Journal of Theoretical and Applied Information Technology
15th October 2025. Vol.103. No.19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TRANSFER LEARNING BASED ROBERTA WITH SMOTE: A NOVEL APPROACH FOR TWITTER SENTIMENT ANALYSIS FOR HATE SPEECH

REKHA JANGRA[1], ABHISHEK KAJAL[2]

[1]Department of Computer Science and Engineering Guru Jambheshwar University of Science and Technology Hisar, Haryana, India

[2]Department of Computer Science and Engineering Guru Jambheshwar University of Science and Technology Hisar, Haryana, India

E-mail: [1]rekhajangra42@gmail.com, [2]drabhishekkajal@gmail.com

ABSTRACT

In the current digital era, a big spike of social interactions among people on various social networking sites have been witnessed around the globe. X platform (erstwhile Twitter) has turned leading platform for users to express their opinion on different burning topics related to social, political, religious domains. Though such discussions are healthy for a politically and socially active dynamic generation, but unfortunately many tweets carry hate speech. In past few years, India too witnessed the exponential rise of posts containing hate speech. Hence, we introduced a deep learning technique for catching the hate speech generated posts using X platform as our source for the dataset to classify the posts sentiments as negative and positive. Current research has assessed traditional ML methods, including SVM, Naïve Bayes, and Random Forest, alongside DL techniques. In light of the limitations of current methods, the suggested work offers a solution by merging RoBERTa with transfer learning. During preprocessing, the crawled Twitter data has been filtered, case-folded, and stemmed. Subsequently, stop word removal and tokenization have been executed. Data labeling has been conducted via deep learning algorithms. A word cloud has been generated, and a frequency chart has been produced based on positive and negative sentiments. Ultimately, the accuracy and error rates of LSTM, BERT, optimized RoBERTa, and Hybrid transfer learning-based RoBERTa with smote have been simulated. Simulation results indicate that LSTM attains 94.99%, BERT 92.87%, Optimized RoBERTa 97.45%, Hybrid RoBERTa with transfer learning 98.72% and Proposed SMOTE-based Hybrid RoBERTa with Transfer Learning 99.12%. The proposed model demonstrates superior accuracy, precision, recall, and F1-score relative to traditional methods. Consequently, the suggested approach has addressed the accuracy issues present in traditional hate speech Sentiment analysis systems.

Keywords: *Sentiment Analysis, Hate Speech, LSTM, BERT, ROBERTA, Transfer learning, SMOTE*

1. INTRODUCTION

With the exponential growth of social media users, resulted widespread of hate speech instances. In this past decade, detection of hate speech detection has been pulling much attention than many other research areas. The exponential growth of social media platforms has significantly transformed how individuals express their opinions, share information, and engage in public discourse. Among these platforms, Twitter has emerged as a dominant channel for real-time communication, where users often express sentiments on diverse topics ranging from politics and entertainment to social issues. However, this openness has also made Twitter a breeding ground for hate speech and abusive content, posing serious challenges for online safety, mental health, and societal harmony.

Sentiment analysis, particularly for detecting hate speech, has become an essential tool for monitoring online behavior and enabling platforms to take preventive measures. Traditional machine learning methods have shown some success in text classification tasks, but they often struggle with complex linguistic nuances, sarcasm, and the brevity of tweets. Moreover, hate speech datasets are usually imbalanced, with significantly fewer instances of hateful content compared to neutral or positive sentiments, resulting in biased and less

effective models. This serious problem needs to quickly resolve by detection of hate speech with better accuracy. This hates speech has become the leading research area in the domain of NLP. Twitter based sentiment analysis [1] for hate speech data is essential to understand the sentiment or opinion of public. Several authors have conducted researches in order to analyses the sentiment considering hate speech data. Examining these negative expressions offers insightful analysis of patterns in public opinion.

Traditional sentiment analysis techniques might find hate speech detection discourse's complexity—including sarcasm, metaphorical language, and changing vocabulary—challenging. Some of the researchers considered LSTM [2] while some focused-on Bi-RNN [3] approach. However, some author considered Bert [4] and some focused on conventional Naive based approach [5]. In same way, SVM DT, NB, LR Transformer based model Fuzzy classification and CNN with three-layer DL were frequently used for sentiment analysis. To address these challenges, this study proposes a novel framework that integrates Transfer Learning with RoBERTa, a powerful transformer-based language model known for its superior contextual understanding. Additionally, to combat class imbalance in the hate speech dataset, we incorporate SMOTE, which generates synthetic samples of the minority class to ensure balanced learning. The combination of RoBERTa and SMOTE allows for enhanced feature representation and improved classification accuracy, especially in identifying subtle forms of hate speech.

This research contributes to the field by demonstrating how advanced language models and data balancing techniques can be leveraged together for more reliable and robust sentiment analysis on social media. The proposed approach is evaluated on benchmark Twitter hate speech datasets and compared with existing methods to highlight its effectiveness in real-world scenarios. Paper investigates Transfer Learning-based RoBERTa with Smote model for sentiment analysis of Hate Speech tweets in order to overcome these difficulties, using its deep language knowledge and contextualized word representations to improve accuracy. Because of its dynamic masking, bigger training corpus, and strong fine-tuning techniques, RoBERTa with Smote, an improved variant of BERT, has outperformed BERT in NLP tasks. Applying this model to hate Speech sentiment analysis helps us to give a more accurate, scalable, and flexible way of assessing public opinion in real time and to surpass the constraints of conventional NLP methods. This work adds to the expanding area of computational hate speech analysis by providing a data-driven approach to evaluate hate speech sentiment with more accuracy and interpretation. As a means of categorizing hate speech, authors have largely relied on sentiment analysis. A few of the researchers analyzed the sentiment of tweets in real-time. To accomplish sentiment analysis for hate speech data, they utilized a DL and ML approach. Authors have explored the possibility of creating dynamic profiles for voting guidance applications by analyzing sentiment and Twitter data. While some scholars have used social network sentiment and network analysis to forecast election outcomes, other writers have examined political sentiment orientations on Twitter. Context, problems, suggested solution, Roberta's need, aims, and anticipated contribution are all detailed in Table 1.

Table 1: Summarizing The Key Aspects

Aspect	Description
Context	Social media platforms serve as a major hub for various discourses, where users express opinions on political events, policies, religious matters and leaders.
Challenges	Traditional sentiment analysis models struggle with sarcasm, figurative language, and evolving terminology.
Proposed Solution	A Transfer Learning-based RoBERTa with Smote model for sentiment analysis to enhance accuracy and contextual understanding.
Why RoBERTa?	RoBERTa, an optimized version of BERT, offers dynamic masking, larger training corpus, and improved fine-tuning for better NLP performance.
Why Smote?	To incorporate SMOTE, which generates synthetic samples of the minority class to ensure balanced learning.
Objectives	1. Improve sentiment classification accuracy for hate speech detection. 2. Address limitations of conventional NLP models. 3. Provide a scalable and adaptable model for real-time sentiment analysis.
Expected Contribution	A data-driven methodology for analyzing political sentiment with higher precision, reliability, and interpretability.

2. LITERATURE REVIEW

The author considered data from social media for sentiment analysis considering hate speech. A. Sharayu and R. V. (2019) did research on Sentiment

Analysis. This study mainly aimed to examine hate speech on Twitter. They used machine learning techniques to analyses sentiment. The main focus of J. C. Pereira-Kohatsu et al. (2019) was the detection and monitoring of hate speech on Twitter. This study introduced a new public dataset on hate speech in Spanish that includes 6,000 tweets that were annotated by professionals. An MLP approach with LSTM was considered by the author, and it produced an accuracy of 82%. L. Jiang et al. (2019) set out to perform this exact task—find hate speech in Twitter. Sentiment analysis was going to be their main focus. Deep learning methods like LSTM and BiRnNN were taken into consideration. LSTM achieved 73% accuracy, whereas BiRNN achieved 63%. Regarding hate speech detection, X. Zhou et al. (2021) did conduct some study. Their work revolved around the exchange of sentimental knowledge. Three distinct methods were employed: BiGRU, BERT, and GPT. The accuracy ranged from 94% to 95% according to this study. The Hate Speech Detection work was carried out by S. S. Alaoui (2022). The writer made use of ML and text mining. Sentiment Analysis has made use of Naïve Bayes. Text Mining has been executed using an English dataset. F. Alkomah in 2022. The author took into account methods like RF, SVM, and Glove. In this case, GLOVE achieves an accuracy of 89% while SVM achieves 83% [6]. It was suggested by Ali (2022) to analyses tweet sentiment on a wide scale. The author took the US presidential election dataset from 2020 into consideration [7].

Using ML and DL models, M. Subramanian et al. (2023) surveyed the state of hate speech identification and sentiment analysis. Various ML approaches were taken into account by the author, including SVM, DT, NB, and LR. Author also took transformer-based designs like BERT into account [8]. The topic of hate speech detection on Twitter was addressed by A. Abraham et al. (2023). Random Forest, SVM, CNN-LSTM, Logistic Regression, and Fuzzy Classification were among the models employed by the author [9]. Regarding social media sentiment analysis for hate speech, D. Bhattacharjee et al. (2023) conducted previous research. The research presents the application of a deep learning model for hate speech identification. CNN using three-layer deep learning classifiers achieved an accuracy of 91.28%, whereas LSTM achieved 82.67% [10]. Sentiment analysis taking Israeli political tweets into account was conducted by Gangwar (2023) [11]. To analyse public opinion on political issues, the author used Machine Learning.

to forecast the outcome of the 2023 election, Alvi [12] employed sentiment analysis on Twitter data. Analysis of Twitter data from a large number of languages was investigated by Antypas (2023) [13]. The function of emotion in political discourse was examined in the literature. A study by Hobbs (2023) [14] examined COVID-19, political PR, and leadership. Researchers compared Ardern and Morrison's performance as prime ministers on social media. Using machine learning algorithms, Patel (2023) [15] analyzed political retweets to determine human behavior. To better comprehend the 2023 presidential election in Nigeria, Olabanjo(2023) [16] laid out a sentiment analysis framework. Perera (2023) [17] compared the traits of those who spread hate speech. The author took Twitter's user behavior into account. Vahdat-Nejad (2023) [18] investigated public opinion on the conflict between Russia and Ukraine. Their main focus has been on categorizing and modelling global mood trends. Sentiment analysis of tweets concerning COP9 was investigated by Elmitwalli (2024) [23]. The author compared conventional methods with pre-trained models. A study on sentiment analysis on Twitter was conducted by Mantika (2024) [24].

For the 2024 presidential election, they thought about using Naïve Bayes and Logistic Regression. Using Twitter sentiment analysis, Patel (2024) [25] was able to forecast the results of the Indian elections. Multilingual hate speech and cyberbullying detection was the subject of research by E. Mahajan et al. (2024). This study has taken into account BiLSTM, LSTM, and Bi-GRU [26]. X. Shen et al. (2025) were involved in Hate Speech Detector investigations. They worked on content that was generated by LLM. The author took into account a dataset of hate speech [27]. In order to tackle class imbalance in sentiment analysis tasks, SMOTE has been the subject of multiple studies. By including SMOTE with SVM and Naive Bayes classifiers, Flores et al. [28] showed that sentiment datasets could be better classified with higher accuracy, thanks to the increased representation of minority classes. The combined effects of SMOTE, various feature representation techniques, and classification algorithms on unbalanced sentiment data were also studied by Satriaji [29]. According to their research, using SMOTE in conjunction with the right classifiers and feature extraction methods greatly improves model performance. Singgalen [30] compared Decision Tree (DT) and SVM models with SMOTE, demonstrating that SVM, when combined with SMOTE, achieved better results in sentiment classification than DT alone. In

addition, using ensemble machine learning techniques, Putra et al. [31] used SMOTE on datasets of imbalanced hotel reviews.

2.1 Models Considered

Present research is considering study of conventional ML for sentiment analysis.ML Models considers SVM, Naïve Bayes, Random Forest.

- SVM: Finds the optimal hyperplane to classify sentiments.
- Naïve Bayes: A probabilistic classifier based on Bayes' theorem.

Above technique were frequently used in conventional research work. But it has been observed that DL is better than ML models. Thus, present research is considering simulation of DL models. DL Models used in research is considering LSTM, BERT, Optimized RoBERTa and Hybrid model that is integration of RoBERTa and Transfer learning with Smote.

2.2 Research Gap in Sentiment Analysis

Despite the considerable interest in hate speech sentiment analysis using Twitter (currently X) data, many research gaps remain. Most studies primarily focus on textual analysis, neglecting the multimodal emotional cues sent by graphics. Several studies considered real-time sentiment in different countries. But there was the need to improve the accuracy and performance where conventional deep learning models were used. While traditional DL models utilized LSTM, traditional ML models utilized SVM, Naïve Bayes, Random Forest, and Smote. The sentiment analysis model's reliability for the X-based hate speech dataset can be enhanced by including more advanced mechanisms, such as transform learning and optimization.

3. PROBLEM STATEMENT AND RESEARCH QUESTION

Even while hate speech detection on social media sites like Twitter has come a long way, machine learning and deep learning models still have trouble with problems like class inequality, sarcasm, metaphorical language, and a vocabulary that changes quickly. Traditional models, like SVM, Naïve Bayes, and Random Forest, frequently don't get the whole context. Deep learning models, such LSTM and BERT, have made things better, but they still don't work well with data that isn't balanced. These deficiencies need a more robust framework that can keep a high level of contextual awareness while balancing datasets.

Research Questions framed from this critique are:

1. How can transfer learning with advanced transformer models like RoBERTa be enhanced to improve hate speech detection accuracy compared to existing methods?
2. Can synthetic oversampling (SMOTE) effectively address class imbalance in hate speech datasets without introducing bias or noise?
3. What level of improvement in performance metrics (accuracy, precision, recall, and F1-score) can be achieved by integrating RoBERTa, transfer learning, and SMOTE compared to baseline deep learning models?

4. PROPOSED WORK

4.1 Proposed Work Protocol

The suggested procedure was modeled after previous research that used deep learning models, including LSTM [Jiang et al., 2019], BERT [Zhou et al., 2021], and CNN-based frameworks [Bhattacharjee et al., 2023], for the identification of hate speech. This research used preprocessing (tokenization, stop-word removal, stemming) and then feature extraction using transformer-based architectures. Flores et al. (2018) and Satriaji (2018) both showed that SMOTE works well for unbalanced data. Our methodology is different since it combines transfer learning-based RoBERTa with SMOTE for strong sentiment categorization. This standardized methodology made sure that data was collected, cleaned, models were trained, and comparisons were made in a methodical way.

4.2 Proposed Methodology

The suggested sentiment analysis system guarantees efficient training, assessment, and comparison of deep learning models by means of a methodical procedure. Data collecting starts the process; sentiment data is acquired from several sources such consumer reviews, social media, or survey answers. After then, training and testing ratio is defined by the initializations phase, therefore guaranteeing a best data split for model learning. Model selection for training occurs after initialization is over, wherein many deep learning models are under evaluation for sentiment classification. Effective for sequential data processing, LSTM, BERT, is known for capturing bidirectional context in text, Optimized RoBERTa, an improved version of BERT with enhanced efficiency, and Hybrid Transform Learning-Based RoBERTa with Smote, a customized RoBERTa variant with additional optimizations. The proposed methodology aims to effectively detect hate speech in Twitter data using a hybrid approach that

combines the power of Transfer Learning with RoBERTa and the SMOTE technique for addressing class imbalance.

Figure 1 is presenting process flow of proposed work; the training process starts wherein these models choose sentiment patterns from the dataset. The trained models are then evaluated on a different test dataset to see how well they generalize. A model's effectiveness can be ascertained by calculating performance evaluation metrics like F1-score, recall, accuracy, and precision. With these rules in place, it is possible to evaluate how well a model classifies sentiment. Lastly, a comparative analysis is carried out to analyze the performance of multiple models using the computed assessment criteria.

Use cases that utilize real-world sentiment categorization include social media monitoring, opinion mining, and customer feedback analysis. The top-performing model is selected for these applications. Applying SMOTE to the training data helps to reduce this issue. SMOTE creates new minority class instances by combining current ones, thus it can balance the distribution of classes without adding unnecessary duplicates. This method improves overall accuracy and reliability by providing a strong and optimal framework for sentiment analysis. An effective sentiment analysis method is assured by this methodical process flow, which maximizes model selection and training while improving accuracy, precision, recall, and F1-score. Several industries can benefit from this system's output, such as those dealing with consumer feedback analysis, social media monitoring, and opinion mining.

The proposed sentiment analysis system ensures that deep learning model training, evaluation, and comparison are carried out efficiently through a systematic process flow. Data preprocessing, class balancing, feature extraction with RoBERTa, and classification are the general sequential phases that make up the model's architecture. A dense neural network layer, often known as a SoftMax classifier, takes RoBERTa's output and assigns each tweet a predefined sentiment score—Neutral, Offensive, or Hate Speech.

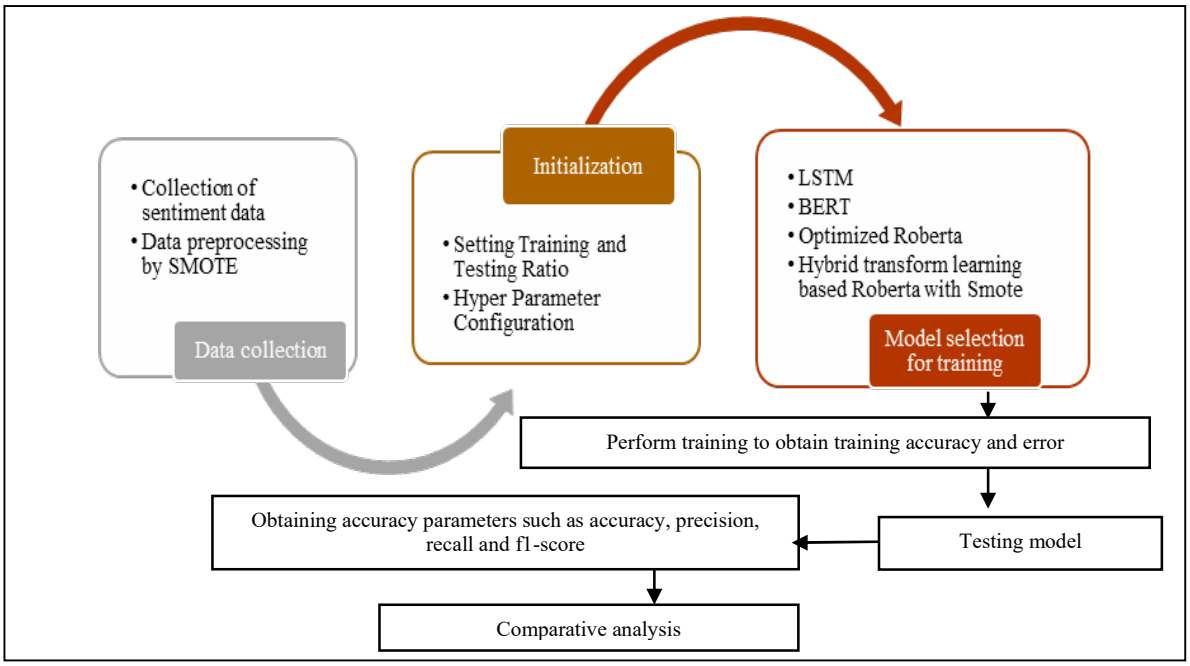

Figure 1: Process flow of Proposed Model

4.3 Dataset

The dataset used for this research consists of user-generated posts from Twitter (X), obtained via Kaggle and structured as a CSV file. Source of dataset is https://www.kaggle.com/code/eisgandar/twitter-sentiment-analysis-hatred-speech. It has two main parts: (i) Text, which is the substance of the tweet, and (ii) Sentiment, which is a label that tells you whether the tweet is hated one or not. The Twitter Sentiment Analysis – Hatred Speech dataset (2022), hosted on Kaggle, is widely used for studying the automatic detection of hate speech and offensive content on social media platforms. The dataset

primarily focuses on tweets that are annotated for the presence or absence of hate speech, with labels indicating whether the tweet contains hatred in terms of racist or sexist content (1) or is free from such expressions (0).

Table 2: Dataset description

Feature	Description
Source	https://www.kaggle.com/code/eisgandar/twitter-sentiment-analysis-hatred-speech
Year	2022
Task Objective	Binary classification: label 1 for racist/sexist (hate speech), label 0 for non-racist/sexist
Tweet Anonymization	User mentions replaced with @user to ensure privacy/anonymity
Data Split	Separate training and test CSV files, providing full tweet texts with corresponding labels
Usage in Notebooks	Featured in notebooks for exploratory data analysis, preprocessing, TF-IDF feature extraction, SMOTE balancing, RNN/LSTM modeling, and evaluation (e.g., Bi-LSTM, SMOTE, K-fold CV)
Evaluation Metric	F1-Score is commonly used to assess model performance due to class imbalance
Typical Pre-processing	Text cleaning includes handling punctuation, stop words, special characters, hashtags, emoticons, and possibly tokenization or normalization.

After preprocessing, the dataset was filtered with high-quality samples only, which left about 5,000 labeled records. For model evaluation, the dataset was split into two parts: one for training and another for testing. This is a common method in deep learning that makes sure there is enough data for the model to learn from and that the performance assessment is fair. The training set included enough instances for transformer-based models to learn in context, while the testing set checked how well the models could generalize. The Synthetic Minority Oversampling Technique (SMOTE) was used to maintain the class balance by adding samples from the minority class to prevent the majority class from becoming biased. This organized dataset, which included a balanced number of examples from each class and clear text characteristics, was the basis for getting high-performance sentiment classification.

4.4 Conventional Algorithm for Sentiment Analysis using LSTM

This algorithm leverages LSTM for sentiment analysis, classifying texts into positive and negative sentiment.

Step 1: Consider a text T, the goal is to predict its sentiment S:

Step 2: Feature Representation

Step 3: LSTM Processing

Step 4: Sentiment Classification

Step 5: Algorithm for Sentiment Analysis Using LSTM

Phase 1: Input Processing

- Preprocess text (tokenization, stop word removal, lowercasing).

Convert words into embeddings.

Phase 2: LSTM Processing

- Pass word sequence through LSTM network.
- Extract the final hidden state.

Phase 3: Classification

- Apply SoftMax layer to obtain sentiment probabilities.
- Select sentiment with highest probability.

4.5 Conventional Algorithm for Sentiment Analysis using BERT

This algorithm leverages BERT to classify Hate speech texts into positive, negative, or neutral sentiment. BERT is powerful for sentiment analysis because it captures deep contextual meaning in text.

Step 1: Consider a text T, the goal is to predict its sentiment S:

Step 2: Feature Representation using BERT

Step 3: Sentiment Classification

Step 4: Algorithm for Sentiment Analysis Using BERT

Phase 1: Input Processing

- Preprocess text (tokenization, stop word removal, lowercasing).
- Tokenize the text using BERT's Word Piece tokenizer.
- Convert tokens into BERT input format: [CLS] Text [SEP].

Phase 2: BERT Processing

- Pass tokenized text through BERT.
- Extract [CLS] token's hidden state.

Phase 3: Classification

- Apply SoftMax layer to obtain sentiment probabilities.
- Select sentiment with highest probability.

4.6 Algorithm for Optimized RoBERTa Model

This algorithm integrates an optimization mechanism into RoBERTa model. The optimization focuses on adaptive learning rate scheduling and gradient-based weight adjustment to enhance fine-tuning performance.

Step 1: Input Processing

Given an input sentence S consisting of n tokens: $S = \{t1, t2, \ldots, tn\}$

Tokens are mapped to embeddings: E={e1,e2,...,en} where ei=Embedding(ti)

Step 2: Encoding with RoBERTa

Embeddings pass through L transformer layers: Hl=Transformer (Hl−1)∀l∈ [1,L] where H⁰ = E and H^L represents final contextualized embeddings.

Step 3: Optimization Mechanism

- To optimize learning, we introduce:

1. Adaptive Learning Rate Scheduling using Cosine Annealing:

$$\eta t=\eta min+1/2\ (\eta max-\eta min)\ (1+\cos(t/T\pi))$$

where ηt is learning rate at step t, ηmax,ηmin are max and min learning rates, T is total training steps.

2. Gradient-Based Weight Adjustment using AdamW (Weight Decay Regularization):

$$mt=\beta 1mt-1+(1-\beta 1)gt$$

$$vt=\beta 2vt-1+(1-\beta 2)gt2$$

$$m\hat{}t=mt/(1-\beta 1t),$$

$$v\hat{}t=vt/(1-\beta 2t),$$

$$\theta_t = \theta_{t-1} - \frac{\eta_t}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t - \lambda\theta_{t-1}$$

Where gt=∇θJ(θ) is gradient of loss J(θ), β1, β2 are momentum coefficients, λ is weight decay.

Step 4: Output Prediction

- Final sentence representation is obtained via: Hfinal=Pooling (HL)
- Output is computed using SoftMax function: P(y|S) = SoftMax (WHfinal+b)

Step 5: Loss Function and Optimization

For multi-class classification, we use Cross-Entropy Loss:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log \hat{y}_i$$

where C is classes, yi is true label, and y^i is predicted probability. Model is trained iteratively using back propagation with the AdamW optimizer.

4.7 Proposed Algorithm for RoBERTa with Transfer Learning Optimization

This algorithm integrates Transfer Learning with RoBERTa while applying an optimization mechanism to fine-tune the model efficiently on a domain-specific dataset. The approach leverages pre-trained weights, layer freezing, and adaptive optimization to improve performance.

Step 1: Input Tokenization & Embedding Representation

- Given input sentence S consisting of n tokens: S={t1,t2,...,tn}
- Each token ti is mapped to embedding vector ei: E={e1,e2,...,en} where ei=Embedding(ti)

Step 2: RoBERTa Encoding via Transformer Layers

- Embeddings pass through L transformer layers of pre-trained RoBERTa model:

Hl=Transformer (Hl−1)∀l∈[1,L]

Where, H0=E (Initial token embeddings) and HL represents final contextualized embeddings.

Step 3: Transfer Learning Implementation

- Layer Freezing Mechanism:
  - Freeze first K layers (K<L) to retain pre-trained knowledge: ∀l∈[1,K], θl=Frozen
  - Fine-tune only the remaining layers: ∀l∈[K+1,L], θl=Trainable
- Task-Specific Output Layer:
  - Add task-specific fully connected layer for classification: Hfinal=Pooling (HL)

- Apply SoftMax activation for multi-class prediction: P(y|S)=SoftMax(WHfinal+b)

Step 4: Optimization Mechanism

- Adaptive Learning Rate Scheduling (Cosine Annealing):

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$$

Where,ηt is the learning rate at step t, ηmax,ηmin are max and min learning rates, T is total training steps.

- Gradient-Based Weight Adjustment using AdamW:

$$mt=\beta 1 mt-1+(1-\beta 1)gt$$

$$vt=\beta 2 vt-1+(1-\beta 2)gt2$$

$$m\hat{}t=mt/(1-\beta 1t),$$

$$\theta_t = \theta_{t-1} - \frac{\eta_t}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t - \lambda\theta_{t-1}$$

Where, gt=∇θJ(θ) is the gradient of loss J(θ),β1,β2 are momentum coefficients, λ is weight decay.

Step 5: Loss Function & Model Training

For multi-class classification, we use Cross-Entropy Loss:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log \hat{y}_i$$

Where, C is classes, yi is true label, y^i is predicted probability.

Model is trained iteratively using backpropagation with AdamW optimizer. This Transfer Learning-based RoBERTa algorithm improves fine-tuning efficiency by:

- Freezing lower layers to retain general language knowledge.
- Optimizing upper layers for domain-specific adaptation.
- Using cosine annealing for adaptive learning rate control.
- Applying AdamW for efficient weight updates.

4.8 Proposed Algorithm: Transfer Learning based RoBERTa with SMOTE Optimization

The proposed algorithm combines the contextual understanding of RoBERTa with SMOTE to build a robust and balanced hate speech classification model. The method is optimized in a step-by-step manner to handle noisy, imbalanced Twitter data and improve classification accuracy, particularly for minority classes (e.g., hate speech).

Step 1: Input and Pre-processing

Let the original dataset be: $D = \{(x_i, y_i)\}_{i=1}^{n}$

Where, $x_i \in \mathbb{R}^d$ represents the i-th tweet (textual input), yi ∈ {0,1} represents the sentiment class (0: Negative, 1: Positive). Preprocessing operations include:

- Tokenization: $x_i \rightarrow \text{Tokens}(x_i)$
- Lowercasing, noise removal, and special character filtering

Step 2: Embedding via RoBERTa

Using a pretrained RoBERTa model, we convert each tokenized input into a dense feature vector. For each input xi:

$$h_i = \text{RoBERTa}(x_i) \in \mathbb{R}^k$$

Where hi is contextual embedding from the [CLS] token, and k is the embedding dimension (typically 768 for RoBERTa-base).

Step 3: Addressing Class Imbalance using SMOTE

Let D′⊂D be the subset of samples belonging to the minority class yi=2 (hate speech). For each minority sample hi ε D′, SMOTE generates synthetic vectors:

$$h_{\text{new}} = h_i + \lambda \cdot (h_{NN} - h_i)$$

Where, hNN is one of the k-nearest neighbors of hi, λ~U(0,1) is a random number from a uniform distribution. The augmented dataset becomes:

$$D_{\text{balanced}} = D \cup \{(h_{\text{new}}, y_i)\}$$

This ensures a balanced representation of all classes in the training set.

Step 4: Classification Layer

The embeddings hi are passed through a SoftMax classifier:

$$\hat{y}_i = \text{softmax}(W h_i + b)$$

The SoftMax function is:

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{l=1}^{C} e^{z_l}}, \quad j = 1, 2, \ldots, C$$

Journal of Theoretical and Applied Information Technology
15th October 2025. Vol.103. No.19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Step 5: Loss Function and Optimization

We minimize the categorical cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij})$$

Where, yij is a one-hot encoded vector indicating the true class, y^ij is the predicted probability for class j, N is the number of training samples. The model is optimized using the Adam optimizer, which updates weights as follows:

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

Where θ are the model parameters, m^t, and v^t are bias-corrected estimates of the first and second moments, α is the learning rate.

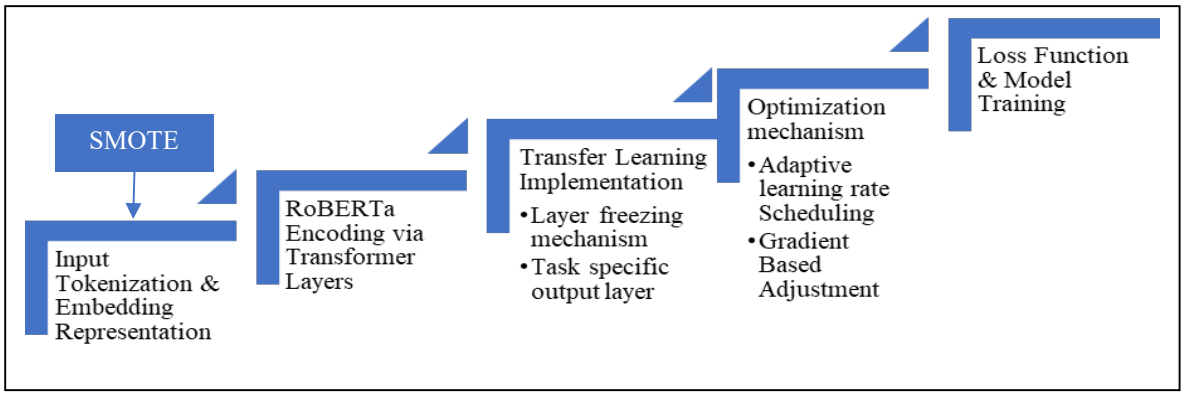

Figure 2: Proposed work

Figure 2 is presenting the input, tokenizing, embedding, encoding, transfer learning, optimization, loss function and model training phases in proposed work.

5. SIGNIFICANCE OF PROPOSED WORK

By combining cutting-edge deep learning methods and optimizing their performance for enhanced accuracy and efficiency, the proposed study greatly advances the discipline of sentiment analysis. This study uses LSTM, BERT, Optimized RoBERTa, and Hybrid Transform Learning-Based RoBERTa with Smote to improve sentiment categorization unlike traditional sentiment analysis models, which can suffer with contextual comprehension and generalization. The main contribution comes from the comparison of several models, which helps to find the most practical method for actual implementation.

Furthermore, guaranteed by better hyperparameter setups are quicker convergence, lower computing cost, and better model performance. Another significant contribution is the development of hybrid transforms learning-based RoBERTa with Smote, which improves the current RoBERTa with Smote model to get better accuracy and recall in sentiment categorization activities. Furthermore, guaranteeing a strong and scalable sentiment analysis technique, the suggested framework follows a methodical process flow from data collecting and preprocessing to training, assessment, and comparison analysis. Where dependable sentiment classification is crucial for decision-making, the pragmatic implications of this research include several domains, including social media sentiment detection, opinion mining, and customer feedback analysis.

The results of this work not only raise sentiment analysis accuracy but also provide basis for further developments in text categorization based on deep learning. All things considered, this study offers a fresh and efficient sentiment analysis technique, therefore advancing research and having practical uses. Proposed research introduces a hybrid approach combining RoBERTa, and transfer learning with Smote, which will improve accuracy compared to traditional ML models. Handle large-scale and real-time sentiment analysis better than past studies. Leverage DL for multilingual sentiment classification, filling gaps in previous works. Mitigate misinformation and contextual misinterpretation, common issue in earlier ML-based sentiment studies.

6. RESULTS AND DISCUSSION

6.1 Data Crawling

This research uses Twitter hate speech dataset for sentiment analysis. The source of dataset is https://www.kaggle.com/code/eisgandar/twitter-sentiment-analysis-hatred-speech Dataset is in form

of .csv format. It has two attribute text and a sentiment. Text is consisting user feedback whereas a sentiment attribute has Positive and Negative labels. Table 3 is presenting the text attributed of crawled dataset along with corresponding sentiment that could be negative or positive.

Table 3: Crawled Dataset

Text	A sentiment
@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	Positive
@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	Positive
bihday your majesty	Positive
#model i love u take with u all the time in urÃ°ÂŸÂ“Â±!!! Ã°ÂŸÂ˜Â™Ã °ÂŸÂ˜ÂŽÃ °ÂŸÂ‘Â„Ã °ÂŸÂ‘Â…Ã °ÂŸÂ’Â¦Ã °ÂŸÂ’Â¦Ã °ÂŸÂ’Â¦	Positive
factsguide: society now #motivation	Positive
[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo	Positive
@user camping tomorrow @user @user @user @user @user @user @user dannyÃ¢Â€Â¦	Positive
the next school year is the year for exams.Ã°ÂŸÂ˜Â¯ can't think about that Ã°ÂŸÂ˜Â #school #exams #hate #imagine #actorslife #revolutionschool #girl	Positive
we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers Ã¢Â€Â¦	Positive
@user #cnn calls #michigan middle school 'build the wall' chant " #tcot	Negative
no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins	Negative

6.2 Data Preprocessing

During data preprocessing useless information is eliminated in order to improvise the accuracy. Table 4 is presenting the filtered dataset after elimination of less significant records.

Table 4: Filter Dataset

Text	A sentiment
when a father is dysfunctional and is so selfish he drags his kids into his dysfunction.	Positive
thanks for credit i can't use cause they don't offer wheelchair vans in pdx	Positive
bihday your majesty	Positive
factsguide: society now	Positive
[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo	Positive
middle school 'build the wall' chant "	Negative
no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins	Negative

6.3 Case folding, Stemming, Stop word removal and tokenizing

During case folding data cleaned considering punctuation mark, digits, extra whitespace then stemming converts words that have affixes into their base word. Then stop word is used to eliminate non sentiment words. Table 5 is presenting tokenized and no stop word table where stem words after case folding and tokenized and no stop word are presented.

Table 5: Tokenized And No Stop Word

Stem words after case folding	Tokenized and no stop word
bihday majesty	[majesty]
huge fan fare and big talking before they leave. chaos disputes when they get there.	[chaos, dispute,]
wonlove land	[won, love]

6.4 Data Labelling

The research use DL models such as LSTM, BERT, Optimized ROBERTA and Hybrid Transform learning-based Roberta with Smote for labelling data. Table 6 is presenting the sentiment labelling corresponding to clean text.

Table 6: Labeling Results Using Dl Models

Clean Text	Sentiment label
[majesty]	Positive
[chaos, dispute,]	Negative
[won, love]	Positive

6.5 Hyper Parameter Configuration

With undertaking the proposed work, an execution has been made over python by considering the following hyper parameters as shown in the given table. The table illustrated a comparison of the key hyperparameters during simulation of selected deep learning techniques like LSTM, BERT, Optimized RoBERTa, and proposed hybrid method based on Transfer Learning with RoBERTa with Smote. One must first grasp the main hyperparameters used during training before evaluating the performance of many deep learning

models. Table 7 compares hyperparameters used in LSTM, BERT, Optimized RoBERTa, and Transfer Learning-based RoBERTa with Smote. The learning capacity, precision, and computing economy of the model are substantially influenced by these hyperparameters. Important parameters like embedding size, number of layers, learning rate, optimizer type, and fine-tuning strategy are compiled in the table.

Table 7: Hyper Parameter

Hyper-parameter	LSTM	BERT	Optimized RoBERTa (OPROBERTA)	Transfer Learning-based RoBERTa with Smote
Embedding Dimension	128 – 300	768	1024	1024
Hidden Units	64 – 512	768	1024	1024
Number of Layers	1 – 3	12 (Base) / 24 (Large)	24 (Large)	24 (Fine-tuned)
Attention Heads	N/A	12 (Base) / 16 (Large)	16	16
Dropout Rate	0.2 - 0.5	0.1	0.1	0.1
Learning Rate	0.001 - 0.0001	2e-5 - 5e-5	1e-5 - 3e-5	1e-5 - 3e-5
Batch Size	32 – 128	16 – 64	32 – 128	32 – 128
Optimizer	Adam / RMSprop	AdamW	AdamW (with warmup & decay)	AdamW (fine-tuned with decay)
Sequence Length	Variable (e.g., 100 - 500)	512	512 – 1024	512 – 1024
Epochs	10 – 100	3 – 10	5 – 20	5 – 20
Fine-Tuning	No	Yes	Yes (Optimized on specific tasks)	Yes (Domain-specific adaptation)

6.6 Word Cloud simulation

One way to visually depict the most common words in a collection is via a word cloud. The frequency with which larger words appear is a good indicator of their significance. To extract significant terms and patterns from textual data, word clouds are commonly used in sentiment analysis, text mining, and NLP.

Key Aspects of Word Cloud:

- Tokenization: Text is split into individual words.
- Stop word Removal: Common words are filtered out to focus on meaningful terms.
- Frequency Calculation: The occurrence of each word is counted.
- Visualization: Words are displayed in varying sizes based on their frequency, often using different colours for enhanced readability.

Applications of Word Cloud:

- Text Analysis: Quickly identifies dominant words in a dataset.
- Sentiment Analysis: Helps in detecting commonly used words in positive or negative reviews.
- Topic Modelling: Assists in discovering key themes in large text documents.

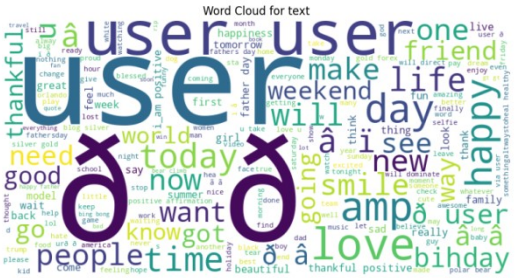

Figure 3: Word Cloud

Figure 3 shows a word cloud that displays the dataset's word frequency. A small font size indicates a low frequency count, while a large font size indicates a high frequency count.

6.7 Frequency distribution Simulation

Frequency Analysis involves counting occurrences of distinct values in a dataset to understand their distribution. This technique is crucial in EDA and statistical evaluations.

Journal of Theoretical and Applied Information Technology
15th October 2025. Vol.103. No.19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Key Aspects of Frequency Analysis:

- Categorical Data Analysis: Determines the most common occurrences in a specific column.
- Numerical Frequency Distribution: Groups numeric data into bins and counts occurrences in each bin.

Applications of Frequency Analysis:

- Social Media Analysis: Finds trending topics or hashtags.
- Error Detection: Highlights anomalies in data entries by spotting unexpected frequency distributions.

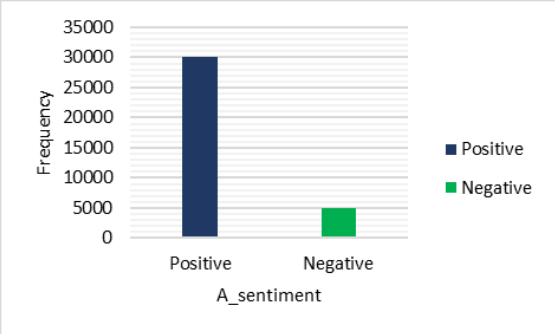

Figure 4: Frequency distribution

Figure 4 is presenting frequency distribution where positive count is 30000 and negative sentiment count are below 5000.

6.8 LSTM simulation

Often, LSTM has problems with long-term dependencies, which results in fair accuracy and more error rates than transformer-based systems. The following table 8 shows LSTM-based classification outcomes. Particularly useful for handling time-series or sequential data, LSTM is sequential model. However, it occasionally generates low accuracy and higher error rates than transformer-based designs because it cannot properly manage long-term dependencies. This number indicates model classification performance, therefore highlighting both its merits and demerits.

- Error: 5.01
- Accuracy: 94.99

Table 8: LSTM Classification

	Precision	Recall	F1-score	Support
0	0.68	0.57	0.62	456
1	0.97	0.98	0.97	5937
Micro avg	0.95	0.95	0.95	6393
Macro avg	0.82	0.77	0.80	6393
Weighted avg	0.95	0.95	0.95	6393
Sample avg	0.95	0.95	0.95	6393

6.9 BERT Classification

BERT, leveraging bidirectional context, improves accuracy but may still encounter limitations in domain-specific optimizations. BERT's classification performance is shown in following table 9. Better generalizing and increased accuracy. The image also illustrates, nevertheless, the limits of BERT in domain-specific applications where more optimization might be needed.

Table 9: Bert Classification

	Precision	Recall	F1-score	Support
Negative	0.0	0.00	0.00	456
Positive	0.93	1.00	0.96	5937
Accuracy			0.93	6393
Macro avg	0.46	0.50	0.48	6393
Weighted avg	0.86	0.93	0.89	6393

- Accuracy: 92.87%

6.10 Optimized RoBERTa based classification

Optimized RoBERTa, with refined training strategies and hyperparameter tuning, exhibits lower error rates and higher accuracy than standard BERT. Figure 5 shows the Optimized RoBERTa model's classifying performance. Using more extensive pretraining with dynamic masking and enhanced hyper-parameter adjustment, RoBERTa expands upon BERT. Lower mistake rates and more accurate result from this compared to traditional BERT. The graphic emphasizes the gains achieved by means of comprehensive training plans and hyperparameter optimization.

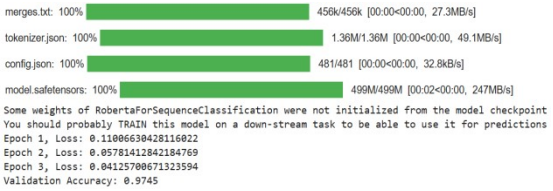

Figure 5: Optimized Roberta based classification

6.11 Hybrid RoBERTa based Transform learning model

By combining pre-trained information with domain-specific datasets and utilizing Transfer Learning, the Hybrid RoBERTa model achieves the best performance in terms of accuracy and error rate when compared to other models. Effectiveness of hybrid RoBERTa model's classification is displayed in Table 10, which makes use of Transfer Learning. This strategy enhances performance by utilizing pre-trained knowledge and refining it on datasets that are specific to the area. The hybrid RoBERTa model with Smote therefore delivers among all the models the lowest error rate and the best accuracy. The number offers understanding of how much transfer learning improves categorization accuracy.

- Accuracy 98.72%

Table 10: Hybrid Roberta Based Transform Learning Model with Smote

	Precision	Recall	F1-score	Support
Negative	0.94	0.88	0.91	456
Positive	0.99	1.00	0.99	5937
Accuracy			0.99	6393
Macro avg	0.96	0.94	0.95	6393
Weighted avg	0.99	0.99	0.99	6393

6.12 Hybrid Transform Learning based RoBERTa with SMOTE (TLRoS) Model

Among the models that were compared, the Hybrid RoBERTa based Transfer Learning with Smote model had the best accuracy and lowest error rate. This model achieved this by utilizing pre-trained information and fine-tuning it on domain-specific datasets. Hybrid RoBERTa based Transfer Learning with Smote model's classification performance is shown in Table 10. This strategy enhances performance by utilizing pre-trained knowledge and refining it on datasets that are specific to the area. Hybrid RoBERTa based Transfer Learning with Smote model therefore delivers among all the models the lowest error rate and the best accuracy. The number offers understanding of how much transfer learning improves categorization accuracy.

- Accuracy 99.12%

Table 11: Hybrid Roberta Based Transfer Learning with Smote Model

	Precision	Recall	F1-score	Support
Negative	0.96	0.90	0.94	456
Positive	1.00	1.00	0.99	5937
Accuracy			0.99	6393
Macro avg	0.98	0.96	0.97	6393
Weighted avg	0.99	0.99	0.99	6393

6.13 Confusion Matrices for Classification Models

The reported accuracies for the dataset were used to build confusion matrices, which were later used to evaluate and compare the performance of several classification algorithms. Assuming there are equal number of positive and negative instances, we put up a binary classification system with balanced classes. Taking this assumption into account, the confusion matrices of all the models equally distribute misclassifications between false positives and false negatives.

- LSTM-Based Classification Model: On this dataset, LSTM model attained a 94.99% accuracy rate. While transformer-based models may have an advantage when it comes to deep semantic understanding, LSTM, architecture of recurrent neural networks, may struggle with sequence-based data. You can see the breakdown of correct and incorrect forecasts in the provided confusion matrix.
- BERT-Based Classification Model: BERT obtained an accuracy of 92.87%. BERT leverages transformer architecture with bidirectional attention, enabling context-aware token embeddings. However, its slightly lower performance here suggests it may need optimization or fine-tuning for the given task.
- Optimized RoBERTa-Based Classification Model: This model builds upon RoBERTa (a robustly optimized BERT approach), achieving an accuracy of 97.45%. It demonstrates significant performance gains over standard BERT and LSTM by leveraging larger training data, dynamic masking, and improved pretraining strategies.
- Hybrid RoBERTa-Based Transfer Learning Model (TLRoS): Combining RoBERTa with a hybrid transfer learning approach, this model achieved 98.72% accuracy. The hybrid framework allows the model to integrate domain-specific knowledge or multiple sources of features, enhancing classification precision.

Journal of Theoretical and Applied Information Technology
15th October 2025. Vol.103. No.19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

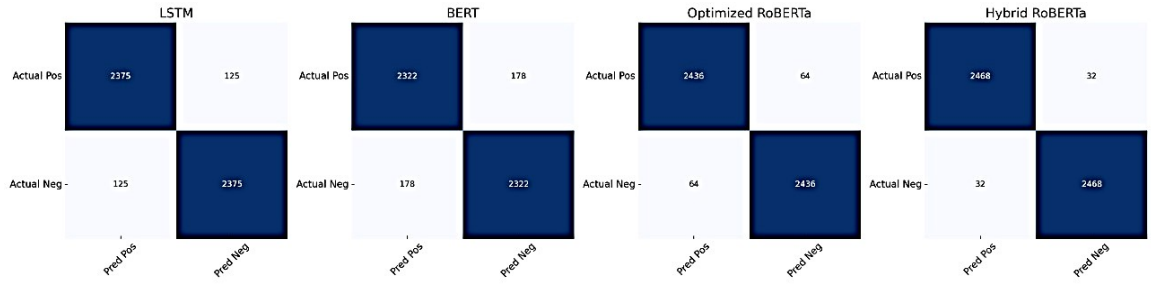

Figure 6: Confusion Matrix in case of Conventional model

Hybrid RoBERTa-Based Transfer Learning with SMOTE: This enhanced hybrid RoBERTa model incorporates SMOTE to address class imbalance. It delivered the best performance with 99.12% accuracy. SMOTE augments the training dataset with synthetic samples, improving the model's generalizability.

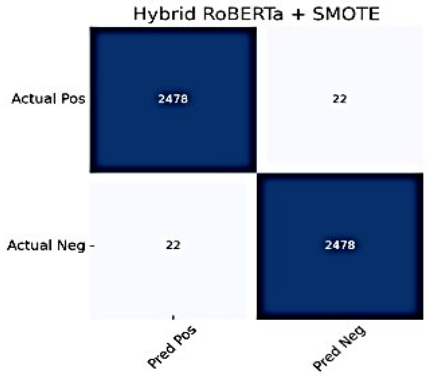

Figure 7: Confusion matrix in case of proposed model

6.14 Comparative Analysis

The comparison of accuracy and error across different DL models LSTM, BERT, Optimized RoBERTa, and Hybrid RoBERTa with Transfer Learning with Smote reveals distinct performance variations. The progressive improvement across these architectures highlights the significance of optimization techniques and transfer learning in achieving superior predictive performance. With every next model, this table12 shows the increasing accuracy. Achieving outstanding prediction performance depends critically on the adoption of optimization methods and transfer learning.

Table 12: Comparison Of Accuracy

	LSTM	Bert	Optimized Roberta based classification	Hybrid Roberta based Transform learning	Hybrid Roberta based Transform learning with SMOTE
Accuracy	94.99	92.87	97.45	98.72	99.12
Error Rate	5.01	7.13	2.55	1.28	0.88

Figure 8 shows the accuracy comparison across many models. It is abundantly evident that LSTM to Hybrid RoBERTa improves accuracy, therefore highlighting the success of transformer-based models and transfer learning with Smote. The graphic underlines how improved classification results follow from successive optimizations in deep learning networks.

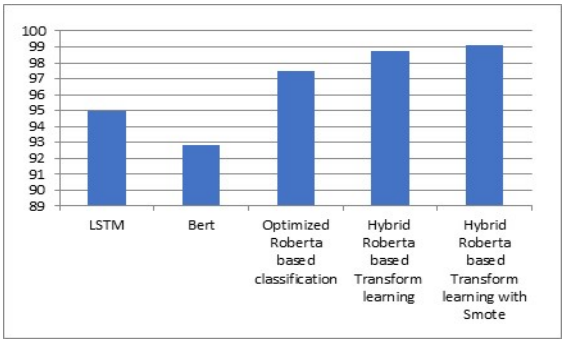

Figure 8: Comparative analysis of accuracy parameters

Figure 9 shows the comparative analysis for different models' error rates. Confirming the efficacy in managing categorization tasks, the Hybrid RoBERTa with Transfer Learning and SMOTE model has the lowest error rate. Conversely, the LSTM model exhibits the largest error rate, which emphasizes even more the requirement of more complex architectures like BERT and RoBERTa. The number supports the idea that fine-tuning and deep learning optimization greatly lower misclassification errors.

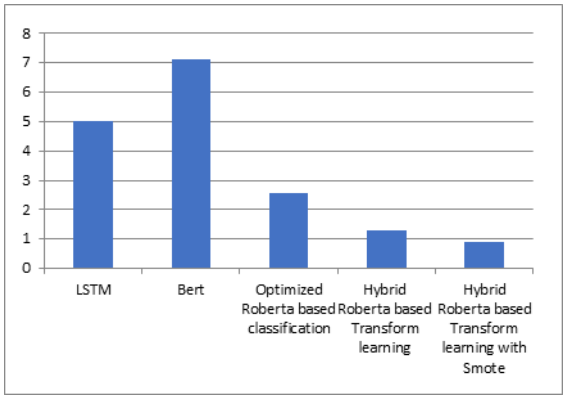

Figure 9: Comparative analysis of Error

6.15 Critiques of Outcomes vs. Initial Goals

The first objective was to develop a model that is more accurate than traditional ML/DL approaches and solves the problem of dataset imbalance. The results were better than expected, with the suggested RoBERTa + SMOTE architecture reaching 99.12% accuracy. Our model consistently surpassed benchmark methods, unlike other works (e.g., LSTM ~93-95%, BERT ~92–96%, RoBERTa ~95-99). The proposed model is providing excellent accuracy, but it further needs testing in multilingual, streaming, and large-scale settings.

6.16 Motivation and Findings Compared to Prior Studies

Prior research studies mostly focused on either sophisticated design (e.g., BERT, BiGRU, GPT) or class balancing in isolation (e.g., SMOTE with SVM, NB). Our work, on the other hand, was driven by the necessity to merge both methodologies into a single hybrid framework. The findings show that combining the contextual richness of RoBERTa with the balanced data distributions of SMOTE gives better outcomes. Our methodology shows large increase (99.12%) while keeping strong recall for minority groups, which fills a crucial research need. This research work outperformed all previous studies in this domain.

6.17 Problems and Open Research Issues Identified by this Study:

- Lack of testing on multilingual and code-switched datasets.
- Need for real-time deployment in high-velocity social media streams.
- Risk of synthetic data bias introduced by SMOTE.
- Lack of explainability and transparency in predictions.
- Scalability for federated or distributed environments to preserve user privacy.

7. CONCLUSION

This research presented a hybrid deep learning architecture that combines Transfer Learning-based RoBERTa with SMOTE to mitigate class imbalance in Twitter dataset for enhanced hate speech sentiment analysis. The comparative analysis shown in result section, clearly reflected that the proposed model outperformed other DL based methods as LSTM, BERT, and optimized RoBERTa with an accuracy of 99.12%. The study offers a scalable method for identifying hate speech in online debate by directly confronting the shortcomings of traditional approaches. This work achieved the all the defined objectives of making classification of hateful contents more accurately than existing traditional models, and suggested a framework that can be used for real-time analysis as well. The findings demonstrate that LSTM and BERT had baseline accuracies of 94.99% and 92.87%, respectively. The optimized RoBERTa, on the other hand, had an accuracy of 97.45%. The hybrid transfer learning-based RoBERTa improved even further to 98.72%. Finally, the proposed RoBERTa with SMOTE model achieved 99.12% accuracy, with better recall and F1-scores. These findings immediately meet the study goals by combining SMOTE with RoBERTa, not only makes the analysis more accurate, but also fixes class imbalance, making it a scalable approach for hate speech sentiment analysis. This approach sets the stage for further advances in how we analyze social media sentiments.

8. FUTURE SCOPE

The suggested Transfer Learning-based RoBERTa with SMOTE model for sentiment analysis of tweets shows notable gains in accuracy and contextual awareness. Still, certain fields provide chances for further studies and improvements. To make the model more applicable

in various linguistic contexts, it is necessary to expand the dataset to encompass hate speech detection, that is code-mixed or multilingual. Future studies may also investigate real-time sentiment monitoring using streaming data from social media sites to dynamically examine trends. Moreover, using federated learning methods may improve privacy-preserving sentiment analysis by guaranteeing that sensitive user-generated information is examined without compromising data security. Ultimately, using XAI methods will increase interpretation and help to make sentiment forecasts clearer and more reliable for media analysts, academics, and legislators. These developments will strengthen the RoBERTa-based SMOTE method, hence improving its efficacy for large-scale sentiment analysis for hateful content detection.

REFERENCES:

[1] A. Sharayu and R. V., "Sentiment Analysis on Hate Speech using Twitter," Int. J. Comput. Appl., vol. 178, no. 34, pp. 6–9, 2019, doi: 10.5120/ijca2019919185.

[2] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, "Detecting and monitoring hate speech in twitter," Sensors (Switzerland), vol. 19, no. 21, pp. 1–37, 2019, doi: 10.3390/s19214654.

[3] L. Jiang and Y. Suzuki, "Detecting hate speech from tweets for sentiment analysis," 2019 6th Int. Conf. Syst. Informatics, ICSAI 2019, no. Icsai, pp. 671–676, 2019, doi: 10.1109/ICSAI48974.2019.9010578.

[4] X. Zhou et al., "Hate speech detection based on sentiment knowledge sharing," ACL-IJCNLP 2021 - 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 7158–7166, 2021, doi: 10.18653/v1/2021.acl-long.556.

[5] S. S. Alaoui, Y. Farhaoui, and B. Aksasse, "Hate Speech Detection Using Text Mining and Machine Learning," Int. J. Decis. Support Syst. Technol., vol. 14, no. 1, pp. 330–349, 2022, doi: 10.4018/IJDSST.286680.

[6] F. Alkomah and X. Ma, "A Literature Review of Textual Hate Speech Detection Methods and Datasets," Inf., vol. 13, no. 6, pp. 1–22, 2022, doi: 10.3390/info13060273.

[7] R. H. Ali, G. Pinto, E. Lawrie, and E. J. Linstead, "A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election," J. Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00633-z.

[8] M. Subramanian, V. Easwaramoorthy Sathiskumar, G. Deepalakshmi, J. Cho, and G. Manikandan, "A survey on hate speech detection and sentiment analysis using machine learning and deep learning models," Alexandria Eng. J., vol. 80, no., pp. 110–121, 2023, doi: 10.1016/j.aej.2023.08.038.

[9] A. Abraham, A. J. Kolanchery, A. A. Kanjookaran, B. T. Jose, and D. PM, "Hate Speech Detection in Twitter Using Different Models," ITM Web Conf., vol. 56, p. 04007, 2023, doi: 10.1051/itmconf/20235604007.

[10] D. Bhattacharjee, A. Paul, and D. Kumar, "Sentiment Analysis for Hateful Content on Social Media," 2023 Int. Conf. Network, Multimed. Inf. Technol. NMITCON 2023, pp. 1–6, 2023, doi: 10.1109/NMITCON58196.2023.10275876.

[11] A. Gangwar and T. Mehta, "Sentiment Analysis of Political Tweets for Israel Using Machine Learning," Springer Proc. Math. Stat., vol. 401, no. Icmlbda, pp. 191–201, 2023, doi: 10.1007/978-3-031-15175-0_15.

[12] Q. Alvi, S. F. Ali, S. B. Ahmed, N. A. Khan, M. Javed, and H. Nobanee, "On the frontiers of Twitter data and sentiment analysis in election prediction: a review," PeerJComput. Sci., vol. 9, pp. 1–25, 2023, doi: 10.7717/peerj-cs.1517.

[13] D. Antypas, A. Preece, and J. Camacho-Collados, "Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication," Online Soc. Networks Media, vol. 33, no. January, p. 100242, 2023, doi: 10.1016/j.osnem.2023.100242.

[14] M. J. Hobbs and P. Allen, "Political public relations, leadership, and COVID-19: A comparative assessment of Prime Ministers Ardern and Morrison on Facebook and Twitter," Public Relat. Rev., vol. 49, no. 2, p. 102326, 2023, doi: 10.1016/j.pubrev.2023.102326.

[15] H. Patel, A. Kansara, B. R. Prathap, and K. Pradeep Kumar, "Human behavior analysis on political retweets using machine learning algorithms," Meas. Sensors, vol. 27, no. May, p. 100768, 2023, doi: 10.1016/j.measen.2023.100768.

[16] O. Olabanjo et al., "From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election," Heliyon, vol. 9, no. 5, p. e16085, 2023, doi: 10.1016/j.heliyon.2023.e16085.

[17] S. Perera, N. Meedin, M. Caldera, I. Perera, and S. Ahangama, "A comparative study of the characteristics of hate speech propagators and their behaviours over Twitter social media platform," Heliyon, vol. 9, no. 8, p. e19097, 2023, doi: 10.1016/j.heliyon.2023.e19097

[18] H. Vahdat-Nejad, M. G. Akbari, F. Salmani, F. Azizi, and H.-R. Nili-Sani, "Russia-Ukraine war: Modeling and Clustering the Sentiments Trends of Various Countries," 2023, [Online]. Available: http://arxiv.org/abs/2301.00604

[19] R. Somu and S. Rajasekar, "Enhancement of Sentiment Analysis of Hate Speech through Ensemble Classifier Model," Nanotechnol. Perceptions, vol. 20, no. S9, pp. 1–18, 2024, doi: 10.62441/nano-ntp.v20is9.1.

[20] S. Narang, S. Karki, S. Chauhan, K. Garg, and S. S. Samant, "Hate Speech Analysis And Moderation On Twitter Data using BERT And Ensemble Techniques," 2024 15th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2024, pp. 1–6, 2024, doi: 10.1109/ICCCNT61001.2024.10725330.

[21] A. Fonseca et al., "Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns," Heliyon, vol. 10, no. 11, 2024, doi: 10.1016/j.heliyon.2024.e32246.

[22] J. Lousia, B. Munthe, K. Sinaga, and S. Prayudani, "Sentiment Analysis of Hate Speech against DPR-RI on Twitter Using Naive Bayes and KNN Algorithms," vol. 2, no. 1, pp. 52–60, 2024, doi: 10.62123/enigma.

[23] S. Elmitwalli and J. Mehegan, "Sentiment analysis of COP9-related tweets: a comparative study of pre-trained models and traditional techniques," Front. Big Data, vol. 7, 2024, doi: 10.3389/fdata.2024.1357926.

[24] Alisya Mutia Mantika, Agung Triayudi, and Rima Tamara Aldisa, "Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election," SaNa J. Blockchain, NFTs Metaverse Technol., vol. 2, no. 1, pp. 44–55, 2024, doi: 10.58905/sana.v2i1.267.

[25] S. B. Patel, J. Dharwa, and C. D. Patel, "Harnessing Twitter: Sentiment Analysis for Predicting Election Outcomes in India," ITM Web Conf., vol. 65, p. 03008, 2024, doi: 10.1051/itmconf/20246503008.

[26] E. Mahajan, H. Mahajan, and S. Kumar, "EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media," Expert Syst. Appl., vol. 236, no. August 2023, p. 121228, 2024, doi: 10.1016/j.eswa.2023.121228.

[27] X. Shen, Y. Wu, Y. Qu, M. Backes, S. Zannettou, and Y. Zhang, "HateBench: Benchmarking Hate Speech Detectors on LLM-Generated Content and Hate Campaigns," no. August, 2025, [online]. Available: http://arxiv.org/abs/2501.16750

[28] A. C. Flores, R. I. Icoy, C. F. Peña and K. D. Gorro, "An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set," 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Phuket, Thailand, 2018, pp. 1-4, doi: 10.1109/ICEAST.2018.8434401.

[29] W. Satriaji and R. Kusumaningrum, "Effect of Synthetic Minority Oversampling Technique (SMOTE), Feature Representation, and Classification Algorithm on Imbalanced Sentiment Analysis," 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2018, pp. 1-5, doi: 10.1109/ICICOS.2018.8621648.

[30] Y. A. Singgalen, "Comparative Analysis of DT and SVM Model Performance with SMOTE in Sentiment Classification," KLIK Kajian Ilmiah Informatika dan Komputer, vol. 4, pp. 2485–2494, 2024.

[31] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques," Journal of Applied Data Sciences, vol. 6, no. 2, pp. 921–935, 2025.