Grenze International Journal of Engineering and Technology, January Issue

Sentiment Classification of Tweets using Auxiliary Feature Integration with CNN and BiLSTM Networks

Roma Rakesh Jain[1], Supriya O Rajankar[2] and Madan B Mali[3]
[1]Research Scholar, Savitribai Phule Pune University, Sinhgad College of Engineering, Pune, India
Email: romajain30@gmail.com
[2-3]Sinhgad College of Engineering, Pune, India
Email: sorajankar.scoe@sinhgad.edu, mbmali.scoe@sinhgad.edu

***Abstract*— Nowadays, social networking sites are on the rise, generating a great quantity of data. Millions of individuals share their opinions on microblogging sites every day because they allow for quick and basic phrases. Opinion and sentiment analysis is an important job for identifying subjective information in social media postings. The purpose of sentiment analysis is to identify the views, feelings, and attitudes expressed in source information. Emotional sensations, such as dread, worry, or trauma, are typically the result of early psychological difficulties that may last a lifetime. Furthermore, individuals talk and share their thoughts on social media, frequently inadvertently expressing their secret feelings in the comments. In tweet sentiment analysis, views in tweets are often classified as positive or negative. This research presents a unique deep learning framework for tweet polarity recognition that integrates Auxiliary Feature Integration, Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) networks. The model uses auxiliary features to improve semantic comprehension, CNN to detect local text patterns, and BiLSTM to retain long-range relationships in both directions. Experimental evaluations on benchmark datasets show that the proposed model outperforms current techniques like CNN_GTO, GNN, KG-CNN, and Ensemble_Lexicon-based approaches in terms of accuracy, precision, and F1-score. The findings validate the model's capacity to handle a wide range of linguistic terms, making it ideal for real-world sentiment analysis applications using social media data. The comprehensive trials illustrate the framework's efficiency and indicate that transformer models outperform standard deep learning models. The framework strikes a compromise between accuracy and computing efficiency, making it appropriate for use in actual applications.**

***Index Terms*— Sentiment analysis, Twitter, Convolutional neural network, Enhanced gorilla optimization algorithm, Supervised learning.**

I. INTRODUCTION

Microblogging websites are simply social media platforms where users may publish brief, frequent updates. Twitter is a popular microblogging service that allows users to read and send messages up to 148 characters in length. Twitter communications are also known as Tweets. It has produced a vast dataset of so-called attitudes, which may be classified using supervised learning [1]-[2]. As a result, tweet sentiment analysis has grown rapidly in the field of natural language processing. A common area of study in natural language processing is tweet polarity recognition, which sorts tweets into three groups: positive, negative, and neutral.

Grenze ID: 01.GIJET.12.1.49

As the number of social media sites has grown quickly in recent years, Twitter has become an important tool for studying public opinion, market behavior, and society trends. Researchers and businesses can learn important things from huge amounts of user data by looking at the emotions people show in tweets. But it's hard to tell how someone is feeling from a tweet because they are short, use loose language, and use slang, acronyms, and symbols a lot. Because of this, finding the meaning of tweets has become an important area of study. Deep learning and natural language processing advances have helped make mood classification models more accurate and reliable. Tweet sentiment analysis is being used in a variety of applications, including decision support and recommendation systems. Sentiment classification is the problem of determining if an opinion-laden textual item (e.g., a product review, a blog post, an editorial, etc.) reflects a favorable or negative opinion about a certain entity. The above situation is a basic example of binary classification, using Positive and Negative as classes. When the Neutral class is included, the job becomes a single-label multi-class (SLMC) classification. Emotional sensations, such as dread, worry, or trauma, are typically the result of early psychological difficulties that may last a lifetime. Furthermore, individuals talk and share their thoughts on social media, frequently inadvertently expressing their secret feelings in the comments.

The vast amount of real-time data created on social media sites such as Twitter has made tweet sentiment analysis for polarity identification more important. Users communicate their thoughts, feelings, and responses to a variety of issues, making tweets an important tool for evaluating public mood. Polarity detection (classifying tweets as positive, negative, or neutral) may give insights into fields like as politics, marketing, crisis management, and public health monitoring. The casual language, slang, abbreviations, and usage of emoticons in tweets provide unique issues that need strong natural language processing (NLP) approaches. Effective sentiment analysis allows firms to monitor public opinion trends, increase consumer engagement, and make data-driven choices [3]. As Medhat et al. [4]. point out, sentiment analysis is crucial for understanding public mood and behavior, especially in short-text forms such as tweets, where shortness increases the difficulty of interpretation.

Tweet sentiment analysis for polarity identification presents various issues owing to the particular qualities of Twitter data. Tweets have a restricted amount of characters, so users utilize casual language, acronyms, emoticons, hashtags, and slang, which complicates precise emotion analysis. Furthermore, the existence of sarcasm, irony, and uncertain context complicates polarity categorization [5]. Code-mixing, multilingual material, and the fast development of language on social media all provide substantial challenges to classic natural language processing (NLP) models. Furthermore, an imbalance in sentiment classes—such as an overrepresentation of neutral or positive tweets—can distort model performance. According to Giachanou and Crestani [6]. the linguistic and structural hurdles make tweet sentiment analysis a non-trivial endeavor, requiring sophisticated and adaptable machine learning approaches specialized to short, noisy text forms.

The rest of this paper is organized as follows: Section 1 contains the Introduction, which offers an introduction of tweet polarity categorization, its relevance, and the reason for using a hybrid deep learning technique. Section 2 describes the Literature Survey, which includes a discussion of current methodologies such as lexicon-based, machine learning, and deep learning models, as well as an identification of research gaps that drive the proposed effort. Section 3 describes the methodology, including the architecture of the proposed model, which combines Auxiliary Feature Embedding with CNN and BiLSTM networks to enhance sentiment analysis. Section 4 describes the Experimental Setup and Results, which include a dataset description, performance metrics, and a comparison assessment with current models to illustrate the usefulness of the proposed system. Finally, Section 5 contains the Conclusion, which summarizes the important results and suggests possible avenues for further study.

II. LITERATURE SURVEY

The lexicon-based and machine learning-based approaches to tweet polarity analysis are the two methodologies that are now in use. The vast majority of learning-based strategies are centered on the production of useful characteristics. Tweets are usually short, loud, and very casual, with slang, acronyms, emojis, and hashtags. This makes it hard for natural language processing systems to accurately record how people feel. It can be hard to put tweets into simple categories because they can be full of cultural connections, humor, comedy, or feeling that is expressed subtly. Also, the small number of characters and steady change in online language make rigid mood lexicons less useful. Because of these issues, as well as the fact that there isn't a lot of labeled data and uneven mood classes, recognizing the direction of tweets is hard but important for getting useful information from social media networks. Classifier ensembles and lexicons were used in the approach that was developed by N. Silva and colleagues [7] for the purpose of automatically classifying the sentiment of tweets. On the basis of a certain query term, tweets are categorized as either positive or negative. A Multi-Layer Perceptron classifier was used by E. Xia and colleagues [8] in order to provide an analysis of the sentiments expressed in tweets during the 2020 United States Presidential Election between Donald Trump and Joe Biden. Enhanced Gorilla Troops Optimization Algorithm is used to enhance the performance of a Convolutional Neural Network (CNN), which is a kind of neural network. Li, F et al. [9] outline the use of this neural network system. In their paper [10]. A. Gutiérrez and colleagues provide a sentiment analysis system for social networks that is based on the autonomic cycle of data analysis tasks paradigm. The purpose of this system is to identify innovation challenges that are present inside enterprises. This self-sustaining cycle is comprised of a collection of tasks that are designed to systematically collect and manage enormous volumes of unstructured data from social media platforms. This makes it simpler to detect difficulties related to innovation via the use of sentiment analysis. A framework for efficient hyperparameter tweaking and a one-of-a-kind stemming approach that makes use of word embeddings to calculate the polarity scores of code-mixed tweets were proposed by Y. Aliyu and colleagues [11]. This technique was developed with the intention of increasing the accuracy of sentiment analysis models.

The researchers Madan, A. et al. [12] carried out sentiment analysis, which is concerned with analyzing words and then extracting hidden information from them. They did this by using a combined and novel framework that incorporates a Lexicon-based approach and a deep learning methodology. A solution that is based on deep learning was developed by D. Madera and colleagues [13]. This solution integrates BERT (both English and multilingual versions) and GraphSAGE, a Graph Neural Network (GNN) model, with techniques that focus on sentiment analysis and natural language processing (NLP). In their study, Wadawadagi, R., and colleagues [14] proposed a two-stage model that included consecutive stages. Both the extraction of aspect words and the categorization of sentiments are determined by aspect words. Aspect words have the potential to capture syntactic and semantic data without the requirement for complex feature engineering. The authors Walha, A., et al. [15] provide a MapReduce Extract-Transform-Load paradigm for the purpose of integrating opinions derived from tremendous volumes of user-generated content data. When it comes to semantically evaluating informal and unstructured texts and translating them into sensations, it underlines the difficulty of striking a balance between several factors, including time, cost, and complexity. An innovative deep learning-based sentiment classification model that is referred to as "Knowledge Graph Convolutional Networks" is presented by Menaouer, B. et al. [16]. This model is able to anticipate and analyze opinions toward the conflict between Russia and Ukraine.

Despite the fact that several methods for recognizing the polarity of tweets have been developed, there are still significant research gaps due to the fact that the discourse on social media is both complicated and constantly evolving. The vast majority of currently available methods are lexicon- or machine learning-based, with a concentration on feature engineering or classifier performance. However, these models often come up short when it comes to handling the informal language, slang, acronyms, emoticons, sarcasm, and context-dependency that are so prevalent in tweets. In spite of the fact that deep learning models like CNNs, BERTs, and hybrid frameworks have improved their performance, they are still unable to accurately capture delicate emotions, especially in tweets that include a combination of languages and codes. In addition, the ability of the systems to be used in genuine applications is hindered by the lack of real-time flexibility, scalability to enormous volumes of streaming data, and domain-specific generalization. For this reason, there is an undeniable need for sentiment analysis frameworks that are more resilient, adaptive, and aware of the environment in which they are being used. These frameworks should be able to manage the linguistic diversity and real-time nature of tweet data in order to properly evaluate polarity.

III. METHODOLOGY

The combination of an Auxiliary BiLSTM and a Dense layer for tweet polarity recognition is an effective way for extracting full sentiment insights from textual data. This design uses the Auxiliary BiLSTM to capture sequential and contextual relationships in tweets by processing input text both forward and backward, allowing the model to understand long-range links and nuanced linguistic patterns. The BiLSTM output is then delivered to a Dense layer, which acts as a feature integrator, combining the high-level temporal embeddings with additional learned representations. This integrated approach improves tweet classification by including semantic context and abstract properties, leading in more accurate and robust sentiment detection. As a result, it is well-suited to dealing with the challenges raised by the short, informal, and widely variable language used in social media communication.

Fig. 1 depicts a tweet polarity detection model that uses text processing and deep learning layers to predict sentiment class. The pipeline starts with raw tweets, which go through a preprocessing step to clean and normalize the content. The processed text is then passed into an embedding layer, which converts the tokens into dense vector representations that preserve semantic meaning. The auxiliary layer enriches these embeddings with extra contextual characteristics before passing them into a 32-unit Bidirectional LSTM (Long Short-Term Memory) layer. This BiLSTM captures sequential dependencies in both forward and backward directions, which improves the model's knowledge of context and word order. To avoid overfitting, the BiLSTM output is sent via a dropout layer (0.25), which disables a fraction of neurons at random during training. Next, the model has a dense layer with 256 neurons and ReLU activation, which enables non-linear transformation and feature learning. This is followed by another thick layer with a single neuron and sigmoid activation, which reduces the output to a probability score reflecting emotion polarity (positive or negative). The last result is the projected class, which represents the tweet's polarity. This architecture is ideal for binary sentiment classification tasks, particularly on informal and context-rich social media data.

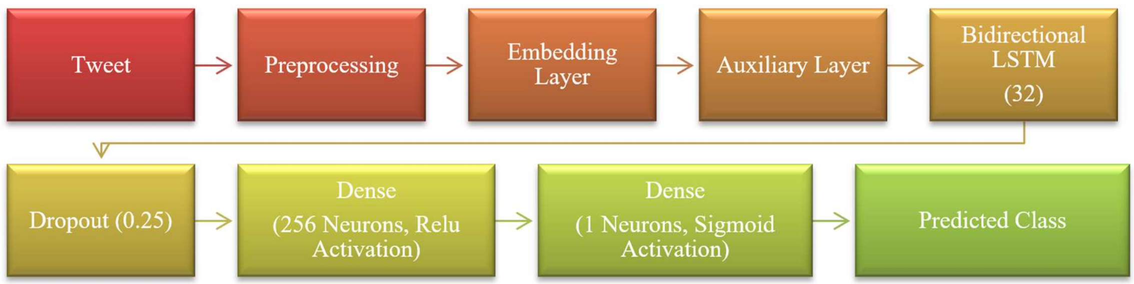

Figure 1. Layers of the hybrid neural network

This article describes the re-implementation of the CNN model, which was created with the goal of acting as a strong baseline. The process of creating a Convolutional Neural Network (CNN) model begins with the input of text data.
$X^{tw} = x_1^{tw}, x_2^{tw}, \dots, x_{tw}^{tw}$ and the terms of this subject $X^{to} = x_1^{to}, x_2^{to}, \dots, x_{to}^{to}$ with an embedding layer, as detailed in Section. Noise was introduced to the embedding layer, and dropout was used to reduce the likelihood of overfitting. In the second phase of the procedure, the phrase vectors of the text were connected with the phrase vectors of the topic, producing a cohesive sentence matrix model $W \in R^{sE}$. Here, s denotes the length of the linked sentences and E represents the size of the embedding layer. The notation A_{ij} or $A[i{:}j]$ is utilised to specify the sub-matrix of A from row i to row j. A convolution operation was applied utilising multiple filters $W \in R^{hE}$, where h represents the size of the filter window and E denotes the dimensions of the embedding layer. The filters employed utilise varying window sizes h, thereby accommodating n-grams of diverse lengths. The output $c_i \in R^{s-h+1}$ 1 of the convolution is generated by the iterative application of filters on the sub-matrices of A where$1 \leq i \leq s - h + 1$. The variable f represents an activation function, while the symbol $'.'$ denotes the dot product between the submatrix and the filters, as in (1). Additionally, $b \in R$ signifies a bias term. Filters are applied sequentially through the sentence matrix A_{ij}, resulting in the generation of a feature map.

$$c_i = f(w.A[i{:}i + h - 1] + b) \qquad (1)$$

$$c = [c_1, c_2, \dots, c_{n-h+1}]$$

Subsequently, max over time pooling was implemented across the various feature maps to extract the maximum value $\tilde{c} = \max\{c_1, c_2, \dots, c_{n-h+1}\}$ from each feature map.
The highest-valued characteristic in each map indicates the most significant features. After dropout is applied, the maximum features gathered from all filters are routed to a fully linked softmax layer. This layer then provides a probability distribution that includes all possible target classes. The Adam optimiser was employed throughout the optimisation procedure. The hyperparameter values set by were utilized to calculate the total number of kernels (filters) and the sizes of the kernels that corresponded to those amounts.
The LSTM block is described by the following (2), (3), and (4), with σ denoting the sigmoid activation function.
Gates:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \qquad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \qquad (3)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4)$$

Input Transform, in (5)

$$c_t = f_t . c_{t-1} + i_t . \tanh(W_c x_t + U_c h_{t-1} + b_c \quad (5)$$
$$\sigma(W_o x_t + U_o h_{t-1} + b_o)$$

State update, as in (6)

$$h_t = o_t . \tanh(c_t) \quad (6)$$

To acquire more information via the input data sequence used in the LSTM, we execute two iterations over it, shown in (7) and (8), one from left to right x_i to x_t, and one from right to left x_t to x_i. A representation is formed by concatenating the forward and backward LSTMs, $\vec{f}$ and $\overleftarrow{f}$, respectively.

$$h_i = \overrightarrow{h_l} || \overleftarrow{h_{l'}} \quad (7)$$

$$h_i \in R^L \quad (8)$$

where L represents the size of the LSTM

IV. Experimental Analysis & Results

Table I and Fig. 2 are a table and chart for tweet polarity classification that clearly shows the higher performance of the proposed model, which combines Auxiliary Feature Integration with CNN and BiLSTM networks. With an accuracy of roughly 0.982, the suggested technique beats all baseline models, demonstrating its ability to capture both semantic and contextual subtleties in tweets. CNN_GTO [9] has somewhat reduced accuracy, indicating that although CNNs using optimization approaches perform well, they may lack deep contextual awareness. KG-CNN [16] and GNN [13] perform reasonably, indicating that graph-based models are good at managing structured connections but may fall short in sequential text interpretation. Ensemble_Lexicon [7] has the lowest accuracy, demonstrating the limits of standard lexicon-based and ensemble approaches for analyzing informal and ambiguous twitter language. The findings show that combining CNN's local feature extraction, BiLSTM's sequential learning, and auxiliary feature enrichment allows the proposed model to generalize more effectively and categorize tweet sentiment with more accuracy and reliability.

TABLE I. Comparison of Various Methods for Accuracy

Method	ACCURACY
Proposed	0.9825
CNN_GTO [9]	0.9725
GNN [13]	0.96625
KGCNN [16]	0.97125
Ensemble_Lexicon [7]	0.9625

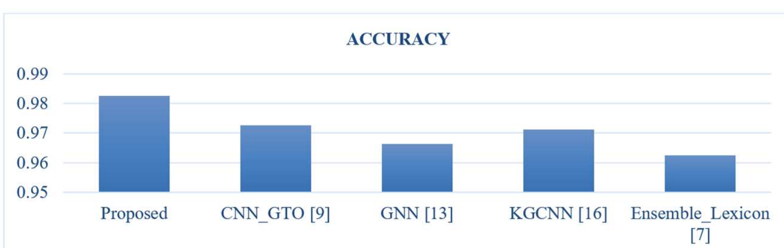

Figure 2. Comparison of Accuracy with a well-known state-of-the-art method

Table II and Fig.3 table and chart for the precision for tweet polarity classification demonstrates the performance of the proposed model, which uses Auxiliary Feature Integration with CNN and BiLSTM Networks. The suggested technique has the greatest accuracy value of about 0.972, demonstrating a good capacity to properly detect important positive or negative attitudes while reducing false positives. This high accuracy is due to the model's ability to use contextual information from BiLSTM, localized feature patterns from CNN, and extra insights from auxiliary features. CNN_GTO [9] and KG-CNN [16] have modest accuracy ratings, indicating a degree of false positives in their classifications. GNN [13] performs marginally worse, indicating a problem in capturing sequential sentiment flow inside tweet content. Ensemble_Lexicon [7] has the lowest accuracy, most likely because to its dependence on predetermined word lists and rules, which may not adapt well to casual, caustic, or confusing tweet content. Overall, the findings show that the proposed hybrid deep learning model is more exact in sentiment classification tasks, making it more suited to applications where eliminating erroneous sentiment predictions is crucial.

TABLE II. COMPARISON OF VARIOUS METHODS FOR PRECISION

Method	PRECISION
Proposed	0.972851
CNN_GTO [9]	0.959276
GNN [13]	0.952489
KGCNN [16]	0.957014
Ensemble_Lexicon [7]	0.947964

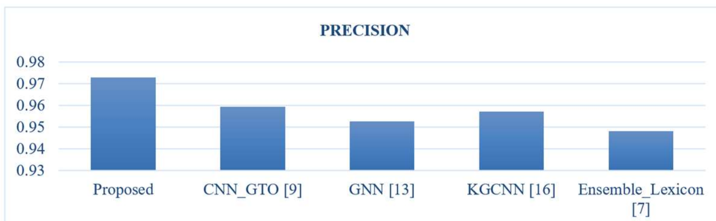

Figure 3. Comparison of Precision with a well-known state-of-the-art method

Table III and Fig.4 are a table and a chart for F1-score for tweet polarity classification, demonstrates the higher performance of the proposed model, which employs Auxiliary Feature Integration with CNN and BiLSTM Networks. With an F1-score of about 0.984, the suggested technique efficiently balances accuracy and recall, proving its robustness in reliably recognizing both positive and negative attitudes while reducing false positives and negatives. This strong F1-score implies that the model is trustworthy and consistent over a wide range of tweet situations, including informal language, acronyms, and sarcasm. CNN_GTO [9] and KG-CNN [16] perform relatively well, although their somewhat lower F1-scores indicate a trade-off between accuracy and recall. GNN [13] has a lower F1-score, most likely owing to constraints in modeling temporal relationships in tweet text. Ensemble_Lexicon [7] has the lowest F1-score, indicating a difficulty in generalizing across varied tweet structures and emotions owing to its static rule-based approach. Overall, the findings show that the suggested hybrid architecture offers the best balanced and effective performance for tweet polarity recognition, making it ideal for real-world sentiment analysis applications.

TABLE III. COMPARISON OF VARIOUS METHODS FOR F1-SCORE

Method	F1-SCORE
Proposed	0.983982
CNN_GTO [9]	0.974713
GNN [13]	0.96893
KGCNN [16]	0.973533
Ensemble_Lexicon [7]	0.965438

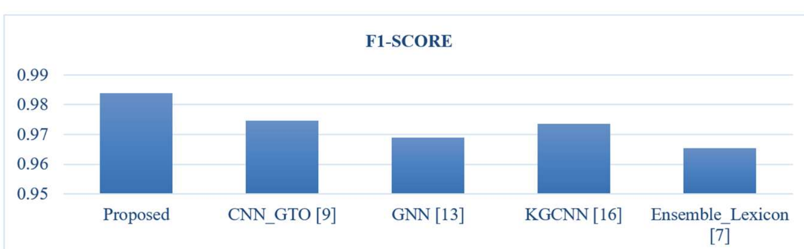

Figure 4. Comparison of F1-Score with a well-known state-of-the-art method

V. Conclusions

The proposed tweet polarity classification model, which combines Auxiliary Feature Integration with CNN and BiLSTM Networks, outperforms current techniques in terms of accuracy with 0.982, precision 0.972, and F1-score 0.984. By combining the sequential context awareness of BiLSTM, the spatial feature extraction capabilities of CNN, and the enhanced semantic comprehension through auxiliary features,the model efficiently captures tweets' complex language subtleties and informality. Experimental findings reveal that the proposed framework outperforms baseline models like CNN_GTO, GNN, KG-CNN, and Ensemble_Lexicon, demonstrating that it is a strong and trustworthy option for social media sentiment analysis.This hybrid architecture is well-suited for real-time and large-scale tweet polarity detection applications, delivering exceptional performance in sentiment classification and generalisation in noisy and dynamic settings.

References

[1] A. Ritter , S. Clark, Mausam, & O. Etzioni, (2011). “Named entity recognition in tweets: An experimental study,” In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP ’11), 2011 ,pp. 1524–1534.

[2] M.Speriosuc, N. Sudan, S. Upadhyay, & J. Baldridge, “Twitter polarity classification with label propagation over lexical links and the follower graph,” In Proceedings of the First Workshop on Unsupervised Learning in NLP (EMNLP ’11), 2011, pp. 53–63.

[3] S. Sweta, “Application of sentiment analysis in diverse domains,” In Sentiment Analysis and its Application in Educational Data Mining. SpringerBriefs in Applied Sciences and Technology. Springer, Singapore, 2024.

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

[5] D.I.Hernández Farias , & P.Rosso, “Irony, sarcasm, and sentiment analysis,” Sentiment Analysis in Social Networks, 2017, pp. 113–128.

[6] A. Giachanou, & F. Crestani, “Like it or not: A survey of Twitter sentiment analysis methods,” ACM Computing Surveys (CSUR), 2016,49(2), 1–41.

[7] N.F.F. da Silva, E.R. Hruschka, & E.R.Hruschka, “Tweet sentiment analysis with classifier ensembles,” Decision Support Systems, 2014,66, 170–179.

[8] E. Xia, H. Yue, & H. Liu, “ Tweet sentiment analysis of the 2020 U.S. presidential election,” In Companion Proceedings of the Web Conference 2021 (WWW’21),2021, pp. 367–371.

[9] F. Li, J. Li, & F. Abza, “Sentiment analysis of tweets employing convolutional neural network optimized by enhanced gorilla troops optimization algorithm,” Scientific Reports, 2025,15, 795.

[10] A.Gutiérrez, J. Aguilar, A. Ortega, & E. Montoya, “Sentiment analysis on social networks for defining innovation problems in organizations,” Technology in Society, 2025, 81, 102804.

[11] Y. Aliyu, A. Sarlan, K.U. Danyaro, A.S Abd Rahman, A.A. Muazu, & M.Y. Abubakar, “Deep learning techniques for sentiment analysis in code-switched Hausa-English tweets,” International Journal of Information Management Data Insights,2025,5(1), 100330.

[12] A. Madan, & D. Kumar, “Real-time topic-based sentiment analysis for movie tweets using hybrid approach,” Knowledge and Information Systems, 2035,67, 2543–2569.

[13] D.P. Madera-Espíndola, Z. Caballero-Domínguez, V.J. Ramírez-Macías, S. Butt, & H. Ceballos, “Detecting sexism in tweets: A sentiment analysis and graph neural network approach,” In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop),2025, pp. 48–54.

[14] R. Wadawadagi, S. Tiwari, & V. Pagi, “Polarity-aware deep attention network for aspect-based sentiment analysis,” Progress in Artificial Intelligence, 2025,14, 33–48.

[15] A. Walha, H. Mallek,F. Ghozzi, & et al., “Tweets opinion analysis integration: ETL modeling with MapReduce,” Cluster Computing,2025,28, 286

[16] B. Menaouer, S. Fairouz, M.B. Meriem, & et al., “A sentiment analysis of the Ukraine-Russia war tweets using knowledge graph convolutional networks,” International Journal of Information Technology, 2025