scientific reports

OPEN

A modernized approach to sentiment analysis of product reviews using BiGRU and RNN based LSTM deep learning models

L. Godlin Atlas[1], Daniel Arockiam[2], Arvindhan Muthusamy[2], Balamurugan Balusamy[3], Shitharth Selvarajan[4,7,8,9]✉, Taher Al-Shehari[5] & Nasser A. Alsadhan[6]

With the advent of Web 2.0 and popularization of online shopping applications, there has been a huge upsurge of user generated content in recent times. Leading companies and top brands are trying to exploit this data and analyze the market demands and reach of their products among consumers using opinion mining. Sentiment analysis is a hot topic of research in the e-commerce industry. This paper proposes such a novel sentence level sentiment analysis approach for mining online product reviews using natural language processing and deep learning techniques. The proposed model consists of various stages like web crawling and collecting product reviews, preprocessing, feature extraction, sentiment analysis and polarity classification. The input reviews are preprocessed using natural language processing techniques like tokenization, lemmatization, stop word removal, named entity recognition and part of speech tagging. Feature extraction is done using bidirectional gated recurrent unit shortly called as BiGRU feature extractor and the sentiments are classified into three polarities such as positive, negative and neutral using a hybrid recurrent neural network based long short-term memory classifier. The specific combination of techniques employed here and applying it to a new kind of online product review is making the proposed model to be novel. Performance evaluation metrics such as accuracy, precision, recall, F measure and AUC are calculated for the proposed model and compared with many existing techniques like deep convolutional neural network, multilayer perceptron, CapsuleNet and generative adversarial networks. The proposed model can be used in a variety of applications like market research, social network mining, recommendation systems, brand analysis, product quality management etc. and is found to generate promising results when compared to prevailing models.

Keywords Sentiment analysis, Product reviews, Deep learning, LSTM, BiGRU, Natural language processing

Sentiment analysis is an especially important field of natural language processing which analyzes the sentiment or emotion that underlies a text[1]. It has been found that sentiment analysis plays a very vital role in making critical business decisions such as launching new products, making changes to a product based on the customers' opinions or even withdrawing them from the market[2]. With the evolution of the Internet and giant growth of information and technology, this concept of sentiment analysis has been made possible. In the initial days when the Internet was first launched it was more of a read only basis. But now having evolved over the years, the Internet is used for both reading and writing purposes. Millions of people all around the world have the freedom to express their thoughts, views, opinions, feedback and share their personal experience in various applications like Facebook, Twitter, Instagram, Dimity, Tumblr etc[3].

[1]Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India. [2]ASET-CSE, Amity University, Gwalior, Madhya Pradesh, India. [3]Shiv Nadar University, Noida, India. [4]Department of Computer Science, Kebri Dehar University, Kebri Dehar, Ethiopia. [5]Computer Skills, Department of Self-Development Skill, Common First Year Deanship, King Saud University, 11362 Riyadh, Saudi Arabia. [6]Computer Science Department, College of Computer and Information Sciences, King Saud University, 12372 Riyadh, Saudi Arabia. [7]Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai, India. [8]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, Rajpura 140401, India. [9]School of Built Environment, Engineering and Computing, Leeds Beckett University, LS6 3HF Leeds, UK. ✉email: shitharths@kdu.edu.et

Today's web is a major source of unstructured data that contains a plethora of information. With the introduction of ecommerce applications, online shopping has gained much popularity[4]. There are numerous websites and online applications available such as Flipkart, Amazon, Myntra, Ajio, ShopClues that sell a range of products from books, gadgets, accessories, groceries and what not. An additional advantage is that all of these websites provide an opportunity for their customers to share both positive and negative feedback about the product that they have bought[5]. The reviews and feedback shared by one customer serves as an input for another customer and influence his decision in many ways. Hence business firms are trying to exploit this user generated information about the product and use it for the betterment of the organization[6]. Sometimes the buyer itself conducts a survey which records the responses and insights of its stakeholders to know the buying interests of the customer in a better manner and enable improvisations. Comments from respective customers serve as a major source of information about the market demand and product trend and hence are considered very valuable.

There are several sources that generate personalized user data such as review sites like epinion, Zdnet, IMDb, CNET, consumer reviews etc.[7]. Reviews are available regarding products, shopping experience, books, tourist spots, movies, food, hotel service, hiring companies, banks, colleges, schools, airlines, transport and several other micro services. It has now become a practice to refer to the Internet and study the reviews of an organization or service before trying it[8]. The central idea behind social networking sites is to make the Internet a user friendly one providing opportunity for the common man to raise his voice for or against anything. The success of initial social media networks such as Orkut has paved way for the explosion of social media today which allows many people to participate in online forums, discussions and debates and anticipate their feelings and attitudes[9].

In the olden days when television, radios, newspapers, and magazines were the only form of mass media, advertisements played an especially important role in business decision making. But nowadays economists agree to the fact that personalized reviews act as a standard or reference for purchasers[10]. It is said that these reviews act as a good source of information for both the sellers and buyers. The producers are able to analyze their marketing strategy without relying upon a third person as they are able to get access directly to the comments made by their customers[11]. It also gives a fair chance to the consumers to communicate their opinions, appreciations, and dissatisfaction about the product to the manufacturers directly so that appropriate corrective measures can be taken. Online shopping has become the norm of the day, which is beneficial for both parties with extra offers and great deals. Offline shopping is very much confined in nature and incurs additional cost both to the producer and the consumer[12].

Existing techniques suffer from the fact that many of them rely upon lexicon-based analysis of sentiment to some extent. Lexicon based analysis depends heavily upon words present in the dictionary. But not always will the customer express his views in a direct manner. Sometimes, they may also register a comment in a sarcastic manner. This will lead to misclassification of review polarity. But the proposed system will be able to identify the exact polarity of the review comment even though it has been written sarcastically with the help of LSTM classifier. Our proposed methodology also perform real time analysis of product reviews and showcases the actual process of execution.

Types and levels of sentiment analysis

Sentiment analysis is a fast-growing field that is much appreciated because of its various applications in business-oriented research. It is also a very hot topic of research that has information related to many other business domains. It can be performed in two ways based on lexicon and machine learning techniques[13]. Lexicon based sentiment analysis can be split into two types called dictionary based and corpus-based sentiment analysis. Dictionary based sentiment analysis is performed based on words present in the dictionary whereas corpus-based sentiment analysis is carried out using words in a statistical and semantic manner. Machine learning approach towards sentiment analysis makes use of a labeled data set which is employed for training the classifiers and testing them thereafter[14]. There is another type of sentiment analysis called the hybrid method which combines both the properties of lexicon based and machine learning based sentiment analysis. Figure 1 below depicts the types of sentiment analysis.

There are three levels of sentiment analysis called the aspect level or feature level sentiment analysis, sentence level sentiment analysis and document level sentiment analysis[15]. Aspect level sentiment analysis is exceptionally fine grained in nature and contains reviews about a particular aspect of a product. For example, the quality of camera of a mobile phone, cooling properties of an air conditioner, mileage provided by an automobile etc. Instead of describing the product as a whole, aspect level reviews contain feedback and opinions about a particular feature of a product[16]. Sentence level sentiment analysis works on a few sentences that describe a product either in a positive manner or in a negative manner.

This is the most widely adopted level of sentiment analysis because product reviews usually contain only a few or more sentences[17]. Document level sentiment analysis, otherwise called text level sentiment analysis, will concentrate on the entire document that contains a review and gives a single classification for the entire document. This is employed particularly in discussions, debates and blogs related to political scenarios, current affairs, or public happenings[18]. Figure 2 illustrates the levels of sentiment analysis.

Sentiment analysis provides an edge for business because of which it has been adopted by retail giants like Amazon and Taobao. The summarized research work contributions are given as follows.

1. To conduct the background study and literature review on sentiment analysis of product reviews using various technologies and compare them.
2. A novel hybrid deep learning model is presented to classify polarity of sentiment using bidirectional gated recurrent unit feature extractor and RNN based long short-term memory classifier.

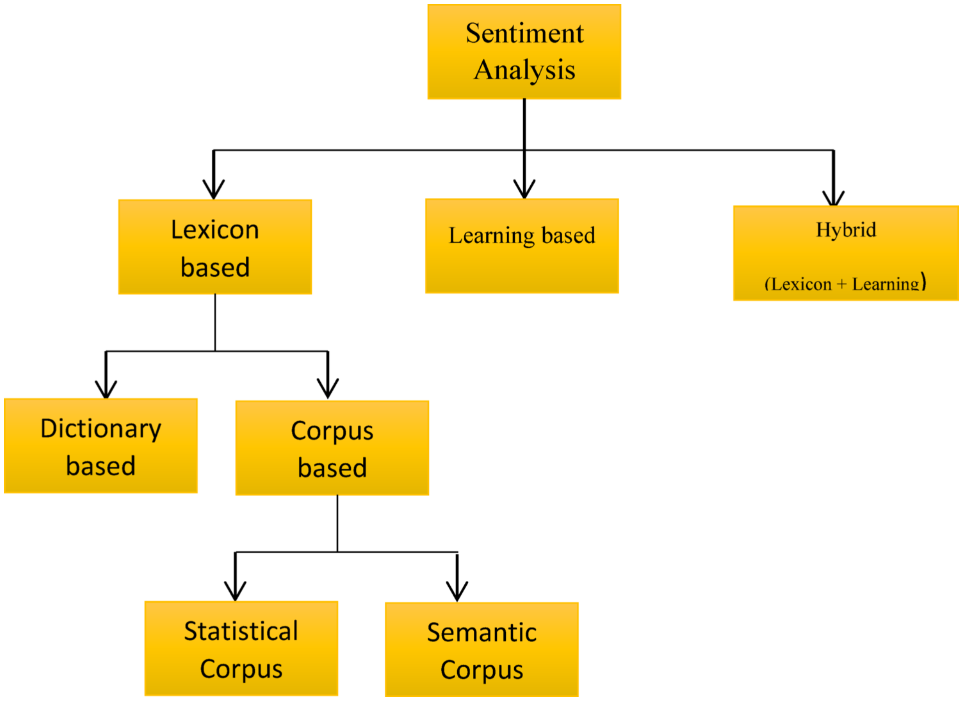

Fig. 1. Types of sentiment analysis.

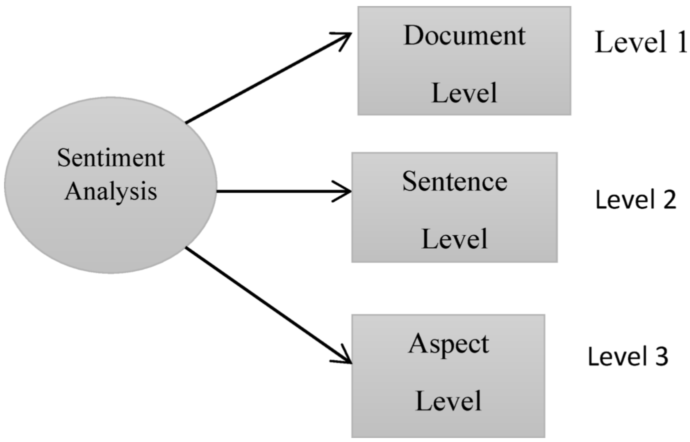

Fig. 2. Levels of sentiment analysis.

3. Experimental analysis using Amazon review scraper tool validates the multi-class sentiment classification performance of proposed model.
4. Comparative analysis with existing classifiers deep convolutional neural network, multilayer perceptron, Capsule Net, and generative adversarial networks are considered to compare with proposed model. Figure 3 illustrates the frame diagram Taxonomy of Sentiment Analysis System. Emotion/Sentiment Intensity his dimension considers the strength or degree of the expressed sentiment, ranging from weak to strong. Discusses intensity ranking and the subjective nature of emotions and sentiments.

Mainly this paper answers the question How can a novel sentence-level sentiment analysis approach be developed for mining online product reviews using natural language processing and deep learning techniques and objective of this research is to propose a novel sentence-level sentiment analysis approach for mining. This approach utilizes natural language processing and deep learning techniques.

This paper is arranged in the following order. Relevant works on sentiment classification are discussed in "Literature survey" section. Materials and methods used for the proposed model are presented in "Materials and methods" section. A detailed description of the proposed model is presented in "Proposed system" section. Experimentation, evaluation results, and discussion are presented in "Experimentation, results and analysis" section. In "Challenges and future directions" section, the final summary is presented with future scope.

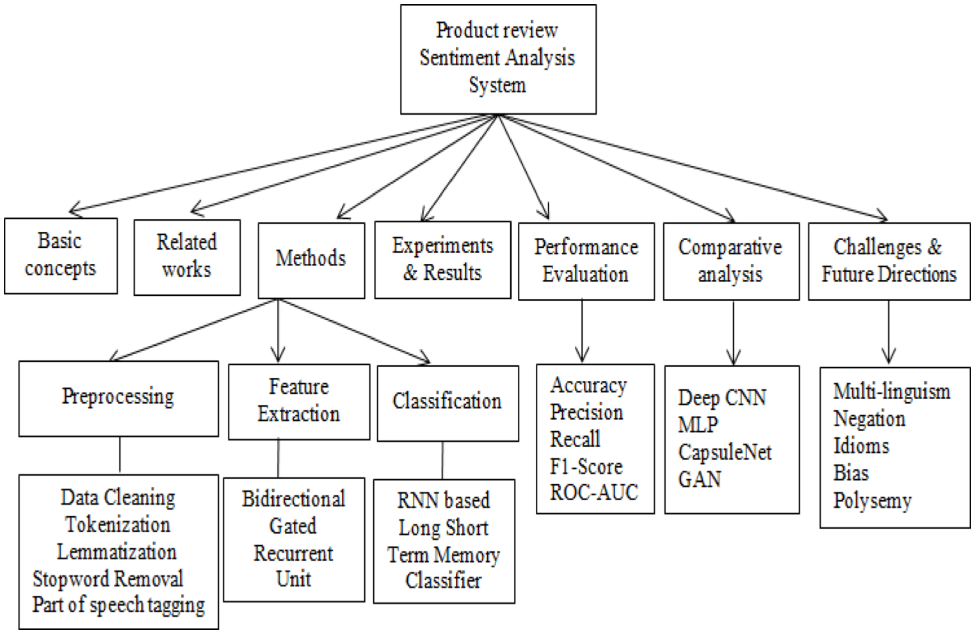

Fig. 3. Taxonomy of sentiment analysis system.

Literature survey

In One particularly promising method mentioned in the manuscript[14], it has used deep learning for performing sentiment analysis on Twitter product reviews. They utilized 93,000 tweets (50,000 positive tweets and 43,000 negative tweets) in the language of English concerning various products technical and non-technical. He described that earlier methods used techniques like word ranking, bag of words etc. which had many problems regarding dimensions and feature vectors. Hence, the work has avoided traditional techniques like those and employed neural language models such as word2vec model, glove2vec model, fastText model, LDA2vec model and doc2vec model. All of these models are based on artificial neural networks and word embedding schemes. A tool called Twitter 4J collected the input data and automatic filtering was done to remove repetitive tweets. Functions such as weighting and aggregating functions were used. Finally, a classifier based on convolutional neural networks and long short-term memory was used to classify the tweets and find their polarity yielding an accuracy of 93.85%. It also used Adam optimizer in addition to the classifier.

Jagdale et al.[19] have used the standard Amazon data set containing reviews related to cameras, laptops, TVs, mobiles, and technical gadgets. The work used preprocessing techniques like tokenization, stop word and punctuation removal, stemming etcetera and employed Naives bayes and support vector machine classifiers and generated scores for every sentence of the review. Naives bayes achieved an accuracy of 98.1% while support vector machine received a low accuracy of 93.54%.

Sentiment analysis was performed to support marketing decisions and product improvisation in[20]. 772 reviews were extracted from FemaleDaily website, half of which was positive and half of which was negative written in the Indonesian language. The work makes use of word embedding techniques such as term frequency and term frequency inverse document frequency. Preprocessing techniques such as tokenization, case reversal, data cleaning and removing URLs were performed on the input data and word2vec model was used as a feature extractor including bag of words, raw term frequency and binary term frequency features. Support vector machine was used as classifier which attained an accuracy of 84%.

Dang et al.[21] have stated that Web 2.0 has many related and dependent data that keeps growing day by day. The work employed a GeForce GTX 2070 GPU model and libraries such as Keras and TensorFlow. Input reviews were obtained from 8 datasets including sentiment140, Cornell book review datasets, IMDb movie reviews etc. possessing a batch size of 4096. They have compared the classification performance of three major techniques such as convolutional neural networks, recurrent neural networks, and deep neural networks. Before classification, preprocessing such as stemming, lemmatization, case conversion, stop word removal, space removal, tag omission etc. were performed. Among the three classifiers it has been concluded that recurrent neural networks performs better than the other two competitors.

In[22], scholars have performed sentiment analysis on reviews regarding airline services enclosing 186 different airlines of 85 countries totaling up to 14,640 tweets from two datasets namely Twitter airline and airline equality. Preprocessing such as tokenization, decoding hypertext markup language was performed, and convolutional neural network based long short-term memory classifier was used for the purpose of classification of polarity. The proposed system was compared with existing techniques such as Logistic regression, Naives bayes, Decision tree, Support vector machine etc. The proposed classifier achieved an accuracy of 91.3%, precision score of 87.8%, recall of 87% and F measure of 87.5%.

In the research paper[23] discusses about the various challenges that confront the field of sentiment analysis and have proposed novel techniques such as Bidirectional Encoder Representations from Transformers (BERT) combined with bidirectional gated recurrent units along with softmax layer. Preprocessing techniques such as padding and pooling were performed and features such as term frequency phrases related to opinions were extracted. Two data sets of COAE2014-task4 and chnSenticorp-HTL-ba-6000 containing several mobile reviews regarding Taobao and sunning brands were employed for executing the proposed model where the dimension of the extracted feature vectors was of the size 768. With the chosen epoch size of 10 and Adam optimizer, the proposed novel classification system achieved an accuracy of 95.5%

A survey of various deep learning technologies that have been used for sentiment analysis has been presented in[24]. The work has divided sentiment analysis into two types based on reviews of service and networks. Service reviews contain opinions about particular product brands whereas network opinion contain debates, discussions regarding political happenings. The study suggest the use of deep learning models for the purpose of sentiment classification when compared to other techniques such as machine learning or neural network models. This is because of the fact that deep learning models have the ability to extract the needed features automatically and also possess more than one hidden layers which represent more functional data than other classifiers.

Ombabi et al.[25] classified 15,100 reviews regarding hotels and movies into positive, negative, or neutral sentiments using deep convolutional neural network combined with support vector machine and long short-term memory classifiers. The hybrid classifier achieves an accuracy of 90.75% after preprocessing the dataset using techniques like stop word removal, case conversions, lemmatization, and tokenization. Feature extraction was done using word embedding schemes such as skipgram and fast text models. It is to be noted that this sentiment classification proposal was executed in the language of Arabic.

The research paper[26] performed sentiment analysis of product reviews and also predicted the future trends of products which can hit the market. This is done with the help of an improved adaptive neuro fuzzy inference system along with deep learning modified neural network classifier. A famous food review data set was chosen for implementing the proposed system and the reviews were divided in the ratio of 80:20 for training the classifier and then consequently testing it. The training data set was classified into three types such as content-based reviews, grade-based reviews, and collaboration-based reviews before further processing. Common preprocessing techniques like spelling corrections, tokenization, word chunking were performed, and SMO algorithm was used as a feature extractor. Keyword frequency and entropy were the major features chosen for classification and the final accuracy obtained was 96.66%.

In Research paper presented in[27], sentiment analysis has been carried out in the Chinese language containing 1,00,000 reviews from Dangdang book site. A robust classifier called as lexicon-based sentiment classifier along with attention based gated recurrent unit and convolutional neural networks shortly called as SLCABG has been proposed and compared with the performance of prevailing techniques such as Naive Bayes, Support vector machine and bidirectional gated recurrent units. The implementation was carried out on Python platform with Jieba package and stop word removal as preprocessing scheme. The proposed model arrives at a final score of 93.5% accuracy, recall of 93.6%, F-measure of 93.3% and precision of 93% indicating better performance than its predecessors.

Zhu et al.[28] presented a review regarding opinion mining[29] and its related techniques. For performing the review, the study has used 31,000 reviews from datasets such as IMDb, Yelp 2013 and TSB utilizing bidirectional gated recurrent unit-based hybrid convolutional neural network classifier. The hyper parameters used were epoch size of 15, batch size of 128, leaky activation function which achieved an accuracy of 90.30%, precision of 90.34%, recall of 90.32% and F-measure also of 90.3%.

A massive 1,93,000 reviews from websites such as Amazon, Yelp, Yi-Jen Mon was used in[30] and filtered only reviews related to mobile phones. Preprocessing techniques such as case conversion, tag removal, tokenization, pooling etc. were performed. The work has compared three different models of long short-term memory with varied hyper parameters and have reviewed the results. For this massive classification of reviews, libraries such as Seaborn, Beautiful soap, TensorFlow and Keras were used.

In Publication[31] extracted information regarding reviews using a web scrapper from the Internet and preprocessed them using part of speech tagging, tokenization, and lemmatization. Feature extraction was done using hybrid mutation earthworm algorithm and classification was done by a newly proposed classifier called as the local search improvised bat algorithm-based Elman neural network. The performance of the proposed classifier was compared with normal Elman neural network and support vector machine. The Elman neural network-based bat optimization algorithm produced an accuracy of 83.55%, f-measure of 86.045%, precision of 84.03% and recall of 85.48%.

Similar deep learning models could also be observed in the literature works[32–35]. Observations from above literature indicate that deep learning models have revolutionized sentiment analysis of consumer reviews, with architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) being widely adopted for their ability to capture semantic patterns and contextual relationships within text data. Transformer-based models, such as BERT have further advanced the field by utilizing large pre-trained networks to understand complex sentiment expressions, demonstrating state-of-the-art performance across diverse datasets and languages [36–38]. These models' ability to process sequential data and their inherent understanding of language structure make them particularly effective for analyzing the vast and varied landscape of consumer feedback.

In this research work[39], we present a hybrid deep learning approach that combines Convolutional Neural Networks and Long Short-Term Memory to perform sentiment analysis on Monkeypox-related tweets. The proposed model, dubbed as "CNNLSTM-Monkeypox", leverages the complementary strengths of CNN and LSTM architectures to capture both local and global features in the tweet text, leading to improve sentiment classification performance. Sentiment analysis on online texts, particularly social media data, has become an essential tool for understanding public sentiment and opinion on various topics. The recent outbreak of the Monkeypox virus has sparked widespread discussion and reactions on social media platforms, making sentiment analysis of Monkeypox-related tweets a valuable research area.

The proposed study[40] CNN-LSTM-based model was trained and evaluated on a dataset of Monkeypox-related tweets collected from the Twitter platform. To capture the dynamic sentiments expressed by the public, the dataset was curated to include tweets from various time periods during the Monkeypox outbreak.

In the research work[41] the model's performance was assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. This[41] publication briefs about core of the proposed approach is the integration of deep neural networks, which are capable of automatically learning salient features from raw speech data. This research paper[42] explores the use of deep neural networks for speech emotion-based sentiment recognition. By leveraging the power of deep learning architectures, the authors aim to develop an effective system that can analyze the emotional content of speech and translate it into meaningful sentiment labels (Table 1).

Materials and methods

This section describes the materials and methods that were employed for the execution of the proposed model.

Materials

50,253 reviews related to fashion products were extracted from the website of Amazon using a web scraping tool called Amazon review scraper. Input reviews were split in the ratio of 75:25 for classifier training and testing. 37,690 reviews were spent for classifier training and 12,563 reviews were used for testing. Five different fashion products are taken for experimentation. They are Bellavita Luxury perfume, Boat flash edition smart watch, Fastrack men casual sunglasses, Bata women sepia sneakers and Lavie Satchel bag. Table 2 shows the sample product reviews of Bellavita Luxury perfume. The dataset used in consists of 50,253 reviews related to fashion products. These reviews were collected from the Amazon website using a web scraping tool known as the Amazon review scraper. The collected data was then divided into training and testing sets with a 75:25 split, resulting in 37,690 reviews for training and the remaining for testing the classifier.

Methods

Preprocessing

This step is very vital as far as sentiment analysis is concerned because not the entire review will be useful in classifying and determining the polarity. Hence it is necessary that the entire review is chunked into parts and preprocessed to bring it to a standard format for the classifier to better understand it and classify it[43]. The techniques for preprocessing employed here are punctuation and whitespace removal, hashtag, and URL omission, named entity omission, tokenization, lemmatization, part of speech tagging and stop word removal. In named entity omission, we remove the name of the reviewer[44]. Tokenization and lemmatization are techniques commonly used in natural language processing. It is often the first processing step that is to be undertaken in any natural language processing task.

Tokenization for the process of splitting up the input text into smaller parts called tokens. There are three types of tokenization available such as sentence tokenization, word tokenization and character tokenization[45]. If the input is a whole text document, then sentence tokenization is used to split the entire text into singular sentences for better understanding. Similarly, if the input text is sentence, then word tokenization method is used so that the entire sentence is split up into separate words. Likewise, character organization splits a single word into

Paper title	Challenges	Applications	Research gap
Personalized emotion detection adapting models to individual emotional expressions[51]	Similarly, in the environment where emotional data is collected and used, we should think seriously about thinking ahead. This is because ethical issues are connected with emotional data collection–in addition to the fact that models may be inaccurate for more diverse populations simply due to bias	Emotion recognition technology has also been extended to read faces. Imagine the potential for a computer program that can identify what kind of mood someone is in even when you are not looking directly at them–whether they have just got up from bed or been talking with friends, for example	In the hospitality industry, as companies continue to use voice-analysis technology to enhance the customer experience, invisible ears and mechanized speech are used to dissect your most personal feelings
Textual emotion recognition: advancements and insights[52]	The methods that the authors propose make use of three major natural language processing tools: sentence extraction, tokenization and PoS tagging. This makes it easier for readers unfamiliar with Sentiment Analysis to understand how the models work	The review summarises state-of- art methodology in textual emotion recognition and proposes new ways to reach this goal. This involves two main branches: techniques for introducing an integrative semantic emotion neural network (SENN) architecture, which creates combining a bidirectional long short-term memory (BiLSTM) network and convolutional neural network (CNN) what can really pull contextual and emotional features out of text	The need for more comprehensive emotional word embedding is pointed out as a direction of improvement, which means that current models have not fully utilized all the emotional depth given to them by text
A review on machine learning and deep learning techniques for textual emotion analysis on social networks[53]	This paper gives a comprehensive overview of techniques for detecting emotion in Natural Language Processing. The main focus is on analyzing texts from social networks. It throws the practice needle across a range of time-honored practices such as SVM, CNN and BERT and uses consistent descriptions to takes a fresh look at Text Classification methods It also points out shortcomings with traditional methods	Natural Language Processing (NLP). Specifically it bears year of analyzing textual data from the likes of Twitter, LinkedIn and Reddit. This serves to show just how valuable emotion analysis is to uncover people's deep-seated feelings and actions when they use these interfaces	The paper conducts a comparison and analysis of multiple machine learning and deep learning models for emotion detection. The techniques covered include Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and BERT (Bidirectional Encoder Representations from Transformers)

Table 1. Details comparison and challenges application in sentimental emotional data.

S. No	Name of the reviewer	Date and place of review	Review rating	Review comment
1.	Sana	Reviewed in India on 12 July 2023	4.0	I was waiting for these but today I received it and when u took smell of all these the fragrance was quite familiar to me. You all can buy these same fragrance in cheaper price and the way better than this
2.	Sahnawaz	Reviewed in India on 20 July 2023	4.0	Nice fragrance it stays 5–6 h easily
3.	Preethi	Reviewed in India on 14 July 2023	4.0	It is totally worthy buying Bellavita luxury perfumes. It is a pack of four fragrances. It is affordable and rich in sense. I suggest buying it. My rating would be 4.5
4.	Anwar Hasan	Reviewed in India on 17 July 2023	5.0	Check out this fantastic women's perfume gift set! It comes with four different mind-blowing fragrances that you'll absolutely love. Once I tried it, I became a huge fan of Bellavita! The best part is that you get to enjoy different luxury perfumes every day, all at a very reasonable price. Not to mention, these fragrances have impressive staying power, lasting for a long time Smelling good and fresh is essential for me, and Bellavita helps me maintain that refreshing scent throughout the day. I receive so many compliments about how wonderful I smell and how great my perfume is. It's truly a game-changer, and I can't recommend it enough. Treat yourself or someone you care about to this delightful perfume set and experience the magic of Bellavita!
5.	Deepak Sharma	Reviewed in India on 15 July 2023	4.0	These perfumes are really good to smell and use. Small in size and can be carried easily in bag or pockets. Smells good but for women
6.	Sagar	Reviewed in India on 18 July 2023	4.0	Not nice fragrance
7.	Mrinal K	Reviewed in India on 11 July 2023	4.0	Nice product and fragrance of each bottle contents are good and lasting
8.	Rahul	Reviewed in India on 5 July 2023	4.0	All the smells are good especially the rose one
9.	–	Reviewed in India on 3 July 2023	4.0	The perfumes have a very soothing and sweet fragrance. They are mild and can be worn daily
10.	Apeksha R. Panji	Reviewed in India on 9 July 2023	4.0	Nothing new

Table 2. Bellavita luxury perfume product sample reviews.

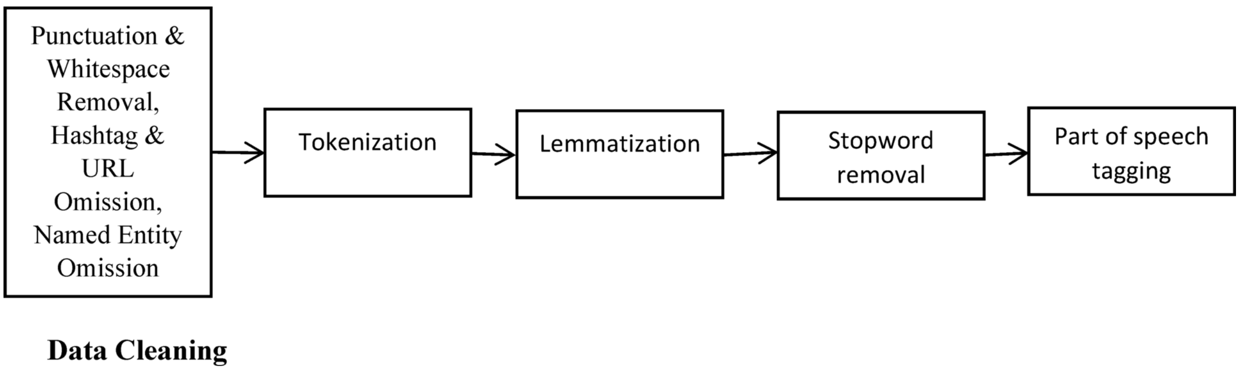

Fig. 4. Data preprocessing pipeline.

single instances of each character. Since the input we use here is a sentence we choose to use word tokenization in the proposed system, where the reviews are split up into individual words for further processing. The advantage of tokenization is that it is easy for the next preprocessing stages to remove unnecessary or irrelevant words and make the input crispier for better performance and results[46]. This task has been automated and is available with many open-source libraries such as natural language toolkit such as Keras, tensor flow, beautiful soap etc.

Lemmatization is the process of omitting the suffix part of the word and making the word more basic and simpler for easier and better understanding of the classifier[47]. It is an especially important process in NLP because it enables accurate and exact identification of the emotion and avoids any misclassification. The next step of preprocessing is stopping word removal which refers to the process of eliminating unimportant words such as I, we, is, was, this, that, here, there etc. These words are called stop words which means they are often used and does not contribute any meaning to the actual text. They are grammatically important but meaningfully not.

The next step after tokenization, lemmatization and stop word removal is part of speech tagging where the tokens created during the process of tokenization are attached to relent part of speech of the language like verb, adverb, adjective, proposition, conjunction etc. This is considered to be the final stage of preprocessing the input reviews which adds a label to the created tokens making them more meaningful[48]. Figure 4 shows the pipeline of proposed preprocessing techniques.

Feature extraction

The proposed system has chosen Bidirectional Gated Recurrent Unit (BiGRU) model to be used as a feature extractor here. This model has been previously employed as feature extractor and has given satisfactory results and hence it has been put to use in the proposed system as well. BiGRU contains two GRUS, one working in the forward direction and the other in the backward direction. A simple Gated Recurrent Unit (GRU) consists of three blocks such as the Recurrent Neural Network (RNN) block, reset block and update block. It contains two gates, namely the reset gate which controls the previous output, and the update gate controls the new input of the current state[49]. It consists of tan functions and sigmoid functions.

It is significant to note that the GRU model overcomes the vanishing gradient problem of the recurrent neural network. The main component of this network is the two gates which control the flow of data in and out of the network and it is also advantageous that the gates remember the output of the previous step. The state of the previous node and training data are taken as input for producing the current node's output. Mathematically in order to obtain the current cell state, the content of update gate is multiplied with that of the previous output. It has fewer inputs and outputs compared to RNN and hence can be trained faster. The reset gate is used for short term memory and the update gate remembers the state of previous nodes for a longer time. The specialty of BiGRU network is the presence of hidden states which extract the most relevant information from the input data set. Figure 5 depicts the architecture of bidirectional GRU model.

BiGRU (Bidirectional Gated Recurrent Unit) layers are a recurrent neural network (RNN) architecture that improves typical GRU performance by processing input sequences in both forward and backward directions. Two GRUs are employed in a BiGRU: one analyzes the input sequence in a forward direction, from start to finish, and the other in a reverse direction, from end to start. A more thorough comprehension of the data is produced by the model's ability to include context from both previous and future phases in the sequence thanks to this bidirectional approach. In tasks like natural language processing, where a word's meaning can depend on both its previous and subsequent words, BiGRU layers are especially helpful.

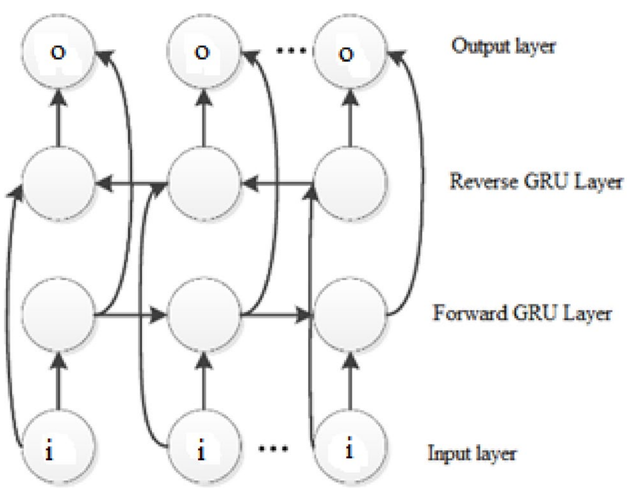

Fig. 5. BiGRU model working.

Pseudocode for BiGRU

```
Input: Product Review data
Preprocessing
Tokenize reviews
Lemmatize tokens
Remove stop words
Perform Part-of-Speech tagging
Initialize gate parameters
Set reset and update gate to default
Select forward GRU
      For current node, do
            Calculate cell state ht
            Cell state = input * previous cell state(ht-1)
            Until all nodes are done
            Select reverse GRU
            Feed training sample
            Monitor model
            Concatenate output
            // BiGRU Model Initialization
            Initialize forward GRU parameters (weights, biases, hidden state size)
            Initialize backward GRU parameters (weights, biases, hidden state size)
            For each review in the dataset:
             Forward Pass:
              For each word vector in the review (from start to end):
               Calculate hidden state using forward GRU equations (incorporating input word vector, previous hidden state, and GRU parameters)
             Backward Pass:
              For each word vector in the review (from end to start):
               Calculate hidden state using backward GRU equations (incorporating input word vector, previous hidden state, and GRU parameters)
             Concatenate forward and backward hidden states for each word
      End for
End
Output: Extracted features
```

Product Review Data: The algorithm begins by taking product review data as input, which is typically a sequence of text data that will be processed by the BiGRU model; Initialize Gate Parameters: The parameters for the gates in the GRU are initialized. These gates control the flow of information through the GRU units and determine how much of the previous state and new input should influence the current state. Set Reset and Update Gate to Default: The reset and update gates are initially set to default values. The reset gate determines how much of the previous state to forget, while the update gate decides how much of the new information to add to the current state. Select Forward GRU:The forward GRU is selected to process the input sequence from the beginning to the end. This part of the BiGRU processes the sequence in the standard left-to-right direction. Calculate Cell State *ht:* At each node, the current cell state *ht* is calculated. The cell state represents the hidden state of the GRU at the current time step and captures information from both the input at this time step and the previous hidden state. Cell State = Input * Previous Cell State *ht – 1*: The current cell state *ht* is computed as a function of the input at the current time step and the previous cell state *ht – 1*. This operation is controlled by the GRU's gates, which decide how much of the previous state to carry forward and how much new information to incorporate. Select Reverse GRU: After the forward GRU has processed the sequence, the reverse GRU is selected. The reverse GRU processes the input sequence in the opposite direction, from the end to the beginning. Feed Training Sample: The reverse GRU processes the same input data but in reverse order. This step allows the model to capture information from the future context relative to each time step, which can be crucial for understanding the sequence. Monitor Model: During the training process, the model is monitored to ensure that it is learning effectively. This typically involves checking metrics like loss and accuracy. Concatenate Output: After both the forward and reverse GRUs have processed the sequence, their outputs are concatenated. This combined output captures information from both directions, providing a richer representation of the input data. Output: Extracted Features: The final output is the set of extracted features from the product review data.

Classification

After feature extraction, the input reviews are then classified using RNN based Long Short-Term Memory (LSTM) classifier. LSTM is a modified kind of neural network which is more advanced than RNN. It contains feedback loops and processes sequential information very well and it also has the capability to remember output of previous states for a longer time and hence preserves long term dependencies. It comprises of three gates namely input gate, output gate, forget gate followed by a SoftMax classification layer. Since the output of the current state depends on the output of the previous states, classification accuracy produced by this classifier is high compared to that of the other classifiers.

LSTMs have the nature of remembering previous outputs and utilizing them for determining the current state's output in order to enhance the performance. It is to be noted that leading technical giants like Apple and Google are making use of LSTM. It basically is dependent upon back propagation technique for its working[50]. The role of input gate is to determine what range of input values should be allowed to enter the network, similarly the forget gate is used to decide what content should be let off by the network and output gate finds out the output of the current state using current input and output of the previous state. Figure 6 shows the architecture of LSTM.

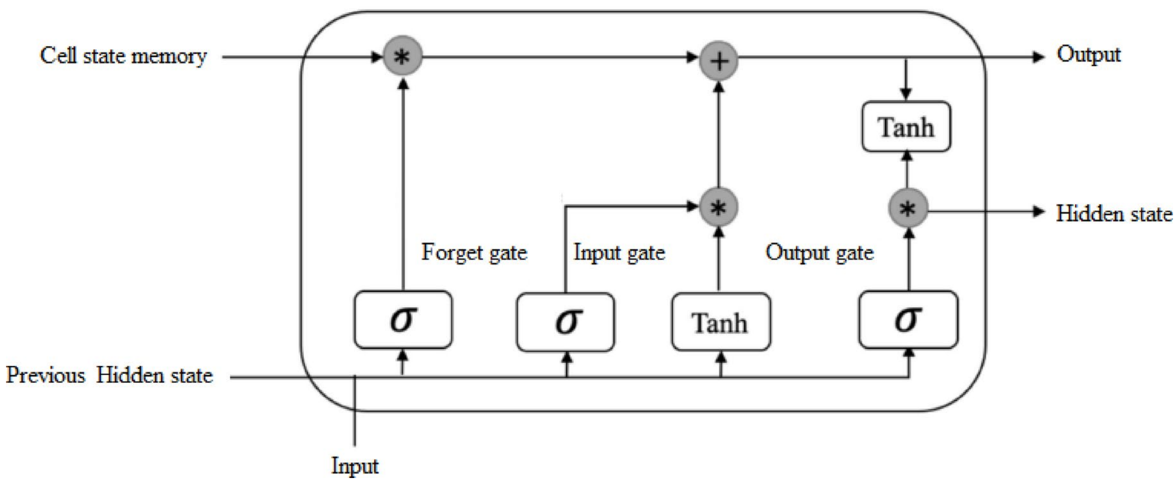

Fig. 6. LSTM gate architecture.

Pseudocode for LSTM

Input: Extracted product features

Set hyper parameters for classifier

Initialize epochs, no. of iterations, learning rate

Set input gate(i), output gate(o)and forget gate(f)to default values

For current node, do

Calculate internal state of cell

*Ct = i*g + f * Ct-1*

Calculate hidden state of cell

*Ht = Tan (Ct) * o*

Repeat for all nodes

Measure error

End for

End

Output: Classified polarity of reviews

Proposed system

The proposed system aims to perform sentence level sentiment analysis and polarize reviews into three categories of positive reviews, negative reviews, and neutral reviews. Reviews from standard online shopping website of Amazon are extracted and used as input. Preprocessing techniques such as spelling correction, punctuation removal, hash tag removal, named entity omission, tokenization, lemmatization, stop word removal and part of speech tagging were employed before feature extraction. Bidirectional gated recurrent unit technique is used as feature extractor and a hybrid recurrent neural network based long short-term memory classifier is used for classifying sentiments underlying the reviews. The process of feature extraction and classification are shown in Fig. 7.

Figure 7 depicts the initial preprocessing stage, where raw product reviews undergo several NLP techniques such as tokenization, lemmatization, stop word removal, named entity recognition, and part of speech tagging. This stage prepares the textual data for feature extraction by converting it into a cleaner, more structured format. The feature aggregation layer combines several feature vectors generated by the preceding layers into a single, unified representation. Following preprocessing, the cleaned reviews are fed into the Bidirectional Gated Recurrent Unit (BiGRU) layer. A word vector matrix is generated by embedding words into a continuous vector space by Word2Vec. The dual pathway architecture of BiGRU processes the text both forwards and backwards, thereby capturing contextual relationships and dependencies that a single direction processing might miss. The output of this stage is a set of high-quality features that represent the core sentiments expressed in the reviews. The extracted features are then input into the LSTM classifier. LSTM is known for its ability to remember long-term dependencies due to its unique gating mechanism consisting of input, forget, and output gates. This capability is crucial for accurately determining the sentiment polarity of product reviews, which may contain complex emotional expressions and nuanced opinions. Finally, the output layer is depicted where the LSTM classifier's analysis is translated into sentiment polarities: positive, negative, or neutral. This stage represents the termination of the process, depicting how the hybrid model directly addresses the challenges of sentiment analysis in e-commerce setting.

The proposed system is also compared with existing techniques such as deep convolutional neural network, multilayer perceptron, CapsuleNet and generative adversarial networks. Figure 8 below shows the proposed model of sentiment analysis in a pictorial format.

The proposed sentiment analysis depicted in Fig. 8 involves several key stages, each chosen for its ability to contribute to the accurate analysis of sentiments in online product reviews. A detailed explanation of each stage and the rationale is given below:

- *Data Collection* The foundation of any sentiment analysis model is the dataset it is being trained on. Collecting a diverse range of online product reviews ensures the model can learn from a wide spectrum of senti-

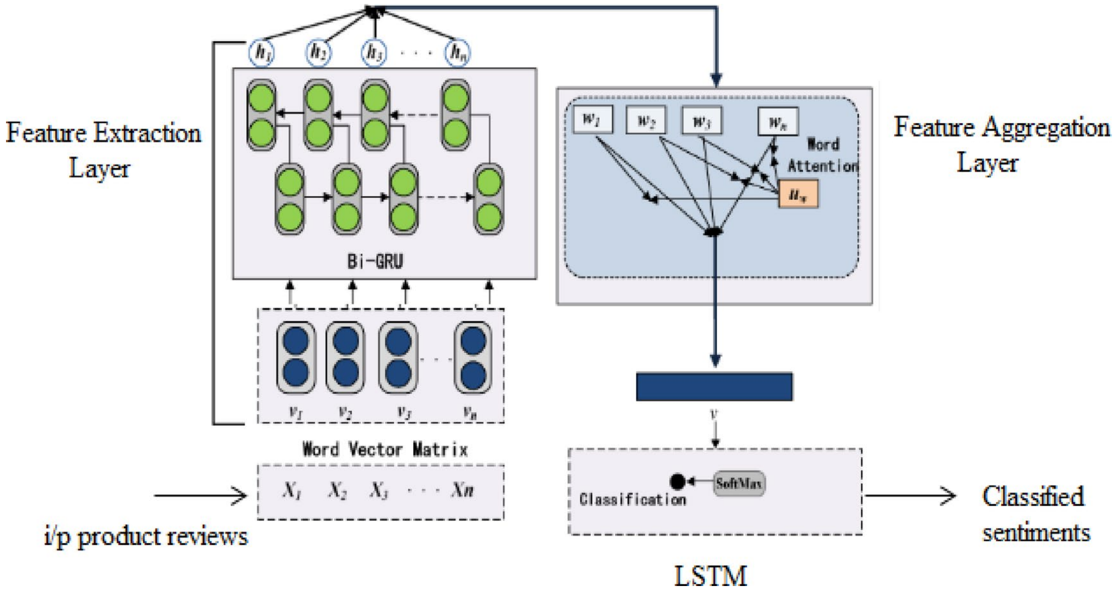

Fig. 7. Working of BiGRU and LSTM.

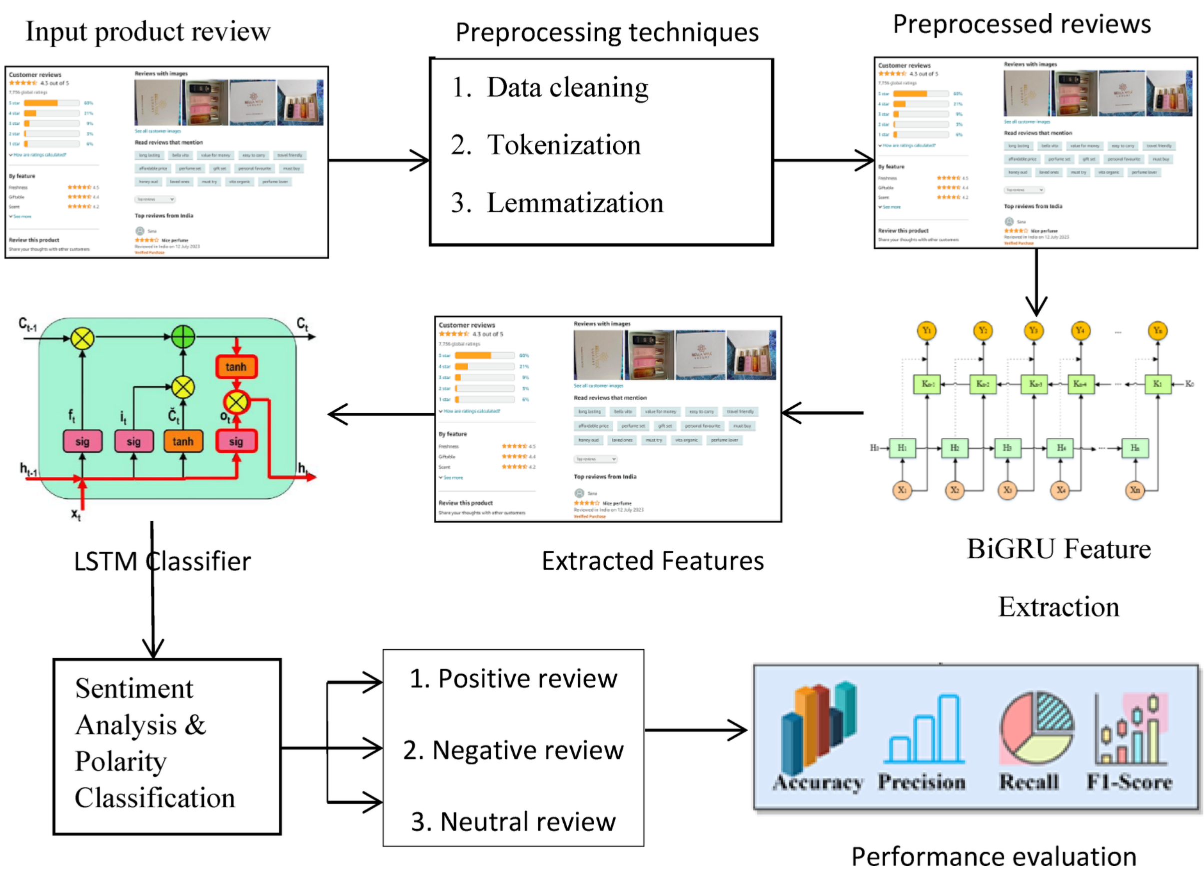

Fig. 8. Proposed sentiment analysis of product reviews.

ments, expressions, and vocabularies. Reviews are scraped from e-commerce platforms, focusing on publicly available, ethically sourced data that reflects varied consumer opinions on products.

- *Data Preprocessing* Raw text data is often noisy and unstructured, necessitating preprocessing to enhance model performance. Cleaning the data helps in removing irrelevant information, reducing complexity, and standardizing the text input for feature extraction. The proposed module includes tokenization (breaking down text into meaningful units), lemmatization (reducing words to their base or dictionary form), removing stop words (filtering out common words that add little semantic value), and named entity recognition (identifying and classifying key information elements).

- *Feature Extraction* Transforming text data into a format understandable by machine learning models is crucial. Effective feature extraction captures the essential attributes of the text that are most indicative of sentiment. The use of Bi-directional Gated Recurrent Units (BiGRU) is chosen for its ability to process text sequences in both forward and reverse directions, capturing contextual dependencies and nuances in language that might be missed when reading text linearly.
- *Model Training and Classification* The core of sentiment analysis is to classify the sentiment of the text accurately. Training a model to recognize and predict sentiment based on extracted features is vital for achieving high accuracy. A hybrid model combining RNN-based Long Short-Term Memory (LSTM) with BiGRU is used. LSTM units are adept at remembering long-term dependencies in text data, crucial for understanding sentiment context. The hybrid model leverages the strengths of both LSTM and BiGRU for improved sentiment classification.
- *Evaluation and Optimization* Assessing the model's performance using appropriate metrics and optimizing based on feedback is essential for refining the analysis capability. Standard evaluation metrics like accuracy, precision, recall, and F1-score are employed to measure the model's effectiveness. Optimization may involve adjusting model parameters, retraining with additional data, or incorporating feedback loops for continuous improvement.
- *Application and Integration* The ultimate goal is to apply the sentiment analysis model in real-world scenarios, integrating it into business workflows for actionable insights. The model is designed to be integrated into existing systems, providing real-time sentiment analysis capabilities for market research, customer feedback monitoring, and strategic decision-making.

Each stage of the proposed approach is carefully designed to build upon the previous one, ensuring a comprehensive sentiment analysis framework. The rationale behind each technique is rooted in maximizing the model's ability to understand, interpret, and classify sentiments accurately, demonstrating its application in enhancing business intelligence and customer insight strategies.

Experimentation, results and analysis

Experimental setup

The input was extracted from Amazon online shopping application with reference to reviews regarding fashion products such as. A web scraping tool called as Amazon reviews scraper is employed for pulling out necessary Bellavita Luxury perfume, Boat flash edition smart watch, Fastrack men casual sunglasses, Bata women sepia sneakers and Lavie Satchel bag reviews from the Internet. The tool is able to extract the reviews with the help of the product URL. It extracts only verified reviews and sorts them based on the time of review generation. Table 3 shows the details of input product reviews.

Figure 9 shows the representation of input product reviews taken for experimentation of proposed system in a graphical manner.

These are ways for handling problems with data shortages: Oversampling: We can increase the number of minority classes in a dataset by making copies or creating new examples, using a method called SMOTE. Under sampling: Cut down the number of examples in the bigger groups to create a balanced collection. However, this might cause some information to be lost. here the product like perfume is 7780 over the bags ,sunglass and watch sampling which is oversampling rate.

The Figs. 9, 10, 11, 12, 13 and 14 below show the product reviews of Bellavita Luxury perfume, Boat flash edition smart watch, Fastrack men casual sunglasses, Bata women sepia sneakers and Lavie Satchel bag respectively from the Amazon website.

A few of the extracted reviews are presented in Tables 3, 4, 5 and 6 below.

This Tables 4, 5, 6, and 7 showcases customer reviews, presumably for sunglasses, collected from Amazon India. The reviews span from April 13th, 2023, to July 5th, 2023. While most reviewers express positive sentiment, awarding 4 or 5-star ratings and praising the quality (e.g., Taufeeq, Chandrakant, Vaibhav wade), one review stands out as highly negative. This particular review cites issues not with the product itself, but with Amazon's return process, describing it as "worst services" due to a delayed product pick-up. This negative experience highlights the importance of efficient customer service and return processes for e-commerce platforms. The final review, by Rashmi, simply mentions "Daily wear," which, without further context, offers limited insight into the product's specific features or performance.

S.No	Product name	No. of reviews
1.	Bellavita luxury perfume	7780
2.	Boat flash edition smart watch	31,429
3.	Fastrack men casual sunglasses	1058
4.	Bata women sepia sneakers	3253
5.	Lavie satchel bag	6830
Total reviews		50,350

Table 3. Input Product review details.

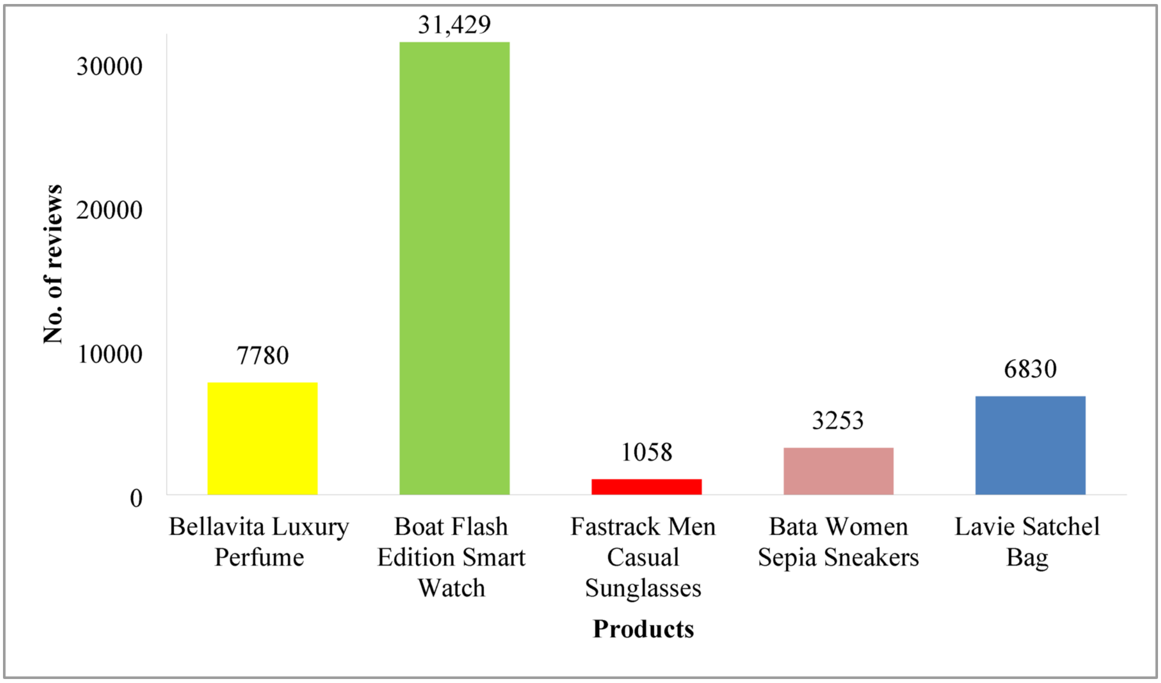

Fig. 9. Input product reviews graphical format.

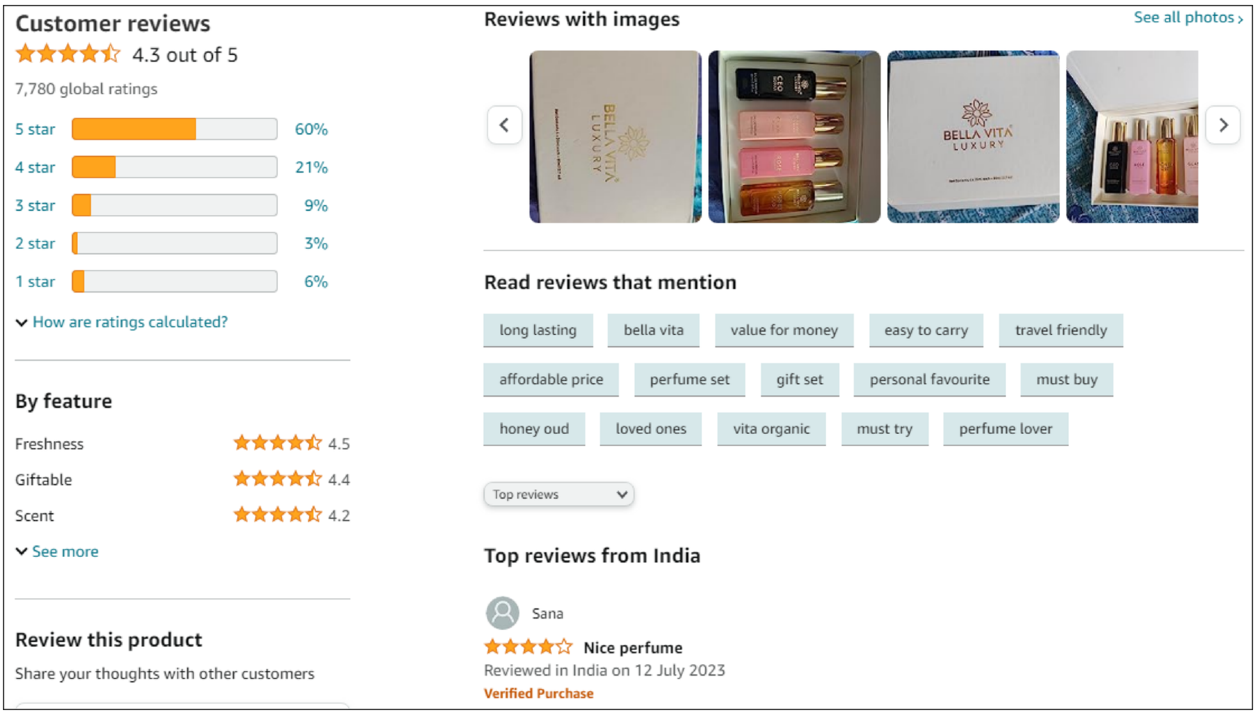

Fig. 10. Bellavita luxury perfume product review.

Preprocessing

After applying data cleaning techniques such as punctuation removal, hash tag and username removal, spelling corrections etc., we begin with the process of tokenization, lemmatization, stop word removal and part of speech tagging. The output of word tokenization technique of natural language processing is given below. word_tokenize is a predefined function of Natural Language Toolkit (NLTK) of python language. The study evaluates the proposed techniques on two datasets of 85 mb data set from Kaggle is used for *word_tokenize (“Good quality as expected. Worth buying. Really usefull for family. The material is artificial leather. OK for money*

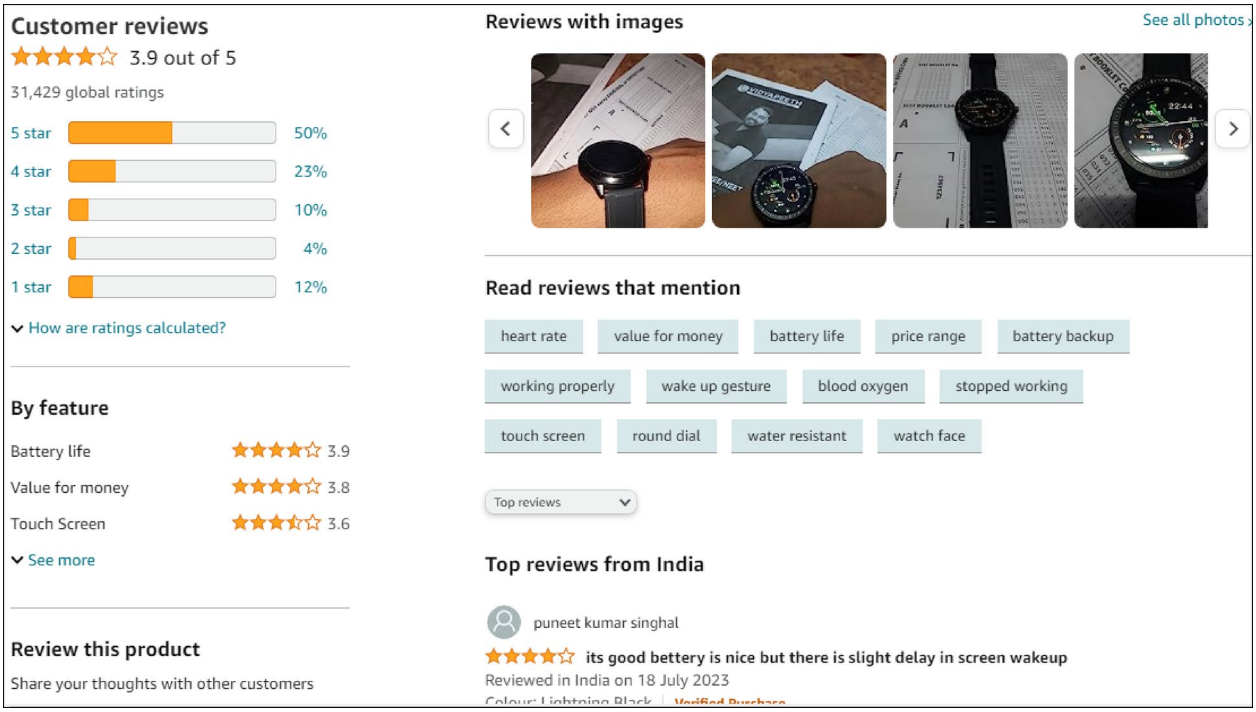

Fig. 11. Customer reviews of boat flash edition smart watch.

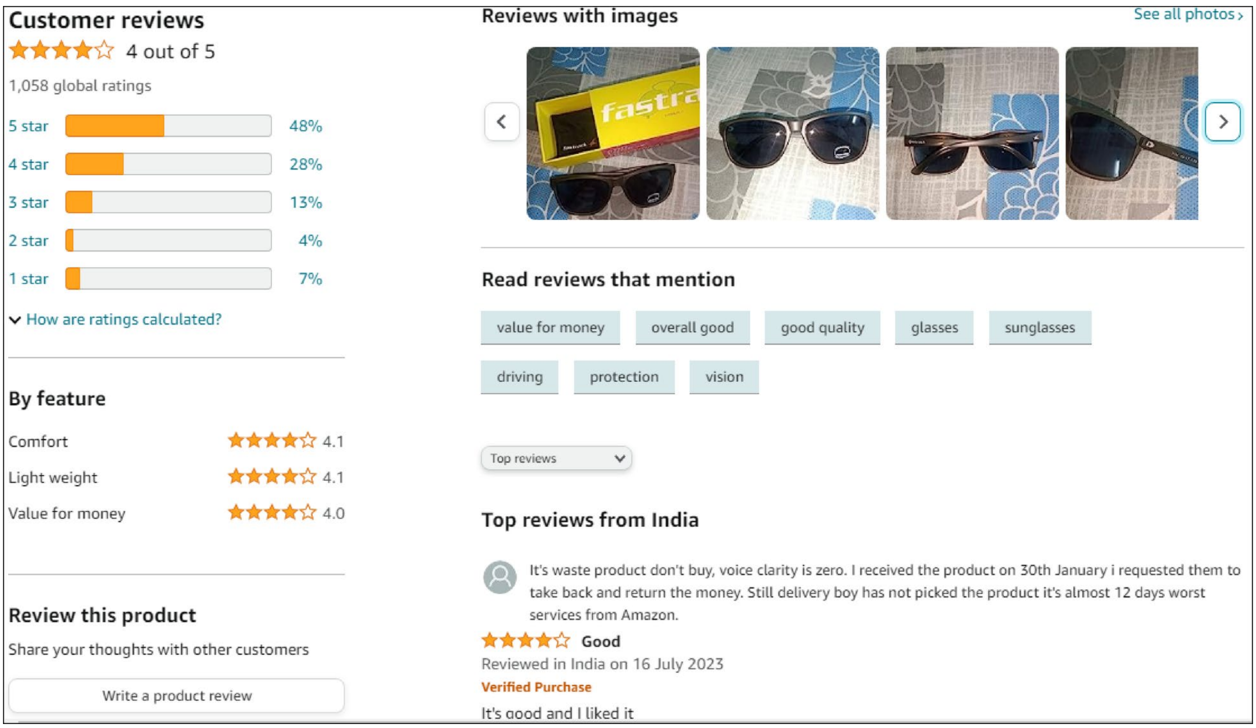

Fig. 12. Fastrack men casual sunglasses product review.

spend. Recommend buying.")=['Good', 'quality', 'as', 'expected', 'Worth', 'buying', 'Really', 'useful', 'for', 'family', 'Material', 'is', 'artificial', 'leather', 'OK', 'for', 'money', 'spend', 'Recommend', 'to', 'buy'].

It can be seen that the output of word tokenization is single words divided from a sentence. This step is very helpful for further processes of sentiment analysis. The process of lemmatization is demonstrated in Fig. 15 below.

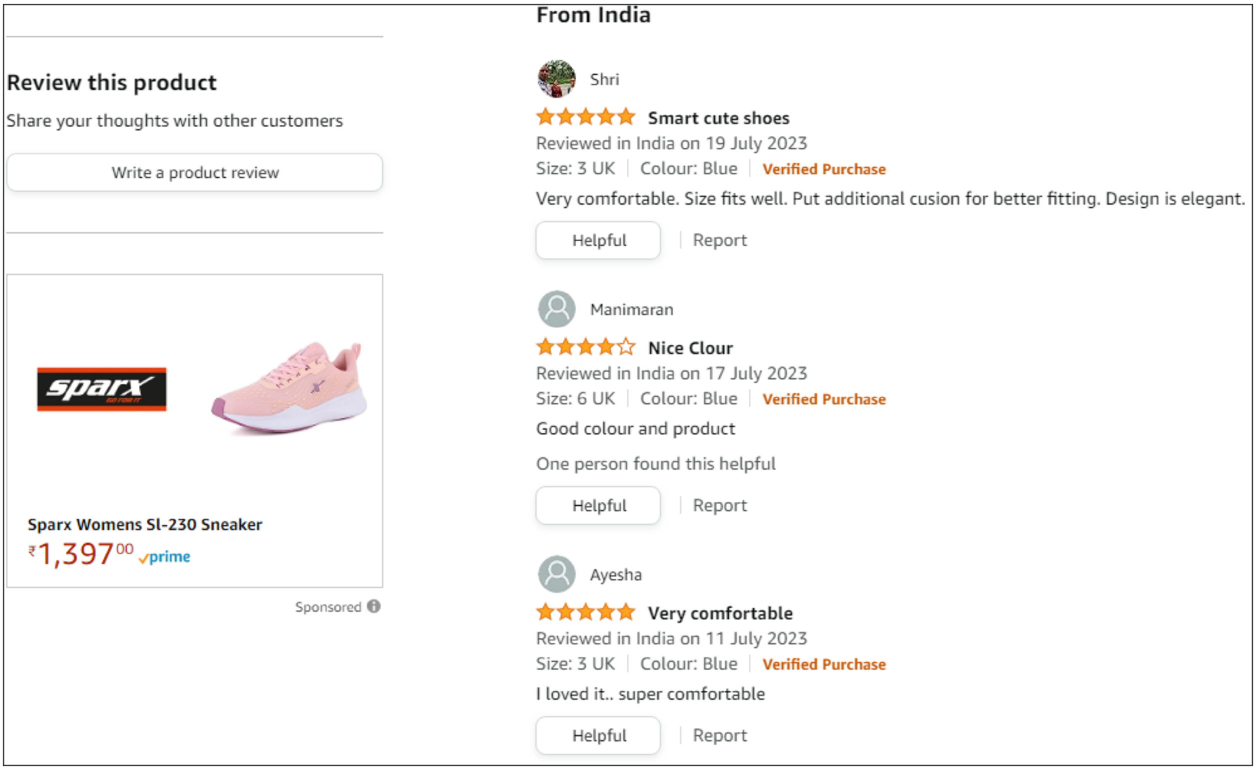

Fig. 13. Reviews of bata women sepia sneakers.

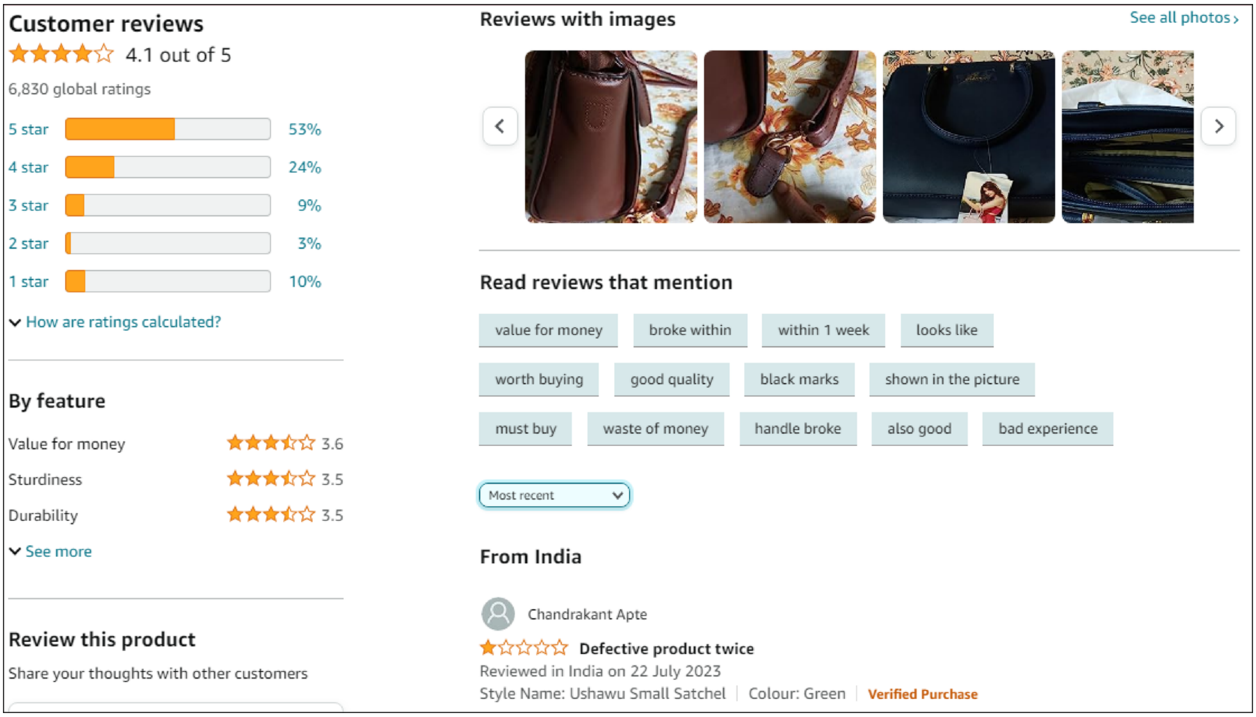

Fig. 14. Lavie Satchel bag product review.

S. No	Name of the reviewer	Date and place of review	Review rating	Review comments
1.	Manoj	Reviewed in India on 17 July 2023	1.0	Laging watch
2.	Anas khan	Reviewed in India on 20 July 2023	5.0	Good working watch best battery work and super stylish
3.	Akash Satsangi	Reviewed in India on 19 July 2023	1.0	The watch stopped working after 8 days. It is not water resistant
4.	Puneet Kumar Singhal	Reviewed in India on 18 July 2023	4.0	works well, I use it as a normal watch not tracking anything specific you can get 20 + days backup if not connected to mobile, only issue is doesn't wake up always or a slight delay in screen wakeup
5.	Rahul	Reviewed in India on 18 July 2023	5.0	Value for money

Table 4. Sample boat flash edition smart watch reviews.

S. No	Name of the reviewer	Date and place of review	Review rating	Review comment
1.	Taufeeq	Reviewed in India on 1 July 2023	5.0	I like it
2.	–	–	–	Its waste product don't buy, I received the product on 30th January. I requested them to take it back and return the money. Still delivery boy has not picked the product it's almost 12 days worst services from Amazon
3.	Chandrakant	Reviewed in India on 31 May 2023	5.0	Sunglasses quality is amazing, feel the quality
4.	Vaibhav wade	Reviewed in India on 13 April 2023	5.0	Nice sunglasses
5.	Rashmi	Reviewed in India on 5 July 2023	4.0	Daily wear

Table 5. Fastrack men casual sunglasses customer reviews sample.

S. No	Name of the reviewer	Date and place of review	Review rating	Review comment
1.	Keerthi. S	Reviewed in India on 25 June 2023	1.0	Don't buy they are not worth buying
2.	Bharathi	Reviewed in India on 22 June 2023	3.0	Not bad
3.	Raana Chatterjee	Reviewed in India on 21 June 2023	1.0	Received and received a damaged and old product
4.	Rupa Kumari	Reviewed in India on 6 July 2023	4.0	Very good
5.	Manimaran	Reviewed in India on 17 July 2023	4.0	Good colour and product

Table 6. Bata women sepia sneakers review samples.

S. No	Name of the reviewer	Date and place of review	Review rating	Review comment
1.	Amazon Customer	Reviewed in India on 20 July 2023	5.0	Good quality as expected. Worth buying. Really useful for family. The material is artificial leather. OK for money spend. Recommend buying
2.	Manali Tomar	Reviewed in India on 18 July 2023	4.0	The bag is not big or small. It's a medium size bag. Overall Quality and look is good. Price worth buying
3.	Kaveri Pramod	Reviewed in India on 18 July 2023	3.0	Nice bag but fake brand!!!
4.	Parveen P	Reviewed in India on 18 July 2023	5.0	The product is ready good…. Great quality, spacious, lot of compartment, looks elegant
5.	–	Reviewed in India on 12 July 2023	3.0	2 months me hi handle nikal kar bahar aa gya h …not worth as price

Table 7. Lavie satchel bag reviews.

Stopwords have been removed from the sample review using the function stop_words of NLTK toolkit.

stop_words ("Good quality as expected. Worth buying. Really useful for family. The material is artificial leather. OK for money spend. Recommend buying") = ['as', 'for', 'is', 'to', 'ok'].

The review after stop word removal becomes like this.

"*Good quality expected. Worth buying. Really useful family. Material artificial leather. money spent. Recommend buy.*"

The output of part of speech tagging after the reviews have been tokenized and lemmatized is represented in Table 8 below.

Feature extraction and classification

Once the review has been preprocessed, the features relevant to the classification process are extracted as explained earlier. Adverbs and adjectives usually contain the sentiment of the text. It is easier to identify the

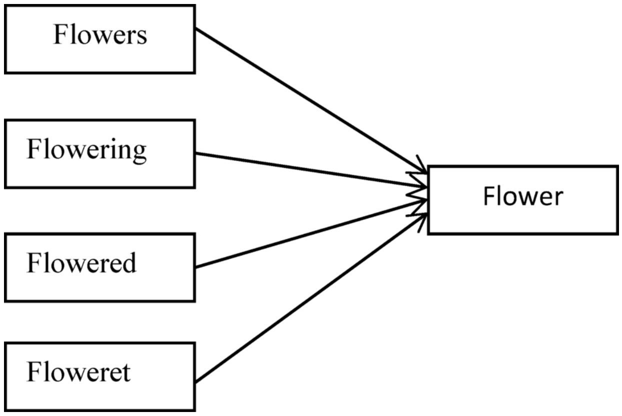

Fig. 15. Lemmatization output.

Good—Adjective	Family—Noun
Quality—Noun	Material—Noun
Expected—Adjective	Artificial—Noun
Worth—Adverb	Leather—Noun
Buying—Verb	Money—Noun
Really—Adverb	Spend—Verb
Useful—Verb	Recommend—Verb
	Buy—Verb

Table 8. Part of speech tagging of input reviews.

polarity when we extract these words as features. Other opinionated words which might express positive or negative comments are also added to the feature vector. To extract the product features from the preprocessed input, BiGRU model is used. The output of BiGRU feature vector is shown below for a sample review.

BiGRU Feature Vector = {good, quality, worth, useful, recommend}

LSTM classifier uses this feature vector and classifies each product review into three types of sentiment such as positive sentiment, negative sentiment, and neutral sentiment. If the reviews are not understandable or has both positive and negative reviews together, review rating values are utilized to calculate the polarity of sentiment. Polarity classification of fashion product reviews done by the RNN based LSTM classifier is portrayed in Table 9.

Performance metrics

Classification metrics such as accuracy, precision, recall, F-score, and AUC are calculated for the proposed system using the Eqs. (1–4) below.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

Table 10 shows the values of performance metrics achieved by the proposed RNN based LSTM classifier.

Figure 16 shows the graphical representation of the performance metrics.

Figure 17 shows the ROC (Receiver Operating Characteristics) curve of the proposed system and the corresponding AUC values of all the five types of products chosen for experimentation. Bellavita Luxury perfume achieves an AUC score of 0.995, Boat flash edition smartwatch attains 0.988 AUC score, Fastrack men's casual sunglasses reaches an excellent AUC of 1.0, Bata women's sepia sneakers achieve 0.991 AUC and Lavie Satchel bag also reaches the maximum AUC score of 1.0. It should be noted that for all five products, the value of AUC is higher than 0.99, which indicates the tremendous classification performance of the proposed model.

S. No	Product reviews	Polarity
1.	I was waiting for these but today I received it and when u took smell of all these the fragrance was quite familiar to me. You all can buy these same fragrance in cheaper price and the way better than this	Negative
2.	Nice fragrance it stays 5–6 h easily	Positive
3.	It is totally worthy buying Bellavita luxury perfumes. It is a pack of four fragrances. It is affordable and rich in sense. I suggest buying it. My rating would be 4.5	Positive
4.	Check out this fantastic women's perfume gift set! It comes with four different mind-blowing fragrances that you'll absolutely love. Once I tried it, I became a huge fan of Bellavita! The best part is that you get to enjoy different luxury perfumes every day, all at a very reasonable price. Not to mention, these fragrances have impressive staying power, lasting for a long time. Smelling good and fresh is essential for me, and Bellavita helps me maintain that refreshing scent throughout the day. I receive so many compliments about how wonderful I smell and how great my perfume is. It's truly a game-changer, and I can't recommend it enough. Treat yourself or someone you care about to this delightful perfume set and experience the magic of Bellavita!	Positive
5.	These perfumes are really good to smell and use. Small in size and can be carried easily in bag or pockets. Smells good but for women	Positive
6.	Not nice fragrance	Negative
7.	Nice product and fragrance of each bottle contents are good and lasting	Positive
8.	All the smells are good especially the rose one	Positive
9.	The perfumes have a very soothing and sweet fragrance. They are mild and can be wear daily	Positive
10.	Nothing new	Neutral
11.	Laging watch	Negative
12.	Good working watch best battery work and super stylish	Positive
13.	The watch stopped working after 8 days. It is not water resistant	Negative
14.	works well, I use it as a normal watch not tracking anything specific you can get 20 + days backup if not connected to mobile, only issue is doesn't wake up always or a slight delay in screen wakeup	Positive
15.	Value for money	Positive
16.	I like it	Positive
17.	Its waste product don't buy, I received the product on 30th January. I requested them to take it back and return the money. Still delivery boy has not picked the product it's almost 12 days worst services from Amazon	Negative
18.	Sunglasses quality is amazing, feel the quality	Positive
19.	Nice sunglasses	Positive
20.	Daily wear	Neutral
21.	Don't buy they are not worth buying	Negative
22.	Not bad	Neutral
23.	Received and received a damaged and old product	Negative
24.	Very good	Positive
25.	Good colour and product	Positive
26.	Good quality as expected. Worth buying. Really usefull for family. The material is artificial leather. OK for money spend. Recommend buying	Positive
27.	The bag is not big or small. It's a medium size bag. Overall Quality and look is good. Price worth buying	Positive
28.	Nice bag but fake brand!!!	Positive
29.	The product is ready good…. Great quality, spacious, lot of compartment, looks elegant	Positive
30.	2 months me hi handle nikal kar bahar aa gya h …not worth as price	Negative

Table 9. Sentiment classification of product reviews.

S. No	Metrics	Values
1.	Accuracy	98.79
2.	Precision	96.64
3.	Recall	98.70
4.	F1-Score	97.43
5.	AUC	0.992

Table 10. Performance metrics of proposed system.

Comparative analysis

Table 11 shows the results achieved by the proposed system and compares it with existing systems like Deep Convolutional Neural Network (DCNN), Multilayer Perceptron (MLP), CapsuleNet (CN) and Generative Adversarial Networks (GAN). In our comparative analysis, we aimed to benchmark the performance of our proposed sentiment analysis model against other prevalent models in the field, namely Deep Convolutional Neural Network (DCNN), Multilayer Perceptron (MLP), CapsuleNet (CN), and Generative Adversarial Networks (GAN). Each of these models represents a significant approach in the domain of natural language processing and sentiment analysis, offering unique perspectives on handling text data for sentiment classification. Deep Convolutional Neural Network (DCNN) is known for its ability to capture spatial hierarchies in data. In

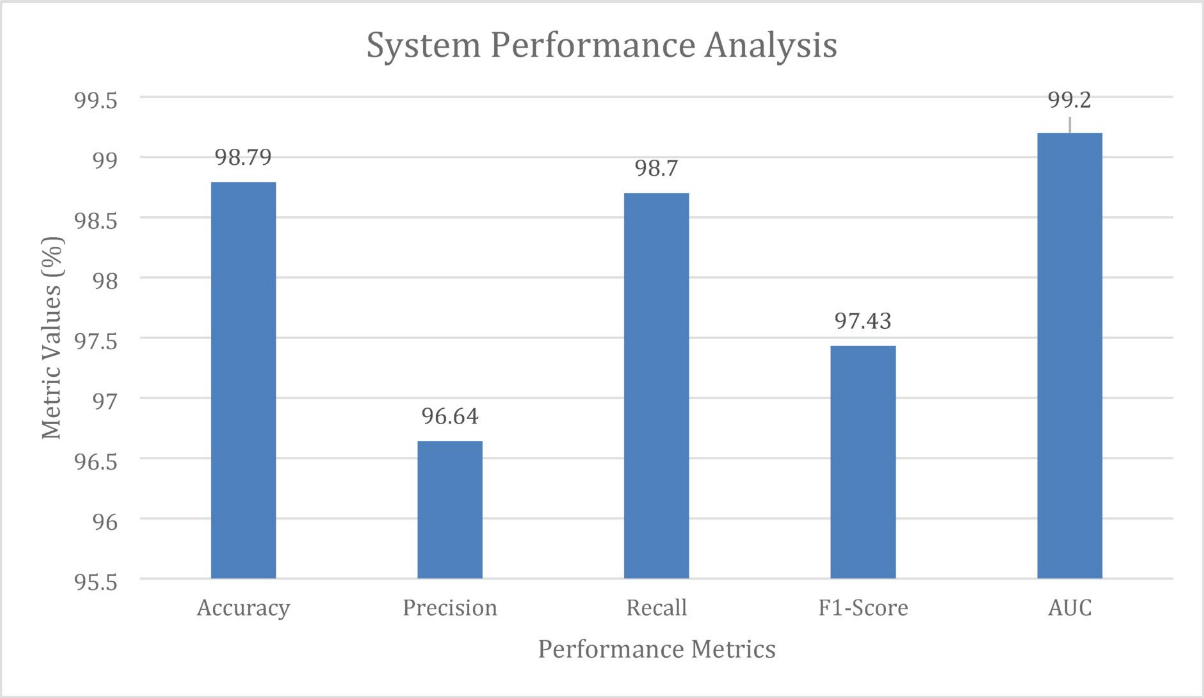

Fig. 16. Proposed system performance.

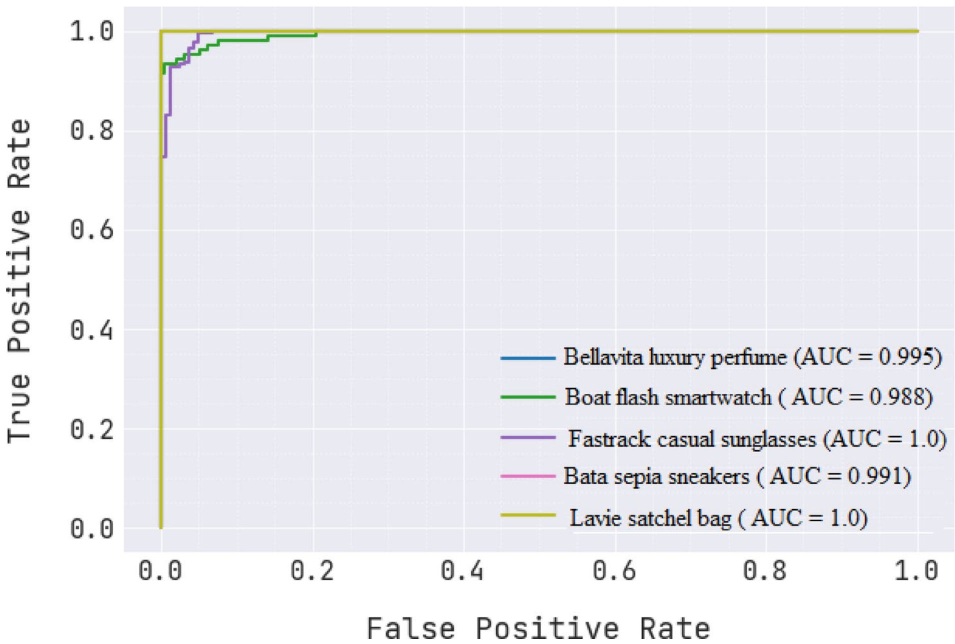

Fig. 17. ROC Curve.

sentiment analysis, DCNNs can effectively identify complex patterns in text data, such as varying n-grams, that are indicative of sentiment. Multilayer Perceptron (MLP) is a basic form of neural network that can capture relationships in high-dimensional data. Its application in sentiment analysis often involves text vectorization followed by the classification of sentiments using the network layers. Capsule Net (CN) is an advanced neural network that aims to capture spatial hierarchies between features in data. In the context of sentiment analysis, Capsule Nets offer an innovative way to understand the complex relationships in textual data, potentially leading to more nuanced sentiment classification. Generative Adversarial Networks (GAN) are known for their ability to generate data and can also be used in sentiment analysis to augment datasets or refine models through adversarial training, enhancing the model's ability to classify sentiments accurately. These four models

S. No	Algorithms	Metrics			
		Accuracy	Precision	Recall	F1-score
1.	DCNN	72.8	87.71	68.79	89.52
2.	MLP	94.71	87.06	90.73	95.15
3.	CN	87.77	88.28	90.79	83.1
4.	GAN	91.7	92.07	93.91	90.67
5.	Proposed model	98.79	96.64	98.70	97.43

Table 11. Comparative performance analysis.

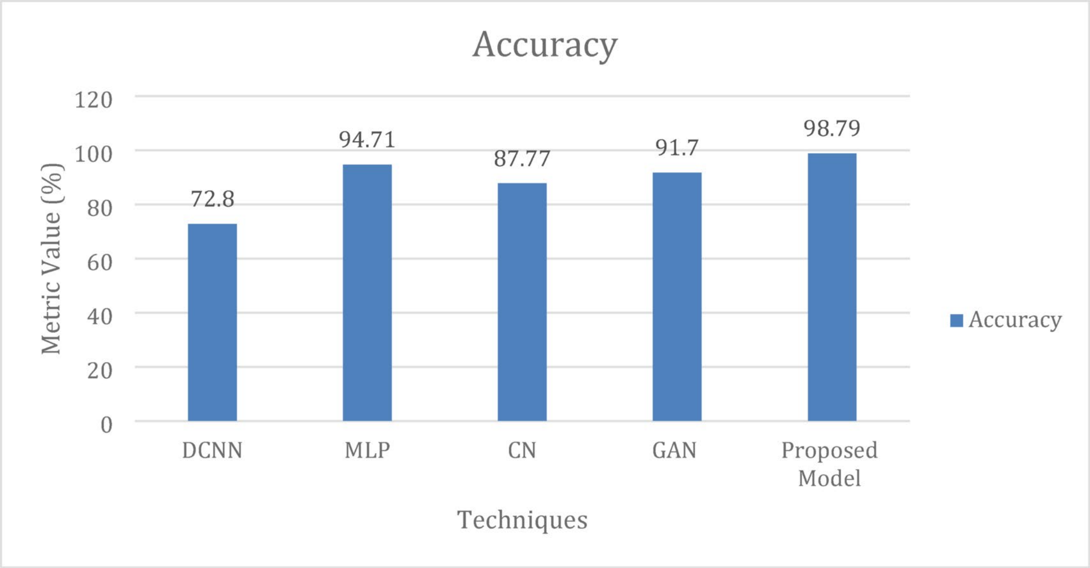

Fig. 18. Comparative analysis of accuracy.

were selected for comparison due to their distinctive contributions and proven effectiveness in the realm of sentiment analysis, which makes them benchmarks in the field. Deep Convolutional Neural Network (DCNN) has demonstrated exceptional capability in capturing spatial hierarchies of text data, making it a robust model for recognizing complex patterns indicative of sentiments. Multilayer Perceptron (MLP), with its simpler neural network structure, has been included due to its versatility and ability to handle high-dimensional data, providing a foundational comparison for neural network-based sentiment analysis. Capsule Net (CN), noted for its advanced ability to understand spatial hierarchies and relationships within text, offers a cutting-edge perspective on text analysis, potentially offering deeper insights into sentiment classification. Lastly, Generative Adversarial Networks (GAN) are included for their innovative approach to data augmentation and model refinement in sentiment analysis, showcasing the potential for using adversarial methods to enhance model performance. Comparing our proposed model against these diverse and significant models allows us to highlight the unique advantages and potential improvements our approach offers within the broader sentiment analysis landscape.

Figure 18 below shows the comparative analysis of performance regarding accuracy in sentiment analysis of product reviews by various classifiers. From the Table 11 and Fig. 18, it can be clearly seen that proposed system performance is better and superior to other existing algorithms for sentiment analysis.

Figure 19 shows the precision- recall analysis of existing and proposed models.

Challenges and future directions

Despite the effectiveness of the existing acceleration methods, there are still many future directions that require being explored. In spite of its growth and popularity, sentiment analysis of product reviews has got its own challenges that need to be addressed in the near future. The most common challenge faced is the presence of regional languages other than commonly used languages like English, Chinese, Arabic, etc. The classifier built cannot be trained in every language that is present. Multinational companies have now opened their horizons for their customers to access and use applications in customized regional languages to promote their products. But this feature has made the analysis of product reviews very challenging. One possible solution is to rely on language translation tools.

Also difficult is addressing biased reviews. Not all product reviews are authentic. Sometimes negative evaluations are made to degrade a product for business reasons. Finding such reviews is difficult. When a company hires biased reviewers, the ratings can be good. Next, sentiment analysis must handle review data with

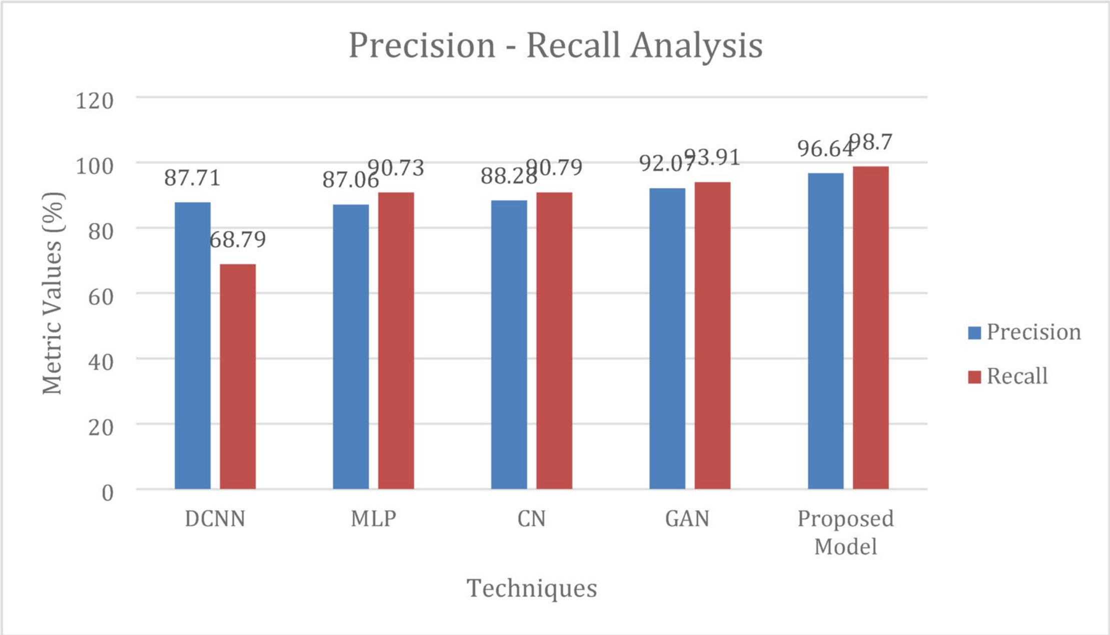

Fig. 19. Precision—recall analysis.

double negations, idioms, and phrases. A double-negative review is positive. However, the sentiment classifier may not find it positive. Similarly, idioms have context-specific meanings. Classifiers may miss context.

Homonyms and the phenomenon of polys provide another major issue. These are the same words with many meanings. Clear knowledge and language context are needed for comprehending these comparable terms.

Quality research may also face a lack of excellent datasets. These are some product evaluation and analysis of sentiment issues. All have possible solutions that we must develop to make this study precise.

The future direction of sentiment analysis is to combine different levels of sentiment analysis. All three levels, such as document level, sentence level and aspect level analysis could be performed in order to attain better accuracy and results. As of now, polarity detection is constrained and applied to only a few sectors such as online shopping, hotels and resorts, food and delivery etc. But there are many areas to which this can be successfully applied.

Conclusion
This research work introduced a novel sentiment analysis model utilizing a hybrid of Bi-directional Gated Recurrent Units (BiGRU) and Recurrent Neural Network (RNN)-based Long Short-Term Memory (LSTM) architectures. Our findings demonstrate superior performance in accurately classifying the sentiment of product reviews across diverse categories, contributing significantly to the domain of natural language processing and e-commerce analytics.

Comparative significance
Compared to existing models, our hybrid approach exhibits enhanced capability in capturing contextual small and long-term dependencies within textual data. This advancement aligns with recent research emphasizing the need for more sophisticated models in sentiment analysis.

Implications and limitations
The practical implications of our research are vast, offering e-commerce platforms and businesses an efficient tool for understanding customer sentiments at a scale. However, the model's performance across languages and extremely subtle sentiments remains an area for future exploration. The inherent complexity of deep learning models also poses challenges in interpretability, urging further research towards explainable AI.

Future directions
Building on this foundation, future research could explore the incorporation of multimodal data analysis to enhance sentiment detection. Expanding the model's applicability to other languages and domains, such as social media analytics, presents an exciting frontier. Additionally, refining the model through the integration of attention mechanisms or transformers could further improve both performance and interpretability. In conclusion, our study not only introduces a potent model for sentiment analysis but also opens several avenues for future research aimed at bridging the gap between advanced machine learning techniques and practical business applications.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Received: 16 October 2023; Accepted: 2 May 2025
Published online: 13 May 2025

References

1. Murthy, G. S. N., Allu, S. R., Andhavarapu, B., Bagadi, M. & Belusonti, M. Text based sentiment analysis using LSTM. *Int. J. Eng. Res. Tech. Res.* **9**(05), 299. https://doi.org/10.17577/IJERTV9IS050290 (2020).
2. Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S. & Nagrath, P. Sentiment analysis using gated recurrent neural networks. *SN Comput. Sci.* **1**, 1–13. https://doi.org/10.1007/s42979-020-0076-y (2020).
3. Al-Natour, S. & Turetken, O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *Int. J. Inf. Manag.* **54**, 102132. https://doi.org/10.1016/j.ijinfomgt.2020.102132 (2020).
4. Ahmed, H. M., Javed Awan, M., Khan, N. S., Yasin, A. & Faisal Shehzad, H. M. Sentiment analysis of online food reviews using big data analytics. *Elem. Educ. Online* **20**(2), 827–836 (2021).
5. Ray, B., Garain, A. & Sarkar, R. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Appl. Soft Comput.* **98**, 106935. https://doi.org/10.1016/j.asoc.2020.106935 (2021).
6. Li, Z., Zou, Y., Zhang, C., Zhang, Q., & Wei, Z. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training. Preprint at arXiv:2111.02194. https://doi.org/10.48550/arXiv.2111.02194 (2021).
7. Basiri, M. E., Nemati, S., Abdar, M., Cambria, E. & Acharya, U. R. ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Futur. Gener. Comput. Syst.* **115**, 279–294. https://doi.org/10.1016/j.future.2020.08.005 (2021).
8. Al-Smadi, M., Talafha, B., Al-Ayyoub, M. & Jararweh, Y. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int. J. Mach. Learn. Cybern.* **10**, 2163–2175. https://doi.org/10.1007/s13042-018-0799-4 (2019).
9. Dong, Y. et al. A sentiment analysis method of capsule network based on BiLSTM. *IEEE Access* **8**, 37014–37020. https://doi.org/10.1109/ACCESS.2020.2973711 (2020).
10. Kumar, A., Srinivasan, K., Cheng, W. H. & Zomaya, A. Y. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Inf. Process. Manag.* **57**(1), 102141. https://doi.org/10.1016/j.ipm.2019.102141 (2020).
11. Kauffmann, E. et al. Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining. *Sustainability* **11**(15), 4235. https://doi.org/10.3390/su11154235 (2019).
12. Bonta, V., Kumaresh, N. & Janardhan, N. A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian J. Comput. Sci. Technol.* **8**(S2), 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037 (2019).
13. Han, Y., Liu, M. & Jing, W. Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. *IEEE Access* **8**, 21314–21325. https://doi.org/10.1109/ACCESS.2020.2969473 (2020).
14. Yan, W., Zhou, L., Qian, Z., Xiao, L. & Zhu, H. Sentiment analysis of student texts using the CNN-BiGRU-AT model. *Sci. Program.* **2021**, 1–9. https://doi.org/10.1155/2021/8405623 (2021).
15. Behera, R. K., Jena, M., Rath, S. K. & Misra, S. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Inf. Process. Manag.* **58**(1), 102435. https://doi.org/10.1016/j.ipm.2020.102435 (2021).
16. Singh, C., Imam, T., Wibowo, S. & Grandhi, S. A deep learning approach for sentiment analysis of COVID-19 reviews. *Appl. Sci.* **12**(8), 3709. https://doi.org/10.3390/app12083709 (2022).
17. Xu, G., Meng, Y., Qiu, X., Yu, Z. & Wu, X. Sentiment analysis of comment texts based on BiLSTM. *Ieee Access* **7**, 51522–51532. https://doi.org/10.1109/ACCESS.2019.2909919 (2019).
18. Imran, A. S., Yang, R., Kastrati, Z., Daudpota, S. M. & Shaikh, S. The impact of synthetic text generation for sentiment analysis using GAN based models. *Egypt. Inform. J.* **23**(3), 547–557. https://doi.org/10.1016/j.eij.2022.05.006 (2022).
19. Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. Sentiment analysis on product reviews using machine learning techniques. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2017* 639–647. https://doi.org/10.1007/978-981-13-0617-4_61 (Springer, 2019).
20. Fauzi, M. A. Word2Vec model for sentiment analysis of product reviews in Indonesian language. *Int. J. Electr. Comput. Eng.* **9**(1), 525. https://doi.org/10.11591/ijece.v9i1.pp525-53 (2019).
21. Dang, N. C., Moreno-García, M. N. & De la Prieta, F. Sentiment analysis based on deep learning: A comparative study. *Electronics* **9**(3), 483. https://doi.org/10.3390/electronics9030483 (2020).
22. Jain, P. K., Saravanan, V. & Pamula, R. A hybrid CNN-LSTM: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents. *Trans. Asian Low-Resour. Lang. Inf. Process.* **20**(5), 1–15. https://doi.org/10.1145/3457206 (2021).
23. Liu, Y., Lu, J., Yang, J. & Mao, F. Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. *Math. Biosci. Eng.* **17**(6), 7819–7837. https://doi.org/10.1016/j.jnlssr.2021.10.003 (2020).
24. Ali, N. M., Abd El Hamid, M. M., & Youssif, A. Sentiment analysis for movies reviews dataset using deep learning models. Int. J. Data Min. Knowledge Management Process (IJDKP) Vol, 9. https://ssrn.com/abstract=3403985 (2019).
25. Ombabi, A. H., Ouarda, W. & Alimi, A. M. Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. *Soc. Netw. Anal. Min.* **10**, 1–13. https://doi.org/10.1007/s13278-020-00668-1 (2020).
26. Sasikala, P. & Mary Immaculate Sheela, L. Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *J. Big Data* **7**, 1–20. https://doi.org/10.1186/s40537-020-00308-7 (2020).
27. Yang, L., Li, Y., Wang, J. & Sherratt, R. S. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* **8**, 23522–23530 (2020).
28. Zhu, Q., Jiang, X. & Ye, R. Sentiment analysis of review text based on BiGRU-attention and hybrid CNN. *IEEE Access* **9**, 149077–149088. https://doi.org/10.1109/ACCESS.2021.3118537 (2021).
29. Shankar, R. S., Rajanikanth, J., Sivaramaraju, V. V., Murthy, K. V. Prediction of employee attrition using datamining. In *2018 IEEE international conference on system, computation, automation and networking (ICSCAN)* 1–8 (IEEE, 2018).
30. Iqbal, A. et al. Sentiment analysis of consumer reviews using deep learning. *Sustainability* **14**(17), 10844. https://doi.org/10.3390/su141710844 (2022).
31. Zhao, H., Liu, Z., Yao, X. & Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Inf. Process. Manag.* **58**(5), 102656. https://doi.org/10.1016/j.ipm.2021.102656 (2021).
32. Iqbal, A. et al. Sentiment analysis of consumer reviews using deep learning. *Sustainability* **14**(17), 10844. https://doi.org/10.3390/su141710844 (2022).
33. Aleem, S. et al. Machine learning algorithms for depression: Diagnosis, insights, and research directions. *Electronics* **11**, 1111 (2022).
34. Alantari, H. J., Currim, I. S., Deng, Y. & Singh, S. An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. *Int. J. Res. Mark.* **39**, 1–19 (2021).

35. Mohbey, K. K. Sentiment analysis for product rating using a deep learning approach, In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India* 121–126 https://doi.org/10.1109/ICAIS50930.2021.9395802. (2021).
36. Silpa, N., Rao, V. M., Subbarao, M. V., Kurada, R. R., Reddy, S. S., Uppalapati, P. J. An enriched employee retention analysis system with a combination strategy of feature selection and machine learning techniques. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)* 142–149 (IEEE, 2023).
37. Shankar, R. S., Priyadarshini, V., Neelima, P., Raminaidu, C. H. Analyzing attrition and performance of an employee using machine learning techniques. In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* 1601–1608. (IEEE, 2021).
38. Jyothirmayee, S., Kumar, V. D., Rao, C. S. & Shankar, R. S. Predicting stock exchange using supervised learning algorithms. *Int. J. Innov. Technol. Explor. Eng.* **9**(1), 4081–4090 (2019).
39. Mohbey, K. K. et al. A CNN-LSTM-based hybrid deep learning approach for sentiment analysis on Monkeypox tweets. *New Gener. Comput.* **42**, 89–107. https://doi.org/10.1007/s00354-023-00227-0 (2024).
40. Meena, G., Mohbey, K. K. & Kumar, S. Monkeypox recognition and prediction from visuals using deep transfer learning-based neural networks. *Multimed. Tools Appl* **83**, 71695–71719. https://doi.org/10.1007/s11042-024-18437-z (2024).
41. Meena, G. & Mohbey, K. K. Sentiment analysis on images using different transfer learning models. *Procedia Comput. Sci.* **218**, 1640–1649 (2023).
42. Choudhary, R. R., Meena, G. & Mohbey, K. K. Speech emotion based sentiment recognition using deep neural networks. *J. Phys. Conf. Ser.* **2236**(1), 012003. https://doi.org/10.1088/1742-6596/2236/1/012003 (2022).
43. Liu, C. et al. Improving sentiment analysis accuracy with emoji embedding. *J. Saf. Sci. Resil.* **2**(4), 246–252. https://doi.org/10.3934/mbe.2020398 (2021).
44. Yang, R., & Edalati, M. Using GAN-based models to sentimental analysis on imbalanced datasets in education domain. Preprint at arXiv:2108.12061. (2021).
45. Mahendhiran, P. D. & Subramanian, K. CLSA-CapsNet: Dependency based concept level sentiment analysis for text. *J. Intell. Fuzzy Syst.* **43**(1), 107–123. https://doi.org/10.3233/JIFS-211321 (2022).
46. Sunitha, D., Patra, R. K., Babu, N. V., Suresh, A. & Gupta, S. C. Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries. *Pattern Recogn. Lett.* **158**, 164–170. https://doi.org/10.1016/j.patrec.2022.04.027 (2022).
47. Pande, S. D. et al. Assessment and recommendation of neural networks and precise techniques for sentiment systems analysis. *J. Ambient. Intell. Hum. Comput.* https://doi.org/10.1007/s12652-023-04643-4 (2023).
48. Tian, J., Slamu, W., Xu, M., Xu, C. & Wang, X. Research on aspect-level sentiment analysis based on text comments. *Symmetry* **14**(5), 1072. https://doi.org/10.3390/sym14051072 (2022).
49. Bhowmik, N. R., Arifuzzaman, M. & Mondal, M. R. H. Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms. *Array* **13**, 100123. https://doi.org/10.1016/j.array.2021.100123 (2022).
50. Muhammad, P. F., Kusumaningrum, R. & Wibowo, A. Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews. *Procedia Comput. Sci.* **179**, 728–735. https://doi.org/10.1016/j.procs.2021.01.061 (2021).
51. Diwakar, M., Saraswoti, S., Umesh, T. & Sanjib, N. Personalized emotion detection adapting models to individual emotional expressions. *Int. J. Innov. Sci. Res. Technol. (IJISRT)* **9**(10), 1932–1937. https://doi.org/10.38124/ijisrt/IJISRT24OCT1478 (2024).
52. Esmail, A. A. et al. Textual emotion recognition: Advancements and insights. In *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt* 295–300 https://doi.org/10.1109/NILES63360.2024.10753241. (2024).
53. Peng, S. et al. A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Netw.* **8**(5), 745–762 (2022).

Acknowledgements

This research was supported by the Researchers Supporting Project number (RSPD2025R846), King Saud University, Riyadh, Saudi Arabia.

Author contributions

Data curation: GAL and DA ; Writing original draft: GAL and DA ; Supervision: SS; Project administration: SS; Conceptualization: AM and BB; Methodology: AM and BB ; Verification: TAS and NAA ; Validation: TAS and NAA; Visualization SS; Resources: SS; Review & Editing: TAS and NAA. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025