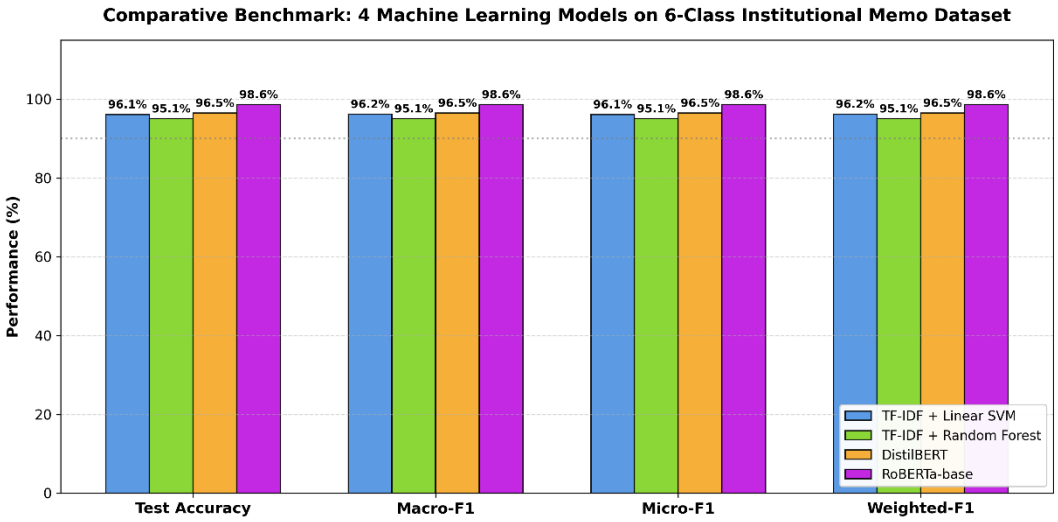
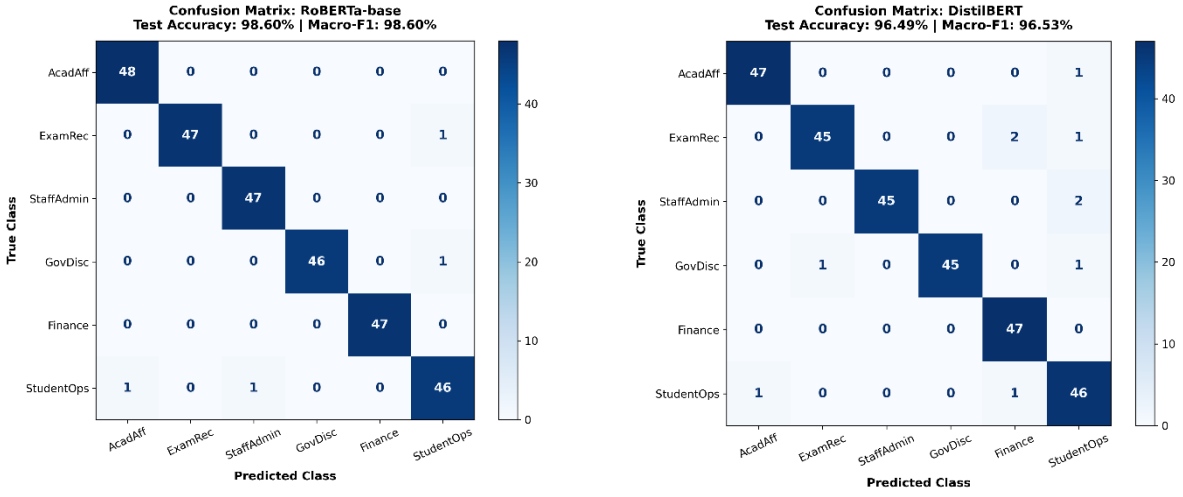


Figure 2. Clustered Metric Performance Comparison Across 4 Evaluated Models

Confusion Matrix Analysis

The analysis of confusion matrices of all four models offers profound insights into department borders in terms of which the misclassification occurred in 285 test set memos. The fine-tuned transformer model architectures showed proximity of the confusion matrix towards complete diagonalization. Specifically, the number of misclassified memos in the hold-out test set for RoBERTa-base is 4 out of 285. Interestingly, RoBERTa-base made correct predictions and achieved 100% recall for each memo in Academic Affairs (48 memos); Staff Administration (47 memos); Finance (47 memos). Furthermore, it recorded zero false positives in Examinations & Records; Governance & Discipline; and Finance (achieved 100% precision in these categories). The four errors are border errors including two memos of Student Services & Operations; one memo of Examinations & Records directed towards Student Services; one memo of Governance & Discipline misclassified in Student Services without misrouting between financial and governance affairs.



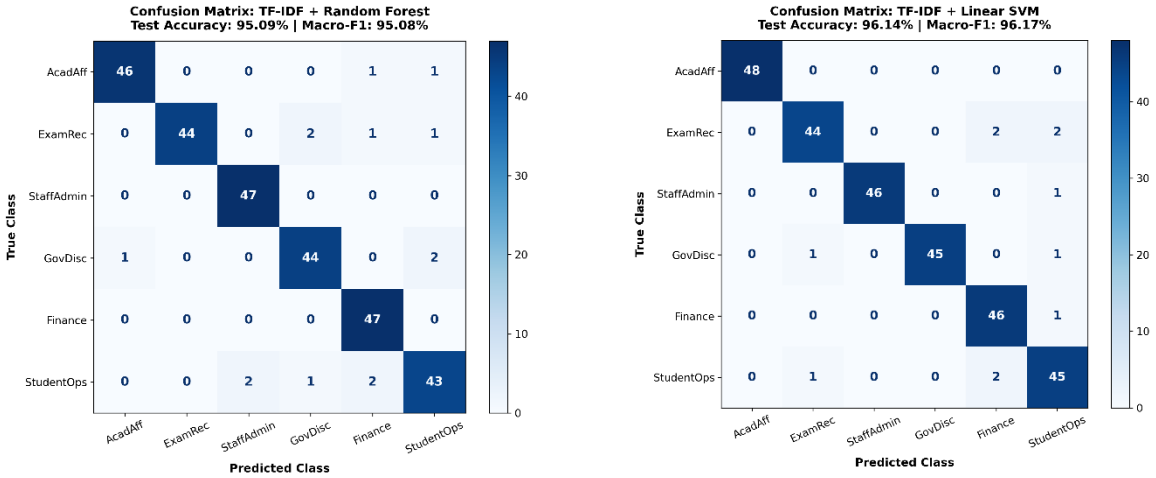


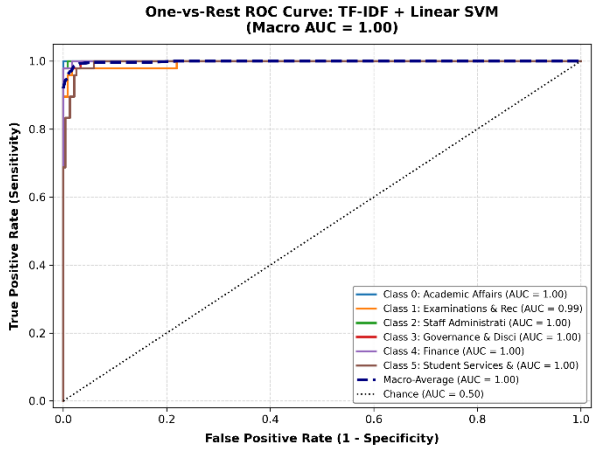
Figure 3. Confusion Matrix for all compared models

Similarly, DistilBERT model maintained high consistency with regards to diagonalization of confusion matrix, recording 10 misclassified memos out of 285-test set samples (275/285 correct predictions). Just like RoBERTa-base, DistilBERT recorded zero mistakes for Finance (47/47 correct predictions; 100% recall) and no errors in Staff Administration and Governance & Discipline (100% precision). The remaining 10 errors included: one error in Academic Affairs; two errors in Staff Administration; two errors in Governance & Discipline; two errors in Student Services; and three errors in Examinations & Records. This distribution highlights the fact that the distilled architecture maintained high precision for categorization while reducing parameter space to approximately half compared to RoBERTa base architecture. The classic models exhibited a rise in errors, mostly in the intersection borders between general student logistics and more specialized departments. The Linear SVM model committed 11 errors (274/285 correct predictions). It however retained separation between the classes with zero errors in Academic Affairs (48/48 correct; 100% recall and precision); zero false positives for Staff Administration and Governance & Discipline (100% precision). The main vulnerability was detected in Examinations & Records department where 4 memos were wrongly classified (2 memos routed to Finance and 2 memos to Student Services) and another 3 memos in Student Services & Operations. Among the four benchmarked models, Random Forest exhibited the highest error rate with a record of 14 errors in categorizing the documents. Similar to DistilBERT and Linear SVM, it managed to achieve 100% recall in Staff Administration (47/47 correct predictions) and 100% recall in Finance (47/47 correct predictions). Nevertheless, the model experienced errors mainly in Student Services where 5 memos were incorrectly categorized leading to reduction in recall rate to 89.58%. Four memos from Examinations & Records were mistakenly classified to Governance, Finance, and Student Services. This pattern reveals inherent limitations in bag-of-words feature representation and decision tree surface in classifying documents involving complex self -attention mechanisms necessary to distinguish students' notices from other transactions.

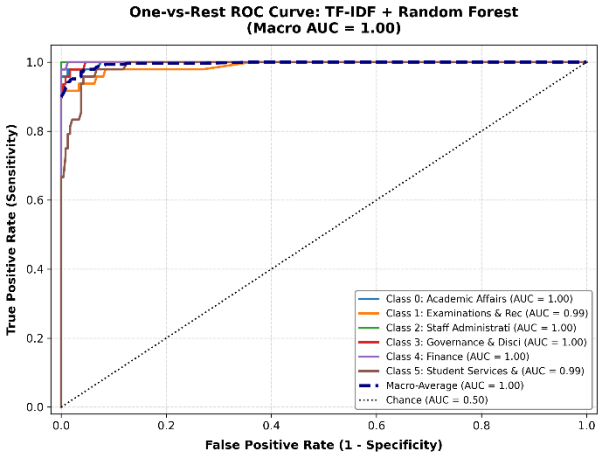
ROC-AUC Analysis

To evaluate the discriminative power of each of the models, multiclass One vs Rest (OvR) Receiver Operating Characteristics (ROC) was used to assess each model's capability in ranking true institutional categories ahead of competing administrative classes regardless of arbitrary argmax decision threshold values. This process of evaluating each model based on arbitrary argmax decision threshold helps identify if the error in classification occurs due to poor ranking or marginal probabilities at close decision boundaries. Figure 4a depicts that the performance of Calibrated Linear SVM produced almost ideal class separability with respect

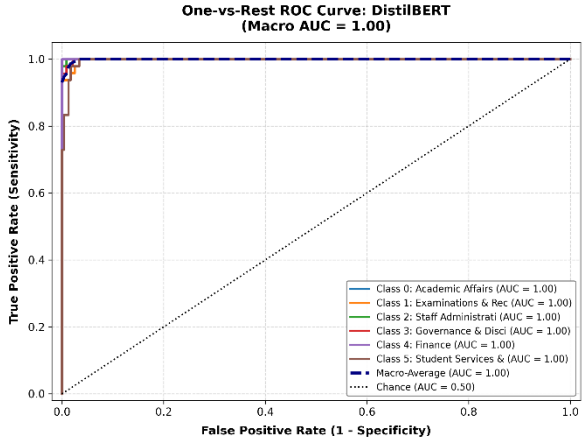
to each administrative classification, with respective AUC values of 1.00 for Academic Affairs, Staff Administration, Governance & Discipline, Finance, and Student Services & Operation, and 0.99 for Examinations & Records, resulting in a macro-average AUC of 1.00. These near-perfect rankings reveal the discriminative capability of the probability distribution estimated via Platt calibration, even if occasionally a few documents were misrouted due to an inaccurate linear decision boundary using the default argmax decision function. Figure 4b illustrates the ROC performance of the random forest classifier with near-perfect class AUC values of 1.00 for Academic Affairs, Staff Administration, Governance & Discipline, and Finance; 0.99 for Examinations & Records and 0.99 for Student Services & Operation, thus achieving a macro average AUC value of 1.00. Slight discrepancies in the ranking for the Student Services and Examination curve could be attributed to the erroneous predictions observed on the confusion matrices, where the bag-of-words splitting strategy encountered difficulties segregating the lexicon ambiguity of routine student circulars from centralized notifications.



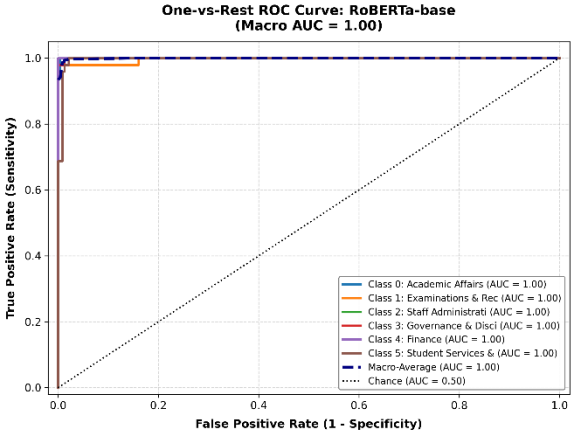
(a)



(b)



(c)



(d)

Figure 4(a-d). ROC Curves for all compared models

The fine-tuned DistilBERT model produced near-perfect ranking with AUC value of 1.00 for all six functional classes, resulting in overall Macro-AUC value of 1.00 (Figure 4c). The near perfect ranking of the ROC analysis suggests that the ten errors produced during the test phase occurred due to marginal probabilistic differences between semantically adjacent administrative categories such as student affairs and academic planning and not confusion among entirely different domains, implying superior bidirectional contextual encoding compared to frequency-based word embeddings. The finest class separability was achieved using the RoBERTa base model, producing a flawless AUC value of 1.00 for all six administrative statutory categories and an overall macro AUC value of 1.00 (Figure 4d).

The ROC curves for RoBERTa rise almost vertically toward the ideal top-left coordinate (0, 1), demonstrating that the model assigns decisively higher confidence scores to correct institutional designations while maintaining a near-zero false-positive rate across all sensitivity thresholds. This perfect class ranking directly corroborates RoBERTa's benchmark-leading test accuracy (98.60%) and its flawless classification of critical administrative domains, such as Finance and Academic Affairs, confirming its robustness for automated production routing in tertiary institutions.

System Functionality Outcomes

Besides being assigned to a class, memos are also automatically routed according to the system's predicted administrative category, with routing protocols, access privileges, and notification frequency tied to that category. This is done to ensure memos are not left unattended. High-priority governance, disciplinary, and personnel memos trigger prompt notifications to the relevant university recipients, with access to sensitive Governance & Discipline and Staff Administration files restricted through a cryptographic random-access token (enc_id), ensuring that only authorized personnel and designated panel members can view the memo's contents. Financial requisitions and payment vouchers are automatically indexed and dispatched toward the Bursary approval queue, while Examination and Records notices and routine Student Services communications are directed into their respective departmental inboxes with standardized update intervals.

Discussion

This research presents comparative results showing that fine-tuned transformers outperform machine-learning baselines on Institutional Memos Classification, while the results make clear the complications of what constitutes "best" when the application environment and accuracy of prediction are considered concurrently. From the experimental findings, RoBERTa-base recorded the highest overall accuracy and macro F1-score (both at 98.60%) when compared against DistilBERT (which achieved 96.49% accuracy and 96.53%). Moreover, RoBERTa-base performed better on the Finance class achieving 100% precision and 100% recall. It also performed very well, recording 100% recall for the Academic Affairs and Staff Administrations classes. Nonetheless, one cannot generalize this assertion as suggesting superiority. At first glance, it seems clear that RoBERTa-base is the better-performing model in comparison with DistilBERT. Nevertheless, other parameters such as latency and size need to be considered here. For instance, RoBERTa-base consumes about 1.45x times more latency (90.95ms vs 62.71ms) and storage space twice larger (478.8MB vs 256.4MB) when comparing to DistilBERT. The characteristics described imply that DistilBERT is the preferred choice for regular campus administration purposes involving memo classification, whereas RoBERTa-base offers assurance of precise routing to the relevant department concerning personnel, finances, and academic affairs which cannot afford any risks of wrong filing of memo directives. Such differences emerge from explicit considerations of efficiency, cost of computation, and particular error distribution pattern per class rather than being a consequence of random fluctuations in experimental outcomes. Prior literature in document classification tasks has prioritized accuracy measures in evaluating models.

The increment in accuracies of the transformer models over the baseline models (RoBERTa-base outperforming Linear SVM and Random Forest with 2.46% and 3.51%, respectively) is consistent with transformer literature applied to similar tasks on administrative documents. The smaller difference noted in this case can reasonably be attributed to the more structured and templated linguistic feature set characteristic of memos, which offer more advantages for n-gram-based TF-IDF representation used in the Linear SVM algorithm. Linear SVM was able to achieve an accuracy rate of 96.14% and macro-AUC of 1.00%.

A careful analysis of error distribution indicates some important practical considerations according to different administrative categories. The traditional methods had concentrated errors within specific boundaries. For instance, Random Forest misclassified 14 documents, with the worst-case scenario recorded in Student Services and Operations (where recall was reduced to 89.58%), while there was also another concentration in Examinations & Records where 4 memos were misrouted. Linear SVM had its major mistake on the boundary of Examinations & Records and Student Services, with 4 exam memos misrouted. From an operational perspective, routing of examination deadlines or students' announcements into the wrong administrative docket may cause process delays. Hence, it is evident that there is a need to consider contextual transformers that pay attention to relational aspects in both “From” and “To” headers along with subject text. For DistilBERT, however, it becomes necessary to analyse whether the error distribution observed is acceptable. DistilBERT committed 10 mistakes among the total test set of 285 institutional memos, while RoBERTa base made only 4 mistakes overall. In each instance, residual errors occurred in a similar pattern within the boundary of the Student Services and Operations category.

However, this pattern could not be attributed to any inherent architectural flaw. Rather, it can be attributed to the uncertain boundaries that exist in practical application space, for instance, an announcement of a multi-purpose welfare notice for a student that involved departmental clearance. For transformer family models, the frequency of high-stakes administrative dimension-related errors such as financial disbursal, disciplinary actions, and performance appraisal was negligible, implying effective recognition of explicit linguistic statutory markers present in those key divisional documents of universities. In essence, this study utilized a primary dataset containing institutional memos and administrative route prediction with respect to 285 test cases obtained from the Department of Computer Science, Akwa Ibom State University.

Several limitations exist within the analysis of results, including that the relative comparison results are drawn based on institutional memos from only one department of the university, consisting of 2,846 samples evaluated against hold-out independent test sets of 285 cases. In this instance, each test sample consists roughly of 2.1% recall of classes. Although sufficient statistical power exists in such large sample sizes for each class, care needs to be taken when interpreting generalization of metrics per class to institutions having divergent administrative nomenclature systems. To examine whether the relative ordering of results is due to the modelling technique itself or whether it is at least partially due to the specific administrative conventions at this institution, the same modelling approach should be extended to multi-institutional memos in future work.

Conclusion

This study presents an AI-based classification system for the automated management of internal memoranda in tertiary institutions, integrating a PHP CodeIgniter web application with an AI/ML classification model. The proposed system classifies memoranda into administrative categories, and the results showed the strength of AI and ML techniques in improving the classification and management of internal memoranda. This study demonstrates the potential of solving the classification task using AI/ML tools and functions in an institutional category of internal memoranda that other works have not addressed, and it lays the foundation for further development of intelligent memo management systems for broader application of AI tools and functions.

References

Abdullahi, A., Alhassan, J. K., & Bashir, S. A. (2025). Document classification in higher education institutions using deep learning: A CNN, RNN, and hybrid CNN-RNN approach. *In Edited Book of Conference Proceedings. Faculty of Science, Federal University of Lafia.* https://doi.org/10.62050/fscp2024.462

Ahmed, H. M., & Sorour, S. E. (2024). Classification-driven intelligent system for automated evaluation of higher education exam paper quality. *Education and Information Technologies*, *29*(15), 19835-19861. https://doi.org/10.1007/s10639-024-12555-9

Ahmed, A. S., Haddad, A. A. A., Hameed, R. S., & Taha, M. S. (2025). An accurate model for text document classification using machine learning techniques. *Ing. des Syst. d'Information*, *30*(4), 913-921. https://doi.org/10.18280/isi.300408

Akpan, S., Akra, U., Asuquo, B., & Udoeka, I. (2017). On a new Lp – Class to establish the relationship between various optimality criteria. *Global Journal of Mathematics*, 10(2), 656-664.

Amensisa, A. D. (2020). Text document categorization using enhanced sentence vector space model and bi-gram text representation model based on novel fusion techniques. *New Media and Mass Communication,* 93. https://doi.org/10.7176/NMMC/93-03

Aubaid, A. M., Mishra, A., & Mishra, A. (2024). Machine learning and rule-based embedding techniques for classifying text documents. *International Journal of System Assurance Engineering and Management,* 15(12), 5637-5652. https://doi.org/10.1007/s13198-024-02555-w

Bosede, F. B. K. F. F., Fasae, K., & Obadare, F. T. O. F. T. (2024). Artificial Intelligence and Communication in Tertiary Institutions: Relevance and Challenges. *Multidisciplinary Journal of Arts and Language Education (MJALE)*, *5*(1), 366–380.

Baimakhanova, A., Zhumadillayeva, A., Avdarsol, S., Zhabayev, Y., Revshenova, M., Aimeshov, Z., &Uxikbayev, Y. (2023). Automatic classification of scanned electronic university documents using deep neural networks with Conv2D layers. *International Journal of Advanced Computer Science and Applications*, *14*(5).

Ekireghwo P.O., Blamah N.V., B.A. Ajayi, Anyachebelu T.K., Salihu A.A., (2025) "An Implementation of Machine Learning-Based Framework for Tertiary Institutions Data Harmonization using Large Language Model,"*Eur. J. Appl. Sc. Eng. Technol., vol. 3(3), pp. 285 302*. DOI: 10.59324/ejaset.2025.3(3).20

Olabanjo, O. (2022). A Novel Web-Based eMemo for Tertiary Institutions in Developing Nations: From Conceptualization to Implementation. *Preprints.* https://doi.org/10.20944/preprints202207.0042.v1

Olarinde, M. O., Egwuche, O. S., Ganiyu, M., & Adeola, A. A. (2022). E-memo management system for organizational internal communication. International Journal of Scientific & Engineering Research, 13(2), 92–100. https://www.researchgate.net/publication/383271703

Onedigbo, M. O., Idara, I. J., Godwin, O. A., & Ifreke, J. U. (2017). SMS-Based Mails Tracking Management System Using Smart Phones. *The Proceedings of the 8th iSTEAMS Multidisciplinary Conference, Caleb University, Lagos, Nigeria*.

Pagapong, E. V., & Buctot, J. D. O. B. (2026). Web application AI-driven memorandum processing and automated recipient notification system using the DONUT (Document Understanding Transformer)

algorithm [Undergraduate thesis, Northeastern Mindanao State University]. Scribd. https://www.scribd.com/document/980944807/Web-Application-for-AI-Driven-Memorandum-1

Penchala, S., Murad, S. A., Roy, I., Gupta, B., & Rahimi, N. (2024). Unveiling text mining potential: a comparative analysis of document classification algorithms. *Proceedings of 39th International Confer*, *98*, 103-115.

Stow, M. T., Ugwu, C., & Onyejegbu, L. N. (2023). Improved Hybrid Model for Classification of Text Documents. *European Journal of Artificial Intelligence* 2(2), 17-23. https://doi.org/10.24018/ejai.2023.2.2.22

Surabhi, C. D., Suraksha, A. S., Kashyap, U., Yashaswini, C. H., & Madhu, C. K. (2026). Machine learning-driven document classification and automated file organization using text analysis. International Journal of Scientific Research in Engineering and Management (IJSREM), 10(15).

Tan, L., Yi, J., & Yang, F. (2024). Improving performance of massive text real-time classification for document confidentiality management. *Applied Sciences*, *14*(4), 1565. https://doi.org/10.3390/ app14041565

Zahoor, K., Bawany, N. Z., & Qamar, T. (2024). Evaluating text classification with explainable artificial intelligence. *IAES International Journal of Artificial Intelligence*, *13*(1), 278-286. https://doi.org/10.11591/ijai.v13.i1.pp278-286

Zhang, P., Ma, Z., Ren, Z., Wang, H., Zhang, C., Wan, Q., & Sun, D. (2024). Design of an Automatic Classification System for Educational Reform Documents based on Naive Bayes Algorithm. *Mathematics*, *12*(8), 1127.