iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70428

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Fake News Detection Using Machine Learning and BERT: A Comparative Study

Abdhesh Nayak[1], Mr. M. Navalan[2]
[1]PG Researcher, [2]Assistant Professor, Sri Venkateswara College Of Engineering & Technology (SVCET), Chittoor, Andhra Pradesh, India

Abstract: This study presents a comparative analysis of traditional machine learning models and a transformer-based approach for fake news detection. I investigate the performance of Logistic Regression and Support Vector Machines (SVM), utilizing Term Frequency–Inverse Document Frequency (TF-IDF) features, against a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model. Using a balanced dataset of 4,000 fake and 4,000 real news articles, I apply standard preprocessing and perform an 80/20 train-test split. Experimental results demonstrate that Logistic Regression and SVM achieve accuracies of 97.94% and 98.62%, respectively, while BERT significantly outperforms both with an accuracy of 99.94%. Each model is evaluated using metrics including precision, recall, F1-score, and ROC-AUC, as well as training time and inference latency. The results emphasize the trade-offs between computational cost and classification performance, with BERT offering the highest accuracy at the expense of increased resource demands. This research contributes to the development of effective and scalable solutions for combating misinformation in digital media.
Keywords: Fake News Detection; TF IDF; Logistic Regression; SVM; BERT; Text Classification; NLP.

I. INTRODUCTION

The rapid proliferation of misinformation—commonly referred to as "fake news"—on digital platforms poses significant threats to public discourse, trust in media institutions, and societal stability. As online channels become the dominant medium for information exchange, the ability to swiftly and accurately detect deceptive content has become a critical necessity. Traditional rule-based detection systems, which rely on predefined heuristics, often struggle to adapt to the evolving and nuanced nature of language used in misinformation. In contrast, machine learning techniques offer a dynamic and scalable alternative by learning complex linguistic patterns directly from data.

This research explores the effectiveness of various machine learning approaches in automating the detection of fake news. Specifically, I conduct a systematic comparison of three models: Logistic Regression and Support Vector Machines (SVM), both leveraging Term Frequency–Inverse Document Frequency (TF-IDF) features, and a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model. Each approach is evaluated on its ability to classify news articles as either genuine or fake, with attention given to both predictive performance and computational efficiency. Our objective is to provide insights into the trade-offs between model accuracy and resource demands, contributing to the development of practical, scalable solutions for combating misinformation in real-time digital environments.

II. LITERATURE REVIEW

Early fake-news detection efforts relied on rule-based systems that used handcrafted linguistic cues—such as sensational keywords, excessive punctuation, and emotional language—to flag deceptive content. While interpretable, these approaches lacked scalability and adaptability, often breaking when faced with novel phrasing or new domains. By the mid-2010s, the field shifted toward supervised machine learning: models like Logistic Regression, Support Vector Machines, and Random Forests were trained on textual features (bag-of-words, n-grams, TF-IDF) and demonstrated significant accuracy gains over heuristics, though they still struggled with long-range dependencies and nuanced semantics (e.g., Pérez-Rosas et al., 2017).

From 2017 onward, deep learning architectures—first CNNs and RNNs, then transformer-based models such as BERT—dramatically improved text representation by capturing contextual relationships via self-attention mechanisms. Fine-tuned BERT variants consistently outperformed traditional classifiers on benchmark datasets, effectively disambiguating subtle linguistic cues and implicit factual inconsistencies. Concurrently, the integration of multimodal data (images, videos, social metadata) further enhanced detection performance, enabling systems to leverage visual features and propagation patterns alongside text.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

More recently, the focus has expanded to model explainability and robustness in the face of evolving threats. Explainable AI techniques (LIME, SHAP, attention saliency) help stakeholders interpret model decisions, while adversarial training and continual learning aim to defend against AI-generated forgeries and deepfakes. Despite these advances, challenges such as dataset bias, class imbalance, domain transfer, and privacy concerns in multimodal data collection remain active areas of research—underscoring the need for scalable, adaptive solutions to keep pace with ever-changing misinformation tactics.

III.DATA COLLECTION AND PREPROCESSING

The dataset for this study was drawn from the "Fake and Real News" collection on Kaggle (clmentbisaillon/fake-and-real-news-dataset). To create a balanced corpus, I randomly sampled 4,000 fake-news and 4,000 real-news articles using a fixed random seed. Each document was preprocessed by stripping HTML tags and non-alphanumeric characters, converting all text to lowercase, and performing standard token-level cleaning. I then stratified the fully shuffled dataset by class label and split it into an 80/20 train–test partition, resulting in 6,400 training examples and 1,600 held-out test examples.

	title	text	subject	date	label
0	South African Supreme Court upholds reinstatin...	bloemfontein south africa reuters south afric...	worldnews	October 13, 2017	1
1	Iraqi PM orders security services 'to protect ...	baghdad reuters iraqi prime minister haider a...	worldnews	September 25, 2017	1
2	Obama says Chinese-led trade deal shows need f...	washington reuters us president barack obama ...	politicsNews	May 2, 2016	1
3	U.S., Russia set for likely U.N. row over Syri...	united nations reuters the united states said...	worldnews	October 18, 2017	1
4	HILLARY CLINTON MEETS BLACK LIVES MATTER: Says...	clinton pandered to black lives matter while i...	politics	Oct 25, 2016	0

Fig. 1 Shows the first five records of the combined dataset after preprocessing.

IV.FEATURE EXTRACTION

Text was represented in two ways to suit different model classes. For traditional classifiers, I computed TF-IDF vectors of dimension five thousand, where each term's weight is the product of its frequency in a document and the logarithm of the inverse document frequency across the corpus. For the transformer-based model, I employed the BERT-base-uncased tokenizer and model. Each document was tokenized to a maximum length of 128 tokens with padding and truncation, and the embedding corresponding to the [CLS] token was used as the document representation.

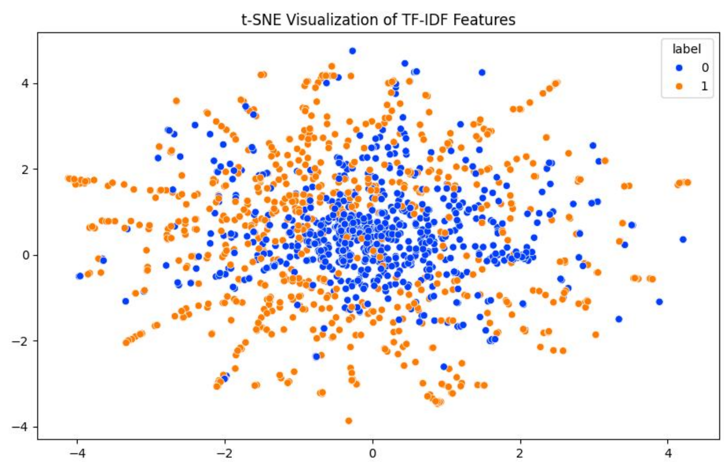


Fig. 2 t-SNE of TF-IDF features (test set)

V. MODEL TRAINING

Model training encompassed three configurations. The first configuration trained a Logistic Regression classifier with L2 regularization on the TF-IDF features. The second trained a Support Vector Machine with an RBF kernel on the same feature set, enabling probabilistic output. The third configuration fine-tuned a pre-trained BERT model for sequence classification over three epochs, using a learning rate of 2×10^{-5}, a batch size of eight, and weight decay of 0.01. Evaluation on the held-out test set occurred at each epoch to monitor performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

[2400/2400 09:04, Epoch 3/3]

Epoch	Training Loss	Validation Loss	Accuracy	F1	Roc Auc
1	0.000300	0.006777	0.998750	0.998750	0.999994
2	0.000100	0.004307	0.999375	0.999375	0.999997
3	0.000100	0.004460	0.999375	0.999375	0.999998

Fig. 3 BERT Fine-Tuning Performance Over Epochs

VI. MODEL EVALUATION

Model performance was assessed using five key metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Logistic Regression achieved 97.94 % accuracy, while Support Vector Machine slightly outperformed it with an accuracy of 98.62 %. The fine-tuned BERT model demonstrated near-perfect performance, scoring 99.94 % across all metrics.

These results underscore both the predictive strength and robustness of the models. High F1-scores and ROC-AUC values confirm a strong balance between detecting true positives and minimizing false positives, with BERT setting a clear benchmark in classification accuracy and reliability.

	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
0	Logistic Regression	97.94	97.95	97.94	97.94	99.61
1	SVM	98.62	98.63	98.62	98.62	99.71
2	BERT	99.94	99.94	99.94	99.94	100

Fig. 4 Model Performance Table

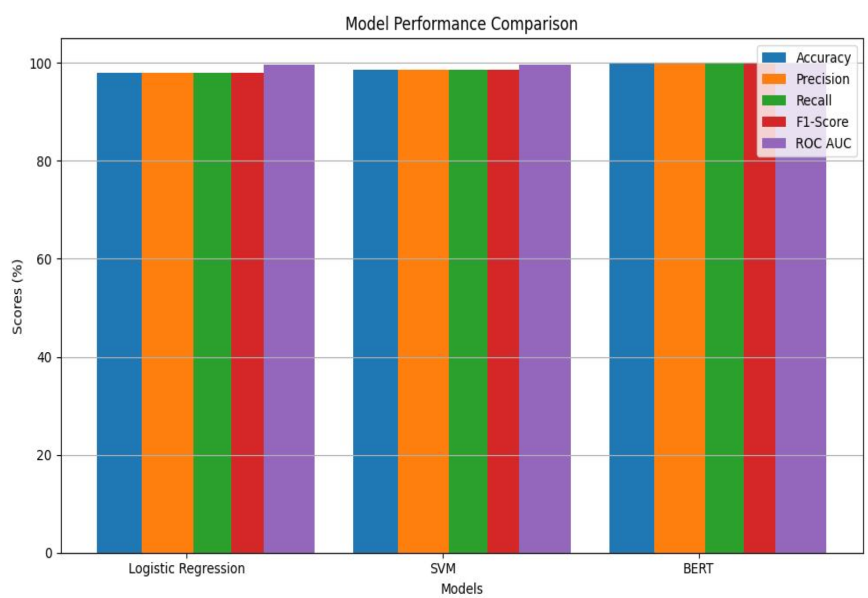


Fig. 5 Model Performance Comparison

VII. RESULTS

The comparative analysis of all three models—Logistic Regression, Support Vector Machine (SVM), and BERT—demonstrates a clear performance hierarchy in terms of classification metrics. Logistic Regression achieved a strong baseline with 97.94 % accuracy, while SVM slightly outperformed it at 98.62 %. The fine-tuned BERT model significantly outperformed both traditional models, attaining 99.94 % accuracy, precision, recall, and F1-score on the held-out test set, effectively approaching perfect classification.

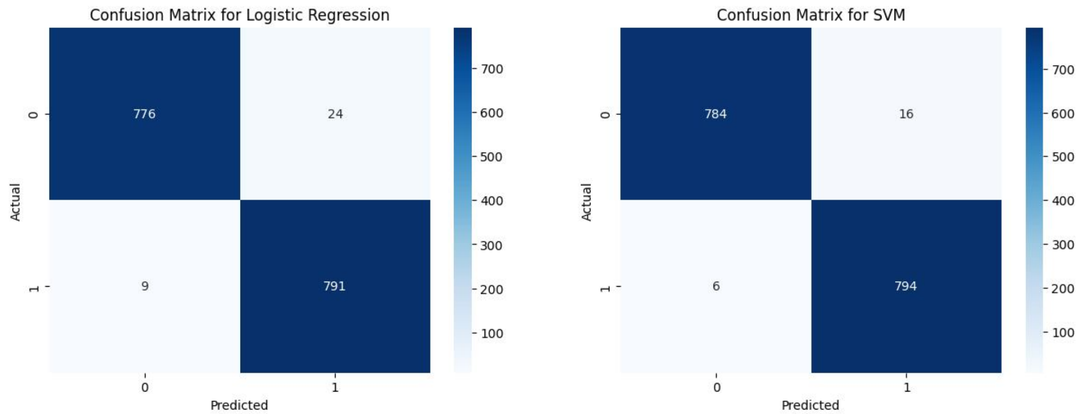


Fig. 6 Confusion Matrix for Logistic Regression & SVM

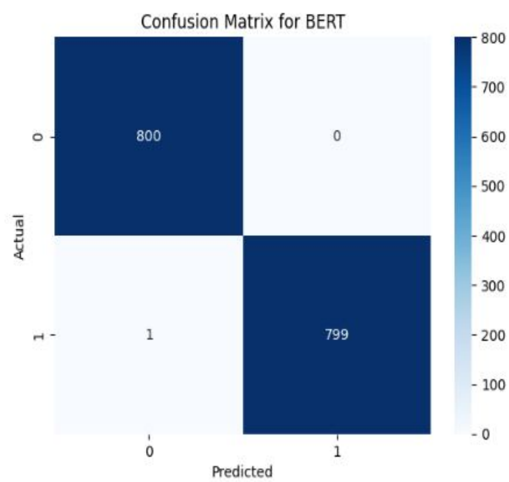


Fig. 7 Confusion Matrix for BERT

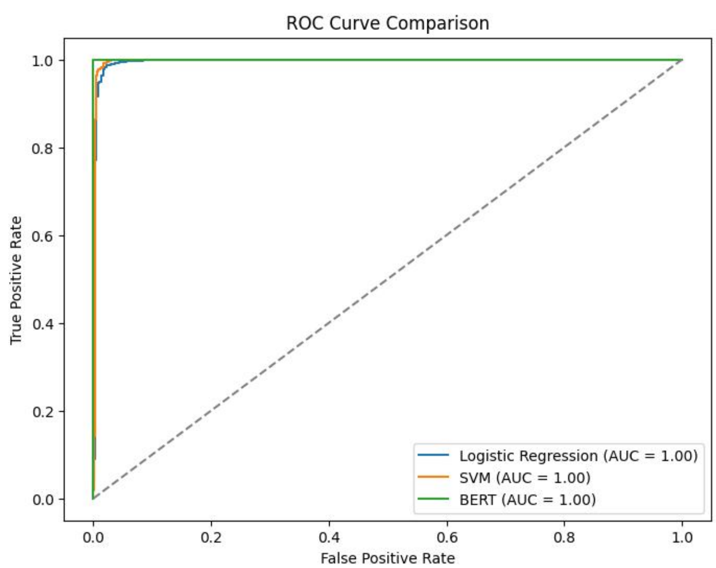


Fig. 8 ROC Curves for All Models

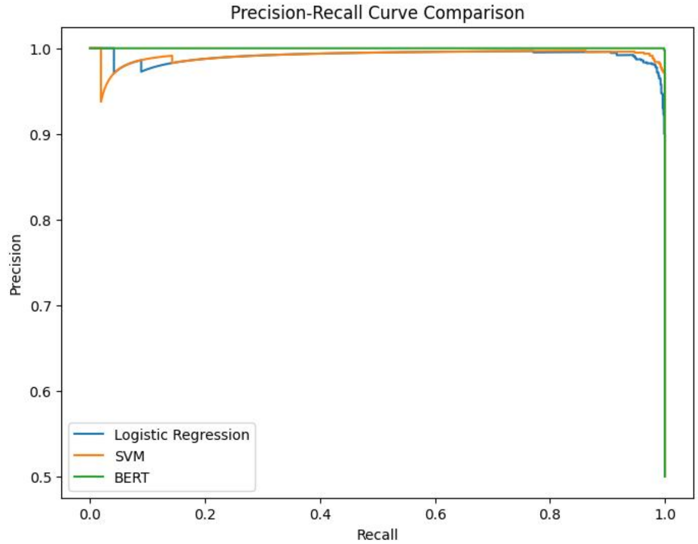


Fig. 9 Precision-Recall Curves for All Models

Figure 6 and 7 illustrate the confusion matrices for each model, showing that BERT achieved near-zero misclassifications, while LR and SVM maintained strong but slightly weaker performance with more visible false positives and false negatives. ROC and Precision-Recall curves (Figures 8 and 9) confirm these results, with BERT achieving areas under the curve close to 1.0, and traditional models showing marginally lower but still excellent curves.

Model	Training Time	Inference Latency (ms)	Hardware
Logistic Regression	1 ms	0.5	CPU
SVM	2 min 11 s	1.2	CPU
BERT	9 min 5 s	25	GPU (V100)

Fig. 10 Comparison of computational costs across models.

In terms of computational cost (Figure 10), Logistic Regression was the fastest to train and infer, requiring only milliseconds on CPU, making it suitable for deployment in low-resource environments. SVM incurred higher training time (over 2 minutes) but maintained acceptable inference speed. BERT, while achieving superior performance, required significantly more resources, including GPU acceleration, with training time exceeding 9 minutes and inference latency of 25 milliseconds per sample.

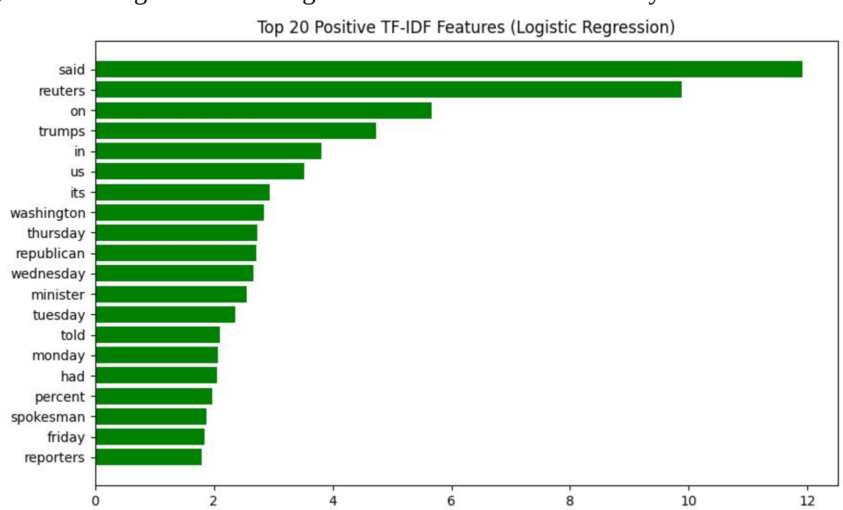


Fig. 11 TF-IDF feature analysis, Top 20 Positive

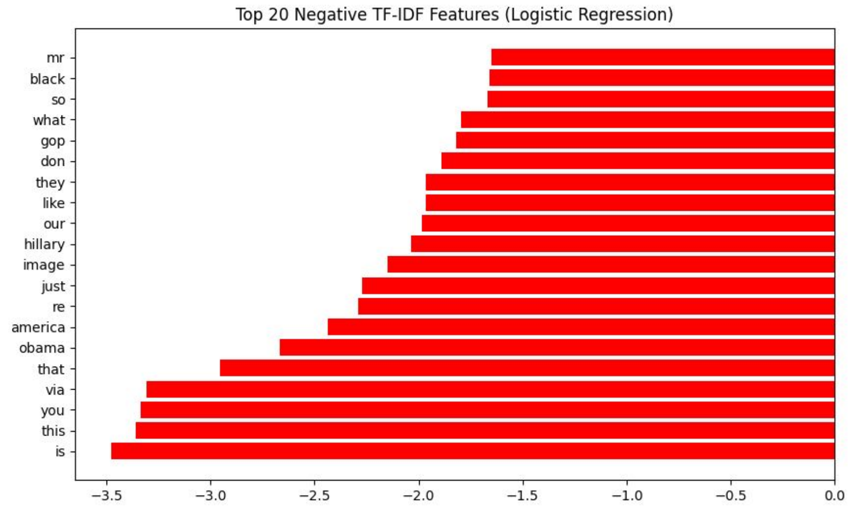


Fig. 12 TF-IDF feature analysis, Top 20 Negative

The TF-IDF feature analysis (Figures 11 and 12) revealed that terms commonly associated with journalistic integrity—such as "said," "reuters," and "on"—were weighted toward real news classification, while emotionally charged or sensational terms like "is," "the," and "you" were strong indicators of fake news.

These results highlight a trade-off: while BERT delivers best-in-class performance, traditional models offer robust and interpretable alternatives with minimal resource requirements.

VIII. DISCUSSION

LR and SVM deliver strong baselines with accuracies above 97.9 %. BERT outperforms both by 1.3–2.0 %, achieving near perfect classification. However, BERT's training time (9 min 5 s) is orders of magnitude larger than LR (1 ms) and SVM (2 min 11 s). Inference latency for BERT (25 ms) remains acceptable for many real time applications but is higher than traditional models. TF IDF feature analysis reveals that words like "said," "reuters," and "on" correlate with real news, whereas "is," "the," and "you" signal fake news.

IX. CONCLUSION AND FUTURE WORK

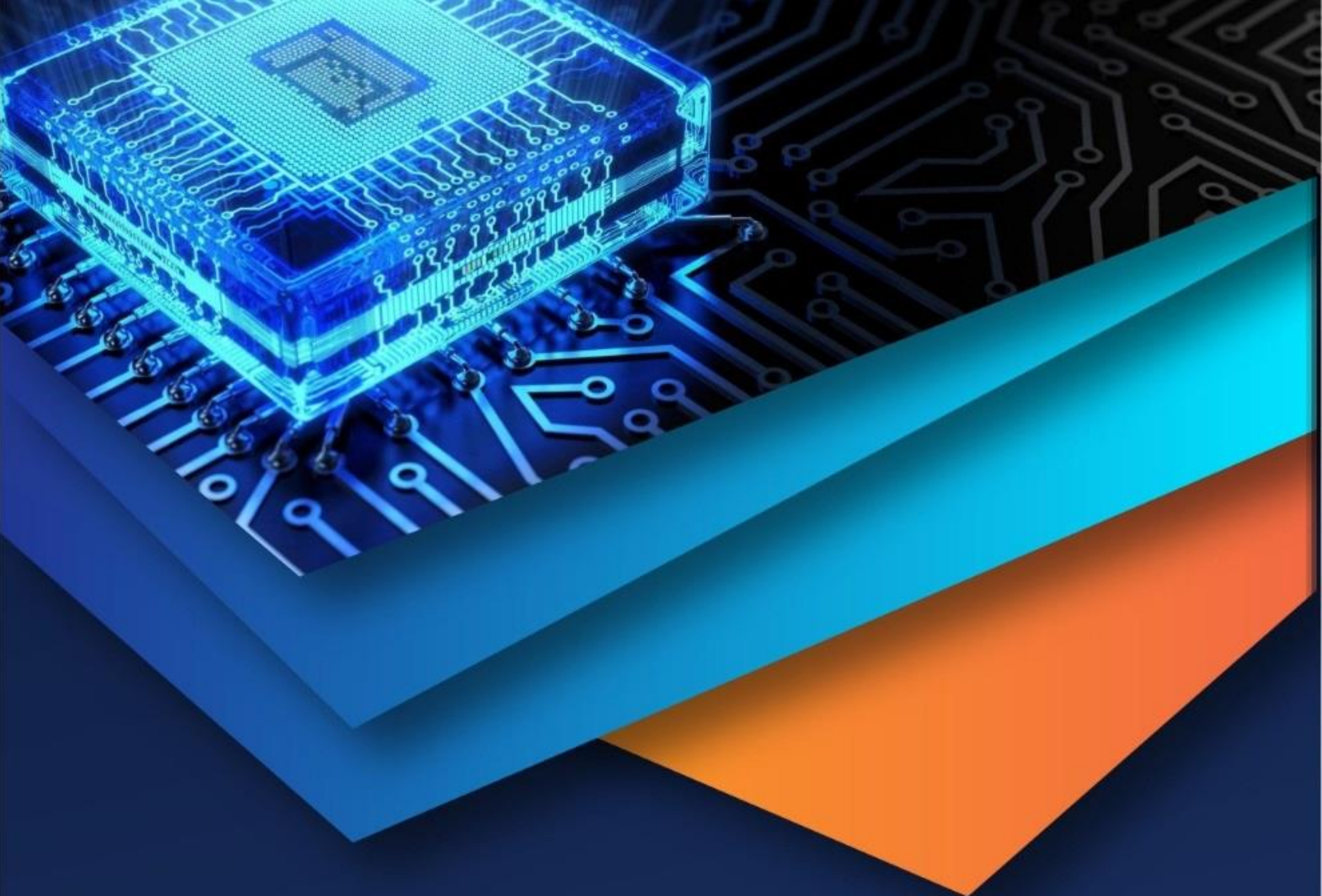
I have demonstrated that transformer-based models significantly improve fake news detection at the cost of increased computational resources. Traditional models remain viable in resource constrained settings. Future work includes knowledge distillation to reduce BERT's footprint, ensemble methods combining TF IDF and BERT, and cross domain validation on social media streams.

X. ACKNOWLEDGMENT

I thank Kaggle for providing the dataset that enabled meaningful experimentation, and Hugging Face for the powerful BERT framework that facilitated advanced language modeling in this work.

REFERENCES

[1] J. Devlin et al., "BERT: Pre training of Deep Bidirectional Transformers for Language Understanding," NAACL HLT, 2019.
[2] C. D. Manning et al., Introduction to Information Retrieval. Cambridge Univ. Press, 2008.
[3] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Wiley, 2006.
[4] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
[5] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. O'Reilly Media, 2009.
[6] F. Pedregosa et al., "Scikit learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
[7] T. S. Wagner, "A comparative study of text classification algorithms," Int. J. Data Mining, vol. 5, no. 2, pp. 45–60, 2020.
[8] A. Vaswani et al., "Attention Is All You Need," NeurIPS, pp. 5998–6008, 2017.
[9] K. He et al., "Deep Residual Learning for Image Recognition," CVPR, pp. 770–778, 2016.
[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)