Sentiment Analysis of Online Product Reviews using ML, DL and XAI Techniques: A Systematic Review

Rahul B. Mannade[1], Praveen C. Shetiye[2]

[1] Department of Information Technology, Government College of Engineering, Aurangabad, Maharashtra 431005, India

[2] Department of Master of Computer Applications, Government College of Engineering, Aurangabad, Maharashtra431005, India

Corresponding Author:

Rahul B. Mannade, Department of Information Technology, Government College of Engineering, Aurangabad, Maharashtra 431005, India

Email: mannade.rahul@gmail.com

ABSTRACT

With the advent of Web 2.0 along with affordable internet data rates as well as smartphones, many people are doing online shopping and writing reviews on shopping sites. These reviews have become increasingly vital for both consumers and sellers. Online customers and their reviews are increasing abruptly. Since customer reviews are unstructured in nature, it is difficult to unsheathe the sentiment from them. The term sentiment analysis (SA) refers to a technique that analyzes and detects the emotions, perspectives, attitudes, and sentiment of individuals hidden in review text. This paper examines and documents previous work in SA related to online product reviews using Machine Learning (ML), Deep Learning (DL), and Explainable Artificial Intelligence (XAI). Study identifies various sentiment analysis levels, approaches, datasets and feature extraction techniques applied in past work. The findings of this review revealed that the most widely used SA approaches for this domain are the Support Vector Machines (SVM) from Machine Learning and the Convolutional Neural Networks (CNN) from Deep Learning. Research in DL approaches can be extended using Hybrid DL models with novel word embedding methods. XAI methods can be used to make opaque DL models more interpretable in turn trustworthy.

Keywords:Sentiment Analysis, Machine Learning, Deep Learning, Explainable Artificial Intelligence, Product Review

1. INTRODUCTION

Availability of reasonable internet data rates nowadays has resulted into hasty growth in online users. Since, last few years, several e-commerce platforms like Flipkart, Amazon, Snapdeal, Myntra, eBay etc. provide online shopping services to consumers. With explosion of low internet charges and low cost smartphones, there is a tremendous growth in number of online buyers. In 2023, the number of digital buyers in world is at 2.64 billion i.e. 33.3% of the population worldwide [1]. It means one among every three persons across the globe is an online shopper. Proliferation of Word Wide Web along with e-commerce, have resulted into online shopping as daily routine of human being [2]. In earlier days, a product reviews were usually written only by experts having domain knowledge [3]. But after the emergence of Web 2.0, scenario is changed. In present days, many people are buying variety of products using online shopping sites and expressing their perspectives about it in the form of reviews. Sometimes, they share the images and/or videos of the purchased product, as well. Common sources of consumer reviews are micro blog posts, tweets and comments left by customers on social sites, micro blogs etc. [4]. Like traditional businesses, in e-Commerce also consumer management is key factor which is measured with consumer engagement. Retention of existing customers and expansion of new customers is very important for any business in order to survive and to make progress in competitive market. Receiving a feedback and reviews from consumers is a positive aspect towards the consumer engagement. Customer's reviews reflect individual's satisfaction and/or dissatisfaction. Review may be positive, negative or neutral [5]. Consumers always attract towards the product having huge number of positive reviews and they tend to detract from product with negative review or "*no reviews*". But review with pessimistic shade are also useful which enables businesses owners to identify the existing flaws and take corrective actions [6]. Customer reviews play a vital role for decision making of other buyers. The online product reviews produce a wealth of information. By analyzing reviews of other buyers, potential shoppers can take quick decision in less time [7]. Form seller's point of view, stakeholder's satisfaction is an important criteria for evaluating consumer's expectations from company [8]. Consumer reviews might help sellers analyzing the factors which were not considered in past, such as delivery time, packing style, humbleness of delivery person, availability of customer support etc. Day by day, the volume of customer's reviews on e-commerce platform is increasing tremendously. With large volume of reviews an individual buyer may not be able to deduce overall sentiment about a certain product [9]. Most of the time, review data is in unstructured form and it is not possible to detect polarity of such huge amount reviews using manual process. Hence, with ceaselessly growing volume of consumer review data, some automated mechanism using ML/DL is required in online shopping domain to analyze and detect the polarity of review data. Hence, sentiment analysis of customer reviews is most important. Sentiment analysis (SA), an alias of opinion mining is the field of research used to extract emotions/sentiment of peoples form textual data. Previous studies [10] [11], [12], [13] and [14] have investigated SA of online product reviews using various methodologies viz. lexicon-based, ML, DL techniques etc., but use of Deep Learning methods with XAI techniques is under explored. In order to fill this research gap, a Systematic Literature Review (SLR) was carried out on Sentiment Analysis of Online Product Reviews using ML, DL and XAI techniques to identify consumer's sentiment from their expressed textual review and interpret the DL model. Rest of the paper is organized as: Section 2 describes methodology used for this review process, in Section 3 results and discussions are presented, section 4 gives key findings and scope for future work and Section 5 concludes the paper.

2. METHODOLOGY

In this review, we used a three-step systematic literature review (refer Figure 1.) process suggested by David Denyer et al. in [15] to conduct the survey of past work in sentiment analysis of product reviews with ML/DL/XAI methods. Irrespective of domain, a systematic literature reviews yields precious contribution in generating useful ideas and detecting possible defects in the existing research [11].

2.1 Review planning

This review was planned with aim to identify the ML/DL techniques applied in sentiment analysis of customer reviews for online product/(s) and identify Explainable Artificial Intelligence (XAI) techniques used in product review domain to make black-box models more interpretable. We have tried to find out the answers to following research questions:

RQ#1 : What are the various ML and DL methods used so far in sentiment analysis of online customer product reviews?

RQ#2 : What are the different datasets used for sentiment analysis of online customer product reviews?

RQ#3 : Is research on DL methods able to pay attention on both 'semantic' and 'emotional 'information of product reviews?

RQ#4 : Are XAI techniques able to build trust in DL methods?

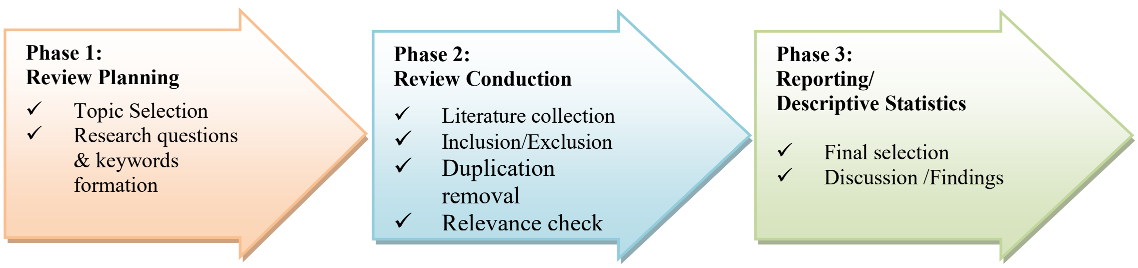


Figure 1. Three-step SLR process used for review process

2.2 Conducting a Review

Initially, we searched research articles on "Sentiment analysis online product reviews using ML, DL and XAI techniques" over the online platforms https://ieeexplore.ieee.org, www.scholar.google.com and https://www.sciencedirect.com, based on the research question constituted in section 2.1. Majority of articles were from ieeexplore.ieee.org. To limit the count of relevant literature, we applied permissible and nonpermissible criteria as mentioned below:

2.2.1 Permissible Criteria – allows a paper for review

- Research papers published during year 2003 to 2023, related to SA in online product reviews domain.
- Research papers demonstrating SA in online product reviews domain using ML techniques.
- Research papers demonstrating SA in online product reviews domain using DL techniques.
- Research papers demonstrating SA in online product reviews domain using XAI techniques.

2.2.2 Nonpermissible Criteria – denies a paper for review

- Research papers without justification of contribution.
- Research papers not relevant to this domain.
- Reviews or survey articles on SA of product reviews without any findings.

At first stage, with keyword search like 'sentiment analysis', 'machine learning', 'deep learning', 'product reviews', 'ML', 'DL', 'XAI', 'Lexicon', 'Aspect' etc., total 84 research papers were selected. By applying inclusion/exclusion criteria and by removing redundant papers, we got 60 articles. Then, we scrutinized selected articles from relevance point of view and finally we come up with 56 numbers of research papers as a final selection for this review (refer Figure 2). Summary of studied articles is shown in tables 1 to 3.

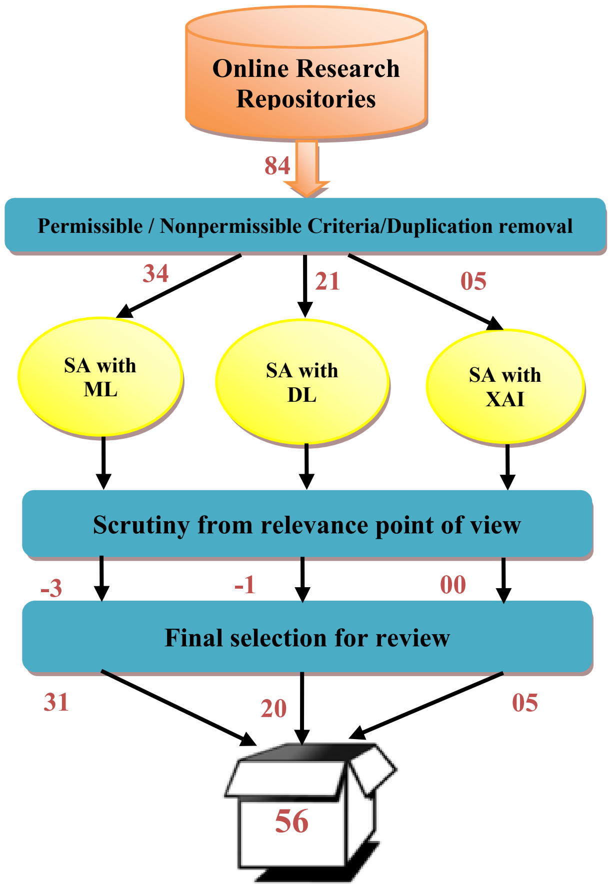


Figure 2. Article selection process

2.3 Descriptive Statistics

Finally, total 56 research articles were collectively selected from different sentiment analysis methods – ML, DL and Explainable Artificial Intelligence (XAI). Out of these, articles on ML were highest one (55%) and articles on XAI were very few (9%). Since, XAI is new emerging research area; it has less publication as compared to other methods. DL has a second largest publication (36%), which indicates that many researches are now working on DL techniques in sentiment analysis (refer Figure 3 and 4).

Research articles/papers published during year last twenty years (2003 to 2023) were selected for this study. We formed 4 groups of this duration (each of 5 years) and we found that at beginning 2003-07, very less research papers (03 no's) on review domain were published and this number grows exponentially. At the recent, during 2018-23, highest no. of research articles (44 no's) were published in our domain. (refer Figure 5).

Method wise statistics also shows somewhat similar scenario. Number of publications on ML methods for SA has exponential growth over the years (2003 – 2023). For first 10 years of this duration, there is no publication on DL methods for SA in our domain. DL methods started emerging from beginning of 2013 and have highest no. of publications (18) in recent five years. Explainable Artificial Intelligence (XAI) is still emerging area in the sentiment analysis domain (refer Figure 6).

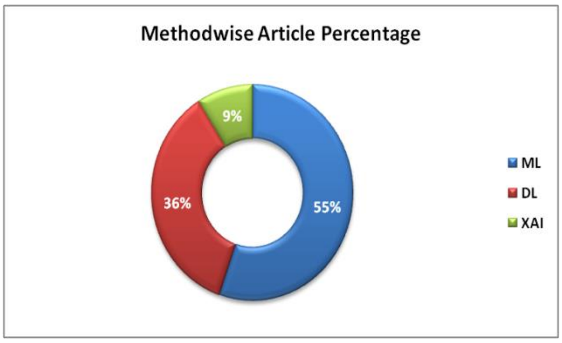


Figure 3. Methodwise Article Percentage selected for review

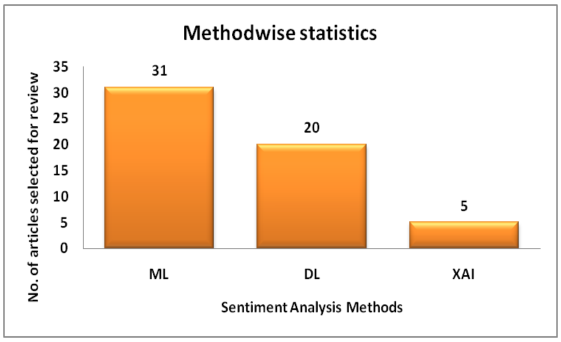


Figure 4. Methodwise No. of publications selected for review

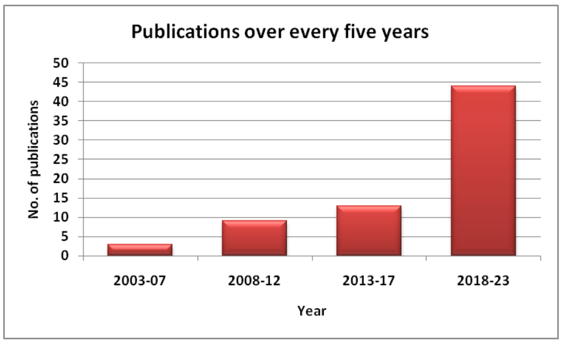


Figure 5. Number of publications every five years ML+DL+XAI (2003 – 2023)

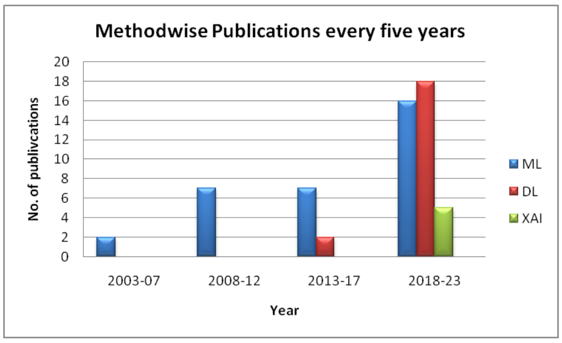


Figure 6. Methodwise Publications over every five years (2003 – 2023)

Table 1. Summary of reviewed articles of ML method

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[16] 2003	Natural Language Processing (NLP)	Topic wise feature extraction based SA with NLP	Reviews dataset :87% Web pages: 93%	Manual validation of data is required to handle semantics accurately
[17] 2005	SVM, Semantic Orientation approach	Compared SVM and SO approach for Chinese reviews.	SVM outperforms over SO with 78.87% accuracy	• Imbalanced and very small dataset Pos reviews: 400 Neg reviews: 130 • Manual data labeling required
[18] 2008	Lexical variation ontology based SVM	An ontology based SVM classifier with ability to classify reviews written using different words(synonyms) that reflects same sentiment	SVM 96% accuracy	Larger BoW degrades efficiency of SVM
[19] 2009	SA with Part of Speech (POS) tagging	A method for assigning a weightage to Parts of Speech(POS) to get higher accuracy	Suggested method improves accuracy over baseline methods.	Method not able to identify and group subjective information in topicwise groups
[20] 2009	Word Association Graph with Random Walk algorithm	A novel approach using Word Association Graph with Random Walk algorithm to classify sentiments in text	New method (SCG) performed better than SVM **SCG:** Precision:0.849 Recall: 0.865 F1 score: 0.857 **SVM:** Precision: 0.821 Recall:0.850 F1 score:0.835	Approach not able to perform well, if corpus contains more positive words at beginning and few negative words at the end where overall sentiment is negative
[3] 2010	POS tagging, C4.S, NN and Bayesian classifier(BC)	SA using ML/DL algorithms with POS tagging	**Precision:** C4.5: 90.7 **NN: 87.0** BC: 81.8 Recall: C4.5: 90.7 **NN: 88.4** BC: 86.0	Small sized dataset, Manual labeling is done only to noun and not to other POS's.

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[21] 2014	Weighted K-Nearest Neighbor (KNN) classifier	A modified KNN classifier that is able to classify weakly, mildly and high polar reviews	Not available	Algorithm is tested only on unigram data, performance on bigram, trigram is not validated.
[22] 2015	Lexicon based approach with Navy Bays (NB)	Aspect based SA using lexicons with NB classifier	Accuracy: 65 Precision: 62.5 Recall: 75 F-Measure: 68.2	Smaller sized imbalanced dataset
[23] 2016	Linear SVM, NB	Sentiment exploration using NB, LSVM and Synthetic words approach on benchmark movie review data	**Accuracy** LSVM: 75 NB: 70	Only single domain (movie) reviews are considered
[24] 2016	SVM, NB, KNN	SA of Arabic reviews with SVM, NB, KNN	**Precision**: SVM: 93.3% NB: 93.87%	• Manual annotation is required for dataset • Small sized dataset Pos reviews: 125 Neg reviews: 125
[25] 2017	Random Forest (RF)	Lexicon based SA approach with RF ensemble	Accuracy: 72%	Method not able to work on large corpus
[26] 2017	SVM, NB	A dictionary-based approach for SA using ML techniques	NB: 98.17% SVM: 93.54%	Method can't perform aspect based SA
[27] 2019	TF-IDF, SVM	Clustering based SA with SVM	Accuracy: 94.63%	Smaller and imbalanced datasets
[28] 2019	SVM	SVM based real time SA of beauty and musical instruments products from Amazon	Recall: 87.88% Precision: 99.98 F1 Score: 93.54	Model works only simple sentences

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[29] 2019	Uni-gram, Bi-gram, Tri-gram and N-gram with SVM	SA using combination of various tokenization methods with SVM	Combination of Iterated-Lovin Stemmer, and parameter C = 1.0 with SVM obtained higher accuracy: 87.70%	• Smaller dataset with manual annotation • Method can't perform well on imbalanced data
[30] 2020	NLP techniques with LR, NB, SVM and RF	Tokenization with NLP techniques and classification of sentiments with ML models - LR, NB, SVM and RF	Random Forest achieved highest accuracy : 95%	Apart from restaurant reviews model is not tested for other domain reviews
[31] 2021	TF-IDF, RF, LR, SVM, KNN, and XGBoost	SA using ML algorithms with TF-IDF vectorizer	KNN obtained highest performance Accuracy:96.25% Precision:0.96 Recall: 0.96	• Model works only on single dataset • POS tagging, synonym analysis not considered
[32] 2022	TF-IDF, Bag of Words (BoW), Vectorizer, RF and three NB algorithms: Multinomial, Complement, and Bernoulli	Featurewise SA of online products TF-IDF, BoW with RF and NB	RF achieves highest accuracy: BoW-> 83.28% TF-IDF-> 82.75%	BERT can be used to improve accuracy
[33] 2022	TF-IDF, G-gram, RF, NB, Linear Regression (LR), Decision Tree (DT)	SA of product reviews using ML algorithms with TF-IDF and N-gram	RF with TF-IDF and N-gram overperforms the others	• Model not tested for neutral polarity • There is scope to extend the work for online shopping sites
[34] 2022	BoW, Bi-gram, N-gram, TF-IDF, WORD2VEC, NB, KNN, SVM, RF	SA with ML algorithms using various feature extraction techniques	SVM overperforms other algorithms with accuracy of 94%	Work can be extended with use of CNN and visualization methods

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[35] 2022	Naïve Bayes, SVM, Decision Tree and Ensemble Classifier	SA of mobile product reviews with ML methods	Ensemble Classifier obtained highest accuracy: 97.63%	Model performance can be improved with deep learning techniques
[36] 2022	KNN, DT, SVM, RF, LR	SA of book reviews in Bangla using ML algorithms	SVM obtained highest accuracy: 94.78	• Deep learning techniques may be tested on model • Only Bangla phrases are considered
[37] 2022	TF-IDF, PCA, Adaboost, XGboost and Gradient boosting	TF-IDF and PCA based SA product reviews using chi square-based feature selection	Proposed model achieved accuracy of 98%	Work can be extended with deep learning models

Table 2. Summery of reviewed articles of DL method

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[38] 2014	PCA, BPN	Hybrid model (PCA+BPN) for SA of product reviews	Model overperformed over plain BPN with:: Precision:84.7 Recall: 85.4 F-Measure: 85.0	Smaller dataset
[39] 2019	CNN, Attention based Bi-Directional Gated Recurrent Unit (Bi-GRU)	A novel method that combines CNN+Attention Based Bi-GRU	Accuracy: 93.5% Precision:93% Recall: 93.6% F-Measure: 93.3	• Method can perform only binary classification(Positive & negative) • Can be extended to emphasize on neutral sentiment

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[40] 2020	CNN, LSTM, Attention mechanism	Attention based CNN-LSTM model that classifies book review data	Accuracy: 85% Precision:90% F-Measure: 90	Model works best only on authors dataset and it is not tested on other data
[41] 2020	CNN-LSTM, TF-IDF	SA with hybrid DL model CNN+LSTM using TF-IDF and weighted GloVe feature extraction method	Proposed method obtained accuracy of 93.85%	--
[42] 2022	SVM, LSTM, CNN, TF-IDF, GloVe	SA of product reviews with ML and DL methods	SVM: 89.44% LSTM: 91.6% CNN: 92.3%	• Work can be extended as combination of LSTM-CNN • Model can be tested on other datasets • Support for other language may be considered
[43] 2022	RNN, LSTM, fastText, CNN, BERT	SA of movie reviews data with various prevailing DL approaches	Lowest Accuracy RNN: 47.87% Highest Accuracy BERT: 91.58	--
[44] 2022	CNN+BiLSTM (BILCNN), Glove, Word2Vec	Hybrid SA approach by combining CNN and BiLSTM with Glove and Word2Vec techniques	Precision: 94.8%	--
[45]	LSTM, BiLSTM, GRU, CNN, MultiFit, XLNet, CamemBERT	Combination of different RNN's with CNN and various word embedding techniques for SA of reviews	GRU+CNN with XLNet outperformed on three different French datasets with accuracy: 92%	Work can be extended with another word embedding methods e.g. GPT and can be tested on other combination of RNN's (GRU+LSTM)

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[46] 2022	CNN, RNN	Use of CNN and RNN for SA of product reviews	Accuracy CNN: 93% RNN: 85%	Work can be extended for multi-model inputs
[47] 2022	LSVM, RF, MNB, Bernoulli NB, Logistic Regression, RNN, LSTM	ML and DL based approach for SA of product review	In ML Models RF got highest accuracy: 91.90% In DL Models RNN-LSTM got accuracy: 97.52%	Model can be extended for multimodal sentimental analysis
[48] 2018	SRN, LSTM, and CNN	Use of DL models to analyze effect of text size on performance of sentiment analysis	Three models obtained accuracy of 78.10%, 85.82% and 87.42% respectively	Model can be fine tuned by applying appropriate hyper parameter tuning for deep learning models
[49] 2019	BoW, Deep Belief Network (DBN), Association rules	A combination of association rules and deep belief network with BoW word embedding technique to identify multiple sentiments that affects customer's buying behavior	Proposed method overperformed the baseline methods with accuracy of above 86%	• Only unigrams are considered for this study. Work can be extended with N-grams • More fine tuned DL methods can adopted • Manually annotated smaller dataset
[50] 2019	CNN, Bi-LSTM, Word2Vec, GloVe, FastText, EWE	A novel approach to detect emotions from text using Bi-LSTM+CNN	Proposed method was tested on 10 datasets and compared with baseline methods and it outperformed over baseline methods.	• **Method can be extended for other tasks like affective computing/sentiment analysis** • Performance cane further improved by using larger word embeddings

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Limitations/Scope
[51] 2020	DLMNN, IANFIS	SA of products review using DMNN and IANFIS	Proposed method over performed the existing methods	• Model is able to identify the sentiment of single word only

Table 3. Summary of reviewed articles of XAI method

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Is method explainable/ interpretable?
[52] 2021	LIME, SHAP, LRP, Grad-CAM, TFIDF, Word2Vec	A novel approach for SA that autogenerate lexicons using XAI methods LIME, SHAP and Grad-CAM	Proposed method overperformed existing ones	Yes
[53] 2022	LSTM, Bidirectional LSTM, Bidirectional GRU-LSTM+CNN, LIME, SHAP	Deep learning based explainable approach using SHAP and LIME for SA of food reviews	LSTM: 96.07% Bi-LSTM: 95.85% Bi-GRU-LSTM-CNN: 96.33% SHAP and LIME revealed contribution of features in prediction	Yes
[54] 2022	LSTM, LIME	SA in Arabic language using LSTM with XAI method-LIME	LSTM obtained accuracy of 79.1% LIME showed how particular word contributed to overall sentiment analysis	Yes

Ref. # & Year of Pub.	SA Technique used	Methodology	Results	Is method explainable/ interpretable?
[55] 2022	Universal Sentence Encoder, Bidirectional Encoder Representations from Transformers, NN, LSTM, TFIDF, BoW, SHAP	SA of text with emoji using LSTM and XAI method – SHAP	LSTM obtained accuracy of 98% SHAP demonstrated model behavior	Yes
[56] 2023	Valence Aware Dictionary for Sentiment Reasoning (VADER), Bi-LSTM, LIME	NLP based explainable SA of social sites data using LIME	LIME overperformed the existing systems designed to attempt interpretability	Yes

3. RESULTS AND DISCUSSION

Research in sentiment analysis of online customer product reviews is going on over several decades and variety of ML/DL techniques have been evolved so far. Tables 1 to 3, shows summery of studied research papers for ML, DL and XAI methods. Results and discussions are presented in this section.

3.1 Sentiment Analysis Levels

Depending upon an entity considered as input for sentiment analysis, SA can be performed at 3 levels: Document, Sentence and Aspect/ Feature level.

3.1.1 Document level

This level of SA examines the whole review text and classifies it as favorable (positive) or unfavorable (negative) [57]. In this type only one product review is processed. If review text contains sentiments of more than one product, then this type fails to capture sentiments of all products. For example, a review text shown in figure 7 contains positive, negative and neutral words. But, its overall sentiment is classified as positive.

3.1.2 Sentence level

In this method, instead of analyzing whole document a single sentence analyzed to determine its polarity shade viz. positive, negative or neutral opinion [26]. Only document level analysis is not sufficient. Because, review document comprises of number of sentences with varying sentiments about the product and document level analysis is not able to focus on all sentiments in review text. Hence, sentence level analysis of review text is necessary. In following figure 7, review text two sentences out of which first have a neutral and second have a negative sentiment.

3.1.3 Aspect level

Aspect level analysis tries to find out what exactly customer likes or dislikes about particular product. This level is also knows as fine-grain sentiment analysis as it analyzes the opinion of customer itself rather than sentence or document. Aspect level analysis mainly concerns with quality, features, and drawbacks of product or service provided by seller. Figure 7 shows examples of various levels of sentiment analysis.

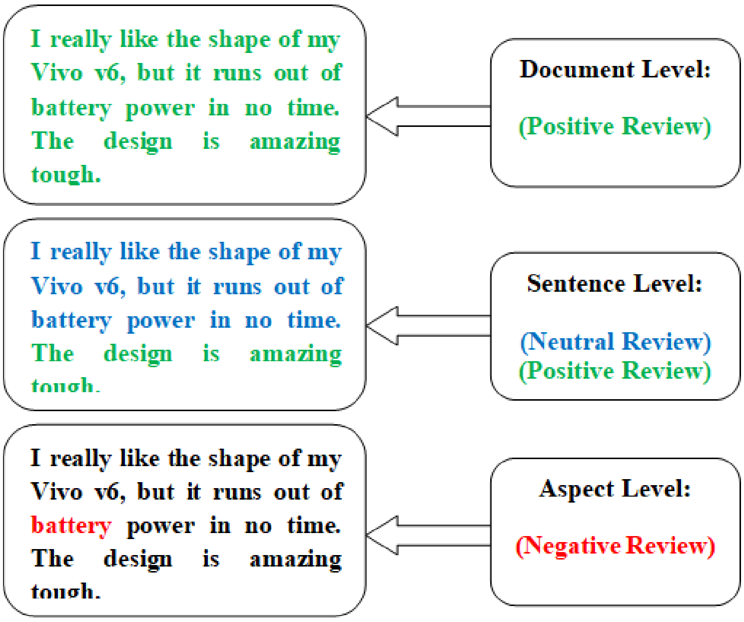
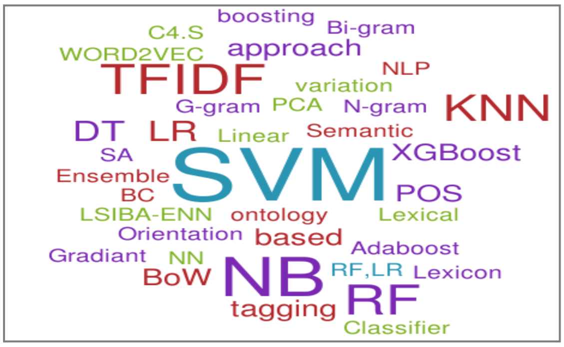


Figure 7. Levels of Sentiment Analysis (Source: [57])



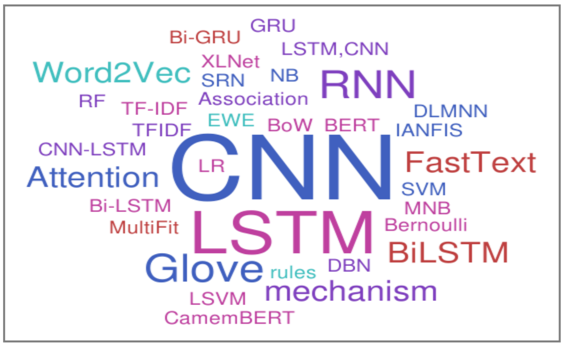


Figure 8. Word Cloud of ML methods for SA   Figure 9. Word Cloud of DL methods for SA

3.2 Sentiment Analysis Approaches

In pursuit of RQ #1 (*What are the various ML and DL methods used so far in sentiment analysis of online customer product reviews?*) mentioned in section 2.1, we reviewed several research papers summarized through tables 1 to 3 and we found various approaches/methods for sentiment analysis as shown in figure 10. Basically, sentiment analysis can be categorized into four basic categories: 1) ML Approaches 2) DL Approaches 3) Lexicon based Approaches 4) XAI based Approaches.

3.2.1 Machine Learning Techniques

Machine learning approaches haven been widely used since long time and these approaches have achieved better accuracy. Machine learning techniques for SA are grouped into two groups: supervised which need labeled data and un-supervised doesn't require any labeled data. For this review, we chose total 31 research papers on ML techniques, for duration of last 20 years (2003 – 2023). Most of the reviewed ML approaches have a smaller sized manually annotated datasets. In [17], authors compared SVM and Semantic Oriented approach for Chinese reviews and they found that SVM outperforms over SO approach with 78.87% accuracy. But dataset used in this work was smaller and it needs manual annotation. Researchers of [20] developed a novel ML method using word association graph and random walk algorithm. Proposed method outperformed the SVM. But, this approach not able to perform well, if corpus contains more positive words at beginning and few negative words at the end, where overall sentiment is negative. For example, "*This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up*" [20]. In this review clearly, sentiment is negative, but due to presence of lot of positive words at beginning of sentence, model will predict it as positive. Hence, there is need to employ some logic information inherent to sentence to improve performance of model. Some researchers [21], [24], [31], [34] and [36] applied various ML algorithms (Decision Trees, Navy Bayes, KNN, Random Forest, Linear Regression, XGBoost, Logistic regression etc.) for sentiment analysis of online produce reviews. Some of these researches stated that their model works well only on single specific dataset and it need to generalize the model so that it can work on other dataset with similar efficiency. Whereas some of these researchers concluded that performance of their models can be improved by applying deep learning techniques.

From review of ML methods, we revealed three aspects: First, most of the datasets used in ML methods for SA are imbalanced data with smaller size and needs manual annotation. Second, there is a need to make use of DL models to improve performance of these models. Third, commonly used ML methods for sentiment analysis are: Decision Trees, Support Vector Machine (SVM), Navy Bayes, Neural Network, Bayesian Network, Linear Regression, Random Forest, XGBoost and KNN. Among these methods, SVM is most popular and widely used ML approach for SA (refer Figure 8).

3.2.2 Lexicon Based Approaches

These approaches make use of data dictionaries like SentiWordNet and SenticNet. In these approaches text is divided into smaller parts called '*tokens/words*' and these tokens are tagged as either positive or negative by referring to the predefined data dictionaries. The whole polarity of the text is computed by summarizing the tagged words. This approach has two types: 1) Dictionary Based Approach which requires the collection of all synonyms & antonyms and 2) Corpus Based Approach which requires creation of massive amount of words from target language. Researchers of [22] and [25] used lexicon based approach with Navy Bayes and Random Forest respectively. In first case, due to small and imbalanced data, NB got accuracy of 65%, whereas in second case RF obtained accuracy of 72%.

3.2.3 Deep Learning Techniques

Deep learning techniques for sentiment analysis are used starting from around 2013 (refer Figure 6). From literature it is found that DL models comprises of several hidden layers and are more accurate than ML models. Some challenging aspects of these models are requirement of large datasets, time consuming and computing intensive training models etc. [53]. Authors in [39] and [53], used Bi-GRU with other DL methods like CNN, LSTM with attention mechanism and results gave an accuracy of 93.5%. In [41] and [42], authors used near about similar approaches with slight difference. In both cases, CNN outperformed. Authors of [43] and [46] applied CNN, RNN and LSTM on review data and found that RNN underperformed than other methods. It means for sequential data RNN doesn't perform well, because it cannot capture long-term memory because of the vanishing gradients problem. But when RNN is used with combination of LSTM it gives better accuracy of 97.52 % in [47]. Work in [39] showed that combination of CNN with Bi-GRU achieved accuracy of 93.5%, but this methodology can perform only binary classification and it does not considered neutral polarity. In [45], researchers used GRU with CNN and achieved high accuracy of 92%. Authors stated that work can be further extended with other feature extraction/word embedding technique like Generative Pre-training Transformer (GPT).

From reviewed research papers, we came to know that widely used DL approaches for SA are: CNN's, RNN's, LSTM's, GRU, BERT etc. Also we created word clouds with [58] for ML and DL methods used so far for sentiment analysis of online customer product reviews (refer Figure 8 and 9). Word clouds also indicated that among ML approaches, SVM is dominant one and other frequently used methods are NB, RF, KNN etc. From DL techniques, CNN is most widely used approach whereas other approaches in DL are LSTM, RNN, GRU, Bi-LSTM etc.

While exploring a literature to answer to our RQ#3 (*Is research on DL methods able to pay attention on both 'semantic' and 'emotional' information of product reviews?*), we found that all reviewed articles of DL methods from [38] to [51] except [50], used various single channel word embedding / feature selection techniques like TF-IDF, BoW, Word2Vec etc. (refer Table 5) that are capable to retain only semantic information of words and they don't pay any attention to emotional information of word. As, the work of only one researcher [50], is able to encompass both semantic as well as emotional aspects of words, we can conclude that there is scope to use multi-channel word embedding that can retain both semantic and emotional information.

3.2.4 XAI Based Techniques

ML and DL approaches for sentiment analysis have shown their excellent performance but outcomes of these models are not interpretable due to their opaque nature. To make these black-box model transparent from explainability point of view, a new research area is emerged recently called Explainable Artificial Intelligence (XAI). XAI approaches have been used over DL models in order to justify outcomes produced by DL models [54]. As, XAI is new emerging research area, we got only 05 research papers on this area for review. Usually, XAI approaches are used to explain the results/predictions of ML/DL models [53] to [56], but in [52], XAI approach is used to auto-generate lexicons and perform pseudo labeling. Most popular XAI approaches used in literature are: LIME - Local Interpretable Model Agnostic explanations, SHAP Shapley Additive exPlanations, LRP - Layer-wise Relevance Propagation and Grad-CAM - Gradient weighted Class Activation Mapping.

In order to validate RQ#4 (*Are XAI techniques able to build trust in DL methods?*), we reviewed articles [52] to [56] . In [52], Hohyun Hwang et al. have developed a novel approach for SA that autogenerate lexicons using XAI methods LIME, SHAP and Grad-CAM and their proposed system over performed existing ones. Anirban Adak et al. in [53], implemented deep learning based explainable approach using SHAP and LIME for SA of food reviews. The model performed well and XAI methods – SHAP, LIME revealed contribution of features in prediction. In [54] and [57], authors developed deep learning approaches with explainabality support using LIME and in both cases, LIME showed how particular word contributed to overall sentiment analysis. Thus, we can derive a conclusion that XAI techniques like LIME and SHAP can be used in DL models to make them more transparent and interpretable which leads to building a trust in DL models.

3.3 Datasets used in literature

In order to quench the thirst of our RQ#2 (*What are the different datasets used for sentiment analysis of online customer product reviews?*), we reviewed total 56 research articles of this domain to identify various datasets used in literature. Table 4, shows the details of reviewed datasets. Most of the datasets are created by authors using review texts publically available on various resources. Some authors have used publically available benchmark datasets for their work. Most sources for datasets are web resources like Twitter, Amazon, Kaggle, IMDB etc.

Table 4. Summary of datasets used in literature

Ref. #	**Dataset Type** **(Author Created/Benchmark)**	**Source/(s) of dataset**
[16]	Authors created	www.cnet.com www.dpreview.com www.epinions.com www.steves-digicams.com www.epinions.com
[17]	Authors created	http://www.mov8.com
[3], [26], [28], [37], [38], [45], [49]	Authors created	www.amazon.com
[41], [44], [45], [54],	Authors created	https://twitter.com/
[33], [35], [42], [44], [46]	Benchmark Dataset	https://www.kaggle.com/

Ref. #	Dataset Type (Author Created/Benchmark)	Source/(s) of dataset
[43], [48]	Benchmark Dataset	https://www.imdb.com/
[31], [36]	Authors created	https://www.daraz.com.bd/

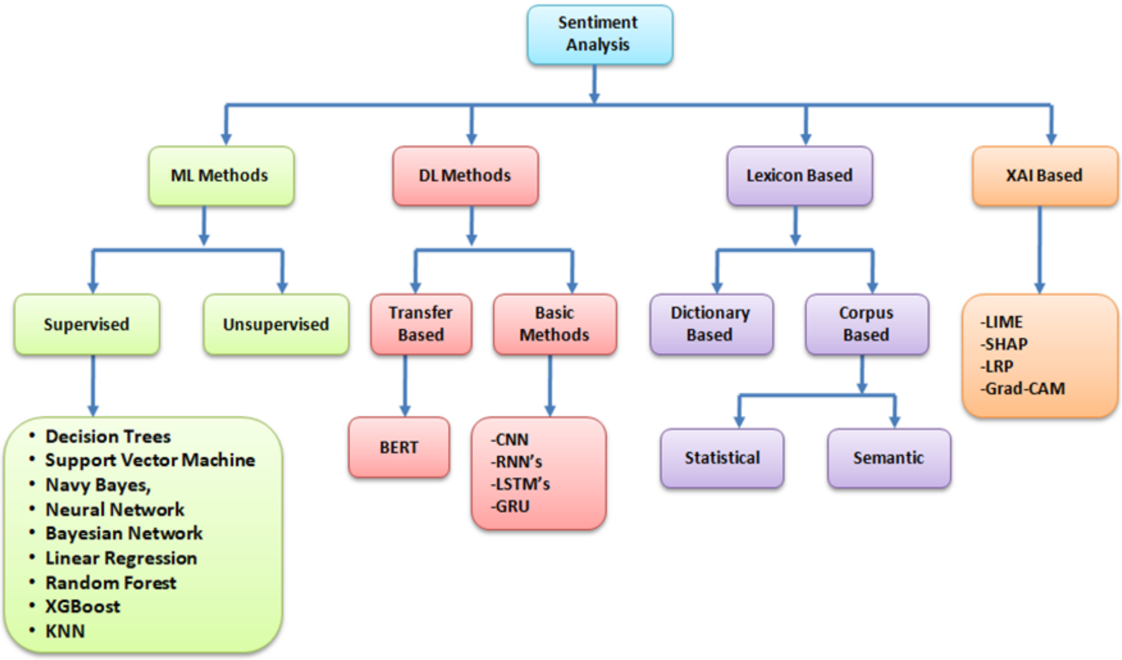


Figure 10. Sentiment Analysis Approaches

Table 5. Word embedding techniques used in literature

Word Embedding Method	Ref. #
TF-IDF	[27], [31], [32], [33], [34], [37], [41], [42],
BoW	[32], [34], [49], [55]
N-gram	[29], [33], [34]
Word2Vec	[34], [44], [50], [52]
Glove	[41], [42], [44], [50]
fastText	[43], [50]

4. KEY FINDINGS AND FUTURE SCOPE

After conducting a three step rigorous review process on total 56 research articles on ML, DL and XAI techniques, we reached at following findings and future perspectives. These findings may prove beneficial to other researchers and/or organizations working in this domain.

- Most widely used ML methods for sentiment analysis of online product reviews are - Decision Trees, Support Vector Machine (SVM), Navy Bayes, Neural Network, Bayesian Network, Linear Regression, Random Forest, XGBoost and KNN. Among these methods, SVM is most popular and widely used ML approach for SA.
- ML models lag in accuracy as compared to DL models. There is a need to make use of DL models to improve performance.
- Most of the datasets used in ML/DL methods for SA are imbalanced, having a smaller size and they need manual annotation. Research can be augmented further to auto-annotate the datasets.
- Majority of the existing research emphasizes only on binary classification (positive or negative) and ignored neutral sentiment in reviews. Classifying a neutral polarity reviews can be explored in future.
- Widely used DL approaches for SA are: CNN's, RNN's, LSTM's, GRU, BERT etc. CNN is the most commonly used approach. Research on sentiment analysis of online product reviews can be extended using hybrid DL models to achieve more accurate results.
- Most of the existing research (except few ones) on DL models for sentiment analysis of online product reviews focuses on only semantic angle of tokens. Hence, there is a need to focus on both semantic and emotional angle of the words/tokens, using novel word embedding techniques.
- DL models are more accurate than ML models, but they are opaque in nature. Therefore, interpretability/explainability support can be provided to these models using XAI techniques like SHAP and LIME.

5. CONCLUSION

This study reviewed and discussed previous research work of ML, DL and XAI methods for sentiment analysis of online customer product reviews. Study highlighted and identified the various sentiment analysis levels, approaches, datasets and feature extraction techniques used in past work. The work unveil that among ML approaches Support Vector Machine is most popular and in DL methods CNN is most widely used approach. Further research in DL approaches can be augmented by applying hybrid DL models with novel feature extraction methods. DL techniques are more accurate but they need explainability support. XAI approaches like LIME and SHAP can be used to make DL models more interpretable/explainable.

REFERENCES

[1] oberlo. www.oberlo.in. [Online]. https://www.oberlo.in/statistics/how-many-people-shop-online

[2] Khin Phyu Phyu, Thi Thi Soe Nyunt Shein, "Sentiment Classification based on Ontology and SVM Classifier," in *Second International Conference on Communication Software and Networks*, 2010, pp. 169 - 172.

[3] Patrawadee Tanawongsuwan, "Product Review Sentiment Classification using Parts of Speech-A Case Study of Textbook Reviews," , 2010, pp. 424 - 427.

[4] R.Sabnis, O.V. Bhattacharyya, S. Lokeshkumar, "A Novel Approach to Extract and Analyse Trending Cuisines on Social Media," *In Lecture Notes on Data Engineering and Communications Technologies; Springer: Cham, Switzerland*, pp. 645 - 656, 2020.

[5] R. Akila, S. Revathi, and G. Shreedevi, "Mining on Food Services Using Topic Modeling and Machine Learning Algorithms," in *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems*, Coimbatore, India, 2020, pp. 1071–1076.

[6] LI Ya'nan, WANG Shuhua HONG Lan, "Improvement of Online Food Delivery Service Based on Consumers' Negative," *Canadian Social Science*, vol. 12, no. 5, pp. 84-88, Dec. 2016.

[7] Wen Shi, Yijun Li Qiang Ye, "Sentiment Classification for Movie Reviews in Chinese by Improved Semantic Oriented Approach," in *International Conference on System Sciences*, Hawaii, 2006, pp. 2 - 5.

[8] Z.K. Chepukaka and F.K. Kirugi, Service Quality and Customer Satisfaction at Kenya National Archives and Documentation Service, 2019.

[9] ZHANG Zi-qiong, LI Yi-jun, YE Qiang1, LAW Rob, "Sentiment Classification for Chinese Product Reviews Using an Unsupervised Internet-based Method," in *International Conference on Management Science & Engineering*, Long Beach, USA, 2008 , pp. 3 - 9.

[10] "A literature survey of sentiment analysis based on Ecommerce," in *Proceedings of the Fifth International Conference on Computing Methodologies and Communication*, 2021, pp. 1767 - 1772.

[11] Rajendra Pamula Praphula Kumar Jain. (2020, Aug.) https://arxiv.org/. [Online]. https://arxiv.org/abs/2008.10282

[12] Enas.M.F.El Houby, Hoda Korashy Mohamed Enas A.Hakim.Khalil, "Deep Learning Approach in Sentiment Analysis: A Review," in *15th International Conference on Computer Engineering and Systems*, 2020.

[13] Ramesh Kumar Ayyasamy, Rehan Akbar Hoong-Cheng Soong, "A Review Towards Deep Learning for Sentiment Analysis," in *International Conference on Computer & Information Sciences*, Universiti Teknologi PETRONAS, 2021, pp. 238 - 243.

[14] Sonali Sharma et. al., "A Critical Review on Sentiment Analysis Techniques," in *3rd International Conference on Intelligent Engineering and Management*, 2022, pp. 741 - 746.

[15] David Denyer, Palminder Smart David Tranfield, "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review," *British Journal of Managemen*, vol. 14, pp. 207-222, 2003.

[16] Tetsuya Nasukawa, Razvan Bunescu, Wayne Niblack Jeonghee Y, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," in *IEEE Computer Society, Third IEEE International Conference on Data Mining*, 2003.

[17] BIN LIN, YI-JUN LI QIANG YE, "SENTIMENT CLASSIFICATION FOR CHINESE REVIEWS: A COMPARISON BETWEEN SVM AND SEMANTIC APPROACHES," in *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, 2005, pp. 2341 - 2346.

[18] Jantima Polpinij and Aditya K. Ghose, "An Ontology-based Sentiment Classification Methodology for Online Consumer Reviews," in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2008, pp. 514 - 524.

[19] FEI SONG CHRIS NICHOLLS, "IMPROVING SENTIMENT ANALYSIS WITH PART-OF-SPEECH WEIGHTING," in *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics*, Baoding, 2009, pp. 1592 - 1597.

[20] Cheng Mingzhi et. al., "A Random Walk Method for Sentiment Classification," in *Second International Conference on Future Information Technology and Management Engineering*, 2009, pp. 327 - 330.

[21] Dr. M. P. Singh, Prabhat Kumar Ankita Shrivastav, "Supervised Semantic Analysis of Product Reviews Using Weighted k-NN Classifier," in *11th International Conferenceon Information Technology*, 2014, pp. 502 - 507.

[22] S.K. Padma Yashaswini Hegde, "Sentiment Analysis for Kannada using Mobile Product Reviews A Case Study," in *IEEE International Advance Computing Conference*, 2015, pp. 822 - 827.

[23] "Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques," in *2nd International*

Conference on Next Generation Computing Technologies, Dehradun, India, 2016, pp. 106 - 111.

[24] Mounir Zrigui Mohamed Ali Sghaier, "Sentiment Analysis for Arabic e-commerce websites," in *978-1-5090-5579-1/16 ©2016 IEEE*, 2016, pp. - 7.

[25] S. K. Padma Yashaswini Hegde, "Sentiment Analysis using Random Forest Ensembler for Mobile Product Reviews in Kannada," in *IEEE 7th International Advanced Computing Conference*, 2017, pp. 777 - 782.

[26] Vishal S. Shirsat and Sachin N. Deshmukh Rajkumar S. Jagdale, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques," *Mallick, P., Balas, V., Bhoi, A., Zobaa, A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_61*, vol. 768, pp. 639 - 647, January 2017.

[27] Aaditya Jain Tulika Dubey, "Sentiment Analysis of Keenly Intellective Smart Phone Product Review Utilizing SVM Classification Technique," in *10th ICCCNT 2019*, IIT - Kanpur, 2019, pp. 1 - 8.

[28] Iqra Urooj, Wu JunSheng and Naqash Azeem Jahanzeb Jabbar, "Real-time Sentiment Analysis On E-Commerce Application," in *IEEE 16th International Conference on Networking, Sensing and Control*, Banff, Alberta, Canada, 2019, pp. 391 - 396.

[29] Adhistya Erna Permanasari and Nining Woro Wijayanti Indriana Hidayah, "Sentiment Analysis on Product Review using Support Vector Machine (SVM)," in *5th International Conference on Science and Technology (ICST)*, Yogyakarta, Indonesia, 2019, pp. 1 - 4.

[30] Narmeen Zakaria Bawany, Soomaiya Hamid Kanwal Zahoor, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," in *21st International Arab Conference on Information Technology*, 2020.

[31] Manoara Begum, Rashed Mustafa Mst. Tuhin Akter, "Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors," in *International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, Dhaka, Bangladesh, 2021, pp. 40 - 44.

[32] U. A. C. Senarathne, N. H. A. C. Madhubhashani, T. C. Liyanage, P. P. G. Dinesh Asanka V. R. Welgamage, "Overall and Feature Level Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques and Web-Based Chrome Plugin," in *International Research Conference on Smart Computing and Systems Engineering (SCSE)*, University of Kelaniya, Sri Lanka, 2022, pp. 205 - 210.

[33] Preethi Salian K, Puneeth B. R, Tanzila Nargis, Supriya Salian Prathyakshini, "A Novel Approach for Classification of Online Product Reviews using various Machine Learning Techniques," in *Proceedings of the Sixth International Conference on Electronics, Communication and Aerospace Technology (ICECA 2022)*, 2022, pp. 878 - 884.

[34] T. Sasikala, S. Santhanalakshmi Sruthi Yarkareddy, "Sentiment Analysis Of Amazon Fine Food Reviews," in *Proceedings of the Fourth International Conference on Smart Systems and Inventive Technology (ICSSIT-2022)*, 2022, pp. 1242 - 1247.

[35] Vibha Pratap Shikha Maurya, "Sentiment Analysis on Amazon Product Reviews," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, 2022, pp. 236 - 240.

[36] Md Forhad Hossain, Taufiq Ahmed, Ananyna Devanath, Shahnaz Akter and Abdus Sattar Rely Das, "Classification of Product Review Sentiment by NLP and Machine Learning ," in *Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2022, pp. 1 - 7.

[37] Dilip Kumar Shaw Joy Gorai, "Two-phase feature extraction with chi square-based feature selection approach for the classification of online product reviews," in *Proceedings of the Seventh International Conference on Communication and Electronics Systems (ICCES 2022)*, 2022, pp. 1433 - 1438.

[38] RM.Chandrasekaran G.Vinodhini, "Sentiment Classification Using Principal Component Analysis Based Neural Network Model," in *ICICES2014*, Chennai, 2014, pp. 1 - 6.

[39] YING LI, JIN WANG and R. SIMON SHERRATT LI YANG, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522 - 23530, February 2019.

[40] Arunavo Dey, "Attention Based LSTM CNN Framework for Sentiment Extraction from Bengali Texts," in *International Conference on Electrical and Computer Engineering (ICECE)*, 2020, pp. 226 - 230.

[41] Onan A., "Sentiment analysis on product reviews based onweighted word embeddings and deep neural networks," *Concurrency and Computation Practice and Experience*, no. e5909. https://doi.org/10.1002/cpe.5909, pp. 1 - 12, June 2020.

[42] Pallak Srivastava and Ananda Kumar S Lakshay Arora, "EVALUATION OF PRODUCT REVIEWS USING DEEP LEARNING CLASSIFIER MODELS," in *13th International Conference on Computing Communication and Networking Technologies, DOI: 10.1109/ICCCNT54827.2022.9984463*, 2022.

[43] Alok Misra and Neha Ujjwal Raj Gaurang Tiwari, "Comparative Classification Performance Evaluation of Various Deep Learning Techniques for Sentiment Analysis," in *8th International Conference on Signal Processing and Communication (ICSC)*, 2022, pp. 304 - 309.

[44] Manikandan. B and Dr. Chakaravarthi. S, "An Efficient Neural Network Model for Sentiment Analysis using Bi-LSTM and CNN," in *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2022, pp. 1 - 7.

[45] H.ANOUN, and L. HASSOUNI N. HABBAT, "Combination of GRU and CNN deep learning models for sentiment analysis on French customer reviews using XLNet model," in *Article accepted for publication in IEEE Engineering Management Review, DOI 10.1109/EMR.2022.3208818*, 2022, pp. 1 - 9.

[46] G. Vijendar Reddy, K. Swapnika, Roshini Akunuri, Harshmeet Kaur Jahagirdar Y. Sri Lalitha, "Analysis of Customer Reviews using Deep Neural Network," in *First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, 2022.

[47] Tanjim Tabassum, Mir Md. Mabrur, Keping Yu, Mohammad Arifuzzaman Joy Chandra Gope, "Sentiment Analysis of Amazon Product Reviews Using Machine Learning and Deep Learning Models," in *International Conference on Advancement in Electrical and Electronic Engineering*, Gazipur, Bangladesh, 2022.

[48] L., Goh, TT. & Jin, D. i, "How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis," *Neural Comput & Applic 32, 4387–4415 (2020). https://doi.org/10.1007/s00521-018-3865-7*, pp. 1 - 29, November 2018.

[49] J.W. Wang, Z. Li, Z.G. Tian, Eric Tsui W.M. Wang, "Multiple affective attribute classification of online customer product reviews:A heuristic deep learning method for supporting Kansei engineering," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 33 - 45, October 2019.

[50] MEIJING LI AND KEUN HO RYU ERDENEBILEG BATBAATAR, "Semantic-Emotion Neural Network forEmotion Recognition From Text," *IEEE Access, DOI: 10.1109/ACCESS.2019.2934529*, vol. 7, pp. 111866 - 111879, Auguat 2019.

[51] P., Mary Immaculate Sheela, L. Sasikala, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS," *Journal of Big Data*, vol. 7, no. 33, pp. 1 - 20, May 2020.

[52] Younghoon Lee Hohyun Hwang, "Semi-Supervised Learning based on Auto-generated Lexicon using XAI in Sentiment Analysis," in *Proceedings of Recent Advances in Natural Language Processing, DOI: https://doi.org/10.26615/978-954-452-072-4_067*, 2021, pp. 593–600.

[53] Biswajeet Pradhan, Nagesh Shukla and Abdullah Alamri Anirban Adak, "Unboxing Deep Learning Model of Food Delivery Service Reviews Using Explainable Artificial Intelligence (XAI) Technique," *MDPI Foods Journal, https://doi.org/10.3390/foods11142019*, vol. 11, pp. 1 - 20, July 2022.

[54] Mohamed Kholief and Ahmed Ahmed Hesham Sedky Youmna Abdelwahab, "Justifying Arabic Text Sentiment Analysis Using Explainable AI (XAI): LASIK Surgeries Case Study," *MDPI's Journal of Information, DOI: https://doi.org/10.3390/info13110536*, vol. 13, no. 535, pp. 1 - 12, November 2022.

[55] Chandrashekar Muniyappa, and Ashutosh Saxena Sirisha Velampalli, "Performance Evaluation of Sentiment Analysis on Text and Emoji Data Using End-to-End, Transfer Learning, Distributed and Explainable AI Models," *Journal of Advances in Information Technology*, vol. 13, no. 2, pp. 167 - 172, April 2022.

[56] Ashish Kumar, Anand Nayyar, Kritika Dewan, Rishika Garg Rachna Jain, "Explaining sentiment analysis results on social media texts through visualization," *Springer Journal Multimedia Tools and Applications, DOI: https://doi.org/10.1007/s11042-023-14432-y*, February 2023.

[57] Dr Naveen Kumar Gondhi Harsheta Pandita, "A literature survey of sentiment analysis based on Ecommerce reviews," in *Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)*, 2021, pp. 1767 - 1772.

[58] [Online]. https://worditout.com/

[59] Zhenghong Liu b, Xuemei Yao c, Qin Yang Huiliang Zhao, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing and Management*, vol. 58, no. 5, pp. 1 - 14, June 2021.