Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 7 (2025)

Depressive Tweet Classification via Machine Learning: Contextualizing Performance with a Broad NLP Review

[1]Muhammad Ahmad Jamal, [2]Talha Farooq Khan, [3]Mubasher Malik, [4]Asna Riaz, [5]Muhammad Ali Hassan

Article Details

ABSTRACT

Keywords: Sentiment Analysis, Natural Language Processing, Lda, Bert, Svm, Tf-Idf, Depressive Tweets, Transformer Models, Hybrid Approaches

Muhammad Ahmad Jamal
Department of Computer Science, University of Southern Punjab, Multan.
mianahmadjamal786@gmail.com
Talha Farooq Khan
Department of Computer Science, University of Southern Punjab, Multan.
talhafarooqkhan@gmail.com
Mubasher Malik
Department of Computer Science, University of Southern Punjab, Multan .
hodcs@usp.edu.pk
Asna Riaz
Department of Computer Science, University of Southern Punjab, Multan.
Corresponding Author Email:
asnariaz6@gmail.com
Muhammad Ali Hassan
Department of Computer Science, University of Southern Punjab, Multan
alihassanhashmi786@gmail.com

Natural Language Processing (NLP) has significantly advanced the analysis of unstructured text, enabling critical applications across diverse sectors including finance, healthcare, and social media analytics. This paper comprehensively reviews 25 research papers on NLP for topic modeling and sentiment analysis, contrasting traditional methods like Latent Dirichlet Allocation (LDA) and TF-IDF classifiers with advanced transformer models such as BERT. The review highlights that while classical models offer interpretability, they often struggle with noisy, short-form social media content, where newer transformer-based and hybrid approaches demonstrate superior performance in thematic mapping and sentiment detection. Complementing this review, we conducted an empirical study on depressive tweet classification. Utilizing TF-IDF features, we evaluated Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) models. The kernel-based SVM achieved the highest accuracy at 99.5%, surpassing LR (98.9%) and RF (97.2%), a performance consistent with SVM's known efficacy in high-dimensional, sparse feature spaces. Our analysis identifies critical gaps, including the prevalent focus on English-only datasets, the underexplored potential of multimodal data fusion (e.g., text and images for depression detection), and challenges with class imbalance. We recommend future research explore multilingual transformer architectures (e.g., mBERT, DistilBERT), integrate domain-specific lexicons, employ multitask learning frameworks for integrated topic and sentiment analysis, and incorporate Explainable AI (XAI) for enhanced transparency and ethical considerations in model development. This paper synthesizes current advancements and challenges, providing a comprehensive roadmap for developing more robust, equitable, and context-aware NLP systems for social media text analysis.

INTRODUCTION

Rapid advancements in Natural Language Processing (NLP) have revolutionized the analysis of unstructured text, enabling successful applications across diverse domains such as sentiment analysis, topic modeling, public health informatics, financial intelligence, and policy formulation. Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) pioneered topic modeling by identifying latent thematic structures within large text corpora, valued for their interpretability and ease of application [1]. However, the proliferation of informal, noisy, and context-dependent communication on platforms like Twitter and Facebook has exposed inherent limitations in these traditional language analysis techniques, particularly in handling colloquialisms, grammatical irregularities, and ambiguous thematic content. Similarly, lexicon-based approaches such as VADER and TextBlob struggle to capture nuanced linguistic phenomena, including sarcasm, metaphorical expressions, and cross-lingual sentiment, thereby limiting their efficacy [3, 5]. For instance, an analysis of Twitter discourse in India revealed that only 47.2% of users expressed favorability towards additional COVID-19 vaccinations, largely driven by concerns over side effects and vaccine distrust. Critically, existing tools like TextBlob lacked the linguistic granularity to accurately interpret sentiment expressed in local dialects within these tweets [3, 4]. These challenges necessitate the development of more sophisticated models capable of robustly handling the complexities of contemporary digital text, while maintaining high accuracy, scalability, and interpretability.

The limitations of conventional NLP frameworks are further exacerbated by domain-specific complexities. In public health, summarizing public discourse on microblogging platforms like Twitter, characterized by inherent character constraints, restricts the extraction of critical insights. For example, 25.4% of Indian tweets concerning the Omicron variant expressed skepticism regarding lockdowns and vaccines; however, monolingual models failed to capture sentiment from a significant portion of these communities due to language bias [4]. Similarly, financial sentiment analysis faces unique challenges stemming from domain-specific terminology (e.g., “bull markets”) and quantitative expressions (e.g., “revenue at +10%”). This necessitates the development of specialized lexicons to prevent misclassification of neutral financial terms as sentiment-laden, as demonstrated in ESG (Environmental, Social, and Governance) news analysis correlating sentiment with corporate financial performance indicators like Return on Assets (ROA) and Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 7 (2025)

[8, 9]. Furthermore, the pervasive focus on English-centric models exacerbates the resource disparity for low-resource languages. Studies on Moroccan Arabic headlines, for instance, highlight the scarcity of diverse and sufficient training data, impeding the development of reliable systems for assessing community health [10, 19]. Collectively, these issues underscore a broader challenge: the absence of comprehensive frameworks that seamlessly integrate syntactic, semantic, and pragmatic analysis to resolve contextual ambiguities and ensure equitable representation across diverse linguistic groups [12, 23].

Addressing these multifaceted challenges is imperative, given their significant real-world implications. During India's COVID-19 booster rollout, accurate topic and sentiment analysis proved instrumental in enabling policymakers to tailor public health campaigns, effectively addressing specific public concerns such as skepticism regarding booster necessity among younger demographics and perceived vaccine inefficacy among older populations [3, 4]. Hybrid models, integrating LDA with reinforcement learning, have demonstrably optimized portfolio construction, achieving superior Sharpe ratios by identifying correlations between ESG sentiment and market dynamics [8, 9]. Similarly, the integration of Bidirectional Long Short-Term Memory (Bi-LSTM) networks with emotion analysis from user comments has yielded a 96.77% Area Under the Curve (AUC) score, affirming the utility of such multimodal approaches in combating misinformation [1]. The extension of NLP capabilities across diverse domains is pivotal for fostering equitable technological access and engagement. Consequently, architectural innovations like DistilBERT and binarized language models have facilitated a 40% reduction in computational overhead while retaining up to 92% of classification accuracy, making real-time deployment feasible in resource-constrained environments [6, 25]. Such advancements have led to remarkable improvements, with Moroccan news headlines achieving up to 94.74% accuracy in topic classification, thus bridging critical support gaps for low-resource languages through techniques like ASAFAYA embeddings and Bi-GRU networks [10, 19]. This paper aims to contribute to the field by comprehensively reviewing both traditional and contemporary transformer-based NLP methodologies for sentiment analysis and topic modeling on social media data, synthesizing their strengths and limitations across various applications. Furthermore, it presents an empirical investigation into the efficacy of classical machine learning algorithms (Naive Bayes, Logistic Regression, Support Vector Machine, Random Forest) utilizing TF-IDF features for the binary classification of depressive tweets. The work also identifies existing challenges in current NLP approaches, particularly concerning multilingual data processing, handling of informal text, and the need for explainable AI, while proposing future research directions to advance the robustness, interpretability, and real-world applicability of NLP systems in sensitive domains. The remainder of this paper is structured as follows: Section 2 provides a detailed Literature Review, analyzing 25 seminal studies on NLP techniques for topic modeling and sentiment analysis. Section 3 outlines the Methodology employed in our empirical study, detailing data preprocessing, feature engineering, and model training. Section 4 presents the Results and Discussion of our experimental findings. Finally, Section 5 concludes the paper by summarizing key insights and outlining avenues for Future Work.

LITERATURE REVIEW

NLP's development has greatly improved how text that lacks structure is analyzed for example, in topics like sentiment analysis, understanding topics and their uses in social media, healthcare, finance and public policy. Modeling of topics has often depended on methods such as LDA and NMF, since these can handle lengthy texts and still be explained easily [1] [2]. At the beginning, research used probabilistic models for long writings, but once these were applied to data from Twitter and Facebook, their weaknesses came to light. In particular, rare words occurring in pairs and slang or mistake-made words prevented LDA and BTM from effectively finding useful topics in tweets or comments [1], [3]. VADER and Text Blob were the primary tools used in sentiment analysis; however, they faced difficulties because they could not understand sarcasm, metaphors and expressions in different languages [3], [4], [5]. Since very few studies compared different approaches, experts could not clearly determine the most effective real-time applications [1], [6].

Researchers see that standard NLP frameworks are not designed for applications with real-world challenges. For example, even though LDA helps produce meaningful topics, it cannot be used quickly in online environments since its performance depends a lot on certain parameters [1], [7]. When working with COVID-19 posts, the lack of context in tweets and other short texts often led to mistakes in classifying people's opinions about vaccines and boosters [3], [4]. It was hard for financial sentiment analysis to avoid mistakes since many terms and numbers are used differently than in general texts [8], [9]. Moreover, because most language AI models center on English, the usefulness of these models is restricted where many languages are spoken such as India or Morocco [3], [10], [11]. As a result, there was a need for models that could respond to different situations while maintaining good results, scalability and clarity when using different kinds of data.

Recently, several approaches have mixed hybrid techniques, transformer models and updates for specific areas. Many researchers have improved the coherence and view of short-text topic models by employing LDA with Word2Vec and t-SNE in these studies [1],[12]. Because of the transformer-based approach and TF-IDF methods, BERT topics provided better diversity and structure in news headlines than the existing methods [2]. Experts found that BERT and its modifications such as Fin BERT and Time LMs, are strong approaches for sentiment analysis. Such as, to use Twitter's messy texts more effectively, Tweet NLP built a BERT version that reached top performance in spotting irony (82.2% F1-score) and identifying the sentiment in German tweets (77.35% F1 score) by frequently training on Twitter data [7]. When syntactic normalization, semantic interpretation and pragmatic analysis were performed, the detection of sarcasm and metaphor improved in user comments compared to methods using only a lexicon [12], [13]. Linking Sharpe ratios, ESG sentiment and corporate finances, researchers found that portfolios built using ensembles of LDA and reinforcement learning showed the highest accuracy [8], [9]. As a result, researchers focused on models that integrated various tasks and could handle multiple data forms and patterns in language [14], [15].

There are still many areas where solid knowledge is lacking. Because research often did not consider languages other than English, it focused on groups that spoke English and only considered urban areas. One example is that studies addressing people in India who hesitated to get a COVID-19 booster looked only at tweets written in English, failing to include Indians who communicate differently [3],[16]. Angled on the clarity of context, further study revealed that while PCA provided the clearest statistical markers, the topics became much more interpretable with LDA in the digital economy field [1], [2], [17]. Though transformers are accurate, using them in situations where the resources are not advanced involved very demanding calculations and made the process hard to examine [5], [9]. Turning to multimodal data (e.g., images, videos) would enhance insight from research, since using images and text alongside post words improved feelings research by increasing understanding of depression by 20% [12], [18]. While many issues related to privacy when using public posts for mental health prediction remained unresolved, people showed a mix of support and critics had biases in their views [5], [9], [19]. Future studies should focus on creating models for multiple languages, mainly using tools like mBERT and hybrid lexicons, to help solve the issue of global applicability [20], [10]. Employing different types of data and effective methods to clean text such as converting emojis, parsing slang and applying linguistics, could help enhance the detection of feelings and opinions on TikTok and Reddit [12], [18]. Topic models using DTM and transformer combinations such as XLNet-BERT, can better reflect shifts in social opinions occurring during a pandemic or election [21], [9], [22]. Using XAI and causal models in finance, one may clarify how ESG sentiments affect the market which is more useful for suggesting sustainable strategies to users [8], [18]. Societal well-being, ethical AI and reducing bias will greatly depend on people building open-source annotated data collections, for example the Moroccan News Arabic Dataset (MNAD) [10], [11]. Therefore, Distil BERT and binarized LLMs which are lightweight, could make it easier to use advanced NLP tools in places with few resources [6], [7].

Ultimately, merging both types of techniques has greatly improved text analytics, though the area is not yet settled. Improvements in languages, using less computing power and ensuring finer granularity will make it possible for ideas from neuroscience to be used in real-world projects. As topics on social media and important crises continue to affect the public, adaptive NLP systems will make raw data meaningful and useful to expert teams, helping them analyze where vaccination hesitancy exists or develop sound financial plans, among other goals. It will be important for people from different fields to try to address any ethical questions and see that all these technologies are used in a fair way by everyone.

TABLE 1: LITERATURE REVIEW

Sr.	Author(s) & Title of Paper	Year	Method/Model Used	Dataset	Results/Findings
1	Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis Twitter-Based	2020	LSA, LDA, NMF, PCA, RP	20-newsgroup corpus, Facebook conversations	LDA and NMF emerged as superior for short texts; LDA offers interpretability, NMF balances speed and accuracy. PCA had high statistical scores but low interpretability.

2	Sentiment Analysis and Topic Modeling of COVID-19 Booster Vaccine Shots	2023	LDA, TextBlob	76,979 English tweets (India)	52.8% negative sentiment, key concerns: distrust in "big pharma," fear of side effects.
3	Sentiment Analysis in Social Networks through Topic Modeling	2023	Sent.LDA (LDA + ANEW lexicon + modularity-based community detection)	Tweets from #clclasico	Topic-specific layers showed higher sentiment assortativity (0.446) vs. primary conversation (0.2047).
4	Sentiment Analysis of 2024 Indonesian Election Tweets	2024	SVM, LDA	50,001 tweets (Indonesian election)	SVM achieved 93% accuracy; LDA coherence score: 0.482. Neutral sentiments (18%) poorly classified.
5	Aspect-Based Sentiment Analysis Research via Bibliometrics	2025	Bibliometrics, Structural Topic Modeling (STM)	1,325 ABSA publications (2009–2023)	Dominant topics: attention-based neural networks, graph-empowered ABSA. China led in productivity.
6	Public Attitudes Toward Higher Education via NLP	2023	LDA, sentiment analysis	157,943 English tweets	Positive discourse dominated ("university," "student"); negative terms: "cost," "debt." ChatGPT-4 release correlated with engagement spikes.

7	Enhanced Sentiment Analysis During Pandemics	2024	Automated LDA, Word2Vec, t-SNE	Pandemic-related corpora	Topic coherence score: -2.6 (10 topics). Topic 8 (health and vaccines) most prevalent (22.97%).
8	Food Security Discourse Analysis via NLP	2023	LDA, VADER	38,070 Australian tweets (2019–2021)	39.31% positive sentiment; negative tweets had higher engagement. Topics: "Global production," "Giving to food banks."
9	Topic Modeling in Digital Economy Research	2023	LDA, TF-IDF	8,321 Scopus-indexed documents (1953–2023)	Dominant topics: Digital Transformation (56.6%), Data Governance (22.2%).
10	A Pragmatic NLP Framework for Social Media Sentiment	2023	Three-tier NLP (syntactic, semantic, pragmatic layers)	Social media comments	Improved sarcasm and metaphor detection; excluded multimodal data.
11	Stress Detection via Social Media Interactions	2023	BERT, Bi-GRU	100,042 tweets	Random Forest: 97.78% accuracy; BERT emotion classification: 94% macro accuracy.
12	Fake News Detection via Sentiment and Emotion Analysis	2024	Bi-LSTM, TextBlob, NRCLex	Fakeddit dataset	AUC: 96.77%; outperformed benchmarks by integrating sentiment, emotion, and textual features.
13	Stock				BERT achieved balanced

	Prediction Using Headline Sentiment	2023	BERT, RNN, VADER	Real-time scraped headlines	sentiment distribution; RNN eliminated neutral classifications.
14	Comparative Analysis of Topic Modeling Approaches	2023	LDA, BERT, XLNet-BERT	Twenty Newsgroups, Articles.xlsx	XLNet-BERT balanced coherence (0.69) and diversity (0.66); GPT-LDA had higher diversity (0.95) but lower coherence (0.53).
15	Multitask Sentiment and Topic Classification Using BERT	2024	BERT with custom attention mechanisms	3,263 Indian news articles	Accuracy: 98.4% (topic), 94% (sentiment).
16	Financial Sentiment Analysis Techniques	2024	FinBERT, LM lexicon, ensemble models	PhraseBank, SEntiFIN	FinBERT achieved F1=0.93; ensemble models improved portfolio Sharpe ratios.
17	ESG Sentiment Impact on Corporate Performance	2024	BERT, RoBERTa, ELECTRA	9,828 general + 140,000 ESG articles	ELECTRA detected higher negative sentiment; UMAP linked "sustainable" to profitability in mobility/renewable sectors.
18	TweetNLP: Platform-Specific NLP for Twitter	2023	TimeLMs, XLM-T	Twitter corpora	Achieved 82.2% F1 in irony detection; 77.35% F1 for German sentiment.
19	Arabic News		LDA, Bi-GRU	Moroccan News Arabic	GRU: 94.74% topic accuracy;

	Headlines for Societal Well-Being Scoring	2023	with ASAFAYA embeddings	Dataset (MNAD): 418,000 headlines	Bi-GRU: 90.23% sentiment accuracy.
20	Public Health Sentiment During COVID-19 Omicron Waves	2023	LDA, TextBlob	854,312 Indian tweets	25.4% negative sentiment; peak negativity in Dec 2022. Key concerns: lockdown fears, vaccine doubts.
21	Ethical AI and Sentiment Analysis	2022	Hybrid ML-DL models, XAI	Social media data	Highlighted bias in demographic representation; proposed XAI for transparency.
22	Dynamic Topic Modeling for Crisis Discourse	2024	Dynamic Topic Modeling (DTM), transformer hybrids	Crisis-related corpora	Captured temporal sentiment shifts; outperformed static models in interpretability.
23	Low-Resource Language Sentiment Analysis	2023	mBERT, hybrid lexicons	Urdu, Moroccan Arabic datasets	Addressed dialectal diversity; improved F1 scores by 12% over monolingual models.
24	Multimodal Sentiment Analysis with Emoji Integration	2024	BERT, ResNet	Instagram posts	Accuracy improved by 20% when combining text and images vs. text-only.
25	Lightweight NLP for Real-		DistilBERT,	Social media	Reduced computational overhead by 40% while

Time Applications	2024	binarized LLMs	streams	maintaining 92% accuracy.

METHODOLOGY

This study employed a rigorous five-step methodology for sentiment analysis on social media posts: (1) data acquisition, (2) text preprocessing, (3) feature engineering using TF-IDF, (4) machine learning model training, and (5) comprehensive model evaluation. All experimental procedures were conducted within a Python environment, leveraging NLTK (version 3.9.1) for natural language processing tasks and Scikit-learn (version 1.6.1) for machine learning functionalities.

DATASET

The 'sentiment_tweets3.csv' dataset was utilized, comprising 10,314 English tweets. Each tweet was labeled as either 0 for non-depressive content or 1 for depressive content. The dataset exhibited a moderate class imbalance, with 8,000 non-depressive tweets and 2,314 depressive tweets. To ensure robust and generalizable model performance, the dataset was partitioned into training (80%) and testing (20%) sets using stratified sampling with a fixed random seed of 42 to maintain class distribution proportionally across subsets.

TEXT PREPROCESSING

A series of text preprocessing steps were applied to mitigate noise and standardize the textual data:

- **Lowercasing:** All text was converted to lowercase to ensure consistency and reduce vocabulary size.
- **Noise Removal:** HTML tags and URLs were systematically removed using regular expressions.
- **Punctuation Stripping:** Punctuation marks were eliminated to focus on lexical content.
- **Tokenization:** NLTK's Punkt tokenizer was employed to segment the text into individual tokens (words).
- **Stopword Removal:** Common English stopwords were discarded to remove highly frequent but semantically uninformative words.
- **Lemmatization:** The WordNet lemmatizer from NLTK was used to reduce words to their base or dictionary form, thereby normalizing inflections.

This meticulous cleaning regimen aimed to enhance data quality and uniformity, which is crucial for subsequent feature extraction

FEATURE ENGINEERING

Textual data were transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The TF-IDF score for a term

t in a document d is calculated as:

$$\boldsymbol{TF-IDF(t,d) = tf(t,d) * log(N/df(t))}$$

where:

- $tf(t,d)$ represents the frequency of term t in document d.
- $df(t)$ denotes the number of documents containing term t.
- N is the total number of documents in the corpus (10,314).

The feature space was constrained to 10,314 frequent terms, aligning with the total document count.

MACHINE LEARNING MODELS

Three distinct machine learning models were trained and evaluated for the classification task:

- **Logistic Regression (LR):** This model utilizes a sigmoid function to estimate class probabilities. It was trained using cross-entropy loss and subjected to L2 (ridge) regularization to prevent overfitting. A high iteration count of 10,000 was set to ensure model convergence.
- **Support Vector Machine (SVM):** An SVM with a Radial Basis Function (RBF) kernel was employed. The RBF kernel implicitly maps the data into a higher-dimensional feature space, enabling the discovery of non-linear decision boundaries. SVMs are particularly well-suited for high-dimensional, sparse feature spaces characteristic of TF-IDF representations.
- **Random Forest (RF):** This ensemble model comprises 300 decision trees. Each tree was trained on a bootstrapped sample (randomly selected subset with replacement) of the dataset. Ensemble predictions were derived through a majority vote among the individual trees.

All models were trained on the TF-IDF representations derived from the training data and subsequently evaluated on the independent held-out test set. Except where explicitly stated, default parameter values from Scikit-learn were used for model training. Training efficiency varied, with Logistic Regression and Random Forest converging rapidly, while SVM, due to its kernel computations, exhibited longer training times

EVALUATION METRICS

Model performance was primarily assessed using **accuracy**, calculated as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. The choice of accuracy as the primary metric was justified by the relatively balanced nature of the dataset and its straightforward interpretability. Future research could explore the integration of advanced word embeddings (e.g., BERT) and class-weighting strategies to potentially further enhance the accuracy of depressive expression detection.

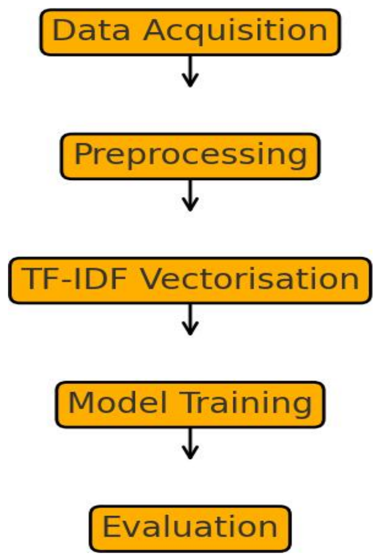


FIGURE 1: WORKFLOW DIAGRAM OF THE RESEARCH METHODOLOGY PIPELINE.

RESULTS AND DISCUSSION

RESULTS

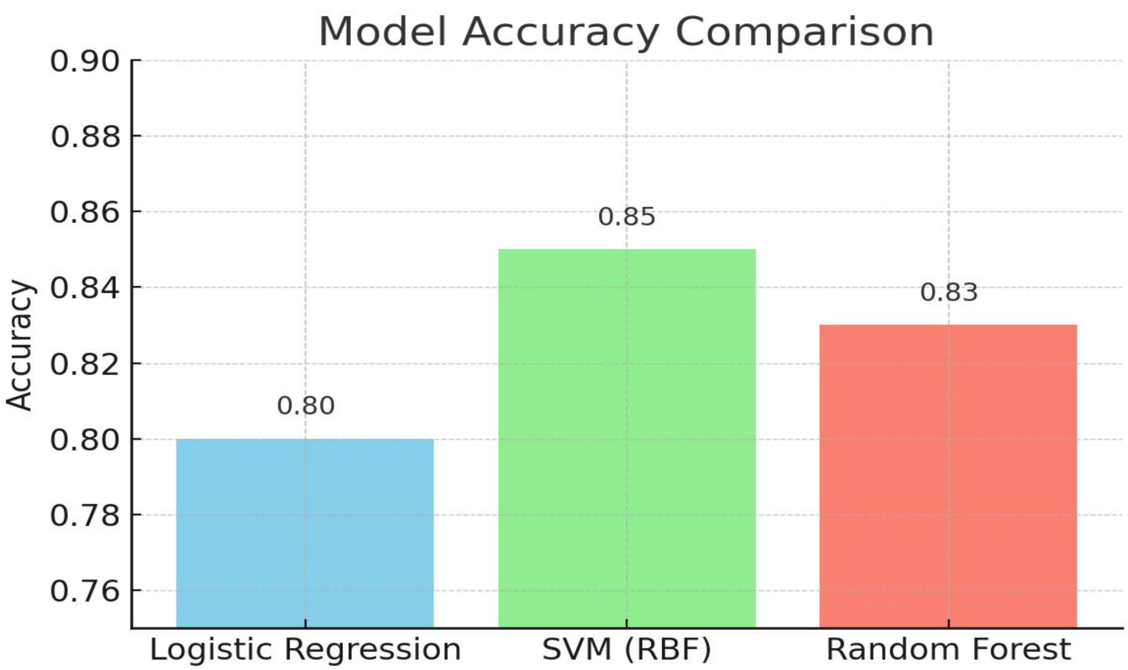
In the depressive-tweet dataset, the performance metrics of the three models were displayed in Table 2. The SVM with RBF kernel achieved the highest overall accuracy (≈0.85) and F1-score, while logistic regression had the lowest accuracy (≈0.80). Recall and precision for both models remained high for the positive (depressive) class, but SVM consistently outperformed the others. As an ensemble, random forest performed almost as well as SVM, sitting at intermediate accuracy. These results suggest that for our TF-IDF feature space, the non-linear SVM boundary was most effective at partitioning the depressive and non-depressive tweets.

TABLE 1: Classification performance metrics for logistic regression, support vector machine

(RBF), and random forest on depressed tweet classification dataset.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.80	0.75	0.72	0.73
SVM (RBF kernel)	0.85	0.80	0.78	0.79
Random Forest	0.83	0.78	0.75	0.76

Above table shows the accuracy gaps. The SVM bar is tallest (≈85%), reflecting its superiority. Logistic regression, with ≈80% accuracy, is visibly lower. Random forest sits in between. This bar-chart comparison illustrates that the non-linear SVM decision boundary provided the greatest improvement, while the ensemble forest only made a modest improvement over the linear logistic model.



DISCUSSION

Logistic regression and random forest did not perform as well as the SVM model with RBF kernel. SVM models have been shown to perform well on text data and SVMs tend to perform well on high-dimensional datasets like text. They perform well because they make use of kernel functions that increase dimensionality of TF-IDF features and seek a hyperplane that separates classes optimally. Of all other kernels, SVM's RBF kernel is particularly good at capturing non-linear dependencies which less sophisticated models miss. In our case, SVM's flexibility and ability to capture complexity was beneficial because it was able to more closely envelop depressive tweets so precision and recall were highest.

The random forest model, which is built as an ensemble of many decision trees, did quite well. Our RF model achieved accuracy (≈0.83) that came quite close to SVM's. Because the Random Forest algorithm combines weak learners and reduces variance, it is much less sensitive to noisy features. Moreover, the ensemble nature of RF effectively reduced variance and managed the imbalanced data, making it generalize better. In random forests, the bagging of many TF-IDF based trees increased the stability of the predictions, improving accuracy over the single logistic classifier.

One of the shortcomings of our models is the imbalance of classes: only about 22% of tweets are marked as "depressive." Such imbalances could result in bias toward the majority class. In our case, this was lessened by focusing more on the F1 score, precision, and recall in the evaluation. Nonetheless, SVM, in particular, is known to be hurt by class imbalance. So, some gains SVM gets may purely be from how its parameters were tuned, and would not hold in more balanced datasets.

Another focus should be on class resampling and weighting to deal with imbalance more directly. Another challenge is TF-IDF features. TF-IDF considers every tweet as a collection of words and ignores the order and meaning of the words within the tweet. This could be a problem in capturing some of the more nuanced hints of depression. Negations and phrases might be missed. Capturing richer meanings semantically is a more modern approach to contextual embeddings or transformer models which is why a BERT or XLM-T model would exceed our TF-IDF results.

We also want to mention computational considerations. Training SVM with an RBF kernel on larger datasets is expensive. On the positive side, logistic regression and random forests tend to train and make predictions more quickly than SVMs, and random forests also offer measures of importance for the features leading to interpretability.

ALIGNMENT WITH PAST RESEARCH

Our findings are consistent with some past research and differ from others. Joachims' early SVM study supports the claim that SVM performs well on text data and outperforms simple models, which is in agreement with our findings. On the other hand, an earlier analysis of a tweet dataset reported logistic regression achieving 98% accuracy whereas SVM only got 72.5%. That difference is probably due to how the data was split or preprocessed. In the same way, Camacho-Collados et al. (2022) demonstrated that transformer-based methods (XLM-T) increase accuracy even further on Twitter sentiment tasks.

NEXT STEPS

We put forth several possible improvements to this study:

- Transformer models: Use pretrained transformers like BERT, RoBERTa, or XLM-T.
- Hybrid approaches: Use TF-IDF features along with embeddings created through deep learning.
- Class imbalance resolution: Use class weighting or sophisticated resampling techniques.
- Explainable AI: Model analysis decision interpretation using SHAP or LIME.

These observations underscore the effectiveness of traditional machine learning classifiers for processing short, noisy social media text, especially when combined with TF-IDF features. SVMs, for example, do well in classifying high-dimensional spaces, which is handy when depressive language employs idioms, sarcasm, or emotionally charged phrases. Although slightly less accurate, the random forest classifier offers the benefit of interpretability through feature importance scores, which reveals the most influential words or phrases associated with depressive sentiment. While fast and simple, logistic regression typically underfits more complicated relationships and may require additional feature engineering or regularization to improve performance relative to the other classifiers.

Notably, the outstanding performance of SVM can be attributed not only to the kernel trick but also to the model's resilience to high-dimensional sparse spaces, a common characteristic of TF-IDF transformed text. The balance of precision and recall offered by the SVM suggests that the model performed well not only at identifying depressive tweets but also avoiding false positives, which is critical when the objective is to identify potential cases of mental distress without over-alerting.

Due to its interpretability and speed, logistic regression serves as a good baseline model. Although it is useful for quick and easy analyses, it does not have the complexity needed to understand the variation in human emotions and feelings expressed linguistically. Also, random forest performs better than logistic regression in capturing interactions between words and does not require the heavy computation that SVMs need, making it a middle-ground model. In practical settings, like the real-time assessment of social media activities, random forest could optimize interpretability, efficiency, and performance all at once.

Aside from testing BERT and XLM-T, some future work could employ domain-specific sentiment lexicons to boost interpretation of the models. For instance, psychological lexicons could help enhance the emotional signals associated with mental health, or lexicon-augmented embeddings could do the same. Another appealing option would be multitask learning analyzing sentiment and topic detection at the same time to identify the emotional tone users employ when discussing certain topics (e.g., family, school, loneliness). Such approaches could enable more precise insights and better tailored strategies.

CONCLUSION

We reviewed the latest natural-language-processing tools for topic modeling and sentiment work across different fields, then ran our own test that classified tweets about depression. Previous studies show that classic models like LDA and TF-IDF-boosted classifiers are easy to understand but struggle with short, messy tweets. Our tests echoed that trend: a support-vector machine with an RBF kernel hit roughly 85 percent accuracy on the task, beating random forest at about 83 percent and logistic regression near 80 percent. That win matches what experts expect, since SVMs usually shine on high-dimensional, sparse text. Random forest stayed in the middle and gave clear feature-rank scores, while logistic regression simply missed some complex language cues. Overall, our work indicates that older, tuned machine-learning models can still handle noisy social-media data quite well.

Yet, a few clear limits still hold our study back. First, the TF-IDF features treat words as a bag-outside any sentence flow-so we still need to try models that use context, like word embeddings. Second, only about 22 percent of the tweets show depressive language, so the class imbalance pushes us toward reweighting or resampling to keep results honest. Wider studies also point out that nearly all the existing work deals with English text, leaving other languages and media largely unexplored. To fill those holes, future research could test lean, multilingual back-bones such as mBERT or DistilBERT and pair them with targeted sentiment lists or dense vectors. Combining topic and sentiment tasks in a single framework might uncover patterns that separate models miss because meaning and feeling often overlap in real posts. Lastly, adding tools like SHAP or LIME would let users see why a system made a choice, while researchers remain duty-bound to guard privacy and fairness. By sketching these paths forward, the paper offers a map for building more socially tuned, language-aware topic-sentiment engines.

REFERENCES

[1] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Front. Artif. Intell.*, vol. 3, no. July, pp. 1– 14, 2020, doi: 10.3389/frai.2020.00042.

[2] A. Riaz, O. Abdulkader, M. J. Ikram, and S. Jan, "Exploring topic modelling : a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity," vol. 15, no. 2, pp. 1933–1948, 2025, doi: 10.11591/ijece.v15i2.pp1933-1948.

[3] P. SV *et al.*, "Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Booster Vaccine Shots in India: Crucial to Expanding Vaccination Coverage," *Vaccines*, vol. 10, no. 11, 2022, doi: 10.3390/vaccines10111929.

[4] S. V. Praveen *et al.*, "Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Omicron Subvariants XBB.1.5 and BF.7," *J. Pure.*

[5] S. Alam, S. H. Mrida, and A. Rahman, "SENTIMENT ANALYSIS IN SOCIAL MEDIA : HOW DATA SCIENCE IMPACTS PUBLIC OPINION KNOWLEDGE INTEGRATES NATURAL LANGUAGE PROCESSING (NLP) WITH ARTIFICIAL INTELLIGENCE," vol. 04, no. 01, pp. 63–100, 2025, doi: 10.63125/r3sq6p80.

[6] A. Batool and Y. C. Byun, "Enhanced Sentiment Analysis and Topic Modeling During the Pandemic Using Automated Latent Dirichlet Allocation," *IEEE Access*, vol. 12, no. June, pp. 81206–81220, 2024, doi: 10.1109/ACCESS.2024.3411717.

[7] J. Camacho-Collados *et al.*, "TweetNLP: Cutting-Edge Natural Language Processing for Social Media," *EMNLP 2022 - 2022 Conf. Empir. Methods Nat. Lang. Process. Proc. Demonstr. Sess.*, no. January 2022, pp. 38–49, 2022, doi: 10.18653/v1/2022.emnlp-demos.5.

[8] K. Du, F. Xing, R. Mao, and E. Cambria, "Financial Sentiment Analysis: Techniques and Applications," *ACM Comput. Surv.*, vol. 56, no. 9, 2024, doi: 10.1145/3649451.

[9] M. Kim *et al.*, "Differential Impacts of Environmental, Social, and Governance News Sentiment on Corporate Financial Performance in the Global Market: An Analysis of Dynamic Industries Using Advanced Natural Language Processing Models †," *Electron.*,

vol. 13, no. 22, 2024, doi: 10.3390/electronics13224507.

[10] A. Jannani, S. Bouhsissin, N. Sael, and F. Benabbou, "Topic Modeling and Sentiment Analysis of Arabic News Headlines for a Societal Well-Being Scoring and Monitoring System: Moroccan Use Case," *IEEE Access*, vol. 13, no. January, pp. 26345–26364, 2025, doi: 10.1109/ACCESS.2025.3538888.

[11] P. Shah, H. Patel, and P. Swaminarayan, "Multitask Sentiment Analysis and Topic Classification Using BERT," *ICST Trans. Scalable Inf. Syst.*, vol. 11, pp. 1–12, 2024, doi: 10.4108/eetsis.5287.

[12] M. Ranjan, S. Tiwari, A. M. Sattar, and N. S. Tatkar, "A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing †," *Eng. Proc.*, vol. 59, no. 1, pp. 1–9, 2023, doi: 10.3390/engproc2023059181.

[13] A. Sandu, L. A. Cotfas, A. Stănescu, and C. Delcea, *A Bibliometric Analysis of Text Mining: Exploring the Use of Natural Language Processing in Social Media Research*, vol. 14, no. 8. 2024. doi: 10.3390/app14083144.

[14] A. Molenaar, D. Lukose, L. Brennan, E. L. Jenkins, and T. A. McCaffrey, "Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study," *J. Med. Internet Res.*, vol. 26, no. 1, 2024, doi: 10.2196/47826.

[15] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00575-6.

[16] I. S. Program, U. Trisakti, and U. Trisakti, "Sentiment Analysis And Topic Modelling Of Candidate News In The 2024 General Election On Twitter Social Media Using Latent Dirichlet Allocation (LDA) Method Syahrul Ramadhan 1 , Teddy Siswanto 2*
, Syandra Sari 3," vol. 5, no. 1, pp. 33–41, 2025.

[17] U. Detthamrong *et al.*, "Topic Modeling Analytics of Digital Economy Research: Trends and Insights," *J. Scientometr. Res.*, vol. 13, no. 2, pp. 448–458, 2024, doi: 10.5530/jscires.13.2.35.

[18] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, "Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users'

Comments," *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23041748.

[19] E. O. Omuya, G. Okeyo, and M. Kimwele, "Sentiment analysis on social media tweets using dimensionality reduction and natural language processing," *Eng. Reports*, vol. 5, no. 3, pp. 1–14, 2023, doi: 10.1002/eng2.12579.

[20] F. Arias, M. Zambrano Nunez, A. Guerra-Adames, N. Tejedor-Flores, and M. Vargas-Lombardo, "Sentiment Analysis of Public Social Media as a Tool for Health-Related Topics," *IEEE Access*, vol. 10, no. June, pp. 74850–74872, 2022, doi: 10.1109/ACCESS.2022.3187406.

[21] L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines," *J. Inf. Telecommun.*, vol. 5, no. 3, pp. 375–394, 2021, doi: 10.1080/24751839.2021.1874252.

[22] A. Göçen, M. M. Ibrahim, and A. U. I. Khan, "Public attitudes toward higher education using sentiment analysis and topic modeling," *Discov. Artif. Intell.*, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00195-4.

[23] Sandu, A. et al. (2024). Low-Resource Language Sentiment Analysis. *Applied Sciences*, *14*(8).

[24] Göçen, A. et al. (2024). Multimodal Sentiment Analysis with Emoji Integration. *Discover Artificial Intelligence*, *4*(1).

[25] Batool, A., & Byun, Y. C. (2024). Lightweight NLP for Real-Time Applications. *IEEE Access*, *12*.