DICTIONARY BASED APPROACH FOR IMPROVING THE ACCURACY OF OPINION MINING ON BIGDATA

A. Angelpreethi[1] & Dr. S. BrittoRameshKumar[2]
[1]Assistant Professor, Department of Information Technology, St. Joseph's College Tiruchirappalli, Tamilnadu, India.
[2]Assistant Professor, Department of Computer Science, St. Joseph's College, Tiruchirappalli, Tamilnadu, India.

Received: September 11, 2018 Accepted: October 23, 2018

ABSTRACT *In this modern world, everything is online. People use to express their feelings in the form of reviews in social media sites like twitter, face book, LinkedIn, You tube etc. user opinions increasing day by day because of the use of Internet. Opinion Mining or Sentiment Analysis plays a vital role to classify the opinions according to the user perceptions. While using internet people use different Slangs in their reviews. To classify these slangs is a challenging task. Therefore, Slang's analysis is crucial for sentiment identification. This paper presents a framework for Slang words classification and scoring of Internet Slangs using SentiWordNet in conjunction with other lexical resources. The comparative results show that proposed system outperforms the existing systems.*

Keywords: *Internet slang, opinion mining, sentiment analysis, Emoticon.*

1. Introduction:

In today's day and age, a person's opinion plays a vital role online. The volume and velocity of online reviews tend to grow rapidly day-by-day. Opinion is a person's perspective about an entity or event. Mining the users' opinions or Opinion Mining (OM) is also called as Sentiment Analysis (SA) [9, 14, 18]. With the quick raise in micro blogging sites such as Twitter, consumers are now relying on these services for their buying options, as are businesses for quick access to customer opinion.In the last few years, Twitter-based sentiment analysis applications have inward substantial interest from online consumers desiring information about a product and companies needing to respond rapidly to user opinions. Due to many features of these data, it is difficult to analyse the opinions using preceding approaches. Consequently, it is an important task to develop a method to extract and analyse twitter data through the automatic classification of tweets as Positive, Negative and Neutral.

A main challenge faced in developing such applications is slang words , Emoticons, Negations, Intensifiers, Domain-specific Words. This paper focuses on Slang words and Emoticons. Slang Words are the words informal and insufficient words, which are used by the users in their posts. Most of the previous works focused slang words and acronyms are same. But the facts that slang words include acronyms. For that, the proposed work separates the slang words into Derivative, Shortened, and Acronyms slang words. The advantage of splitting these slang words is to save the classification time. Emoticons are nothing but a graphical cue, which is used to express people's opinion in a graphical format. This research aims to improve the performance of Twitter-based sentiment classification by introducing Derivative Slang Classifier (DSC), Shortened Slang Classifier (SSC), Acronym Slang Classifier (ASC), Emoticon Classifier (EC), and Sentiwordnet Classifier (SWN).

The overall performance of the proposed work is tested with twitter data and it produces results with improved accuracy, precision, recall and F-Measure. This paper is organized as section 2 review of Literature, section 3 describe the proposed method.methodological diagram is presented. Section 4 , Results are discussed and section 5 conclusion.

2. Review of Literature:

OM plays an essential task in which it obtains opinions stated by its users via social media. There are three levels of opinion mining. Document level, Sentence Level and Aspect Level.

Taboada et al (2014)proposed a method is based on subjectivity of the text, while considering intensifiers, negations and opinionated words. They used Movie reviews and camera reviews using unsupervised lexicon based method and got 78.74% Accuracy.

zhenhua et al (2014)model has noticed the phenomenon of polysemy in single character emotional word in Chinese and discusses single character and multi character emotional word separately. This model can deal

with various kinds of short text data.They used Micro blog of COAE 2013, Movie reviews, product reviews using unsupervised method and got 58.40% accuracy.

Masud et al. (2014) projected a lexicon-centric approach that combines different lexicons and dictionaries for the sentiment analysis of tweets.Their main importance was on the accurate classification of sentiments with respect to slang in the tweets. Their proposed method has differentmodules, namely, (a) tweetcapturing and filtering, (b) subjectivity detection, and (c) sentiment scoring. Many lexicon s, such as the opinion lexicon, WordNet, SWN, andemoticon repositories are supported these modules.

Ribeiro, Weigang, and Li (2015) proposed a unified approach for performing Tweet-based sentiment analysis. The proposed approach iscomposed of four modules: (a) data collection, (b) noise reduction, (c) lexicon generation, and (d) sentiment classification. They introduced four algorithms to implement the aforementioned modules. The results obtained from the experiments conducted on the "iPhone 6" dataset demonstrated that the proposed approach is more effective than similar methods. However, to achieve more efficacy, further experiments arerequired on larger datasets with tweet scraping in streaming mode.

Liu et al (2015)aims at the integration of different emotional clues into a unified model and trains on both tagged and untagged datasets by switching the propagation phenomenon alternately. The experiments conducted on multiple datasets demonstrated the effectiveness of the proposed approach.

Gelbukh et al (2015) proposed a convolutional multiple kernel learning-based method for the sentiment analysis of shortmultimedia content, such as text and audio and video clips. Features are extracted from the multimedia content by applying activation values inthe inner layer of a deep convolutional neural network model. The results show that an improvement of about 14% is achieved over the baselinemethods.

Alani et al (2016) proposed a lexicon-supported technique for Twitter-based data analysis that captures the sentimentclass of words in numerous contexts and updatesthe sentiment scores accordingly. Their approach is based on co-occurrence word signals indifferent domains at both the tweet and entity levels. The proposed approach was evaluated using three Twitter datasets and achieved 4–5%higher accuracy than the comparison methods.

Asghar et al (2017) proposed hybrid based slang words classification. They used many classifiers to detect slangs, acronyms. In addition, they used SWN for sentimentLimited work is carried out in slang word detection. All the existing work treats slang words are Acronyms. Lack of slang classification is a Limitation of the Existingwork. This work focuses on Slang words classification based on the usage and scoring of such words using SWN.

One of the main reasons for slang usage is related to time constraints. There are many works concerning the user's sentiment posted on social media networks with the focus on classifying the opinion as positive, negative and neutral opinions.

3.Proposed Opine_Lexi Approach:

Lexicon based approach entirely plays a major part to decide the classes such as positive, negative and neutral. In lexicon, based approach is to extract and handle the sentiment as no slang words. The most of the authors have given suggestions to handle the slang words but none has properly handled or created any lexicon dictionary for the internet slang words. The sentiments are as followed in many dictionaries, which are named as lexicon-based dictionaries, which are (1) Bing Liu's Opinion Lexicon (2) MPQA Subjectivity Lexicon (3) Senti Word Net Lexicon (4) Semantic Evaluation (SemEval). Figure 1 shows the methodological diagram of lexicon based slang words and emoticon classification.

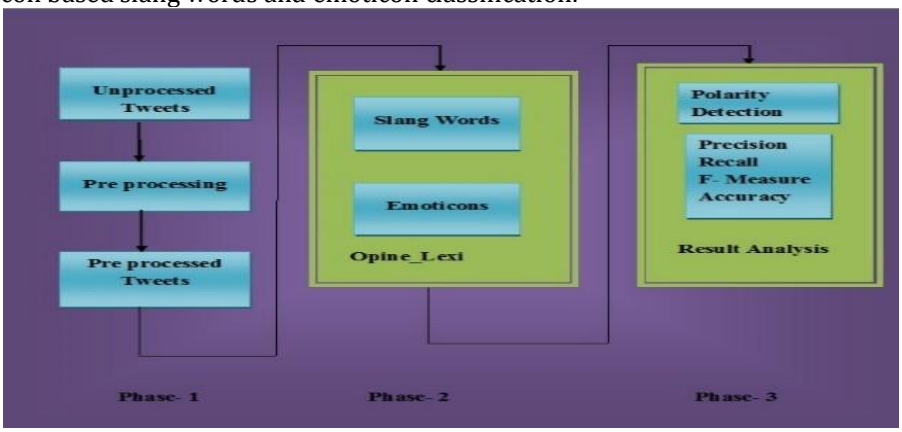

Figure1. Methodology diagram for slang words and emoticon classification

The proposed system performs and classifies the slang words and emoticons using 5 classifiers namely

1. Derivative slang Classifier(DSC)
2. Shortened slang Classifier(SSC)
3. Acronym slang Classifier(ASC)
4. Emoticon Classifier(ESC)
5. SentiWordNet Classifier(SWNC)

classify the tweets as positive, negative and neutral. In third Phase the tweets are classified as positive, negative and neutral and results are discussed in the discussion section of this article. The frequently used polarity measures such as precision, recall, accuracy and F-score are calculated to measure the accuracy of the Opine_Lexi approach.

The proposed work consists of three Phases.

First phase the collected tweets are preprocessed and stored in the desired format (.txt and .csv). The preprocessing step includes tokenization, Stemming and lemmatization,Removal of irrelevant content, Transliteration, and POS Tagging. These processes are briefly explained in the following subsections.

In Second Phase Opine_Lexi Approach has been proposed for handling the Slang words and Emoticons and are briefly explained in the proposed work section. A mathematical equation has been framed for Sentiment Score Computation (SSC) to First phase the collected tweets are preprocessed and stored in the desired format (.txt and .csv). The preprocessing step includes tokenization, Stemming and lemmatization,Removal of irrelevant content, Transliteration, and POS Tagging. These processes are briefly explained in the following subsections.

In Second Phase Opine_Lexi Approach has been proposed for handling the Slang words and Emoticons and are briefly explained in the proposed work section. A mathematical equation has been framed for Sentiment Score Computation (SSC) to classify the tweets as positive, negative and neutral.In third Phase the tweets are classified as positive, negative and neutral and results are discussed in the discussion section of this article. The frequently used polarity measures such as precision, recall, accuracy and F-score are calculated to measure the accuracy of the Opine_Lexi approach.

3.1 Phase 1:

3.1.1Unprocessed Tweets:

This module can gain the data from the users tweets using twitter API. These tweets contain many attributes, which include username, date and time, location, re tweet status, re tweet count.

3.1.2 Pre- Processing:

After collecting Data from the twitter the data needs to be pre-processing. Data pre-processing is an essential task in opinion mining during this phase, the opinion mining system prepares collected data for further processing. This involves six steps.

a. Tokenization.

Tweets are splited or segmented into individual words, or tokens.

b.Lowercasing.

Characters are converted to lower case to ease the process of matching words in tweets.

c. Stemming and lemmatization.

To more facilitate word matching, words in tweets are converted to their root word for example; “placing,” “placed,” and “placement” is all converted to “place.”

d. Removal of irrelevant content.

Punctuation and stop words, which are irrelevant for opinion mining, are removed to improve system response time and effectiveness.

e. Transliteration.

To address the issue of use of mixed Language in tweets the text is transliterated using Google Transliterate API.

f. POS Tagging:

POS Tagging is also called Word Category Disambiguation and Grammatical Tagging. The Process ofclassifying the words into their Parts of Speech and tokenize them accordingly.The Algorithm for Pre-Processing is depicted in Figure 2.

Input: Raw Tweets
Output: Pre-processed Tweets

1. Begin
2. For each tweet t in dictionary Do
3. Tokenize each word in a tweet
4. For each token belongs to t then
5. Remove stop words
6. if (2<length of the token>10)
7. Go to step 9
8. Transform cases
9. Ignore token
10. Perform POS tagging
11. Handle Negations

Figure 2. Pre- processing Algorithm

3.2 Phase 2: Slang Words classification and Emoticon Detection

After Pre-processing the pre -processed tweets filters non- slang words separately to save the running time of a detection process. Non- slang word is nothing but the word is found in an opinion lexicon or Dictionary. If the word is not found, then it is considered as Slang word.

Slang Words are further classified based on the usage, root meaning, and Pronunciation. Slang words are classified as Derivative Slang, Shortened Slang, and Acronym Slang. Each slag word has its own classifiers.

Unigram based Feature selection is applied to the pre-processed tweets which splits the words individually with the identification of whitespace. It checks the incoming word matches with the any one of the dictionary either slang or emoticon dictionary. Suppose the word is not found the word equivalent semantic word like grt into great, n8 into night, :-(into happy. Else it identifies a word slang word or Emoticon but if the word is not in a dictionary then it will be inserted into the dictionary with the equivalent meaning.

3.2.1 Derivative Slang Dictionary (DSD):

Slang words that have been derived from other words or have their root meaning in another word, which gives a similar meaning when used is called aDerivative Slang words. Figure 3 shows the Derivative Slang Classifier.

SPC= Strongly Positive Category
PC= Positive Category
SNC= Strongly Negative Category
NC= negative Category

$$\textbf{Derivative slang Classifier(DSC)} = \begin{cases} 1 & \text{if } W_i \in \text{SPC} \\ 0.5 & \text{if } W_i \in \text{PC} \\ -0.5 & if\ W_i \in \text{SNC} \\ -1 & if\ W_i \in \text{NC} \\ 0 & otherwise \end{cases}$$

Figure 3 Derivative Slang Classifier

3.2.2 Shortened Slang Dictionary(SSD):

Words which are formed in a way when pronounced as a word it will sound as same as the root word, but although with minimal amount of a combination of letters, numerals and combination of both figure 4 shows the Shortened Slang Classifier.

$$\textbf{Shortened slang Classifier(SSC)} = \begin{cases} 1 & \text{if } W_i \in \text{SPC} \\ 0.5 & \text{if } W_i \in \text{PC} \\ -0.5 & if\ W_i \in \text{SNC} \\ -1 & if\ W_i \in \text{NC} \\ 0 & otherwise \end{cases}$$

Figure 4 Shortened Slang Classifier

3.2.3Acronym Slang Dictionary (ASD):

Acronym words are the words formed from the initial letter or letters of each of the consecutive parts or major parts of a compound term. figure 5 shows the Acronym Slang Classifier.

$$\textbf{Acronym slang Classifier(ASC)} = \begin{cases} 1 & \text{if } W_i \in \text{SPC} \\ 0.5 & \text{if } W_i \in \text{PC} \\ -0.5 & if\ W_i \in \text{SNC}\ z \\ -1 & if\ W_i \in \text{NC} \\ 0 & otherwise \end{cases}$$

Figure 5 Acronym Slang Classifier

3.2.4 Emoticons Dictionary (ED):

As public progressively more use emoticons in text in order to convey, strain, or disambiguate their sentiment, it is crucial for automatic sentiment analysis tools to correctly account for such graphical cues for sentiment. Emoticons are used to express people's sentiment.Figure 6 shows the Emoticon Classifier

$$\textbf{Emoticon Classifier(EC)} = \begin{cases} 1 & \text{if } e_i \in \text{SPC} \\ 0.5 & \text{if } e_i \in \text{PC} \\ -0.5 & if\ e_i \in \text{SNC} \\ -1 & if\ e_i \in \text{NC} \\ 0 & otherwise \end{cases}$$

Figure 6 Emoticon Slang Classifier

3.2.5 Sentiwordnet:

Senti word net (SWN) is a lexical resource for opinion mining. SWN has wide range of words with sentiment scores. SWN has more than 60,000 synsets obtained dynamically from Word Net. Each of the words is tagged with three types of sentiment scores namely positive, negative and neutral. The score range in the interval 0.0 to 1.0 for each word.

$$\mathbf{P_{score}(W_i, P)} = \frac{1}{S}\sum_{i=1}^{n} Pscore(W_i)$$

$$\mathbf{N_{score}(W_i, P)} = \frac{1}{S}\sum_{i=1}^{n} Nscore(W_i)$$

$$\mathbf{Neu_{score}(W_i, P)} = \frac{1}{S}\sum_{i=1}^{n} Neu\_score(W_i)$$

Figure 7 Average scores

S= Total number of synsets of a term for a corresponding POS.
P= Part of Speech tagger like adjective, noun, adverb, verb

There may be numerous meanings of a word in a specific grammatical type in SWN. To disambiguate numerous meanings in a particular type we compute the average of the positive, negative and neutral scores which is depicted in figure 7.

After finding the average for different synsets of a word W_i in a POS category, three scores are obtained namely positive, negative and neutral. Figure 7 shows the mathematical expression for finding the average for different synset values and Figure 8 shows the senti word Net based classifier.

Wi is positive if the scores of negative and neutral must be minimum. Same way negative polarity also computed based on comparing other two that is positive and neutral. Neutral score is computed if the scores of positive and negative are equal orneutral score is greater than the positive and negative.Based on the opine_Lexi algorithm the tweet will be classified as Positive, negative and neutral. SWN classifier is depicted in Figure 8.

$$\mathbf{SWN}_{\mathbf{Senti}_{\mathbf{Score}}(\mathbf{W_i},\mathbf{P})} = \begin{cases} \mathbf{P}_{\mathbf{score}}(\boldsymbol{Wi},\mathbf{P}) & \text{if } (\text{pscore}(\text{W}_\text{i},\text{P}) > nscore(wi,\text{p}))\wedge(\text{pscore}(\text{wi},\text{p}) > neu_{score\ (wi,\text{p})}) \\ \mathbf{N}_{\mathbf{score}}(\mathbf{W_i},\mathbf{P}) & \text{if } (\text{Nscore}(\text{wi},\text{p}) > Pscore(wi,\text{p}))\wedge(\text{Nscore}(\text{wi},\text{p}) > neu_{score\ (wi,\text{p})}) \\ \mathbf{Neu}_{\mathbf{score}}(\boldsymbol{Wi},\mathbf{P}) & \text{else} \end{cases}$$

Figure 8 SentiWordNet based classifier

Overall Tweet sentiment using SWN

$$= \begin{cases} \boldsymbol{Positive} \sum_{i=1}^{n}(\boldsymbol{DSC}+\boldsymbol{ASC}+\boldsymbol{SSC}+\boldsymbol{EC}+\boldsymbol{SWN\ sentiscore}) > 0 \\ \boldsymbol{Negative} \sum_{i=1}^{n}(\boldsymbol{DSC}+\boldsymbol{ASC}+\boldsymbol{SSC}+\boldsymbol{EC}+\boldsymbol{SWN\ sentiscore}) < 0 \\ \boldsymbol{Neutral} \sum_{i=1}^{n}(\boldsymbol{DSC}+\boldsymbol{ASC}+\boldsymbol{SSC}+\boldsymbol{EC}+\boldsymbol{SWN\ sentiscore}) = \mathbf{0} \end{cases}$$

Figure 9 Overall Tweet sentiment using SWN

```
APPROACH: Opine_Lexi
INPUT: Raw Tweets
OUTPUT: Sentiment Score { P,N,Neu} Where P: Positive Score, N: Negative
Score and Neu: Neutral Score.
t← term, T←Tweet  ED ← Emoticon Dictionary  SSD← Shortened Slang
Dictionary  DSD ← Derivative Slang Dictionary  ASD ← Acronym Slang
Dictionary OD← Opinion Dictionary
    1.  Begin
    2.  While (t€ T) Do
    3.  Call pre-processing (t)
    4.  Search each t € ( OD v  ED v SSD v DSD v ASD )
    5.  If ( t € OD )
    6.  Subjective
    7.  Assign score using SWN senti_score(subjective tweet)
    8.  Else
    9.  Objective tweet
    10. End if
    11. For each t € T
    12. If Word found in DSD
    13. Execute classification using Derivative slang classifier
    14. If Word found in SSD
    15. Execute classification using Shortened slang classifier
    16. If Word found in ASD
    17. Execute classification using Acronym slang classifier
    18. If Word found in ED
    19. Execute classification using Emoticon classifier
    20. Execute classification using SWN classifier
    21. Classify t using SWN classifier and affirm positive, negative and
        neutral
    22. End while
```

Figure 10Opine_Lexi Approach

4. RESULTS AND DISCUSSIONS:

The proposed Opine_Lexi Algorithm approach is to evaluate the sentiment knowledge on big data using Lexicon based Approach. Tweets are collected from Twitter using Tweepy. Tweepy is an open source and easy to use python library for accessing the Twitter API. After collecting, the tweets are pre-processed and store it in a HDFS. HDFS is a Hadoop Distributed File System. It is designed to store very large datasets. In Pre-processing tweets are tokenized and store it in a HDFS. Tweets are filtered based on SWN. Because SWN have only formal words. Based on the occurrence of the SWN, tweets are filtered by slang words, and

emoticons. Each term of the tweets are searched whether it belongs to opinion dictionary, or Emoticon Dictionary or Derivative Slang Dictionary or Shortened Slang Dictionary or Acronym Slang Dictionary. If a tweet is matched with shortened slang dictionary, then it perform the classification using shortened slang classifier. Then the tweets are scored based on SWN classifier. Based on the Score Tweets are classified like positive, negative and neutral.

For Example

"After having mobile phone I am having the worst depression Damn☹"

Here Derivative Slang Classifier score = -0.16477

Acronym Slang Classifier Score =0

Shortened Slang Classifier Score=0

There is no slang expressed in this review.

Emoticon Score = -0.5 negative Emoticon

Opinion score: Worst – (-0.75) and depression – (0.85 neutral score)

SWN Senti score = -0.16477+-0.75+0.85+ (-0.5) = -0.56477(negative)

Therefore, the overall polarity of this sentence is Negative. Without Emoticon, the overall score is positive but actually, it is a negative sentence. Therefore, Emoticon can take a vital part in polarity classification.

S.No	AUTHOR	ACCURACY (%)
1	Zhenhua et al	58.40%
2	Alexander et al	59.50%
3	Hussam et al	64.27%
4	Ayushi et al	67.04%
5	Saprativa et al	68.46%
6	Edison et al	68.75%
7	Masud et al	81.00%
8	Proposed	**83.38%**

Table1 comparison of results.

Tweets	Total
Derivative slang	70,288
Shortened Slang	56,446
Acronym Slang	68,130
Emoticon	72,320

Table2. Total number of slangs and Emoticons

Classifiers	class	Results (in percentage)			
		Precision	Recall	F-Measure	Accuracy
DSC	Positive	84.59	82.65	83.60	59.24
	Negative	68.54	56.24	61.78	
	Neutral	16.48	65.58	26.34	
SSC	Positive	54.28	63.33	58.45	57.21
	Negative	64.87	35.58	45.95	
	Neutral	65.54	63.31	64.40	
AC	Positive	98.54	46.25	62.95	64.00
	Negative	78.82	81.64	80.20	
	Neutral	31.85	97.24	48.00	
EC	Positive	95.56	25.46	40.21	27.71
	Negative	12.46	16.12	14.05	
	Neutral	13.56	98.45	22.83	
SWNC	Positive	63.98	60.87	31.19	57.24
	Negative	55.48	57.54	56.49	
	Neutral	78.20	85.50	81.68	

Table3. Individual Results of the confusion Matrix

	P	R	F	A
Opine_Lexi	0.8370	0.8227	0.8298	0.8338

Table 4. Proposed work results

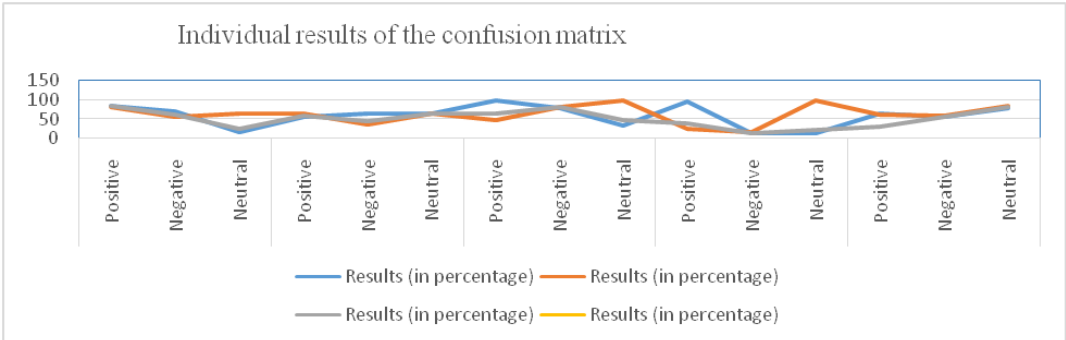

Figure 11. Results of the confusion matrix

Table 1 displays the comparision of existing works results and Table 2 depicts the Data used in this work and Table 3 displays the Individual Results of the confusion matrix. And Figure 11 Shows the diagramatic representation of the results of confusion matrix and figure 12 dispays the pictorical representation of Proposed work results.

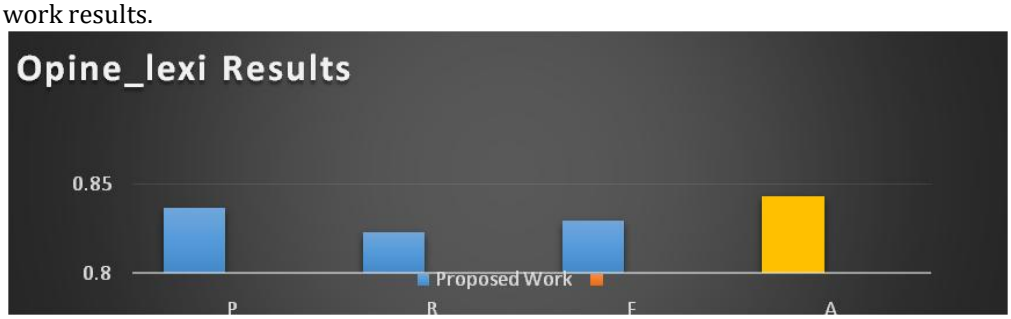

Figure 12. Proposed Work

5. Conclusion:

This paper focuses on the adoption of various sentiment classifiers using hybrid classification scheme. The proposed approach is based on unigram model, which produces better accuracy than the existing work. The proposed work consists of Derivative slang classifier, Shortened classifier, Acronym Classifier and Emoticon Classifier. These classifiers are getting polarity value of each synset using SWN. Finally, SWN aggregate the value by dividing number of synset values with sentiment class. Opine_Lexi Algorithm produces improved results than the existing work. In future, this work will be implemented with Domain specific words.

References:

1. B. Liu, "Sentiment analysis and subjectivity." Handbook of natural language processing, vol. 2, pp. 627–666, 2010.
2. MaiteTaboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede "Lexicon-Based Methods for Sentiment Analysis", Association for Computational Linguistics vol 37, pp 267-307,2011.
3. Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers", 2012, pp.7-140.
4. Anna Jurek, Maurice D. Mulvenna and Yaxin Bi, Improved lexiconbased sentiment analysis for social media analytics Science direct, Published: 9 December 2015.
5. James Spencer and Gulden Uchyigit, Sentimentor: Sentiment Analysis of Twitter Data. School of Computing, Engineering and Mathematics. University of Brighton.
6. CataldoMusto, Giovanni Semeraro, Marco Polignano, "A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts", Department of Computer Science, University of Bari Aldo Moro, Italy
7. A Kowcika and Aditi Guptha "sentiment Analysis for social media" ,International journal of advanced research in computer science and software engineering,216-221,Volume 3,Issue 7, july 2013.
8. Vinodini and RM.Chandrashekaran, "sentiment analysis and opinion mining: A survey", International journal of advanced research in computer science and software enginnering,283- 294, Volume 2,Issue 6, june 2012.

9. Anna Jurek, Maurice D. Mulvenna and Yaxin Bi, Improved lexiconbased sentiment analysis for social media analytics Science direct, Published: 9 December 2015.
10. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, 2002, pp. 79–86.
11. D.-H. Pham and A.-C. Le, "A Neural Network based Model for Determining Overall Aspect Weights in Opinion Mining and Sentiment Analysis," Indian Journal of Science and Technology, vol. 9, no. 18, 2016.
12. Kundi, F. M., Ahmad, S., Khan, A., &Asghar, M. Z. (2014). "Detection and scoring of internet slangs for sentiment analysis using SentiWordNet", Life Science Journal, 11(9), 66–72.
13. Muhammad ZubairAsghar, FazalMasudKundi, Shakeel Ahmed, Aurangzeb Khan, Furqan khan " T-SAF: Twitter Sentiment Analysis Framework using a hybrid classification scheme", Wiley Expert systems, 2017 pp-1-19.
14. FazalMasud Kundi1, Shakeel Ahmad1, Aurangzeb Khan2, Muhammad ZubairAsghar, "Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet" Life Science Journal 2014.
15. A.Angelpreeethi, Dr.S.Britto Ramesh Kumar, " An Enhanced Architecture for Feature based opinion mining from product reviews", World Congress on Computing and Communication Technologies , IEEE, 2017 pp 89-92.
16. Umar Farooq, Tej Prasad Dhamala, Antoine Nongaillard, YacineOuzroutand Muhammad Abdul Qadir, "A Word Sense Disambiguation Method for Feature Level Sentiment Analysis", 9th International Conference on Software, Knowledge, Information Management and Applications,IEEE, 2015.
17. Vibhor Varshne1 and Rupali Sunil Wagh, "Weighted Sentiment Score Formulation Using Sentence Level Sentiment Density for Opinion Analysis" International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 2 (2017), pp. 285-298.
18. Jie Li; LirongQiu," A Sentiment Analysis Method of Short Texts in Micro blog", IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), pp 776-779, 2017.
19. A.Angelpreethi, Dr.S.Britto Ramesh Kumar," Opinion Mining on Hybrid Approach: A Perspective", International Journal of scientific research in computer science and management studies, volume 7, issue 5 (sep 2018)
20. Alexander Pak and Patrick Paroubek, 2010. Twitter as a corpus for sentiment analysis and opinion mining. In the Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), pp: 1320-1326.
21. Ameeta Agrawal, Aijun An "Kea: Sentiment Analysis of PhrasesWithin Short Texts", Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 380–384, Dublin, Ireland, August 23-24, 2014.
22. Dibakar ray, " Lexicon based sentiment analysis on twitter Data", International Journal for Research in Applied Science & Engineering Technology,Volume 5, Issue X, October 2017.