

Mining of Customer Review Feedback Using Sentiment Analysis for Smart Phone Product



P. Suresh and K. Gurumoorthy

Abstract With the fast growth of e-commerce, a large number of products are sold online, and a lot more people are purchasing products online. People also give feedback of product purchased in the form of reviews. The user-generated reviews for products and services are largely available on the Internet. To extract the valuable understanding, classification of reviews is required from a huge set of feedbacks which have been converted into positive and negative sentiments. The process of sentiment analysis (SA) has mined the attitude, opinions and emotions spontaneously from text, speech and database via natural language process (NLP). It contains feedback review about product, product features or some sentiment emotional views on the product given by the customer. In this research work, feedback from the customer which is associated with smart phones is taken from [amazon.com](https://www.amazon.com) in order to predict the rating of the product given by the user feedback using SA. Feedback review of the customers has been collected from [amazon.com](https://www.amazon.com), and this research work had nearly 4000 customer feedback reviews based on related categories, namely, ID of the product, name of the product, name of the brand, rating, review of the product and vote based on review. This kind of analysis will be helpful for the customers to identify the better product with quick analysis and find the unspoken commodity. Perhaps the e-commerce sector will increase revenue by giving discounts on unique tacit goods.

Keywords Customer review · Feedback · Sentiment analysis · Smart phone · Product

P. Suresh (✉)

Department of Computer Science, Salem Sowdeswari College (Govt-Aided),
Salem, Tamil Nadu, India

K. Gurumoorthy

Department of Computer Science, Periyar University, Salem, Tamil Nadu, India

1 Introduction

Nowadays huge information, reviews or opinions are getting stored in the websites of social media or e-services in the form of raw data. In recent years, customers prefer to get the product through online. Therefore the prospective buyers choose the right products and large number of data collected from customers in the form of feedback. These research works softly provide opinionated words which assist the e-commerce business to recognize areas that need to be improved. In order to implement the proper methods, the raw data is required, and various methods are either related to adverbs, nouns, verbs or adjectives. However, a recent study shown in SA has the combination of adjectives and adverbs which are stronger than adjectives alone, but none of the research has focused on all the possible combinations of adverbs, adjectives and verbs. This paper presents the theoretical analysis of some well-known methods or proposal of SA. The data is really helpful, as well as knowledge for businesses looking to understand suggestions on their products or services. In fact, it is beneficial for consumers to be helpful with companies, the ratings and opinion stripped from them. For instance, reviews of hotels in a city that helps a consumer search for a good hotel to stay a city. Similarly, product ratings help other people determine whether the phone is worth buying or not. Similarly, phone reviews facilitate different users to choose whether or not the mobile phone is worth for money or not [1, 2]. This methodology incorporates various algorithms for evaluating and making sense out of the corpus of data. SA uses the manipulation of natural language to remove the particular knowledge from the data [3, 4]. The key component takes consumer reviews as input in NLP system and then separated by tokenizer into token. A series of characters combined in a text is called a symbol, while a semantic unit for processing is identified. The tokenizer contains punctuation marks, icons, words, etc. Consumers who help with an additional level of quality segregation for the product may convert an expression into word-level tokens that have been executed to generate rules that provide word counts and even ranking. In this analysis, consumer reviews related to smart phones were obtained from amazon.com in order to estimate the product ranking based on user feedback using SA. The organizations of this paper is described as follows: Section 2 defines the associated review about method based on SA; Section 3 defines the proposed methodology based on customer review data collection, data preprocessing, SA and frequency of review rating; and Sect. 4 discusses the conclusions.

2 Literature Review

In this section, many researches have been done using SA. The area of text-based classification was not much of a research work done to classify the sentences or word related to feedback rating.

[5] describes various mobile phone reviews based on SA which can be obtained by learning various posts given from various numbers of users that can classify the smart phones. [6] discusses to collect feedback on Tokopedia's quality of service on-line analysis over several months of observations. Because of its high-level precision, the Naïve Bayes classification technique is applied, which facilitates large data processing. The outcome showed that the element of reliability and personalization needed more focus because they have a strong negative feeling. [7] suggests a continuous sentiment analysis (CSA) system for repetitive study of customer emotions emphasizing the intent of one such effort to catch the tone of the message. This 'sentiment analysis' approach is relatively a recent technique which uses NLP to provide meaning to the plentiful data available at hand. [8] explores a novel approach by trusting the comments on social media to build on a specific topic. The proposed solution includes a list of the words used to construct training dependent on knowledge of positive words and negative words. Originally data is obtained from web networks, namely, Amazon, Flipkart and eBay, etc., along with collecting special attributes from the information gathered and then applying them to vector and value set. This research study is carried out step by step, explaining the feedback, based on interpretation of SA. [9] proposed a methodology that used reviews from many customers who visit different hotels, book rooms and order food. This can be achieved using SVM algorithms, logistic regression and Naive Bayes. [10] proposes a machine learning (ML) model for SA and compares some popular ML approaches in the context of sentiment classification. The classifier efficiency is calculated in terms of precision. [11] explores how text analysis methods can be used to investigate based on various tweet language patterns and message volumes on Twitter into some of the details in a series of posts. The experimental tests reveal that the current classifiers for machine learning are more effective and accomplish better in terms of precision. [12] proposed detailed process descriptions of sentiment polarity categorization. Experiments were done with positive findings on both sentence-level and review-level categorizations. [13] discusses the long short-term memory (LSTM) classifier provides the best results in classifying comments with POS-tagged lexicon features into positive and negative review. [14] illustrates that mathematical approaches are frequently mixed with conventional linguistic laws and definitions. [15] discusses the study of emotions relevant to the field of education and gamification of learning. Naïve Bayes (NB) is the better classifier in which the results based on accuracy showed better results compared to the disagreement group when performing 1000 students for testing, the agreement group in learning to use gamification, that may improve student's evaluation in learning. [16] suggests to introduce a multimodal sentiment prediction framework from various modal sources, namely, images, text and audio, that can interpret the projected emotions and combine them to understand the student's community emotions in a classroom. This system includes a digital microphone device that records the student's live video and audio streams during a lecture. [17] had used lexicon-based approach from Twitter posts to implement the SA mechanism. In this paper SentiCircles and lexicon-based methods are proposed which have been described primarily on the logical semantics that expresses the word-oriented sentiment. There are three separate databases

which are Stanford Sentiment Gold Standard (STS)-Gold, Obama-McCain Debate (OMD) and Health Care Reform (HCR) tested by the proposed process. [18] suggested a Twitter data SA method. This paper describes the few methods to do text-based SA using lexicon-based methods. They dealt with three separate databases, AlchemyAPI, OpenCalais and Zemanta. [19] describes the approach of hybrid model of SA which is based on learning and lexicon. These can define emotions and polarity of the opinion which can be obtained with better accuracy of 75%. [20] suggested a technique of sentiment analysis through the study of restaurant domain customer feedback. In addition, creating the rule base to classifier by predicting the polarity of the review used by priority based algorithm. For incremental instance counts, the analysis performs well by k-nearest neighbourhood (K-NN) create. [21] suggested the a priori algorithm is a basic association rule mining (ARM) algorithm that can be used to mine frequent item sets and their related rules. An improved a priori algorithm is used to prune the subset and classify the more regular item collection, resulting in a better smart phone range. [22] has addressed the research of AI has attained an excellent level with a sublevel of ML and deep learning application with a minimal method that is proceeding to concrete future business.

The above study has to identify polarity of words using the analysis of SA techniques and several NLP concepts to linguistic the tokenized sentence and words.

3 Research Methodology

Most of the business establishments have done a 'market basket analysis' to evaluate the user input feedback on their mobile phone and buyer motive. According to an instance, a person is buying a mobile with the best battery consumption feature in the basket. Later, he switches over to better front camera features instead of considering the battery life. Moreover, in the current techniques, there is no consideration or intimation why the user or customer had switched over to another feature like front camera features instead of battery consumption feature. The existing techniques may not predict exactly, whereas the implicit rule interference algorithm is used to identify the kind of featured mobile that has been purchased by the person based on the basket data alone. To evaluate the explicit and implicit model, the present research work considers smart phone feedback analysis-based rule mining with SA. This work aims to progress a recommendation algorithm that is built from an explicit and implicit analysis based on laws of association. This paper discusses the use of NLP for deep learning as seen in Fig. 1.

Fig. 1 Block diagram of proposed methodology

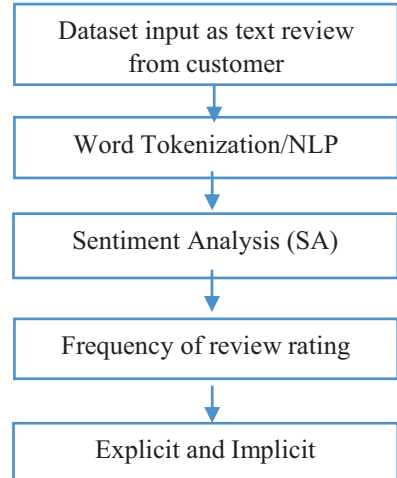


Table 1 Features involved in the data set

Feature	Description
Prod_ID	Identification number of the product
Prod_Name	Name of smart phone
Brd_Name	Company name
Rating	Customer rating scale between 1 and 5
Rview	Customer feedback provided for every smart phone
Rview_vote	Number of people providing vote who found the review helpful

3.1 Input Dataset/Feedback

The data collection employed contains consumer ratings of smart phones obtained from amazon.com. The buyer agrees a recommendation on a scale from 1 to 5 and gives its individual opinion according to the overall experience about the product. The mean value is calculated over all ratings in order to arrive at the final ranking. Other visitors may also mark yes or no to a comment that has added benefit to the review and reviewer depending on their helpfulness. In this report, we looked at over 4000 user interface reports on amazon.com for cell phones. The dataset collected from '<http://www.kaggle.com>' are outlined in Tables 1 and 2 with the following attributes from amazon.com based on the category 'Cell Phone'.

Table 2 Feedback and rating from online shopping customer based on product ID

S. no.	Prod_ID	Prod_Name	Brd_Name	Rating	Rview	Rview_vote
1	PID1	Galaxy SPHD700	Samsung	5	I feel so lucky to have found	1
2	PID2	Apple iPhone 5c 16 GB Green	Apple	5	Awesome description, condition and seller Very pleased, thank you	0
3	PID3	Lenovo A850	Lenovo	4	A very good phone, excellent hardware, very good performance and high compatibility	6
4	PID4	Nokia Asha 302	Nokia	3	Shipped quickly and was exactly what I expected	2
5	PID5	BlackBerry Bold 9650	BlackBerry	4	I liked	0
6	PID6	Huawei Honor 6 Plus	Huawei	5	Excellent	1
7	PID7	5530 XpressMusic	Nokia	4	The phone is pretty good	2
8	PID8	8330 BlackBerry	BlackBerry	4	Good	0
9	PID9	ACER LIQUID E700 TRIO	Acer	4	It works well	1
10	PID10	Acer Liquid M220	Acer	5	This is the best budget phone	0

3.2 Data Preprocessing: Word Tokenization and NLP

Once the analysis text is imported, it is treated as individual consumer input, which is tokenized and analysed by NLP to generate the appropriate relationship. However, NLP has helped in categorising this method as a regulated programme of natural language that may be apprehensive in relation to machine and human language from the computer science with deep learning. The vast amount of text was analysed and dealt with using NLP predictive analysis. This is a deep learning technique that includes stemming, chunking data and removing stop terms. The benefit of NLP in generating emotion words by segregating the words in terms of noun and even the paragraph and sentences are tokenized and chunked in deciding the sentences as positive or negative. As a result, NLP is often used as a translator when translating from one language to another. It can produce low noise, resulting in reliable results. NLP is used to feed consumer reviews as an input and is then separated into tokens by a tokenizer. A character's series has been paired with organisation containing punctuation marks, icons, special characters, sentences and other elements,

resulting in the modification of a sentence into different words using word tokenization. This study looked at how the Natural Language Toolkit (NLTK) is calculated and used with Python and how it is used to help and understand the shape of sentences as well as their context. According to this approach, the analysis must change the consumer input expressed in the attribute of review language, as well as the text of unstructured data to structured data. To begin, all NLP tasks use data from part of speech to find nouns, verbs, adjectives and roots for each word in the sentence from the analysis document. This suggested chunking NLP algorithm aids in distinguishing sentiment terms such as adverb, noun and adjective that are used as a function that can reflect high accuracy in the analysis document.

Chunking NLP Algorithm for Extracting the Required Terms

Step: 1 Get the information you need in terms of tokenization requirements
`Defextract_NN`

Step: 2 assign `grmr = r ""`

Step: 3 NBAR: # Adjectives and Nouns, Noun during terminated words

```
{<NN.*>*<NN.*>}
```

Step:4 NP: {<NBAR>}# connected with, above, in/of/etc..

Step:5 recognising the adverbs, nouns, and adjectives that are used in the summary document to express one's viewpoint {<NBAR><IN><NBAR>}

“”

Step:6 parsing the partial syntactic structure of a sentence Chkr as `nlk.RegexParser(grmr)`

```
om = set()
```

Step:7returning over tokenization of specified character `cnk` as `tokmizerfractory` for this chunker

Step:8 for `tree` in `cnk.subtrees(filt = lambda t:t.label() == 'NP');`

```
om.add(''.join([child in tree.leaves()for child[0]]))
```

```
returnom
```

Step:9 `sub= []`

```
forsentenc in data;
```

```
    #extract predefine the structure of sentences along its meaning NN (sentenc)
```

Step:10concat method in the string class as `Sub.append(extract_NN(sentenc))`

```
print (sub)
```

3.3 Sentiment Analysis

Customer feedbacks are evaluated in this proposed work using SA which has been received from the website. Before charging the money, the customer needs feedback about the company. At the moment to read all the suggestions has not possible which was provided by the customer in the website. However all kinds of product analysis or feature analysis present in the companies are available with new information. Therefore, all kinds of essential inputs that have been provided by the customers are possibly to be missed. Thus the organized review rating frequency has assisted to resolve previous challenges. Then word count has been calculated from the extraction of all tokenized words based on SA. These can be obtained by deep learning. The easiest way to interpret the reviews using an SA along with word count is to figure out the feedback rating. Hence, the rating can be based on the reviews given by the customer. After the SA output has been received, the consumer should make a quicker and minimized attempt to read the feedback as the decision. The analysis terms are equipped using document frequency (DF) or inverse document frequency (IDF) and have been used for the sake of deciding the phrase count displayed in Table 3.

Table 3 Calculate word count

PROD_ID	Product name	Brand name	Rating	Reviews	Review votes	Word counts
0	PID1 Galaxy SPHD700	Samsung	5	I feel so Lucky to have found this used (phone...	1.0	{ 'after': 1, 'again': 1, 'all': 1, 'an': 1, 'a...
1	PID1 Galaxy SPHD700	Samsung	4	Nice phone, nice up grade from my pantach revu...	0.0	{ 'an': 1, 'and': 2, 'android': 1, 'are': 1, 'b...
2	PID1 Galaxy SPHD700	Samsung	5	Very pleased.	0.0	{ 'please': 1, 'very': 1}
3	PID1 Galaxy SPHD700	Samsung	4	It works good but it goes slow sometimes but i...	0.0	{ 'but': 2, 'goes': 1, 'good': 2, 'it': 3, 'its...
4	PID1 Galaxy SPHD700	Samsung	4	Great phone to replace my lost phone The only....	0.0	{ 'adjust': 1, 'again': 1, 'am': 1, 'but': 1, '...

Algorithm for Sentiment Analysis

Step 1: CountVectorizer() converts a collection of text documents to a matrix of token counts

Step 2: assign a shorter name for analyze

Step 3: analyzer = vectorizer.build_analyzer() #which tokenizes the string

Step 4: If the string is not empty, tokenize it and proceed.

If analyzer(s):

d = {}

Step 5: Count the vocabularies and convert to an array.

Step 6: item() transforms the dictionary's (word, index) tuple pairs

Step 7: For k, v in vc.items()

D → index:word

For index, i in enumerate (w[0]);

C → word :count

Return C

Step 8: dF1 = dF → document frequency

dF['Rating'] . value_counts(). To_frame()

Step 9: color dF1[Rating] #Rating 4 higher → positive, Rating 2 lower → negative, Rating 3 → neutral

These reviewed word count to form a selected emotional word has been vectorized and gets associated with particular customer. The positive and negative feedback based on selected words – ['best', 'good', 'love', 'amazing', 'impressive', 'super', 'glad', 'fantastic', 'funny', 'wonderful', 'extraordinary', 'awesome', 'bad', 'boring', 'unhappy', 'never', 'upset', 'sad', 'terrible', 'disappointment', 'poor', 'confused', 'hard' and 'hate'] – are illustrated in Fig. 2.

In this work the frequency of the review rating is from 1 to 5. Very positive, positive, neutral, poor and really bad are the ranking ranges that are used. In a score, for example, very good = 5 stars and very negative = 1 star, all of which are translated onto a five-star grade. It indicates the overall rating scale with corresponding product_ID is illustrated in Fig. 3.

The buying ranking amount in this 3D chart depicts the measurement of both the overt and tacit relationships based on consumer input (Fig. 4). Based on the average value of buyer purchasing number of items being associated with mixture of goods and object infrequent, the X axis represents the product ID, the Y axis represents ranking and the Z axis represents the number of sales in Fig. 4.

```
Selected Words
great      8919
love      3305
happy     1150
awesome   921
bad       787
dtype: int64
```

```
ProductID
PID74      1448
PID420     1237
PID56      1069
PID37      1007
PID64       919
Name: PROD_ID, dtype: int64
```

Selected Words



Class Names



All Words in the Reviews



Fig. 2 Plot the frequency of sentimental words

This kind of analysis may be helpful for e-commerce business to improve the sales and identify the implicit product based on providing offers for particular implicit products.

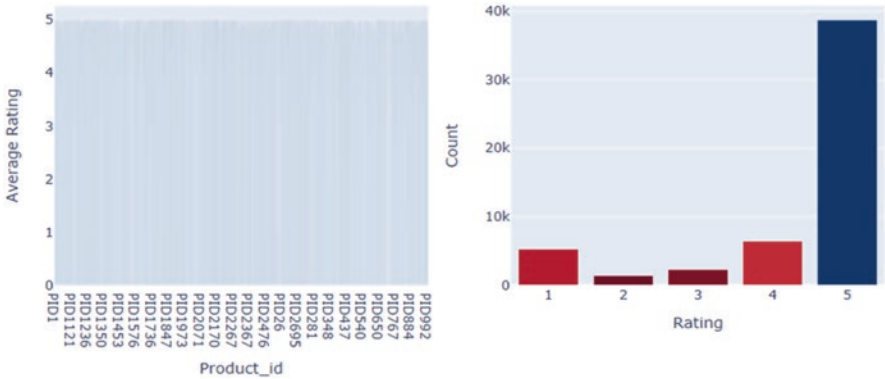


Fig. 3 Average rating based on product ID

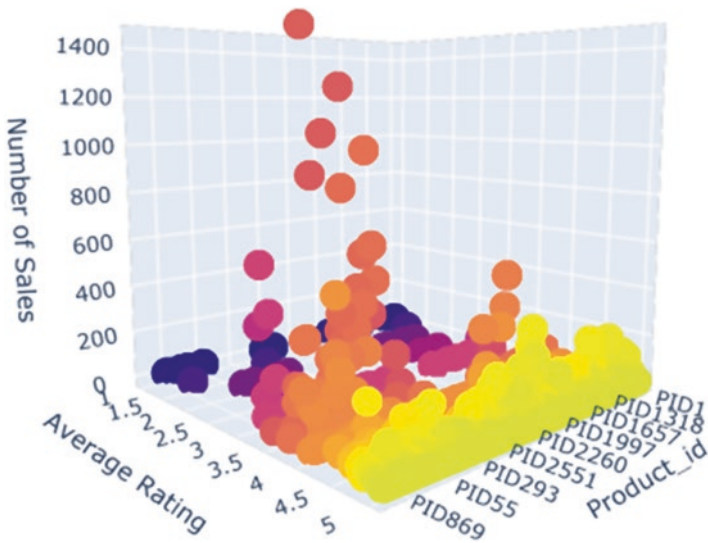


Fig. 4 3D plot of the number of sales, average rating and product ID

4 Conclusion

The most difficult task for the user is to choose the best cell phone while shopping online due to app features that are difficult to justify. Consumer reviews and ratings, on the other hand, can suggest phone quality to the customer, but there is a lag in determining the exact device feature quality based on ranking. Many businesses claim that their business success depends solely on customer satisfaction. Therefore, scientists are encouraged to find better solutions for SA. Consequently, the aim of this project was to use SA to meet the needs of customers with their review text reviews. This SA works for natural language processing (NLP), which helps

tokenize text for word counts. As a result, the word count is applied to the emotion terms as well as the consumer rating based on product ID to decide which mobile phone is better. This method of research has to be boosting the sales by defining the indirect product and offering approach for the various implicit products. In future research work, proposed system is made to evaluate the train and test dataset of SA with various classification techniques for justifying the accuracy level of qualified model.

References

1. Kumari, U., Sharma, A.K., Soni, D.: Sentiment analysis of smart phone product review using SVM classification technique. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS 2017), pp. 1469–1474 (2018)
2. Shaheen, M.: Sentiment analysis on mobile phone reviews using supervised learning techniques. *Int. J. Mod. Educ. Comput. Sci.* **11**(7), 32–43 (2019)
3. Kamalapurkar, D., Bagwe, N., Harikrishnan, R., Shahane, S., Manisha, G.: Phone recommender: sentiment analysis of phone reviews. *Int. J. Eng. Sci. Res. Technol.* **6**(5), 212–217 (2017)
4. Pankaj, P., Pandey, P., Muskan, M., Soni, N.: Sentiment analysis on customer feedback data: Amazon product reviews. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 320–322 (2019)
5. Chawla, S., Dubey, G., Rana, A.: Product opinion mining using sentiment analysis on smartphone reviews. In: International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Sept 20–22, 2017
6. Sari, P.K., Alamsyah, A., Wibowo, S.: Measuring e-Commerce service quality from online customer review using sentiment analysis. *J. Phys. Conf. Ser.* **971**(1), 012053 (2018)
7. Moorjani, G., Sadath, L.: Sentiment analysis-a tool for data mining in big data analytics. *Int. J. Innov. Technol. Explor. Eng.* **8**(9), 2125–2131 (2019)
8. Kaur, H., Kaur, P.: Dimensionality reduction in sentiment analysis using colony–support vector machine. *Int. J. Innov. Technol. Explor. Eng.* **8**(8), 2791–2797 (2019)
9. Sowmya, K., Monika, K., Radha, M., Vijay Kumar, V.: Customer review rating analysis using opinion mining. *Int. J. Innov. Technol. Explor. Eng.* **8**(7), 2444–2447 (2019)
10. Bordoloi, M., Biswas, S.: Sentiment analysis of product using machine learning technique: a comparison among NB, SVM and MaxEnt. *Int. J. Pure Appl. Math.* **118**(18), 71–83 (2018)
11. Ganagavalli, K., Mangayarkarasi, A., Nandhinisri, T., Nandhini, E.: Sentiment analysis of twitter data using machine learning algorithm. *J. Comput. Theor. Nanosci.* **15**(5), 1644–1648 (2018)
12. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data.* **2**(1), 5 (2015)
13. Thakur, P., Shrivastava, R.: Sentiment analysis of tourist review using supervised long short term memory deep learning approach. *Int. J. Innov. Res. Comput. Commun. Eng.* **7**(2), 592–604 (2019)
14. Mostafa, L.: Student sentiment analysis using gamification for education context. In: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019. AISI 2019. Advances in Intelligent Systems and Computing, vol. 1058, pp. 329–339. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-31129-2_30
15. Sharma, A., Mansotra, V.: Multimodal decision-Level group sentiment prediction of students in classrooms. *Int. J. Innov. Technol. Explor. Eng.* **8**(12), 4902–4909 (2019) ISSN: 2278-3075
16. Saif, H., He, Y., Fernandez, M., Alani, H.: Contextual semantics for sentiment analysis of Twitter. *Inf. Process. Manag.* **52**, 5–19 (2016)

17. Le, B., Nguyen, H.: Twitter sentiment analysis using machine learning techniques. In: *Advanced Computational Methods for Knowledge Engineering*, pp. 279–289. Springer, Cham (2015)
18. Appel, O., Chiclana, F., Carter, J., Fujita, H.: A hybrid approach to sentiment analysis. In: *IEEE Congress on Evolutionary Computation (CEC)*, pp. 4950–4957. IEEE (2016)
19. Mohan, A., Manisha, R., Vijayaa, B., Naren, J.: An approach to perform aspect level sentiment analysis on customer reviews using sentiscore algorithm and priority based classification. *Int. J. Comput. Sci. Inf. Technol.* **5**(3), 4145–4148 (2014)
20. Thakkar, K., Goyal, A., Bhattacharyya, B.: Emergence of deep learning as a potential solution for detection, recovery and de-noising of signals in communication systems. *Int. J. Intell. Netw.* **1**, 119–127 (2020)
21. Andersen, D.L., Ashbrook, C.S.A., Karlborg, N.B.: Significance of big data analytics and the internet of things (IoT) aspects in industrial development, governance and sustainability. *Int. J. Intell. Netw.* **1**, 107–111 (2020)
22. Devikanniga, D., Ramu, A., Haldorai, A.: Efficient diagnosis of liver disease using support vector machine optimized with crows search algorithm. *EAI Endorsed Trans. Energy Web.* **20**(29), 164177 (2018). <https://doi.org/10.4108/eai.13-7-2018.164177>