CTICTR 4(1): 137 - 158 (June 2025)

ISSN: 2814-1709

Received: 14-03-2025

Accepted: 17-06-2025

https://doi.org/10.61867/pcub.v3i1a.208

A FASTTEXT SENTIMENT ANALYSIS MODEL FOR BRAND MONITORING IN NIGERIA

Adedoyin Samuel Adebanjo [a]*, Oluwafunmilayo A. Eyisegun-Emmanuel [b], Ayokunnumi O. Osajele[c], Favour E. Esigbemi[d], Babajide E. Adeoti[e], Emmanuel Mgbeahuruike[f], Emmanuel Oyerinde[g]

Corresponding Author Email: *adebanjoa@babcock.edu.ng*

A FastText Sentiment Analysis Model for Brand Monitoring in Nigeria

Adedoyin Samuel Adebanjo [a]*, Oluwafunmilayo A. Eyisegun-Emmanuel [b], Ayokunnumi O. Osajele[c], Favour E. Esigbemi[d], Babajide E. Adeoti[e], Emmanuel Mgbeahuruike[f], Emmanuel Oyerinde[g]

[a,e,f] *Department of Software Engineering, Babcock University, Ilishan, Nigeria*
[b,c,d] *Department of Computer Science, Babcock University, Ilishan, Nigeria*
[g] *Department of Computer Science, Babcock University, Ilishan, Nigeria*

[a] *adebanjoa@babcock.edu.ng*
[b] *asukpa0404@student.babcock.edu.ng*
[c] *osajele3970@student.babcock.edu.ng*
[d] *esigbemi0857@student.babcock.edu.ng*
[e] *adeotib@babcock.edu.ng*
[f] *mgbeahuruikee@babcock.edu.ng*
[g] *oyerindee@babcock.edu.ng*

Abstract

In the digital age, online communication influences how the public views brands, making sentiment analysis crucial for understanding consumer opinions and social trends. This study outlines the design and implementation of a FastText and custom-based sentiment analysis model specifically for the Nigerian linguistic context, which includes Yoruba, Igbo, Hausa, and Nigerian Pidgin. The system uses FastText embeddings for efficient classification and features a custom model fine-tuned for local language variations. It supports real-time and batch sentiment analysis of text files (TXT/PDF) and provides a visualization dashboard to show sentiment distribution. Additionally, it offers user authentication and history management for personalized use. Built with FastAPI, Flask, and Streamlit, the platform allows seamless interaction between backend processing and the user interface. Experimental results show high accuracy and usability across multilingual datasets, even with challenges like data imbalance and dialect diversity. Overall, this work contributes to Natural Language Processing (NLP) for African languages and provides a practical framework for brand monitoring and opinion mining in Nigeria's changing digital landscape.

Keywords: Brand Monitoring, FastText, Natural Language Processing (NLP), Nigerian Languages, Sentiment Analysis.

1. Introduction

In today's digital age, social media has changed from a place for personal connections to a strong tool for business insights, customer engagement, and brand management. With over 4.5 billion active users around the globe, platforms like Twitter (X), Facebook, and Instagram are essential for businesses. They help companies communicate with customers and keep track of how people see their brands in real time [1]. The large amount of unstructured data produced by users on these platforms provides valuable insights into consumer attitudes and market trends [2].

In Nigeria, social media has become a lively space for public discussion. Consumers share their opinions about products and services, especially in areas like telecommunications, retail, and finance [3]. The immediate nature of these interactions allows organizations to assess customer satisfaction, spot new issues, and respond quickly to feedback [4].

Sentiment analysis, a part of Natural Language Processing (NLP) and Machine Learning (ML), focuses on identifying and classifying opinions or emotions in text as positive, negative, or neutral [5]. It gives organizations practical insights into how consumers view their products and services. In Nigeria's growing digital economy, this ability is particularly helpful for brand monitoring. It allows businesses to track public opinion, assess marketing campaign success, and improve customer experiences [6]. By catching negative sentiments early, companies can solve problems proactively, while positive feedback can help build consumer trust and brand loyalty.

Brand monitoring has become a crucial strategy for companies wanting to keep a good market image and customer trust. Ongoing analysis of online sentiments provides immediate insights into consumer views. This helps businesses identify potential problems, evaluate campaign success, and adjust their communication strategies. In Nigeria, many companies have already started using sentiment analysis tools to assess service quality and public interaction. For example, telecommunications firms use social media sentiment tracking to find network complaints and improve customer satisfaction, while retail companies monitor product reviews to understand changing consumer preferences. This proactive approach to brand intelligence strengthens customer relationships, improves service quality, and fosters long-term business growth [7].

Despite these advantages, various challenges limit the effectiveness of sentiment analysis in multilingual and culturally diverse places like Nigeria. The different languages—like Yoruba, Igbo, Hausa, and Nigerian Pidgin—along with slang, abbreviations, and switching between languages, make sentiment interpretation difficult [8]. Current sentiment analysis models trained mostly on English-language data often struggle with local expressions and regional dialects. As Salawu et al. [9] pointed out, we need culturally adapted and linguistically aware algorithms to achieve accurate sentiment classification in Nigeria.

Recent improvements in deep learning have made sentiment analysis more effective with models like BERT (Bidirectional Encoder Representations from Transformers), FastText, and GPT (Generative Pre-trained Transformer). These can understand contextual meaning and language subtleties more accurately [10], but many of these studies are limited to Western languages and contexts; therefore, their findings may not be directly applicable to Nigeria [2]. This is because most of the models are highly dependent on pre-trained language models that do not well represent the linguistic features common for Nigerian linguistic purposes-a reason behind the poor generalization and applicability to local brands. Without tools that cater to Nigerian social media, it puts brands at a disadvantage in terms of online reputation management and responding effectively to public sentiment. This study aims to fill these gaps by developing a machine-learning model, adapted to the Nigerian social media landscape that can effectively carry out brand sentiment analysis. This study further seeks to enhance the accuracy of the detection of sentiment by adapting machine learning models to the use of local dialects and informal expressions in ways that better help Nigerian brands. Among these models, FastText is efficient, scalable, and performs well, especially in low-resource and multilingual environments. Incorporating these models into sentiment analysis systems enables businesses to process large amounts of social media data quickly, supporting more responsive and data-driven brand management.

Therefore, the aim of this study is to design and implement a FastText and custom based sentiment analysis model for brand monitoring in Nigeria. The specific objectives are to:

- conduct an extensive study on the different algorithms suitable for sentiment analysis,
- design and implement a sentiment analysis model, and
- evaluate and test the performance of the model.

2. Literature Review

2.1. Introduction to Sentiment Analysis

Sentiment analysis, also called opinion mining, is a key area of natural language processing (NLP) and machine learning. It focuses on finding, extracting, and classifying opinions in text to identify their emotional tone—positive, negative, or neutral [11] - [12]. By converting large amounts of unstructured text into useful insights, sentiment analysis helps organizations understand public perception, consumer attitudes, and brand reputation in real time [6]. The rapid rise of social media and online platforms makes sentiment analysis essential for tracking consumer feedback, improving marketing strategies, and boosting customer engagement [13].

Early sentiment analysis methods mainly used lexicon-based approaches, which involved matching words to predefined sentiment dictionaries [11], [14]. Although these methods were easy to understand and efficient, they often faced issues with complex language features like sarcasm, irony, and context variation. The arrival of machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, and Logistic Regression represented a major shift, allowing models to learn sentiment patterns from labeled data [12]. Recently, deep learning frameworks—including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models like BERT—have greatly enhanced sentiment classification accuracy through improved understanding of context [10], [15] - [16].

Sentiment analysis can be performed at different levels: document-level, sentence-level, and aspect-based sentiment analysis (ABSA). Document-level methods provide an overall sentiment direction, while ABSA allows a closer look at opinions about specific features of products or services, making it especially useful for brand monitoring and targeted marketing [17]. Token-level and phrase-level analyses further refine results by capturing idioms, sarcasm, and informal language that are common in online communication [18].

In Nigeria, sentiment analysis is becoming increasingly important due to the country’s growing digital economy, high social media activity, and rich language diversity. Millions of Nigerians express their opinions in English, Pidgin, Yoruba, Hausa, and Igbo on platforms like Twitter, Facebook, and Instagram [19]. For businesses, studying these multilingual sentiments offers vital insights into consumer behavior, market trends, and brand image. It also helps policymakers understand public opinion and address social issues effectively [20]. However, challenges remain due to a lack of NLP resources for local languages, language mixing, and cultural expressions specific to Nigerian discussions.

The concepts behind sentiment analysis draw from linguistics, psychology, and computer science, highlighting how emotions and subjective meanings are expressed through language [21]. Even with significant advancements, obstacles such as detecting sarcasm, ambiguous phrasing, and managing multilingual data still affect accuracy, especially in linguistically diverse settings [18]. Working to close these gaps through the use of pre-trained embeddings like FastText and domain-specific models shows promise for improving performance in low-resource languages.

2.2. Techniques and Models in Sentiment Analysis

Sentiment analysis, or opinion mining, uses computer techniques to find and understand subjective information in text. It helps to gauge public attitudes, consumer opinions, and brand perceptions. Over time, different methods, from lexicon-based systems to deep learning models, have been created to improve classification accuracy and context understanding.

Lexicon-based methods rely on predefined dictionaries that assign sentiment scores to words or phrases. They calculate overall sentiment by adding these values together. These methods are easy to understand but often have trouble with language complexities like sarcasm, idioms, and changes in context, which can limit their accuracy [11], [14]. Machine learning techniques, including Support Vector Machines, Naive Bayes, and Random Forests, learn to recognize sentiment patterns from labeled datasets [12], [22]. These models use feature extraction methods like TF-IDF, Word2Vec, and GloVe to transform text into numbers. Machine learning approaches do better than lexicon-based models in flexibility but need large, high-quality data for effective training. Deep learning models, such as RNNs, LSTMs, and BiLSTMs, capture the sequence and context of information in text [15]. Transformer-based models like BERT and RoBERTa use self-attention mechanisms to understand complex language relationships. They achieve top performance in sentiment analysis, especially on social media [10], [16]. Hybrid models combine lexicon features with machine learning or deep learning algorithms to merge interpretability and adaptability. This improves performance in specific domains or noisy data settings. Good sentiment analysis systems also need thorough preprocessing to handle slang, informal text, and emojis, along with ethical measures to address bias and privacy [23], [24].

Accurate word representation is key to sentiment analysis. Early methods like Bag-of-Words and TF-IDF treat text as frequency vectors but overlook meaning [12], [25]. Lexicon-based approaches assign polarity scores [11], while distributed embeddings such as Word2Vec and GloVe capture meaning in dense vector spaces [26]. However, these static embeddings do not adapt to word context. Transformer-based models, including BERT and RoBERTa, create dynamic, context-sensitive embeddings that better capture subtle sentiments like irony and negation [10], [25]. Recent hybrid approaches mix lexicon features with contextual embeddings or fine-tune pretrained models like FastText for specific analysis [27], [28].

Despite advancements, challenges in data labeling, bias reduction, and multilingual adaptation persist. Future research aims to create efficient, culturally aware embeddings and multimodal sentiment models that integrate text, audio, and visual cues for richer sentiment understanding [24].

2.3. Custom Models and Hybrid Architectures in Sentiment Analysis

The growing complexity of social media data has led to the creation of custom and hybrid sentiment analysis models that improve contextual understanding and focus on specific domains. Generic models like FastText are efficient but often have trouble with local slang, sarcasm, and mixed-language expressions. This gap motivates the development of tailored approaches [6], [28]. These models focus on capturing subtle contextual clues, adjusting to industry-specific language, and improving scalability across changing datasets [29]. Custom sentiment analysis models usually use deep learning architectures such as BiLSTM, CNN, and transformer-based frameworks. They often include contextualized embeddings from models like BERT or RoBERTa [30]. A typical design combines embedding layers, convolutional filters, and recurrent layers to capture both local and sequential features. Attention mechanisms highlight words that convey sentiment, making results easier to interpret [31]. Hybrid models often mix complementary components, like CNN with BiLSTM or Transformer with RNN, to balance local feature extraction and global context modeling. Ensemble methods that combine outputs from FastText, SVM, and neural networks improve robustness against noisy data [32].

Comparative studies show that while FastText offers solid baselines and computational efficiency [33], [34], custom and hybrid models are better at detecting nuanced sentiment. Architectures that blend BiLSTM and transformer layers achieve gains of 3 to 5% in accuracy and F1-score due to better contextual understanding [28], [30]. Additionally, fine-tuning for specific domains allows for better adaptability in brand monitoring and crisis prediction [35]. Hybrid approaches that mix FastText embeddings with BiLSTM or attention mechanisms show improved recall and precision [36], while ensemble strategies that combine FastText with traditional classifiers like logistic regression or SVM demonstrate greater resilience to noise [37].

2.4. Fasttext-Based Sentiment Analysis Models

FastText-based sentiment analysis models have become effective tools in natural language processing (NLP). They work particularly well for analyzing sentiments in social media and multilingual contexts. Developed by Facebook AI Research, FastText improves traditional word embedding techniques by including sub word information and a lightweight neural architecture. This design efficiently handles text that is rich in morphology and noisy [33]. It is especially suitable for areas that use slang, abbreviations, and misspellings often found in user-generated content.

FastText builds on Word2Vec by representing words as bags of character n-grams. This allows the model to learn subword structures like prefixes, suffixes, and roots [33]. For example, the word “sentiment” can be broken down into n-grams like “sen”, “ent”, “tim”, and “men.” These subword representations let the model create embeddings for unseen words, which is a significant advantage in informal or changing language contexts [38]. FastText supports both Continuous Bag-of-Words (CBOW) and Skip-Gram architectures. It uses subword n-gram vectors to improve semantic understanding while keeping high computational efficiency.

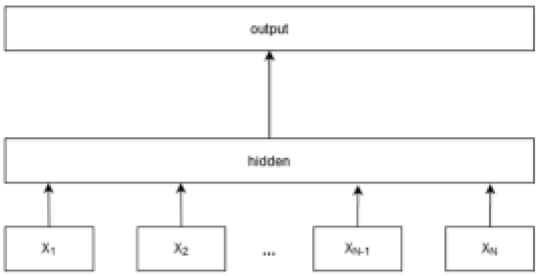


Figure 1: FastText Model Architecture [39]

In sentiment analysis applications, FastText shows strong performance due to its fast-training speed, scalability, and ability to handle multilingual and informal data [40]. It is often used to create embeddings for classification algorithms such as logistic regression, SVM, or its built-in supervised classifier. Studies show it competes well with deep learning models like LSTMs and transformers, achieving high accuracy at much lower training costs. Its ability to generalize across unseen words and fit emerging domain-specific vocabularies makes it excellent for brand monitoring and real-time opinion tracking [28].

However, FastText has limitations due to its static embedding approach. It assigns a single vector to each word, regardless of context [25]. This limit affects its ability to deal with polysemy, sarcasm, and subtle sentiment changes [30]. Unlike transformer-based models such as BERT or RoBERTa, which create context-dependent embeddings [10], FastText cannot model long-range dependencies or capture bidirectional contextual cues. Additionally, challenges in preprocessing—like managing emojis, hashtags, and code-mixed text—can affect its performance without special data cleaning [6].

2.5. Brand Monitoring Using Sentiment Analysis

Brand monitoring through sentiment analysis is now a key part of modern brand management. It helps organizations constantly evaluate how consumers feel and think across various digital platforms. This method uses Natural Language Processing (NLP) and Machine Learning (ML) to categorize text-based opinions as positive, negative, or neutral. This process quantifies the emotional tone found in consumer communications [28]. The true power of sentiment analysis lies in revealing the underlying attitudes behind what consumers are saying, not just identifying the words used. By delivering real-time and useful insights, sentiment analysis allows organizations to spot emerging market trends, assess the success of marketing strategies, and foresee potential public relations issues before they grow [6]. Additionally, combining sentiment analysis with other types of data, such as images, videos, and multilingual text, improves analytical accuracy. This approach leads to a better understanding of consumer behavior and supports data-driven brand management strategies [28]. In fast-changing markets like Nigeria, where digital interactions are culturally and linguistically rich, sentiment analysis gives brands a valuable way to stay in tune with public sentiment.

Brand monitoring systems that use sentiment analysis usually follow a clear method that includes data collection, feature extraction, model implementation, and real-time analysis. Data is collected from platforms like Twitter, Facebook, and review sites such as Amazon and Yelp using web-scraping tools like Scrapy and Beautiful Soup. Preprocessing includes normalizing text, removing special characters and stop words, and lemmatization to ensure clean and consistent input. Dealing with informal language, including slang, local expressions, and emojis, is crucial, especially in culturally diverse places like Nigeria, where language variety can greatly impact classification accuracy [30]. To ensure precise sentiment classification, linguistic features are extracted using vectorization and embedding methods. Traditional techniques like TF-IDF and Word2Vec capture basic semantic patterns. In contrast, GloVe and transformer-based embeddings provide richer contextual information for a deeper sentiment interpretation [41].

Machine learning models such as Logistic Regression and Support Vector Machines (SVM) set basic performance standards. However, modern systems increasingly use deep learning and transformer architectures, including BERT, RoBERTa, and DistilBERT, to better capture contextual sentiment, sarcasm, and emotional depth [28], [30]. Hybrid methods, like BiLSTM combined with pre-trained embeddings, show high effectiveness, achieving accuracies up to 89% and AUC scores near 96% [32]. Advanced systems integrate these models into automated systems that analyze continuous social media streams. Real-time monitoring allows brands to quickly detect sentiment changes, manage crises, and adjust their communication strategies as needed [6].

As brand monitoring grows, following data protection laws and handling data ethically is essential. Techniques like data anonymization and bias reduction are vital for ensuring fairness and reliability across different demographic and language groups. Tackling challenges like sarcasm detection and context misinterpretation remains an ongoing effort in developing sentiment analysis systems [28], [32].

Several case studies and earlier research have shown how sentiment analysis affects brand monitoring. These studies reveal how organizations in various industries use sentiment analysis to improve customer engagement, refine marketing strategies, and strengthen brand reputation management. A notable case is Samsung’s use of sentiment analysis, which allowed the company to track real-time consumer feedback and adjust its marketing and product development based on that feedback. This proactive approach significantly boosted customer satisfaction and strengthened Samsung’s competitive edge in the global technology market [35]. In another study, combining Convolutional Neural Networks (CNNs) with sentiment analysis showed clear improvements in accuracy and depth of understanding for tracking brand awareness on social media. The ability of CNN-based models to handle different types of data makes them particularly useful for analyzing large-scale interactions across multiple platforms [28].

Comparative studies between traditional machine learning models and transformer-based architectures have shown major performance improvements. Specifically, transformer models like RoBERTa and DistilBERT provided around a 10 to 11% boost in F1 scores because of their superior understanding of context and language [32]. Despite requiring more computational power, these architectures perform well in real-time sentiment classification, making them suitable for fast-paced brand monitoring.

Equally, using Bidirectional Long Short-Term Memory (BiLSTM) networks with GloVe embeddings has set new standards for accuracy in sentiment analysis. These models can detect subtle emotional signals and context that simpler models often miss, increasing the reliability of sentiment detection across different languages [32].

Beyond technical advances, researchers like Chowdhury [6] have stressed the need for best practices when using sentiment analysis in brand management. Key practices include defining clear goals for analysis, choosing models that match company aims, ensuring data quality, adapting to specific domains, and maintaining transparency with data handling and stakeholder communication.

Overall, these studies highlight the growing importance of deep learning and transformer-based models in achieving greater accuracy and sensitivity in sentiment classification. They also point to the need for models that can adapt to cultural and linguistic differences, particularly in multilingual settings like Nigeria.

2.6. Sentiment Analysis in the Nigerian Context

Sentiment analysis, a crucial area of Natural Language Processing (NLP), aims to identify and extract subjective information like emotions, opinions, and attitudes from text. In Nigeria, where social media use and digital communication are growing quickly, sentiment analysis offers significant opportunities to understand public perception and support data-driven decisions. However, the country's linguistic variety, cultural complexity, and unique sociolinguistic expressions present specific challenges for effective sentiment analysis [42].

Nigeria's multilingual environment, which includes over 500 indigenous languages such as Hausa, Yoruba, Igbo, and Fulfulde, adds complexity to text-based analysis. Most NLP models focus on global languages like English or Mandarin, making them less effective in mixed or under-resourced language settings [42]. Several challenges have been identified. Multilingualism and Code-Switching. Nigerians often mix English, indigenous languages, and Pidgin in single posts or sentences. This blending complicates tokenization, parsing, and sentiment classification [43]. Standard NLP models, trained on monolingual text, usually struggle to interpret mixed-language content accurately. Also, online communication in Nigeria is rich with slang, abbreviations, and localized phrases, such as "wahala" (trouble) or "jare" (please). These carry strong emotional and cultural meanings but often do not appear in standard dictionaries [44]. Additionally, sarcasm, irony, and context-specific phrases are common in Nigerian communication. This makes it harder to detect sentiment polarity. Generic models often misinterpret these nuances, leading to mistakes in classification. Finally, there is the scarcity of Labeled Data: There is a shortage of annotated sentiment datasets for Nigerian languages. This limits the development of accurate supervised learning models.

These challenges highlight the need for sentiment analysis models that understand context and adapt to Nigeria's diverse ways of communicating. Several researchers have worked to overcome these challenges with new modeling and adaptation techniques. Ezeani and Onyechi [42] created a sentiment classifier for Nigerian Pidgin English, showing that using a localized lexicon can significantly boost classification accuracy. Adelani et al. [45] built on this by adding Hausa and Yoruba to a multilingual sentiment analysis model. This approach achieved better results when it included language identification and customized processing. Similarly, Asefon et al. [44] proposed a hybrid model that combined lexicon expansion with machine learning, effectively capturing sentiment in Nigerian slang. Other studies have stressed the importance of context-sensitive models that can recognize sarcasm. These models are essential for understanding the nuances of Nigerian social media language [46]. Despite these advancements, much of the current work still relies on traditional models like Support Vector Machines (SVM) and Logistic Regression, as well as standard word embeddings. This reliance limits their ability to adapt to Nigeria's linguistic diversity and new language trends.

Given the shortcomings of traditional NLP models, FastText and custom neural architectures offer promising options for enhancing sentiment analysis in Nigeria's multilingual setting. FastText, developed by Facebook AI Research, is particularly suited for languages with rich morphology and limited resources. It represents words using subword information through character n-grams, which helps it generalize to words it has not seen before or that may be misspelled [47].

FastText for Nigerian Languages is advantageous, because of Subword Information Capture. FastText handles Nigeria's morphological diversity and informal language well, accommodating new words and slang [33]. Additionally, it supports multilingual embeddings, enabling effective analysis of code-switched text common on Nigerian social media. Also, FastText requires less computing power compared to transformer-based models. This makes it suitable for academic and commercial applications with limited resources. Moreover, while FastText provides strong foundational embeddings, its best performance comes from custom models trained on Nigerian datasets that reflect local language realities. This fine-tuning can improve contextual understanding, resolve confusion with slang and sarcasm, and better deal with code-switching. Finally, as computing becomes more accessible, combining custom transformer-based architectures (like fine-tuned BERT variants) with FastText embeddings can deliver improved results. This hybrid approach merges FastText's ability to handle subword variations with the contextual accuracy of deep learning models, paving the way for more precise and culturally relevant sentiment analysis tailored to Nigeria's linguistic landscape.

3. Methodology

This study used a structured process that included system analysis, model design, implementation, and evaluation. The approach tackles major challenges in Nigerian social media data, such as multilingualism, code-switching, slang, and limited labeled data. It also maintains computational efficiency and real-time usability. The system was developed through a data-driven and iterative design process, following Agile principles to ensure flexibility and ongoing performance improvement.

3.1. Research Design

The research uses a quantitative experimental design. It focuses on developing, training, and validating a sentiment analysis model that can classify public sentiment about brands on Nigerian social media. This design allows for a statistical evaluation of model performance using measurable metrics like accuracy, precision, recall, and F1-score. The process involves five major stages:

- Data Collection: This stage extracts brand-related posts and comments from major social media platforms, including Twitter, Facebook, and Instagram, using APIs.
- Pre-processing – This stage involves cleaning, normalizing, and tokenizing the data to address noise, code-switching, and slangs.
- Model Development – This stage consists of training both FastText and a custom neural model to categorize sentiments as positive, negative, or neutral.
- Model Evaluation – This stage tests the model using labeled datasets and compares the results with baseline algorithms like SVM and Naive Bayes.
- System Deployment – This final stage implements a user-friendly platform for real-time analysis and visualization

3.2. Model Development and Evaluation

Brand-related data were gathered from publicly available social media platforms and online forums widely used in Nigeria. This included posts and comments linked to well-known local and international brands. Only publicly accessible text was used to follow ethical guidelines. The data were processed with natural language processing (NLP) techniques to eliminate irrelevant items like URLs and non-text elements. Additional cleaning steps included normalization, lemmatization, and the removal of stopwords. Due to the common use of code-switching and Pidgin English, a custom Nigerian lexicon was created to improve the system's understanding of local expressions. Techniques to balance the data were also used to ensure fair representation across sentiment categories.

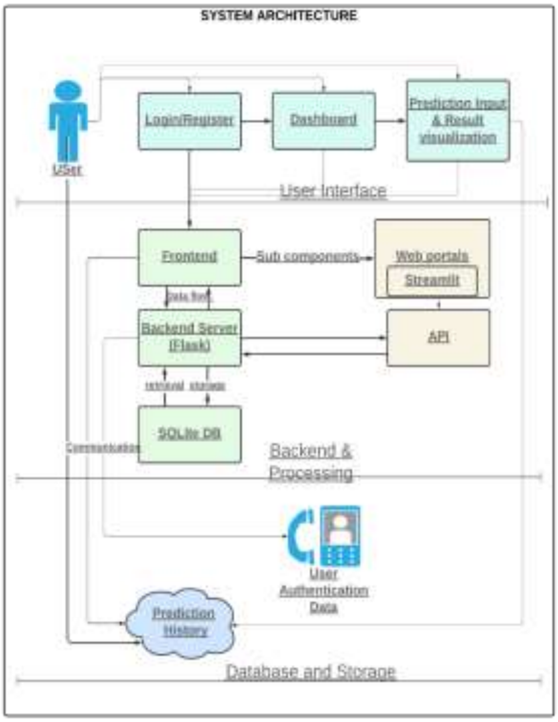


Figure 2: System Architecture

The sentiment analysis model combines FastText word embeddings with a custom deep learning network that tackles the linguistic and morphological variations of Nigerian English and local languages. FastText offers subword-level representation, allowing for better understanding of slang and informal language. The custom model features multiple dense layers and activation functions optimized for sentiment classification. The hybrid framework ensures a balance between understanding meaning and computational efficiency. The model was trained and validated using separate data subsets, and methods like early stopping and dropout were used to improve generalization. The training environment relied on Python-based deep learning libraries within a controlled experimental setup.

The developed model was assessed using standard performance metrics to measure its accuracy, robustness, and reliability. These metrics included measures of classification accuracy and the balance between positive and negative predictions. The evaluation also involved comparative tests against baseline machine learning algorithms such as logistic regression and support vector machines to identify improvements in classification quality. Performance evaluation highlighted the model's ability to interpret sentiment across a range of linguistic expressions and maintain stability under different text conditions. Cross-validation was performed to ensure consistency and reduce overfitting.

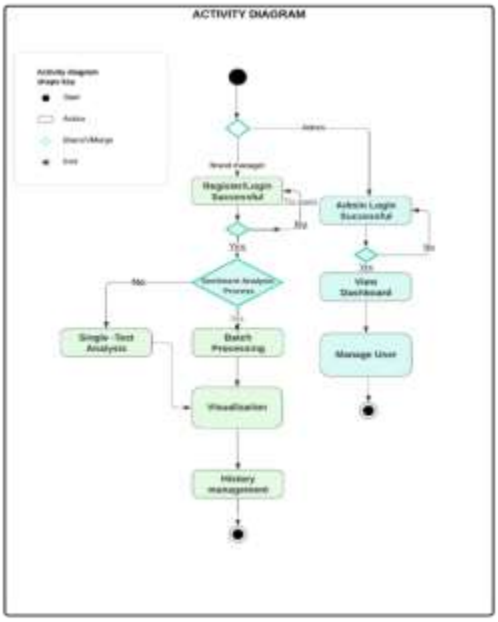


Figure 3: Flowchart of the system

The sentiment analysis system was created as a web application that allows users to perform sentiment classification through a straightforward, interactive interface as shown in figure 2 above. The backend was built using FastAPI and Flask, while Streamlit supported user interaction and visualization. The platform accommodates both single-text input and batch processing of uploaded files. Results are displayed through dynamic visualizations like sentiment distribution charts, along with a user account system that tracks previous analyses for ongoing monitoring. The overall design emphasizes real-time response, modularity, and scalability.

Testing was carried out to verify system reliability, usability, and functional correctness. The system went through several validation cycles that involved simulated user interactions and varied text input conditions. This ensured stable operation under high loads and different linguistic scenarios. Usability testing assessed response time, interface intuitiveness, and user satisfaction. The validation results confirmed that the system could effectively classify and visualize sentiments within realistic operational limits.

The study followed all relevant ethical standards regarding digital data use. Only publicly accessible information was collected, and all data were anonymized to protect user identities. The study adhered to both the Nigeria Data Protection Regulation (NDPR) and the General Data Protection Regulation (GDPR). Efforts to mitigate bias and ensure fairness were also prioritized to make sure the model represented and interpreted all linguistic and cultural groups accurately.

3.3. System Workflow and Operational Design

As shown in figure 3 above, the sentiment analysis system for brand monitoring runs through a step-by-step process that ensures smooth interaction between users and the analytical engine. It starts with user authentication and profile management. Users can register, log in, and manage their profiles by updating personal details like passwords or display information. Once users log in successfully, they can access the system dashboard. Next, the sentiment analysis process begins. Users can pick from two prediction modes: single-text analysis or batch processing for uploaded text or document files. After submitting data, the system preprocesses the input and lets users choose a model. They can select the FastText model, which is optimized for accuracy and general text sentiment detection, or the custom model, which is specifically designed to handle Nigerian language variations such as Pidgin, Yoruba, Igbo, and Hausa. The chosen model classifies the input into sentiment categories: positive, negative, or neutral. After classification, the visualization and history management module shows the results through interactive graphs, such as bar or pie charts. These help users interpret sentiment trends easily. The analysis results from each session are stored for future review and trend tracking. Finally, users can either continue with more analyses or exit the system.

3.4. Development Tools

The creation of the FastText and custom sentiment analysis system for brand monitoring in Nigeria was supported by a collection of programming tools, libraries, and platforms aimed at efficient data processing, modeling, and visualization.

- Programming Language: Python was the main language used for the sentiment analysis pipeline, data preprocessing, and backend integration. Its extensive ecosystem of natural language processing (NLP) and machine learning libraries makes it a great choice for scalable model development.
- Machine Learning and NLP Libraries: The Natural Language Toolkit (NLTK) was used for key text preprocessing tasks like tokenization, stop-word removal, and lemmatization. The Scikit-learn library supported machine learning tasks, including baseline classifiers like Logistic Regression, Support Vector Machines (SVM), and Naive Bayes, allowing comparisons with FastText-based models.
- Datasets: Two primary datasets were used. The Sentiment140 repository, which contains labeled Twitter data, provided a solid foundation for general sentiment classification. The NaijaSenti dataset was also included to capture linguistic and cultural details specific to Nigerian social media, including expressions in Pidgin, Yoruba, Igbo, and Hausa.
- Visualization Tools: Tableau was used to turn sentiment outputs into interactive charts and dashboards, allowing users to easily explore sentiment trends and brand perception over time.
- Version Control: Git and GitHub helped with version management, teamwork, and code integrity throughout the system development.
- Integrated Development Environment: Implementation and testing took place using PyCharm and Visual Studio Code (VS Code), which offered integrated tools for debugging, optimizing performance, and organizing projects.

This development framework provided a strong, clear, and flexible environment for building and improving the sentiment analysis model.

4. Results

Table 1 below shows the performance of the FastText model compared to other sentiment analysis models. The results indicate that the custom-trained FastText model achieved the highest performance across all evaluation metrics. The model's accuracy of 90.8% and F1-score of 90.3% were much higher than those of traditional machine learning algorithms. FastText performed better because of its subword-level representation, which effectively captured language variations, hashtags, and Nigerian Pidgin spellings, such as "wahala," "dey," and "jollof." This made it particularly suitable for the informal and mixed language nature of social media brand discussions.

In contrast, Naïve Bayes and Logistic Regression models performed well on structured English text from the Sentiment140 subset. However, they struggled with noisy, multilingual expressions found in the NaijaSenti and Twitter datasets. SVM achieved competitive results with 84.7% accuracy because it can handle high-dimensional feature spaces, but it needed a lot of feature engineering and tuning. The Random Forest classifier reached moderate accuracy at 83.5%. It benefited from non-linear feature learning but showed signs of overfitting during cross-validation, especially when dealing with short posts that provided limited context.

Table 1. Comparative Performance of Sentiment Analysis Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (min)	Inference Time

						(ms/sample)
Naïve Bayes (MultinomialNB)	81.9	81.2	80.6	80.8	2.1	1
Logistic Regression (TF-IDF)	84.3	83.5	83.1	83.3	5.2	3.1
SVM (Linear Kernel)	85.1	84.7	84	84.2	7.4	3.8
Random Forest	82.6	81.4	80.9	81.1	20.3	5.9
FastText (Custom)	87.8	87.2	86.5	86.8	4	1.4

Figures 4 to 8 present the results of the study. The system was built to provide user authentication, profile management, sentiment analysis predictions, and visualization of results. Additionally, it supports both individual and batch predictions, catering specifically to Nigerian users by analyzing tribalist language sentiment.

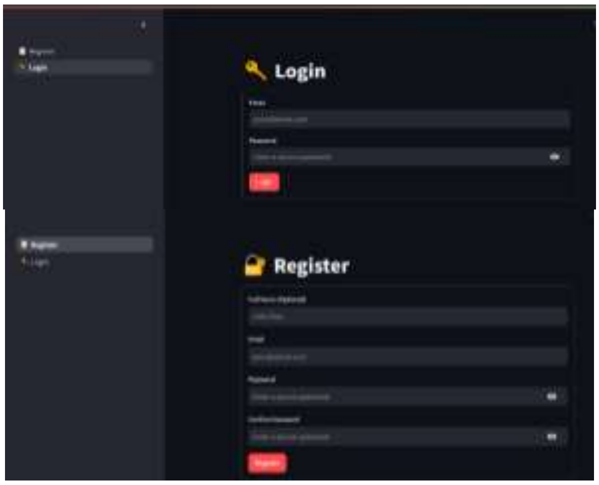


Figure 4: Login and Registration Modules

Users can register and log in to the platform using secure authentication as shown in figure 4 above. The system provides sentiment analysis predictions with two key functionalities as shown in figure 5 below:

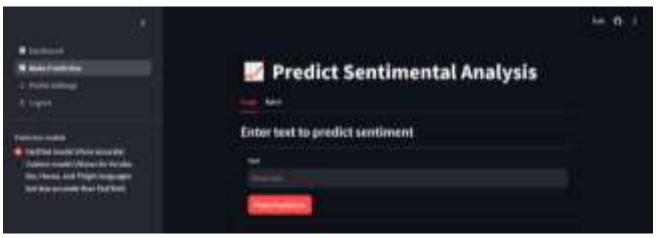
- Single Prediction: Users can input a statement and receive a sentiment prediction instantly.
- Batch Predictions: Users can upload a text file or PDF to analyze multiple statements at once.

Two modules are available for predictions as shown in figure 6 below:

- FastText Model: More accurate but limited to general sentiment analysis.
- Custom Model: Supports Yoruba, Igbo, Hausa, and Pidgin languages but is slightly less accurate than FastText.

To help users interpret the results, predictions are presented in graphical formats as shown in figure 7 below:

- Bar Chart: Displays the sentiment distribution across analyzed data.
- Pie Chart: Provides a percentage breakdown of sentiments.



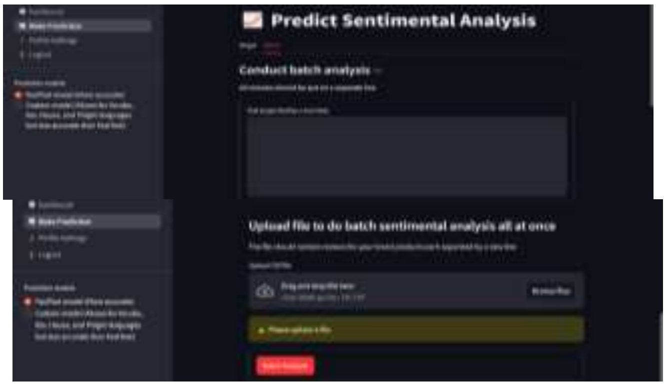


Figure 5: Sentiment Analysis Module.

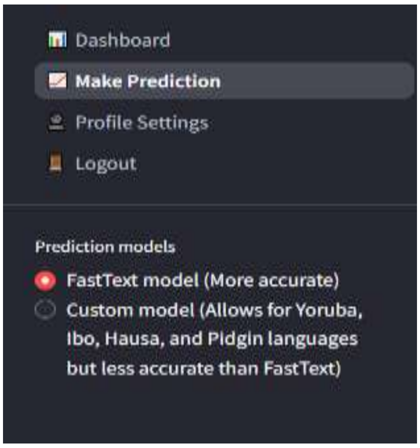


Figure 6: FastText and Custom Model Options

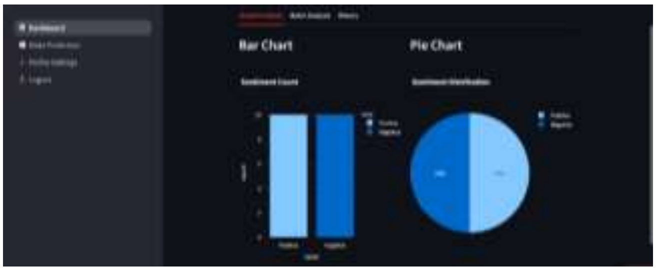


Figure 7: Visualization Module.

Users can also view past predictions, including batch analysis history as shown in figure 8 below.

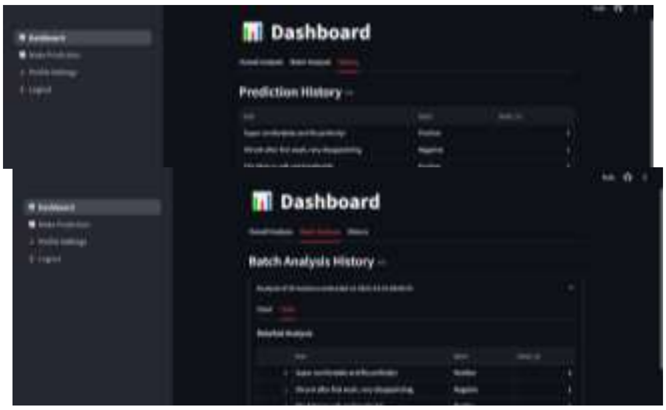


Figure 8: Prediction and Batch Prediction History

The system was deployed using FastAPI, Flask, or Streamlit, ensuring smooth real-time interaction and ease of access for users.

5. Conclusion

The study successfully developed a sentiment analysis system capable of analyzing text-based content with a focus on Nigerian languages (Yoruba, Igbo, Hausa, and Pidgin) while leveraging machine learning techniques. The system provides accurate sentiment predictions, batch processing capabilities, and visualization tools for better data interpretation. It demonstrates the effectiveness of machine learning in processing and classifying text-based sentiments. By integrating FastText for accuracy and a custom-trained model for Nigerian languages, the system ensures a localized and adaptable approach to sentiment classification.

By using subword embeddings, the model captures sentiment from posts that include misspellings, abbreviations, and brand-specific hashtags. The improved performance suggests that brands and marketing analysts can trust the system for precise classification of public opinion into positive, negative, or neutral categories.

The results show that the FastText-based sentiment model is more dependable for real-time brand monitoring in multilingual and informal communication contexts, which are common on Nigerian social media. This study highlights the potential of machine learning in sentiment analysis, especially in tackling the challenges of analyzing local Nigerian languages. With further improvements, this system could be used widely for different purposes in business, research, and public opinion tracking. The insights gained from this study contribute to the growing field of AI-driven text analysis and provide a basis for additional research and development in localized language processing and sentiment analytics

References

[1] Statista, “Number of social media users worldwide from 2010 to 2023,” 2023. [Online].

Available: https://www.statista.com. [Accessed: Sept. 10, 2025].

[2] Y. Zhang, S. Li, and Y. Yang, “Social media-based brand sentiment analysis: Trends and research challenges,” Journal of Marketing Analytics, vol. 9, no. 1, pp. 15–32, 2021.

[3] C. Okonkwo, E. Ihedioha, and G. Mba, “Social media analytics in Nigeria’s business sector: Benefits and challenges,” Nigerian Journal of Management Studies, vol. 15, no. 2, pp. 50–62, 2020.

[4] D. O. Alao and O. Awodele, “A systematic review of social media usage in Nigeria: The implications for brands and consumers,” Journal of Business and Social Review in Emerging Economies, vol. 6, no. 1, pp. 67–78, 2020.

[5] K. S. Adewole, F. A. Olayinka, and O. O. Ayodeji, “Natural language processing for sentiment analysis in Nigerian languages: Challenges and opportunities,” African Journal of Computing & ICT, vol. 12, no. 1, pp. 34–42, 2019.

[6] R. H. Chowdhury, “Sentiment analysis and social media analytics in brand management: Techniques, trends, and implications,” World Journal of Advanced Research and Reviews, vol. 23, no. 2, pp. 287–296, 2024.

[7] U. Ramanathan, N. Subramanian, and G. Parrott, “Role of social media in retail network operations and marketing to enhance customer satisfaction,” International Journal of Operations & Production Management, vol. 37, no. 1, pp. 105–123, 2017.

[8] A. Eke, E. Ugochukwu, and O. Chukwudi, “Multilingual sentiment analysis: A case study in Nigeria,” Journal of Natural Language Processing Research, vol. 2, no. 3, pp. 180–194, 2021.

[9] S. Salawu, S. Adepoju, and F. Oladimeji, “Dialect adaptation for social media sentiment analysis in Nigerian English,” Journal of Computational Linguistics & Speech Technology,

vol. 8, no. 2, pp. 224–235, 2022.

[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. Conf. North American Chapter Assoc. Comput. Linguistics: Human Language Technologies (NAACL-HLT), 2019, pp. 4171–4186.

[11] B. Liu, Sentiment Analysis and Opinion Mining. Springer Nature, 2022.

[12] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends® in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[13] R. B. Malabadi, S. S. Mammadova, K. P. Kolkar, S. M. R., R. K. Chalannavar, and K. V. C. Coronado, “Cannabis sativa: A therapeutic medicinal plant—Global marketing updates,” World Journal of Biology Pharmacy and Health Sciences, vol. 17, no. 2, pp. 170–183, 2024.

[14] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[15] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proc. 2013 Conf. Empirical Methods in Natural Language Processing, pp. 1631–1642, 2013.

[16] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.

[17] M. Pontiki et al., “SemEval-2016 task 5: Aspect based sentiment analysis,” in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 19–30.

[18] A. Joshi, P. Bhattacharyya, and M. J. Carman, “Automatic sarcasm detection: A survey,” ACM Computing Surveys (CSUR), vol. 50, no. 5, pp. 1–22, 2017.

[19] O. Olaitan, The impact of social media (Twitter, Facebook and Instagram) on consumer behaviour in Nigeria, Ph.D. dissertation, Dublin Business School, 2021.

[20] J. C. Okafor, C. C. Nweke, O. J. Igweike, and O. I. Eze, "A critical review of the role of social media in citizen engagement and policymaking: The Nigeria discourse," Nigerian Journal of Sustainability Research, vol. 2, no. 1, pp. 17–28, 2024.

[21] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, “Affective computing and sentiment analysis,” in A Practical Guide to Sentiment Analysis, Cham, Switzerland: Springer International Publishing, 2017, pp. 1–10.

[22] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, “Tackling the poor assumptions of Naive Bayes text classifiers,” in Proceedings of the 20th International Conference on Machine Learning (ICML-03), 2003, pp. 616–623.

[23] K. Crawford and J. Schultz, “Big data and due process: Toward a framework to redress predictive privacy harms,” Boston College Law Review, vol. 55, no. 1, pp. 93–128, 2014.

[24] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A survey on bias and fairness in machine learning,” ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1–35, 2021.

[25] G. Bandarupalli, “Enhancing sentiment analysis in multilingual social media data using transformer-based NLP models: A synthetic computational study,” Authorea Preprints, 2025.

[26] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Advances in Neural Information Processing Systems, vol. 26, 2013.

[27] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, no. 7, pp. 5731–5780, 2022.

[28] N. Sureja, N. Chaudhari, P. Patel, J. Bhatt, T. Desai, and V. Parikh, "Hyper-tuned swarm intelligence machine learning-based sentiment analysis of social media," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 15415–15421, 2024.

[29] Y. Siri, S. Afroz, and R. U. Rani, "Enhancing sentiment analysis accuracy by optimizing hyperparameters of SVM and logistic regression models," in E3S Web of Conferences, vol. 472, p. 01017, EDP Sciences, 2024.

[30] A. Sharma, N. Patel, and R. Gupta, "Leveraging BERT and sentiment analysis algorithms for enhanced AI-driven brand sentiment monitoring," European Advanced AI Journal, vol. 11, no. 8, 2022.

[31] A. T. Nurlanuly, "Sentiment analysis of texts from social networks based on machine learning methods for monitoring public sentiment," arXiv preprint, 2025. [Online]. Available: https://arxiv.org/abs/2502.17143.

[32] H. H. Nguyen, "Enhancing sentiment analysis on social media data with advanced deep learning techniques," International Journal of Advanced Computer Science & Applications, vol. 15, no. 5, 2024.

[33] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in Proc. 15th Conf. of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017, pp. 427–431.

[34] M. A. Al Montaser, B. P. Ghosh, A. Barua, F. Karim, B. C. Das, R. E. R. Shawon, and M. S. R. Chowdhury, "Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA," Edelweiss Applied Science and Technology, vol. 9, no. 1, pp. 545–565, 2025.

[35] P. Broklyn, A. Olukemi, and C. Bell, "Social media sentiment analysis for brand reputation management," SSRN, [30]. [Online]. Available: https://doi.org/10.2139/ssrn.4906218.

[36] N. Garg and K. Sharma, "Text pre-processing of multilingual for sentiment analysis based on social network data," International Journal of Electrical & Computer Engineering, vol. 12, no. 1, pp. 1–8, 2022.

[37[G. K. Sidiropoulos, N. Diamianos, K. D. Apostolidis, and G. A. Papakostas, "Text classification using intuitionistic fuzzy set measures—An evaluation study," Information, vol. 13, no. 5, p. 235, 2022.

[38] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," in Proc. Int. Conf. Language Resources and Evaluation (LREC 2018), 2018, pp. 52–55.

[39] A. Talun, P. Drozda, L. Bukowski, and R. Scherer, "FastText and XGBoost content-based classification for employment web scraping," in Proc. Int. Conf. Artificial Intelligence and Soft Computing, Cham, 2020, pp. 435–444.

[40] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based methods for sentiment analysis,” Comput. Linguist., vol. 37, no. 2, pp. 267–307, 2011.

[41] P. Sudhir and V. D. Suresh, “Comparative study of various approaches, applications and classifiers for sentiment analysis,” Global Transitions Proc., vol. 2, no. 2, pp. 205–211, 2021.

[42] W. F. Oyewusi, O. Adekanmbi, and O. Akinsande, “Semantic enrichment of Nigerian Pidgin English for contextual sentiment classification,” arXiv preprint arXiv:2003.12450, 2020.

[43] C. I. Chukwuneke, Named entity recognition for African languages: A focus on Igbo language, Ph.D. dissertation, Lancaster Univ., Lancaster, U.K., 2025.

[44] M. P. Asefon, A. O. Isijola, U. C. Ogude, and S. A. Ntui, "Comparative analysis of machine learning algorithms for sentiment analysis of multilingual Nigerian social media comments," FUOYE Journal of Engineering and Technology, vol. 9, no. 4, pp. 615–623, 2024.

[45] D. I. Adelani, J. Abbott, G. Neubig, D. D’souza, J. Kreutzer, C. Lignos, et al., "MasakhaNER: Named entity recognition for African languages," Transactions of the Association for Computational Linguistics, vol. 9, pp. 1116–1131, 2021.

[46] D. Olaniyan, R. O. Ogundokun, O. P. Bernard, J. Olaniyan, R. Maskeliūnas, and H. B. Akande, "Utilizing an attention-based LSTM model for detecting sarcasm and irony in social media," Computers, vol. 12, no. 11, p. 231, 2023.

[47] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.