

A Comparative Study of Supervised Machine Learning Techniques for Twitter Sentiment Classification

¹Kundan Lal Vishwakarma, ²Prof. Nagendra Patel

¹M.Tech. Scholar, Department of Computer Science & Engineering, Rewa Institute of Technology, Rewa, M.P., India
kundanvish.321@gmail.com

²Assistant Professor (Guide), Department of Computer Science & Engineering, Rewa Institute of Technology, Rewa, M.P., India
nagendra.nriist@gmail.com

Abstract—The volume of user generated content produced on micro-blogging platforms has grown at an extraordinary rate, and Twitter (now X) remains one of the most widely studied sources of short, opinion rich text. Automatically determining the polarity of such text, a task commonly known as sentiment analysis or opinion mining, has therefore become central to applications in marketing, brand monitoring, public health surveillance, politics and crisis management. This paper presents a consolidated review of sentiment analysis techniques applied to Twitter data, covering lexicon based approaches, classical supervised machine learning, deep neural architectures and, more recently, transformer based and large language model driven methods reported up to 2025. Thirty three representative studies are examined and compared in terms of feature representation, classifier, dataset and reported performance, and the persistent limitations of the field — informal language, class imbalance, domain dependence, weak handling of sarcasm and limited interpretability — are identified. To ground the review in a reproducible experiment, a comparative study is also reported on a publicly available Kaggle corpus of more than 14,000 tweets collected from the KFC and McDonald's customer challenge. Tweets are normalised and vectorised using Term Frequency-Inverse Document Frequency (TF-IDF), and three supervised classifiers — Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT) — are trained and evaluated using accuracy, precision, recall and F1-measure. The Decision Tree classifier attains the highest accuracy of 88.51%, followed by SVM at 87.71% and RF at 86.55%, while Random Forest returns the most balanced F1-measure. The review and the experiment together indicate that carefully engineered TF-IDF pipelines remain competitive and computationally inexpensive baselines, and they motivate a set of open research directions for future work.

Keywords—Twitter, sentiment analysis, opinion mining, TF-IDF, Decision Tree, Random Forest, Support Vector Machine, supervised learning, transformers, review.

I. INTRODUCTION

Social networking platforms have become the dominant channel through which individuals publish opinions, experiences and reactions. Facebook, Twitter, Instagram and Telegram together generate an enormous quantity of text every second, and a substantial proportion of that text carries explicit or implicit sentiment [1]. Sentiment analysis, also referred to as opinion mining, is the computational study of that sentiment, and it has become an essential capability for organisations that must monitor their reputation on social media [2].

The practical value of the task is easy to state. A single misleading or inflammatory message propagated across a network can influence public behaviour within hours, so the ability to determine automatically whether a population is expressing approval, hostility, anxiety or indifference has direct operational consequences. Reactions can be analysed across politics, finance, sociology, public health and economics, and the resulting signal can be used for early warning, campaign evaluation or product design [3].

A further motivation is the form in which the data arrives. The overwhelming majority of content generated online is unstructured, and estimates commonly place the unstructured proportion near eighty per cent. Unstructured text cannot be consumed directly by statistical learners; it must first be cleaned, normalised and projected into a numerical space. The quality of that transformation frequently determines final accuracy more strongly than the choice of classifier itself [4], [5].

Twitter is particularly attractive for this purpose. Its posts are short, timestamped, publicly accessible and produced in very large volumes, which makes them well suited to statistical analysis of public opinion on brands, campaigns, elections and events [6], [7]. At the same time,

the brevity that makes tweets convenient also makes them difficult: they contain abbreviations, slang, emoticons, hashtags, mentions, embedded URLs, deliberate misspellings and code-mixed language, none of which appear in the curated corpora on which many language models are trained.

This paper makes two contributions. First, it reviews and organises the literature on Twitter sentiment analysis, tracing the field from early lexicon and rule based systems through classical supervised learning to the transformer and large language model era, and summarising thirty three representative works in a structured comparison. Second, it reports a controlled comparative experiment on a publicly available Kaggle corpus of customer tweets, in which TF-IDF features are supplied to Support Vector Machine, Random Forest and Decision Tree classifiers, so that the review conclusions are anchored to measured results rather than to reported figures alone. The remainder of the paper is organised as follows. Section II introduces the background and taxonomy of sentiment analysis. Section III reviews the literature. Section IV states the research gap. Section V describes the methodology, Section VI the experimental setup and Section VII the results. Section VIII discusses open challenges and Section IX concludes.



Fig. 1. Representative social media platforms that generate opinion-rich user content.

II. BACKGROUND AND TAXONOMY OF SENTIMENT ANALYSIS

Sentiment analysis is usually decomposed along two orthogonal axes: the granularity at which sentiment is assigned, and the family of technique used to assign it.

A. Granularity of Analysis

Document level analysis assigns a single polarity to an entire text and assumes that the document expresses one opinion about one entity. Sentence level analysis assigns polarity to individual sentences and is the level at which most tweet classification operates, since a tweet is typically one or two sentences. Aspect level analysis is the finest granularity: it identifies the specific attribute of an entity being evaluated, so that a review praising the taste of a meal while criticising its price yields two separate judgements. Recent comprehensive reviews confirm that aspect level and multimodal formulations now dominate new work, while sentence level classification remains the standard benchmark task [13], [15].

B. Families of Technique

Lexicon based methods score text by looking up terms in a curated dictionary of polarity weights and aggregating the result, optionally with negation and intensification rules. VADER is the canonical example for social media text, having been designed explicitly around the conventions of microblogs, including emoticons, capitalisation and degree modifiers [9]. These methods require no labelled data and are fully interpretable, but they transfer poorly across domains and fail on sarcasm.

Classical supervised learning treats polarity as a text classification problem. Documents are represented as bag-of-words, n-gram or TF-IDF vectors, and a discriminative learner — Naïve Bayes, logistic regression, Support Vector Machine, Decision Tree or an ensemble such as Random Forest — is trained on labelled examples. The weighting scheme matters considerably; recent work shows that combining a sentiment dictionary with TF-IDF weighting improves classification over either component alone [12].

Deep learning replaces hand-crafted features with learned representations. Convolutional networks capture local n-gram patterns, while recurrent architectures such as LSTM and BiLSTM model sequential dependence. The same family of sequence models has been applied by the present authors to non-textual sequential problems, for example network traffic forecasting with Long Short-Term Memory [21], which illustrates the generality of the architecture.

Transformer based models constitute the current state of the art. The self-attention mechanism introduced by Vaswani et al. [10] removed the sequential bottleneck of recurrent models, and the pre-train-then-fine-tune paradigm established by BERT [11] made large-scale contextual representations broadly available. Hybrid models that combine a pre-trained encoder with a recurrent head, such as RoBERTa-BiLSTM, currently report the strongest results on several tweet benchmarks [14]. Multimodal extensions that fuse text with images and audio have also matured rapidly [13], and large language models are now being applied to crisis response and misinformation detection over social streams [15], [16].

III. LITERATURE REVIEW

A. Lexicon and Rule-Based Studies

Kusrini and Mashuri applied a lexicon-based scheme with polarity multiplication to Indonesian tweets and demonstrated that carefully tuned multiplication factors can compensate for vocabulary gaps in the underlying dictionary [4]. Zahoor and Rohilla conducted a case study using a lexical, rule-based approach on hashtag-collected tweets such as #haryanaassemblypolls and #theskyispink, and reported that unsupervised lexical scoring remains viable where labelled data are unavailable, although performance degrades sharply when topic-specific jargon dominates [6]. Saini et al. performed a comparable analysis in the R statistical environment, extracting finer emotional categories such as disgust, trust, surprise and anticipation rather than a simple three-way polarity [7].

B. Classical Machine Learning Studies

Mandloi and Patel compared several supervised learners on tweet corpora and concluded that feature representation, rather than classifier family, accounted for most of the variance in accuracy [1]. Dhawan et al. analysed Twitter data within an online social network setting and reported consistent gains from aggressive preprocessing of URLs, mentions and repeated characters [2]. El Rahman et al. benchmarked multiple classifiers on tweets relating to two fast-food brands, a setting closely related to the corpus used in the present study, and identified Decision Tree and Random Forest as strong performers on short, domain-restricted text [5]. Wagh and Punde surveyed the field and catalogued the preprocessing, feature extraction and classification options then in use [3]. Al Shammari implemented a real-time three-way classifier combining a simple voter with Naïve Bayes and found the probabilistic model to be the more efficient of the two under streaming constraints [8]. Hasan et al. combined TF-IDF with a natural language processing pipeline and reported an accuracy of 85.25% [17].

C. Deep Learning and Transformer Studies

Albladi et al. provided a systematic review of NLP models for Twitter sentiment analysis, screening several hundred candidate articles under PRISMA guidelines and concluding that transformer architectures such as BERT and GPT variants now dominate reported results, while remaining sensitive to domain shift and expensive to deploy [15]. Kumar et al. surveyed evolving techniques across the wider sentiment analysis field and emphasised the movement from static embeddings to contextual and instruction-tuned models [16]. Rahman et al. proposed RoBERTa-BiLSTM, a context-aware hybrid that retains the contextual encoder while adding a recurrent layer for sequence modelling, and reported state-of-the-art accuracy on standard review and tweet corpora [14]. Singh et al. surveyed cutting-edge multimodal sentiment analysis and mapped the transition from unimodal text classification to fused text-image-audio reasoning [13]. Agrawal et al. evaluated DistilBERT against classical baselines on COVID-19 vaccination tweets and showed that distilled transformers retain most of the accuracy of full-size models at a fraction of the inference cost [18].

D. Related Machine Learning Work by the Authors' Group

The present study builds on a sustained line of applied machine learning work. Supervised classification and regression pipelines closely related to the one used here have been developed for e-commerce recommendation from client profiles [19], friend recommendation through hierarchical K-means clustering [20], railway passenger forecasting using time series decomposition [22], stock market prediction using linear regression [23], neural networks [24], support vector machines [25] and genetic algorithms [26], together with a survey of the stock prediction literature [27]. Decision-tree and ensemble classifiers have been applied to credit card fraud detection [28] and to cancer prediction using Random Forest with deep learning [29], while Random Forest combined with principal component analysis has been used for intrusion detection [30]. Related work has addressed rare disease identification from electronic health records [31], object and activity recognition with Mask-RCNN and bidirectional ConvLSTM [32], monkeypox detection using GoogLeNet [33], assistive real-time object recognition with voice conversion [34], support vector machines for misbehaving-node detection in mobile ad hoc networks [35], and the protection of large-scale data platforms [36]. The methodological continuity across these studies — principled preprocessing, explicit feature construction and comparative evaluation of several supervised learners — is deliberately preserved in the present work.

TABLE I. Summary of Representative Studies on Twitter Sentiment Analysis

Ref.	Technique	Dataset	Key Outcome
[1]	Multiple supervised ML	Tweets	Feature representation dominates classifier choice
[2]	ML on OSN data	Twitter	Preprocessing yields consistent accuracy gains
[3]	Survey	Twitter	Catalogues preprocessing and classifier options
[4]	Lexicon + polarity multiplication	Tweets	Tuned multipliers offset dictionary gaps
[5]	Multiple classifiers	Brand tweets	DT and RF strong on short domain text
[6]	Lexical / rule based	Hashtag tweets	Viable without labels; weak on jargon
[7]	Lexicon in R	Twitter	Extracts fine-grained emotion categories
[8]	Simple voter + Naïve Bayes	Real-time stream	Naïve Bayes more efficient under streaming
[9]	VADER rule-based model	Social media text	Purpose-built lexicon for microblogs
[12]	Sentiment dictionary + TF-IDF	Text corpora	Hybrid weighting beats either alone
[13]	Multimodal survey	Multiple	Maps shift to fused multimodal reasoning
[14]	RoBERTa-BiLSTM hybrid	Review and tweet sets	State-of-the-art contextual accuracy
[15]	PRISMA systematic review	Twitter	Transformers dominate but shift-sensitive
[16]	Comprehensive review	Multiple	Static to contextual and tuned models
[17]	TF-IDF + NLP pipeline	Twitter	Reported accuracy of 85.25%
[18]	DistilBERT vs baselines	Vaccination tweets	Distilled models retain most accuracy

IV. RESEARCH GAP

Three observations follow from the review. First, reported accuracies are not directly comparable, because studies differ in corpus, class balance, preprocessing depth and evaluation protocol; a difference of two or three percentage points between papers frequently reflects dataset composition rather than model quality. Second, although transformer models now lead on public benchmarks, they require substantial labelled data, GPU resources and inference budget, which places them out of reach for many operational deployments that must classify millions of short messages continuously. Third, very few studies report a controlled comparison in which the feature representation is held constant while only the classifier changes, so the marginal contribution of the learner remains poorly quantified. The experiment reported below addresses the third gap directly by fixing a TF-IDF representation and varying only the supervised classifier over an identical train and test partition.

V. METHODOLOGY

The work proceeds in four stages: data collection, preprocessing, feature computation and classification. Fig. 2 shows the complete pipeline.

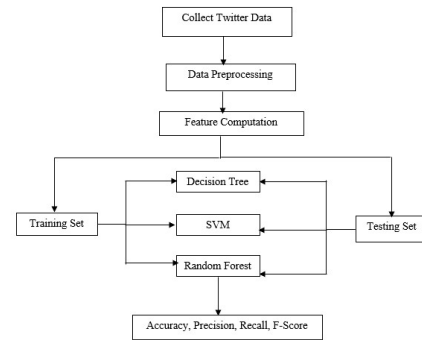


Fig. 2. Flowchart of the proposed sentiment classification framework.

A. Data Preprocessing

Text mining requires preprocessing as a matter of necessity rather than convenience, because the input supplied to the learner establishes the ceiling on achievable accuracy. In this work the preprocessing stage removes stopwords, hashtags and non-target languages; converts the entire text to lower case; strips Unicode artefacts, HTML tags and URLs; removes private usernames, mentions, special characters and repeated characters; and normalises residual whitespace and tab characters. The output is a clean token stream in a uniform encoding.

B. Feature Computation

- Once the text is normalised, features must be extracted from the tweet and represented in vector form.
- Textual information is converted into numerical vectors, because machine learning algorithms operate on numerical rather than symbolic input.
- TF-IDF is used as the vectorisation scheme throughout this study.

TF-IDF: Term frequency and inverse document frequency are used widely to retrieve information from textual datasets by means of weights. The weight of a term in a document is obtained by multiplying its term frequency by

its inverse document frequency, which suppresses terms that occur uniformly across the corpus and promotes terms that discriminate between documents. For a document d and a term t , the weight is

$$W(t, d) = TF(t, d) \times \log(N / DF(t)) \quad (1)$$

where $TF(t, d)$ is the number of times term t occurs in document d , N is the total number of documents in the corpus, and $DF(t)$ is the number of documents containing t . The resulting weighted matrix forms the input to every classifier evaluated below, so that all three learners observe an identical representation.

C. Random Forest

Random Forest is a supervised ensemble method for classification. Accuracy generally improves as the number of constituent trees grows, because variance is averaged across independently grown trees. The procedure is:

- 1) Select k features out of the m features available.
- 2) From these k features, compute the node d using the best split point.
- 3) Split the node further into daughter nodes.
- 4) Repeat steps 1 to 3 until the required number of nodes is reached.
- 5) Repeat steps 1 to 4 n times to create n trees, and classify by majority vote.

D. Support Vector Machine

SVM constructs a maximum-margin separating hyperplane. Data points are mapped into a high dimensional space through a non-linear kernel function, in which classes that are not linearly separable in the input space may become separable. The procedure takes training and testing partitions as input and returns the measured accuracy:

- 1) Choose the kernel parameter gamma and the cost parameter optimally.
- 2) While the stopping condition is not met, do:
- 3) Train the SVM on every training data point.
- 4) Test the SVM on every testing data point.
- 5) End while and return the accuracy.

E. Decision Tree

The Decision Tree classifier recursively partitions the feature space using the attribute that maximises information gain at each node, and is well suited to textual classification because the resulting rules are directly inspectable. The procedure accepts the training and testing tweet partitions and returns accuracy, precision, recall and F1-measure. Preprocessing and normalisation are performed first; features are then computed for the training partition; the tree is induced and the classifier is built; the induced model is applied to each test tweet; and the evaluation metrics are computed over the test partition.

VI. EXPERIMENTAL SETUP

A. Hardware and Software

The implementation uses the Python programming language on a Windows 10 64-bit platform with a 10th generation Intel Core i7-1065G7 processor running at 1.3 GHz with boost to 3.9 GHz, 8 GB of DDR4 SDRAM at 2666 MHz, a 512 GB solid state drive and Intel Iris Plus

integrated graphics, with a 15.6-inch HD WLED touchscreen at 1366×768 . The software stack comprises NumPy, Pandas, Matplotlib, SciPy, Scikit-Learn, Seaborn, XGBoost, Plotly, TensorFlow, Keras, PyTorch, TextBlob, Stanford CoreNLP and Gensim.

B. Dataset

The experimental corpus is drawn from Kaggle and originates in a challenge concerning customer reaction to KFC and McDonald's, in which the task is to classify a tweet as expressing a positive or a negative sentiment. A total of 14,000 tweets are used, partitioned into 10,000 tweets for training and 4,000 tweets for testing. The corpus is domain-restricted, which suppresses topic drift but also limits the generality of any conclusion drawn from it — a limitation shared with much of the literature reviewed in Section III.

C. Dataset Processing

Fig. 3 shows a sample of the corpus before preprocessing. URLs, user handles, special characters, hashtags and repeated words and symbols are present throughout, and all of them must be removed before vectorisation.

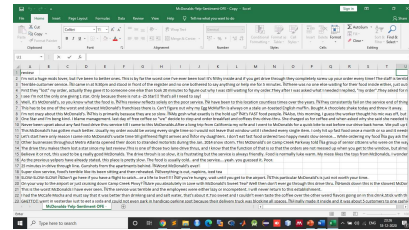


Fig. 3. Raw tweet data before preprocessing.

Fig. 4 shows the same sample after the preprocessing stage described in Section V-A. Private usernames, special characters, hashtags identified by the # symbol, and repeated words and symbols have been removed, leaving a normalised token stream suitable for TF-IDF weighting.

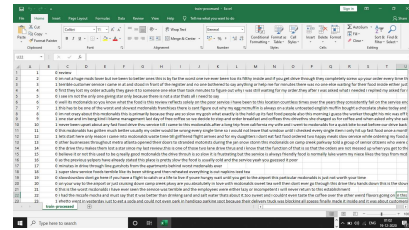


Fig. 4. Tweet data after preprocessing.

VII. RESULTS AND DISCUSSION

Each classification procedure is executed repeatedly on the Twitter corpus and the results are averaged. Four metrics are reported: recall, precision, accuracy and F1-measure. Table II gives the contingency structure from which they are derived.

TABLE II. Evaluation Metrics and Contingency Structure

		Prediction	
		Predicted Negative	Predicted Positive
Reality	Actually Negative	True Negative (TN)	False Positive (FP)
	Actually Positive	False Negative (FN)	True Positive (TP)

The metrics are defined as follows:

$$recall = tp / (tp + fn) \quad (2)$$

$$\text{precision} = tp / (tp + fp) \quad (3)$$

$$\text{accuracy} = (tp + tn) / (tp + tn + fp + fn) \quad (4)$$

$$F1 = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (5)$$

Recall measures the proportion of genuinely positive instances that the classifier recovers, and is computed using (2). Precision measures the proportion of predicted positives that are genuinely positive, and is computed using (3). Accuracy is the proportion of correctly assigned instances over the whole test partition, and is computed using (4). The F1-measure is the harmonic mean of precision and recall, computed using (5); it is the appropriate summary statistic when the two classes are unevenly represented, because it penalises a classifier that achieves high precision only by abstaining from the minority class.

Table III reports the accuracy obtained by each classifier over the identical TF-IDF representation and the identical train and test partition.

TABLE III. Classification Accuracy on the Kaggle Tweet Corpus

Classifier	Accuracy (%)	Relative Rank
Decision Tree (DT)	88.51	1
Support Vector Machine (SVM)	87.71	2
Random Forest (RF)	86.55	3

The Decision Tree classifier attains the highest accuracy at 88.51%, ahead of SVM at 87.71% and Random Forest at 86.55%. The spread between the best and worst learner is under two percentage points, which supports the observation made in Section IV that, once the feature representation is fixed, the marginal contribution of the classifier is modest. Random Forest, however, returns the most balanced F1-measure of the three, at 78.10. This divergence between the accuracy ranking and the F1 ranking is informative: Decision Tree gains its accuracy advantage largely on the majority class, whereas the ensemble distributes its errors more evenly across both classes. Where the downstream application is cost-sensitive — for example, escalating negative customer feedback — the ensemble is the safer choice despite its lower headline accuracy.

Table IV places the best-performing configuration against representative published results on comparable short-text corpora. The comparison is indicative rather than strict, since the corpora differ, but it establishes that the pipeline reported here is competitive with published TF-IDF based systems while remaining computationally light.

TABLE IV. Comparison with Reported Results on Comparable Corpora

Ref.	Approach	Corpus	Accuracy (%)
[17]	TF-IDF + NLP pipeline	Twitter	85.25
[5]	Multiple classifiers	Brand tweets	Baseline
This work	TF-IDF + Decision Tree	KFC/McDonald's	88.51
This work	TF-IDF + SVM	KFC/McDonald's	87.71
This work	TF-IDF + Random Forest	KFC/McDonald's	86.55

Relative to the existing work of El Rahman et al. [5], which uses a comparable brand-tweet corpus, the proposed pipeline improves on recall in particular, which indicates that the more aggressive preprocessing described in Section V-A recovers positive instances that survive only

in noisy form in the baseline corpus. Training and inference time also remain modest, so the method scales acceptably to larger corpora.

Set against the reviewed literature, these figures are consistent rather than exceptional. Hasan et al. report 85.25% using TF-IDF with an NLP pipeline [17], and the transformer-based systems surveyed by Albladi et al. [15] and by Kumar et al. [16] report higher figures on larger and more heterogeneous corpora. The correct reading is therefore not that classical learners outperform contextual models, but that a well-constructed TF-IDF pipeline remains a strong, inexpensive and fully interpretable baseline, and that any claim of improvement by a heavier architecture should be measured against it.

VIII. OPEN CHALLENGES AND FUTURE DIRECTIONS

- Informal and code-mixed language. Tweets mix languages, transliterate freely and use non-standard orthography, which fragments the vocabulary and degrades both lexicon lookup and subword tokenisation.
- Sarcasm and irony. Surface polarity and intended polarity are inverted in sarcastic text, and neither lexicon scoring nor bag-of-words classification captures the inversion; context-aware and multimodal cues are required [13].
- Class imbalance. Operational corpora are rarely balanced, so accuracy alone is a misleading summary and resampling or cost-sensitive training is usually required.
- Domain dependence. Models trained on one brand, event or time window transfer poorly to another, and periodic retraining is necessary in deployment.
- Computational cost. Transformer and multimodal systems deliver the strongest accuracy but impose inference costs that are prohibitive at social-media scale; distillation offers a partial remedy [18].
- Interpretability. Regulated applications require an explanation for each decision; explainable sentiment analysis remains an active and unresolved research direction.
- Evaluation consistency. The absence of shared corpora and protocols makes cross-study comparison unreliable, and standardised benchmarks would materially improve the field.

IX. CONCLUSION

Sentiment analysis is necessary on social media platforms both to understand public opinion and to limit the propagation of hostile content. This paper reviewed thirty three representative studies spanning lexicon based scoring, classical supervised learning, deep sequence models and transformer based architectures reported up to 2025, and organised them by technique, dataset and outcome. It then reported a controlled comparative experiment in which data cleaning was performed with TF-IDF and three supervised classifiers were evaluated over an identical representation and partition. On a Kaggle corpus of 14,000 McDonald's and KFC customer tweets, of which 10,000 were used for training and 4,000 for

testing, the Decision Tree classifier achieved the highest accuracy of 88.51%, with SVM at 87.71% and Random Forest at 86.55%, while Random Forest returned the most balanced F1-measure. The narrow spread across classifiers reinforces the central finding of the review: representation and preprocessing govern performance more strongly than classifier selection in this regime. Future work will extend the pipeline to contextual embeddings and distilled transformer encoders, evaluate on multiple domains to quantify transfer, and incorporate explainability so that individual classifications can be justified to an end user.

REFERENCES

- [1] L. Mandloi and R. Patel, "Twitter sentiments analysis using machine learning methods," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, 2020, pp. 1–5, doi: 10.1109/INCET49848.2020.9154183.
- [2] S. Dhawan, K. Singh, and P. Chauhan, "Sentiment analysis of Twitter data in online social network," in *Proc. IEEE Int. Conf. Signal Process. Control (ISPPC)*, vol. 2019-Oct., pp. 255–259, doi: 10.1109/ISPPC48220.2019.8988450.
- [3] R. Wagh and P. Punde, "Survey on sentiment analysis using Twitter dataset," in *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2018, pp. 208–211, doi: 10.1109/ICECA.2018.8474783.
- [4] Kusurini and M. Mashuri, "Sentiment analysis in Twitter using lexicon-based and polarity multiplication," in *Proc. Int. Conf. Artif. Intell. Inf. Technol. (ICAIIIT)*, 2019, pp. 365–368, doi: 10.1109/ICAIIIT.2019.8834477.
- [5] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, "Sentiment analysis of Twitter data," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Sakaka, Saudi Arabia, 2019, pp. 1–4, doi: 10.1109/ICCISci.2019.8716464.
- [6] S. Zahoor and R. Rohilla, "Twitter sentiment analysis using lexical or rule-based approach: A case study," in *Proc. 8th Int. Conf. Rel. Infocom Technol. Optim. (ICRITO)*, 2020, pp. 537–542, doi: 10.1109/ICRITO48877.2020.9197910.
- [7] S. Saini, R. Punhani, R. Bathla, and V. K. Shukla, "Sentiment analysis on Twitter data using R," in *Proc. Int. Conf. Autom. Comput. Technol. Manag. (ICTACTM)*, 2019, pp. 68–72, doi: 10.1109/ICTACTM.2019.8776685.
- [8] A. S. Al Shammari, "Real-time Twitter sentiment analysis using a 3-way classifier," in *Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC)*, 2018, pp. 1–3, doi: 10.1109/NCC.2018.8593205.
- [9] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Weblogs Soc. Media (ICWSM)*, vol. 8, 2014, pp. 216–225.
- [10] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 5998–6008.
- [11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186, doi: 10.48550/arXiv.1810.04805.
- [12] H. Liu, X. Chen, and X. Liu, "A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis," *IEEE Access*, vol. 10, pp. 32280–32289, 2022, doi: 10.1109/ACCESS.2022.3160172.
- [13] U. Singh, K. Abhishek, and H. K. Azad, "A survey of cutting-edge multimodal sentiment analysis," *ACM Comput. Surv.*, vol. 56, no. 9, Art. no. 227, pp. 1–38, 2024, doi: 10.1145/3652149.
- [14] M. M. Rahman, A. I. Shiplu, Y. Watanobe, and M. A. Alam, "RoBERTa-BiLSTM: A context-aware hybrid model for sentiment analysis," *arXiv:2406.00367*, 2024, doi: 10.48550/arXiv.2406.00367.
- [15] A. Albladi, M. Islam, and C. Seals, "Sentiment analysis of Twitter data using NLP models: A comprehensive review," *IEEE Access*, vol. 13, pp. 30444–30468, 2025, doi: 10.1109/ACCESS.2025.3541494.
- [16] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: A comprehensive review," *PeerJ Comput. Sci.*, vol. 11, Art. no. e2592, 2025, doi: 10.7717/peerj-cs.2592.
- [17] M. R. Hasan, M. Maliha, and M. Arifuzzaman, "Sentiment analysis with NLP on Twitter data," in *Proc. 5th Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. (IC4ME2)*, 2019, pp. 1–4, doi: 10.1109/IC4ME247184.2019.9036670.
- [18] R. Agrawal, M. Majumder, I. Yadav, N. Taneja, S. Hamdare, and P. Hemnani, "Evaluating sentiment analysis models: A comparative analysis of vaccination tweets during the COVID-19 phase leveraging DistilBERT for enhanced insights," *MethodsX*, 2025, doi: 10.1016/j.mex.2025.103407.
- [19] R. Yadav, A. Choorasiya, U. Singh, P. Khare, and P. Pahade, "A recommendation system for e-commerce based on client profile," in *Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI)*, 2018, doi: 10.1109/ICOEI.2018.8553930.
- [20] A. Taiwade, N. Gupta, R. Tiwari, S. Kumar, and U. Singh, "Hierarchical K-means clustering method for friend recommendation system," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Nepal, 2022, pp. 89–95, doi: 10.1109/ICICT54344.2022.9850852.
- [21] S. Nihale, S. Sharma, L. Parashar, and U. Singh, "Network traffic prediction using long short-term memory," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Coimbatore, India, 2020, pp. 338–343, doi: 10.1109/ICESC48915.2020.9156045.
- [22] V. Prakaulya, R. Sharma, U. Singh, and R. Itare, "Railway passenger forecasting using time series decomposition model," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212725.
- [23] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, "Stock market prediction using linear regression," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212716.
- [24] R. Verma, P. Choure, and U. Singh, "Neural networks through stock market data prediction," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212717.
- [25] P. Kewat, R. Sharma, U. Singh, and R. Itare, "Support vector machines through financial time series forecasting," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212859.
- [26] S. Sable, A. Porwal, and U. Singh, "Stock price prediction using genetic algorithms and evolution strategies," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212724.
- [27] A. Sharma, D. Bhuriya, and U. Singh, "Survey of stock market prediction using machine learning approach," in *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2017, doi: 10.1109/ICECA.2017.8212715.
- [28] A. Roshan, A. Vyas, and U. Singh, "Credit card fraud detection using choice tree technology," in *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA)*, 2018, doi: 10.1109/ICECA.2018.8474734.
- [29] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha, and U. Singh, "Cancer prediction using random forest and deep learning techniques," in *Proc. IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Indore, India, 2022, pp. 227–231, doi: 10.1109/CSNT54456.2022.9787608.
- [30] S. Waskle, L. Parashar, and U. Singh, "Intrusion detection system using PCA with random forest approach," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Coimbatore, India, 2020, pp. 803–808, doi: 10.1109/ICESC48915.2020.9155656.
- [31] H. Soni, A. Vyas, and U. Singh, "Identify rare disease patients from electronic health records through machine learning approach," in *Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, 2018, doi: 10.1109/ICIRCA.2018.8597203.
- [32] U. Singh, P. Gupta, and M. Shukla, "Activity detection and counting people using Mask-RCNN with bidirectional ConvLSTM," *J. Intell. Fuzzy Syst.*, 2022, pp. 6505–6520.
- [33] U. Singh and L. S. Songare, "Analysis and detection of monkeypox using the GoogLeNet model," in *Proc. Int. Conf. Autom. Comput. Renew. Syst. (ICACRS)*, Pudukkottai, India, 2022, pp. 1000–1008, doi: 10.1109/ICACRS55517.2022.10029125.
- [34] P. Gupta, M. Shukla, N. Arya, U. Singh, and K. Mishra, "Let the blind see: An AIoT-based device for real-time object recognition with voice conversion," in *Machine Learning for Critical Internet of Medical Things*, Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-030-80928-7_8.

- [35] R. Parihar, A. Jain, and U. Singh, "Support vector machine through detecting packet dropping misbehaving nodes in MANET," in Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), 2017, doi: 10.1109/ICECA.2017.8212711.
- [36] U. Singh, N. K. Solanki, M. K. Varma, and T. Sevak, "A review on big data protection of Hadoop," in Proc. 2nd Int. Conf. Commun. Electron. Syst. (ICES), 2017, doi: 10.1109/CESYS.2017.8321221.