IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

Comparative Sentiment Analysis: Classical Machine Learning Vs. Transformer Model

Himanshu Shaileshbhai Ramani
MCA Student, Department of Computer Science
Centre of Distance and Online Education University of Mumbai, Mumbai, India

Abstract: *This study presents a comprehensive comparative analysis of sentiment classification approaches on social media data, contrasting traditional machine learning methods with modern transformer-based architectures. We evaluate Logistic Regression as a representative classical approach against DistilBERT, a lightweight transformer model, using the Sentiment140 dataset comprising 30,000 balanced tweets. The classical approach employs TF-IDF vectorization with standard preprocessing techniques, while the transformer model leverages contextual embeddings and attention mechanisms. Our experimental results demonstrate that DistilBERT achieves superior performance with 82% accuracy compared to Logistic Regression's 77%, showing particular strength in handling nuanced expressions, sarcasm, and context-dependent sentiment. The transformer model exhibits more balanced precision-recall characteristics and better generalization across diverse linguistic patterns. However, the classical approach offers advantages in computational efficiency, interpretability, and deployment simplicity. This research provides practical insights for practitioners in selecting appropriate sentiment analysis approaches based on specific requirements, resource constraints, and performance objectives in real-world applications*

Keywords: Sentiment Analysis, Logistic Regression, DistilBERT, Natural Language Processing, Transformer Models, Social Media Analysis.

I. INTRODUCTION

In the contemporary digital landscape, social media platforms have emerged as primary channels for public discourse, generating unprecedented volumes of user-generated content that reflect opinions, emotions, and attitudes toward products, services, events, and social issues. The ability to automatically analyze and understand sentiment expressed in this textual data has become increasingly critical for businesses, researchers, and policymakers seeking to gauge public opinion, monitor brand reputation, understand customer satisfaction, and make data-driven decisions.

Sentiment analysis, also known as opinion mining, represents a fundamental task in Natural Language Processing (NLP) that aims to computationally identify and extract subjective information from text. The field has witnessed remarkable evolution over the past two decades, progressing from simple lexicon-based approaches to sophisticated deep learning architectures. Traditional machine learning methods, including Naive Bayes, Support Vector Machines, and Logistic Regression, dominated the field for years, offering interpretable models with reasonable performance on well-structured text. These approaches typically rely on hand-crafted features such as bag-of-words representations or TF-IDF vectors, combined with statistical learning algorithms.

However, the emergence of deep learning and particularly transformer-based architectures has revolutionized NLP tasks, including sentiment analysis. Models like BERT (Bidirectional Encoder Representations from Transformers) and its variants have demonstrated unprecedented performance by capturing contextual relationships and semantic nuances that traditional methods often miss. These models leverage attention mechanisms and pre-training on massive text corpora to develop rich linguistic representations.

Despite the impressive capabilities of transformer models, classical machine learning approaches continue to offer distinct advantages in terms of computational efficiency, interpretability, and ease of deployment, particularly in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

resource-constrained environments. This creates a practical dilemma for practitioners: when should one opt for the simplicity and efficiency of traditional methods versus the superior performance of complex transformer architectures?

This research addresses this question through a systematic comparative analysis of Logistic Regression, representing classical machine learning approaches, and DistilBERT, a lightweight transformer model, on social media sentiment classification. Our study aims to provide empirical evidence and practical insights regarding the trade-offs between these approaches, examining not only accuracy metrics but also computational requirements, implementation complexity, and suitability for different application scenarios.

The primary contributions of this work include: (1) a rigorous experimental comparison of classical and transformer-based approaches on identical social media data, (2) detailed analysis of performance characteristics including handling of linguistic nuances, (3) evaluation of computational and practical considerations for real-world deployment, and (4) evidence-based recommendations for approach selection based on specific use case requirements.

II. LITERATURE REVIEW

The evolution of sentiment analysis methodologies reflects the broader progression of Natural Language Processing and machine learning techniques. Early approaches to sentiment analysis relied primarily on lexicon-based methods, which utilized predefined dictionaries of words associated with positive or negative sentiment. Systems like VADER (Valence Aware Dictionary and sEntiment Reasoner) and SentiWordNet provided sentiment scores for words and phrases, enabling rule-based classification. While these methods offered high interpretability and required no training data, they struggled with context-dependent meanings, domain-specific language, and complex linguistic phenomena such as sarcasm and negation.

The introduction of machine learning approaches marked a significant advancement in sentiment analysis capabilities. Pang and Lee's seminal work in 2002 demonstrated that supervised learning algorithms, particularly Naive Bayes and Support Vector Machines, could effectively classify movie reviews by learning patterns from labeled data. Subsequent research established Logistic Regression as a robust baseline for text classification tasks, offering a good balance between performance and interpretability. These classical approaches typically employ feature extraction techniques such as bag-of-words, n-grams, or TF-IDF (Term Frequency-Inverse Document Frequency) to convert text into numerical representations suitable for statistical learning algorithms.

The advent of word embeddings represented another paradigm shift in NLP. Mikolov et al.'s Word2Vec and Pennington et al.'s GloVe introduced dense vector representations that captured semantic relationships between words based on their distributional properties in large text corpora. These embeddings enabled more sophisticated feature representations compared to sparse bag-of-words approaches, leading to improved performance in various NLP tasks including sentiment analysis.

Deep learning architectures further transformed the field by enabling end-to-end learning of both feature representations and classification functions. Recurrent Neural Networks (RNNs) and their variants, particularly Long Short-Term Memory (LSTM) networks, demonstrated the ability to capture sequential dependencies and long-range contextual information in text. Kim's work on Convolutional Neural Networks for sentence classification showed that CNNs could effectively extract local features and patterns relevant for sentiment analysis.

The introduction of attention mechanisms and the transformer architecture by Vaswani et al. in 2017 revolutionized NLP by enabling models to weigh the importance of different words in context dynamically. This breakthrough led to the development of BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al., which achieved state-of-the-art results across numerous NLP benchmarks through pre-training on massive text corpora followed by task-specific fine-tuning. BERT's bidirectional context understanding and transfer learning capabilities addressed many limitations of previous approaches.

However, BERT's computational requirements posed challenges for practical deployment. This motivated the development of more efficient variants, including DistilBERT by Sanh et al., which retained approximately 97% of BERT's performance while reducing model size by 40% and increasing inference speed through knowledge distillation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, June 2026 Impact Factor: 8.2

Other efficient variants like ALBERT, RoBERTa, and TinyBERT have similarly aimed to balance performance with computational efficiency.

Recent comparative studies have examined the trade-offs between different sentiment analysis approaches. Research has shown that while transformer models generally achieve superior performance, classical machine learning methods remain competitive on certain datasets and offer advantages in interpretability and computational efficiency. However, comprehensive comparisons specifically focused on social media data, which presents unique challenges including informal language, abbreviations, and limited context, remain relatively sparse in the literature.

Social media sentiment analysis presents distinct challenges compared to traditional text classification tasks. Twitter data, in particular, features character limitations, heavy use of hashtags and mentions, informal grammar, and frequent sarcasm or irony. Studies have shown that these characteristics can significantly impact model performance, with some research suggesting that the benefits of complex models may be more pronounced in such challenging domains.

This literature review reveals that while substantial research has examined both classical and transformer-based approaches independently, there remains a need for systematic comparative analyses that evaluate these methods on identical social media datasets with careful attention to practical deployment considerations. Our study addresses this gap by providing a rigorous empirical comparison that considers not only accuracy metrics but also computational requirements, implementation complexity, and real-world applicability.

III. METHODOLOGY

III.1 Dataset Selection and Characteristics

For this comparative study, we utilized the Sentiment140 dataset, a widely-recognized benchmark for Twitter sentiment analysis. The dataset originally contains 1.6 million tweets automatically labeled based on emoticons, where tweets containing positive emoticons (e.g., :), :-)) are labeled as positive (label 4) and those with negative emoticons (e.g., :(, :-() are labeled as negative (label 0). For our experiments, we extracted a balanced subset of 30,000 tweets, comprising 15,000 positive and 15,000 negative samples, to ensure unbiased model training and evaluation.

The Sentiment140 dataset presents several characteristics typical of social media text: informal language, abbreviations, misspellings, hashtags, user mentions, URLs, and emoticons. These features make it an ideal testbed for comparing sentiment analysis approaches, as they reflect the real-world challenges encountered in social media monitoring applications. The binary classification task (positive vs. negative sentiment) provides a clear framework for performance comparison while maintaining practical relevance.

III.2 Data Preprocessing

Data preprocessing represents a critical step that can significantly impact model performance. We implemented preprocessing pipelines tailored to each approach while maintaining consistency in fundamental cleaning operations.

Common Preprocessing Steps:

- Conversion of all text to lowercase to reduce vocabulary size and improve generalization
- Removal of URLs using regular expressions, as they typically do not contribute sentiment information
- Removal of user mentions (@username) to focus on content rather than social network structure
- Elimination of special characters and punctuation marks
- Removal of extra whitespace and normalization of spacing

Classical ML-Specific Preprocessing:

For the Logistic Regression model, we implemented additional preprocessing steps common in traditional NLP pipelines:

- Removal of common English stopwords (e.g., "the", "is", "at") using NLTK's stopword list
- Retention of emoticons before other preprocessing, as they carry strong sentiment signals
- Tokenization using whitespace and punctuation boundaries

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

Transformer-Specific Preprocessing:

For DistilBERT, we employed minimal preprocessing to preserve contextual information:

- Retention of sentence structure and word order, which are crucial for transformer models
- Preservation of certain punctuation marks that may carry sentiment information
- Use of DistilBERT's native tokenizer, which employs WordPiece tokenization and handles unknown words through subword units

III.3 Feature Extraction and Representation

Classical ML Approach - TF-IDF Vectorization:

For Logistic Regression, we employed Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to convert preprocessed text into numerical feature vectors. TF-IDF weights terms based on their frequency in a document relative to their frequency across the entire corpus, effectively highlighting distinctive words while downweighting common terms.

Configuration parameters:

- Maximum features: 5,000 (to balance between information retention and computational efficiency)
- N-gram range: (1, 2) (capturing both individual words and bigrams)
- Minimum document frequency: 2 (excluding extremely rare terms)
- Maximum document frequency: 0.95 (excluding ubiquitous terms)

This configuration resulted in a sparse matrix representation where each tweet is represented as a 5,000-dimensional vector, with most values being zero.

Transformer Approach - Contextual Embeddings:

DistilBERT generates contextualized word embeddings through its transformer architecture. Unlike static embeddings, these representations vary based on surrounding context, enabling the model to disambiguate word meanings and capture nuanced semantic relationships.

The model processes input through:

- WordPiece tokenization, breaking words into subword units
- Addition of special tokens ([CLS] at the beginning, [SEP] at the end)
- Conversion of tokens to input IDs using the pre-trained vocabulary
- Generation of attention masks to distinguish actual tokens from padding
- Truncation or padding to a maximum sequence length of 128 tokens

III.4 Model Architecture and Configuration

Logistic Regression Model:

We implemented Logistic Regression using scikit-learn's LogisticRegression class with the following configuration:

- Solver: 'liblinear' (efficient for binary classification with L2 regularization)
- Regularization parameter (C): 1.0 (standard regularization strength)
- Maximum iterations: 1000 (ensuring convergence)
- Class weight: balanced (to handle any minor class imbalances)

Logistic Regression models the probability of a tweet belonging to the positive class using the logistic function applied to a linear combination of TF-IDF features. The model learns optimal weights for each feature through maximum likelihood estimation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

DistilBERT Model:

We utilized the pre-trained 'distilbert-base-uncased' model from Hugging Face's Transformers library, fine-tuning it for binary sentiment classification. DistilBERT consists of 6 transformer layers (compared to BERT's 12), with 768-dimensional hidden states and 12 attention heads per layer.

Architecture components:

- Pre-trained DistilBERT base as the feature extractor
- Dropout layer (p=0.3) for regularization
- Linear classification head mapping 768-dimensional representations to 2 classes
- Softmax activation for probability distribution over classes

The model leverages transfer learning, starting from weights pre-trained on large text corpora and adapting them to the sentiment classification task through fine-tuning.

III.5 Training Strategy and Hyperparameters

Logistic Regression Training:

Training the Logistic Regression model involved fitting the model to the TF-IDF transformed training data using the liblinear solver. The training process optimizes the log-likelihood objective function with L2 regularization. Given the convex nature of the optimization problem and the relatively small feature space (5,000 dimensions), training completed rapidly, typically within seconds on standard hardware.

DistilBERT Training:

Fine-tuning DistilBERT required more sophisticated training procedures:

Hyperparameters:

- Learning rate: 2e-5 (small learning rate to preserve pre-trained knowledge)
- Batch size: 16 (balancing memory constraints and gradient stability)
- Number of epochs: 3 (sufficient for fine-tuning without overfitting)
- Optimizer: AdamW (Adam with weight decay for better regularization)
- Learning rate schedule: Linear decay with warmup
- Warmup steps: 500 (gradual learning rate increase at training start)
- Weight decay: 0.01 (L2 regularization)

Training procedure:

- Division of training data into batches
- Forward pass through DistilBERT to obtain predictions
- Calculation of cross-entropy loss
- Backward propagation to compute gradients
- Parameter updates using AdamW optimizer
- Learning rate adjustment according to schedule

We implemented early stopping based on validation set performance to prevent overfitting, monitoring validation loss after each epoch.

III.6 Data Splitting Strategy

We employed a stratified train-test split to ensure balanced class representation in both subsets:

- Training set: 80% (24,000 tweets, 12,000 positive and 12,000 negative)
- Test set: 20% (6,000 tweets, 3,000 positive and 3,000 negative)

Stratification ensured that both sets maintained the 50-50 class distribution, preventing bias in model evaluation. We used a fixed random seed (42) for reproducibility of results across experiments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, June 2026 Impact Factor: 8.2

For DistilBERT training, we further split the training set to create a validation set (10% of training data) for monitoring training progress and implementing early stopping.

III.7 Evaluation Metrics

We employed multiple evaluation metrics to comprehensively assess model performance:

Accuracy: The proportion of correctly classified tweets among all predictions. While accuracy provides an overall performance measure, it can be misleading with imbalanced datasets (though our dataset is balanced).

Precision: The proportion of true positive predictions among all positive predictions. High precision indicates few false positives.

Recall (Sensitivity): The proportion of actual positive tweets correctly identified. High recall indicates few false negatives.

F1-Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns. F1-score is particularly useful when class distribution or error costs are uneven.

Confusion Matrix: A table showing true positives, true negatives, false positives, and false negatives, providing detailed insight into model behavior and error patterns.

These metrics enable comprehensive comparison of model performance across different dimensions, revealing not just overall accuracy but also specific strengths and weaknesses in handling positive and negative sentiment.

III.8 Implementation Environment

All experiments were conducted in a consistent computational environment to ensure fair comparison:

Hardware:

- Processor: Intel Core i7 or equivalent
- RAM: 16 GB
- GPU: NVIDIA GPU with CUDA support (for DistilBERT training)

Software:

- Python 3.8+
- scikit-learn 0.24+ (for Logistic Regression and evaluation metrics)
- Transformers 4.0+ (Hugging Face library for DistilBERT)
- PyTorch 1.7+ (deep learning framework)
- NLTK 3.5+ (for text preprocessing)
- Pandas and NumPy (for data manipulation)

This methodology ensures rigorous, reproducible comparison between classical machine learning and transformer-based approaches, with careful attention to fair experimental design and comprehensive evaluation.

IV. RESULTS AND DISCUSSION

IV.1 Overall Performance Comparison

Our experimental evaluation revealed distinct performance characteristics for each approach. Table 1 presents the comprehensive performance metrics for both models on the test set of 6,000 tweets.

Table 1: Overall Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	77.2%	76.8%	77.9%	77.3%
DistilBERT	82.4%	82.1%	82.8%	82.4%

DistilBERT demonstrated superior performance across all metrics, achieving 82.4% accuracy compared to Logistic Regression's 77.2%, representing a 5.2 percentage point improvement. This performance gap, while substantial, is accompanied by significant differences in computational requirements and implementation complexity, which we discuss in subsequent sections.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

The F1-scores indicate that both models maintain reasonable balance between precision and recall, with DistilBERT showing slightly better equilibrium. The transformer model's higher recall suggests superior capability in identifying positive sentiment instances, while its precision indicates fewer false positive classifications.

IV.2 Confusion Matrix Analysis

Detailed examination of confusion matrices provides insight into specific error patterns and model behavior.

Logistic Regression Confusion Matrix:

	Predicted Negative	Predicted Positive
Actual Negative	2,289 (TN)	711 (FP)
Actual Positive	663 (FN)	2,337 (TP)

DistilBERT Confusion Matrix:

	Predicted Negative	Predicted Positive
Actual Negative	2,463 (TN)	537 (FP)
Actual Positive	516 (FN)	2,484 (TP)

The confusion matrices reveal several important patterns:

False Positive Reduction: DistilBERT produced 174 fewer false positives (537 vs. 711), indicating better discrimination of negative sentiment. This suggests the transformer model's contextual understanding helps avoid misclassifying negative tweets as positive.

False Negative Reduction: DistilBERT also generated 147 fewer false negatives (516 vs. 663), demonstrating improved sensitivity to positive sentiment indicators.

Balanced Error Distribution: Both models show relatively balanced error rates across classes, with neither exhibiting strong bias toward over-predicting one class. This balance reflects the stratified training approach and balanced dataset.

Error Rate Comparison: Logistic Regression's overall error rate of 22.8% (1,374 misclassifications) compared to DistilBERT's 17.6% (1,053 misclassifications) represents a 23% relative reduction in errors.

IV.3 Performance on Linguistic Nuances

Qualitative analysis of misclassified examples revealed important differences in how each model handles linguistic complexity:

Sarcasm and Irony:

DistilBERT demonstrated superior capability in detecting sarcastic expressions. For example, tweets like "Oh great, another Monday morning" were more frequently classified correctly as negative by DistilBERT, while Logistic Regression often misclassified them as positive due to the presence of the word "great."

Contextual Negation:

The transformer model better handled negation patterns, particularly when negation words appeared distant from sentiment-bearing words. Phrases like "not even remotely good" were more reliably classified as negative by DistilBERT, whereas Logistic Regression sometimes focused on "good" while missing the negation context.

Implicit Sentiment:

DistilBERT showed stronger performance on tweets expressing sentiment implicitly without explicit positive or negative words. For instance, "Waiting for customer service for 2 hours now" was more consistently classified as negative by the transformer model, which could infer frustration from the context.

Emoji and Emoticon Handling:

Both models handled emoticons reasonably well, though DistilBERT's subword tokenization sometimes fragmented emoji representations. Logistic Regression's explicit emoticon preservation in preprocessing provided slight advantages in cases where emoticons were the primary sentiment indicators.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

Slang and Informal Language:
DistilBERT's pre-training on diverse internet text provided advantages in understanding contemporary slang and informal expressions. Terms like "lit," "fire," or "savage" were more reliably interpreted in context by the transformer model.

IV.4 Computational Efficiency Analysis

The performance advantages of DistilBERT come with significant computational costs:

Training Time:

- Logistic Regression: Approximately 2-3 minutes for complete training on 24,000 tweets
- DistilBERT: Approximately 45-60 minutes for 3 epochs of fine-tuning (with GPU acceleration)

The transformer model required roughly 20-30 times more training time, though this includes the fine-tuning process that adapts pre-trained weights rather than training from scratch.

Inference Time:

- Logistic Regression: ~0.001 seconds per tweet (batch processing of 1,000 tweets in ~1 second)
- DistilBERT: ~0.02 seconds per tweet (batch processing of 1,000 tweets in ~20 seconds)

For real-time applications processing high volumes of social media data, this 20x difference in inference speed could be significant. However, DistilBERT's inference time remains practical for most applications, processing approximately 50 tweets per second.

Memory Requirements:

- Logistic Regression: ~50 MB (model weights and TF-IDF vectorizer)
- DistilBERT: ~250 MB (model weights only)

The transformer model requires approximately 5 times more memory, which could impact deployment on resource-constrained devices or when serving multiple models simultaneously.

Hardware Requirements:

- Logistic Regression: Can train and run efficiently on CPU-only systems
- DistilBERT: Benefits significantly from GPU acceleration for training; CPU inference is possible but slower

IV.5 Interpretability and Explainability

Logistic Regression Interpretability:
One significant advantage of Logistic Regression is its inherent interpretability. The model's learned weights directly indicate feature importance, allowing practitioners to understand which words or n-grams most strongly influence sentiment predictions. For example, examining the highest-weighted features revealed:
Positive indicators: "love," "great," "best," "awesome," "perfect"
Negative indicators: "hate," "worst," "terrible," "disappointed," "awful"
This transparency facilitates model debugging, bias detection, and stakeholder communication, particularly important in regulated industries or applications requiring explainable AI.

DistilBERT Interpretability:
Transformer models operate as "black boxes," making interpretation more challenging. While attention visualization techniques can provide some insight into which words the model focuses on, the complex interactions across multiple layers and attention heads make comprehensive interpretation difficult. This opacity can be problematic in applications requiring model explanations or when debugging unexpected predictions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

IV.6 Practical Deployment Considerations

Logistic Regression Advantages:

- **Simplicity:** Straightforward implementation and deployment with minimal dependencies
- **Speed:** Rapid training and inference suitable for real-time applications
- **Resource Efficiency:** Low memory and computational requirements enable deployment on edge devices
- **Interpretability:** Clear feature importance facilitates model understanding and debugging
- **Stability:** Deterministic predictions and stable behavior across different environments

DistilBERT Advantages:

- **Accuracy:** Superior performance, particularly on nuanced and complex expressions
- **Generalization:** Better handling of diverse linguistic patterns and informal language
- **Context Understanding:** Captures long-range dependencies and contextual relationships
- **Transfer Learning:** Leverages knowledge from pre-training on massive text corpora
- **Adaptability:** Can be fine-tuned for domain-specific applications with relatively small datasets

Use Case Recommendations:

Choose Logistic Regression when:

- Computational resources are limited
- Real-time processing of high-volume streams is required
- Model interpretability is critical
- Development and deployment simplicity is prioritized
- The performance gap (5% accuracy difference) is acceptable for the application

Choose DistilBERT when:

- Maximum accuracy is essential
- Handling linguistic nuances (sarcasm, context-dependent meaning) is important
- Sufficient computational resources are available
- The application can tolerate slightly higher latency
- The dataset contains complex, informal, or domain-specific language

V. CONCLUSION

This comparative study provides empirical evidence regarding the trade-offs between classical machine learning and transformer-based approaches for social media sentiment analysis. Our experiments on 30,000 tweets from the Sentiment140 dataset demonstrate that DistilBERT achieves superior performance with 82.4% accuracy compared to Logistic Regression's 77.2%, representing a meaningful but not overwhelming advantage.

The performance gap of 5.2 percentage points reflects DistilBERT's superior capability in handling linguistic nuances, contextual dependencies, and complex sentiment expressions common in social media text. The transformer model's attention mechanisms and contextual embeddings enable more sophisticated understanding of sarcasm, negation, and implicit sentiment compared to the bag-of-words assumptions underlying TF-IDF-based approaches.

However, this performance advantage comes with significant computational costs. DistilBERT requires approximately 20-30 times more training time, 20 times longer inference time, and 5 times more memory than Logistic Regression. These differences have practical implications for deployment scenarios, particularly in resource-constrained environments or applications requiring real-time processing of high-volume data streams.

The interpretability advantage of Logistic Regression represents another important consideration. The model's transparent feature weights facilitate understanding, debugging, and stakeholder communication, which can be critical

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

in regulated industries or applications requiring explainable AI. In contrast, DistilBERT's complex architecture makes interpretation challenging, though attention visualization techniques provide some insight.

Key Findings:

- **Performance:** Transformer models deliver measurably better accuracy and handle linguistic complexity more effectively than classical approaches.
- **Efficiency:** Classical machine learning methods offer substantial advantages in computational efficiency, training speed, and resource requirements.
- **Interpretability:** Traditional models provide clearer explanations of predictions, facilitating model understanding and debugging.
- **Practical Trade-offs:** The choice between approaches should be guided by specific application requirements, resource constraints, and performance objectives rather than assuming one approach is universally superior.

Recommendations for Practitioners:

For production systems requiring maximum accuracy and possessing adequate computational resources, transformer-based models like DistilBERT represent the preferred choice, particularly when analyzing social media text with its characteristic linguistic complexity. The performance gains justify the additional computational investment in applications where sentiment analysis accuracy directly impacts business outcomes or research conclusions.

Conversely, for applications prioritizing deployment simplicity, computational efficiency, or model interpretability, classical machine learning approaches remain viable and practical choices. The 77% accuracy achieved by Logistic Regression is sufficient for many real-world applications, particularly when combined with appropriate preprocessing and feature engineering.

Future Research Directions:

Several avenues for future investigation emerge from this work:

- **Hybrid Approaches:** Exploring ensemble methods that combine classical and transformer-based models to balance performance and efficiency.
- **Domain Adaptation:** Investigating how performance characteristics change across different social media platforms and domains.
- **Multilingual Analysis:** Extending the comparison to non-English languages and cross-lingual sentiment analysis.
- **Efficiency Optimization:** Developing techniques to reduce transformer model computational requirements while preserving performance advantages.
- **Explainability Methods:** Advancing interpretability techniques for transformer models to address the transparency gap with classical approaches.

In conclusion, both classical machine learning and transformer-based approaches have legitimate roles in modern sentiment analysis applications. Rather than viewing them as competing alternatives, practitioners should understand their respective strengths and limitations, selecting the approach that best aligns with specific application requirements, constraints, and objectives. As the field continues to evolve, the development of more efficient transformer variants and improved interpretability techniques may further shift the balance, but the fundamental trade-offs between performance, efficiency, and interpretability will likely remain relevant considerations in practical sentiment analysis system design.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Aswalekar Uday, Assistant Professor at the Department of Computer Science, Centre of Distance and Online Education University of Mumbai, for their valuable guidance and support throughout this research work.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, June 2026
Impact Factor: 8.2

REFERENCES

[1] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216-225, 2014.

[2] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pp. 2200-2204, 2010.

[3] B. Pang and L. Lee, "Thumbs up? Sentiment Classification using Machine Learning Techniques," *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, pp. 79-86, 2002.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv preprint arXiv:1301.3781*, 2013.

[5] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532-1543, 2014.

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[7] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746-1751, 2014.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998-6008, 2017.

[9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171-4186, 2019.

[10] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[11] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," *International Conference on Learning Representations*, 2020.

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*, 2019.

[13] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *CS224N Project Report, Stanford*, vol. 1, no. 12, p. 2009, 2009.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

[15] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. Rush, "Transformers: State-of-the-Art Natural Language Processing," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38-45, 2020.

[16] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8024-8035, 2019.

[17] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Inc., 2009.