

RESEARCH

Open Access



Analysis of customer reviews with an improved VADER lexicon classifier

Kousik Barik¹ and Sanjay Misra^{2,3*}

*Correspondence:
sanjay.misra@ife.no; sanjay.misra@hiof.no

¹ Department of Computer Science, University of Alcala, Madrid, Spain

² Department of Applied Data Science, Institute for Energy Technology, Halden, Norway

³ Department of Computer Science and Communication, Østfold University College, Halden, Norway

Abstract

Background: The importance of customer reviews in determining satisfaction has significantly increased in the digital marketplace. Using sentiment analysis in customer reviews has immense potential but encounters challenges owing to domain heterogeneity. The sentiment orientation of words varies by domain; however, comprehending domain-specific sentiment reviews remains a significant constraint.

Aim: This study proposes an Improved VADER (IVADER) lexicon-based classification model to evaluate customer sentiment in multiple domains. The model involves constructing a domain-specific dictionary based on the VADER lexicon and classifying doevIEWS using the constructed dictionary.

Methodology: The proposed IVADER model uses data preprocessing, Vectorizer transformation, WordnetLemmatizer-based feature selection, and enhanced VADER Lexicon classifier.

Result: Compared to existing studies, the IVADER model accomplished outcomes of accuracy of 98.64%, precision of 97%, recall of 94%, f1-measure of 92%, and less training time of 44 s for classification.

Outcome: Product designers and business organizations can benefit from the IVADER model to evaluate multi-domain customer sentiment and introduce new products in the competitive online marketplace.

Keywords: Multi-domain sentiment analysis, Improved VADER (IVADER), Customer reviews, Lexicon-based dictionary

Introduction

Extracting ideas and attributes from web-based customer reviews is increasingly relevant in the digital marketplace [1]. While purchasing a product, the customer enquires about the item's popularity, features, usefulness, and value addition. Online review attracts new prospective customers by furnishing facts to support purchasing decisions, and subsequently, organizations benefit from future product enhancement. Studies have demonstrated that more than half of customers across various industries, including computer hardware, sports and fitness, and tourism, search for product reviews and other relevant information before purchasing [2]. The sentiment data can be more efficient and practical due to the massive review data and the need for

online processing [3]. Opinion mining is popular in social media analysis, web mining, and data mining. With the evolution of Natural Language Processing (NLP), text analysis, and applied linguistics, the sentiments articulated in customer reviews are decoded, evaluated, and classified [4].

Researchers have desired to apprehend customer sentiments accurately. Rule-based, machine learning (ML), and combination techniques are the three basic methodologies used in sentiment analysis [5]. The rule-based method incorporates the dictionary-based approach. The ML category includes conventional methods like conditional random fields, and the methods for Deep Learning (DL) are covered. Several studies have effectively included traditional ML approaches with DL methods in text sentiment analysis by developing the sentiment dictionary in recent years [6]. The quantity and quality of labeled data are crucial to ML effectiveness. The more complex the neural network, the more data is required. Furthermore, the quality of word embedding influences sentiment classification results as it assesses the connection between word vectors in vector space [7].

In low data quality, sentiment lexicons are an essential tool in most sentiment analysis techniques compared to ML approaches. There has been a significant interest in reviewing lexicon-based sentiment analysis methods. There are several approaches to creating a sentiment dictionary. Emotional vocabulary includes words or phrases that express a specific emotion and its strength and polarity. Dictionary-based and corpus-based techniques are examples of lexicon-based methods. The words are collected manually and annotated in the classic unigram emotion lexicon. SentiWordNet, VADER, TextBlob, and others are widely used as dictionaries. Although these approaches provide a basis for building a sentiment lexicon, the lexicon contains few sentiment terms, and coverage of these words is limited. The researchers recommended expanding the standard sentiment lexicons to include more terms expressing different emotions. Instead, expanded n-gram lexicons provide the polarity of n-grams without quantitatively describing the emotional potency of these terms. Furthermore, these systems can not develop new emotion words due to the limitations of the core lexicon [8].

Existing studies have demonstrated that distinct domains have unique sentiment words that often play a significant role in sentiment analysis. Because a supervised ML classifier trained in one domain can not learn invisible emotion words, it may perform inadequately when tested with data from another domain. VADER is a lexicon-based sentiment analysis (SA) for many text types [9]. Compared to TextBlob, VADER is a widely used and faster lexicon-based paradigm [10]. Conventional lexicons are less accurate because they are based on emotional ratings of words [11]. Because customer reviews come from various sources, analyzing them using the same vocabulary across different industries is not viable. To address this issue, domain-specific sentiment dictionaries that rely on properties of the target domain should be developed. However, sentiment analysis of customer reviews must adequately address the construction of a domain-specific VADER lexicon. This work pinnacles the design of an INVADER model using a VADER lexical dictionary in multiple domains. The significant contributions to the proposed work are as follows:

1. The preprocessing stage is conducted to extract missing or incompatible information of significance that may arise from errors driven by humans or computers and to ensure data consistency.
2. The problem of domain heterogeneity in sentiment analysis is addressed by developing domain-specific dictionaries. It allows precise sentiment classification by considering the different sentiment directions of words.
3. The proposed IVADER model is evaluated in the multi-domain dataset. The model is also validated by changing data samples. The IVDER model is further validated with a different dataset and compared to existing studies.
4. The proposed model yielded less training time in classifying sentiment with reduced loss and error rate, symbolizing stability.

The remaining paper is formulated as follows. "[Related works](#)" section concerns the related works. "[Proposed work](#)" section illustrates the proposed work, dataset, and methodology. The performance evaluation is illustrated in "[Results](#)" section. The discussion, threats to validity, study limitations, and future studies are exhibited in "[Discussion](#)" section. Finally, the paper is concluded in "[Conclusion](#)" section.

Related works

ML and DL approaches to sentiment analysis

Korovkinos et al. [12] presented a study on sentiment analysis that used SVM classification with a smaller training set. SVM takes less time to run but must fine-tune its parameters to improve sentiment analysis accuracy. Zhou et al. [13] applied text sentiment analysis and fuzzy mathematics to assess online users' repurchase intention. In an emotionally calculated model, customer satisfaction, trust, and marketing efforts are harbingers and motivators for recurring purchases. Sun et al. [14] suggested an approach to analyzing electronic word-of-mouth (eWOM) items based on fine-grained sentiment analysis from online customer evaluations. They demonstrate a context-aware, feature-based sentiment analysis system that can use massive user evaluations on social media platforms. They used a semi-supervised fuzzy product ontology mining technique to glean semantic information from positive or negative online customer evaluations. Alharbi et al. [15] proposed a method that leverages CNN and user behavior information in customer reviews. However, the quality of the training data has an impact on the CNN model. Wang et al. [16] proposed a model based on a Multi-head self-awareness-base based sentence-to-sentence attention network1 (S2SAN) for sentiment analysis across multiple domains. However, the accuracy of their sentiment analysis does not reach a high level. Behera et al. [17] presented a co-LSTM technique for sentiment categorization reviews in several fields using two DL structures, i.e., CNN and LSTM. Their proposed convolutional layer model may not capture the sequential dependency of the words, which may lead to a reduction in sentiment classification accuracy.

Sentiment analysis identifies the emotional recommendation of a string of language and is fundamentally used to comprehend customers' attitudes, ideas, and feelings. Priyadarshini et al. [18] suggested a novel deep neural network model based on grid search and CNN with Long Short-Term Memory (CNN-LSTM) for sentiment analysis.

Styawati et al. [19] proposed a sentiment analysis model using the SVM technique and the word2vec text embedding. Word2vec is used as a feature extraction approach to represent words as vectors. The word2vec employed has a skip-gram model. Kayıkcı et al. [20] employed the SenDemonNet to show how the population deemed the newly imposed demonetization regulation. The primary goal is to select weighted features using the hybrid Forest-Whale Optimisation Algorithm (F-WOA) for the best classification results. With the help of these characteristics, the Heuristic Deep Neural Network (HDNN) is used for classification, and the proposed FOA and WOA are used to tweak the DNN's parameters for the highest accuracy rate. Nasfi et al. [21] used a hybrid generative-discriminative technique that combined Fisher kernels with generalized inverted Dirichlet-based Hidden Markov Models (HMM) to enhance recognition performance in textual analysis. They provide a technique that combines SVM's discriminative approach with generative HMMs.

Sagarino et al. [22] presented a study on sentiment analysis in product evaluations using Shopee data. They preprocessed the data, and the VADER algorithm is used to annotate it. Multinomial Naive Bayes (MNB) and SVM are used to analyze the consumers' sentiments. Benarafa et al. [23] developed a method to enhance K-Nearest Neighbours (KNN) to address the Implicit Aspect Identification problem (IAI). They employed improving KNN distance computation using WordNet semantic relations to help the IAI challenge. Jain et al. [24] suggested a Bidirectional Encoder Representations from Transformers (BERT) based Dilated Convolutional Neural Network (BERT-DCNN) framework. They used BERT as a pre-trained language model to build word embeddings. Three consecutive layers of a DCNN layered with a global average pooling layer aid in fine-tuning the model. The proposed BERT-DCNN model accomplishes dimensionality reduction and incorporates an expansion of associated dimensions while preventing information degradation.

Sentiment analysis using lexicon-based methods

Li et al. [25] presented a study on imbalanced text sentiment classification, but the performance could be much better on imbalanced data. Xing et al. [26] presented a strategy to train a sentiment classifier using existing sentiment lexicons, including Opinion Lexicon, SentiWordNet, the dictionary of Loughran & McDonald, and SenticNet. This technique did not consider information about overlap and conflicts between lexicons and achieved an accuracy of 77.9%. Deng et al. [27] studied a unique hierarchical supervision methodology for higher-level categorization to build a topic-adaptive sentiment lexicon (TaS). However, an unbalanced dataset poses a challenge to TaSL performance. Dey et al. [28] proposed a study using Cross-D Vectorize as a collection of three functional areas for cross-domain sentiment categorization. They extracted sentiment unigrams, intensifiers, and negators from the dataset using the VADER algorithm, but domain-specific features were not considered.

Frangidis et al. [29] presented a study on movie screenplay reviews to determine whether the reviewer's emotional response to a movie can predict the movie's rating reliably. They did not consider Word2vec and N-grams. VADER and NRC are used for sentiment analysis. Wook et al. [30] employed the Lexicon-based Vader as a dictionary to establish the division of terms in student responses. However, the feature selection is

considered, and a high negative value of 79.4% is achieved. Moussa et al. [31] proposed a framework using the VADER. They did not consider opinion mining and cross-domain datasets. Lee et al. [32] presented a study of lexicon-based review sentiment analysis methods to determine if the review systems are based on actual customer experiences, satisfaction, and opinions. Sharma et al. [33] proposed a domain-specific lexicon-based sentiment analysis using the SentiDraw framework to determine polarity. However, the overall accuracy of the model of different datasets could be higher. Beigi et al. [34] proposed a domain-specific sentiment lexicon study for unsupervised domains using the Multilayer Perceptron (MLP) technique. They used different vocabularies are used in each domain.

Hasanati et al. [35] presented a quantitative analysis of fine-grained sentiment analysis. They used review data from Twitter related to the issue of the COVID-19 vaccination and utilized the SVM algorithm. Text preparation is accomplished during the modification step to organize the dataset. They assigned sentiment classifications to the data set using a lexicon-based approach. Juanita et al. [36] used a lexicon-based approach to apprehend textual information in a collection of user sentiments. They used Naive Bayes and the Lexicon-Based Model. The study benefits from the optimal model for sentiment analysis on e-marketplaces. Tahayna et al. [37] presented an augmentation procedure to enhance categorizing idiomatic tweets with small training sets. They assess the performance of an embedding model that has been fine-tuned for classification. Thangavel et al. [38] proposed a lexicon-based method to apply enhanced algorithms for sentiment analysis using tweet data. A lexicon-based technique and framework have been presented for multimodal sentiment analysis of text compiled from audio, pictures, and videos. Ojeda-Hernández et al. [39] presented a study using the Formal Concept Analysis (FCA) technique for sentiment analysis to build classification dictionaries. This technique enables the generation of bespoke dictionaries suited to the particular data and activities, unlike other approaches that depend on pre-defined lexicons. Yue et al. [40] presented a study using a collaborative neural network (CAN) for sentiment classification using multi-domain. They used two types of datasets, i.e., amazon and JD. The first issue is that the accuracy the suggested model may attain is constrained by the precision of the UDA technique chosen in phase 1. Badr et al. [41] proposed an Unsupervised DomainAdaptation with Source Preservation (UDA-SP) model for sentiment analysis. It is achieved by understanding expressions of shared and distinctive features comprehended from different networks. Geethapriya et al. [42] presented a method using a spectral clustering technique to map domain-specific word sentiment classification. The features are additionally filtered by incorporating synonyms and replacing negative polarity phrases with suitable antonyms. Due to the high number of characteristics produced, implicit context-specific feature selection needs to be addressed The summary of existing studies is presented in Table 1.

Based on the existing studies, the efficiency of traditional VADER encyclopedias could be more accurate since most of them are based on sentiment values. Additionally, using the lexicon to analyze customer reviews from different domains is challenging. A domain-specific sentiment lexicon can be constructed based on the target domain type to address the existing issues of multi-domain sentiment analysis. This motivated us to create a domain-specific VADER lexicon dictionary-based sentiment

Table 1 Summary of existing studies

Sr. No	References	Description	Techniques	Dataset	Limitations
1	Styawati et al. [19]	Sentiment analysis on online transportation service applications reviews of users	Word2vec text embedding model, SVM	Online customer reviews in transportation systems	Grab app gets a lower performance, with 87% accuracy achieved
2	Kayıkçı et al. [20]	To understand the public opinion on the recently implemented demonetization policy using the proposed SenDemonNet	F-WOA, HDNN	Twitter sentiment reviews	This work does not consider sarcasm tweets; it considers fewer amounts of tweets
3	Nasfi et al. [21]	Sentiment analysis on movie reviews and product reviews	Gaussian-based HMM, SVM	Amazon products and IMDb movie reviews	The hybrid model, especially when using SVMs, may make it susceptible to overfitting
4	Sagarino et al. [22]	Sentiment analysis on product reviews as customer recommendations in Shopee Philippines	VADER and MNB	Shopee customer reviews	Generalizability to other e-commerce platforms or languages may be limited, and low accuracy may be achieved
5	Benarafa et al. [23]	Sentiment analysis of electronic products and restaurant reviews	Implicit Aspect Identification (IAI) and KNN	Electronic products and restaurant reviews	There is a risk that it might overfit the WordNet data
6	Sharma et al. [33]	Domain-specific sentiment analysis using product star ratings	SentiDraw framework	Cornell Movie Reviews Dataset and Large Movie Review Dataset	The average accuracy of the model was below 90%
7	Beigi et al. [34]	Domain-specific feature creation and sentiment classification	Domain-Independent Lexicon and Multilayer perception	Amazon multi-domain customer reviews	Each domain utilizes a different range of vocabulary from one another
8	Hasanati et al. [35]	Tweets about the COVID vaccination are subject to fine-grained sentiment analysis	SEMMA and SVM	Twitter customer reviews	Not considered domain-specific words relating to COVID-19 and vaccinations
9	Juanita et al. [36]	Sentiment analysis on E-Marketplace users' opinions	Lexicon-Based Model and Naive Bayes model	Online customer reviews	Overfitting to the training data
10	Thangavel et al. [37]	Lexicon-based procedure to implement tweets sentiment analysis	Lexicon based framework	Twitter customer reviews, STS-Gold Dataset	considers fewer tweets
11	Tahayna et al. [38]	To improve the classification of idiomatic tweets with tiny training samples	Lexicon-based approach and Pliable augmentation technique	Twitter customer reviews	16% error rate to handle larger datasets efficiently
12	Ojeda-Hernández et al. [39]	Sentiment analysis to create dictionaries for classification	Formal Concept Analysis (FCA)	Twitter dataset	FCA can be computationally intensive, especially when dealing with large datasets or complex concept lattices

Table 1 (continued)

Sr. No	References	Description	Techniques	Dataset	Limitations
13	Yue et al. [40]	Multi-domain sentiment analysis based on neural network	Attention Neural Networks	Amazon and JD customer reviews	computing complexity is higher
14	Badr et al.[41]	Labeled sentiment analysis in an unsupervised domain	Unsupervised DomainAdaptation with Source Preservation	Amazon, FDU-MTL, and Spam dataset	Implicit context-specific feature selection can be addressed
15	Geethapriya et al. [42]	An improved method in cross-domain sentiment classification	CDSARFE approach	Amazon multi-domain customer reviews	Not considered implicit feature extraction

analysis and enhance multi-domain performance. This study aims to address the problem of domain heterogeneity in sentiment analysis. The proposed model is also validated by varying the sample in the multi-domain dataset. Furthermore, the model is tested with a different dataset to validate its applicability.

Proposed work

This section demonstrates the workflow of the proposed model, depicted in Fig. 1. The multi-domain customer review dataset is collected and processed to reduce the complexity of the dataset. TF-IDF vectorization transformation is applied to extract domain-specific features. The Wordnet lemmatizer-based feature selection method is used to select more significant domain-specific features based on lemmatization and cosine similarity between features and the synset of the WordNet database. An IVADER lexicon classification method employs domain-specific VADER lexicon-based dictionary construction and classification of reviews into positive and negative categories using the constructed dictionary as a reference. The proposed model is evaluated using performance metrics.

Dataset description

This study uses a publicly available multi-domain sentiment dataset (<https://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html>) for sentiment analysis based on amazon.com. It comprises four domains, i.e., books, DVDs, electronics, and kitchens, and thousands of reviews in each domain. This study randomly used 8,000 reviews in each domain, 1000 positive and 1000 negative reviews. The datasets are downloaded from the site in tar.gz format. The Python environment and the NLP toolkit are used to extract the data. The labeled samples are considered, and unlabeled data are not considered. The description of the dataset is presented in Table 2.

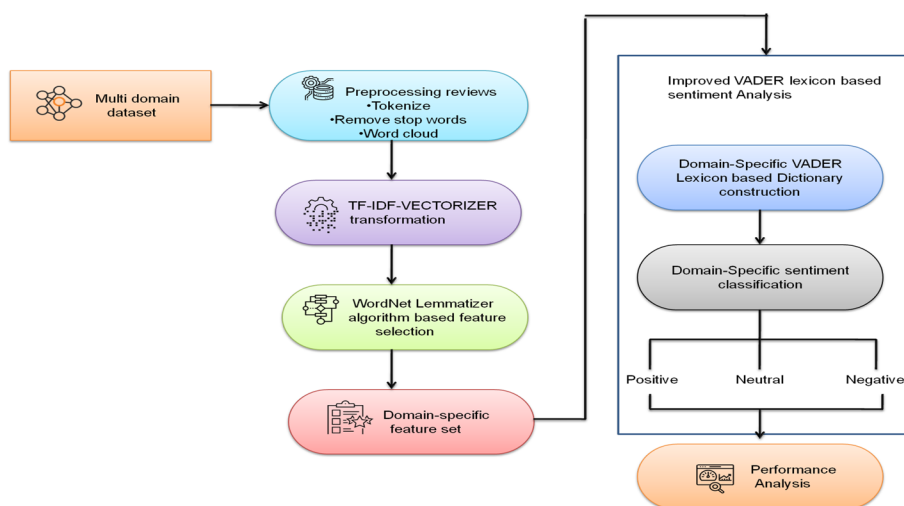


Fig. 1 The workflow of the proposed model

Table 2 Dataset description

Dataset		Number of positive reviews	Number of negative reviews	Total reviews
Multi-Domain Sentiment Dataset	Electronics	1000	1000	2000
	Books	1000	1000	2000
	DVD	1000	1000	2000
	Kitchen	1000	1000	2000

Pre-processing

The collected reviews are unstructured text, preprocessed to remove irrelevant data. HTML tags, URLs, punctuation marks, symbols, numbers, and spaces are stripped from the record because they do not represent sentiments. Removing noise and ambiguities is a critical step in the preprocessing stage. The stemming method reduces the words to their root form [43]. In addition, review preprocessing includes tokenization, i.e., breaking a sentence into separate sentences, and stopping word removal, i.e., removing prepositions, articles, and connectors. A word cloud visualization is performed to analyze the frequency of terms in a domain. Spelling errors, unusual characters, and unrelated words are causes of noise in sentiment analysis. The stop-word removal method eliminates words that are used often, such as adverbs, conjunctions, prepositions, and articles. The density of the datasets is decreased by removing these terms. It also contains the most often-used terms, such as "they," "she," "but," "if," "he," and "we," among others. The stop words are removed. A word cloud for sentiment lexicons is utilized to see the terms often used in the lexicon and how they relate to various sentiments. It is accomplished by generating a grammatical dependency graph that reflects the semantic relationships between keywords in the reviews. The graph is used to create word clouds that are clustered. Each clustered word cloud in a domain has reviews with a closer semantic relationship. The steps performed in the preprocessing stages are demonstrated in Fig. 2.

Term frequency inverse document frequency (TF-IDF) vectorizer transformation

The preprocessed reviews are sent as input to the TF-IDF vectorization model for domain-specific feature extraction. The TF-IDF vectorizer technique extracts characteristics more pertinent to the domain. A domain-specific feature vector is created using the TF-IDF vectorization technique. The TF represents the frequency of a word or phrase appearing in the domain [44]. Because the length of domains varies, it is used for a given term to be repeated more often in more prolonged than shorter domains. The term frequency is divided by the total number of words normalized in the domain using Eq. 1.

$$TF_{w_n} = \frac{g_{w_n}^{d_m}}{T_{d_m}} \tag{1}$$

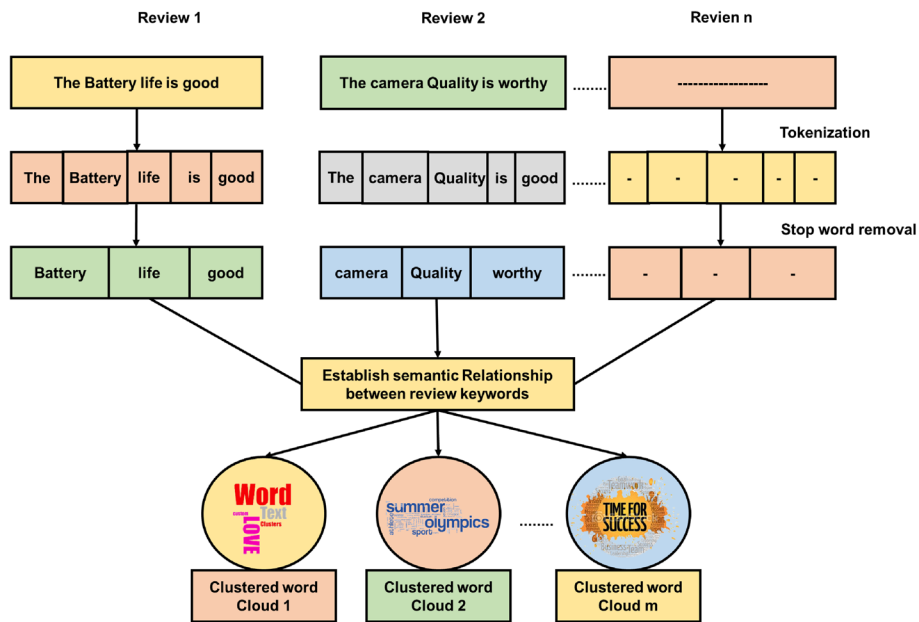


Fig. 2 Steps involved in review preprocessing

where w_n is the n th word in domain (d_m), $m = 1$ to 4 (since four domains are considered), T_{d_m} means the total number of words in domain (d_m), $g_{w_n}^{d_m}$ defines the number of times the word (w_n) occurs in the domain (d_m).

Inverse Document Frequency (IDF) estimates the essential term is to customer review over the whole domain. The IDF of a word (w_n) belonging to domain d_m is estimated using Eq. 2.

$$IDF_{w_n} = \log \left(\frac{T_{d_m}}{N_{w_n}} \right) \tag{2}$$

where N_{w_n} means the number of domains having the word (w_n).

In certain words, TF-IDF is the combination of TF and IDF values. Each document in the TF-IDF [44] is represented as a vector containing TF-IDF values for each word in the file. TF-IDF of the word (w_n) belonging to domain d_m is estimated using Eq. 3.

$$(TF - IDF)_{w_n} = TF_{w_n} \cdot IDF_{w_n} \tag{3}$$

This approach determines the relative importance of characteristics within each domain. The term's relevance to the sentiment analysis domain increases with the term's TF-IDF score. The words with higher TF-IDF scores are identified as paramount domain-specific characteristics for sentiment categorization.

WordnetLemmatizer algorithm-based feature selection

The essential domain-specific features are selected from the domain-specific features extracted with TF-IDF using the Wordnet Lemmatizer algorithm. This algorithm involves two stages, namely, lemmatization and cosine similarity analysis. Certain extracted features have similar meanings but are expressed in different word forms. The lemmatization method is used in the first phase to combine domain-specific characteristics represented in several word forms into a single fundamental form. This initially led to a small reduction in features [45]. Based on the computation of the cosine similarity between elements recovered in this study and the WordNet database, the crucial domain-specific characteristics are selected from this list of features. This study analysis selects features similar to WordNet as critical domain-specific features. Lemmatization is a technique of combining different word forms having the same meaning into a single word. By removing both affixes, lemmatization brings terms back to their roots.

Example: Two features, namely "heating" and "heat," have similar word sense. Hence, the feature "heating" can be lemmatized into the feature "heat."

Domain-specific features expressed in different word forms with similar meanings can be lemmatized into root words. Then, this list is the input for the second stage of the Wordnet Lemmatizer algorithm. WordNet is a database organized into multiple sets of synonyms called synsets. Each Synset is a collection of phrases with similar literal meanings. The generated feature list is compared with Synset for sentiment analysis. The cosine similarity between an extracted domain-specific feature and a synset from WordNet is calculated using Eq. 4.

$$CS_{f_n^{d_m} \rightarrow W} = \frac{f_n^{d_m} \cdot W}{|f_n^{d_m}| \cdot |W|} \quad (4)$$

where CS means the cosine similarity, $f_n^{d_m}$ means the nth feature of the domain (d_m), and W is the synset vector of WordNet.

The domain-specific features of similar Synsets in WordNet are identified and added as significant features for the domain representation. The steps to create a domain-specific feature set are illustrated in Algorithm 1.

Algorithm 1 Domain-specific feature set construction

```

Input: Preprocessed Review Set  $[R^{d_m}]$  where  $m=1$  to 4
Output: Domain-specific Feature Set

Let domains of the dataset be  $d_1, d_2, d_3,$  and  $d_4$ 

for ( $m=1, m \leq 4, m=m+1$ )
{
  Let  $[R^{d_m}]$  be a set of preprocessed words.  $[R^{d_m}] = \{w_1^{d_m}, w_2^{d_m}, \dots, w_x^{d_m}\}$ 
  for ( $b=1, b \leq x, b=b+1$ )
  {
    Find  $g_{w_b}^{d_m}$ , the number of times the word ( $w_b^{d_m}$ ) occurs in the domain ( $d_m$ )
    Compute term frequency ( $TF_{w_b}^{d_m}$ ) and inverse document frequency ( $IDF_{w_b}^{d_m}$ )
    Estimate  $(TF - IDF)_{w_b} = TF_{w_b} \cdot IDF_{w_b}$ 
  }
  Select words with higher (TF-IDF) values from  $[R^{d_m}]$ 
}

for ( $m=1, m \leq 4, m=m+1$ )
{
  Construct an initial domain-specific feature set  $[S^{d_m}]$ 

   $[S^{d_m}] = \text{Lemmatize}([R^{d_m}])$  /* Lemmatize function combines words expressed in different word forms having the same meaning into a single word */

  Let  $[S^{d_m}] = \{f_1^{d_m}, f_2^{d_m}, \dots, f_y^{d_m}\}$ 
  for ( $i = 1, i \leq y, i = i + 1$ )
  {
    Establish the Cosine similarity between  $f_i^{d_m}$  and the synset of WordNet.
  }
  Select features with higher cosine similarity values from  $[S^{d_m}]$ 
  Assign selected features into the final domain-specific feature set  $[[F^{d_m}]]$ 
}

End

```

Sentiment analysis using improved VADER lexicon classifier

The sentiment is represented differently in diverse areas. The classification of domain-specific reviews into positive, negative, and neutral is accomplished in this stage using an IVADER Lexicon classification algorithm. A sentiment lexicon, or polarity lexicon, consists of words with associated values representing their sentiment polarities. Dictionaries are useful for assessing sentiment analysis because they furnish significant information. The conventional VADER lexicons are modeled with only the sentiment scores with different parameters, including sentiment score, valence, and entropy. The IVADER Lexicon classification method includes two main stages. The first stage

is to create a dynamic domain-specific VADER lexicon-based dictionary by computing the overall polarity score of this model, including sentiment score and entropy from domain-specific features. The second level is a classification of reviews based on the sentiment score of reviews, which is determined using a domain-specific VADER lexicon-based dictionary as a reference. Algorithm 2 portrays the pseudocode for an improved domain-specific sentiment analysis depending on the VADER lexicon. The stages of domain-specific sentiment evaluation with IVADER are explained below. Figure 3 pictures the model of the IVADER-based domain-specific sentiment classification.

The significant feature sets of each domain are obtained from the previous step. In each domain, the polarity score of each significant sentiment feature is calculated using Eq. 5.

$$P_{f_n}^{d_m} = \left\{ \begin{array}{l} +1, \text{ if } \text{prob}(\text{pos}_{f_n}) > \text{prob}(\text{neg}_{f_n}) \\ +1, \text{ if } \text{prob}(\text{pos}_{f_n}) < \text{prob}(\text{neg}_{f_n}) \end{array} \right\} \tag{5}$$

where $P_{f_n}^{d_m}$ is the polarity score of an nth feature of domain (d_m), $\text{prob}(\text{pos})$ means the probability of a positive sense of the feature, and $\text{prob}(\text{neg})$ defines the probability of a negative sense of the feature.

In each domain, the entropy of each significant sentiment feature is calculated using Eq. 6.

$$E_{f_n}^{d_m} = \sum_{n=1}^n SS_{f_n} \log(SS_{f_n}) \tag{6}$$

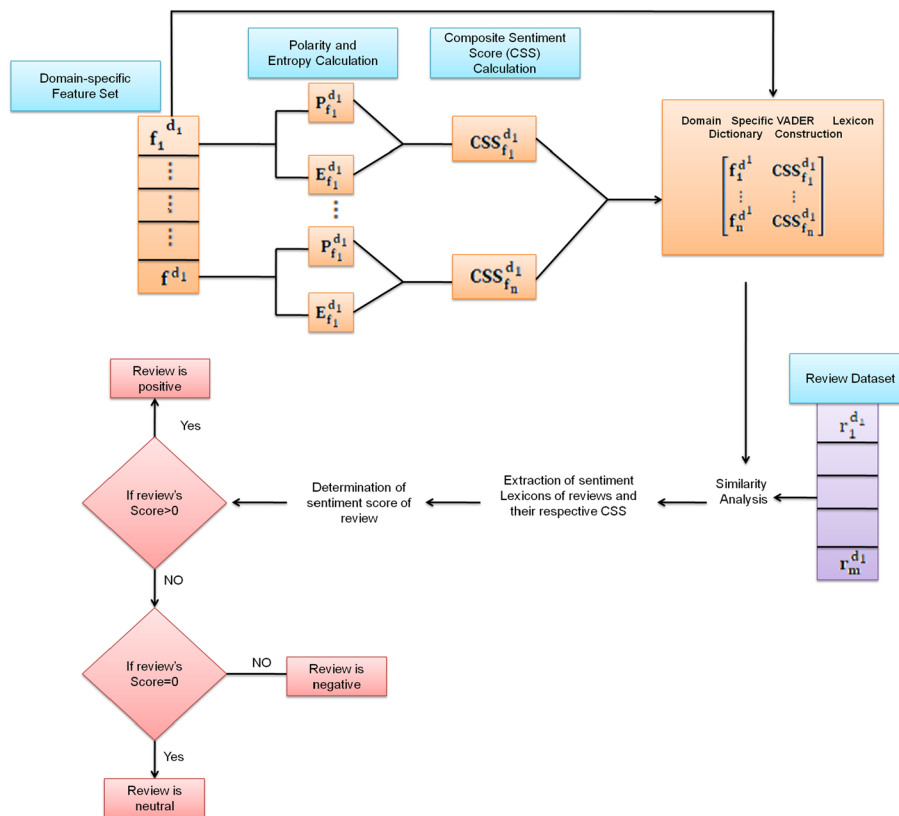


Fig. 3 The model of IVADER lexicon-based sentiment analysis

where $E_{f_n}^{d_m}$ is the entropy of the nth feature of the domain (d_m), and SS is the semantic similarity between a pair of features.

The methods of computing the semantic similarity between two sentences used in this study are a) word vector similarity [46], b) WordNet similarity [46], and c) Word order similarity [47]. Once the semantic similarity (SS) between pairs of features is computed, this information is used to calculate entropy for sentiment analysis using Eq. 7. The entropy measure quantifies the uncertainty or diversity in the sentiment expressed in a text, considering the relationships between terms.

$$Entropy = - \sum_n^{i=1} P_i \cdot \log_2 P_i \tag{7}$$

Where N is the number of sentiment classes (e.g., positive, negative, neutral). P_i is the probability of a term belonging to sentiment class i , which can be estimated based on the semantic similarity scores obtained.

The semantic similarity (SS) between each significant sentiment feature (f_n) and other features in the same domain (d_m) is determined. The semantic similarity score is calculated for each attribute using the natural logarithm (log). This logarithmic transformation emphasizes the variety of unpredictability in the semantic connections between the features. These logarithmic values are added for each important emotion feature in the domain. The entropy of the sentiment features in the domain indicates the varied or uncertain sentiment expressions inside that particular domain. For example, semantic similarity for feature f_1 of the domain (d_m) is estimated based on formulating a pair of features, as shown in Eq. 8.

$$\begin{pmatrix} f_1^{d_m}, f_2^{d_m} \\ f_1^{d_m}, f_2^{d_m} \\ \vdots \\ \vdots \\ \vdots \\ f_1^{d_m}, f_n^{d_m} \end{pmatrix} \tag{8}$$

The composite sentiment score of each significant sentiment feature in each domain is determined using Eq. 9. This composite sentiment score includes the polarity score and entropy of a feature.

$$CSS_{f_n}^{d_m} = P_{f_n}^{d_m} \cdot E_{f_n}^{d_m} \tag{9}$$

where $CSS_{f_n}^{d_m}$ is the composite sentiment score of the nth feature of the domain (d_m).

The polarity score establishes a feature’s sentiment orientation (positive or negative). The entropy details the consistency or variability of sentiment expressions associated with that characteristic. A composite score is generated that balances sentiment direction and variability when we compute the CSS by multiplying these values. The CSS values of all essential sentiment characteristics in a domain can impact the final sentiment categorization. Sentiments are classified based on high CSS values, high polarity scores, and low entropy, which influence sentiment categorization significantly.

A domain-specific lexicon-based VADER dictionary is constructed with a list of significant sentiment lexicons and their corresponding polarity score for each domain. Each

score in a given domain is sent as input to this newly constructed, dictionary-based, domain-specific VADER Lexicon Classifier. The similarity between a review and a list of domain-specific sentiment features is analyzed to extract the sentiment dictionaries available in a review. The overall rating sentiment in a given area is determined using Eq. 10.

$$SS_{r_j}^{d_m} = \frac{\sum_{l=1}^a CSS_{t_l}}{a_{r_j}} \quad (10)$$

where $SS_{r_j}^{d_m}$ is the overall sentiment score of j th review of domain (d_m), a_{r_j} means the total number of extracted sentiment lexicons in j th review of domain (d_m), CSS_{t_l} means the composite sentiment score of sentiment lexicon (t_l) in j th review.

This study classified sentiment as if a review's overall composite sentiment score is more significant than zero; the review is categorized as positive. The classification is considered neutral if the overall composite sentiment score is zero. If a review's overall composite sentiment score is less than zero, the review is classified as negative. The examples concerning the sentiment classification of a review are described below.

Example 1: Positive Sentiment-Camera quality is good, and the resolution of photos is high.

In this example, the domain is electronics. Sentiment lexicons in this review are "good" for the feature "camera quality" and "high" for the feature "photo resolution." The composite sentiment scores of "good" and "high" lexicons obtained from the electronic domain-specific VADER lexicon dictionary are 0.75 and 0.8, respectively. The total number of extracted sentiment lexicons is two. The entire emotion score of the above review is estimated to be 0.775 using Eq. 9. This calculated score is larger than 0, indicating that the review presented above is Positive.

Example 2: Neutral Sentiment-The book had standard examples and average explanations.

In this example, the domain is a book. Sentiment lexicons in this review are "standard examples" for the feature "book" and "average explanations" for the feature "book." The composite sentiment score of the "standard examples" and "average explanations" lexicons obtained from the book domain-specific VADER lexicon dictionary are 0.50 and 0.50, respectively. The total number of extracted sentiment lexicons is two. The overall sentiment score of the above review is estimated to be 0.50 using Eq. 9. This estimated score is approximately 0, meaning the above review is neutral.

Example 3: Negative Sentiment-The battery life is terrible, and the phone frequently crashes.

In this example, the domain is electronics. Sentiment lexicons in this review are "terrible" for the feature "battery life" and "frequently crashes" for the feature "phone." The composite sentiment scores of "terrible" and "frequently crashes" lexicons obtained from the travel domain-specific VADER lexicon dictionary are -0.75 and -0.80, respectively. The total number of extracted sentiment lexicons is two. The overall sentiment score of the above review is estimated to be -0.775 using Eq. 9. This estimated score is less than 0, meaning the above review is negative.

Algorithm 2 IVADER Lexicon-based sentiment analysis

Input: Domain-specific Sentiment feature Set $[F^{d_m}]$, Where $m=1$ to 4

Output: Sentiment Classification of Review

Let domain of dataset be d_1, d_2, d_3, d_4

Stage 1: Construction of Domain-specific VADER lexicon-based dictionary ($Dict^{d_m}$)

For($m=1, m \leq 4, m=m+1$)

{

Let $[F^{d_m}] = \{f_1, f_2, \dots, f_n\}$ Where $n \rightarrow$ number of selected sentiment features

For($i=1, i \leq n, i=i+1$)

{

Compute Polarity score ($P_{f_i}^{d_m}$) of feature (f_i)

Determine Entropy ($E_{f_i}^{d_m}$) of feature (f_i)

Calculate the composite sentiment score of feature (f_i) using the equation $CSS_{f_i}^{d_m} = P_{f_i}^{d_m} \cdot E_{f_i}^{d_m}$

Add $f_i^{d_m}$ and $CSS_{f_i}^{d_m}$ to $Dict^{d_m}$

}

Domain-specific VADER lexicon-based dictionary ($Dict^{d_m}$) is finally constructed.

}

Stage 2: Sentiment Classification of reviews using constructed $Dict^{d_m}$

For($m=1, m \leq 4, m=m+1$)

{

Let Review Set $[R^{d_m}] = \{r_1, r_2, \dots, r_j\}$ Where $j \rightarrow$ number of reviews in domain (d_m)

For($b=1, b \leq j, b=b+1$)

{

Extract sentiment lexicons (t) from review (r_b) using $Dict^{d_m}$ as a reference.

Determine the total number of extracted sentiment lexicons (a_{r_j}) in r_b

Compute overall sentiment score ($SS_{r_j}^{d_m}$) of review (r_b)

If ($SS_{r_j}^{d_m} > 0$)

Then r_b is a positive review

Else

{

If ($SS_{r_j}^{d_m} = 0$)

```

    Then ri is a neutral review
Else
    Then ri is a negative review
}
}
}
End

```

Results

This section discusses the performance analysis of the IVADER lexicon-based sentiment analysis model in multiple domains. The experimental setup is executed on a single system running 64-bit Windows 11 and an Intel Pentium CPU with 32 GB of RAM and 1 GB SSD, the dataset on the Python interface. The dataset is categorized into 70% training and 30% testing datasets. The performance indicators used in the comparative analysis are accuracy, precision, recall, f-measure, specificity, and error presented using Eqs. 11, 12, 13, 14, 15, 16 and 17.

$$\text{Accuracy (A)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$\text{F - Measure (F)} = 2 \times \frac{\text{P} \times \text{R}}{\text{P} + \text{R}} \quad (14)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15)$$

$$\text{AUC} = (\text{Percent Concordant} + 0.5 * \text{Percent Tied})/100 \quad (16)$$

$$\text{Error} = (\text{Approximate Value} - \text{Exact Value})/\text{Exact Value} \times 100 \quad (17)$$

where TP represents the proportion of reviews accurately categorized as positive and positive overall.

TN represents the percentage of evaluations that are both genuinely and appropriately categorized as negative.

FP depicts the percentage of negative evaluations that were mistakenly categorized as positive.

FN depicts the percentage of positive evaluations that were mistakenly categorized as negative.

Accuracy is measured as the proportion of accurately identified reviews. Figure 4 illustrates the accuracy-based performance evaluation of the proposed IVADER model in four domains considered in this study, i.e., electronics, DVDs, books, and kitchens. The average accuracy of IVADER (98.64%) in the multi-domain sentiment dataset is higher than that of conventional lexicon models, namely CAN [40], SentiDraw [33], SL-MLP [34], UDA-SP [41] and CDSARFE [42]. A comparative analysis of different sentiment classifier models across multi-domains is demonstrated in Table 3. As a result, our practices outperform traditional techniques in terms of accuracy.

Loss is a metric to assess the fitness of data in a sentiment classification model. It evaluates the error of the model on the dataset. Figure 5 depicts the comparative analysis of training and testing loss for the proposed IVADER model. It signifies that as the epoch increases, training and testing loss decreases. The testing and training loss are closer, so the model performs satisfactorily classifying reviews in both the testing and training stages.

An accuracy and error plot indicates the model performance throughout the training phase. Figure 6 illustrates the accuracy and error plot of the proposed model and indicates that training error is reduced and performs better accuracy by increasing the epochs.

Precision is a metric to determine how appropriately a classifier performs in multi-domains. It is calculated as the percentage of positively anticipated responses divided by the total amount of responses categorized as positive. Figure 7 depicts the precision-based performance analysis of the proposed model and comparative analysis with existing studies, i.e., CAN [40], SentiDraw [33], SL-MLP [34], UDA-SP [41], and CDSARFE [42]. The average precision of the IVADER is 97% in the multi-domain sentiment dataset, which is better than in existing studies. The higher precision for the proposed model signifies that the number of reviews that are actually negative in a specific domain but incorrectly classified as positive is lower.

Recall is the proportion of the same expectation of positive feedback to all the positive feedback. Figure 8 portrays the recall-based performance evaluation of the proposed model in a multi-domain dataset comparative analysis with existing studies. The average recall of the proposed IVADER model is 94% and higher compared with existing studies. The higher recall for IVADER represents that the number of positive reviews in a specific domain but incorrectly classified as negative is less than other methods.

The F-measure indicates the harmonic average of precision and recall and combines recall and precision. The average F-measure of the proposed IVADER model is 92% in the multi-domain sentiment dataset, which is superior to existing studies, illustrated in Fig. 9.

Specificity is the proportion of correctly identified negative reviews to the dataset's total number of negative reviews. The performance study of existing lexicon-based models in multi-domain datasets, i.e., electronics, DVDs, books, and kitchens, is depicted in Fig. 10. The average specificity of the proposed IVADER model is 96%, higher than

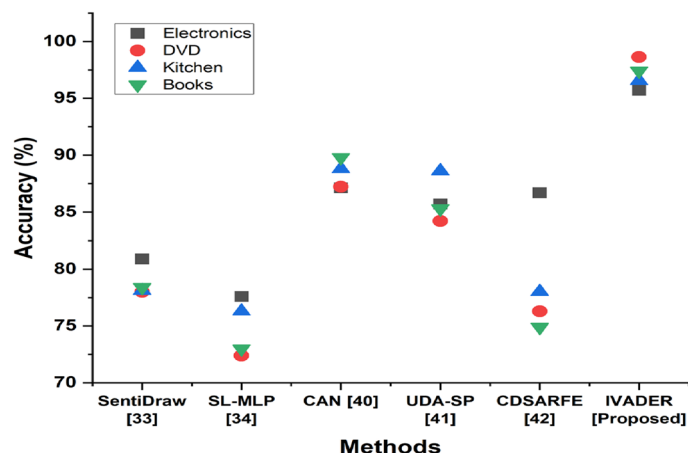


Fig. 4 Accuracy-based performance analysis in multi-domains

existing studies considered, namely SentiDraw [33], SL-MLP [34], CAN [40], UDA-SP [41] and CDSARFE [42]. The higher specificity for the IVADER model implies that the number of negative reviews in a specific domain incorrectly classified as positive is significantly lower. The number of correctly classified negative reviews in each domain compared to other methods is high.

The Receiver Operating Characteristic (ROC) curve is a performance measure for classification problems at different threshold settings. It is a probability curve, and the area under the curve (AUC) indicates how efficiently the model can distinguish between positive and negative valuation classes. Figure 11 portrays the ROC curve of different lexicon-based sentiment classifiers. A higher AUC is noted for the IVADER model than existing studies, which signifies that the proposed model performed superior at distinguishing between positive and negative reviews in each domain.

The time required to classify the reviews in each domain using the created domain-specific lexicon-based dictionary as a reference is the training time measured in seconds. Figure 12 pictures the training time for the proposed model and different existing lexicon-based models in multi-domains, i.e., electronics, DVDs, books, and kitchens. The average training time of the proposed IVADER model on the multi-domain sentiment dataset is 44 s, lower than existing studies. This signifies that the proposed model can effectively classify sentiment reviews in less time.

The proposed IVADER model is tested with lower-size data samples in each domain, i.e., 250, 500, and 750. This evaluates the model’s validity in lesser datasets if the vocabulary is well-established and tailored. The comparative evaluation of the model with four parameters is demonstrated in Fig. 13. It symbolizes that the proposed model performs adequately in small-size data samples in multi-domain sentiment analysis.

Further, the data samples are randomly selected to validate and impact the proposed model. First, 10,000 samples have been used, i.e., 2500 reviews in each domain. Second, 20,000 samples have been selected randomly, i.e., 5000 reviews in each domain, as illustrated in Table 4. The outcome of the IVADER model is presented in Fig. 14. It signifies the stability of the proposed model. Figure 15 the performance of the INVADER model, which symbolizes stability.

Table 3 Comparative analysis of various sentiment classifier models in multiple domains

Methods/Domain	Electronics	DVD	Kitchen	Books	Average
SentiDraw [33]					
A (%)	80.9	78	78.1	78.4	78.85
P (%)	60	63	65	58	61.5
R (%)	62	65	70	72	67.25
F (%)	66	68	58	55	61.75
SL-MLP [34]					
A (%)	77.6	72.4	76.3	73	74.825
P (%)	55	58	66	81	65
R (%)	42	44	46	48	45
F (%)	45	48	53	58	51
CAN [40]					
A (%)	87.13	87.23	88.8	89.8	88.24
P (%)	75	65	72	70	70.5
R (%)	78	75	76	68	74.25
F (%)	73	75	77	80	76.25
UDA-SP[41]					
A (%)	85.7	84.2	88.6	85.3	85.7
P (%)	85	88	82	86	85
R (%)	81	83	84	86	81
F (%)	80	83	85	86	80
CDSARFE [42]					
A (%)	86.7	76.29	78.01	74.9	86.7
P (%)	83	86	87	81	83
R (%)	80	87	85	81	80
F (%)	81	82	84	86	81
IVADER (proposed)					
A (%)	95.7	98.64	96.55	97.41	98.64
P (%)	97	93	96	91	97
R (%)	91	94	92	93	94
F (%)	93	92	90	91	92

The proposed model is also validated with a different dataset, i.e., Sentence Polarity (<https://www.cs.cornell.edu/people/pabo/movie-review-data/>). It consists of 5,331 positive and 5,331 negative reviews. The performance of the IVADER is evaluated using five parameters, i.e., accuracy, precision, recall, f1-score, and specificity. The comparative analysis of the IVADER model is demonstrated using two different datasets, depicted in Fig. 16. It denotes that the IVADER model yielded better outcomes in both datasets, as exhibited in Table 5.

Discussion

Developing a sentiment lexicon for datasets with limited resources is time-consuming and expensive [48]. The performance of ML algorithms fails when used with multi-domain datasets [49]. This study proposed an IVADER-based model for multiple-domain sentiment analysis. A domain-specific feature set is constructed, as exhibited in Algorithm 1. The domain-specific aspects of similar Synsets in WordNet are determined and added as significant features for the domain expression. The

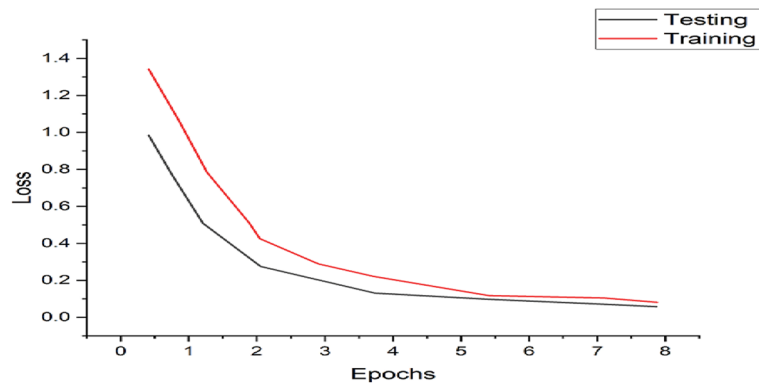


Fig. 5 Comparative assessment of the training and testing losses for IVADER

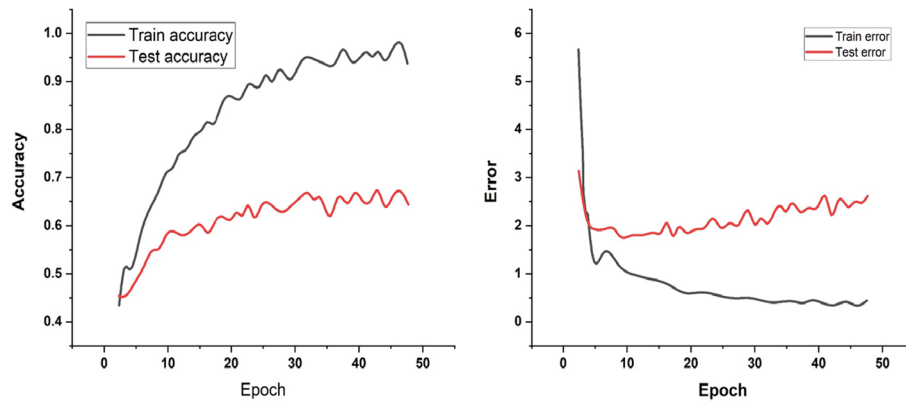


Fig. 6 Comparative analysis of training and testing accuracy and error plot for IVADER

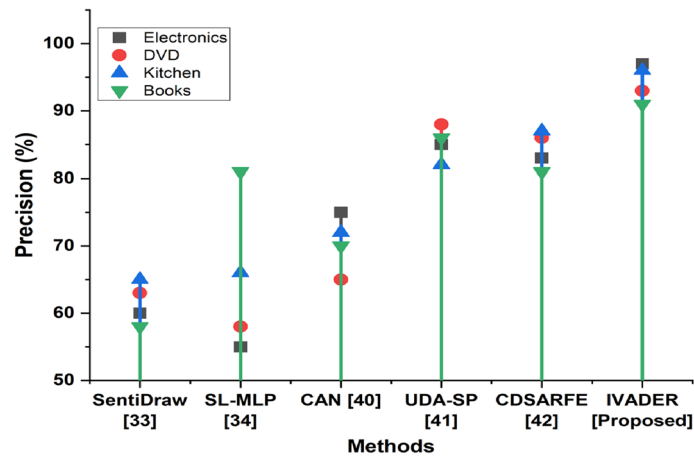


Fig. 7 Performance evaluation of several models using precision in multi-domains

classification technique of the IVADER lexicon contains two significant steps. In the initial phase, a dynamic domain-specific VADER lexicon-based dictionary is built by calculating the overall polarity value.VADER lexicon-based dictionary by calculating

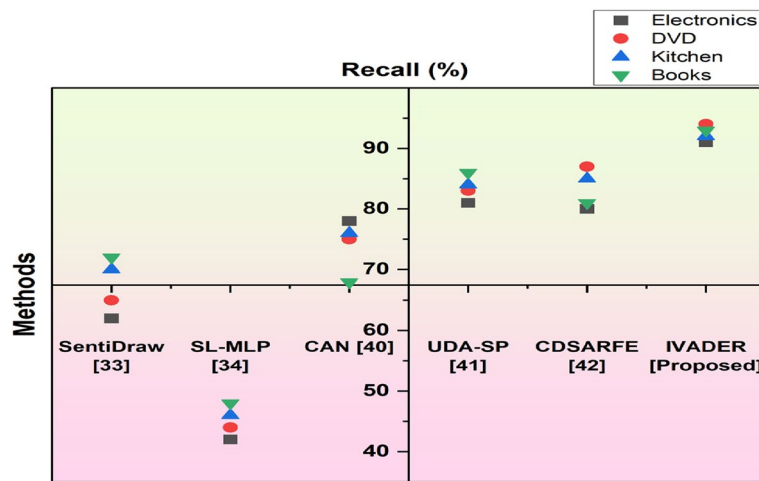


Fig. 8 Recall-based performance analysis of different models in multi-domains

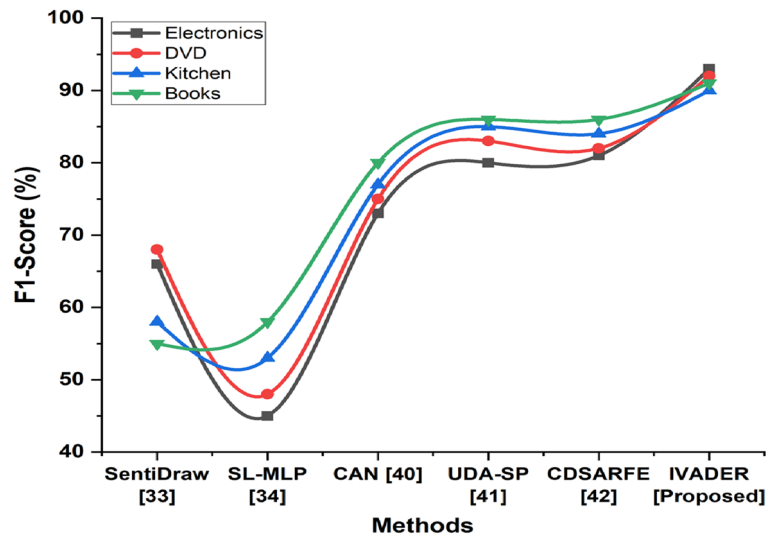


Fig. 9 F-measure-based performance analysis of different models in multi-domains

the overall polarity score. The second phase is a classification of reviews based on sentiment scores constrained by a lexicon-based domain-specific VADER dictionary. An IVADER lexicon-based sentiment analysis model is presented in Fig. 3 and Algorithm 2. A domain-specific lexicon-based VADER dictionary is created with a list of important sentiment lexicons and their connected polarity score for each domain. IVADER Lexicon-based sentiment analysis.

Four different domain datasets are considered in this study to validate the proposed model. Different evaluation parameters, i.e., accuracy precision, recall, f-measure, and specificity, are employed to evaluate the model’s performance. The IVADER model achieves an accuracy of 98.64%, as demonstrated in Fig. 4, and compared with existing studies, i.e., SentiDraw [33], SL-MLP [34], CAN [40], UDA-SP [41] and CDSARFE [42]. The IVADER model achieved a precision of 97%, recall of 94%, f-measure of 92%,

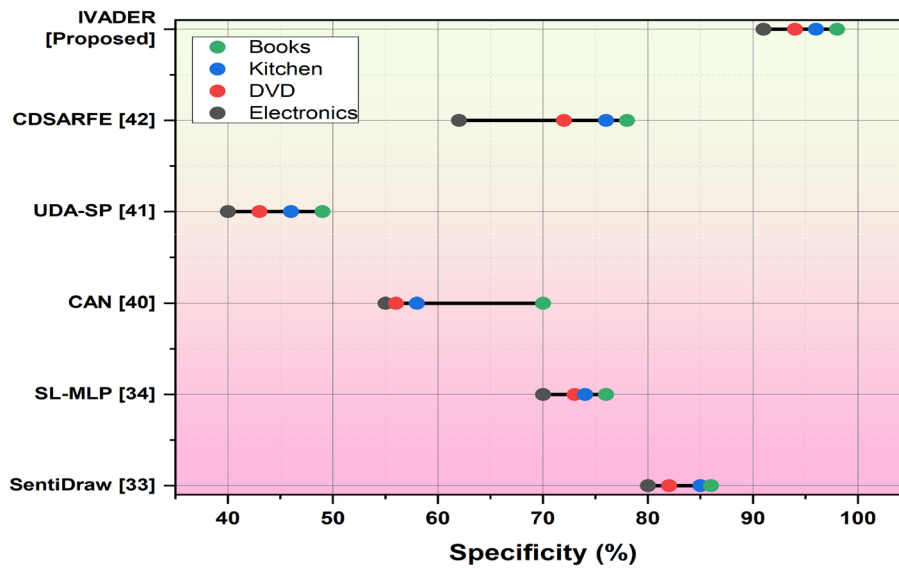


Fig. 10 Specificity-based performance analysis of different models in multi- domains

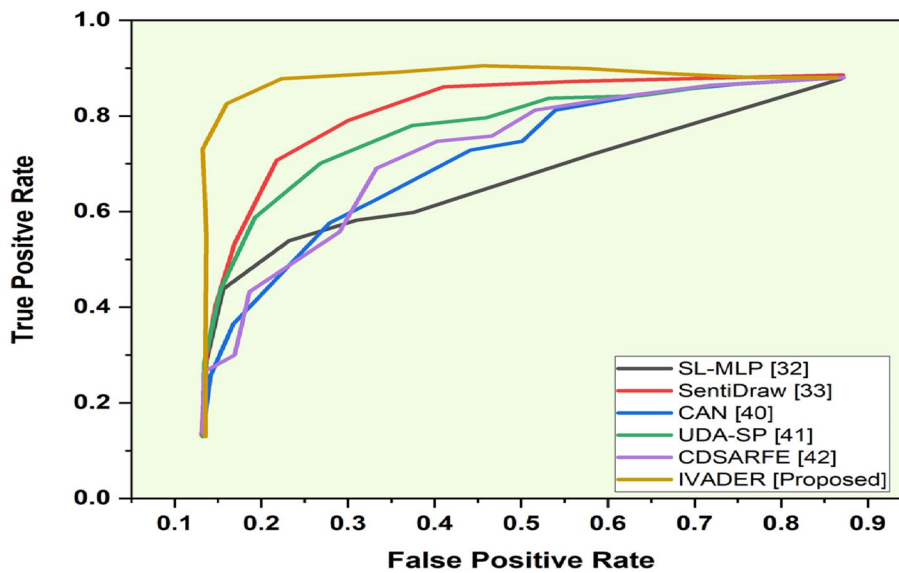


Fig. 11 ROC curve of various lexicon-based sentiment classifiers

and specificity of 96%, illustrated in Figs. 7,8,9 and 10, respectively, compared with existing studies. A higher AUC is identified for the IVADER model, which symbolizes superior at distinguishing between positive and negative reviews in each domain, as presented in Fig. 11. The average training time of the proposed model on the multi-domain sentiment dataset is 44 s, lower than existing studies, as pictured in Fig. 12. The proposed model can effectively classify reviews in less time. The comparative analysis of training and testing loss for the IVADER model is exemplified in Fig. 5, which indicates that training and testing loss decreases as the epoch increases. Figure 6 demonstrates the accuracy and error plot of the proposed model and demonstrates that training error is decreased and achieves better accuracy by expanding the

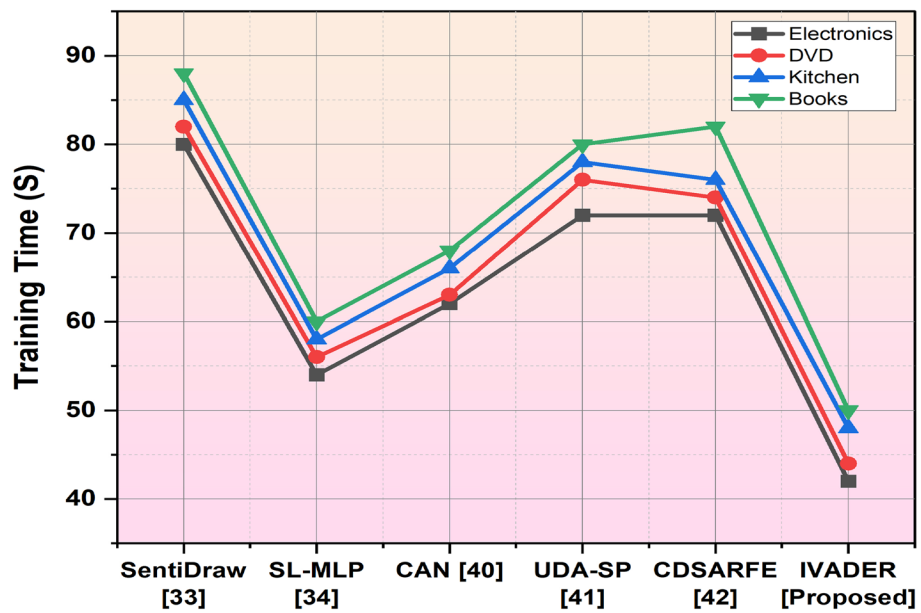


Fig. 12 Training time-based performance analysis of different models in multi-domains

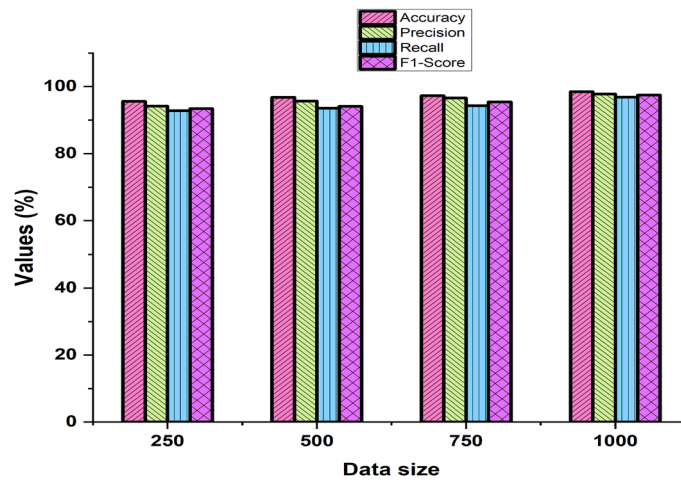


Fig. 13 Accuracy, precision, recall, and f1-score-based performance analysis of different sizes of dataset

Table 4 Data size

Dataset		Data size	
		Number of positive and negative reviews	Number of positive and negative reviews
Multi-Domain Sentiment Dataset	Electronics	2500	5000
	Books	2500	5000
	DVD	2500	5000
	Kitchen	2500	5000

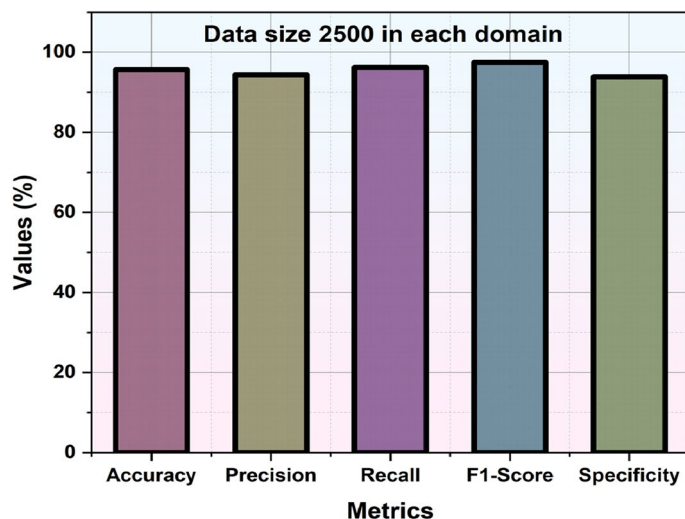


Fig. 14 Accuracy, precision, recall, f1-score and specificity with size 2500

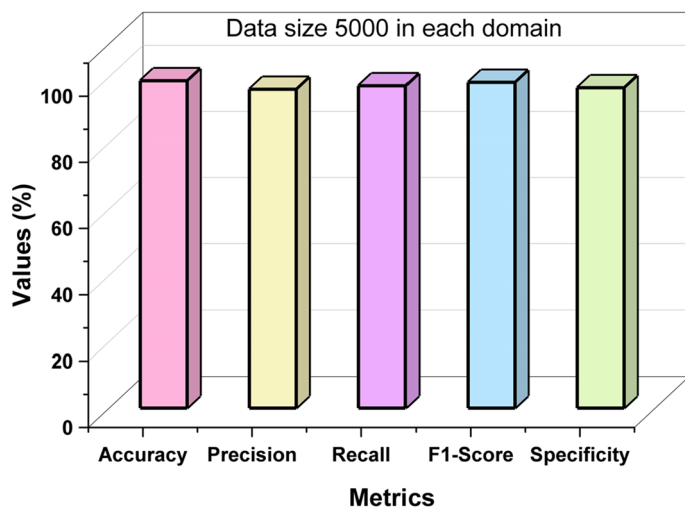


Fig. 15 Accuracy, precision, recall, f1-score, and specificity with size (5000)

epochs. This outcome is due to the provision of composite sentiment scores, including polarity and entropy, for lexicons from this dictionary. This constructs IVADER especially well suited for processing customer reviews in multi-domains with different characteristics than conventional methods.

Although existing lexicon-based models perform better in classifying customer reviews into four domains, They have certain limitations. The parameter selection method is difficult with CAN, and its computing complexity is higher [40]. The first issue is that the accuracy the suggested model may attain is constrained by the precision of the UDA technique selected in phase 1. The second problem is that the number of epochs for training C in phase 3 must be chosen automatically or semi-automatically rather than experimentally [41]. Due to the high number of characteristics produced,

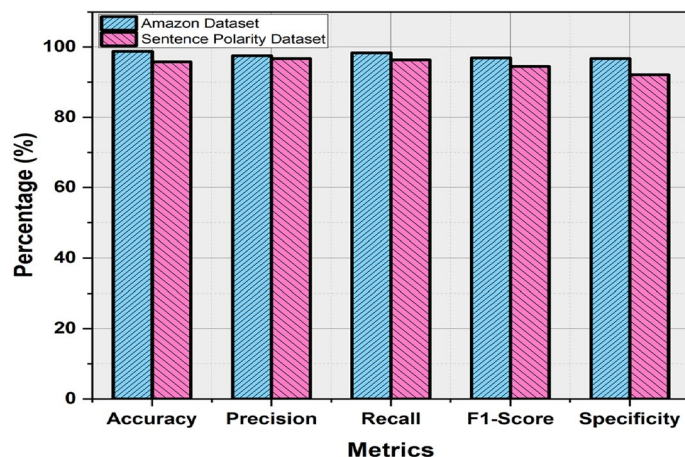


Fig. 16 Comparison of two datasets, Amazon and Sentence Polarity datasets

Table 5 Comparison of two different datasets

	Percentage (%)	
	Amazon dataset	Sentence polarity dataset
Accuracy	98.79	95.83
Precision	97.54	96.71
Recall	98.35	96.35
F1-Score	96.89	94.5
Specificity	96.73	92.15

implicit context-specific feature selection must be addressed. The threshold also varies for implicit characteristics, which impacts how well the system works. The shortcoming of this approach is the omission of implicit feature extraction [42].

The INDAER model is also tested by reducing the dataset size in four domains, as exemplified in Fig. 13. Additionally, the IVADER model is validated by varying the dataset into two stages, i.e., 2500 samples each and 5000 samples each, as illustrated in Figs. 14, 15, and Table 4. Further, the proposed model is validated using different datasets, and a comparative study is depicted in Fig. 16 and Table 5. Based on the analysis, it signifies that the IVADER model performed better and is comprehensively validated.

Threats to validity

The primary threats related to this work are product bias, promotional offers, inconsistent reviews, and incorrect data extraction. In addition, the ratings are collected from just one e-commerce platform and applied to the proposed framework. Technological advances and the resulting consumer demand may change over time concerning competing products. Other parameters like characteristic product conversions will be an exciting study area for different requirements. Considerable efforts have been made in

data acquisition, preprocessing, and feature selection. Appropriate studies are carefully selected based on selection criteria.

Limitations of the study

The ambiguity of words and sentences in context is a limitation. Although VADER has a dictionary of words with a certain sensation, the meaning of a word might vary based on the context. Improved versions could offer contextual enhancements, which still need to be addressed. While VADER attempts to identify negations, it could have problems with complex negation patterns or double annulments that call for a more in-depth understanding. Like irony, sarcasm detection is challenging for VADER and depends on contextual indications and techniques.

Future work

Improved techniques for sarcasm, irony, and other subtle expressions of emotion may improve the model's capacity to handle additional complex text types, specifically on social media platforms. Incorporating sentiment analysis would enable Improved Vader to identify a wider range of sentiments represented in text, proceeding beyond positive and negative sentiments.

Conclusion

The validation sentiment can vary with the change of domain. Therefore, a domain-specific construction of the sentiment lexicon is required to improve sentiment classification. Conventional VADER lexicons only consider the polarity of dictionaries. This study proposes an IVADER lexicon-based model for multi-domain sentiment analysis. The classification technique of the proposed model possesses two significant steps: a dynamic domain-specific VADER lexicon-based dictionary and a classification of reviews based on sentiment scores constrained by a lexicon-based domain-specific VADER dictionary. The performance of the IVADER model is evaluated using four multi-domain datasets. The IVADER model accomplished an accuracy of 98.64%, a precision of 97%, a recall of 94%, an f1-measure of 92%, and a specificity of 96% compared with existing studies. Further, the IVADER model takes less training time, i.e., 44 s, in classifying sentiment. Furthermore, the model is validated by varying the dataset's size and evaluated on a different dataset.

Acknowledgements

Authors are thankful to their departments for providing research environments and facilities to perform this research.

Author contributions

Both authors contributed equally.

Funding

Open Access funding provided by Østfold University College, Halden, Norway.

Availability of data and materials

This study uses a publicly available multi-domain sentiment dataset (<https://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html>).

Declarations

Ethics approval and consent to participate

'Not applicable'. No material is used to get ethical approval also no Table or Figure or any copyright material is taken from any sources.

Consent for publication

Authors give consent for publication.

Competing interests

Authors do not have any financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Received: 2 January 2023 Accepted: 14 December 2023

Published online: 07 January 2024

References

- Alharbi NM, Alghamdi NS, Alkhamash EH, Al Amri JF. Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews. *Math Prob Eng*. 2021. <https://doi.org/10.1155/2021/5536560>.
- Xia H, Yang Y, Pan X, Zhang Z, An W. Sentiment analysis for online reviews using conditional random fields and support vector machines. *Electron Commer Res*. 2020;20(2):343–60.
- Tang F, Fu L, Yao B, Xu W. Aspect based fine-grained sentiment analysis for online reviews. *Inf Sci*. 2019;488:190–204.
- Huang M, Xie H, Rao Y, Liu Y, Poon LK, Wang FL. Lexicon-based sentiment convolutional neural networks for online review analysis. *IEEE Transactions on Affective Computing*; 2020.
- Ghiassi M, Lee S. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Syst Appl*. 2018;106:197–216.
- Yang L, Li Y, Wang J, Sherratt RS. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*. 2020;8:23522–30.
- Li W, Zhu L, Shi Y, Guo K, Cambria E. User reviews: sentiment analysis using lexicon integrated two-channel CNN–LSTM family models. *Appl Soft Comput*. 2020;94: 106435.
- Du M, Li X, Luo L. A training-optimization-based method for constructing domain-specific sentiment lexicon. *Complexity*. 2021. <https://doi.org/10.1155/2021/6152494>.
- Al-Natour S, Turetken O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *Int J Inf Manag*. 2020;54: 102132.
- Kumar S, Yadava M, Roy PP. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Inf Fusion*. 2019;52:41–52.
- Naresh Kumar KE, Uma V. Intelligent sentiment-based lexicon for context-aware sentiment analysis: optimized neural network for sentiment classification on social media. *J Supercomput*. 2021;77(11):12801–25.
- Korovkinas K, Danėnas P, Garšva G. SVM accuracy and training speed trade-off in sentiment analysis tasks. In *International Conference on Information and Software Technologies*. Springer, Cham, 2018, pp. 227–239
- Zhou Q, Xu Z, Yen NY. User sentiment analysis based on social network information and its application in consumer reconstruction intention. *Comput Hum Behav*. 2019;100:177–83.
- Sun Q, Niu J, Yao Z, Yan H. Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Eng Appl Artif Intell*. 2019;81:68–78.
- Alharbi ASM, de Doncker E. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cogn Syst Res*. 2019;54:50–61.
- Wang P, Li J, Hou J. S2SAN: A sentence-to-sentence attention network for sentiment analysis of online reviews. *Decis Support Syst*. 2021;149: 113603.
- Behera RK, Jena M, Rath SK, Misra S. Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. *Inf Process Manag*. 2021;58(1): 102435.
- Priyadarshini I, Cotton C. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. *J Supercomput*. 2021;77(12):13911–32.
- Styawati S, Nurkholis A, Aldino AA, Samsugi S, Suryati E, Cahyono RP. Sentiment analysis on online transportation reviews using Word2Vec text embedding model feature extraction and support vector machine (SVM) algorithm. In *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*. IEEE, 2022, pp. 163–167
- Kayıkçı Ş. SenDemonNet: sentiment analysis for demonetization tweets using heuristic deep neural network. *Multi-media Tools Appl*. 2022;81(8):11341–78.
- Nasfi R, Bouguila N. Sentiment Analysis from User Reviews Using a Hybrid Generative-Discriminative HMM-SVM Approach. In: *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Cham: Springer International Publishing, 2022; pp. 74–83
- Sagarino VMC, Montejo JIM, Ceniza-Canillo AM. Sentiment analysis of product reviews as customer recommendations in shopee philippines using hybrid approach. In *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)* (). IEEE, 2022; pp. 1–6
- Benarafa H, Benkhalifa M, Akhloufi M. WordNet semantic relations based enhancement of KNN model for implicit aspect identification in sentiment analysis. *Int J Comput Intell Syst*. 2023;16(1):3.
- Jain PK, Quamer W, Saravanan V, Pamula R. Employing BERT-DCNN with a sentic knowledge base for social media sentiment analysis. *J Ambient Intell Humaniz Comput*. 2023;14(8):10417–29.
- Li Y, Guo H, Zhang Q, Gu M, Yang J. Imbalanced text sentiment classification using universal and domain-specific knowledge. *Knowl-Based Syst*. 2018;160:1–15.

26. Xing FZ, Pallucchini F, Cambria E. Cognitive-inspired domain adaptation of sentiment lexicons. *Inf Process Manag.* 2019;56(3):554–64.
27. Deng D, Jing L, Yu J, Sun S, Ng MK. Sentiment lexicon construction with hierarchical supervision topic model. *IEEE/ACM Trans Audio Speech Lang Process.* 2019;27(4):704–18.
28. Dey A, Jenamani M, Thakkar JJ. Cross-D-vectorizers: a set of feature-spaces for cross-domain sentiment analysis from consumer review. *Multimedia Tools Appl.* 2019;78(16):23141–59.
29. Frangidis P, Georgiou K, Papadopoulos S. Sentiment analysis on movie scripts and reviews. In: *IFIP International Conference on Artificial Intelligence Applications and Innovations.* Springer, Cham, 2020, pp. 430–438
30. Wook M, Razali NAM, Ramli S, Wahab NA, Hasbullah NA, Zainudin NM, Talib ML. Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback). *Educ Inf Technol.* 2020;25(4):2549–60.
31. Moussa ME, Mohamed EH, Haggag MH. A generic lexicon-based framework for sentiment analysis. *Int J Comput Appl.* 2020;42(5):463–73.
32. Lee SW, Jiang G, Kong HY, Liu C. A difference of multimedia consumer's rating and review through sentiment analysis. *Multimedia Tools Appl.* 2021;80(26):34625–42.
33. Sharma SS, Dutta G. SentiDraw: using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination. *Inf Process Manag.* 2021;58(1): 102412.
34. Beigi OM, Moattar MH. Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. *Knowl-Based Syst.* 2021;213: 106423.
35. Hasanati N, Aini Q, Nuri A. Implementation of support vector machine with lexicon based for sentiment ANALYSIS ON TWITTER. In: *2022 10th International Conference on Cyber and IT Service Management (CITSM).* IEEE, 2022, pp. 1–4
36. Juanita S, Adiyarta K, Syafrullah M. Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naïve Bayes Model. In: *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).* IEEE, 2022, 379–382
37. Thangavel P, Lourdasamy R. A lexicon-based approach for sentiment analysis of multimodal content in tweets. *Multimedia Tools Appl.* 2023. <https://doi.org/10.1007/s11042-023-14411-3>.
38. Tahayna B, Ayyasamy RK, Akbar R, Subri NFB, Sangodiah A. Lexicon-based non-compositional multiword augmentation enriching tweet sentiment analysis. In: *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS).* IEEE, 2022, pp. 19–24
39. Ojeda-Hernández M, López-Rodríguez D, Mora Á. Lexicon-based sentiment analysis in texts using formal concept analysis. *Int J Approximate Reasoning.* 2023;155:104–12.
40. Yue C, Cao H, Xu G, Dong Y. Collaborative attention neural network for multi-domain sentiment classification. *Appl Intell.* 2021;51(6):3174–88.
41. Badr H, Wanas N, Fayek M. Unsupervised domain adaptation with post-adaptation labeled domain performance preservation. *Mach Learn Appl.* 2022;10: 100439.
42. Geethapriya A, Valli S. An enhanced approach to map domain-specific words in cross-domain sentiment analysis. *Inf Syst Front.* 2021. <https://doi.org/10.1007/s10796-020-10094-5>.
43. Berdyugina D, Cavallucci D. Automatic extraction of inventive information out of patent texts in support of manufacturing design studies using Natural Languages Processing. *J Intell Manuf.* 2023;34(5):2495–509.
44. Akuma S, Lubem T, Adom IT. Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets. *Int J Inf Technol.* 2022;14(7):3629–35.
45. Nayak S, Sharma YK. A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis. *Decis Anal J.* 2023;8: 100289.
46. Ahmad F, Faisal M. A novel hybrid methodology for computing semantic similarity between sentences through various word senses. *Int J Cogn Comput Eng.* 2022;3:58–77.
47. Haque S, Eberhart Z, Bansal A, McMillan C. Semantic similarity metrics for evaluating source code summarization. In: *Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension, 2022;* pp. 36–47.
48. Alemneh, G.N., Rauber, A. and Atnafu, S., 2019, May. Dictionary based amharic sentiment lexicon generation. In: *International Conference on Information and Communication Technology for Development for Africa.* Springer, Cham, pp. 311–326
49. Aziz AA, Starkey A. Predicting supervised machine learning performances for sentiment analysis using contextual-based approaches. *IEEE Access.* 2019;8:17722–33.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.