(Internship Report)

Duration: 15 June to 01 Aug, 2021

Extracting Emotion Quotient of Information Virality over Twitter Data

Student's Name: Pawan Kumar and Reuben Eappen Reji
National Institute of Technology, Surathkal, India
{pawan.181ee133,reubeneappenreji.181ee136}@nitk.edu.in

Mentor's Name: Dr. Vikram Singh
Computer Engineering Dept., National Institute of Technology, Kurukshetra, India
viks@nitkkr.ac.in

Abstract. In social media platforms, a viral information or trending term (hashtags) draws attention. The viral information asserts the impact of user content towards topic/terms for the duration. In a real-time sentiment analysis, these viral terms could deliver potential insights for the analysis and decision-support. The traditional sentiment analysis tools or strategy generates the level of predefined sentiments over social media content for the defined duration, and lacks in the extraction of emotional impact created by the same. In these settings, it is multifaceted to precisely estimate the emotional quotient viral information creates.

In this project, a novel algorithm is designed (i) to extract the sentiment and emotions quotient of a current viral information over twitter data, (ii) to comparative analysis over two co-occurring trending/viral inf0rmartion, (iii) also in-depth analysis of potential twitter text data via UI. The generated emotion quotients and micro-sentiment reveals several valuable insights of a viral/trending topic, and assists in decision support. A use-case analysis over real-time extracted data asserts significant insights, as generated sentiments and emotional effects reveals co-relations caused by viral/trending information. The algorithm delivers an efficient, robust, and adaptable solution for the sentiment analysis also.

Keywords: Big Data, Emotional Quotient, Sentiment Analysis, Twitter.

1 Introduction

The traditional social media platforms, e.g., Twitter, Facebook, etc. cater to the global users and list their personal information and media. The heterogeneous user data is often utilized for deriving common sentiments or trending information. The trending or viral information primarily harnesses the global content shared co-related to a particular topic and keyword (e.g., hashtags).

For Example: Viral information related to 'Covid-2019' may cause significant emotional effects on the citizens in current time. The cause-effect analysis over twitter may assist in administration (Healthcare offices) to track the sources/persons to take precautionary measures proactively. Similarly, for spectrum of applications where these estimated statistics may play a significant role:

- Sentiment analysis on social media is extensively used in the Stock market and crypto market to see people present state of mind and potential of Panic sell.
- Rough emotional understanding of the city/state/country/and the world as well is there on websites. [https://sentiment-sweep.com/]
- Sentiment analysis gives the Image of practically any object (Person/county/commodity).
- There are models of simulating Political elections using Sentiment analysis as a basis. [There are many Models of US elections and Indian Elections.
- The government could utilize to track civil riot origins before they become uncontrollable, etc.

A naive user or new user usually refers to this trending or viral list of information to see the most occurring or contributory piece of information. In this process, a user simply refers to the viral information or hashtag and explores the related term over the Twitter API, without cognitive awareness of the emotional effect of viral information. A piece of viral information may have a list of information that may trigger the emotional effect on the user and lead to emotional splits or swings on the choice of information. User assistance is pivotal for the user, which may assist the user to showcase the emotional effects that viral information may carry. Though social media platforms offer limited or no functions or aspect-related views on the API for the generic user.

Typically, the designed algorithm for the sentiment analysis and EQ statistics could serve several pivotal objectives, as asserted by the experimental analysis also. The sentiment and EQ statistics generated could be utilized in several application areas: decision-making, advertising, public administrations, etc. Though, generating these statistics for real-time published data from the twitter data is a complex and multifaceted computing task.

1.1 Motivation and Research Questions

The sentiment analysis is a complex computing task, mainly due to the semantic correlation that exists between the user-generated data and targeted level of sentiment. The task becomes multifaceted, primarily, when it is aimed for deriving the 'emotion effect' caused by the sentiments, as a micro-level. In these settings, a strategy could be the need of the hours that acquire the real-time twitter data and deliver the insights.

We have conceptualized the following research questions (RQs) are for the smooth conduct of work in this project:

RQ-I: What are the key twitter data elements/features to extract emotional quotients?
RQ-II: How to estimate the emotional effects caused by viral/trending information?
RQ-III: How to estimate the sentiment levels and co-related to EQ of viral information?
RQ-IV: What could be the advantages in application domains of derived 'emotional effects' and how it could deliver significant insights.

Above RQs are the focus of work conducted and eventually are discussed within the experimental analysis. The detailed discussions on the RQs are over the designed system and its outcome asserts the feasibility of the sentiment and emotional quotient-based analysis over the twitter data for real-time decision making.

1.2 Contributions and Outline

The key contribution of the project work is to design a robust and adaptive sentiment model. An extensive design strategy for the creation of the end-to-end framework is elaborated. The user-interface (UI) is plugged with the designed strategy for the user interaction with the generated analysis and futurist data play. The other contributions of the work are as follows:

- **(i)** An interaction user interface for social media analytics. The UI is portable and could be easily utilized for the underlying analysis.
- **(ii)** The system generates the real-time statistics (emotional and sentiment polarity) for a submitted emotional value 'as query' or to current viral/trending information.
- **(iii)** The design and development of work outlines the importance of several features of social media data for sentiment analysis, such as: subjectivity, polarity of statements and the emotions expressed in a statement.
- **(iv)** The experimental assessment asserts the overall accuracy of the designed system, as 89% and 90 % accuracy respectively for sentiment and emotional quotient generations. The accuracy on intrinsic levels for each is also significant and satisfactory.
- **(v)** The feasibility analysis of generated statistics for real-time analysis and decision-making is uncovered using 04 use-cases, also outlines the key features of social media data for the purpose.

The project report is organized as follows: section 2 lists the relevant research efforts to the sentiment and emotional statistics. The proposed scheme for the objective is elaborated in the section 3, internals of the designed computational works are also described using working example equations and pseudo code. Further, 3.4 elaborate the use-case of the designed system. Section 4 explains the experimental assessment of the designed algorithm based on the several traditional and advanced performance metrics and measures. The conclusion and future work is listed under section 5.

1.3 Methods and Material

The user data correlated to submitted query (keyword/hashtag/ trending/viral information) is the key data to be processed to achieve the desired sentiment score and emotional quotients (EQ) over the UI. The initial preprocessing in social media data (twitter data) is achieved through traditional tools, e.g., LDA, Semantic Analysis (LSA), etc. All these powerful tools are employed for the feature's extraction and text pre-processing prior to statistics generations and visualizations. The designed prototype interacts with twitter API for each corresponding user data request.

The user interface (UI) is designed using *Python library Tkinter* and statistics are using *Python library Matplotlib*for extracted tweet objects for a user request. The real-time extraction of tweets objects using *API* and further cleaned and stored into *Pandas Data Frame*. The number of tweets for extraction is related to user input, as for each user input it is related. For the various input, we have experimented with 1 to 1600 number of tweets extracted on real-time basis on the prototype.

2 Related Work

In recent years, numerous research efforts have been for the development of algorithms and approaches to generate the sentiment scores or levels over the user created content on social networks. Though, research efforts related to generation of sentiments and further extraction of emotional quotation at micro-level on viral or trending information is limited. In efforts related to work conducted in this paper falls fewer than two heads:

2.1 Sentiment Analysis on Viral or Trending Information

In recent years, much research has been conducted on sentiment analysis to develop systems that are more reliable and provide better accuracy. The core task of sentiment analysis is the automatic identification of opinionated text in documents [20]. Previous research used both rule based and statistical machine learning approaches for opinion mining and sentiment analysis [24]. In this section, we briefly discuss some techniques on sentiment analysis and their applications.

Ibrahim et al. [34] presented a detailed survey of different techniques used for opinion mining and sentiment analysis. Turney [38] suggested an unsupervised algorithm, which uses semantic orientation of the phrases for classification of reviews. The lexicon-based approach determines the polarity or sentiment using some function of opinion words in the document or sentence [4, 36]. Esuli et al. [5] developed the SentiWordNet lexicon which contains opinion strength for each term. The feasibility of SentiWordNet lexicon for sentiment classification of documents has been assessed by Ohana [22]. Hamouda et al. [9] used the SentiWordNet Lexicon to classify reviews. A dictionary based technique is proposed by Fei et al. [6] to identify aspects of a review by considering adjectives only.

Pang and Lee [23, 25] used NB, ME and SVM for sentiment analysis of movie reviews by taking into account some special features like unigrams, bigrams, and a combination of both (i.e. unigram and bigram), including POS and positional information with unigram, and adjectives. It was evident from their experiment that feature presence provides greater accuracy than feature frequency. For small feature space, NB performs better than SVM. However, when feature space is increased, SVM outperforms NB classifier. Bikel et al. [1] implemented a subsequence kernel based voted perceptron and evaluated its performance with standard SVM. The authors observed that the increase in the number of false positives with the increase in the number of true positives is far less in aforesaid scheme when compared with the bag-of-words based SVM, where the trend is almost linear. The model shows resiliency over the fact that a smooth continuum is observed for intermediate star rating reviews even when trained only on the extreme one- and five-star rating reviews. Further, microblog sentiment analysis can be evaluated through various classifiers in two phases under two different settings. The first phase involves separation of subjective and objective documents using various classifiers. In the second phase, these filtered documents are tagged as positive or negative by the classifiers

2.2 Emotional Classifier on Viral or Trending Information

Emotion detection is one of the natural language processing and text analytics fields used to discover people's feelings in written texts. It has been applied in different tracking systems, including disasters and social media monitoring.

As social media become important and ubiquitous for social networking, they support backchannel communications and allow for wide-scale interaction [38]. Social media content can be used to build tracking systems in many applications. The tracking of sentiment on news entities over time [39] is one of the systems that attract researchers in monitoring socio-politics issues. Sentiment-spike detection has been presented in [40]. The authors used Twitter data and analyzed the sentiment towards 70 entities from different domains. Tracking health trends from social media has been studied in [41]. The authors introduced an open platform that uses crowd-sourced labeling of public social media content.

Tourism is another domain for which tracking systems have been built; specifically, tracking systems have been used to monitor tourist comments being written on social media [42]. The authors used a sentiment analysis lexicon to track the opinion of the tourists in Tunisia. Monitoring people's opinions before or during elections is useful to track and analyze the campaigns using Twitter data [43]. The study identified the topics that were most causal for public opinion, and show the usefulness of measuring Information 2021, 12, 86 5 of 21 public sentiments during the election. Tracking and monitoring earthquake disasters from Weibo Chinese social media content was proposed in [44]. The authors focused on how to detect disasters from massive amounts of data on a micro-blogging stream. They used sentiment analysis to filter the negative messages to carry out incident discovery in a post-disaster situation.

To summarize, building a tracking system on top of social media content is very useful for governments and decision-makers. Most of the proposed work has been done for English data with fewer contributions in Arabic. The research related to the emotions and symptoms in Arabic text did not focus on the monitoring and tracking of emotions and symptom evolution over time. That motivated us to conduct our research on Arabic data to help health authorities, governments, and decision-makers to understand people's emotions during the COVID-19 pandemic

3 Proposed Strategy

The traditional social media platform, e.g., Twitter, Facebook, etc. cater to the global users and list their personal information and media. The heterogeneous user data is often utilized for deriving the common sentiments or trending information. The trending or viral information primarily harnesses the global content shared co-related to a particular topic and keyword (e.g., hashtag).

A naive user or new user usually refers to this trending or viral list of information to see the most occurring or contributory piece of information. In this process, a user simply refers to the viral information or hashtag and explores the related term over the twitter API, without cognitive awareness of the emotional effect of viral information. The viral or trending information may have a list of information which may trigger the emotional effect on the user and lead to the emotional splits or swing on the choice of information. User assistance is pivotal for the user, which may assist the user to showcase the emotional effects that viral information may carry. Though, social media platforms offer limited or no functions or aspect related views on the API for the generic user.

3.1 Conceptual Framework

A novel strategy for the real-time generation of emotional quotients of viral/trending information on twitter is designed. Figure 1 illustrates the internal computing blocks and their interactions for the intended objectives. The proposed framework begins with a traditional data collection over twitter API. The data extraction is driven

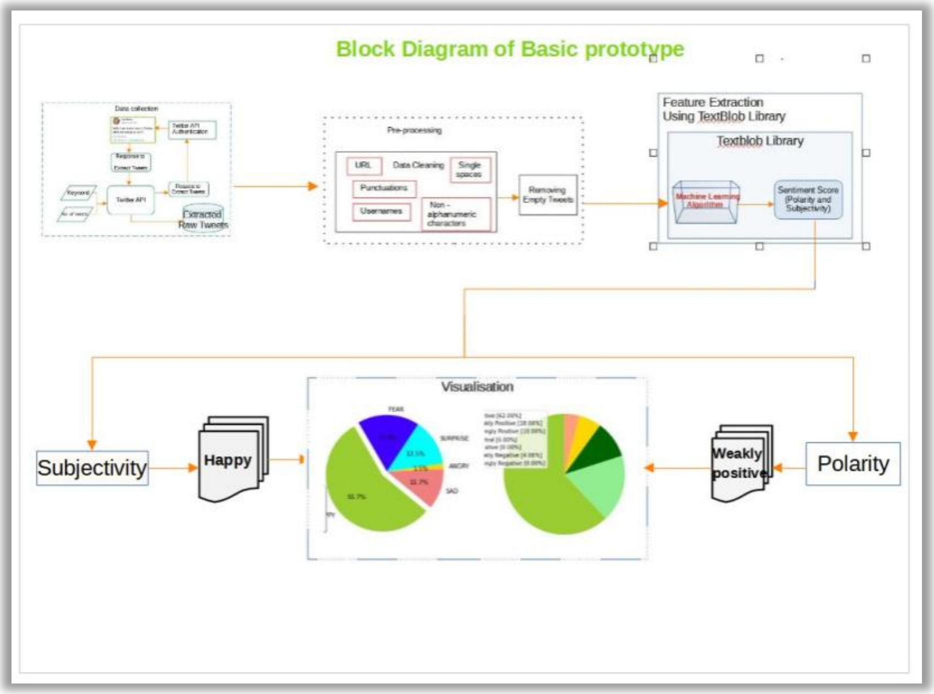


Fig. 1. Conceptual framework of proposed strategy

by the user inputs, e.g. keyword/hashtags/, number of tweets, and duration. The retrieved tweets from the API, are now to be stored in a temporary storage for later text-processing and feature extraction. The local twitter data storage is also connected to the 2nd computing clock 'text pre-processing', each tweet extracted must go through local text processing and further supplied to the 'feature-extraction'. Further, a small computing thread is kept within the 'feature extraction' computing block for the estimation '*sentiment score (SC)*' and '*emotional quotient (EQ)*' co-located twitter data objects.

3.2 Generating Sentiment level (SL) and Emotional Quotient (EQ)

The aim of the designed system is to generate the sentiment statistics and emotional quotients for user submitted inputs. A prospective user (e.g., naïve, programmer, decision maker, business analytics, admin, Govt. official, etc.) submit the data request over the provided GUI, using user defined keywords, number of tweets of interest and name of emotional (optional). The designed system, evaluates the both statistics over a real-time.

The pre-processing stage, each extracted tweet is divided into tokens; every token is assigned a probability (for e.g., T_{prob}= happy, sadetc.) of it being associated with one of the sentiments. The python sentiment analysis is conducted using NLTK library. The probability score (WGT_{Prob}.) is weighted a value, as to account for the fact that there are many fewer negative tweets as there are positive and neutral ones. Additionally, token below a threshold count is filtered since are assumed to not appear enough to be used confidently, and similarly, words which are found to carry too little information are also removed. The latter is determined using the H_{10} entropy defined as,

$$\mathrm{H}_{10}(\mathrm{token}) = -\sum_{s \epsilon \mathrm{sentiment}} (\mathrm{p}_{(s|\mathrm{token})} \log_{10} \mathrm{p}_{(s|\mathrm{token})}) \ldots\ldots\ldots\ldots\ldots..(1)$$

The measures for *positive*, *neutral* and *negative* emojis are found. Finally, as given that a tweet is composed of several words; all the different features are aggregated /summed for each word, as to obtain an overall *tweet_value* (T_v), normalized by the *tweet_length*(T_{ct}).The *positive score* (s^+) and *negative scores*(s^-) for each tweet are determined as average of the both scores. The following notions are used for the purpose:

$$\mathrm{s}^+ = \frac{\sum_{i\epsilon t} \mathrm{pos\_score}_{\,i}}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots...(2)$$

$$\mathrm{s}^- = \frac{\sum_{i\epsilon t} \mathrm{neg\_score}_{\,i}}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3)$$

$$\text{Overall Sentiment score (SC)} = (s^+ - s^-) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(4)$$

To extract emotion from a tweet, the topical words (bigram) are taken from tweet message using item response theory and are categorized according to its unsupervised nature of the features. Topic proportion gives the sentiment, Idea behind the algorithm is to find those terms that relate to a topic sentiment with respect to the topic sentiment lexicon.

Lexicon Based Approach: Our tweets message isto be tested on topic models. Because the entire document is the mixture of one or more topics which are estimated using the parameter estimation technique, this allows the users to find the text that is relevant to the topic through the use of a particular keyword. The lexicon approach measures the sentiment of a group of document corpus with the help of dictionary of words and its associated polarity scores in the training corpus and all such words of the documents are compared to the word usage in the lexicon. There are many ways generally for the lexicon to offer best chance to successfully estimate sentiment:(i)preassembly domain specific lexicon, (ii) dictionary based lexicon, and (iii)corpus based lexicon.

The dictionary of lexicons elements is added externally to the corpus for the purpose of enhancing the preassembled lexicons. In the lexicon-based sentiment analysis, it is sufficient to simply count the term frequency of every document relevant to our topic of interest. The conditional probabilities of each and every lexical token in the vocabulary were computed with the help of training sample using the following equation:

$$P(w\,|+) = \frac{M_w}{|N^+|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$$

$$P(w\,|-) = \frac{M_w}{|N^-|} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(6)$$

The score of positive and negative sentiment is coded as N^+, N^-.For each sentence message "*m*," the log likelihood ratio is calculated using the following equation:

$$S_m = \sum_{i=1}^{n} \log(\frac{P(w_i|-)}{P(w_i|+)}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(7)$$

where w is the lexical unit of the dictionary and$n$ is the number of words and collocations included in the dictionary, which are found in the sentence "*m*."M_w is the set of messages containing lexical token "*w*."Next, the designed algorithm for the proposed system is described. The algorithm is primarily designed to estimate the values of polarity for each level of sentiment and emotion quotient for the user input.

Algorithm: Extracting SL and EQ of Tweet_objects

Input: Topic name (T_{kw}), No. of Tweets (ToI), and Emotion _name (Emo)

Output: Tweet_list, EQ, SL

Step1: Cleaning up the tweets*/*Here we clean up unnecessary characters from the tweets such as links,@,#,etc.*/*

```
    df['Tweet'] = df['Tweet'].apply(cleanUpTweets)
all_tweets=df['Tweet'].tolist()
df = df.drop(df[df['Tweet'] == ' '].index)
```

Step2: Gathering and Storing Emotions*/*Using text2emotion emotion values is gathered and stored for each tweet*/*

```
    for text in lt:
dict = te.get_emotion(text)
for key in dict:
        if (key == "Happy"):
Happy.append(dict[key])/*Similartly also for "Sad","Fear" etc*/
```

Step 3: Finding the dominating Emotion */*For each tweet, store the dominating emotion values */*

```
      for i in range(0,NoOfTweets):
maxel=max(Happy[i],Sad[i],Angry[i],Fear[i])
            if((maxel==Happy[i])and(maxel!=0)):
            Happy2.append(all_tweets[i])      /*Similarly also for Sad2, Fear2 etc*/
```

Step4: Finding the overall percentage of each emotion after analyzing each tweet on the topic

*/*All the values in each emotion column are added and for each emotion percent is calculated by dividing summation of values for that emotion by the total sum*/*

```
happysum = sum(Happy)
totalsum = zip(Happy, Sad, Angry, Surprise, Fear)
 for x in totalsum:
        f or y in x:
            total = total + y
happypercent = (happysum / total) * 100/*Similarly the percentage of other emotions is calculated*/
```

Step5: Polarity analysis*/*For each tweet, polarity is analyzed& values are classified into three groups */*

```
      for tweet in self.tweets:
        analysis = TextBlob(tweet.text)
        polarity += analysis.sentiment.polarity
          if (analysis.sentiment.polarity== 0 neutral += 1
elif (analysis.sentiment.polarity> 0 and analysis.sentiment.polarity<0.3):wpositive += 1
elif (analysis.sentiment.polarity>0.3 and analysis.sentiment.polarity<=0.6): positive += 1
elif (analysis.sentiment.polarity> 0.6 and analysis.sentiment.polarity<= 1):spositive += 1
elif (analysis.sentiment.polarity> -0.3 and analysis.sentiment.polarity<= 0):wnegative += 1
elif (analysis.sentiment.polarity> -0.6 and analysis.sentiment.polarity<= -0.3): negative += 1
elif (analysis.sentiment.polarity> -1 and analysis.sentiment.polarity<= -0.6):snegative += 1
```

3.3 System Use-Cases for Sentiment and Emotional Analytics

The designed system is plugged with an interactive user interface (UI) for several types of users. The aim of UI is to offer estimated statistics and processed data for the different decision making and analytics purposes. Several are several use-cases of the designed system listed during the design phase, e.g. related viral information to an emotional value, comparing the emotional causes of more than one viral/trending information, etc. The first use-case from the naive user is a '*basic search of the sentiment and emotional statistics*', as shown in Figure 2. Here, a different part of UI supports the various cognitive tasks of the user query for the real-time analytics over the twitter data extraction.

For a user input 'keyword/hashtag or emotional name', the system extracts the related tweets and prepares intermediate statistics to be shown in the graphical scheme. Part B of the UI illustrates both information in pie chart view with scores 'as % values', here *Part A* lists down the tweets of the user interest. In *Part B*, the feature of compares and explores offers a new dimension of the designed prototype. Here a user may be interested to compare effects of two trending/viral information or may delve into the deeper view of these information and related tweets.

Second related use-case is an exploration into the relevant other viral/trending information for a user query. The utilization of the designed system for the exploration within the deep down the relevant tweets for a user's emotional effects. The matching viral information and tweets may be easily extracted and utilized for the analytics. Third use-case of the system is a thread capacity, to deliver a matching list of current trending/viral information's for and matching viral the system for the exploration within the tweets data for an input, 'emotional value'. The viral/trending topics may be extracted with the presence of the same emotional quotient values. Figure 4(a) and (b) illustrates the similar view.

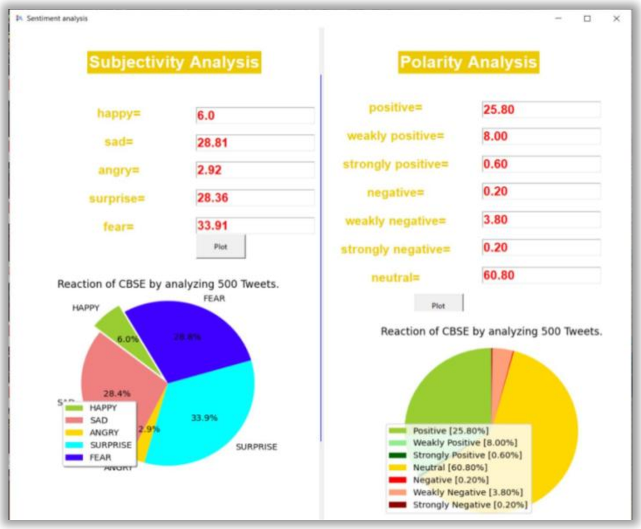


Fig.2. created User Interface (UI) for the analytics and Interactions

Fourth potential use-case of the designed system is a systematic comparative view between more than one user inputs (trending and viral information) and its emotional effects and further exploration on the generated tweet text.

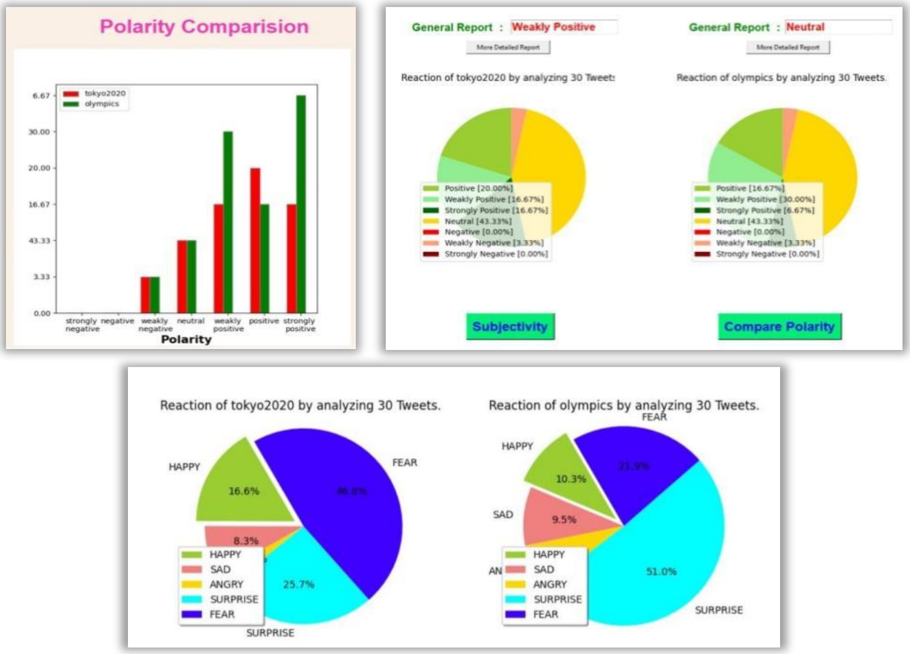


Fig.3. User Inputs: ‘tokyo2020' and ‘olympic’ for 100: (a) Polarity comparison and analysis, and (b) emotional quotient view of each information.

3.4 Working Example

The Olympics is the world's biggest sporting event. There is a lot of emotion attached with it.Every player gives their best to win for his country. They train to their fullest with sacrifices and dedication. This is the story of most athletes reaching the Olympic stage.If an athlete wins a medal, it's well and good, we here will illustrate the other side.

The example we will study express the sentiment of an athlete of India after not acquiring the shiny medal, and compare his sentiment with that of a winning athlete from India, but as this comparison will not give much significance, we will also compare him with contenders from other countries which are sports friendly in nature (e.g., Russia, Japan, United States).We took men's single badminton player *Sai Praneeth*, who lost both his initial matches and is out of the tournament.

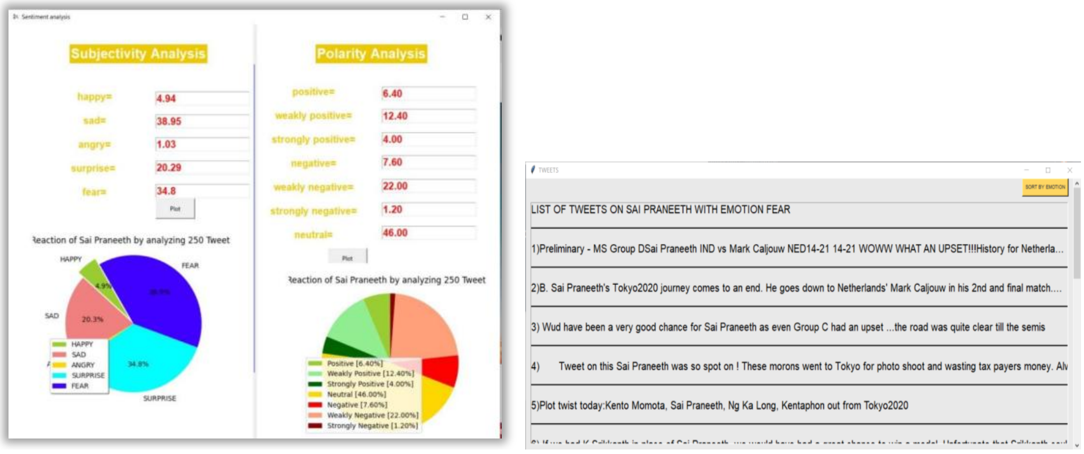


Fig.4 (a) Input '*Sai Praneeth*' with 250 number of Tweets (b) Estimated*Subjectivity* and *Polarity* Analysis

In Polarity positives [22.8%] is less than negative's [30.8%]. Fear and Surprise dominate in subjectivity.Next, we compare *Sai Praneeth* with the other viral information as '*MirabaiChanu*'.Polarity graph is interesting to observed as, opposite to both players, in figure 5 (a). In subjectivity *Happiness* surrounds MirabaiChanu with some sadness, for Sai Praneeth Surprise, Fear, sadness is mostly visible.Now we compare Sai Praneeth to KentoMomota ["presently World Rank 1"] a player from Japan who also lost in his 2nd round of men's badminton singles.

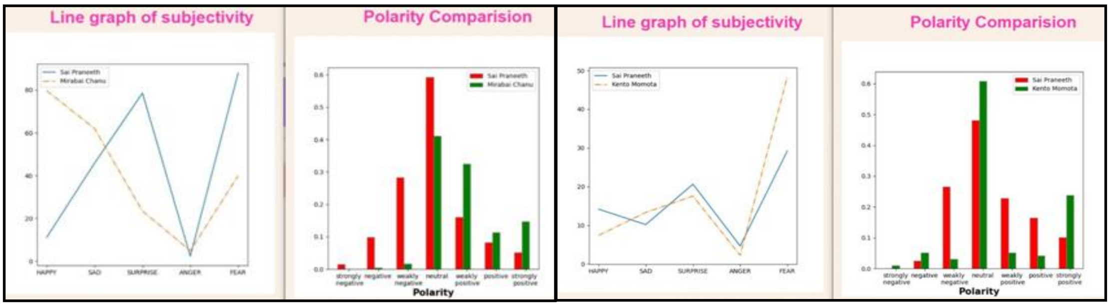


Fig.5 Subjectivity and polarity comparison (a) for '*Sai Praneeth'* and '*MirabaiChanu*', (b) for '*Sai Praneeth*' and '*KentoMomoto*'.

In subjectivity analysis, the graph almost correlates, only significant difference is found in "fear", where higher fear is surrounding *KentoMomota*.On Polarity Graph, *KentoMomota* has a lot of Positivity where *Sai Praneeth* has a trace of weak negativity in the aura of tweets.

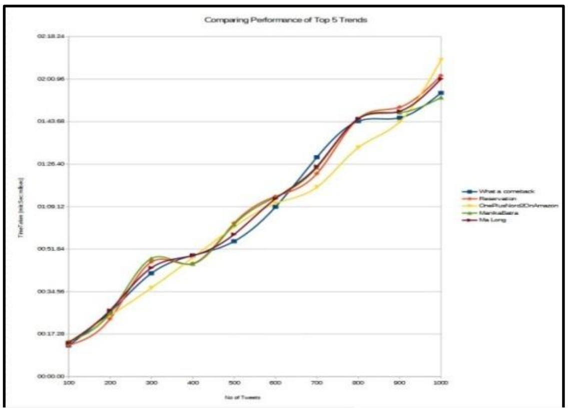
4 Performance Assessment and Evaluation

4.1 Data Settings

The experimental setup includes software used *Jypter notebook*, *Visual Studio* and *PyCharm*. Twitter API isused for the real-time data extraction, at instance we have extracted 1 to 800 tweets for Comparing 1 to 1600 tweets for analytics. There are several libraries, e.g. *tweepy,re,text2emotion,textblob, pandas, numpy, matplotlib, sys,csv* and *Tkinter* (for GUI).The hardware components includes, 2 PCs with specification as: AMD Ryzen 5 2500U processer with Radeon Gfx 2.00 GHz,RAM 8GB and another with processor of Intel Core i5(8250U)CPU @ 1.60GHz, Intel UHD graphics 620, 12 GB RAM).

4.2 Performance on Sentiment and Emotion Quotient (EQ) estimation

The performance evaluation of the designed system for the real-time data processing to estimate the underlying statistics outlines several insights, specific to system's feasibility and its viability for just-in-time decision making and analysis. We have planned employed two key parameters for the overall performance evaluation, e.g. query length (*QL*), Number of tweet of interests (*ToI*), andPerformance (processing time). Here *QL*defined as the dimension of the keyword/hashtag, i.e No. of characters in the user input*keyword/hashtag*. For a user input '*Tokyo2020'*, *QL is* 9. Similarly,*ToI* value directs system to extract at least these many recent tweets from API on in real-time basis,e.g. ‘*Tokyo2020’* with*ToI*value20 fetches recent 20 tweets at the time.



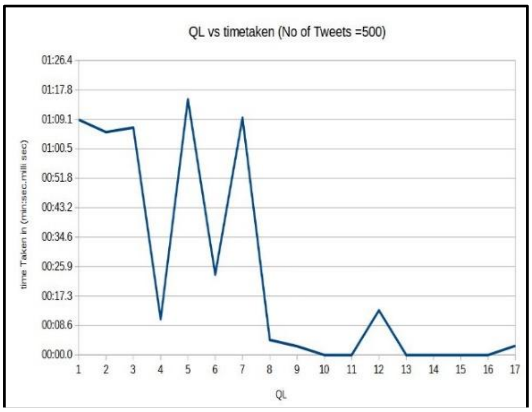


Fig.6.AnalyzingSL and EQ polarity for multiple viral Topics. **Fig.7**. Overall performance for the user input ‘*QL*’

Figure 6 illustrates the overall processing time for the five different viral topics (on 30 July, 2021). The computation estimation is based on number of tweets in coverage. Whereas, figure 7 depicts the overall processing time taken in designed algorithm for varying query length (QL), as input size. Here, we have consider different viral/trending topics appearing , as to ensure the variable QL and observed the processing time patterns for at least recent 500 tweets on each viral/trending topics.

4.3. Accuracy on Sentiment and Emotional Quotient generation

The evaluation of accuracy in the detection of accurate emotion quotient and sentiment levels for each user specified inputs is pivotal for the analysis. The performance statistics for the estimation of accuracy on specified sentiment-level are listed in table 1. A brief comparison metric on 497 tweets sample is used to validate the accuracy indicates the precisely predicted.

Table1. Accuracy on estimation of each specified *Sentiment -le*vels.

			Predicted		
			Positive	Negative	Neutral
			234	***108***	***155***
Actual	Positive	***181***	143	14	24
	Negative	***177***	54	86	37
	Neutral	***139***	37	8	94

The overall accuracy (in %) on the prediction of positive, negative, and neutral is 79%,48.58%,67.62% respectively, The designed algorithm delivered overall 64.98% of accuracy in the indication of inherent sentiment of user input and relevant tweets.

At the next level, prediction of accurate emotion levels is 66.66%,15%,38.88%,26.66%, and 25% for Happy, Sad, Fear, Surprise and Anger respectively. With overall accuracy delivered is 34.61%.

Table2. Accuracy on estimation of each specified Emotion.

			Predicted				
			Happy	Sad	Fear	Surprise	Anger
			56	***24***	***22***	***12***	***16***
Actual	Happy	***33***	22	5	2	2	2
	Sad	***40***	15	6	8	6	5
	Fear	***18***	6	2	7	0	3
	Surprise	***15***	5	5	1	4	0
	Anger	***24***	8	6	4	0	6

4.4 Overall Retrieval Performance

Further, the set of traditional retrieval measures are adapted for the evaluation of the system performance, e.g.*Precision, recall and f-measure*. The precision is adapted in its fundamental notion, as a measure of 'precisely matched results to the user input', and recall as a measure '*closely relevant result to the user query*'. F-measure is a geometric mean of precision and recall. Table 2 lists the precision, recall and f-measure indicators for the designed system, when experimented with varying degree of user input (Query length and Tweets of interest). Firstly, three traditional indicators are shown for the estimation of the sentiment of the user input at instance.

Table 3.Precision, Recall and F-measure metrics for both *Sentiment* and *EQ estimates*

Sentiment type	Metric (scores)	
Positive	Precision	0.611
	Recall	0.389
	F-measure	0.475
Negative	Precision	0.796
	Recall	0.224
	F-measure	0.351
Neutral	Precision	0.606
	Recall	0.240
	F-measure	0.344

Emotion type	Metric (scores)	
Happy	Precision	0.393
	Recall	0.259
	F-measure	0.312
Sad	Precision	0.250
	Recall	0.113
	F-measure	0.156
Fear	Precision	0.318
	Recall	0.108
	F-measure	0.161
Surprise	Precision	0.333
	Recall	0.048
	F-measure	0.084
Anger	Precision	0.375
	Recall	0.063
	F-measure	0.107

5 Conclusion

The data generated over the various Social media platforms trigger significant changes on the public sentiment and emotional flux. In this project, a novel algorithm is designed to estimate the sentimental and emotional quotient of viral or trending information in real-time. The project work carried is in line with current need to textual emoticons mining in several real-life application scenarios. The novel algorithm for the estimation of sentiment and emotion quotients of a user requested input or may be current viral /trending information over the twitter. The current approach builds a corpus of tweets and related fields where each tweet is classified with respective emotion based on lexicon approach. The systematic evaluation asserts the significance of delivered statistics for the user input 'viral information', and its usability. The feasibility analysis of generated statistics for real-time analysis and decision-making is uncovered using 04 use-cases, also outlines the key features of social media data for the purpose.

References

[1] Bikel, D. M., & Sorensen, J. (2007). If we want your opinion. In: International conference on semantic computing (ICSC 2007) (pp. 493–500). doi:10.1109/ICSC.2007.81.

[2] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15–21. doi:10.1109/MIS.2013.30.

[3] Chen, R., & Xu, W. (2016). The determinants of online customer ratings: A combined domain ontology and topic text analytics approach. Electronic Commerce Research. doi:10.1007/s10660- 016-9243-6.

[4] Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 international conference on web search and data mining (pp. 231–240). doi:10.1145/1341531.1341561.

[5] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of 5th language resources and evaluation (Vol. 6, pp. 417–422).

[6] Fei, G., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2012). A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In: Proceedings of 24th international conference on computational linguistics (p. 309).

[7] Feldman, R., Fresco, M., Goldenberg, J., Netzer, O., & Ungar, L. (2007). Extracting product comparisons from discussion boards. In: Seventh IEEE international conference on data mining (ICDM 2007) (pp. 469–474). doi:10.1109/ICDM.2007.27. A model for sentiment and emotion analysis of unstructured... 197 123

[8] Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. Proceedings of the international conference on weblogs and social media (ICWSM), 7(21), 219–222.

[9] Hamouda, A., & Rohaim, M. (2011). Reviews classification using sentiwordnet lexicon. In: World congress on computer science and information technology.
[10] Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. In: Proceedings of the 21st national conference on artificial intelligence (Vol. 2, pp. 1331–1336).
[11] Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. Expert Systems with Applications, 42(11), 4999–5010. doi:10. 1016/j.eswa.2015.02.007.
[12] Li, J., Fong, S., Zhuang, Y., & Khoury, R. (2015). Hierarchical classification in text mining for sentiment analysis of online news. Soft Computing, 20, 3411–3420. doi:10.1007/s00500-015-1812-4.
[13] Li, Y., Qin, Z., Xu, W., & Guo, J. (2015). A holistic model of mining product aspects and associated sentiments from online reviews. Multimedia Tools and Applications, 74(23), 10177–10194. doi:10. 1007/s11042-014-2158-0.
[14] Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2, 627–666.
[15] Liu, B. (2011). Opinion mining and sentiment analysis. In: Web data mining: Exploring hyperlinks, contents, and usage data (pp. 459–526). doi:10.1007/978-3-642-19460-3_11.
[16] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.
[17] Liu, P., Gulla, J. A., & Zhang, L. (2016). Dynamic topic-based sentiment analysis of large-scale online news. In: Proceedings of the 17th international conference on web information systems engineering (pp. 3–18). doi:10.1007/978-3-319-48743-4_1.
[18] Ma, Y., Chen, G., & Wei, Q. (2017). Finding users preferences from large-scale online reviews for personalized recommendation. Electronic Commerce Research, 17(1), 3–29. doi:10.1007/s10660- 016-9240-9.
[19] Mo, S. Y. K., Liu, A., & Yang, S. Y. (2016). News sentiment to market impact and its feedback effect. Environment Systems and Decisions, 36(2), 158–166. doi:10.1007/s10669-016-9590-9.
[20] Montoyo, A., Martı´Nez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decision Support Systems, 53(4), 675–679. doi:10.1016/j.dss.2012.05.022.
[21] Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015). Text mining of newsheadlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. Expert Systems with Applications, 42(1), 306–324. doi:10.1016/j.eswa.2014.08.004.
[22] Ohana, B. (2009). Opinion mining with the sentwordnet lexical resource. M.Sc. dissertation, Dublin Institute of Technology.
[23] Pang, B., & Lee, L. (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on association for computational linguistics (p. 271). doi:10.3115/1218955.1218990.
[24] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. doi:10.1561/1500000011.
[25] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing (Vol. 10, pp. 79–86). doi:10.3115/1118693.1118704.
[26] Parkhe, V., & Biswas, B. (2016). Sentiment analysis of movie reviews: Finding most important movie aspects using driving factors. Soft Computing, 20(9), 3373–3379. doi:10.1007/s00500-015- 1779-1.
[27] Peng, J., Choo, K. K. R., & Ashman, H. (2016). Astroturfing detection in social media: Using binary n-gram analysis for authorship attribution. In: Proceedings of the 15th IEEE international conference on trust, security and privacy in computing and communications (TrustCom 2016) (pp. 121–1286).
[28] Peng, J., Choo, K. K. R., & Ashman, H. (2016). Bit-level n-gram based forensic authorship analysis on social media: Identifying individuals from linguistic profiles. Journal of Network and Computer Applications, 70, 171–182. doi:10.1016/j.jnca.2016.04.001.
[29] Peng, J., Detchon, S., Choo, K. K. R., & Ashman, H. (2016). Astroturfing detection in social media: A binary n-gram-based approach. Concurrency and Computation: Practice and Experience. doi:10. 1002/cpe.4013. 198 J. K. Rout et al. 123
[30] Pro¨llochs, N., Feuerriegel, S., & Neumann, D. (2015). Enhancing sentiment analysis of financial news by detecting negation scopes. In: Proceedings of the 48th Hawaii international conference on system sciences (HICSS) (pp. 959–968). doi:10.1109/HICSS.2015.119.
[31] Robinson, R., Goh, T. T., & Zhang, R. (2012). Textual factors in online product reviews: A foundation for a more influential approach to opinion mining. Electronic Commerce Research, 12(3), 301–330. doi:10.1007/s10660-012-9095-7.
[32] Rout, J., Dalmia, A., Choo, K. K. R., Bakshi, S., & Jena, S. (2017). Revisiting semi-supervised learning for online deceptive review detection. IEEE Access, 5(1), 1319–1327. doi:10.1109/ ACCESS.2017.2655032.
[33] Rout, J., Singh, S., Jena, S., & Bakshi, S. (2017). Deceptive review detection using labeled and unlabeled data. Multimedia Tools and Applications, 76(3), 3187–3211. doi:10.1007/s11042-016- 3819-y.
[34] Sadegh, M., Ibrahim, R., & Othman, Z. A. (2012). Opinion mining and sentiment analysis: A survey. International Journal of Computers & Technology, 2(3), 171–178.

[35] Song, L., Lau, R. Y. K., Kwok, R. C. W., Mirkovski, K., & Dou, W. (2017). Who are the spoilers in social media marketing? Incremental learning of latent semantics for social spam detection. Electronic Commerce Research, 17(1), 51–81. doi:10.1007/s10660-016-9244-5.

[36] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267–307. doi:10.1162/COLI_a_00049.

[37] Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. Expert Systems with Applications, 36(7), 10760–10773. doi:10.1016/j.eswa.2009.02.063.

[38] Turney, P.D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417–424). doi:10.3115/1073083.1073153.

[39] Wang, D., Li, J., Xu, K., & Wu, Y. (2017). Sentiment community detection: Exploring sentiments and relationships in social networks. Electronic Commerce Research, 17(1), 103–132. doi:10.1007/ s10660-016-9233-8.

[40] Zheng, L., Wang, H., & Gao, S. (2015). Sentimental feature selection for sentiment analysis of chinese online reviews. International Journal of Machine Learning and Cybernetics. doi:10.1007/ s13042-015-0347-4.

[41] Alves, A.L.F. A spatial and temporal sentiment analysis approach applied to Twitter microtexts. J. Inf. Data Manag. 2015, 6, 118.

[42] Chaabani, Y.; Toujani, R.; Akaichi, J. Sentiment analysis method for tracking touristics reviews in social media network. In Proceedings of the International Conference on Intelligent Interactive Multimedia Systems and Services, Gold Coast, Australia, 20–22 May 2018.

[43] Contractor, D. Tracking political elections on social media: Applications and experience. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

[44] Bai, H.; Yu, G. A Weibo-based approach to disaster informatics: Incidents monitor in post-disaster situation via Weibo text negative sentiment analysis. Nat. Hazards 2016, 83, 1177–1196.

<table>
<tr><th>Name & Signature of Student’s</th><th>Name, Designation, Address and Signature of the Mentor</th></tr>
<tr><td>P Kumar

Pawan Kumar (181EE133)
Electrical Engineering Dept.
National Institute of Technology, Surathkal,
Karnataka, India
Email: pawan.181ee133@nitk.edu.in
Contact: <add></td><td rowspan="2">Date: 13 Jan 2021
Dr Vikram Singh
Asstt. Prof.
Computer Engg. Dept.,
NIT Kurukshetra
viks@nitkkr.ac.in, 88160 86589

Dr. Vikram Singh
Assistant Professor
Computer Engineering, Dept.,
National Institute of Technology, Kurukshetra,
Haryana-136119, India
Email: viks@nitkkr.ac.in
Contact: 08816086589</td></tr>
<tr><td>Reuben.

Reuben Eappen Reji (181EE136)
Electrical Engineering Dept.
National Institute of Technology, Surathkal,
Karnataka, India
Email: reubeneappenreji.181ee136@nitk.edu.in
Contact: 7012338043</td></tr>
</table>