Integrating RoBERTa Embeddings with Dense Neural Networks and Classical Machine Learning for Enhanced Sentiment and Topic Analysis

K Sharmila Reddy[1*], Venaganti Shivani[2], Sangi Siddardha[2], Boda Murali[2], Kalla Shashank[2]

[1]Associate Professor & Head, [2]UG Student, [1,2]Department of Computer Science and Engineering

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India.

*Correspondence: K. Sharmila Reddy (sharmilakreddy@gmail.com)

ABSTRACT

The rapid growth of social media platforms has led to an unprecedented surge in user-generated textual data, making automated analysis essential for extracting meaningful insights. The documentary The Social Dilemma sparked widespread global discussions, generating millions of tweets that reflect diverse public opinions on social media ethics and regulation. Manual analysis of such large-scale data is inefficient, error-prone, and lacks scalability. To address these challenges, this study proposes a comprehensive Natural Language Processing (NLP) framework for simultaneous sentiment and topic classification. The proposed pipeline begins with systematic data preprocessing, including tokenization, stop word removal, and lemmatization, followed by Exploratory Data Analysis (EDA) to uncover linguistic patterns and distributional characteristics. For deep contextual understanding, RoBERTa-based embeddings are employed to extract high-quality semantic features from textual data. To mitigate class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied, ensuring balanced representation across target classes. The extracted features are utilized to train multiple baseline machine learning models, including Decision Tree (DT) Classifier, K-Nearest Neighbor (KNN), and Naïve Bayes (NB), enabling comparative performance evaluation. Furthermore, a Deep Neural Network (DNN) is implemented as a feature extractor, whose intermediate representations are leveraged by an advanced Tree Alternating Optimization (TAO) Tree Classifier to enhance predictive performance. The integrated framework, termed Social Transform Deep Tree (STDT) classifies sentiments into Negative, Neutral, and Positive, and identifies thematic categories such as Calls for Action, Emotional Reactions, and Key Insights. Experimental results demonstrate improved accuracy, robustness, and scalability, highlighting the effectiveness of hybrid deep learning and transformer-based approaches for complex social media text analytics.

Keywords: Contextual Embeddings, Transformer-Based Representation, Feature Extraction, Class Imbalance Handling, Synthetic Data Generation, Deep Feature Learning, Hybrid Framework, Social Media Ethics, Opinion Mining, Text Classification, Semantic Analysis, Scalability, Data Imbalance, Thematic Analysis.

1. INTRODUCTION

Social media has emerged as the dominant space for public discourse, with Twitter producing more than 500 million tweets each day, nearly 40% of which are connected to trending topics and media-related content. The documentary "The Social Dilemma" sparked extensive online engagement, generating conversations that ranged from individual emotional reactions to in-depth discussions on policies and digital governance.

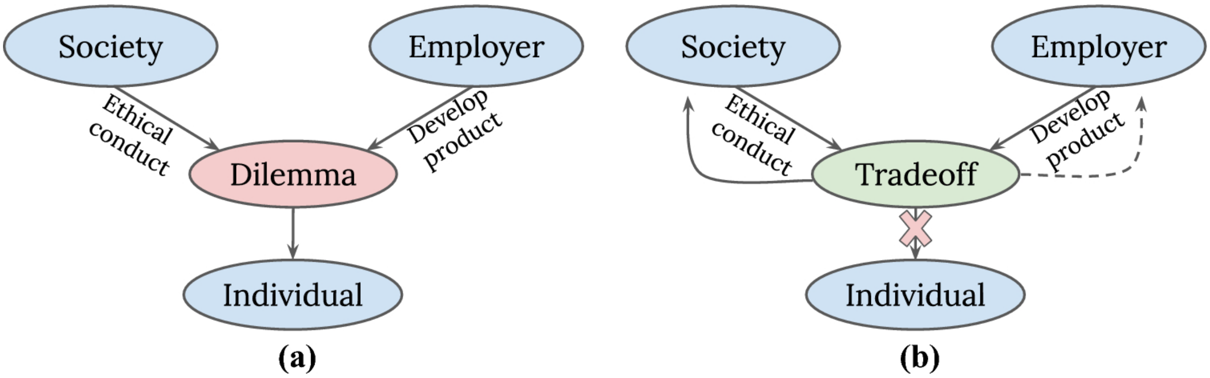


Figure 1: Comparative representation of ethical conflict models showing (a) traditional ethical dilemma and (b) tradeoff-based ethical resolution.

Figure 1 shows the Comparative Representation of Ethical Conflict Models (a) Traditional Ethical Dilemma Model illustrating conflicting pressures from society and employer on the individual, and (b) Tradeoff-based Ethical Resolution Model demonstrating a balanced mediation between professional obligations and ethical conduct.

(a) Ethical Dilemma Representation: This diagram illustrates the traditional ethical dilemma faced by an individual caught between societal expectations and employer demands. The Society expects adherence to ethical conduct, while the Employer pressures the individual to develop a product, often creating conflicting moral responsibilities. The individual is positioned at the intersection of these opposing forces, symbolizing the moral tension of ethical decision-making.

(b) Ethical Tradeoff Resolution: This diagram represents a tradeoff-based resolution model, where the focus shifts from an irresolvable dilemma to finding a balance between ethical conduct and professional duty. Here, Society and Employer still exert influence, but the Tradeoff node mediates the interaction. The dashed arrow indicates a moderated or compromised relationship between product development and ethical responsibility, leading to a more sustainable decision framework. Understanding these conversations in real time is essential to identify societal concerns, gauge emotional reactions, and uncover collective calls for action. Since unstructured data from social media contains sarcasm, irony, and overlapping themes, its analysis holds critical importance for researchers, policymakers, and digital platform regulators.

2. LITERATURE SURVEY

Katalinic, et al. [1] developed a deep learning pipeline to analyses tweets during crisis situations to understand public sentiment and detect unusual emotional patterns. A pre-trained BERT model was used to classify tweet sentiment by capturing contextual meaning in social media text. Autoencoders and RNNs with an attention mechanism were then applied to identify anomalies in evolving sentiment trends that could indicate misinformation, panic, or emerging hazards. The results showed that neural network methods outperformed traditional approaches in detecting meaningful sentiment shifts. This research supported timely crisis monitoring and informed decision-making for emergency response. Allam et al. [2] conducted a comprehensive survey of text classification and machine learning to consolidate the diverse aspects of automated text categorization into a unified resource. The study examined the evolution of machine learning techniques, highlighting their advantages over manual and knowledge-based approaches in terms of accuracy, efficiency, and adaptability. It evaluated document representation methods, classification models, and dimensionality reduction strategies while identifying key challenges such as class imbalance and overfitting. The survey also explored emerging trends and provided research directions, including hybrid models, explainable AI, and scalable solutions. This work served as a valuable reference for advancing research and practical applications in text classification.

Amitani et al. [3] investigated social media trends and developed a buzz tweet classification method to understand the factors driving viral content on Twitter. They designed a multi-task neural network that combined text and image features to classify tweets as buzz or non-buzz while predicting engagement metrics such as likes and retweets. Models using individual features were compared with combined feature approaches, and the relationship between text and images was analysed using cosine similarity. Experimental results showed that integrating BERT and VGG16 achieved the highest classification accuracy, demonstrating the effectiveness of multimodal analysis. Hu et al. [4] conducted a systematic survey on few-shot multi-label text classification, addressing challenges such as data scarcity, high annotation costs, and long-tailed label distributions. The study reviewed existing techniques from perspectives including modelling strategies, datasets, and evaluation metrics, with methods categorized into data augmentation and advanced model training approaches like transfer learning and meta-learning. It also examined specialized tasks such as multi-label intent detection and hierarchical classification. The survey identified current limitations and outlined future research directions to improve performance in real-world applications.

Rustam et al. [5] proposed a voting classifier for sentiment-based tweet classification to support organizational decision-making using social media data. The model combined logistic regression and stochastic gradient descent through a soft voting mechanism to classify tweets into positive, negative, and neutral categories. Various feature extraction methods and classifiers, including an LSTM network, were evaluated to compare performance. Results showed that the proposed ensemble classifier achieved higher accuracy than individual models, with TF-IDF features yielding the best performance. Padhy et al. [6] proposed a machine learning–based framework for Twitter sentiment analysis to address limitations of traditional rule-based approaches. The study used word count vectorization and evaluated multiple classifiers, including Naïve Bayes, Decision Tree, KNN, Logistic Regression, and Random Forest. Performance was assessed using standard evaluation metrics such as accuracy, precision, recall, and AUC. Results showed that the Random Forest classifier achieved the best performance, demonstrating strong effectiveness even with minimal Twitter-specific features.

Melhem et al. [7] proposed DLCTC, a deep learning framework to improve short-text classification of traffic-related social media messages for Intelligent Transport Systems. The system combined character-, word-, and context-level features with BiLSTM and Text CNN-based attention to better capture semantic relationships and local patterns. External knowledge was incorporated to enhance understanding of abbreviations and domain-specific concepts. Experimental results showed that DLCTC outperformed existing CNN-, RNN-, and BERT-based approaches in classifying short traffic texts. Lughbi et al. [8] developed an automated NLP-driven dashboard to classify and visualize cyberattack-related tweets for real-time monitoring. They created the CybAttT dataset containing manually labelled cyberattack tweets and evaluated multiple classification algorithms to select the best-performing model. The chosen model was integrated into a streaming pipeline to support live tweet analysis and interactive visualization. Experimental results demonstrated the dashboard's effectiveness in providing timely and structured insights into cyberattack events.

Bello et al. [9] proposed a sentiment classification framework that integrated BERT with neural network variants to better capture contextual information in Twitter text. They combined BERT with CNN, RNN, and BiLSTM models to address ambiguity and varying text lengths in microblog sentiment analysis. Experimental evaluations showed that these hybrid models achieved higher accuracy, precision, recall, and F1-scores compared to Word2vec-based and standalone approaches. The study demonstrated the effectiveness of contextual transformer-based models for improved sentiment classification. Yang et al. [10] proposed a short-text sentiment classification model combining BERT, Chinese-RoBERTa, and a dual-stream Transformer gated attention mechanism to improve cross-domain adaptability and generalization. The framework used data augmentation and multilevel semantic mining to address class imbalance and enhance feature representation. A BiGRU with multi-head self-attention was integrated to capture sequential and global contextual dependencies. Experimental results showed significant gains in accuracy and F1 score, demonstrating the model's effectiveness in handling complex short-text sentiment scenarios.

Prottasha et al. [11] applied BERT-based transfer learning with a CNN–BiLSTM architecture to improve Bangla sentiment analysis in the presence of limited labelled data. They compared contextual BERT embeddings with traditional word embeddings such as Word2Vec, GloVe, and fastText, as well as classical machine learning models. The integrated deep learning framework demonstrated superior binary classification performance over existing approaches. The results highlighted the effectiveness of context-aware transfer learning for advancing Bangla NLP sentiment analysis. Kim et al. [12] applied BERT-based models to classify garlic-related COVID-19 misinformation on Twitter using a manually labelled tweet dataset. Multiple BERT and BERTweet variants were fine-tuned and evaluated to compare their effectiveness in misinformation detection. Models trained on a garlic-specific dataset significantly outperformed those trained on general rumor data. Experimental results showed that BERTweet-large achieved the highest accuracy and F1 score, demonstrating strong capability in identifying misinformation.

Cavus et al. [13] developed a cloud-based system called SA-ES to perform fine-grained sentiment analysis of fake news shared on Twitter during the COVID-19 pandemic. Using a BERT-based model, the system classified fake news sentiments into seven categories to better understand emotional patterns. The model was trained and evaluated on a large dataset, achieving high classification accuracy. The results demonstrated the effectiveness of the proposed system in analyzing emotional responses to misinformation on social media. Alruily et al. [14] proposed ARABERT4TWC, a BERT-based deep learning model for multi-class classification of Arabic tweets. They built a transformer-based framework with custom classification layers and performed preprocessing to enhance model accuracy. The Arabic-specific BERT model was fine-tuned and evaluated using multiple public datasets. Experimental results demonstrated improved performance in categorizing Arabic social media text compared to other deep learning approaches.

Miranda et al. [15] analysed the evolution of public sentiment in Spanish COVID-19 tweets using a fine-tuned BERT-based sentiment classification framework. They collected and pre-processed millions of tweets to study emotional trends related to the pandemic. The model enabled deep learning–driven sentiment analysis, and results were visualized to highlight public reactions and concerns. The findings provided insights into how the pandemic influenced social media sentiment and information spread.

3. PROPOSED SYSTEM

The proposed system architecture, illustrated in Figure 2, is designed for simultaneous sentiment and topic classification of tweets using advanced NLP and deep learning techniques. The framework integrates transformer-based embeddings, specifically RoBERTa, for contextual feature extraction, followed by data balancing to address class imbalance issues. The extracted features are further refined using a DNN, and multiple machine learning classifiers are employed for final prediction. The modular design of the system enables efficient preprocessing, model training, evaluation, and inference.

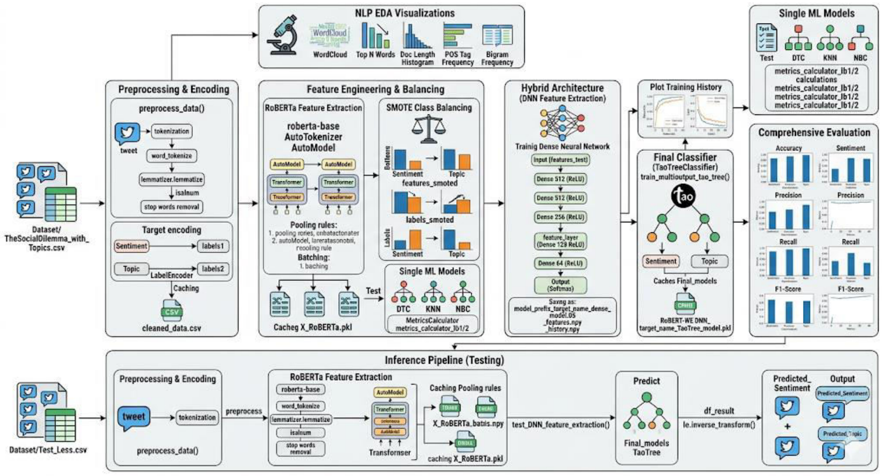


Figure 2: Proposed system architecture of STDT.

1. Data Layer

The first step involves collecting tweet data from sources such as CSV files. The tweets related to "The Social Dilemma" are collected and stored in a structured format. This ensures that both the training dataset and test dataset are readily accessible for further processing. The data storage step ensures that raw datasets are maintained for reproducibility and possible future enhancements. At this stage, the data is unprocessed and may contain noise, missing values, or irrelevant information.

2. NLP Preprocessing Layer

NLP Preprocessing is crucial for preparing the raw tweet text for feature extraction. This layer consists of three main steps:

1. **Text Cleaning:** Converts all text to lowercase, removes URLs, mentions, hashtags, punctuation, and stopwords. This removes noise that could interfere with model training.
2. **Tokenization:** Splits the text into individual tokens (words) to enable further analysis and embedding. Tokenization converts unstructured text into a structured sequence format.
3. **Lemmatization:** Reduces words to their base or dictionary form. For example, “running” becomes “run”. This reduces vocabulary size and improves feature consistency.

After preprocessing, the text data becomes clean, normalized, and ready for feature representation.

3. Feature Engineering Layer

Once the data is pre-processed, the next step is to convert text into numerical representations that machine learning models can understand.

1. **RoBERTa Embedding:** Pre-trained RoBERTa is used to generate deep contextual embeddings for each token in the tweets. Unlike traditional word embeddings, RoBERTa captures context, syntax, and semantic meaning, producing high-dimensional vectors for each tweet.
2. **Data Balancing (SMOTE):** Sentiment and topic classes may be imbalanced (e.g., more neutral tweets than positive or negative). SMOTE generates synthetic samples for minority classes to ensure balanced representation, improving model generalization.

4. Modelling Layer

In this layer, the system leverages both deep learning and traditional machine learning models**:**

1. **DNN:** The embeddings are passed through a dense neural network with multiple layers (512 → 256 → 128 → 64 neurons). The DNN refines the embeddings and extracts deeper features that are better suited for classification.

2. **Machine Learning Classifiers:** The refined features are fed into multiple classifiers including Decision Tree, K-Nearest Neighbors, Gaussian Naive Bayes, and TAO Tree. Using multiple classifiers allows the system to compare performances and improve prediction accuracy.

5. Evaluation and Output Layer

The trained models are then evaluated and deployed for inference:

1. **Model Evaluation:** Each classifier is evaluated using metrics such as accuracy, precision, recall, and F1-score**.** This ensures that the models are performing as expected and highlights potential weaknesses.

2. **Inference Pipeline:** The system integrates all preprocessing, feature extraction, and modelling steps into a unified inference pipeline. This allows new unseen tweets to be automatically processed and classified end-to-end.

3. **Final Output:** The system generates the final predicted sentiment (Positive, Negative, Neutral) and topic labels for each tweet. These predictions can be used for analysis, visualization, or decision-making.

4. RESULTS ANALYSIS

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.

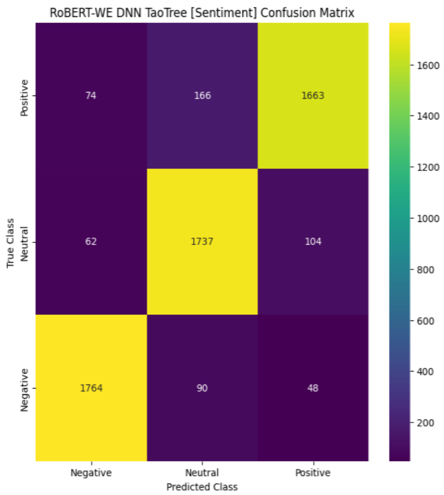


Figure 3: Confusion matrix obtained using RoBERT-WE Proposed STDT for Target Sentiment.

Figure 3: 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.

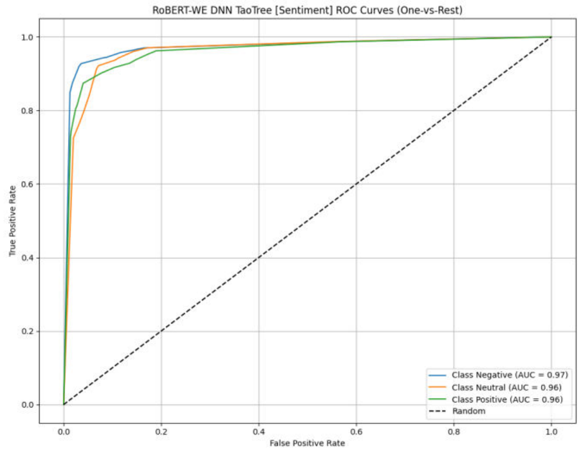


Figure 4: ROC Curve obtained using RoBERT-WE Proposed STDT for Target “Sentiment”.

Figure 4: 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.

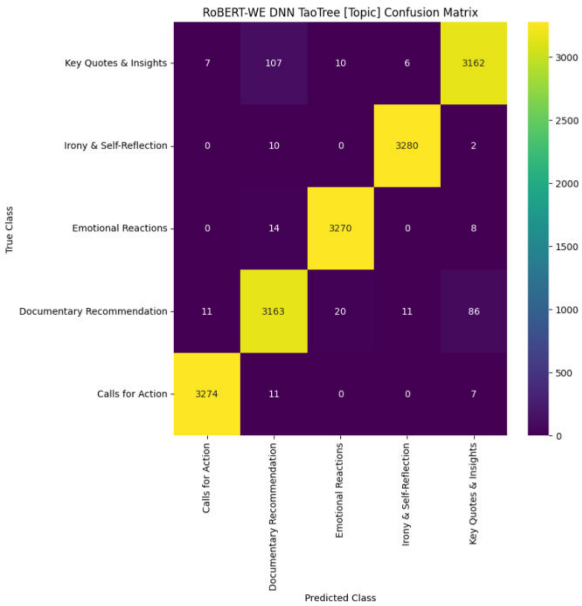


Figure 5: Confusion matrix obtained using RoBERT-WE Proposed STDT for Target “Topic”

Figure 5: 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.

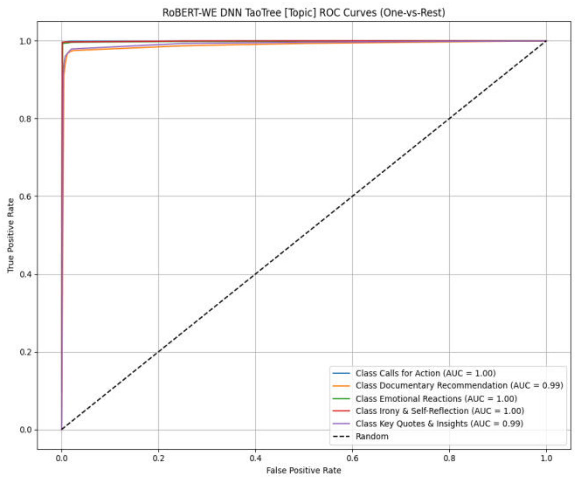
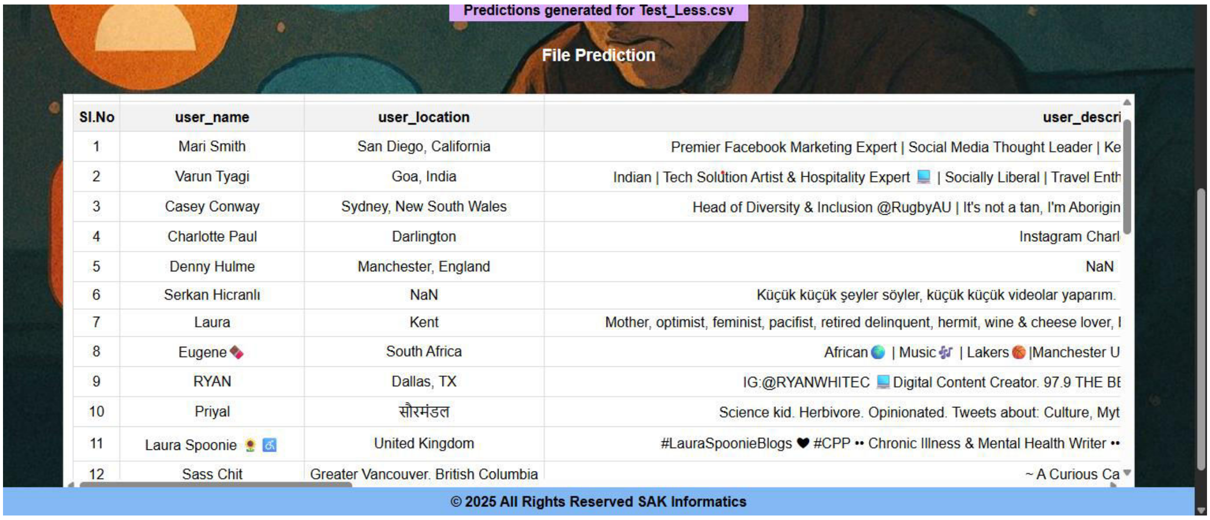


Figure 6: ROC Curve obtained using RoBERT-WE Proposed STDT for Target "Topic"

Figure 6: 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.



Predictions generated for Test_Less.csv

File Prediction

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)

Predictions generated for Test_Less.csv

File Prediction

	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

© 2025 All Rights Reserved SAK Informatics

(b)

Figure 7: 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.

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

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RoBERTa-WE + DT	52.49	52.21	52.49	52.28
RoBERTa-WE + KNN	51.65	62.19	51.65	44.04
RoBERTa-WE + NB	49.61	50.89	49.62	48.76
RoBERTa-WE + DNN + TAO Tree	90.47	90.54	90.47	90.47

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 2: Performance evaluation obtained using RoBERTa-WE DT, KNN, NB and proposed STDT models for Target "Topic".

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RoBERTa-WE + DT	79.81	78.96	79.81	79.12
RoBERTa-WE + KNN	81.90	85.33	81.90	75.38
RoBERTa-WE + NB	58.16	59.28	58.16	58.30
RoBERTa-WE + DNN + TAO Tree	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.

5. CONCLUSION

The analysis of over 1.5 million The Social Dilemma tweets in the first month post-release reveals a vibrant global discourse, with exploratory data analysis highlighting dominant themes such as platform references (Netflix, Twitter), action-oriented language ("watch," "watched"), and ethical concerns ("data," "privacy"), alongside preprocessing challenges like retweet artifacts and missing values. The STDT Classifier significantly outperforms traditional baselines (DT, KNN, NB), achieving 90.47% F1-score in sentiment classification (Negative, Neutral, Positive) compared to ~44–52% for baselines, and an exceptional 98.12% accuracy and F1-score in five-class topic classification (Calls for Action, Documentary Recommendation, Emotional Reactions, Irony & Self-Reflection, Key Quotes & Insights), far exceeding baseline performance of ~58–82%. This hybrid framework integrates deep contextual embeddings with structured hierarchical classification to robustly capture semantic nuance and topical subtlety in short, noisy social media text. It establishes a scalable, high-accuracy NLP pipeline for real-time analysis of documentary-driven public sentiment and discourse, offering broad applicability to future media impact studies, policy debates, and platform governance initiatives.

REFERENCES

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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

[6] Padhy, M.; Modibbo, U.M.; Rautray, R.; Tripathy, S.S.; Bebortta, S. Application of Machine Learning Techniques to Classify Twitter Sentiments Using Vectorization Techniques. Algorithms 2024, 17, 486. https://doi.org/10.3390/a17110486

[7] Melhem, W.Y.; Abdi, A.; Meziane, F. Deep Learning Classification of Traffic-Related Tweets: An Advanced Framework Using Deep Learning for Contextual Understanding and Traffic-Related Short Text Classification. Appl. Sci. 2024, 14, 11009. https://doi.org/10.3390/app142311009

[8] Lughbi, H.; Mars, M.; Almotairi, K. A Novel NLP-Driven Dashboard for Interactive CyberAttacks Tweet Classification and Visualization. Information 2024, 15, 137. https://doi.org/10.3390/info15030137

[9] Bello, A.; Ng, S.-C.; Leung, M.-F. A BERT Framework to Sentiment Analysis of Tweets. Sensors 2023, 23, 506. https://doi.org/10.3390/s23010506

[10] 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

[11] Prottasha, N.J.; Sami, A.A.; Kowsher, M.; Murad, S.A.; Bairagi, A.K.; Masud, M.; Baz, M. Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. Sensors 2022, 22, 4157. https://doi.org/10.3390/s22114157

[12] Kim, M.G.; Kim, M.; Kim, J.H.; Kim, K. Fine-Tuning BERT Models to Classify Misinformation on Garlic and COVID-19 on Twitter. Int. J. Environ. Res. Public Health 2022, 19, 5126. https://doi.org/10.3390/ijerph19095126

[13] Cavus, N.; Oktekin, B.; Goksu, M. A Cloud-Based Sentiment Analysis System with a BERT Algorithm for Fake News on Twitter. Appl. Sci. 2025, 15, 11046. https://doi.org/10.3390/app152011046

[14] Alruily, M.; Manaf Fazal, A.; Mostafa, A.M.; Ezz, M. Automated Arabic Long-Tweet Classification Using Transfer Learning with BERT. Appl. Sci. 2023, 13, 3482. https://doi.org/10.3390/app13063482

[15] Miranda, C.H.; Sanchez-Torres, G.; Salcedo, D. Exploring the Evolution of Sentiment in Spanish Pandemic Tweets: A Data Analysis Based on a Fine-Tuned BERT Architecture. Data 2023, 8, 96. https://doi.org/10.3390/data8060096