International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Aspect Based Sentiments from Tweets using Co-Ranking Multi-Modal Natural Language Processing Methodologies

M. Kanipriya, R. Krishnaveni, M. Krishnamurthy, S. Bairavel

Abstract: ***Now-a-days people interest to spend their time in social sites especially twitters to post lot of tweets in every day. The posted tweets are used by many users to get the knowledge about the particular applications, products and other search engine queries. With the help of the posted tweets, their emotions and sentiments are derived which are used to get opinion about particular event. Lot of traditional sentiment detection system that has been developed but they failed to analyze huge volume of tweets and online contents with temporal patterns were also difficult to analyze. To overcome the above issues, the co-ranking multi-modal natural language processing based sentiment analysis system was developed to detect the emotions from the posted tweets. Initially, tweets of different events are collected from social sites which are processed by natural language procedures such as Stemming, Lemmatization, Part-of-speech tagging, word segmentation and parsing are applied to get the words related to posted tweets for deriving the sentiments. From the extracted emotions, co-ranking process is applied to get the opinion effectively related to particular event. Then the efficiency of the system is examined using experimental results and discussions. The introduced system recognize the sentiments from tweets with 98.80% of accuracy.***

Keywords: ***Twitter, tweets, sentiment analysis, emotion, co-ranking multi-modal natural language processing based sentiment analysis system, stemming, Lemmatization, Part-of-speech tagging, word segmentation and parsing***

I. INTRODUCTION

Sentiment analysis [1] is one of the most important things in text mining and social site based communication. The text mining concept utilizes various techniques to determine the neutral, negative and positive comments from the list of comments. Now-a-days most of the people utilize social sites such as facebook, twitter, instagram and so on. Among the several sites, twitter [2] is one of the common social sites to convey their opinion about particular incident, political thoughts, public actions in political problems and business decision. There is lot of tools [3] such as revealed context, engenuity, social mention, steam crab and meaning cloud are used to examine the twitter comment and predict the sentiment effective manner. In addition to this, python and R [4] also utilize the twitter dataset for predicting people sentiments, emotion effectively. Even though these tools are fail to detect the sentiment 100% of accuracy with specific time. Along with this, these tools are sometimes creating complexity while managing large volume twitter data. For overcoming the above discussed issue, several earlier detection techniques [5] such as Latent Dirichlet Allocation (LDA),Latent Semantic Analysis (LSA), support vector machine, naïve Bayes, apriori, FP-growth, maximum entropy approaches are used to predict the sentiment from twitter data.

This mentioned algorithm is utilized by several authors to predict the sentiment from the twitter dataset. Few authors' opinions are discussed in this part to get the knowledge, idea, templates; thoughts are get to develop the effective sentiment analysis system. (Shihab Elbagir et al., 2018) [6] developing twitter sentiment analysis system by using scikit learn and machine learning techniques. The twitter data is collected from NLTK corpora which is examined successfully to extract the various features. The extracted twitter features are processed with the help of several classifiers such as SGD classifier, bernouslli NB, multinomial NB, linear SVC, SVC and NuSVC approach. The discussed classifiers recognize the twitter sentiments up to 75% of accuracy compared to the prediction tools. (Md Shad Akhtar et al., 2016) [7] Detecting sentiments from twitter data using hybrid convolution neural networks. Initially, twitter data is collected from websites which are processed and several twitter features are derived. From the derived features, optimized twitter data is selected with the help of multi objective optimization approach. The selected features are processed, trained using support vector machine. The trained features are successfully helps to recognize the sentiment word effectively. Then the efficiency of the system is evaluated using English and Hindi twitter dataset. Then the introduced system attains 77.16% of accuracy while analyzing various twitter comments.

(Imane El Alaoui et al., 2018)[8] Analyzed public opinion from social data using effective sentiment analysis technique.

Manuscript published on January 30, 2020.
* Correspondence Author

M. Kanipriya,* Research Scholar, Hindustan Institute of Technology & Science, Chennai, India,

R. Krishnaveni, Professor, Hindustan Institute of Technology & Science, Chennai, India,

M. Krishnamurthy, Professor, K.C.G College of Technology, Chennai, India,

S. Bairavel, Assistant Professor, K.C.G College of Technology, Chennai, India,

Initially, the public opinion data is collected from large social sites. The collected data is processed and user opinions are derived based on the polarity and hashtag related contents are derived from comment.

From the extracted information, various sentiments are classified using adaptive analysis technique. During the Excellency examination process, US 2016 election related comments are taken to classify the positive, negative comments effectively. (Ahmed Sulaiman et al.,2019)[9] introducing the enhanced approach called convolution neural network for detecting sentiment from twitter data. The system utilizes the semeval 2016 workshop related two twitter database. After collecting data, effective features are derived which are processed with the help of multi-layer convolution network. The network successfully predicts the positive and negative tweets. Then the efficiency of the system is compared with the several baseline models such as support vector machine and naïve Bayes. Among the various models, deep learning based prediction process provides effective result while predicting twitter sentiments.

According to the discussion, twitter sentiments are predicted effectively by created automatic system. The prediction accuracy and amount of data used in the research work is still minimum. So, the system efficiency need to be improved also large volume of data has to be utilized to enhance overall sentiment prediction efficiency. For this requirement, natural language processing tool [10] is used to examine the each word in the tweet also it utilizes large volume of data effectively. Then the introduced approach uses different bench mark of dataset that perfectly recognize the positive, neutral and negative comments successfully. Finally, the discussed system is developed using python tool and efficiency of the system is determined using different performance metrics. Depending on the discussion, the rest of the paper is arranged as follows, section I analyze the basic introduction, different author opinion regarding sentiment analysis has been discussed. Section II discusses the co-ranking multi-model natural language processing approaches based sentiment analysis process, section III evaluates the efficiency of co-ranking multi-modal natural language processing approaches related sentiment analysis and outcome of the work is discussed in section IV.

II. RANKING MULTI-MODAL NATURAL LANGUAGE PROCESSING APPROACHES BASED SENTIMENT ANALYSIS SYSTEM

In this section discusses about the detailed working process of Co-Ranking Multi-Model Natural Language Processing Approaches Based Sentiment Analysis Process. During the analysis process, the system must be used large volume of data because the traditional system has specific twitter dataset [11]. According to the discussion, in this work large volume of dataset [12] such as awesome twitter data is used to predict the sentiment effectively. The captured data is processed by different processing steps which are depicted in figure 1.

A. Twitter Data Set

The first step of the work is data collection, in this work twitter data is gathered from awesome twitter dataset [13]. As discussed earlier, the developed sentiment prediction system needs to utilize large amount of dataset. So, the utilized awesome dataset consists of large volume of twitter data such as movie twitting, chase tweets, ego tweets, new year resolution tweets, crowd flower gender tweets, weather sentiment tweets and so on. In addition to this, awesome dataset consists of several user twitter data namely, max plank

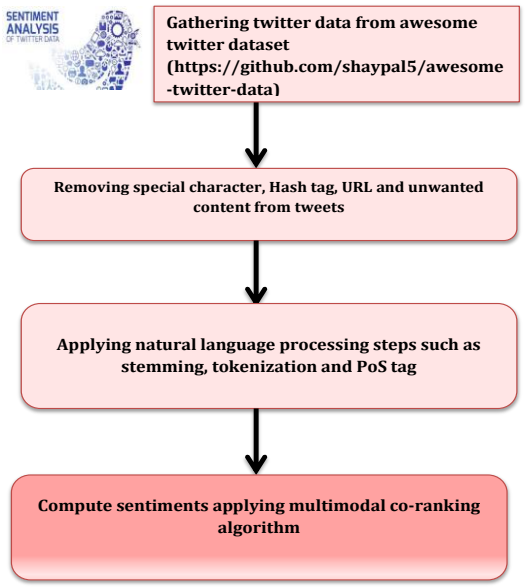

Figure 1: Co-Ranking Sentiment Analysis System

twitter data(1,755,925,520 tweets), social group tweets(41m tweets), higger tweets (456K tweets) and ego twitter (80k). So, the collected tweets consists of almost both positive, negative and neutral tweets which are analyzed using natural language processing and co-ranking algorithm to predict the sentiments effectively. From the collected twitter data [14], several unwanted tweets information which need to be eliminated for further processing. The sample tweets are depicted in table I with relevant tweet id.

Table- I: Sample Tweets

11	I must think about positive..
12	thanks to all the haters up in my face all day! 112-102
13	this weekend has sucked so far
14	jb ustr showing in ustralia any more!
15	ok I it you win.
16	<-------- This is the way I feel right now...
17	awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter. http://myloc.me/27HX
18	Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate
19	HUGE roll of thunder just now...SO scary!!!!
20	# I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. @shaunamanu is happy in the meantime.
21	Very sad about Iran.
22	wompppp wompp
23	You're the only one who can see this cause no one else is following me this is for you because you're pretty awesome
24	<---Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut down. Now it's all lost *lays in fetal position*

25	... Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place!!!!!!!!!!! I Think I Re-Ripped something !!! Yeah THAT !!
26	BoRinG): whats wrong with him?? Please tell me..... :-/

B. Multi-Modal Natural Language Processing Based Twitter Preprocessing

The next step of the work is multi-modal natural language processing based sentiment tweets analysis. The detailed structure of multi-model-based twitter sentiment analysis process is shown in figure 2. During the analysis process, the developed system not only analyze the text tweets but also examines the multi-modal comments [15] such as images, visual data and so on.

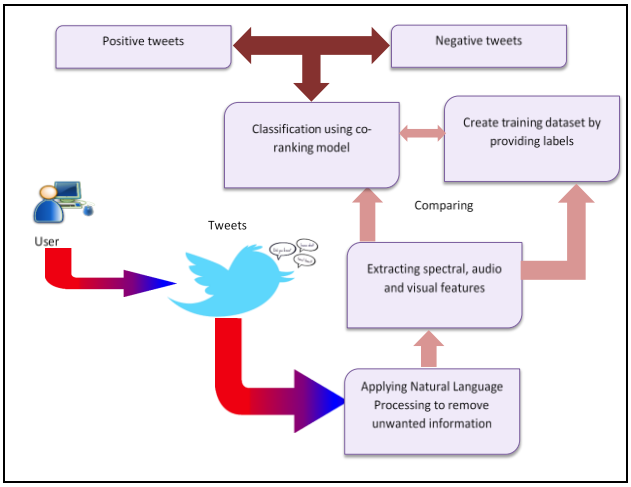

Figure 2: Sentiment analysis multi-modal Architecture Diagram

For example, the twitter data [16] consists of several video links, images for expressing their feeling in particular situation. From the comment decision has been taken to business discussion, political activity etc. First collect tweet from above discussed dataset, which is analyzed continuously for eliminating the unwanted characters [17] from tweets. The tweets consists of various symbols such as /,@, #, and URL which are unwanted while predicting sentiment details. First the URL needs to be removed from the gathered tweet, then the tag of the tweet need to be eliminated. The tag is identified using @ symbol. Then the # symbol must be removed from the unwanted data. This process eliminates the unwanted character from tweet successfully. Further the tweet is analyzed continuously to performing stemming process [18] which identifies the root word of the given tweet. The root word is predicted by removing prefix, suffix, and infix from the given tweet. Then the open NLP tool is applied to the tokenization process that successfully segment the given tweet into parts of speech (PoS), word boundary segmentation, and entity extraction. The extracted words are maintained in the table that used to predict the negative word from the given tweet. If the word having the negative version, then assign value as 0 and remaining neutral and positive tweets are having value 1. Based on the above discussion, the natural language processing based preprocessed sentiment tweet is described as follows.

Sample tweet:

 # I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. @shaunamanu is happy in the meantime.

Removing unwanted character, URL, /,@ from tweet I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. is happy in the meantime.

NLP processing, stemming and PoS process

Growing-grow, cut, happy, time.

Happy having positive tweet and provide vector as 1.

Based on the discussion, the removed tweets are maintained with root word for deriving the various features from the tweets. More ever, the tweet having image, video as the comment so different features are derived for predicting sentiments effectively.

C. Sentiment tweets related Feature Extraction

The third step of the work is deriving the feature [20] from the preprocessed tweets. The features are extracted from tweets based on the comma, interjection, words, question mark, punctuations. These semantic features are examined from the tweets, once, the tweet having any negative comment it does not have any positive tweets. After identifying semantic features, number of negative (NE) and positive (PO) words are identified with its ratio that is estimated as follows.

$$pone(t) = \frac{PO-NE}{PO+NE} \quad (1)$$

From the eqn (1), emotional related word value is set as 0, if the tweet does not contain any emotional tweet word. Then the slang of the tweet is estimated to determine the negative tweets, and the analyzed tweets are grouped according to the three classes such as positive(0) and negative (1) tweets, neutral. From the derived classes, training set [21] is formed because it is used to detect the sentiment related tweets easily. If the tweets having any audio related comments, phonetic property of the audio is computed. Then the spectral centroid of the audio tweet is estimated as follows,

$$centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (2)$$

In eqn (2), x(n) is denoted as the audio weighted frequency value, n is the number of bin and f(n) is center frequency value of audio. Then the spectral flux feature is extracted from the audio signal comment. The spectral flux feature identifies the power spectrum of the audio signal. For extracting the various audio features, opensmile tool is used in this research work. Further the visualization features are derived from the image and video related comment [22]. Open face tool is used to derive the facial expression and visualization features. Based on the derived features, the label is provided to the tweet which is treated as the training features. According to the discussion, the formed training feature set is depicted in table II.

Depending on the table II value, the incoming new testing tweets are examined to predict the sentiment effectively. Then the twitter related emotions are derived with the help of co-ranking algorithm that is discussed in the section D.

D. Sentiment Classification using Co-ranking Algorithm

The last step of this work is to classification of sentiments from the tweets which are done by applying co-ranking algorithm [23]. The incoming new tweets are processed by above discussed (section B) natural language processing techniques. The NLP technique effectively removes the unwanted characters from the tweets. Then the different features are extracted which are in text format that is covert into graph format. The extracted words are treated as node and construct the graph for predicting the sentiment analysis.

Table- II: Trained feature labeled tweets

Tweet ID	Labeled Tweets	Tweet
11	0	I must think about positive..
12	1	Thanks to all the haters up in my face all day 112-102
13	0	this weekend has sucked so far
14	0	jb isnt showing in Australiaanymore
15	0	ok thats it you win.
16	0	This is the way i feel right now...
17	0	awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter.
18	1	Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate
19	0	HUGE roll of thunder just now SO scary
20	0	I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. is happy in the meantime.
21	0	Very sad about Iran.
22	0	womppp wompp
23	1	You're the only one who can see this cause no one else is following me this is for you because you're pretty awesome
24	0	Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut down. Now it's all lost lays in fetal position
25	0	Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place I Think I Re-Ripped something Yeah THAT
26	0	BoRinG whats wrong with him Please tell me
27	0	can't be bothered. i wish i could spend the rest of my life just sat here and going to gigs. seriously.
28	0	Feeling like shit right now. I really want to sleep, but noo I have 3 hours of dancing and an art assignment to finish.
29	1	goodbye exams, HELLO ALCOHOL TONIGHT
30	0	I didn't realize it was THAT deep. Geez give a girl a warning at least

The related semantic words are analyzed and formed as edges for this purpose, senti wordnet is used for compute the word polarity value each node or word. The computed sent word allocates synset score to word as negativity, positivity and objectivity. From the computed polarity value, highest value is shown as bias and weights are determined between the edges. The edge value having negative word, then the weight value must be negative bias. According to the discussion signature words are computed from tweets, then the nodes are treated as word and form as the graph. Then the link between the edges is computed using ranking model to construct entire graph. The weights are computed according to the positive and negative words. Then the graph coefficient value is generate and computed as follows. First, graph opinion target value is estimated as follows.

$$C_t^{k+1} = (1-\mu) \times M_{to} \times C_o^k + \mu \times I_t, \tag{3}$$

Then the word opinion candidate value is computed

$$C_o^{k+1} = (1-\mu) \times M_{to}^T \times C_t^k + \mu \times I_o, \tag{4}$$

In eqn (3 and 4) C_0^k, C_t^k is represented as the confidence of opinion words and confidence of opinion target value, association of weights denoted as M_{to}^T and I_0, I_t represented as prior confidence of opinion word and target candidates. After computing the candidate value, negative and positive score for every word.

$$s_i^{(0)} = \begin{cases} s_i^{(0)} \Big/ \sum_{j \in D_{neg}^L} (-s_j^{(0)}), & if \;\; s_i^{(0)} < 0 \\ s_i^{(0)} \Big/ \sum_{j \in D_{pos}^L} s_j^{(0)}, & if \;\; s_i^{(0)} > 0 \end{cases} \quad i = 1, \ldots, n \tag{5}$$

We give "-1" to si(0) if di's label is "negative", and "1" if "positive". So we obtain the initial sentiment score vector S(0) for both domain data. This process is repeated and the computed output value is compared with the training set to predict the sentiment about the tweet is determined effectively. According to the discussion the algorithm steps are discussed as follows,

Algorithm for sentimental analysis

Step 1: Collect the twitter data from dataset

Step 2: Removing unwanted content from tweets using natural language processing techniques.

Step 3:Different type of features are need to be extracted from the tweets.

i) Identifying positive, negative features and their ratio is computed as follows,

$$pone(t) = \frac{PO - NE}{PO + NE}$$

ii audio phonetic property is applied to extract the audio based feature using above spectral centroid.

$$centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

iii Then spectral flux features are extracted from the audio signal power spectrum.

iv opensmile tool is applied to get the facial expression features from the video or image tweets.

Step 4: Arrange the feature and form the training feature set and label feature (0 for positive and 1 for negative).

Step 5: New incoming features are classified by applying co ranking algorithm, which form the graph while classifying the sentiments.

Step 6: The extracted features are compared with training set to recognize the sentiments from tweets.

Step 7: Repeat the process for every new incoming tweets.

Then the efficiency of the system is evaluated using the experimental analysis which is discussed in section III.

III. RESULTS AND DISCUSSION

The excellence of Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) Based Sentiment Analysis is discussed in this section. During this process, twitter data is collected from the awesome twitter dataset and the system is developed using the Python. The python based developed system utilizes the discussed (section II) discussed procedure. The successfully multimodal natural language processing based unwanted character elimination process improves the overall sentiment analysis process. During this process, 70% of twitter data is uses as training and 30% if data is uses as testing data which helps to determine the excellence of the system. The twitter sentiment analysis system efficiency [24] is determined using different parameter such as precision, recall, accuracy, f-score and ROC metrics. The co-ranking multimodal prediction system successfully selects the sentiment using training value. The efficiency of the system is analyzed using precision metrics that is computed as follows.

$$Precision = \frac{True\ positive}{True\ positive + false\ positive} \quad (6)$$

Table III: Precision

Methods	Number of Tweets									
	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500
CRMMNLP	98.78	98.83	98.84	98.76	98.97	99.03	98.93	98.76	98.84	98.09
LDA	97.24	97.32	97.53	97.98	97.12	97.35	97.52	97.65	97.34	97.41
LSA	96.87	96.53	96.62	96.78	96.83	96.29	96.37	96.45	96.53	96.86
FP-Growth	95.78	95.67	95.81	95.84	95.67	95.03	95.82	95.48	95.72	95.92
Apriori	94.19	94.83	94.76	94.64	94.97	94.71	94.91	94.83	94.79	94.83

Eqn (6), the successful prediction of right sentiment feature which is measured using true positive and predicting negative tweets. The incorrect prediction of positive and negative sentiment tweets is measured using false positive value. The CRMMNLP based obtained precision value is depicted in table III.

The table III demonstrates that the precision value of Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) compared with different twitter data sentiment analysis process such as linear discriminate analysis (LDA) [25], Latent Semantic Analysis (LSA) [26], Fp-growth [27] and apriori [28]. The discussed methods are effectively identified and select the sentiments from twitter data effectively. The introduced CRMMNLP method selects the right sentiment data from large volume of data using effective construction of graph effectively. Then the obtained result is shown in figure 2.

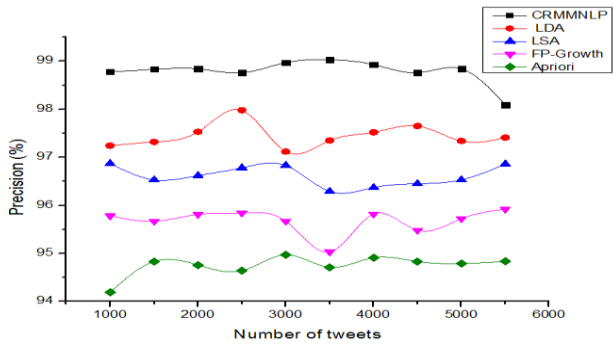

Figure 2: CRMMNLP-Precision value

Table IV: Recall

Methods	Number of Tweets									
	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500
CRMMNLP	97.368	97.65	97.87	97.95	97.43	97.89	97.63	97.85	97.84	97.68
LDA	96.45	96.76	96.83	96.47	96.95	96.82	96.65	96.37	96.18	96.27
LSA	95.36	95.12	95.38	95.52	95.47	95.68	95.79	95.39	95.69	95.12
FP-Growth	94.57	94.59	94.67	94.87	94.38	94.76	94.60	94.72	94.75	94.14
Apriori	93.21	93.67	93.76	93.80	93.65	93.6	93.96	93.58	93.89	93.18

The figure 2 demonstrates that the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) precision value on the awesome twitter data.

According to the graphical analysis, CRMMNLP method attains higher precision value (98.78%) compared to other sentiment prediction methods such aslinear discriminate analysis (LDA) (97.44%), Latent Semantic Analysis (LSA)(96.61%), Fp-growth (95.67%) and apriori (94.74%).

Along with the prediction value, recall of the sentiment prediction system is determined which is computed as follows.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (7)$$

In above eqn (7) effectively detected negative tweets from collection of tweets are named as true negative and incorrectly recognition of positive and negative tweets are named as false negative. According to the analysis computed recall value of different tweets are depicted in table IV.

The table IV demonstrates that recall value Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) compared with different twitter data sentiment analysis process such as linear discriminate analysis (LDA), Latent Semantic Analysis (LSA), Fp-growth and apriori. The effective removal of unwanted data helps to predict the particular sentiment relevant data that used to build the graph with accurate manner. The effective construction of graph helps to analyze huge volume of data effectively. Then the related graphical representation of recall is depicted in figure 3.

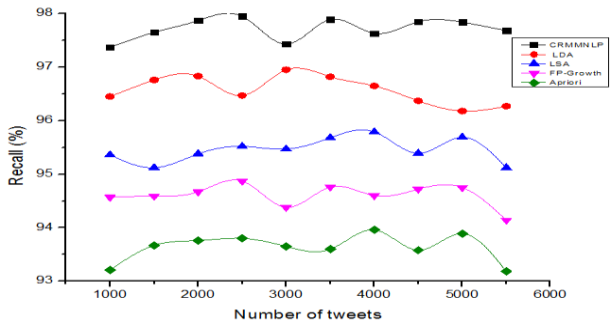

Figure 3: CRMMNLP-Recall

The figure 3demonstrates that the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) recall value on the awesome twitter data. According to the graphical analysis, CRMMNLP method attains higher recall value (97.71%) compared to other sentiment prediction methods such aslinear discriminate analysis (LDA) (96.57%), Latent Semantic Analysis (LSA)(95.45%), Fp-growth (94.60%) and apriori (93.63%). The successful prediction of positive and negative sentiments, the precision and recall value is improved. From the analysis overall efficiency of the system is determined using f-score value which is computed using eqn (8).

$$F - score = 2.\frac{precision.recall}{precision + recall} \quad (8)$$

Based on the eqn (8), f-score value is computed and depicted in table V

Table V F-Score

Num. of tweets	CRMM NLP	LDA	LSA	FP-Growth	Apriori
1000	98.074	96.845	96.115	95.175	93.7
1500	98.24	97.04	95.825	95.13	94.25
2000	98.355	97.18	96	95.24	94.26
2500	98.355	97.225	96.15	95.355	94.22
3000	98.2	97.035	96.15	95.025	94.31
3500	98.46	97.085	95.985	94.895	94.155
4000	98.28	97.085	96.08	95.21	94.435
4500	98.305	97.01	95.92	95.1	94.205
5000	98.34	96.76	96.11	95.235	94.34
5500	97.885	96.84	95.99	95.03	94.005

The effective extraction of tweets semantic features, audio and visualization features used to determine the word class whether it belongs to positive or negative one. Along with this, the co-ranking model effectively compute the candidate rank value of entire words present in the tweets successfully that used to improve overall sentiment prediction which is depicted in table 5. The analysis clearly tells that Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) approach attains high f-score value compared with different twitter data sentiment analysis process such as linear discriminate analysis (LDA), Latent Semantic Analysis (LSA), Fp-growth and apriori. Then the related graphical representation of f-score value is shown in figure 4.

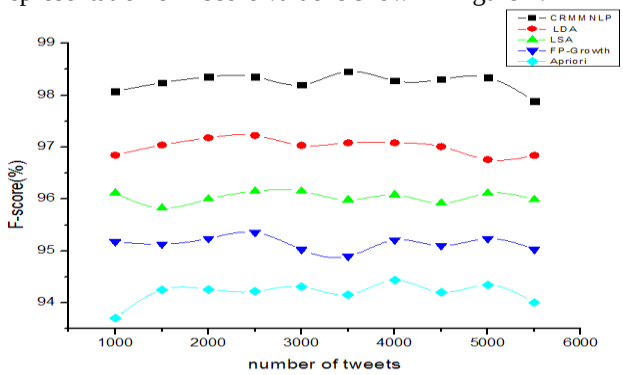

Figure 4:F-Score

The figure 4 demonstrates that the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) f-score value on the awesome twitter data. According to the graphical analysis, CRMMNLP method attains higher f-score value (98.24%) compared to other sentiment prediction methods such as linear discriminate analysis (LDA) (97.01%), Latent Semantic Analysis (LSA)(96.03%), Fp-growth (95.13%) and apriori (94.18%). The successful prediction of positive and negative sentiments, enhance prediction accuracy. According to the analysis, CRMMNLP method got effective accuracy on both testing and training process. Then the overall accuracy is depicted in table 6.

Table VI: Sentiment prediction overall accuracy

Methods	Sentiment prediction overall accuracy		
	Training Efficiency	Testing Efficiency	Entire Efficiency
CRMMNLP	98.57	99.04	98.805
LDA	97.86	97.84	97.85
LSA	96.35	96.18	96.265
FP-Growth	95.67	95.71	95.69
Apriori	94.28	94.45	94.365

Table VI demonstrates the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) based sentiment analysis overall accuracy (testing and training). Asusal CRMMNLP method attains maximum accuracy compared to other methods. Then the resultant accuracy is shown in figure 5.

The figure 5 demonstrates that the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) overall accuracy on the awesome twitter data. According to the graphical analysis, CRMMNLP method attains maximum accuracy (98.80%) compared to other sentiment prediction methods such aslinear discriminate analysis (LDA) (97.85%), Latent Semantic Analysis (LSA)(96.26%), Fp-growth (95.69%) and apriori (94.36%). Along with this, receiver operating characteristic curve (ROC) is analyzed to get the exact accuracy of CRMMNLP method based sentiment analysis process.

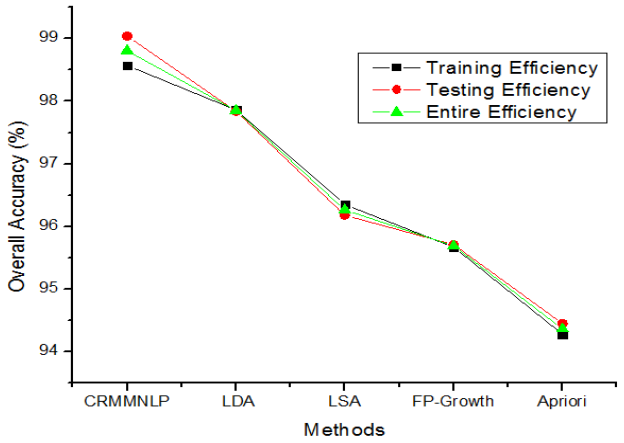

Figure 5: Sentiment Prediction Overall Accuracy

Then the ROC value is computed using false positive rate which is measured using eqn (9).

$$False\ positive\ rate = \frac{False\ positive}{false\ positive + true\ negative} \quad (9)$$

Based on the computation obtained ROC value is shown in figure 6.

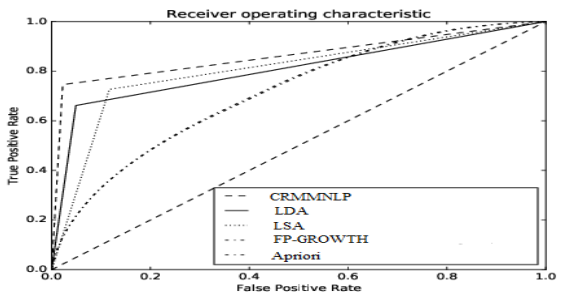

Figure 6: CRMMNLP-ROC curve

Thus the figure 6 depicted that the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) successfully recognize sentiments from tweets with accurate manner compared to traditional prediction methods.

IV. CONCLUSION

Thus the paper analyze the Co-Ranking Multi-Model Natural Language Processing Approaches (CRMMNLP) based sentiment prediction system. Then introduced system ability to handle the large volume of twitter dataset. So, the twitter data is collected from awesome twitter dataset which consists of several tweets. The collected information is processed by multi modal based natural language processing approach which effectively eliminates the unwanted characters from the given tweets. After that stemming operation is applied to get the root word of the tweet and the relevant features are extracted. The extracted features polarity value is computed to identify the negative and positive classes of words which are used to predict the sentiment about specific test tweets. Finally, incoming new tweets words are identified to form the graph which is done by using co-ranking model computation process. The polarity value is used to compare with the training table to classify the sentiments about tweets. Then the described system is developed using python tool and system ensures the highest prediction accuracy (98.8%) compared to other methods. In future the sentiment prediction system is improved using the optimized machine learning techniques.

REFERENCES

1. M.R.Huq, A. Ali and A. Rahman, “Sentiment Analysis on Twitter Data using KNN and SVM”, IJACSA International Journal of Advanced Computer Science and Applications, vol. 8, no. 6, pp. 19- 25, Jan. 2017
2. M.E. Larsen, T.W. Boonstra, P.J. Batterham, Bridianne O'Dea, C. Paris, and H. Christensen, “We Feel: Mapping emotion on Twitter, IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4,pp. 2168-2194, Feb. 2015
3. A. Kumar, V. Dabas, “A Social Media Complaint Workflow Automation Tool using Sentiment Intelligence”. Lecture Notes in Engineering and Computer Science: Proc. of The World Congress on Engineering 2016, pp.176-181
4. N. Wang, B. Varghese and P.D. Donnelly, “A Machine Learning Analysis of Twitter Sentiment to the Sandy Hook Shootings”, in Proc. 12th Int. IEEE Conf. e-Science, USA, 2016, pp. 3-312.
5. L. Bing and K.C.C. Chan, “Fuzzy Logic Approach for Opinion Mining on Large Scale Twitter Data”, in Proc. 7th In
6. Shihab Elbagir, Jing Yang, “Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn”, Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, Article No. 57.
7. Md Shad Akhtar, Ayush Kumar, Asif Ekbal, Pushpak Bhattacharyya, “A Hybrid Deep Learning Architecture for Sentiment Analysis”, Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers,pages 482–493, Osaka, Japan, December 11-17 2016.
8. Imane El Alaoui, GahiRochdi MessoussiYouness ChaabiAlexis TodoskoffAbdessamad Kobi, “A novel adaptable approach for sentiment analysis on big social data”, Journal of Big Data, December 2018, 5:12.
9. Ahmed Sulaiman M.AlharbiabElisede Doncker, “Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information”, Cognitive Systems Research, Volume 54, May 2019, Pages 50-61.
10. Hong G. (2005) Relation Extraction Using Support Vector Machine. In: Dale R., Wong KF., Su J., Kwong O.Y. (eds) Natural Language Processing – IJCNLP 2005. IJCNLP 2005. Lecture Notes in Computer Science, vol 3651. Springer, Berlin, Heidelberg
11. Yadollahi, Ali; Shahraki, Ameneh Gholipour; Zaiane, Osmar R. (25 May 2017). "Current State of Text Sentiment Analysis from Opinion to Emotion Mining". ACM Computing Surveys. 50 (2): 1–33. doi:10.1145/3057270
12. Ellis, Joseph G.; Jou, Brendan; Chang, Shih-Fu (12 November 2014). "Why We Watch the News". Why We Watch the News: A Dataset for Exploring Sentiment in Broadcast Video News. ACM. pp. 104–111. doi:10.1145/2663204.2663237. ISBN 9781450328852.
13. https://github.com/shaypal5/awesome-twitter-data
14. M. S. Neethu ; R. Rajasree, “Sentiment analysis in twitter using machine learning techniques”, Computing, Communications and Networking Technologies in IEEE, 2013.

15. Poria, Soujanya; Peng, Haiyun; Hussain, Amir; Howard, Newton; Cambria, Erik (October 2017). "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis". Neurocomputing. 261: 217–230. doi:10.1016/j.neucom.2016.09.117.
16. Perez Rosas, Veronica; Mihalcea, Rada; Morency, Louis-Philippe (May 2013). "Multimodal Sentiment Analysis of Spanish Online Videos". IEEE Intelligent Systems. 28 (3): 38–45. doi:10.1109/MIS.2013.9.
17. Mohamed Zakaria Kurdi (2017). Natural Language Processing and Computational Linguistics: semantics, discourse, and applications, Volume 2. ISTE-Wiley. ISBN 978-1848219212.
18. Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. Detecting offensive language in social media to protect adolescent online safety. In Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust, SOCIALCOM-PASSAT '12, pages 71–80, Washington, DC, USA, 2012. IEEE Computer Society. doi:10.1109/SocialComPASSAT.2012.55.
19. Kam-Fai Wong ; Wenjie Li ; Ruifeng Xu ; Zheng-sheng Zhang, "Introduction to Chinese Natural Language Processing", Synthesis Lectures on Human Language Technologies in IEEE,2008.
20. Domínguez M.A., Infante-Lopez G. (2008) Searching for Part of Speech Tags That Improve Parsing Models. In: Nordström B., Ranta A. (eds) Advances in Natural Language Processing. GoTAL 2008. Lecture Notes in Computer Science, vol 5221. Springer, Berlin, Heidelberg.
21. Yoon Kim, Yacine Jernite, David Sontag, "Character-Aware Neural Language Model", Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).
22. Pereira, Moisés H. R.; Pádua, Flávio L. C.; Pereira, Adriano C. M.; Benevenuto, Fabrício; Dalip, Daniel H. (9 April 2016). "Fusing Audio, Textual and Visual Features for Sentiment Analysis of News Videos". arXiv:1604.02612 [cs.CL].
23. Geetika Gautam ; Divakar Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis", Contemporary Computing in IEEE,2014.
24. Powers, David M W (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). Journal of Machine Learning Technologies. 2 (1): 37–63
25. Ying Ma, Ke Qin, and Shunzhi Zhu, "Discrimination Analysis for Predicting Defect-Prone Software Modules," Journal of Applied Mathematics, vol. 2014, Article ID 675368, 14 pages, 2014. https://doi.org/10.1155/2014/675368.
26. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing ContextualPolarity in Phrase-Level Sentiment Analysis," Proc. HumanLanguage Technology Conf. and Conf. Empirical Methods in NaturalLanguage Processing, pp. 347-354, 2005.
27. Yi Zeng, Shiqun Yin, Jiangyue Liu, and Miao Zhang, "Research of Improved FP-Growth Algorithm in Association Rules Mining," Scientific Programming, vol. 2015, Article ID 910281, 6 pages, 2015. https://doi.org/10.1155/2015/910281.
28. Li T., Luo D. (2014) A New Improved Apriori Algorithm Based on Compression Matrix. In: Luo X., Yu J.X., Li Z. (eds) Advanced Data Mining and Applications. ADMA 2014. Lecture Notes in Computer Science, vol 8933. Springer, Cham

AUTHORS PROFILE

Ms. Kanipriya M, received B.E degree in Computer Science and Engineering from Anna University Chennai and M.E. degree in Computer Science and Engineering from Anna University Chennai, India, in 2010 and 2012, respectively. She is now pursuing Ph.D. degree in Information and Communication Engineering at Hindustan Institute of Technology and Science Chennai. She has published research papers in International journals and conferences including reputed journals. Her research interests includes large database mining techniques, incremental data mining techniques, Big Data and Security, Image Processing and Machine Learning and social network analysis and mining.

Dr.Krishnaveni R. is currently working as professor at Hindustan Institute of Technology and Science, Chennai. She is a Ph.D holder from Anna University. She completed her B.E degree from Bharadhiyar University and M.E degree from Madras University. She has published around 30 research papers in International journals and conferences including Springer, IFIP, JCTN, IJNS and many other reputed journals. Her areas of research interests include Biometrics Security, Web Security, Information Security, Networking and Security, Cloud Computing and Security, Big Data and Security, Cyber Security, Image Processing and Machine Learning.

Dr. M Krishnamurthy is currently working as a Professor and Head in the Department of Computer Science and Engineering, KCG College of Technology. He leads a research group of Computer Science and Engineering faculty in KCG. He has published more than 46 papers in refereed international conferences and journals. His primary work and research interests include large database mining techniques, incremental data mining techniques, and social network analysis and mining. He also received prestigious award of "Best Java Architect" and "Distinguish Faculty in Data Mining" from Cognizant Technology Solutions and Venous International Foundation, Chennai in the year 2000 and 2016 respectively. He served as reviewer on several international journals including Knowledge and Information Systems (KAIS) Journal, International Journal of Physical Sciences and China Communications. International Journal of Knowledge and Information Systems. International Journal of Physical Science. Journal of King Saud University (Elsevier Editorial System). International Journal of Advanced Intelligence Paradigms (IJAIP). International Journal of Advanced Intelligence Paradigms (Inderscience) and ACM Computing Science.

Mr. S Bairavel, received B.E degree in Computer Science and Engineering from Manonmanium Sundaranar University Tirunelveli and M.E. degree in Computer Science and Engineering from Anna University Chennai, India, in 1997 and 2005, respectively. He is now pursuing Ph.D. degree in Information and Communication Engineering at Anna University Chennai. His research interests includes large database mining techniques, incremental data mining techniques, Big Data and Security, Image Processing and Machine Learning and social network analysis and mining.