BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

# MACHINE LEARNING FOR SENTIMENT ANALYSIS OF TEXTUAL DATA : *A LITERATURE-BASED CRITICAL REVIEW*

**Divyanshi Singh**

Student, B.E. (Artificial Intelligence and Data Science)
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India
Savitribai Phule Pune University (SPPU)

**Prof. Neelam Jain**

Assistant Professor, Seminar Guide
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India

**Dr. Vishal Puranik**

Assistant Professor, Seminar Guide
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India

**Abstract**

Today, there is an abundance of opinions regarding products, services, institutions, and public events on online platforms that cannot be analyzed by any team of readers. Sentiment analysis solves this problem by classifying a given piece of text as positive, negative, or neutral. Machine learning algorithms are used for creating such classification models. This paper is a theoretical literature review of machine learning applications to sentiment analysis of text. It describes the standard process of text preprocessing (tokenization, stop-word removal, stemming, lemmatization, Bag-of-Words and TF-IDF), and gives the mathematical foundation of four popular text classifiers: Naive Bayes, Logistic Regression, Support Vector Machine and Random Forest. The reviewed studies are contrasted according to the representation of text, the classifier used, and the limitations mentioned. The trend to use deep learning and pre-trained transformers is described. The evidence from social media, e-commerce, and customer support is summarized with its limitations. Four persistent challenges of sentiment analysis are discussed: dependency on context, sarcasm and figurative language, multilingual and code-mixed texts, and noisy unstructured text. No machine learning model was created, and no experiment was conducted; thus, no performance results are given. The review concludes that classical models still hold their ground as cost-effective baselines for large and well-formed corpora, that transformers deal better with context at a higher computational cost, and that there is no solution that works well in all four cases.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

**Keywords:** Sentiment Analysis, Machine Learning, Natural Language Processing, Text Classification, Opinion Mining, TF-IDF, Naive Bayes, Support Vector Machine, Social Media Analytics.

## Introduction

Online writing about one's personal experience is a routine practice. Product reviews, blog entries, forum discussions, support tickets and surveys all comprise continuous written expression of opinion, and any organization would want to understand what the stream is saying about them. Human analysis is not possible for such large amounts of data. Sentiment analysis or opinion mining is the branch of natural language processing that addresses the issue by treating the evaluative content of a text as measurable quantity [1], [2]. At its simplest, the task consists in determining the sentiment expressed in a text as positive, negative or neutral.

The first automatically built systems were based on dictionaries of words and manually written rules. One important milestone was reached when Pang et al. applied standard classification algorithms to manually labelled movie reviews and realized that the classifier could better evaluate previously unseen reviews than human-composed word lists [3]. Supervised machine learning took over since then, and the range of methods used evolved from the linear classifiers over bag-of-words features to neural networks and pretrained transformers [4], [5]. The evolution has complicated things further. Studies vary in their datasets, label sets, and evaluation metrics used, so one cannot objectively compare different statements of this kind [6].

This means that the issue at hand involves an analysis of current machine learning approaches to sentiment analysis of text, and specifically how they classify opinions into polarity classes, how they deal with large amounts of unstructured text, and how they fail when it comes to complex language, slang, sarcasm, and context. The four main challenges that have been highlighted in the seminar synopsis that led to the writing of this paper are: context, sarcasm and figurative language, multilingual and code-switched text, and noisy unstructured data. These will be discussed one by one below.

This paper is theoretical in nature. The paper does not design any systems, collect any data, conduct any experiments and does not provide any new figures of accuracy. The contributions are: (i) a concise formal model of the conventional approach to sentiment analysis; (ii) comparative study of preprocessing methods, feature extraction and classifiers; (iii) a synopsis of literature available on social networking, e-commerce, customer support; (iv) critical analysis of open issues and future research directions; and (v) a list of 34 references.

There are four research questions that lead the investigation. **RQ1:** how does the selection of preprocessors and features affect the performance of machine learning classification algorithms for sentiment analysis? **RQ2:** how do the Naïve Bayes algorithm, Logistic Regression, Support Vector Machine, and Random Forest compare with each other and with Deep Learning and Transformer Models? **RQ3:** how is sentiment analysis used, and what is the strength of evidence in each area? **RQ4:** what are the existing challenges and research gaps?

## Review Approach and Scope

This is a narrative critical review, not a systematic one, and no review protocol was registered. The base of references comprises the eight sources cited in the synopsis (surveys and application studies, from 2022 to 2026), as well as additional sources only when it was required by the analysis: the original

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

machine learning papers on sentiment classification, the description of the algorithms and the term weighting approach analyzed, popular benchmarks, and related topics such as sarcasm, code-mixed text and pre-trained language models. Recent sources are predominantly surveys and therefore their conclusions do not present controlled experiments but rather characterize the state of the art; they constitute secondary literature. The aspect-based approaches, emotion recognition, and multimodal sentiment analysis are out of the scope of the paper. Reference numbers are according to the IEEE style and are assigned in the order of appearance.

## Background and Fundamentals

### Task Definition and Levels of Analysis

Denote by x the text and Y a label set, which is usually {positive, negative, neutral}. The classifier is trained on a labeled dataset D = {(xi, yi)} and predicts a label with the highest estimated probability for the unseen text as follows:

$$\hat{y} = \text{argmax}_{y \in Y} P(y \mid x) \quad (1)$$

The process can be performed on the whole document, on individual sentences or on the aspect level, when the opinion is attributed to a certain characteristic of the product, like the battery of a mobile phone [1]. Approaches are usually classified into lexicon-based, machine learning and hybrid techniques [7], [8]. Lexicon-based techniques aggregate polarity scores of the words in the dictionary and correct them for negations and intensifiers [9], while rule-based approaches, including VADER, are adapted to the informal tone of social media [10]. Machine learning techniques train the mapping from labeled data, and they are the main focus of this paper.

### The Standard Pipeline

A standard machine learning process consists of five steps (Figure 1). The text is collected, cleaned, and tokenized; transformed into numeric features; input into a classifier trained on labeled data; and finally evaluated on held-out data. Each step is associated with a choice which will affect the result, and a bad choice at an early stage, for example dropping a negation word, can never be undone.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

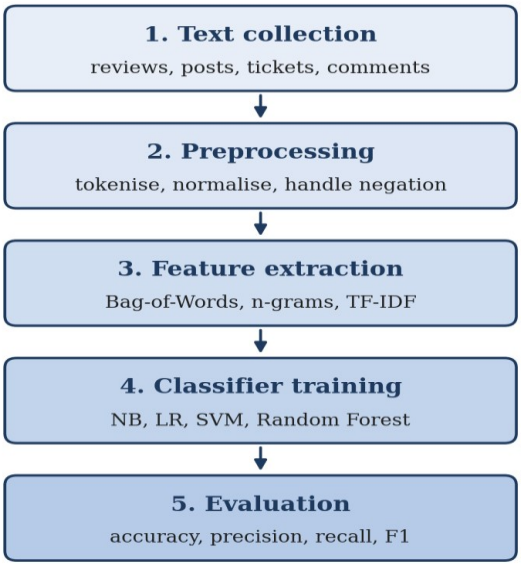



Fig. 1. Conceptual pipeline for machine learning based sentiment analysis. This is an explanatory illustration prepared by the author and does not depict an implemented or tested system.

**Text Preprocessing**

Tokenisation is the process of breaking down the text into words or sub-words. Normalisation entails converting the text into lowercase and sometimes replacing user mentions, links and character repetitions with placeholders. Stop-word removal entails removing function words that occur frequently with the aim of reducing the feature space. However, it should be noted that standard stop-word lists contain words such as "not" and "no" which bear polarity, hence the need for careful inspection of a stop-word list before its implementation. Stemming involves reduction of a word to its root through the removal of suffixes using algorithms such as Porter's [11], while lemmatisation involves mapping a word to its lemma. Two things should be borne in mind in relation to sentiment: emoticons/emojis may be useful polarity indicators and hence should not be eliminated blindly, and negations have to be taken care of explicitly.

**Feature Extraction**

The classical classifiers require fixed length vectors of numbers as input data. With the Bag-of-Words representation, all documents have vectors based on the vocabulary that contain frequencies of terms or just presence of a particular term, as was found out by Pang et al. [3] with respect to movie reviews when presence of a term performed equally well as frequency. The n-gram approach extends the vocabulary with combinations of several terms in order to account for the simple negation like "not good" and it was used successfully in classifying sentiments [12]. The TF-IDF measure accounts for [13]:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \log( N / \text{df}(t) ) \quad (2)$$

In the formula above, tf(t, d) represents the frequency of occurrence of word t in document d; N is the total number of documents and df(t) denotes the number of documents in which t occurs. Words occurring in most documents have very small weights. All the three representations are sparse and ignore

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

word order except within the n-gram context window. Word embeddings like word2vec map words to a dense space, where similar words appear close by [14].

**Classification Algorithms**

*The Naive Bayes* model works based on Bayes rule while assuming that the words in a text become independent of each other after conditioning the class as follows [15]:

$$P(y \mid x) \propto P(y) \prod_i P(w_i \mid y) \quad (3)$$

Using the multinomial formulation of naive bayes, the word probabilities get estimated based on counts during the training phase using additive smoothing:

$$P(w \mid y) = ( n_{w,y} + \alpha ) / ( n_y + \alpha|V| ) \quad (4)$$

Where $n_{w,y}$ refers to the occurrence count of word w in class y, $n_y$ represents the occurrence count of all the words in that particular class. Also, V stands for vocabulary and $\alpha$ corresponds to the additive smoothing constant. The training process is inexpensive hence its popularity as a benchmark technique.

A logistic regression approach is taken such that the probability of positive class gets determined based on the sigmoid of a linear function:

$$P(y = 1 \mid x) = 1 / ( 1 + \exp( -(w^T x + b) ) ) \quad (5)$$

The linear parameters, w, get determined through minimization of cross-entropy loss in most cases.

*A Support Vector Machine (SVM)* attempts to find the separating hyperplane with maximum margin [16]. In its soft-margin formulation, it solves

$$\min \tfrac{1}{2}\|w\|^2 + C \sum_i \xi_i \quad \text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \ \xi_i \geq 0 \quad (6)$$

where C controls the penalty for margin violations and $\xi_i$ are slack variables. SVMs are suited for handling the very large numbers of sparse features found in text [17].

*Random Forest* builds many decision trees, each built on a bootstrap sample of the training set and a random subset of features considered at each split, and makes predictions using majority voting [18]:

$$\hat{y} = \text{mode}\{ h_1(x), h_2(x), ..., h_T(x) \} \quad (7)$$

where $h_t$ is the t-th tree and T is the number of trees.

**Literature Review and Related Work**

The reviewed studies fall into four groups, discussed in turn below, and Table 1 lists them in order of publication.

**Foundations of Machine Learning for Sentiment**

The Pang and Lee survey defined much of the terminology that has been used since, including document-level classification, subjectivity, and the utility of opinion-based applications [2]. In the experiments conducted by Pang et al., Naive Bayes, maximum entropy, and SVM classifiers were compared on movie reviews. The SVM classifier performed best, achieving about 83% accuracy, while it was observed that

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

sentiment classification is more difficult than topic classification because opinions are often expressed in an indirect way [3]. Wang and Manning found that the ranking of simple models is dependent on the dataset, where Naive Bayes worked well on short texts, SVMs on long texts, and NBSVM worked well on all datasets [12].

**Recent Surveys and Application Studies**

According to Kavitha et al., there is an example of the use of NLP preprocessing techniques and machine learning in sentiment analysis of social media texts, but context-dependent and informal language reduces reliability [19]. Palanivinayagam et al. analyzed 224 papers on text classification using machine learning techniques and found the most popular models, datasets and metrics for this task, but since each study differs in all three, comparisons become difficult [6]. Khan et al. use SVM and Random Forest classifiers for sentiment after preprocessing and feature extraction, but traditional machine learning models may have difficulties with processing complex context-dependent language [20]. Suryawanshi discusses the evolution of models from traditional learners to recurrent neural networks and then to transformers and notices that the latter need more computational resources and better training data [21]. Mao et al. generalize to large language models and pre-trained models and identify open issues in datasets, preprocessing, models and evaluation [22]. Methods like SVM, decision trees, and word embeddings in social media and e-commerce are discussed by Nawrocka et al., where noisy data and domain-specific language persist as major issues [23]. In general terms, Muaz et al. describe the existing surveys and identify the major trends, applications, challenges, and research areas [24]. The Twitter-based classification from 2026 reveals the practical side of things: Classification was done with Naive Bayes using Bag-of-Words and TF-IDF features successfully; however, slang, abbreviations, and misspellings affected accuracy [25].

**Deep Learning and Transformers**

The use of neural models changed the paradigm from feature engineering to representation learning. Applying convolutional networks to word embeddings yielded good performance at the sentence level, without any further tuning [26], while LSTM networks handle word ordering through gating recurrent memory [27]. Stanford Sentiment Treebank and its accompanying recursive neural network allowed for the investigation of phenomena like negation at the phrase level [28]. The transformer architecture later superseded recurrence with self-attention [29], and pre-trained models like BERT allow fine-tuning for sentiment analysis using a small amount of labeled data [5]. Reviews on the same line of research concur that the progress in contextual modeling comes with increased computational and data demands [4], [21].

**Evolution of the Approaches**

Figure 2 maps out the major methods within a timeline framework. In fact, each new step in the development was built upon the previous one: lexical approaches are applicable in absence of labelled data sets, while classical methods prove highly competitive on big well-constructed corpus samples.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

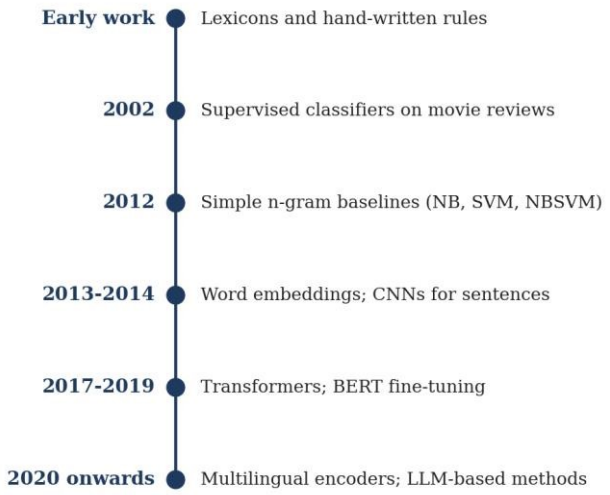



Fig. 2. Conceptual evolution of sentiment analysis approaches. This is an explanatory illustration prepared by the author.

**Table 1:** Literature matrix of the reviewed studies (oldest to newest)

| Study | Method / focus | Finding (as reported) | Limitation |
|---|---|---|---|
| Pang et al. (2002) [3] | NB, maximum entropy and SVM on movie reviews | Learned classifiers beat human-chosen word lists; SVM best; presence at least as good as frequency | Single domain; modest corpus |
| Wang and Manning (2012) [12] | NB, SVM and NBSVM with unigram and bigram features | NB stronger on short text, SVM on long text; bigrams help sentiment | Linear models; limited context |
| Devlin et al. (2019) [5] | Pre-trained bidirectional transformer, fine-tuned | Pre-training plus fine-tuning gives contextual representations for text classification | High compute; needs fine-tuning labels |
| Kavitha et al. (2022) [19] | NLP preprocessing and ML on social media text | Shows how NLP and ML extract sentiment from social media | Informal, context-dependent language lowers reliability |
| Palanivinayagam et al. (2023) [6] | Systematic review of 224 ML text-classification papers | Maps common models, datasets and evaluation metrics | Different datasets and metrics block direct comparison |

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

| Study | Method / focus | Finding (as reported) | Limitation |
|---|---|---|---|
| Khan et al. (2024) [20] | SVM and Random Forest with preprocessing and feature extraction | Traditional learners can classify sentiment from text | May struggle with complex context |
| Suryawanshi (2024) [21] | Survey: traditional ML, RNNs, transformers | Traces the progression and compares applications at different text levels | Advanced models need more compute and good data |
| Mao et al. (2024) [22] | Review of ML, deep learning, LLMs and pre-trained models | Charts the evolution; names challenges in data, preprocessing, models, evaluation | Context, multilingual data and LLMs remain open |
| Nawrocka et al. (2025) [23] | Review of SVM, decision trees, embeddings | Examines uses in social media and e-commerce | Noisy data; language differs by platform |
| Muaz et al. (2026) [24] | Survey of surveys on ML-based sentiment analysis | Maps trends, applications, challenges, opportunities | Broad; methods not tested experimentally |
| Twitter study (2026) [25] | Naive Bayes with BoW and TF-IDF | Effective on large-scale tweet classification | Slang, abbreviations, misspellings reduce accuracy |

## Comparative Analysis

### Preprocessing and Feature Choices

The comparison between the methods discussed in Section II is provided in Table 2. As far as sentiment is concerned, there is no such thing as neutrality because preprocessing operations that are good for topic classification can actually destroy the terms which are used to detect sentiment [4], [7]. Based on the analysis of the literature on the subject matter, several rules can be formulated. It seems reasonable to choose presence-based or TF-IDF features for unigrams and bigrams as an effective way to start [3], [12]. Preprocessing needs to depend on the platform since social media are less formal than the reviews [19], [23].

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

**Table 2:** Comparison of preprocessing and feature techniques

| Technique | What it does | Benefit | Risk for sentiment |
|---|---|---|---|
| Tokenisation | Splits text into units | Required first step | Naive rules split emoticons, hashtags and contractions badly |
| Stop-word removal | Drops very frequent words | Smaller feature space | Lists may contain negations that flip polarity |
| Stemming [11] | Strips suffixes by rules | Merges word variants cheaply | Over-stemming joins unrelated words |
| Lemmatisation | Maps to dictionary form | More accurate than stemming | Slower; needs language resources |
| Bag-of-Words | Counts or presence of words | Simple and effective [3] | Word order and context lost |
| N-grams | Adds word sequences | Captures short negation [12] | Feature count grows quickly |
| TF-IDF [13] | Down-weights common terms | Highlights discriminative words | Sparse; carries no word meaning |
| Embeddings [14] | Dense word vectors | Related words lie close | One vector per word, regardless of context |

**Classifiers**

Comparison of the four classifiers is given in Table 3. Sources that were analyzed in this paper did not declare any winners among them. SVMs were the top classifier during the initial comparison [3] and work well with sparse texts [17], while Naive Bayes classifier was better at analyzing short texts [12] and, being integrated with TF-IDF feature set, can still be applied efficiently to a stream of tweets [25]. According to Khan et al., SVM and Random Forest are complimentary classifiers [20]. Choice of the best classifier relies on the text size and the amount of data available, among other factors.

**Table 3:** Comparison of classical classifiers for sentiment

| Algorithm | Basic idea | Strengths | Weaknesses |
|---|---|---|---|
| Naive Bayes [15] | Bayes' rule with word independence | Very fast; works with little data; good on short text [12] | Independence assumption ignores word interactions |
| Logistic Regression | Linear score passed through a sigmoid | Probabilistic output; inspectable weights | Linear boundary; relies on good features |
| SVM [16] | Maximum-margin hyperplane | Strong on high-dimensional sparse text [17]; best of three in [3] | Sensitive to $C$ and kernel; no direct probabilities |

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

| Algorithm | Basic idea | Strengths | Weaknesses |
|---|---|---|---|
| Random Forest [18] | Many trees, majority vote | Captures non-linear patterns; resists overfitting | Less efficient on very sparse, high-dimensional text; larger models |

### Classical and Transformer-Based Models

Table 4 provides a qualitative assessment of the differences. Transformer architectures manage the problem of context and word ordering more effectively but require greater amounts of computational resources and labeled data to fine-tune [21], [22]. It is important to note that this is qualitative analysis, not quantitative. A proper quantitative comparison requires similar settings across reviewed experiments, something that is rarely found [6].

**Table 4:** Classical machine learning versus transformer models

| Aspect | Classical ML | Transformer-based |
|---|---|---|
| Representation | Sparse counts or TF-IDF; word order mostly ignored | Contextual vectors; sensitive to word order |
| Labelled data | Works with modest amounts | Pre-training reduces need, but fine-tuning still needs labels |
| Compute | Low; trains on a CPU | High; GPUs for training, costly inference |
| Interpretability | Feature weights can be inspected directly | Harder to explain |
| Context, negation | Local only, through n-grams | Better, according to the surveys [21], [22] |
| Many languages | Separate resources per language | Multilingual encoders exist [30] |

### Evaluation Practices

Accuracy is the ratio of correct predictions; however, accuracy might give a misleading result for imbalanced datasets because an incorrect prediction that always classifies the most common class can still yield a high score. Hence, precision, recall, and F1 are also provided for each class:

$$P = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP}) \tag{8}$$

$$R = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN}) \tag{9}$$

$$F_1 = 2PR / (P + R) \tag{10}$$

where TP, FP and FN are true positives, false positives and false negatives respectively. Taking the average of F1 per class (macro-F1) puts all the classes on an equal footing. Some examples of benchmark data sets include movie reviews [31], Stanford Sentiment Tree Bank [28] and tweets annotated by distant supervision through emoticons [32]. These data sets vary with respect to length, domain as well as the quality of labeling, so that a performance measure in one data set cannot be applied directly to another, and diversity of data sets and measures is the main reason for non-comparability among research studies [6].

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

### Applications

This part gives a summary of where machine learning based sentiment analysis can be applied, and Table 5 gives a compact form of those major applications areas. There is not much evidence in the selected articles about the results of implementation of the proposed systems.

### Social Media Analytics

Posts and comments provide an extremely rich and dynamic source of public opinion. Distant supervision using emoticons enabled the construction of large tweet corpora without labeling [32], while classical approaches remain applicable for tweet classification on a large scale [25]. The reviews also highlight the fact that language varies across platforms and evolves quickly [19], [23].

### E-Commerce and Customer Experience

F-The likes and dislikes of consumers for a specific product or service are captured via their reviews and rating which helps businesses understand recurring problems with what they are offering and to improve it [1], [23]. Reviews have more detail than the content of any social media post since there is more thought and time involved while writing [1].

### Conversational AI and Customer Support

Frustration or dissatisfaction signals can be linked to chatbots as well as any support tool, so they can detect problems and prioritize cases immediately. There is no proof from scientific research of the benefits of using the tool since it was only mentioned as an innovative trend that should be used in the future. The small amount of information included in chats and reliance on previous dialogue make context management crucially important.

### Public Opinion, Brand and Market Monitoring

Organisations are guided by opinions on brands, policies, and events by means of the news, forums, and social media [1], [8]. The people who post their opinions on social media are not a random selection of the entire population; therefore, conclusions drawn from these results apply only to the selected sample.

**Table 5:** Summary of application settings

| Setting | Typical text | Typical use | Caveat |
|---|---|---|---|
| Social media | Tweets, comments | Tracking public opinion and trends [25], [32] | Slang, noise, platform drift |
| E-commerce | Product reviews | Satisfaction and recurring problems [23] | Domain-specific vocabulary |
| Customer support | Chat messages, tickets | Detecting frustration; prioritising cases | Little controlled evidence in retained sources |
| Brand and market monitoring | News, forums | Reputation and opinion tracking [1] | Sampling bias |

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

### Challenges and Research Gaps

Below we explore the four issues presented in the synopsis above along with two additional issues, evaluation, and resource cost, which were discovered while reviewing the literature. In each case, the discussion highlights what is known, what is partially solved, and what is still an open issue (Table 6).

### Context Understanding

One same word may have a different meaning and sentiment in different contexts – "unpredictable" is considered positive for thriller plots, but negative for the car's steering ability. Therefore, the classifiers are domain-specific [1], [2]. The Bag-of-Words models neglect word ordering and n-gram models only account for local negation [12]. On the other hand, contextual models, like BERT, manage most of this problem by embedding every word in its sentence [5]; however, context understanding is still listed among unsolved issues, especially in multidomain scenarios and long documents [20], [22].

### Sarcasm and Figurative Language

Sarcasm and irony contradict the actual meaning of the phrase, and hence, the literal opposite of the words used becomes misleading to the classifier. It is noted in a study on automatic sarcasm detection that most of the techniques rely on a difference between sentiment and context, and also require external clues beyond the sentence [33]. Usage of emojis, slang, and hyperbole makes things even more ambiguous [10]. What is lacking is reliable detection using contextual knowledge and labelled data because it is subjective to decide whether the text is sarcastic or not.

### Multilingual and Code-Mixed Text

Most sentiment resources and benchmark datasets are in English. Users often combine multiple languages into one sentence, such as Hindi combined with English in roman script, which causes problems with vocabulary-based features and standard tokenizers. Code-mixed SemEval-2020 dataset was created for the Hindi-English and Spanish-English tweet classification challenge to motivate research in this area [34], and multilingual encoders like XLM-R generalize across one hundred languages [30]. Nonetheless, multilingual datasets are still cited as one of the open challenges by Mao et al. [22]. The problem is relevant to multilingual countries like India, where code-mixing of languages is commonplace, and datasets in many language pairs and variations of transliteration spelling are required.

### Noisy and Unstructured Data

The social media text has misspellings, acronyms, hashtags, emoticons, elongated words, and sentence fragments. The rule-based approaches are equipped to handle some of these issues [10]. Informal language use in social media data has been shown to reduce the performance of classical models [19], [23], [25]. The normalisation process is useful, but it removes important information like letter repetitions or capitalization which signify strength of feeling.

### Evaluation and Reproducibility

Scores cannot easily be compared between different works due to using various corpora, splitting and metrics [6]. A common set of benchmarks and reporting of variability across several runs along with macro-averaged metrics could solve this problem.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

### Computational Costs and Data Requirements

The best context-sensitive approaches are computationally costly and require high-quality training data [21]. This becomes an issue in practice when applying context-sensitive models in small organizations and in low-resource languages.

| Gap | Established or partly solved | What remains open | Direction |
|---|---|---|---|
| Context | Local negation via n-grams [12]; contextual encoders [5] | Domain shift; long texts [22] | Domain adaptation; efficient fine-tuning |
| Sarcasm | Detection approaches surveyed [33] | Needs context beyond the sentence | Conversation-level and world-knowledge cues |
| Multilingual, code-mixed | Shared task [34]; multilingual encoders [30] | Few labelled datasets; transliteration | New datasets; robust handling of mixed scripts |
| Noisy text | Rule-based tools [10]; normalisation | Spelling variation; emojis; loss of intensity cues | Noise-robust preprocessing |
| Evaluation | Many benchmarks [28], [31], [32] | No common protocol [6] | Shared benchmarks and reporting |
| Resource cost | Surveys of costs [21] | Affordable context-aware models | Lighter models |

**Table 6:** Research gaps, existing progress and open problems

### Future Directions

These shortcomings provide pointers in a few possible future research directions. First, one is making context-aware systems more cost-effective, in order to enable smaller organizations achieve context awareness comparable to transformer architectures at reduced computational expenses [21]. The second direction is domain adaptation, which will decrease the need to relabel datasets in case of novel categories of products or other platforms. The third is better processing of sarcasm taking advantage of conversation-level context [33], and the fourth is creation of labeled resources for code-mixed and under-resourced languages [34]. Tolerant pre-processing able to preserve intensity markers [5], common benchmarking framework and report standards [6], as well as systems hybridizing sentiment lexicons and learned methods [7], [9] are the last two items of the list.

The issue of responsible use cannot be overlooked either. One can assume that training datasets reflect certain societal biases of the communities of origin, and processing personally identifiable posts gives raise to privacy concerns. Such questions are outside the scope of the retained references but must be considered in practice nevertheless.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

## Conclusion

This paper has reviewed machine learning approaches for sentiment analysis of texts, including the text processing pipeline, the formalization of four common classifiers, the evidence from recent surveys and applications, and open issues. Four conclusions may be drawn from this paper. First, preprocessing and feature selection are as important as the choice of classifier, and measures which are useful for topic classification may harm sentiment. Second, there is no classical algorithm which works everywhere; it depends on the length of the text, the number of documents and features, so simple algorithms such as Naive Bayes and SVM remain relevant. Third, transformer-based algorithms work well with context, but they require more computational resources and data. Fourth, context, sarcasm, multilingual and code-mixed text, and noisy data are still partially solved, and no universal solution which covers all these issues exists.

There are limitations to the present review. It is based on the literature, provides performance numbers as stated by the original authors, and several of the recent sources are surveys rather than experiments. One of the next steps to be done could be the empirical comparison of the discussed algorithms on one dataset, which includes code-mixed text.

## References

[1] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1-135, 2008.

[3] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2002, pp. 79-86.

[4] L. Zhang, S. Wang and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining Knowl. Discov.*, vol. 8, no. 4, art. e1253, 2018.

[5] 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.

[6] A. Palanivinayagam, C. Z. El-Bayeh and R. Damaševičius, "Twenty years of machine-learning-based text classification: A systematic review," *Algorithms*, vol. 16, no. 5, art. 236, 2023.

[7] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093-1113, 2014.

[8] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14-46, 2015.

[9] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267-307, 2011.

[10] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Weblogs Social Media (ICWSM)*, 2014, pp. 216-225.

[11] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130-137, 1980.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[12] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguist. (ACL), Short Papers*, 2012, pp. 90-94.

[13] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513-523, 1988.

[14] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space," arXiv:1301.3781, 2013.

[15] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in *Proc. AAAI-98 Workshop Learn. Text Categorization*, 1998, pp. 41-48.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, 1995.

[17] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. Eur. Conf. Mach. Learn. (ECML)*, 1998, pp. 137-142.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, 2001.

[19] M. Kavitha, B. B. Naib, B. Mallikarjuna, R. Kavitha and R. Srinivasan, "Sentiment analysis using NLP and machine learning techniques on social media data," 2022.

[20] T. A. Khan et al., "Sentiment analysis using support vector machine and random forest," 2024.

[21] N. S. Suryawanshi, "Sentiment analysis with machine learning and deep learning: A survey of techniques and applications," 2024.

[22] A. Mao, C. Liu and Y. Zhang, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," 2024.

[23] A. Nawrocka et al., "The application of machine learning in sentiment analysis of social media and e-commerce: A review of methods and techniques," 2025.

[24] S. A. Muaz, A. Y. Gital and U. Ali, "A survey of surveys on sentiment analysis using machine learning," 2026.

[25] "Twitter sentiment analysis using machine learning techniques," 2026.

[26] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, 2014, pp. 1746-1751.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[28] R. Socher et al., "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. EMNLP*, 2013, pp. 1631-1642.

[29] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[30] A. Conneau et al., "Unsupervised cross-lingual representation learning at scale," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguist.*, 2020, pp. 8440-8451.

[31] A. L. Maas et al., "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguist.: Human Lang. Technol.*, 2011, pp. 142-150.

[32] A. Go, R. Bhayani and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Rep., Stanford Univ., 2009.

[33] A. Joshi, P. Bhattacharyya and M. J. Carman, "Automatic sarcasm detection: A survey," *ACM Comput. Surv.*, vol. 50, no. 5, art. 73, 2017.

[34] P. Patwa et al., "SemEval-2020 Task 9: Overview of sentiment analysis of code-mixed tweets," in *Proc. 14th Workshop Semantic Eval. (SemEval)*, 2020, pp. 774-790.