



## Twitter Sentiment Analysis using Aspect-based Bidirectional Gated Recurrent Unit with Self-Attention Mechanism

Mohan Kumar Antharasanahalli Venkataramaiah<sup>1\*</sup> Nandakumar Ambuga Narayana Achar<sup>2</sup>

<sup>1</sup>*Department of Computer Science & Engineering, Visvesvaraya Technological University, India*

<sup>2</sup>*Department of Computer Science & Engineering, City Engineering College, India*

\* Corresponding author's Email: mohankumarav31@gmail.com

---

**Abstract:** Nowadays, social media sites and applications such as Twitter, YouTube, Facebook and blogs gained much attention, because organizations use this huge information to monitor the user opinions. Therefore, the researchers focused on sentiment analysis to help organizations for better production. But, the existing studies concentrated only on document-level in the analysis, where the details of the sentiments are not considered. In order to perform a fine-grained analysis, this research study uses the aspect-based sentiment analysis on Twitter data for finding the opinions of users. In general, a raw tweet contains stop words, Uniform Resource Locator (URL), emoji that are reduced in the stage of pre-processing. The polarities of pre-processed tweets are identified, then two important feature extraction techniques are used to extract the useful information. The specific sentiments with various aspects of services are extracted by using aspect-based feature extraction, which are given as an input for Bi-directional Gated Recurrent Unit (BGRU) for classification. The learning rate of BGRU is further improved by incorporating the self-attention layers. To test the efficiency of the proposed method, the experiments are conducted on real-time collected tweets in terms of accuracy, precision, recall and f-score. The results showed that the developed method achieved 80.4% accuracy, 81% precision, 80% of recall and f-score for 1000 dataset length when compared with traditional techniques: Long-short term Memory (LSTM) and standard GRU.

**Keywords:** Aspect-based analysis, Document-level analysis, Gated recurrent unit, Opinions of user, Sentiment analysis, Self-attention layers, Social media sites.

---

### 1. Introduction

The user's opinions or emotions are extracted and identified over social sites, websites of merchant or manufactures, various blogs, discussion forums, etc. is defined as Sentiment Analysis (SA) or Opinion Mining (OM) [1]. In this research study, Twitter microblogs is considered for the SA, because it helps end-users to think regarding real-world situations or express their inner feelings. In addition, people show their opinions on movies, experiences in products or services, vast amount of topics, political candidates, sports news or just expressing their current mood [2, 3]. But, the most challenging task in SA is identifying the sentiments in Twitter data, because tweets contain informal language and usually short in length (i.e. posts are limited to 140 characters). Due to limited

characters of tweets, people find a new way to express themselves, for instance, users use a lot of abbreviations, URLs, slang or any specific terms in Twitters such as user mentions and hashtags [4, 5]. In the last few years, the researchers have studied the Twitter-based SA using various feature extraction techniques, as well as machine learning methods to find the final summary of products or services and respond quickly to users' opinions [6]. Hence, a unified method must be developed for extracting and analyzing the Twitter data through automatic tweets classification with emphasis of domain-specific paradigm [7, 8]. In general, there are three levels are present in SA, where one class is assigned to the whole document is defined as document-level. Each sentence is assigned to a class is known as sentence-level and finally, the various aspects of entities are

identified by feature or aspect level and also used to evaluate a separate polarity for each tweet [9].

The research study focused on third level SA, which is a fine-grained model that extracts the expression of users using different aspects of the entity. The third-level process contains the extraction of opinions and aspects, classification of those extracted aspects into similar classes and finally, polarities are identified [10, 11]. The three main subtasks are categorized in aspect-based SA, namely the extraction of aspects, analysis of sentiment lexicons and summarization of opinions [12]. During the first task, the opinion words polarity is identified by using sentiment lexicon analysis [13], where the presentation of extracted aspects and the polarity of their opinion words is illustrated in opinion summarization. Among the three-levels, aspect-based extraction task is the most challenging and important task, which is widely studied by many researchers [14]. The different algorithms, namely natural language processing, maximum entropy, neural network and decision tree are proposed for automatic sentiment classification for reviewing the documents [15, 16]. But, the existing techniques didn't focus on the aspect features for effective classification, where aspect-features are used to determine the fine-grained polarity towards a specific tweet. The aim of this research study is to develop an effective aspect-based BGRU with self-attention layer for Twitter classification. The term frequency-inverse document frequency (TF-IDF) and Latent Dirichlet allocation (LDA) are used to extract the features, where aspects-based extraction technique is used for finding the orientation of subjective terms present in the text. Then, the self-attention layer is used to increase the learning rate of BGRU for predicting the SA. The method is validated by collecting tweets using Twitter API on Zomato services in terms of accuracy, precision, recall and f-score.

The research paper is organized as follows: Section 2 presents the survey of existing techniques with its limitations. The explanation of proposed aspect-based BGRU methodology is given in Section 3, where the validation of aspect-based BGRU method is presented on the collected tweets in terms of various parameters in Section 4. Finally, the conclusion of the work with future development in Section 5.

## 2. Literature review

In this section, a discussion of existing techniques is presented, where these techniques are implemented

for SA. In addition, the limitation of existing techniques is also illustrated.

Z. Jianqiang, G. Xiaolin, and Z. Xuejun, [17] developed a Depth Convolution Neural Network (DCNN) to classify the sentiment of user on Twitter tweets. A sentiment feature set was formed by combining the word embedding with n-gram features and sentiment polarity score features. A DCNN model obtained these features set as an input for identifying the sentiment classification labels. The data sparseness problem was reduced by catching the context sentiment information and retain the word order information. The experiments were conducted on five datasets by means of important parameters, namely f-measure, precision, recall and accuracy. The performance of DCNN was affected, because the parameter values were too large and unlabelled microblog messages were used to extract the inputs.

M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, [18] implemented some hybrid techniques such as an emoticon classifier, an improved domain-specific classifier, a slang classifier and SentiWordNet (SWN) classifier for tweets classification. The problem of incorrect sentiment classification was solved by these hybrid techniques called T-SAF. The performance of the hybrid classifier was boosted by classifying the slang expressed in tweets, emoticons detection and classification of emoticons using an enhanced emoticon dictionary. The validation was conducted on the twitter dataset in three domains such as laptop, health and automobile. The parameters such as recall, precision, accuracy and f-measure were used to test the efficiency of hybrid techniques against traditional techniques, namely Support Vector Machine (SVM), Random Forest (RF) and so on. However, the insufficient automatic scoring of domain-specific words increased the error rate, because a lookup operation was not performed in SWN.

A. P. Rodrigues, and N. N. Chiplunkar, [19] classified the sentiments using Hybrid Lexicon-Naive Bayesian Classifier (HL-NBC) method. In real time, the method worked with labelled and large scale datasets, where lexicon model was used to filter the unwanted tweets. The method collected the tweets using Apache Flume, which were related to sports news, reviews of mobiles, Indian political issues and reviews of movies. Due to fault-tolerant mechanism, the HL-NBC worked in an efficient manner by means of execution time, precision, accuracy, recall and f-measure. But, the activities of the SA were misled by sarcastic tweets that lead to wrong classification, because the HL-NBC didn't consider how to filter the sarcasm sentiments.

L. Wang, J. Niu, and S. Yu, [20] predicted the polarities of sentiments by implementing the SentiDiff that considered the inter-relationships between textual information and sentiment diffusion patterns of Twitter messages. A sentiment reversal was investigated to analyze the sentiment diffusion of tweets. The method collected the tweets and retweets from Beijing Intelligent Starshine Information Technology Corporation, China. The efficiency of SentiDiff was tested by using various parameters such as classification accuracy and Area under the Precision-Recall Curve (PR-AUC). The combination of textual and sentiment diffusion information had a negative influence on SA tasks that lead to poor performance on the sentiment reversal prediction model.

Y. Zhang, D. Song, P. Zhang, X. Li, and P. Wang, [21] designed a Quantum-inspired Sentiment Representation (QSR) model to capture the sentiment information which was central to the SA. According to the subjective expression, namely adverbs and adjectives, sentimental phrases were extracted by QSR that were matched with the designed sentiment patterns. The representation of documents was successfully illustrated by integrating the density matrices with sentiment information. The experiments were conducted on two datasets such as Sentiment140 Twitter and Obama-McCain Debate (OMD) datasets by means of F1-score, precision, recall and accuracy. However, the QSR classification accuracy decreased with less number of sample size.

L. Terán, and J. Mancera, [22] investigated the 2017 Ecuadorian national elections and designed the Dynamic Voting Advice Applications (DVAA) for dynamic candidate profiling generation. During the construction of candidate and party profiles, the biased recommendations were given by using dynamic elements of VAA. In profile generation, the similarities level and variation between each element were shown by the final results. The tweets were collected from three main cities of Ecuador i.e. Quito, Guayaquil, and Cuenca and Root Mean Square Error

Rate (RMSE) was used to test the DVAA's performance. However, the method manually reviewed the tweets and insufficient to recover all the combined semantics. In addition, the interaction of users was limited (i.e. only 118 users) and minimum information were provided by DVAA for final recommendation.

The above existing techniques didn't concentrate on aspect-based feature extraction for effective SA. In order to overcome this issue, the aspect-based BGRU method developed the techniques based on aspect-based extraction to identify the sentiment on collected tweets. The next section explains the description of the proposed methodology.

### 3. Proposed methodology

There are five steps available in the process of SA, where the collection of tweets using Twitter API is the first step. The particular tweets are analyzed and classified using text analytics and natural language processing, because the dataset contains a vast amount of data with various words and slangs. In the second step, those collected tweets are analyzed to clean the non-textual emoticons, contents, numerical values and important information are extracted [23]. The emotions are detected from the extracted tweets in the third step then, the opinions and its reviews are inspected. Finally, the classifier identifies the sentiment from the trained tweets as either "positive" or "negative" [24]. The key objective of the SA is to present the output, where the unstructured text is converted into meaningful information. In this research study, SA is carried out on collected tweets [25]. All the above five steps are processed in this study to extract the particular text features from the collected tweets. Then, the proposed algorithm based on aspect-based extraction techniques are used to classify the sentiment on extracted data. Fig. 1 shows the working procedure of aspect-based BGRU methodology.

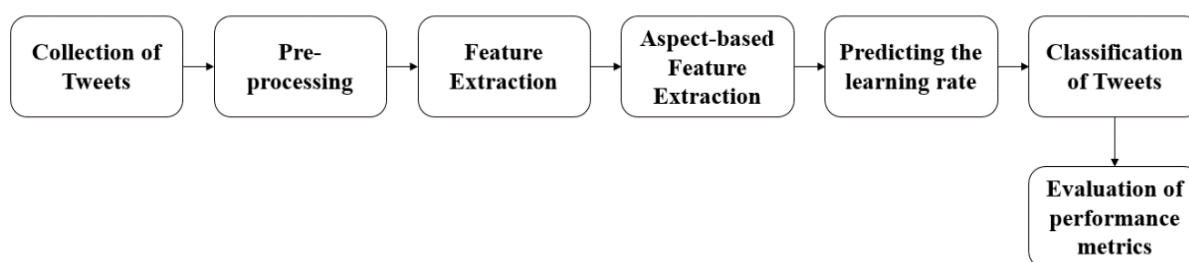


Figure.1 Working procedure of proposed methodology

In general, the computational study of opinions, emotions and sentiments expressed in the text is defined as SA. In this research, the novel classifier that is GRU based classifier is used for SA and identification of emotions in Twitter messages. There are two major parts presents in this system, namely classifier and feature extraction. Initially, the features are extracted by using LDA and TF-IDF, then the important features based on aspects are extracted by aspect-based feature extraction technique. Finally, the combination of self-attention layer with BGRU is designed to classify the tweets. The phases of this aspect-based BGRU architecture is given as follows,

- Initially, the tweets are collected from the Python Tweepy software package. Then pre-processing is performed over the acquired twitter data.
- The adequate features are extracted from the twitter data is processed by using the aspect based feature extraction technique.
- The self-attention is combined with BGRU classifier is used in the research study for classifying the given inputs as positive, neutral or negative.

This combination of feature extraction and classifier gives a better result for classifying sentiment. The description of the modules used in this aspect-based BGRU architecture is given as follows.

### 3.1 Tweets collection and pre-processing

The Python Tweepy software package, i.e. <https://pypi.python.org/pypi/tweepy> is used to collect the information on a Zomato service from Twitter. The collected tweets are stored in the SQL Server 2016, where those tweets are classified into either positive, negative or neutral classes and stored in the database of SQL Server 2014. In addition, the retweets data are also considered for this study, where Table 1 shows the detail description of collected datasets.

In the next step, these collected tweets undergo for pre-processing step to remove the stop words, punctuations, unwanted information and so on. Normally, real-time Twitter contains more noises and

Table 1. Database description

Tweets about Topics / Service	Zomato
Total Number of Tweets	4449
Total Number of Re-tweets	1111
Total Number of Positive Tweets	1288
Total Number of Negative Tweets	1134
Total Number of Neutral Tweets	2027
Total Number of Users	3137
Length of tweets	5000

no sentiment information namely URL, hashtags (#), irregular words, short words, account id (@), etc. In tweet cleaning step, this no sentiment information is removed, where if features are identified as slang term or abbreviation, then it should be removed by using slang filtering. The python-based NLTK tokenizer is used to perform the tokenization operation, where the tweets are categorized into chunks of texts. In order to identify and remove those tokenized words from the collected tweets, a stop words' predefined list is used.

After the pre-processing, the final results are the vectorized form of tweets and its keywords without any unwanted information, which is used as an input to extract the features. However, the tweets polarities must be identified before the extraction of important features. In order to improve the classification performance, this research study determined the tweets polarities by using text blob [26] method, where the traditional techniques identified the polarities only after extracting the useful information. The basic features of natural-language processing are comes with text, blob that are used for identifying the subjectivity calculation and tweets polarities. In this proposed method, the sentences of tweets are categorized into each letter and then, polarities are determined, where the existing techniques identified the polarities by using the full sentences of tweets. Hence, the speed of the proposed method is improved for final classification of tweets.

### 3.2 Feature extraction

In the next step, the features are generated from the pre-processed tweets, where overall scoring of the words, tag count in text, score and frequency of both positive and negative words are presented in the feature extraction technique. In this research study, two various kinds of features extractions are carried out, i.e. aspect-based feature extraction and TF-IDF based extraction with LDA. The below section provides the different types of extraction.

#### 3.3.1. Feature extraction based TF-IDF and LDA

The words are matched with the documents that are related to the queries by using TF-IDF, which is a simple and efficient algorithm. Then, a highly related documents are returned for a specific query by TF-IDF approach. The mathematical Eq. (1) and (2) shows the formula for finding the terms that frequently appear in the documents.

$$\text{Term frequency} = \frac{\text{No. of times term (t) appears in a document}}{\text{Total No. of terms in a document}} \quad (1)$$

Table 2. Parameter notation

Parameter	Description
$\pi$ and $\mu$	Parameters of Dirichlets
$\mathfrak{K}$	Topic variables
$N$	Number of words
$x$	Topic assignment of each word
$y$	Observed words
$M$	Number of documents
$D$	Corpus Generation
$p(\cdot)$	Probability of Corpus
$x^T$	Input vector of BGRU
$h^T$	Hidden vector of BGRU
$y^T$	Output vector of BGRU
$\odot$	Element-wise multiplication
$r$	Reset gate
$z$	Update gate
$W$	Weight of the gate
$m$	Memory state
$t$	Time step
$\hat{c}_t$	Present state of single memory cell
$In_t$	Input gate activation function
$For_t$	Forget gate activation function
$Out_t$	Output gate activation function
$S$	Logistic sigmoid function

$$\text{Inversedocument frequency} = \log \frac{\text{Total no.of documents}}{\text{No.of documents with term } (t)} \quad (2)$$

After the feature extraction, the extracted tweets are initialized for three different categories includes positive, neutral and negative by using LDA technique [27]. It is also defined as a probabilistic topic model, where the labelling of a random mixture of latent topics is carried out for every word. These random mixtures are used to determine the underlying structure of the observed data. The two-phase process is generated for words that are presented in every document, where the distributed topics are selected for each document is conducted in the initial phase. Table 2 presents the notations of the equations used in the research study.

In LDA, the two parameters include  $\pi$  and  $\mu$  are examined for three layered representations during the generation of corpus, then the investigation of topic variables for each document are conducted. In order to obtain the variables of word level, each document word is investigated, where the generative process of LDA is defined as joint distribution. The determination of the random variable for probability density function's  $k - \text{dimensional dirichlet}$  is explained by Eq. (3). Moreover, the calculation of joint distribution for topic mixture and corpus probability is explained in the Eqs. (4) and (5).

$$p(\mathfrak{K}|\pi) = \frac{\Gamma(\sum_{i=1}^k \pi_i)}{\prod_{i=1}^k \Gamma(\pi_i)} \mathfrak{K}_1^{\pi_1-1} \dots \dots \mathfrak{K}_k^{\pi_k-1} \quad (3)$$

$$p(\mathfrak{K}, x, y|\pi, \mu) = p(\mathfrak{K}|\pi) \prod_{n=1}^N p(x_n|\mathfrak{K}) p(y_n|x_n, \beta) \quad (4)$$

$$p(D|\pi, \mu) = \prod_{d=1}^M \int p(\mathfrak{K}_d|\pi) \times (\prod_{n=1}^{N_d} \sum_{x_{dn}} p(x_{dn}|\mathfrak{K}_d) p(y_{dn}|x_{dn}, \mu)) d\mathfrak{K}_d \quad (5)$$

Where,  $p(\cdot)$  probability of corpus, topic variables of document-level are illustrated as  $\mathfrak{K}$ , parameters of dirichlet are denoted as  $\pi$  and  $\mu$ ,  $M$  is used to represent the document,  $N$  is defined as number of words, generation of corpus is illustrated as  $D$ , topic assignment of each word is illustrated as  $x$  and the observed word is presented as  $y$ . In a document, the inference of the posterior distribution for hidden variable is estimated by using the LDA model, which is also defined as the problem of intractable. The individual weighted values for three classes are stored in a dictionary, where these weighted values are used for the testing phases while coordinated with the dictionary.

### 3.3.2. Aspect-based feature extraction

The sentiment of an overall text is identified by using the SA, where different aspects of each text are determined by aspect-based SA. In other words, the specific sentiments with various aspects of service or products are used to identify the input into either positive or negative sentiments. The main aim of the aspect based feature extraction is to determine the subjective term orientation in the text and extract the opinions from the text, which is used to identify whether the text has positive or negative aspects. The aspect-based feature extraction technique is categorized into four important tasks that uses the customer reviews as input and performed the four tasks to provide the final output. In order to determine the associated opinion of each aspect, the entities (i.e. opinions and aspects) of the product are mined in the first task. According to the similarity values, the aspects are grouped in the second task and then, the most popular aspect sentences are chosen in the third task. At last, an opinionated summary is generated in the final task, which depends on product aspects. The following Table 3 shows the overall percentage of three classes for different dataset length. Fig. 2 shows the total number of positive, negative and neutral tweets for 3000 dataset length.

Table 3. Percentage of all three classes

Percentage of Classes	Dataset Length				
	1000	2000	3000	4000	5000
Positive	41.40 %	34.30 %	31.04 %	29.40 %	28.95 %
Negative	14.40 %	17.00 %	20.83 %	49.58 %	45.56 %
Neutral	44.20 %	48.70 %	48.13 %	21.03 %	25.48 %

Sentiment Analysis for 3000 dataset length

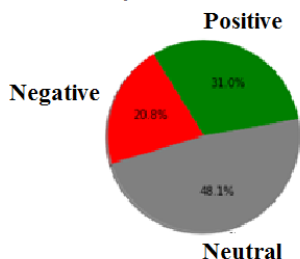


Figure.2 Pie-chart of three classes.

The final step is to classify the sentiments from these inputs, which are illustrated in the next section. The pseudo code of proposed algorithm is given in Algorithm 1.

**Algorithm 1: Pseudo-code of Aspect-based BGRU with self-attention layer.**

```

1. Procedure: BGRU
2. Class:
BGRU[tfidf_features,Senti_Analysis,Collected_Tweets]
3. Tfidf ← tfidf_vectorizer
   fit_transform(Collected_Tweets) // Extracting the Aspect features using TF-IDF
4. SentiAnalysis ← Analysis(Collected_Tweets) // Calculate Sentiment
   (X_train,X_test,Y_train,Y_test)
   ← train_test_split(Aspect_features, Senti_Analysis, test size = 0.3) // Splits the data into 70% for training and 30% for testing
5. X_train = sequence.pad_sequences(X_train, max_length=1000)
   X_test = sequence.pad_sequences(X_test, max_length=1000)
6. def model:
//BGRU Model Initialization
   Model ← Sequential()
   Adding SelfAttention Layer
   ← SeqSelfAttention(attention_activation = 'softmax')

```

```

Model Compile ← model.compile('Nadam', 'sparse_categorical_crossentropy' metrics = ['accuracy'])
ModelFitting ← model.fit(X_train, batch_size = batch_size, n_epochs = 5, validation_data = [X_test, Y_test])
Model Prediction
← model.prediction_classes(Y_test)
7. End function
8. def Sentiment:
   Sentiment=
   TextBlob(Collected_Tweets)
   If Sentiment.polarity is less than 0:
   return 1
   else if Sentiment.polarity is equal to 0:
   return 0
   else:
   return -1
9. End function
10. End Class

```

**3.3 Classification of Tweets**

The final step is to classify the tweets as either positive or negative by using BGRU with self-attention layer. A special kind of Recurrent Neural Network (RNN) is developed in this work called GRU network to classify the tweets in SA. The GRU solves the problem of vanishing gradients that normal RNN unit suffers from, they do it by implementing a memory cell within their network, and this enables them to store data from early within the sequence to be used later within the sequence. When compared with the standard LSTM model, the final GRU model is simple and faster. For instance, a lot of time is saved with limited performance of the GRU model, when training the big data.

Consider the sequence of input vectors as  $x^1, x^2, \dots, x^T$  the corresponding hidden activations are computed by GRU as  $h^1, h^2, \dots, h^T$  and final output as vector sequence is illustrated as  $y^1, y^2, \dots, y^T$ . The element-wise multiplication of two vectors is denoted by  $\odot$ , the hidden state  $h$  and output vector  $y$  are computed by single layer GRU. Fig. 3 shows the diagrammatic representation of BGRU.

The following Eq. (6-10) shows the computations:

$$r^t = \sigma(W_r x^t + U_r h^{t-1} + b_r) \tag{6}$$

Where  $r$  is reset gate and  $W$  is weight of the gate

$$z^t = \sigma(W_z x^t + U_z h^{t-1} + b_z) \tag{7}$$



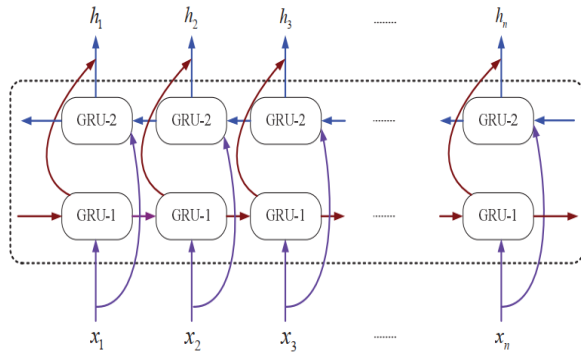


Figure.3 Block Diagram of BGRU

Where  $z$  is update gate

$$\bar{h}^t = \tanh(W_h x^t + U_h (r^t \odot h^{t-1}) + b_h) \quad (8)$$

$$h^t = z^t \odot h^{t-1} + (1 - z^t) \odot \bar{h}^t \quad (9)$$

Where  $h$  is hidden state

$$y^t = \sigma(W_o h^t + b_o) \quad (10)$$

Where, element-wise sigmoid functions are depicted as  $\sigma(\cdot)$ ,  $y^t \in [0,1]^L$  and the number of tweets is represented as  $L$ . The probability of a certain tweets happening at time  $t$  is illustrated by each dimension of  $y^t$ . Likewise, dimension of  $d$  can be described as a hidden state of the GRU, then  $r^t, z^t, h^t, \bar{h}^t \in \mathbb{R}^d$ . In this study, a BGRU is exploited, since it makes full use of the context information from both directions. The end-to-end training is carried out by using Binary Cross Entropy loss (BCE) loss, which is illustrated as Eq. (11).

$$\text{loss}(y^t, \hat{y}^t) = -\frac{1}{L} \sum_{i=1}^L [\hat{y}_i^t \log(y_i^t) + (1 - \hat{y}_i^t) \log(1 - y_i^t)] \quad (11)$$

Where, binary indicator of tweets is defined as  $\hat{y}^t \in \{0,1\}^L$ . Here, the learning efficiency of the BGRU classifier is enhanced by incorporating the self-attention layer.

### 3.3.1. Self-attention mechanism

The two important decisions namely maintenance of current input of extracted features and previous state are effectively handled by GRU, which sets an adaptive gating mechanism. Then, the unclassified sentence is used as an input and obtained the implicit vector by encoding the sentences in many time steps. Let assume the sample sequence as  $S =$

$\{x_1, x_2, \dots, x_t\}$ , where the length of the input sentence is denoted as  $l$ , then, a training of the model is carried out by GRU. The hidden states as  $h_t$  and memory state as  $m_t$  are updated for each time step  $t$ . the following Eq. (12-14) are shown as follows:

$$\begin{pmatrix} In_t \\ FO_t \\ Out_t \\ \hat{c}_t \end{pmatrix} = \begin{pmatrix} S \\ S \\ S \\ T \end{pmatrix} W \cdot [h_{t-1}, i_t] \quad (12)$$

$$m_t = FO_t \odot m_{t-1} + In_t \odot \hat{c}_t \quad (13)$$

$$h_t = Out_t \odot T(m_t) \quad (14)$$

From the equations, input is described as  $i_t$  and single memory cell's present state is denoted as  $\hat{c}_t$  memory cell. The results of input gate activate function is illustrated as  $In_t$ , results of forget gate activation function is presented as  $FO_t$  and finally, the results of the output gate activate function is denoted as  $Out_t$ . Moreover, the function of common logistic sigmoid as  $S$  that is widely used as represents  $\tanh$  and activate the function. In a sentence, the most important semantic information is captured without using any extra knowledge by focusing on the words that has a decisive effect on classification, which is the major contribution of the BGRU with attention mechanism. Therefore, the final output of the experiment is used to determine whether the tweet is positive, neutral or negative. The validation of aspect based BGRU in terms of various parameters is given in below section.

## 4. Result and discussion

In this section, the experiments are conducted on collected tweets to test the efficiency of proposed aspect-based BGRU by means of error rate includes RMSE and MSE, accuracy, recall, precision and f-score. The computer with Intel Core i5 processor, 8GB RAM with 2.2 GHz is used to implement the proposed method in Python 3.7.3 for the experimental setup. The proposed method considered 70% of collected data as training and remaining 30% data as testing data in the validation process. The existing techniques are validated on either publicly available dataset or collected dataset about politics, but this research study worked on collected tweets of Zomato service. Therefore, the proposed aspect-based BGRU validation are based on quantitative and qualitative analysis, which are explained as below.

### 4.1 Parameter evaluation

The performance of the proposed method is tested by using the major parameters namely F-measure, RMSE, accuracy, MSE, precision and recall. The Eq. (15) shows the mathematical equation for precision, which is used to measure the true positive tweets, where a portion of positive tweets are determined by recall for a given class  $c_i$  is explained in Eq. (16).

$$Precision = \frac{|Tweets\ correctly\ classified\ to\ class\ c|}{|Total\ tweets\ in\ class\ c|} \tag{15}$$

$$Recall = \frac{|Tweets\ correctly\ classified\ to\ the\ class|}{|Total\ tweets\ in\ class\ c|} \tag{16}$$

In order to determine the SA, the measurement of statistical variability and random errors are used to define the overall accuracy, which is explained in Eq. (17).

$$Accuracy = \frac{|Total\ number\ of\ correctly\ classified\ tweets|}{|Total\ number\ of\ tweets|} \tag{17}$$

The mathematical Eq. (18) provides the score calculated by considering the recall and precision of the test and then, accuracy is measured.

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{18}$$

In addition to this parameter, the error rate includes RMSE and MSE are calculated as shown in the Eq. (19-20).

$$MSE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_t - y_t|^2 \tag{19}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^T |\hat{y}_t - y_t|^2}{T}} \tag{20}$$

### 4.2 Quantitative analysis of proposed aspect-based BGRU method

In this below section, the validation of the proposed method is carried out on various dataset length in terms of different parameters. Initially, the proposed aspect-based BGRU method is tested for 1000 dataset length, which are illustrated in Table 4.

Table 4. Validation results of proposed method for 1000 dataset length

Parameters (%)	Without Aspect level based features extraction method		With Aspect level based features extraction method	
	BGRU without self-attention Layer	BGRU with self-attention Layer	BGRU without self-attention Layer	Proposed BGRU with self-attention Layer
Accuracy	62.00	77.20	72.40	80.40
Precision	58.0	77.0	73.0	81.0
Recall	62.0	77.0	72.0	80.0
F-Score	56.0	77.0	73.0	80.0
MSE	0.81	0.49	0.54	0.42
RMSE	0.90	0.70	0.70	0.65

Initially, the proposed method is tested with aspect-level features and without aspect-level features, where self-attention mechanism is used to improve the learning rate of BGRU. The BGRU with self-attention layer achieved less accuracy (i.e. 77.20%) while validating without aspect-based features, where the proposed method achieved high accuracy (i.e. 80.40%), when BGRU is incorporated with aspect-based features. Likewise, the BGRU achieved an accuracy of 72.40% than BGRU without aspect-based features (i.e. 62.00%). As like accuracy, the proposed BGRU with self-attention achieved higher precision, recall and f-score than BGRU without self-attention layer in terms of aspect-based feature level. From the Table 3, it is clearly stated that BGRU without aspect-level features and self-attention layers provides poor performance in terms of all parameters. This is because aspect-level features classify the tweets based on specific terms in sentiments than whole sentiments. Table 5 shows the validated results of the proposed method for 2000 dataset length.

When the length of dataset increases, the proposed method provided better performance in terms of all parameters. For instance, the aspect-based BGRU with self-attention achieved 88% f-score for 2000 dataset length, the same method achieved only 80% f-score for 1000 dataset length. This shows that the length of dataset does not affect the performance of proposed method. In addition, the aspect-based BGRU with self-attention achieved 0.25% MSE and 0.50% RMSE, where the aspect-based BGRU without self-attention achieved high MSE (i.e. 0.49%) and



Table 5. Performance of proposed method for 2000 dataset length.

Parameters (%)	Without Aspect level based features extraction method		With Aspect level based features extraction method	
	BGRU without self-attention Layer	BGRU with self-attention Layer	BGRU without self-attention Layer	Proposed BGRU with self-attention Layer
Accuracy	64.0	86.0	75.83	87.68
Precision	53.0	86.0	77.0	88.0
Recall	64.0	86.0	76.0	88.0
F-Score	57.0	86.0	76.0	88.0
MSE	0.68	0.25	0.49	0.25
RMSE	0.83	0.51	0.70	0.50

Table 6. Results of proposed method for 3000 dataset length

Parameters (%)	Without Aspect level based features extraction method		With Aspect level based features extraction method	
	BGRU without self-attention Layer	BGRU with self-attention Layer	BGRU without self-attention Layer	Proposed BGRU with self-attention Layer
Accuracy	68.22	88.45	82.88	88.67
Precision	72.0	88.0	83.0	89.0
Recall	68.0	88.0	83.0	89.0
F-Score	63.0	88.0	83.0	89.0
MSE	0.69	0.23	0.34	0.20
RMSE	0.83	0.48	0.58	0.45

RMSE (i.e. 0.70%) for 2000 dataset length. This is because, the learning rate of BGRU is highly improved by using self-attention layer, which correctly predicted the tweets into positive, negative and neutral. The following Table 6 shows the results of proposed method for 3000 dataset length by means of MSE, RMSE, accuracy, precision and recall.

In terms of without aspect-level features, BGRU without self-attention achieved low precision (i.e. 72%) with high RMSE (i.e. 0.83%). But, BGRU with self-attention achieved 88% precision and 0.48% RMSE, this is due to the incorporation of self-attention layer in BGRU. The same BGRU with self-attention method achieved 89% precision and 0.45% RMSE due to the presence of aspect-based features. When aspect-based feature extraction techniques are used, the BGRU correctly classified the tweets into positive, negative and neutral using specific terms in the sentiments. However, there is a slight difference occurs in BGRU with aspect-based features for all parameters, when it is processed without self-attention layer. For instance, the BGRU without self-attention method achieved only 83% precision and 0.58% RMSE in terms of aspect-based feature extraction technique. In addition, the proposed method achieved 88.67% accuracy for 3000 dataset length and only 87.68% accuracy for 2000 dataset length. This shows that the performance of proposed aspect-based BGRU is not affected when the length

of dataset is increased. Not only accuracy, the other parameters such as RMSE, MSE, precision, recall and F-score are also improved in the proposed method. The experimental results of proposed method for 4000 tweets length is illustrated in Table 7.

In terms of without self-attention layer, the BGRU achieved recall of 76%, when aspect-level features are not considered. But, the same BGRU achieved recall of 79%, while incorporated with aspect-level features. The adequate features are extracted by aspect-level feature extraction techniques, which will automatically improve the performance of BGRU. In terms of with self-attention layer, the performance of BGRU achieved 88% recall, where BGRU shows poor performance in the analysis of without self-attention layer and it is proved that self-attention mechanism also plays major role in BGRU performance. In addition, the error rate includes MSE and RMSE are decreased when BGRU is incorporated with aspect-based features and self-attention mechanism. For instance, the BGRU method achieved only 0.23% of MSE and 0.48% of RMSE, where the BGRU method without aspect-based features and self-attention mechanism achieved 0.56% of MSE and 0.75% of RMSE. Table 8 provides validated results of the proposed method for 5000 dataset length.

Table 7. Performance of proposed method for 4000 dataset length

Parameters (%)	Without Aspect level based features extraction method		With Aspect level based features extraction method	
	BGRU without self-attention Layer	BGRU with self-attention Layer	BGRU without self-attention Layer	Proposed BGRU with self-attention Layer
Accuracy	75.83	88.42	79.08	88.75
Precision	75.0	89.0	80.0	89.0
Recall	76.0	88.0	79.0	89.0
F-Score	75.0	88.0	79.0	89.0
MSE	0.56	0.24	0.53	0.23
RMSE	0.75	0.49	0.73	0.48

Table 8. Validated results of proposed Method for 5000 dataset length

Parameters (%)	Without Aspect level based features extraction method		With Aspect level based features extraction method	
	BGRU without self-attention Layer	BGRU with self-attention Layer	BGRU without self-attention Layer	Proposed BGRU with self-attention Layer
Accuracy	77.13	88.68	79.40	88.79
Precision	78.0	89.0	80.0	89.0
Recall	77.0	88.0	79.0	89.0
F-Score	77.0	88.0	79.0	89.0
MSE	0.52	0.24	0.48	0.21
RMSE	0.71	0.49	0.69	0.46

Table 9. Comparative analysis of proposed BGRU with existing techniques

Parameters (%)	Without Aspect level based features extraction method				With Aspect level based features extraction method			
	Standard LSTM	BLSTM	Standard GRU	BGRU	Standard LSTM	BLSTM	Standard GRU	Proposed BGRU
Accuracy	49.30	62.70	62.30	77.20	77.20	78.8	78	80.4
Precision	24	64	67	77	77	80	79	81
Recall	49	63	62	77	77	79	78	80
F-score	33	63	57	77	76	79	78	80
MSE	0.88	0.78	0.75	0.49	0.48	0.44	0.48	0.42
RMSE	0.94	0.88	0.87	0.70	0.69	0.66	0.69	0.65

When the length of dataset increases, the performance of proposed aspect-based BGRU achieved better performance in terms of accuracy, precision, recall and f-score and it degrades in terms of RMSE and MSE. For instance, the error rate of proposed method is highly decreased from the dataset length from 1000 to 3000, where the length increases, the proposed BGRU achieved high RMSE and high MSE. This is due to the proportion of the total number of positive, negative and neutral tweets in the dataset length of 4000 and 5000. Table 3 clearly stated that the total number of positive and neutral tweets is highly decreased in 4000 and 5000 dataset length, which increases the error rate of proposed aspect-based BGRU.

In addition, the validation of precision, recall and f-score are stable from the dataset length of 3000 to 5000, i.e. the proposed method achieved 89% of

recall, precision and f-score. But, accuracy shows slight variation for all the dataset length. The proposed aspect-based BGRU achieved 80.40% accuracy for 1000 dataset length and 88.79% accuracy for 5000 dataset length. In the next section, the qualitative analysis of proposed method when compared with tradition techniques namely standard LSTM, Bi-directional LSTM as BLSTM and standard GRU.

#### 4.3 Qualitative analysis of proposed aspect-based BGRU method

In the below section, a validation of proposed method with other existing techniques is compared, which are depicted in Table 9. Here, the research

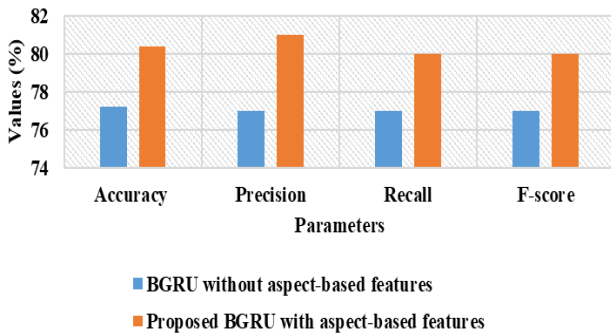


Figure.4 Graphical illustration of BGRU with- and without-aspect based features

Table 10. Comparative Analysis of Proposed Method in terms of Positive, Negative and Neutral Tweets

Parameters	BGRU without Aspect-based features			Proposed BGRU with Aspect-based features		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Accuracy	75	68	72	93.19	92.51	92.51
Precision	87	65	74	89.00	89.00	92.00
Recall	52	94	47	96.00	87.00	82.00
F-Score	65	77	58	88.00	92.00	87.00

study considered BLSTM, standard LSTM, GRU and BGRU only with self-attention mechanism for validation process. The existing techniques namely standard LSTM, GRU and BLSTM are implemented on the collected tweets of Zomato services to test the efficiency of proposed aspect-based BGRU.

Here, the extensive experiments are conducted only on 1000 dataset length to test the efficiency of proposed method with other existing techniques. The results stated that the proposed BGRU achieved better performance than other techniques, while incorporated with aspect-based feature extraction technique. For instance, the standard LSTM achieved 77.20% accuracy, 77% f-score and recall, where the proposed BGRU achieved 80.4% accuracy, 80% f-score and recall, when aspect-based feature extraction is included. In addition, the RMSE and MSE of proposed BGRU method is highly reduced than other existing techniques. When compared with BGRU without aspect-based technique, the proposed BGRU with aspect-based technique provided higher performance in terms of all parameters with less error rate. Fig. 4 shows the graphical representation of accuracy, precision, recall and F-score for BGRU with- and without-aspect based feature extraction technique.

Even the standard LSTM achieved better performance, when incorporated with aspect-based features. The results stated that the standard LSTM without aspect-based features achieved only 24% precision and 0.94% RMSE, where the same technique achieved 77% precision and 0.69% RMSE while aspect-based features are included in the analysis. This showed that aspect-based features extract only the adequate information from the collected tweets rather than all sentiments. The proposed aspect-based BGRU technique achieved 80% of recall and f-score with 0.42% MSE, while comparing with standard LSTM, BLSTM and GRU. In the analysis of aspect-based features, LSTM and GRU achieved nearly 77% of recall and f-score with 0.48% MSE. From the Table 8, it is clearly stated that the proposed aspect-based BGRU achieved high performance, because it automatically concentrates on the words for capturing the most important semantic information in sentence without additional knowledge. When trained with big data, the proposed method solves the gradient problem and stored the data within the sequence that makes faster than existing techniques. Table 10 provides the experimental results of proposed BGRU for each class: Positive, Negative and Neutral for the dataset length 5000, where Fig. 5, 6 and 7 shows the graphical representation for the analysis for positive, negative and neutral tweets.

From the Table 10, it is clearly stated that the proposed BGRU with aspect-based features achieved better performance in terms of all parameters for every class. For instance, the proposed BGRU achieved nearly 92%-93.19% accuracy for all three classes, where the BGRU without aspect-based features achieved only 68%-75% accuracy. The proposed BGRU improved the F-score nearly 40-48% of neutral tweets, while compared with BGRU without aspect-based features. This is because of extracting the specific terms in the tweets accurately by using the aspect-based feature extraction technique. The aspect-level features in BGRU with self-attention layer enhanced the discriminating power of classification in tweets, which is proven by the validated results. Moreover, the aspect-based BGRU reduced the challenges on the dimensionality of tweets and spam message's classification. However, the proposed method is insufficient to identify the sentiment of conflicting, complex and compound tweets. This developed method should be extended with some other techniques to classify the compound sentiments and conflicting tweets.

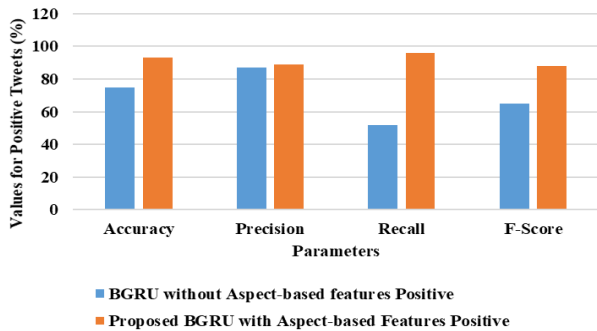


Figure.5 Analysis of Proposed Method for Positive Tweets

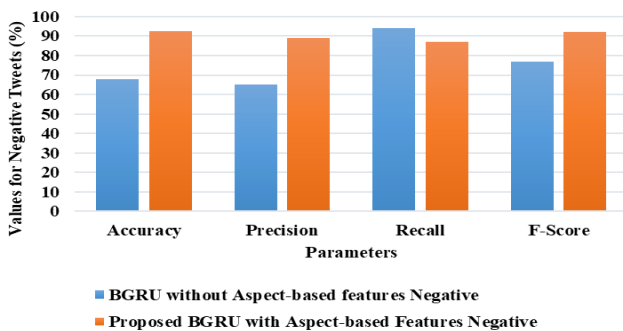


Figure.6 Performance of Proposed BGRU for Negative Tweets

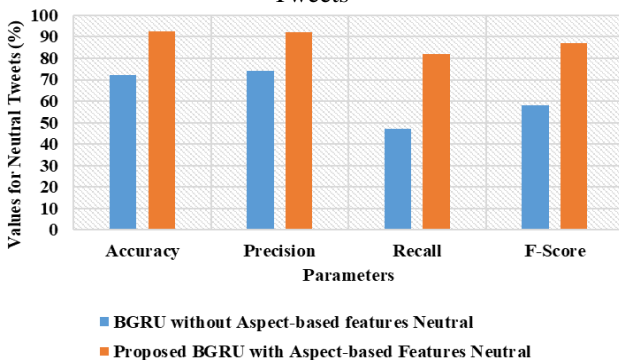


Figure.7 Comparative analysis of Proposed BGRU for Neutral tweets

Table 11. Performance Analysis of Proposed Method with Existing QSR Technique

Parameter Metrics (%)	Dataset Length			
	1000		3000	
	QSR [21]	Proposed BGRU	QSR [21]	Proposed BGRU
Accuracy	65.28	<b>80.40</b>	71.85	<b>88.67</b>
Precision	65.33	<b>81.00</b>	71.08	<b>89.00</b>
Recall	65.20	<b>80.00</b>	70.67	<b>89.00</b>
F-Measure	65.26	<b>80.00</b>	70.83	<b>89.00</b>

#### 4.4 Comparative analysis of proposed method with existing techniques

In this section, the performance of proposed BGRU with aspect-based feature is compared with

Table 12. Comparative Analysis of Proposed Method with Existing Techniques

Parameter Metrics (%)	Dataset Length			
	2000		5000	
	DCNN [17]	Proposed BGRU	HL-NBC [19]	Proposed BGRU
Accuracy	87.39	<b>87.68</b>	82.00	<b>88.79</b>
Precision	88.00	<b>88.00</b>	75.00	<b>89.00</b>
Recall	87.32	<b>88.00</b>	72.00	<b>89.00</b>
F-Measure	87.66	<b>88.00</b>	75.00	<b>89.00</b>

existing techniques namely DCNN [17], HL-NBC [19] and QSR [21] in terms of accuracy, precision, recall and F-measure. Initially, Table 11 shows the comparative analysis for the data length 1000 and 3000, where QSR technique is compared with proposed method.

From the Table 11, the existing QSR techniques provide poor performance for the dataset length 1000 and 3000. For instance, the QSR achieved nearly 65% and 71% of accuracy and precision for 1000 and 3000 dataset length, where the proposed method achieved nearly 80% and 89% of accuracy and precision for the same dataset length. If the dataset length increases, the QSR fails to achieve better performance, which degrades the performance. Table 12 provides the comparative results of the proposed method with DCNN and HL-NBC for the dataset length 2000 and 5000.

From the Table 12, it can be conclude that the proposed method slightly increased the performance of sentiment analysis than DCNN method for the dataset length 2000. This is because, DCNN used the unlabelled data for extracting the information from the tweets and pre-processing technique is not conducted by the DCNN approach. The proposed method removed the unlabelled data from the collected tweets, which increases the performance of BGRU with aspect-based features. Moreover, the existing HL-NBC is compared with the proposed approach for the dataset length 5000. The sarcastic tweets are present in the HL-NBC that degrades the performance and failed to filter those tweets. The proposed BGRU technique effectively filters the sarcastic tweets and achieved nearly 88% to 89% of accuracy, precision, recall and F-Measure. This section clearly shows that the proposed BGRU with aspect-based features performs better performance than the existing techniques for various dataset lengths.

#### 5. Conclusion

In the past few years, the researchers focused on SA, because it is a challenging and interesting task

when compared to other tasks. Twitter is a micro-blogging services, which helps the organization to evaluate the customer opinions about products or services by analyzing their posts, reviews and blogs. In this research study, an aspect-based BGRU is developed to analyze the sentiment of users from the collected tweets. Initially, the TF-IDF and LDA is used to extract the important features of the pre-processed tweets. Then, the orientation of subjective terms is extracted by aspect-based extraction. Finally, the BGRU classified those tweets into “positive”, “negative” and “neutral” tweets, where the learning rate of BGRU are highly improved by incorporating with self-attention layer. The experiments are conducted on collected tweets using Twitter API to validate the efficiency of proposed aspect-based BGRU. The validated results stated that the method achieved 87.68% accuracy, 88% of precision, recall and f-score, 0.50% RMSE and 0.25% MSE for 2000 dataset length. The proposed aspect-based BGRU method focused only on Zomato tweets and several words are extracted by using aspect-based technique that leads to high training time and may be information loss. Therefore, the new methodologies are designed to analyze the sentiments of dynamic datasets with less training time as a future work.

### Conflicts of Interest

The authors declare no conflict of interest.

### Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1<sup>st</sup> author. The supervision, and project administration, have been done by 2<sup>nd</sup> author.

### References

- [1] T. A. Rana, and Y. N. Cheah, “Aspect extraction in sentiment analysis: comparative analysis and survey”, *Artificial Intelligence Review*, Vol. 46, No. 4, pp. 459-483, 2016.
- [2] H. S. Manaman, S. Jamali, and A. AleAhmad. “Online reputation measurement of companies based on user-generated content in online social networks”, *Computers in Human Behavior*, Vol. 54, pp. 94-100, 2016.
- [3] M. Z. Asghar, A. Khan, F. Khan, and F. M. Kundi, “Rift: a rule induction framework for twitter sentiment analysis”, *Arabian Journal for Science and Engineering*, Vol. 43, No. 2, pp. 857-877, 2018.
- [4] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, and I. Chorbev, “Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages”, *Multimedia Tools and Applications*, Vol. 77, No. 24, pp. 32213-32242, 2018.
- [5] A. V. Mohan Kumar and A. N. Nandakumar, “Sentimental Analysis using Robust Hierarchical Clustering Algorithm for Opinion Mining on Movie Reviews-based Applications”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 8, No. 8, pp. 452-457, 2019.
- [6] G. S. Nijaguna and T. Kampalappa, “Imbalance Data Classification Using Local Mahalanobis Distance Learning Based on Nearest Neighbor”, Vol. 1, No. 2, pp. 1-9, 2020.
- [7] A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, and M. Hazim, “Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms”, *IEEE Access*, Vol. 7, pp. 83354-83362, 2019.
- [8] N. Indurkha and F. J. Damerau, “Handbook of natural language processing”, *Chapman and Hall/CRC*, 2010.
- [9] K. Elshakankery and M. F. Ahmed, “HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis”, *Egyptian Informatics Journal*, Vol. 20, No. 3, pp. 163-171, 2019.
- [10] A. Pak and Patrick Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining”, *LREc*. Vol. 10, No. 2010, pp. 1320-1326, 2010.
- [11] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis”, *Mining text data*, Springer, Boston, MA, pp. 415-463, 2012.
- [12] G. Somprasertsri and P. Lalitrojwong, “Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization”, *J. UCS*, Vol. 16, No. 6, pp. 938-955, 2010.
- [13] E. Cambria, R. Speer, C. Havasi, and A. Hussain, “Senticnet: A publicly available semantic resource for opinion mining”, *In 2010 AAAI Fall Symposium Series*, 2010.
- [14] M. Bouazizi and T. Ohtsuki, “A pattern-based approach for multi-class sentiment analysis in Twitter”, *IEEE Access*, Vol. 5, pp. 20617-20639, 2017.
- [15] S. Haccianella, A. Esuli, and F. Sebastiani, “SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining”, In: *Proc. of the Seventh conference on*

*International Language Resources and Evaluation*, 2010.

- [16] F. H. Khan, U. Qamar, and S. Bashir, "A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet", *Knowledge and Information Systems*, Vol. 51, No. 3, pp. 851-872, 2017.
- [17] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis", *IEEE Access*, Vol. 6, pp. 23253-23260, 2018.
- [18] M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme", *Expert Systems*, Vol. 35, No. 1, pp. e12233, 2018.
- [19] A. P. Rodrigues and N. N. Chiplunkar, "A new big data approach for topic classification and sentiment analysis of Twitter data", *Evolutionary Intelligence*, pp. 1-11, 2019.
- [20] L. Wang, J. Niu, and S. Yu, "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis", *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [21] Y. Zhang, D. Song, P. Zhang, X. Li, and P. Wang, "A quantum-inspired sentiment representation model for twitter sentiment analysis", *Applied Intelligence*, pp.1-16, 2019.
- [22] L. Terán and J. Mancera, "Dynamic profiles using sentiment analysis and twitter data for voting advice applications", *Government Information Quarterly*, 2019.
- [23] C. Gokulnath, M. K. Priyan, E. V. Balan, K. R. Prabha, and R. Jeyanthi, "Preservation of privacy in data mining by using PCA based perturbation technique", In: *Proc. of the 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, pp. 202-206, 2015.
- [24] P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran, R. Jidhesh, and T. Vadivel, "Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things", *Cluster Computing*, Vol. 22, No. 4, pp. 7733-7744, 2019.
- [25] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system", *Future Generation Computer Systems*, Vol. 82, pp. 375-387, 2018.
- [26] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts", *Mathematical and Computational Applications*, Vol. 23, No. 1, pp. 11, 2018.
- [27] F. Colace, M. De Santo, L. Greco, V. Moscato, and A. Picariello, "Probabilistic approaches for sentiment analysis: Latent dirichlet allocation for ontology building and sentiment extraction", In: *Proc. of the Sentiment Analysis and Ontology Engineering*, Springer, Cham, pp. 75-91, 2016.