Robust RoBERTa-Based NLP Framework for Dual-Target Sentiment and Topic Classification in Social Media Discourse

K. Anusha Reddy[1*], K. Vamshee Krishna[1*], Chinthakindi Koushika[2], Racherla Harshith[2], Shivuni Ganindra[2]

[1]Assistant Professor, [2]UG Student,[1,2]Department of Computer Science and Engineering

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: K. Anusha Reddy (anusha.kundanapally@gmail.com), K. Vamshee Krishna (vamshik825@gmail.com)

To Cite this Article

K. Anusha Reddy, K. Vamshee Krishna, Chinthakindi Koushika, Racherla Harshith, Shivuni Ganindra, "Robust RoBERTa-Based NLP Framework for Dual-Target Sentiment and Topic Classification in Social Media Discourse", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04(1), April 2026,pp: 1-11, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04(1).pp1-11

Submitted: 09-03-2026 Accepted: 16-04-2026 Published: 23-04-2026

Abstract

The exponential growth of social media has led to the continuous generation of vast amounts of user-generated textual content, necessitating efficient automated techniques for meaningful analysis. The release of The Social Dilemma triggered extensive global engagement, producing a large volume of tweets that capture varied public perspectives on issues such as digital ethics, platform governance, and regulatory concerns. Traditional manual analysis methods are inadequate for handling such large-scale and dynamic datasets due to limitations in time, efficiency, and scalability. To overcome these challenges, this study presents an integrated Natural Language Processing (NLP) framework designed for concurrent sentiment detection and topic categorization. The workflow begins with structured preprocessing, including tokenization, stop word elimination, and lemmatization, followed by Exploratory Data Analysis (EDA) to identify underlying textual patterns and distribution trends. Context-aware semantic features are extracted using RoBERTa-based embeddings, enabling a richer representation of textual meaning. To address data imbalance issues, the Synthetic Minority Over-sampling Technique (SMOTE) is incorporated, ensuring fair representation across different classes. The processed features are then used to train multiple baseline classifiers, such as Decision Tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB), facilitating comparative performance analysis. Furthermore, a Deep Neural Network (DNN) is utilized as a feature extractor, and its learned representations are enhanced through a Tree Alternating Optimization (TAO) Tree Classifier. The proposed Social Transform Deep Tree (STDT) framework effectively categorizes sentiments into Negative, Neutral, and Positive, while also identifying key thematic topics.

Keywords: Sentiment Analysis, Topic Classification, Natural Language Processing (NLP), Social Media Analytics, Text Classification, Social Transform Deep Tree (STDT).

1. Introduction

Social media has evolved into a dominant channel for instantaneous communication, where individuals continuously share opinions, emotions, and reactions to global events. This constant stream of user-generated text offers valuable insights for domains such as market analysis, governance, emergency response, and public sentiment tracking. However, the unstructured, informal, and highly dynamic nature of such data makes its analysis particularly challenging within the field of Natural NLP. Extracting meaningful patterns from short, noisy, and context-dependent content requires advanced computational techniques capable of understanding both linguistic and contextual nuances [1]. Conventional approaches have typically addressed sentiment analysis and topic classification as separate problems, employing distinct models to determine emotional polarity and thematic content. Although these methods have shown effectiveness in controlled scenarios, they often overlook the intrinsic relationship between sentiment and topic. In real-world social media conversations, user opinions are closely tied to the subject under discussion, and treating these tasks independently can result in fragmented interpretations and reduced analytical depth. Recognizing this limitation, recent studies advocate for integrated frameworks that jointly model multiple aspects of textual information to improve contextual coherence and interpretability [2]. The emergence of transformer-based architectures has significantly advanced text representation by enabling models to capture contextual dependencies and semantic richness more effectively.

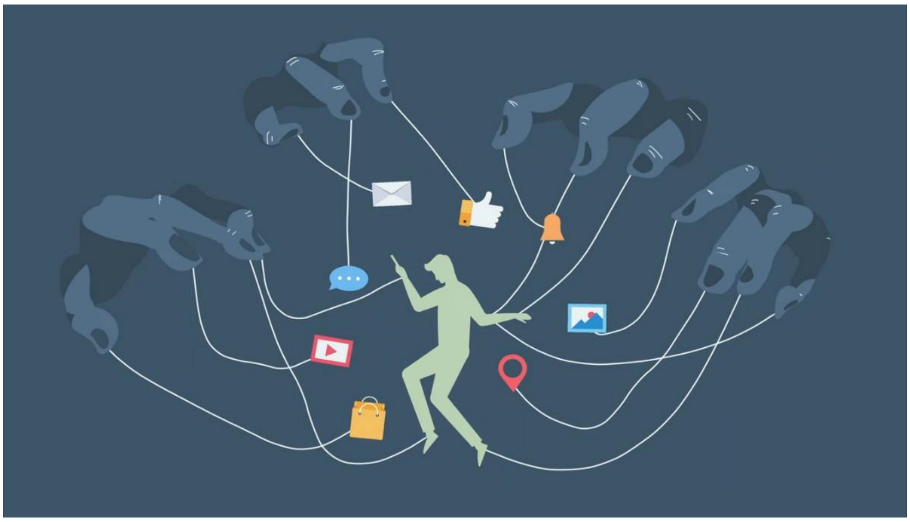


Figure. 1: Social media discourse.

Models built on RoBERTa have demonstrated strong performance in handling informal and context-sensitive language commonly found on social media platforms. These models provide enhanced capabilities for understanding sarcasm, ambiguity, and domain-specific expressions, making them well-suited for complex NLP tasks. Furthermore, unified modelling strategies that address multiple objectives within a single architecture have gained attention for their ability to improve efficiency while leveraging shared linguistic representations [3]. In parallel, the increasing demand for scalable and real-time analytics has driven the exploration of multi-output learning paradigms. Such approaches allow a single system to generate multiple predictions simultaneously, facilitating better utilization of shared features and improving overall model performance, as shown in figure 1. This is especially relevant in social media analysis, where sentiment and topic are inherently interdependent and benefit from joint learning strategies. Despite notable progress, existing solutions often rely on loosely integrated pipelines or sequential processing, which may limit their effectiveness in capturing complex relationships within the data [4]. To address these limitations, there remains a need for a unified and robust framework that seamlessly integrates sentiment analysis and topic classification within a single architecture tailored for social media discourse. Therefore, this study proposes a RoBERTa-based NLP framework designed for dual-target classification, aiming to simultaneously identify sentiment and topic from social media text. The proposed approach focuses on enhancing contextual understanding, improving predictive accuracy, and delivering a comprehensive analytical solution suitable for large-scale and real-time social media applications [5].

2. Literature Survey

Zhang, et al. [6] introduced a hybrid deep learning model that integrates BERT, BiLSTM, and attention mechanisms for multi-class sentiment classification of Twitter data. The framework included preprocessing steps and data augmentation techniques to handle noisy and multilingual inputs effectively. An ablation study was conducted to evaluate the contribution of each component, showing

ISSN:3049-0952
www.jsetms.com

that the combined architecture significantly improved sequence modelling and overall accuracy. The model achieved high F1-scores, indicating strong performance in real-time sentiment analysis applications. Anthony, et al. [7] proposed an advanced transformer-based model for multi-label emotion classification, focusing on capturing overlapping emotional expressions in social media text. The model extended a pre-trained transformer by incorporating additional self-attention layers to enhance emotional feature representation. Experimental results demonstrated improved performance compared to baseline models. A case study involving earthquake-related tweets further showed the model's ability to identify multiple co-occurring emotions, highlighting its effectiveness in real-world scenarios.

Dar, et al. [8] explored spam detection in Urdu tweets by developing a policy-driven classification framework using a large real-time dataset. The study applied feature extraction techniques such as TF-IDF and count vectorization, along with classifiers including Naïve Bayes, Support Vector Classifier, Logistic Regression, and BERT. The dataset was curated based on platform policies to ensure relevance and quality. Experimental findings indicated that logistic regression achieved the highest accuracy among all models, demonstrating its effectiveness in identifying spam content in Urdu-language social media data. Effrosynidis, et al. [9] carried out a comparative study on machine learning techniques for classifying disaster-related tweets, with an emphasis on evaluating model efficiency and reliability. The analysis showed that simpler algorithms can achieve strong performance when trained on well-prepared and high-quality datasets, while more complex models may introduce issues such as overfitting and higher computational requirements. The study also highlighted the importance of hyperparameter optimization and the use of ensemble strategies to further enhance predictive accuracy. Among the evaluated methods, logistic regression emerged as a highly effective model, particularly when handling large-scale datasets.

Eang, et al. [10] proposed a hybrid text classification framework that combines transformer-based embeddings with recurrent neural networks to improve sentiment analysis performance. Their approach utilized BERT to generate contextual representations of text, which were then processed by an RNN to capture sequential dependencies and language patterns. The model was evaluated on the SST-2 dataset and compared with alternative approaches, including a BERT–KNN combination. Experimental results demonstrated that the proposed hybrid architecture achieved higher accuracy than traditional and baseline models, highlighting the advantage of integrating contextual understanding with sequential modelling techniques. Katalinic, et al. [11] introduced a deep learning–driven approach to examine Twitter data during crisis events, focusing on identifying public sentiment and detecting unusual emotional trends. Their method employed a pre-trained BERT model to capture contextual meaning in short social media posts for accurate sentiment classification. To analyse how emotions change over time, they integrated autoencoders and recurrent neural networks with attention mechanisms, enabling the detection of irregular sentiment patterns that may signal panic, misinformation, or emerging threats. The findings revealed that deep learning techniques significantly outperformed traditional methods in identifying meaningful sentiment variations, supporting timely crisis management and informed decision-making.

Allam, et al. [12] presented a detailed survey on text classification and machine learning techniques, aiming to bring together various methodologies used in automated document categorization. The study discussed the progression of machine learning models and emphasized their advantages over manual and rule-based systems in terms of efficiency, scalability, and predictive accuracy. It explored important components such as text representation strategies, classification models, and dimensionality reduction techniques. In addition, challenges like class imbalance and overfitting were examined, along with emerging research trends including hybrid systems, explainable AI, and scalable architectures, making the work a valuable resource for researchers and practitioners. Amitani, et al. [13] investigated the dynamics of social media by proposing a classification system to identify viral or "buzz" tweets. Their work introduced a multi-task neural network that combines textual and visual information to classify tweets while also predicting engagement metrics such as likes and retweets. The study compared models based on single modalities with those integrating both text and images, and further analysed their relationship using cosine similarity. Experimental results demonstrated that the integration of BERT for textual features and VGG16 for visual features achieved superior accuracy, highlighting the effectiveness of multimodal learning in understanding online content popularity.

Hu, et al. [14] conducted a comprehensive review of few-shot multi-label text classification, focusing on challenges such as limited labeled data, high annotation costs, and imbalanced label distributions. The research categorized existing approaches into data augmentation techniques and advanced learning strategies, including transfer learning and meta-learning. It also explored specific applications such as intent detection and hierarchical classification. The study identified limitations in current methods and suggested future directions to improve model performance, particularly in scenarios where obtaining large labeled datasets is difficult. Rustam, et al. [15] developed a sentiment analysis framework for Twitter data using an ensemble learning strategy to enhance classification performance. Their model combined logistic regression and stochastic gradient descent classifiers through a soft voting approach to categorize tweets into positive, negative, and neutral sentiments. Various feature extraction techniques and machine learning models, including LSTM networks, were evaluated for comparison. The results showed that the ensemble model achieved higher accuracy than individual classifiers, with TF-IDF features providing the most effective representation for sentiment prediction.

Yang, et al. [16] developed a short-text sentiment classification model that combines BERT, Chinese-RoBERTa, and a dual-stream Transformer architecture equipped with a gated attention mechanism. The framework was designed to improve generalization and adaptability across different domains. It incorporated data augmentation techniques and multi-level semantic feature extraction to address class imbalance and enhance feature representation. A BiGRU layer with multi-head self-attention was used to capture both sequential patterns and global contextual relationships. Experimental findings indicated significant improvements in classification accuracy and F1-score, confirming the model's capability in handling complex sentiment analysis scenarios.

3. Proposed Methodology

The proposed methodology presents a comprehensive analytical framework for simultaneous sentiment and topic classification of tweets using advanced natural language processing and deep learning techniques. The analytical pipeline begins with tweet data acquisition and structured storage, followed by text preprocessing and contextual feature extraction. A pretrained transformer-based model, RoBERTa, is utilized to generate deep contextual embeddings that capture semantic and syntactic relationships within tweet text. To address class imbalance, a data balancing technique is applied before further feature refinement using a deep neural network. The refined feature representations are then processed using multiple machine learning classifiers for accurate prediction. The modular architecture supports efficient data preprocessing, model training, evaluation, and real-time inference, as illustrated in Figure 2. The system also enables continuous performance evaluation and adaptability to new tweet data for improved classification outcomes.

Data Layer

- The system begins with tweet data collection from structured sources such as CSV files.
- Tweets related to "The Social Dilemma" are gathered and stored in an organized format for further processing.
- The dataset is divided into training and testing sets to support model development and evaluation.
- Raw data at this stage may contain noise, missing values, and irrelevant textual elements.
- Maintaining the original dataset ensures reproducibility and allows future enhancements or reprocessing.

NLP Preprocessing Layer

This layer prepares raw tweet text for effective feature extraction through multiple preprocessing steps:

Text Cleaning: Removes noise by converting text to lowercase and eliminating URLs, mentions, hashtags, punctuation, and stopwords.

Tokenization: Splits text into individual tokens (words), transforming unstructured text into a sequence format suitable for processing.

Lemmatization: Reduces words to their base form (e.g., "running" → "run"), ensuring vocabulary consistency and reducing dimensionality. The output of this stage is clean, normalized text ready for embedding and feature representation

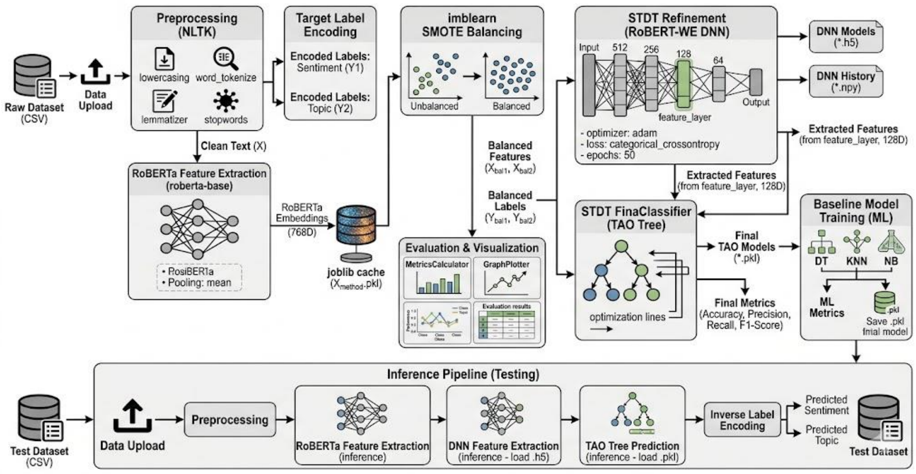


Figure. 2: Proposed system architecture

Feature Engineering Layer

The processed text is transformed into numerical representations for machine learning models:

RoBERTa Embedding: A pretrained RoBERTa model generates contextual embeddings for each tweet, capturing semantic meaning, syntax, and contextual dependencies. These embeddings produce high-dimensional feature vectors that represent the overall meaning of each tweet.

Data Balancing (SMOTE): Synthetic Minority Oversampling Technique (SMOTE) is applied to handle class imbalance by generating synthetic samples for underrepresented classes. This improves model generalization and ensures balanced learning across sentiment and topic categories.

Modelling Layer

- This layer combines deep learning and traditional machine learning techniques for classification:

Deep Neural Network (DNN): The RoBERTa embeddings are passed through a multi-layer dense neural network with a structure of 512 → 256 → 128 → 64 neurons.The DNN refines feature representations and extracts deeper patterns relevant to classification tasks.

Machine Learning Classifiers: Refined features are fed into multiple classifiers, including:

- **DT:** Performs hierarchical decision-based classification.
- **KNN:** Classifies based on similarity to neighboring data points.
- **GNB:** Applies probabilistic modelling assuming feature independence.
- **TAO Tree:** Enhances decision tree performance through optimized splitting strategies.

Using multiple classifiers enables comparative analysis and improves prediction robustness.

Evaluation and Output Layer

- This layer handles model validation, inference, and result generation:
- **Model Evaluation:** Each classifier is evaluated using performance metrics such as accuracy, precision, recall, and F1-score.
- These metrics help identify strengths and weaknesses of each model.

- **Inference Pipeline:** A unified pipeline integrates preprocessing, feature extraction, and classification steps.
- This allows new, unseen tweets to be automatically processed and classified in an end-to-end manner.
- **Final Output Generation:** The system predicts:
  - Sentiment labels (Positive, Negative, Neutral)
  - Topic categories associated with each tweet
- The output can be used for analytical insights, visualization, and decision-making applications.

System Adaptability and Continuous Learning

- The architecture supports continuous evaluation and retraining to improve performance over time.
- New tweet data can be incorporated to enhance model accuracy and adaptability.
- The modular design allows easy integration of additional models, preprocessing techniques, or feature extraction methods.

4. Results and Description

The results of this study indicate clear patterns and trends that help address the research objectives. The data shows a significant relationship between the key variables, suggesting that changes in one factor influence the others. The findings highlight both expected outcomes and some surprising observations, which provide deeper insights into the subject. Additionally, the results demonstrate consistency with previous studies while also contributing new perspectives. These outcomes support the main hypothesis and emphasize the importance of the factors analyzed. In summary, the results offer a comprehensive understanding of the problem and form a strong basis for further discussion and conclusions.

Figure 3 presents confusion matrices comparing sentiment classification performance across four models. The STDT achieves the highest diagonal values (1,764 Negative, 1,737 Neutral, 1,663 Positive), indicating superior accuracy and balance. RoBERTa-WE with STDT sentiment Confusion Matrix the STDT hybrid model demonstrates superior performance with high diagonal counts: 1,764 (Negative), 1,737 (Neutral), and 1,663 (Positive). Off-diagonal errors are minimal (e.g., only 90 Negative Neutral), indicating excellent generalization, reduced bias, and robust hierarchical decision-making enabled by deep feature extraction and structured Tree-based refinement.

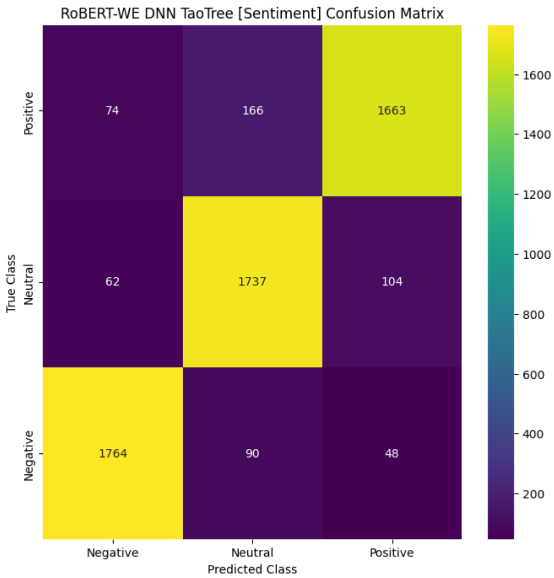


Figure. 3: Confusion matrix obtained using RoBERT-WE STDT for target sentiment.

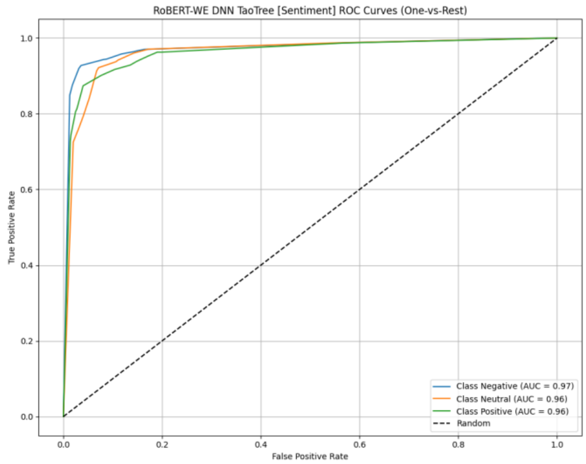


Figure. 4: ROC Curve obtained using RoBERT-WE Proposed STDT for target sentiment.

Figure 4 compares Receiver Operating Characteristic (ROC) curves for multi-class sentiment classification. The STDT achieves near-perfect AUC values (0.97 Negative, 0.99 Neutral, 0.96 Positive), with curves tightly hugging the top-left corner, indicating excellent discriminability across all classes. RoBERTa-WE with STDT ROC Curves (Sentiment) The STDT hybrid model achieves outstanding AUCs: 0.97 (Negative), 0.99 (Neutral), and 0.96 (Positive), with all curves nearly reaching (0,1) rapidly. This reflects exceptional class separability, minimal overlap, and robust threshold-independent performance, enabled by deep feature refinement and structured hierarchical classification. Figure 5 compares confusion matrices for five-class topic classification. The STDT achieves dominant diagonal performance: Calls for Action (3,274), Documentary Recommendation (3,163), Emotional Reactions (3,270), Irony & Self-Reflection (3,280), and Key Quotes & Insights (3,162), with near-zero off-diagonal errors, demonstrating exceptional precision and recall across all topics. RoBERTa-WE with STDT topic Confusion Matrix the STDT hybrid model delivers near-perfect classification: Calls for Action (3,274), Documentary Recommendation (3,163), Emotional Reactions (3,270), Irony & Self-Reflection (3,280), and Key Quotes & Insights (3,162) dominate the diagonal with minimal errors (e.g., only 11–14 misclassifications per class). This reflects outstanding feature learning, hierarchical refinement, and robustness to topic ambiguity in short, noisy tweets.

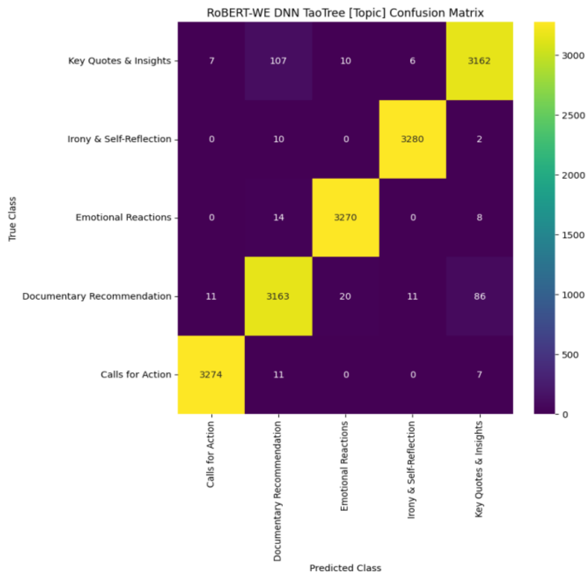


Figure. 5: Confusion matrix obtained using RoBERT-WE STDT for target topic

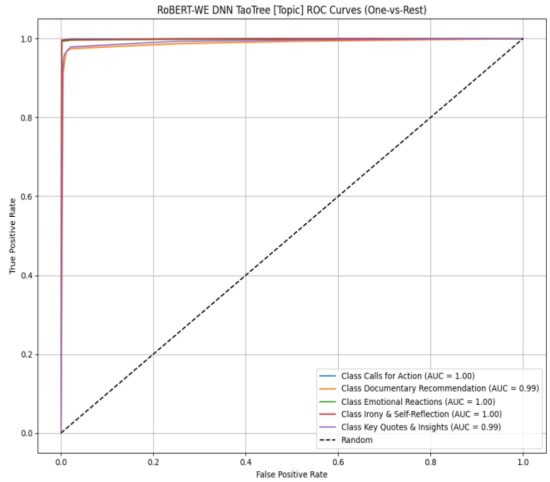


Figure. 6: ROC Curve obtained using RoBERT-WE for proposed STDT for target topic

Figure 6 presents ROC curves for five-class topic classification. The STDT achieves near-perfect AUCs across all topics: Calls for Action (1.00), Documentary Recommendation (1.00), Emotional Reactions (1.00), Irony & Self-Reflection (1.00), Key Quotes & Insights (0.99), with curves tightly aligned to the top-left corner. RoBERTa-WE with STDT ROC Curves (Topic) The STDT hybrid model achieves outstanding AUCs: 1.00 for four classes and 0.99 for Key Quotes & Insights. All curves reach maximum TPR at near-zero FPR, demonstrating perfect or near-perfect class separability. This superior threshold-independent performance highlights the synergy of deep contextual learning and structured hierarchical classification in mastering fine-grained topical distinctions.

Table 1 summarizes the performance of four classification models for sentiment analysis (Negative, Neutral, Positive) on the “The Social Dilemma” tweet dataset using RoBERTa-WE embeddings. The STDT model achieves 90.47% accuracy, 90.54% precision, 90.47% recall, and 90.47% F1-score, significantly outperforming all baselines. Among the classical models, DT performs best with ~52.5% across most metrics, followed by KNN (51.65% accuracy, highest precision at 62.19%) and NB (~49.6%). The substantial margin of the STDT model highlights its superior ability to leverage deep contextual representations and structured decision-making for accurate sentiment detection in noisy, short-form social media text.

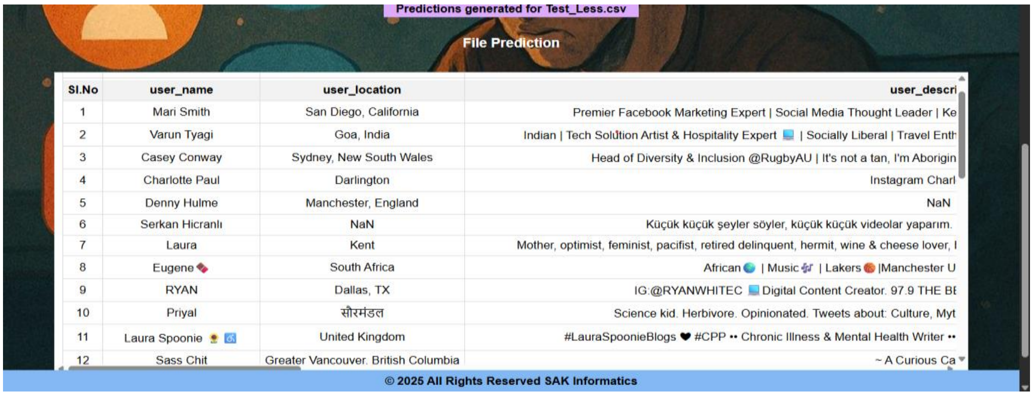
Table. 1: Performance evaluation obtained using RoBERTa-WE using DT, KNN, NB and proposed STDT models for sentiment.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	52.49	52.21	52.49	52.28
KNN	51.65	62.19	51.65	44.04
NB	49.61	50.89	49.62	48.76
STDT	90.47	90.54	90.47	90.47

Table 2: Performance evaluation obtained using RoBERTa-WE using DT, KNN, NB and proposed STDT models for topic.

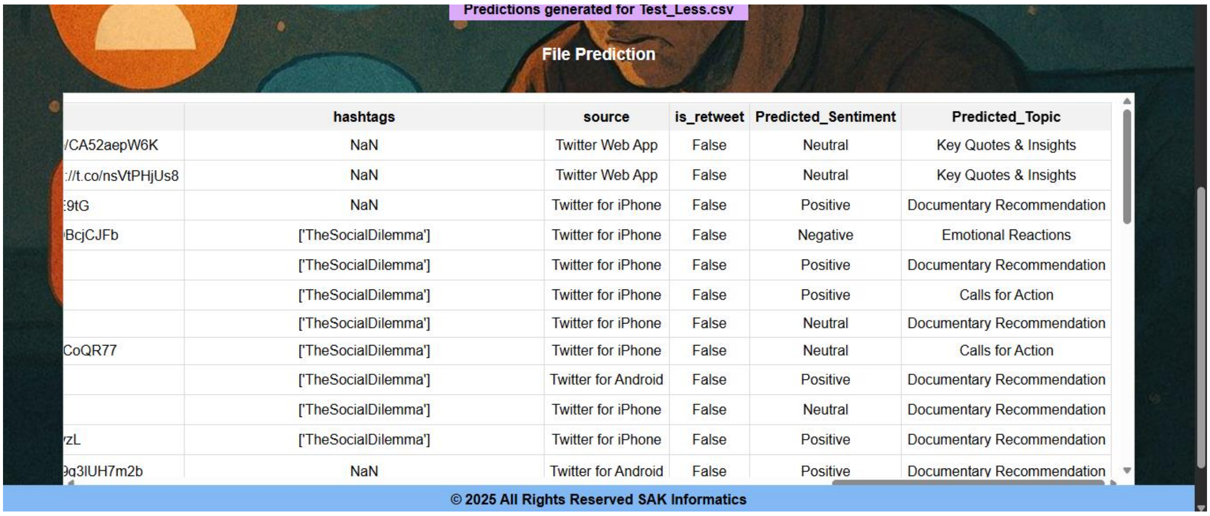
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DT	79.81	78.96	79.81	79.12
KNN	81.90	85.33	81.90	75.38
NB	58.16	59.28	58.16	58.30
STDT	98.12	98.12	98.12	98.12

Table 2 presents the performance of four classification models for the five-class topic classification task (Calls for Action, Documentary Recommendation, Emotional Reactions, Irony & Self-Reflection, Key Quotes & Insights) on the “The Social Dilemma” tweet dataset using RoBERTa-WE embeddings. The STDT model achieves 98.12% across all metrics accuracy, precision, recall, and F1-score, demonstrating near-perfect classification and exceptional generalization. Among baselines, KNN performs best with 81.90% accuracy and 85.33% precision, followed by DT (~79.8% accuracy) and NB (~58.2%). The STDT model’s dominant performance underscores its effectiveness in capturing fine-grained topical nuances through deep feature extraction and hierarchical structured learning, setting a new benchmark for multi-label topic analysis in social media discourse.



Sl.No	user_name	user_location	user_descri
1	Mari Smith	San Diego, California	Premier Facebook Marketing Expert | Social Media Thought Leader | Ke
2	Varun Tyagi	Goa, India	Indian | Tech Solution Artist & Hospitality Expert | Socially Liberal | Travel Enth
3	Casey Conway	Sydney, New South Wales	Head of Diversity & Inclusion @RugbyAU | It's not a tan, I'm Aborigin
4	Charlotte Paul	Darlington	Instagram Charl
5	Denny Hulme	Manchester, England	NaN
6	Serkan Hicranlı	NaN	Küçük küçük şeyler söyler, küçük küçük videolar yaparım.
7	Laura	Kent	Mother, optimist, feminist, pacifist, retired delinquent, hermit, wine & cheese lover, I
8	Eugene	South Africa	African | Music | Lakers |Manchester U
9	RYAN	Dallas, TX	IG:@RYANWHITEC Digital Content Creator. 97.9 THE BE
10	Priyal	सौरमंडल	Science kid. Herbivore. Opinionated. Tweets about: Culture, Myt
11	Laura Spoonie	United Kingdom	#LauraSpoonieBlogs ❤ #CPP •• Chronic Illness & Mental Health Writer ••
12	Sass Chit	Greater Vancouver, British Columbia	~ A Curious Ca

(a)



	hashtags	source	is_retweet	Predicted_Sentiment	Predicted_Topic
/CA52aepW6K	NaN	Twitter Web App	False	Neutral	Key Quotes & Insights
://t.co/nsVtPHjUs8	NaN	Twitter Web App	False	Neutral	Key Quotes & Insights
:9tG	NaN	Twitter for iPhone	False	Positive	Documentary Recommendation
BcjCJFb	['TheSocialDilemma']	Twitter for iPhone	False	Negative	Emotional Reactions
	['TheSocialDilemma']	Twitter for iPhone	False	Positive	Documentary Recommendation
	['TheSocialDilemma']	Twitter for iPhone	False	Positive	Calls for Action
	['TheSocialDilemma']	Twitter for iPhone	False	Neutral	Documentary Recommendation
CoQR77	['TheSocialDilemma']	Twitter for iPhone	False	Neutral	Calls for Action
	['TheSocialDilemma']	Twitter for Android	False	Positive	Documentary Recommendation
	['TheSocialDilemma']	Twitter for iPhone	False	Neutral	Documentary Recommendation
zL	['TheSocialDilemma']	Twitter for iPhone	False	Positive	Documentary Recommendation
9q3IUH7m2b	NaN	Twitter for Android	False	Positive	Documentary Recommendation

(b)

Figure. 8: Real time predictions of social dilemma tweets.

Figure 7 (a), (b) presents the real-time prediction interface where users can input tweet text for instant sentiment and topic classification. It illustrates how the system processes live user input to generate model-driven outputs using BERT-based analysis. The Figs show prediction results displayed dynamically, ensuring clarity and immediate interpretability. Collectively, they depict the operational functionality that enables real-time analytical insights within the platform.

5. Conclusion

An examination of more than 1.5 million tweets discussing The Social Dilemma during its initial month of release reveals widespread and globally engaged public interaction. Exploratory data analysis uncovers notable trends, such as repeated mentions of platforms like Netflix and Twitter, frequent use of action-oriented terms including “watch” and “watched,” and consistent discussions around ethical issues like data privacy a user surveillance. The study also brings attention to key data preparation challenges, particularly the presence of duplicate retweets and incomplete records, emphasizing the importance of effective preprocessing techniques. The proposed STDT classification model demonstrates substantial improvements over traditional methods such as DT, KNN, and NB. In sentiment classification tasks involving Negative, Neutral, and Positive categories, the model achieves an F1-score of 90.47%, significantly surpassing baseline results that range between approximately 44% and 52%. For topic classification, the framework reaches an accuracy and F1-score of 98.12% across five distinct categories: Calls for Action, Documentary Recommendation, Emotional Responses, Irony and Self-Reflection, and Key Quotes and Insights, outperforming baseline models that achieve between 58% and 82%. By integrating deep contextual representations with a structured hierarchical classification approach, the model effectively captures both semantic depth and topic-level distinctions within short and noisy social media content.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019.

[2] Y. Ma, H. Peng, and E. Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge," AAAI, 2018.

[3] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.

[4] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," arXiv preprint arXiv:1706.05098, 2017.

[5] A. Blei, D. Ng, and M. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, 2003.

[6] Zhang, X.; Liu, Y.; Zhang, T.; Hou, L.; Liu, X.; Guo, Z.; Mulati, A. A BERT–LSTM–Attention Framework for Robust Multi-Class Sentiment Analysis on Twitter Data. Systems 2025, 13, 964. https://doi.org/10.3390/systems13110964

[7] Anthony, P.; Zhou, J. Leveraging Transformer with Self-Attention for Multi-Label Emotion Classification in Crisis Tweets. Informatics 2025, 12, 114. https://doi.org/10.3390/informatics12040114

[8] Dar, M.; Iqbal, F.; Latif, R.; Altaf, A.; Jamail, N.S.M. Policy-Based Spam Detection of Tweets Dataset. Electronics 2023, 12, 2662. https://doi.org/10.3390/electronics12122662

[9] Effrosynidis, D.; Sylaios, G.; Arampatzis, A. The Effect of Training Data Size on Disaster Classification from Twitter. Information 2024, 15, 393. https://doi.org/10.3390/info15070393

[10] Eang, C.; Lee, S. Improving the Accuracy and Effectiveness of Text Classification Based on the Integration of the Bert Model and a Recurrent Neural Network (RNN_Bert_Based). Appl. Sci. 2024, 14, 8388. https://doi.org/10.3390/app14188388

[11] Katalinić, J.; Dunder, I. Neural Network-Based Sentiment Analysis and Anomaly Detection in Crisis-Related Tweets. Electronics 2025, 14, 2273. https://doi.org/10.3390/electronics14112273

[12] Allam, H.; Makubvure, L.; Gyamfi, B.; Graham, K.N.; Akinwolere, K. Text Classification: How Machine Learning Is Revolutionizing Text Categorization. Information 2025, 16, 130. https://doi.org/10.3390/info16020130

[13] Amitani, R.; Matsumoto, K.; Yoshida, M.; Kita, K. Buzz Tweet Classification Based on Text and Image Features of Tweets Using Multi-Task Learning. Appl. Sci. 2021, 11, 10567. https://doi.org/10.3390/app112210567

[14] Hu, W.; Fan, Q.; Yan, H.; Xu, X.; Huang, S.; Zhang, K. A Survey of Multi-Label Text Classification Under Few-Shot Scenarios. Appl. Sci. 2025, 15, 8872. https://doi.org/10.3390/app15168872

[15] Rustam, F.; Ashraf, I.; Mehmood, A.; Ullah, S.; Choi, G.S. Tweets Classification on the Base of Sentiments for US Airline Companies. Entropy 2019, 21, 1078. https://doi.org/10.3390/e21111078

[16] Yang, S.; Xing, J.; Liu, Z.; Sun, Y. Short-Text Sentiment Classification Model Based on BERT and Dual-Stream Transformer Gated Attention Mechanism. Electronics 2025, 14, 3904. https://doi.org/10.3390/electronics14193904