Advances In Multidisciplinary & Scientific Research
A Multidisciplinary & Interdisciplinary Journal

Vol. 12 No. 2, June , 2026

Development of a Sentiment Analysis AAUA Newsletter Using Deep Learning Technique

Aliyu, E.O. (PhD)
Department of Computer Science,
Adekunle Ajasin University,
Akungba Akoko, Ondo State, Nigeria
E-mail: olubunmi.aliyu@aaua.edu.ng
ORCID ID: http://orcid.org/0000-0001-7278 -3452

ABSTRACT

The identification and extraction of subjective information from textual data is vital for refining institutional communication strategies and improving decision-making processes. While prior studies have concentrated on social media and movie reviews, institutional newsletters, especially within the Nigerian educational context remain largely unexplored. This paper presents the development of a Sentiment Analysis of Adekunle Ajasin University, Akungba-Akoko (SAAAUA) newsletter using a deep learning approach. Twenty-two newsletters published between 2022 and 2024 were collected. Text was extracted using PyMuPDF for digital PDFs and EasyOCR for scanned documents, segmented into paragraphs and refined through noise removal and length normalization. A semi-automatic labelling pipeline produced 196 neutral, 62 positive and 16 negative samples. The BERT-base-uncased model was fine-tuned with a dense classification layer and softmax output, trained using the AdamW optimizer with class-weighted cross-entropy loss and early stopping. The model demonstrated exceptional performance on the Neutral class with (97% precision and recall) and moderate performance on the underrepresented negative class. These findings demonstrate the feasibility of applying transformer-based models to niche institutional corpora and establish a baseline for future natural language processing (NLP) research in Nigerian university communication.

Keywords: Sentiment Analysis, BERT, Deep Learning, Natural Language Processing, Institutional Communication, AAUA Newsletter, Text Classification.

Aims Research Journal Reference Format:
Aliyu E. O. (2026): Development of a Sentiment Analysis AAUA Newsletter Using Deep Learning Technique.
Advances in Multidisciplinary Research Journal. Vol. 12 No. 2, Pp 59-66..
www.isteams.net/aimsjournal. dx.doi.org/10.22624/AIMS/V12N2P5

1. INTRODUCTION

Sentiment analysis is the computational identification of opinions, attitudes, and emotions within textual data. It has become a cornerstone of modern Natural Language Processing (NLP) research. Institutional newsletters represent a unique category of text that has received comparatively little academic attention. Unlike social media posts or product reviews, newsletters from educational institutions convey carefully structured, official information, yet they carry underlying tones that reflect the institution's culture, priorities, and responses to challenges. Understanding these tones programmatically can offer meaningful insights for administrators and policymakers. The Adekunle Ajasin University, Akungba-Akoko (AAUA) publishes a periodic newsletter, The Highlights, which serves as a primary vehicle for communicating academic events, policy updates, staff appointments, and university achievements to its community.

Between 2022 and 2024, twenty-two editions of this newsletter were released, collectively forming a valuable corpus for sentiment study. However, these newsletters existed in mixed formats; some as digitally generated PDFs and others as scanned image-based documents; presenting a non-trivial text extraction challenge. Traditional lexicon-based approaches to sentiment analysis rely on fixed vocabulary lists and syntactic rules, making them brittle when applied to domain-specific, formal institutional language (Liu, 2012; Medhat et al., 2014). Machine learning methods such as Naive Bayes and Support Vector Machines (SVMs) improved upon these limitations by learning patterns from labelled data but require hand-crafted feature engineering and lose contextual information across longer text spans (Pang & Lee, 2008). The emergence of transformer architecture, most notably BERT (Devlin et al., 2019) has revolutionised NLP by enabling models to capture bidirectional contextual relationships across entire sequences without relying on explicit feature engineering.

This study therefore investigates the application of *BERT-base-uncased* for sentiment analysis of Adekunle Ajasin University Akungba-Akoko (SAAAUA) newsletter paragraphs into three categories: positive, neutral, and negative. The aim is to demonstrate that transformer-based deep learning can be effectively adapted to small, domain-specific institutional corpora, establishing a computational tool that can support university communications monitoring and analysis.

The remainder of this paper is organised as follows: Section 2 reviews related literature on sentiment analysis and deep learning models; Section 3 describes the dataset, preprocessing pipeline, and learning algorithm; Section 4 presents the experimental setup and results; and Section 5 concludes with implications and recommendations for future research.

2. LITERATURE REVIEW

Sentiment analysis has been studied extensively across diverse domains. Pang et al., (2002) pioneered machine-learning-based sentiment classification, demonstrating that supervised classifiers could outperform human-defined lexicons on movie review data. Subsequent surveys by Liu (2012) and Medhat et al., (2014) systematised the field into document-level, sentence-level, and aspect-level tasks, noting the growing importance of fine-grained and multi-class classification frameworks. The application of deep learning to NLP accelerated substantially with Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which addressed the vanishing gradient problem and enabled better modelling of sequential data (Zhang et al., 2018). However, these architectures remained sequential and struggled to capture long-range dependencies efficiently. The introduction of attention mechanisms and, subsequently, the Transformer architecture by Vaswani et al., (2017) marked a pivotal shift.

Devlin et al., (2019) proposed Bidirectional Encoder Representations from Transformers (BERT), pre-trained on large-scale corpora using masked language modelling and next-sentence prediction. BERT achieved state-of-the-art results on eleven NLP benchmarks, including the GLUE benchmark (80.5% average accuracy), MultiNLI (86.7%), and SQuAD v1.1 (93.2% F1). Its bidirectional training allows the model to incorporate both left and right context simultaneously which is a critical advantage over unidirectional predecessors. Alaparthi and Mishra (2021) subsequently adapted BERT specifically for sentiment analysis, confirming that fine-tuning on domain-specific data yields superior performance compared to generic models. Tan et al., (2023) conducted a systematic review of sentiment analysis approaches, categorising them into classical machine learning, deep learning, and ensemble techniques, and highlighted that transformer models consistently outperformed earlier methods, particularly on imbalanced and short-text datasets.

In the educational domain, Troussas et al., (2020) applied sentiment analysis to student feedback, demonstrating that automated sentiment detection could inform curriculum improvements and enhance learning experiences. Krouska et al., (2020) extended this to social media analysis for student performance evaluation in higher education. Notably, most studies within the Nigerian academic context focus on social media data, leaving the analysis of formal institutional communications underexplored.

Mao et al., (2024) conducted a comprehensive systematic review of sentiment analysis methods up to 2023, noting that challenges persist in handling imbalanced class distributions, domain-specific corpora, and low-resource languages. Their findings directly motivated the methodological decisions in this study, particularly the adoption of class-weighted loss functions and stratified sampling to address AAUA's imbalanced newsletter dataset. The most recent frontier involves multimodal fusion models such as TF-BERT (Hou et al., 2025), though text-only approaches remain the standard for formal written documents.

3. METHODOLOGY

3.1 Dataset Representation and Preprocessing

The primary dataset comprised 22 issues of AAUA's The Highlights newsletter published between 2022 and 2024, collected in PDF format. These documents presented significant format heterogeneity: while some were digitally generated text-based PDFs, the majority were scanned or image-based PDFs requiring optical character recognition (OCR). A dual-extraction pipeline was implemented using PyMuPDF for text-based files and EasyOCR for scanned documents. Initial extraction across all 22 newsletters yielded approximately 657 raw paragraphs from 22 source files.

Text preprocessing involved the following sequential steps: (1) removal of repeated headers, footers, page numbers, and OCR artefacts; (2) segmentation of documents into paragraphs to preserve contextual coherence while maintaining manageable sequence length for transformer input; (3) minimum-length filtering, eliminating paragraphs with fewer than five words; and (4) chunking of excessively long paragraphs (maximum 195 words post-processing) to comply with BERT's 512-token constraint. The refined dataset consisted of approximately 590 usable paragraphs across 18 newsletters (four newsletters contributed no valid text after cleaning) with an average paragraph length of approximately 37 words.

Sentiment labelling followed a semi-automatic pipeline. First, the CardiffNLP Twitter RoBERTa-base model was applied as a weak supervision step to generate preliminary labels across three categories: positive, neutral and negative. These automatically generated labels were then manually reviewed and corrected where necessary with a focus on removing misclassified paragraphs that resulted from OCR noise or ambiguous institutional language. The resulting labelled dataset consisted of 196 neutral paragraphs (72.6%), 62 positive paragraphs (23.0%) and 16 negative paragraphs (5.9%), reflecting the inherently informational and supportive tone of official university communications. Table 1 summarises the final dataset composition.

Table 1: Final Labelled Dataset Composition (N = 274 after semi-automatic labelling)

Attribute	Neutral	Positive	Negative
Paragraphs	196 (72.6%)	62 (23.0%)	16 (5.9%)

3.2 Learning Algorithm and Evaluation Measures

BERT-base-uncased (Devlin et al., 2019) is a 12-layer, 110-million-parameter transformer encoder pre-trained on the English Wikipedia and BooksCorpus using two self-supervised objectives: Masked Language Modelling (MLM) and Next Sentence Prediction (NSP). Its bidirectional attention mechanism allows each token to attend to every other token in the input simultaneously, producing rich contextual embeddings that are highly transferable to downstream tasks. For sentiment classification, the model was adapted by appending a dense classification head to the [CLS] token; a special token prepended to every input sequence whose final hidden state serves as an aggregate sequence representation. A fully connected linear layer projected this 768-dimensional embedding into three output logits (negative, neutral, positive), followed by a softmax activation to produce class probabilities. The architecture is presented in Figure 1.

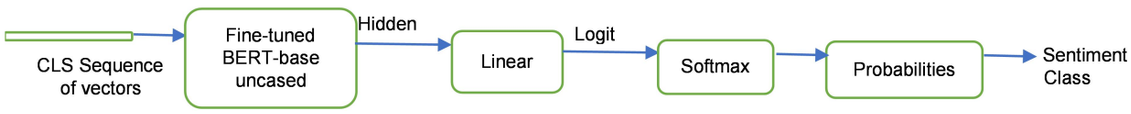

Figure 1: AAUA Newsletter Sentiment Architecture

Training complexity for BERT is $O(n^2 \cdot d)$ per layer and per sequence (where n is sequence length and d is model dimension) making it computationally intensive. However, fine-tuning on a small domain-specific dataset is considerably more efficient than pre-training from scratch. The 128-token maximum sequence length constraint adopted here reduced computational load while accommodating the average paragraph length in the corpus.

3.2.1 Training and Applying the BERT Algorithm

The dataset was split into 80% training (n = 219) and 20% testing (n = 55) subsets using stratified sampling to ensure proportional representation of all three sentiment classes across both partitions. Tokenisation used BERT's WordPiece tokeniser, which decomposes words into sub-word units; enabling the model to handle rare or domain-specific terminology not seen during pre-training. Each paragraph was tokenised with truncation at 128 tokens, padding for shorter sequences, and conversion to PyTorch tensors.

The fine-tuning configuration employed the AdamW optimiser with a learning rate of 5×10^{-5} ; a standard value for BERT fine-tuning that balances rapid convergence with stable training dynamics. Class-weighted cross-entropy loss was applied to address class imbalance, assigning higher penalty weights to minority classes (particularly negative), thereby encouraging the model not to default to majority-class predictions. Early stopping based on validation loss prevented overfitting on the small training set. Evaluation was performed using accuracy, precision, recall, F1-score, macro-average and weighted-average F1-score across all three sentiment classes. A confusion matrix was generated to visualise class-level prediction patterns as shown in Figure 2.

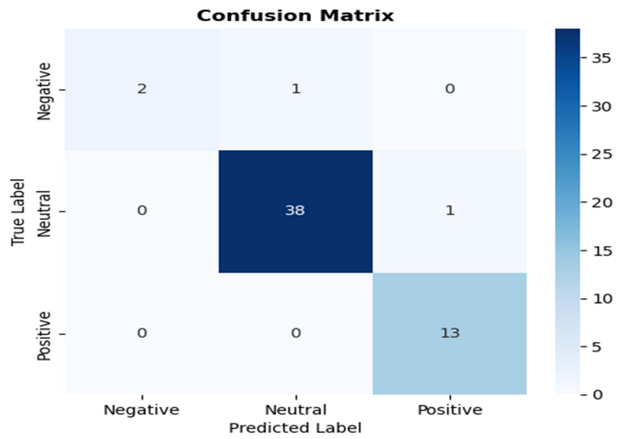

Figure 2: Confusion Matrix of Test Predictions

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Experimental Setup

All experiments were implemented in Python within a Jupyter Notebook environment. The primary deep learning framework was PyTorch, with the Hugging Face Transformers library used to access and fine-tune the *BERT-base-uncased* model. Supporting libraries included Pandas and NumPy for data manipulation, Matplotlib and Seaborn for visualisation, NLTK and scikit-learn for preprocessing and label encoding, PyMuPDF for text-based PDF extraction and EasyOCR for scanned document processing. The model was initialised using *BertForSequenceClassification.from_pretrained('bert-base-uncased', num_labels=3)*. Sentiment labels were encoded numerically (Negative = 0, Neutral = 1, Positive = 2). A custom *SentimentDataset* class wrapped tokenised inputs and labels for use with *PyTorch DataLoaders*, which handled batching and shuffling during training. The training configuration used a batch size of 16, a maximum of 10 epochs with early stopping patience of 2 epochs, and class weights inversely proportional to class frequencies.

4.2 Results and Discussion

The fine-tuned *BERT-base-uncased* model achieved an overall accuracy of 96% on the held-out test set (n = 55). However, since accuracy works well when classes are balanced, the study used precision and recall for F1-score calculation since it is useful for imbalanced binary tasks. Table 2 presents a detailed breakdown of class-level performance metrics.

Table 2: Model Performance by Sentiment Class on the Test Set (n = 55)

Sentiment Class	Precision (%)	Recall (%)	F1-Score(%)	Support (n)
Negative	67.0	80.0	72.9	3
Neutral	97.0	97.0	97.0	39
Positive	100.0	96.0	98.0	13
Macro Avg	88.0	91.0	89.3	-
Weighted Avg	96.0	96.0	95.9	55

The model demonstrated exceptional performance on the Neutral class; the dominant category in the dataset, correctly classifying 38 of 39 neutral samples (97% precision and recall). For the Positive class, performance was equally strong, with all 13 positive samples correctly identified (100% precision, 96% recall). These results confirm that BERT's bidirectional contextual embeddings effectively captured the linguistic markers of neutral informational reporting and positive celebratory language characteristic of institutional newsletters.

Performance on the Negative class, however, was comparatively weaker. With only three negative samples in the test set, a direct consequence of the severe class imbalance (16 negative paragraphs total, representing just 5.9% of the dataset): the model misclassified one negative sample as Neutral, yielding a recall of 67%. Precision for the negative class was 100%, indicating that when the model predicted negative, it was correct; the limitation lay in detecting all negative instances. This is consistent with findings by Tan et al. (2023) and Mao et al. (2024), who observed that minority-class recall is the primary performance bottleneck in imbalanced sentiment datasets. The visualisation representation is presented in Figure 3.

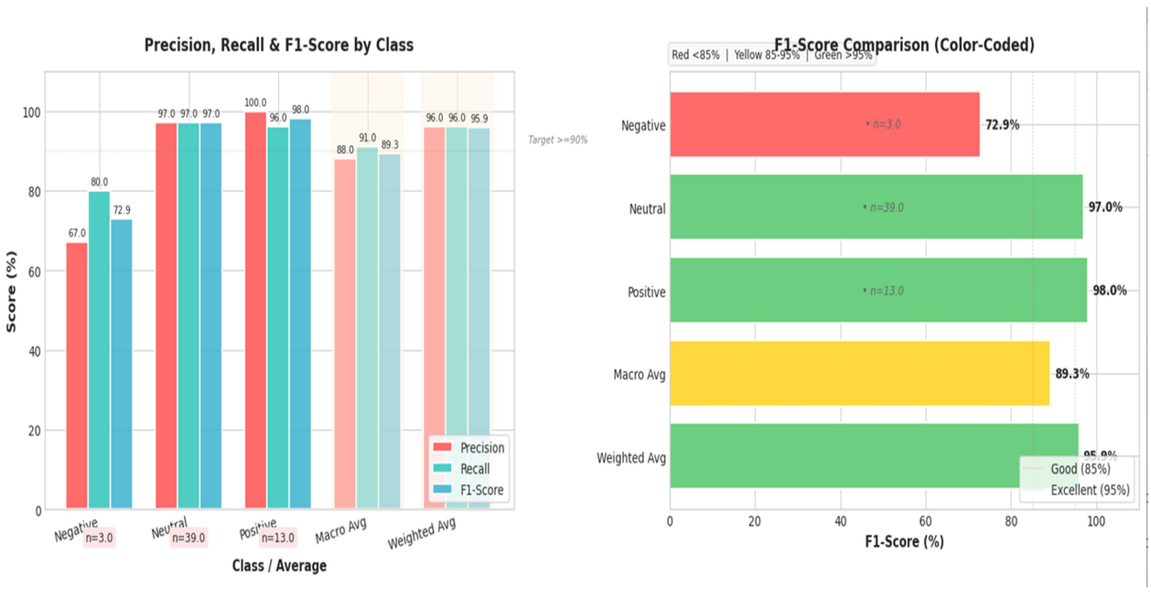

Figure 3: Sentiment Classification Metrics

Comparing these results to the baseline established by Devlin et al., (2019), the fine-tuned model achieved 96% accuracy on the AAUA domain-specific task; substantially above BERT's average GLUE benchmark score of 80.5% and comparable to its best task-specific results (for instance, SQuAD v1.1 F1 of 93.2%). This disparity is expected because fine-tuning on a targeted domain concentrates the model's representational capacity on a constrained vocabulary and style. However, it is important to note that Devlin's benchmarks evaluated general language understanding, not sentiment classification specifically, making direct comparison illustrative rather than strictly equivalent. In comparison with related studies, Alaparthi and Mishra (2021) reported sentiment classification accuracies of 90–94% using BERT on general-domain datasets. Troussas et al. (2020) achieved similar accuracy ranges in educational feedback sentiment analysis. Our 96% accuracy on a significantly smaller and more specialised corpus with limited negative samples suggests that BERT's pre-trained representations are robust even when fine-tuned on sparse domain-specific data.

This aligns with Howard and Ruder (2018) argument for universal language model fine-tuning as a cost-effective strategy for low-resource NLP. The results also have substantive implications for institutional communication management. The predominance of neutral sentiment (72.6%) corroborates the expected informational register of official university newsletters. The meaningful proportion of positive sentiment (23.0%) reflects the newsletter's communicative function of celebrating institutional achievements and community milestones. The scarcity of negative content (5.9%) reflects a genuine absence of critical discourse in the newsletters. This distinction is a valuable insight for administrators seeking to understand how institutional tone is projected to stakeholders.

5. CONCLUSION

This paper presented a BERT-based-uncased deep learning framework for sentiment analysis of AAUA institutional newsletters. A dataset of 22 newsletters (2022–2024) was preprocessed using a hybrid OCR and text-extraction pipeline, segmented into approximately 590 paragraphs and labelled through a semi-automatic process into 196 neutral, 62 positive, and 16 negative samples. The BERT-base-uncased model was fine-tuned using class-weighted cross-entropy loss and early stopping, achieving an overall accuracy, precision and recall of 96% on the held-out test set. The study makes two primary contributions. First, it establishes the first known application of deep learning sentiment analysis to Nigerian university institutional communications, demonstrating the feasibility of the approach for niche domain-specific corpora. Second, it introduces a replicable preprocessing pipeline for mixed-format PDF corpora; combining PyMuPDF and EasyOCR that can be adapted for similar institutional datasets across African higher education institutions.

However, limitations include the small dataset size, significant class imbalance (particularly in the negative class) and the single-institution scope. Future research should prioritise expanding the corpus across multiple Nigerian universities, employing data augmentation strategies such as back-translation and synonym replacement to address class imbalance and exploring more advanced transformer variants such as RoBERTa or DistilBERT. A deployed dashboard offering real-time sentiment monitoring of institutional communications would represent a practical extension of this work. Integration with multilingual models could further extend applicability to institutions communicating in indigenous Nigerian languages alongside English.

REFERENCES

Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. Journal of Marketing Analytics, 9(2), 118–126.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186). Association for Computational Linguistics.

Hou, J., et al. (2025). Tensor-based fusion BERT for multimodal sentiment analysis (TF-BERT). ScienceDirect.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In Proceedings of ACL 2018 (pp. 328–339).

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. E. (2019). Text classification algorithms: A survey. Information, 10(4), 150.

Krouska, A., Troussas, C., & Virvou, M. (2020). Sentiment analysis of social media for evaluating students' performance in higher education. IEEE Transactions on Learning Technologies, 13(3), 603–612.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. Morgan & Claypool Publishers.

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University — Computer and Information Sciences.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP 2002 (pp. 79–86).

Tan, K. L., Lee, C. P., Lim, K. M., et al. (2023). A survey of sentiment analysis: Approaches, datasets and future research. Applied Sciences, 13(7), 4550.

Troussas, C., Krouska, A., & Virvou, M. (2020). Sentiment analysis of educational feedback for course improvement. Journal of Educational Computing Research, 58(2), 342–363.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser L., & Polosukhin, I. (2017).
Attention Is All You Need; 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. https://doi.org/10.1002/widm.1253