# An Enhancement of the Public Sentiment Analysis on Social Networking by Improving Sentiment Analysis Tools

**Osamah Ali Mohammed Ghaleb[1*]**      **Anna Saro Vijendran[2]**

[1]*Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, 641006, India*
[2]*School of Computing, Sri Ramakrishna College of Arts and Science, Coimbatore, 641006, India*
* Corresponding author's Email: Osama.ghalib85@gmail.com

**Abstract:** One of the most widely used medium of communication over internet is social media networking where users can share and discuss their opinions. In such medium, it is highly important to track the sentiments of people to make wise decision. A combination of sentiment analysis and sentiment variation methods is used to track the variation of in public sentiments. TwitterSentiment and SentiStrength tools are two state-of-the-art sentiment analysis tools. The main aim of this paper is to improve sentiment analysis using Improved TwitterSentiment and SentiStrength in order to achieve more accurate sentiment labeling result, thus will improve tracking and interpretation of public sentiments. TwitterSentiment has been enhanced using Semantic Role Labeling (SRL), which extract semantic arguments and roles and Kullback-Leibler Divergence (KLD) for labeling the sentiments based on the extracted roles, and this is referred to as Improved TwitterSentiment. Further, the issues in SentiStrength are resolved using Natural Language Processing (NLP) technique and Maximum entropy classifier, and this is referred to as Improved SentiStrength, which aims to extracts the features of poor grammars in short informal tweets and classifying sentiments accurately. The results of both the Improved TwitterSentiment and Improved SentiStrength are then considered to label the sentiments. Thus, the performance of the sentiment analysis enhanced through the combination of Improved TwitterSentiment and Improved SentiStrength. This paper proposed three improved methods. For experimental study purposes dataset has been collected from Twitter about two key targets namely "Narendra Modi" and "Apple". The performance of the proposed sentiment method is found to be efficient with high accuracy of 0.87, high precision of 0.89, high recall of 0.85 which has been outperformed the existing methods.

**Keywords:** Sentiment analysis, Improved TwitterSentiment, Improved SentiStrength, Semantic role labeler, Natural language processing.

## 1. Introduction

Sentiment analysis, it otherwise called opinion mining, characterized as one of Natural Language Processing (NLP) fields for the most part used to track the cleverness of people in general about the particular item [1]. It investigates of individuals' opinions, sentiments, assessments, evaluations, demeanours and feelings towards products, for example, items, services, associations, etc. Sentiment analysis is for the most part concentrates on distinguishing and classifying opinions which are utilized to express either positive or negative sentiments. In current years, this field has turned into a more dynamic territory of research because of the significance of clients' opinions in basic decision-making process.

Alongside the improvement of online networking like web journals, Twitter, and so forth, people and associations are progressively using the substance in these media with a specific end goal to settle on educated choices. Regularly, microbloging, social and audit destinations incorporate a lot of opinion content which can't without much of a stretch deciphered in gathering postings. It is hard to recognize the important destinations, separate and abridge the opinions. Thus, having a programmed

opinion mining frameworks would empower separating and investigating the opinions on the web. Sentiment analysis is one of the branches of Natural Language Processing, which is utilized to recognize the passionate extremity of a given content in a web. Regularly the extremity of the sentiment is arranged into three classes to be specific as positive extremity, negative extremity and impartial extremity. Distinctive levels of sentiment analysis are presented [2], many examinations dissected the sentiments utilizing those levels, and common levels that are studied by many authors are document level [3], sentence level [4], and entity level and aspect level sentiment analysis [5]. The sentiment analysis on long range interpersonal communication clients gives a most financial and compelling method for uncovering general opinion, which is utilized to settle on educated choices in various areas.

Many research studies and modern applications have been concentrating on the territory of public sentiment following and demonstrating. In any case, there is no exploration is performed to decide the primary driver behind the public sentiment variations successfully [6]. Knowing the reasons of sentiments varieties will empower to mine valuable bits of knowledge, and giving essential basic decision making data. It is more essential to locate the conceivable explanations for the varieties of public sentiments. For the most part finding the correct reason for the sentiment variety is testing undertaking for number of reasons controlled by [6]; it ought to be noticed that the developing subjects that are talked about in the variety time frame are thought to be unequivocally identified with the explanations for the sentiment varieties. Be that as it may, mining rising points in Twitter still a testing one. The sentiment varieties of the public sentiments can be controlled by sentiment analysis and sentiment variety strategies. In the sentiment analysis process, the sentiments were classified by both the TwitterSentiment and SentiStrength instruments. TwitterSentiment is exceptionally disposed in light of the fact that it misinterprets the non-neutral opinions as a neutral, in this way it influencing the precision of sentiment analysis. Further, SentiStrength leaves short and casual sentence structure content for sentiment analysis. In this way it is profoundly vital to improve the TwitterSentiment and SentiStrength by resolved the addressed issues keeping in mind the end goal to perform sentiment analysis effectively.

As higher efficiency is not achieved from the existing tools for sentiment analysis, this paper attempts to resolve the drawback of TwitterSentiment by introducing Semantic Role Labeling (SRL) and Kullback-Leibler Divergence (KDL) classifier inspired by Ponti [7]. The motivation for the proposed approach is that analysis of social media user opinions provides rich structured information about the individuals involved in the communication; thus enhancing identification of the common specified opinion of the maximal users. The detection of these opinions can help in determining the current mood of customers or common people regarding a product, service or event, as the tweet is viewed all over the world. The improved models of TwitterSentiment and SentiStrength are developed by utilizing the efficient labeling and feature extraction processes. This method of improving the performance of existing tools has also resulted in better sentiment analysis. Our work is aimed to improve the existing tools with the purpose of enhancing the accuracy of sentiment classifications. Thus, enabling more accurate analysis and interpretation of the sentiment variation provide important information for decision making in different domains. In the proposed approach, the SRL identifies semantic arguments and roles in a sentence. Based on the argument type, the class of the sentiments is assigned using the KLD method. The identified semantic arguments enhance the classification process which classifies the tweets more efficiently into accurate classes. This process called as Improved TwitterSentiment. In addition to that, the drawback of SentiStrength is resolved using NLP and Maximum entropy technique. The NLP technique extracts the features of poor grammar text and gives it as input to the maximum entropy technique to learn the class of the sentiment. The NLP process accompanied with the classifier enhances the sentiment classification of SentiStrength with the consideration of the high end grammar for general as well as short sentences. Finally, the sentiments are labelled based on the results of combining both Improved TwitterSentiment and Improved SentiStrength. Comparison of these models of sentiment analysis with those of the existing ones will be helpful in illustrating the performance enhancement.

The remainder of this paper is organized as: section 2 presents some of best known research works in the sentiment analysis concepts. Section 3 explains the process of the proposed sentiment analysis models while section 4 illustrates the performance evaluation through experimental results. Section 5 makes a conclusion about this paper.

## 2. Related work

Intelligent representation framework for sentiment analysis named as SentiView [8] analyses public sentiments to know about popular topics. This should be possible via looking and connecting frequent words in the public sentiments content data. The SentiView framework mines and models the varieties of sentiments on public topics. Keeping in mind the aim to represent the sentiments and to envision the progressions of relationship among socioeconomics of intrigue and various characteristics, a period changing helix alongside a trait astrolabs is utilized. The relationship outlines the relationships of enthusiasm among various members. Be that as it may, the difficulties in keeping up a relationship delineate distinctive members prompted advancement of semi-regulated theme versatile sentiment classification (TASC) show [9] for dynamic Tweet sentiment classification. This model arranges sentiments as positive, negative or unbiased utilizing a classifier, which is based on normal components and blended names from different topics. This model decreases the pivot misfortune to adjust unlabeled information and components, including the point versatile elements. Both content and non content elements in Tweet were separated and isolated into two perspectives for co-training. The TASC algorithm was utilized to refresh point versatile components in light of the synergistic choice of unlabeled information. It chose more dependable tweets which enhanced the tweet sentiment classification.

Sentiment Analysis by Collective Inspection (SACI) strategy [10] is a lexicon based unsupervised technique which is utilized to separate the aggregate sentiments independent of individual classification. A transition graph is developed among the terms of a post set which is utilized as a part of SACI. SACI depends on an earlier classification of terms in a change chart with respect to their parts in uniting opinions. In the change diagram, the implied subsets of posts and aggregate opinions are depicted by navigating all ways of progress chart. In any case, the SACI technique abuses just terms and the relationship between these terms for sentiment analysis. This opened the portal for a novel technique to enhance target subordinate Twitter sentiment classification [11]. This strategy utilized both setting mindful methodologies and target subordinate methodologies for sentiment classification. The objective ward approach is utilized to join syntactic elements which are made utilizing words which are linguistically associated with the given focus in the tweet. This procedure

chooses if the sentiments relate to the objective the outcomes exactness is extremely poor. So authors of [12] proposed a novel half breed meta-heuristic technique in light of the mix of K-means and cuckoo hunt to decide the ideal group head from the sentiment substance of Twitter dataset. It adjusted the irregular introduction of cuckoo seek by the arrangement which was acquired from K-implies calculation, which in turn enhanced the execution of sentiment analysis. The significant disadvantage in this strategy is that change in precision was still not acquired to the degree desired.

A lexicon based approach known as SentiCircles [13] was developed to improve accuracy of sentiment analysis. The SentiCircles considered the co-event patterns of words in tweets that catch the tweet semantics and at the same time refreshed their pre-assigned polarity and strength in sentiment dictionaries. This approach successfully distinguished sentiments at the both tweet level and sentiment level. It offers static and settled earlier sentiment polarities of words with thinking about the settings of tweet words. In this approach, MPQA or SentiWordNet is utilized for sentiment labelling. Another lexicon based approach was presented in [14] with the solution for polarity shift problem. However, it failed to detect the neutral words in the tweets and also it performs only sentence level analysis, which is the major drawback of this method. PoliTwi system [15] was designed to tackle this drawback for detection of emerging political topics in Twitter. The main aim of this framework was quick detection of rising topics on political thoughts at a beginning time of the conversation. This framework was reached out by a sentiment analysis part that distinguishes the polarity of topics, set apart by hashtags. These hashtags are known as sentiment hashtags. Encourages, a social chart was developed for rising political topics that contain enhanced information, for example, polarity. As learned from various researches hybrid model outperforms individual models.

A hybrid approach [16] for Twitter sentiment analysis that utilizing n-gram demonstrate for highlight extraction and Support Vector Machine (SVM) algorithm or a dynamic Artificial Neural Network has been presented. In this approach, a Twitter specific lexicon was created by presenting an approach for managed highlight lessening utilizing measurable analysis and n-grams. This approach augmented this reduced Twitter-specific lexicon with brand-specific terms for brand-related tweets. The fundamental point of this approach was to order each tweet into a sentiment class. LDA collocation show [17] is another strategy created to

take care of issues in programmed sentiment analysis of Chinese weibo writings. The LDA show implies the content with word order flow. Furthermore, an unsupervised subject sentiment blend calculation was proposed to characterize the Chinese weibo messages consequently. The reliance relationship was isolated into subject and protest utilizing reliance parsing and it was utilized to outline six sorts of reliance models by assessing feelings words and questions. This model is utilized to separate the passionate components from Chinese Weibo messages naturally. In any case, it has low F-measure value for automatic sentiment analysis.

A system [18] for aspect-based sentiment analysis was proposed utilizing a sentiment sentence compression model that compresses redundancy in sentiment sentences and retains the polarity-related information to maintain the polarity of the original sentences. Through this strategy, over natural and unconstrained sentences of sentiments could be packed into more formal forma, consequently rendering it less demanding to parse the sentence after the pressure procedure, it viably investigates the sentiments in view of rich syntactic elements. Be that as it may, the issue of naturalness could in any case influence the execution of the syntactic parser. This can be corrected via automatic generation of sentiment dictionary known as SentiFul [19]. The SentiFul is extended through direct antonym and equivalent word relations, inferences, hyponym connection and intensified with natural lexical units. In view of this information, procedures for finding new sentiment passing on words were depicted. This procedure can likewise be utilized for analysis of sentiments other than English language. The standards of SentiFul were not approved. Consequently, the specialists returned to Multi class sentiment classification [20], in view of the Support Vector Machine (SVM) algorithm and one-Vs-one (OVO) technique. The relative ability weight was dictated by a proposed enhanced OVO system which is prepared in light of K-closest neighbour. The preparation writings are changed over into term include vectors, following which the imperative elements for multi class classification were chosen utilizing Information Gain (IG) algorithm. The preparation includes vectors of each match of sentiment classes are prepared by paired SVM classifiers. A confidence score matrix is developed based on the consequences of various SVM classifiers and it is utilized to distinguish the sentiment class of test content. In [21], authors utilized SVM with genetic algorithm to improve twitter sentiment analysis. However, the SVM has
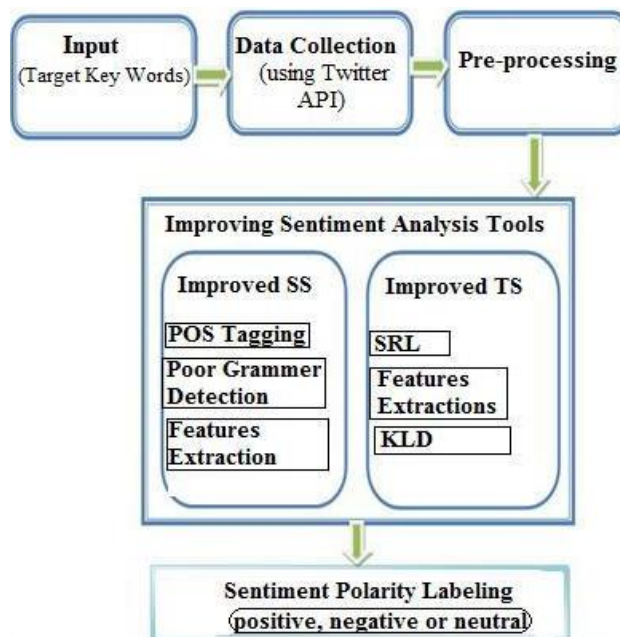


Figure.1 General model of the proposed work methodology

the major disadvantage of speed and size both in testing and training data, which degrades the performance.

From the literature two important points can be noticed, i) Twitter sentiment analysis has been widely studied ii) still exists many drawbacks exists with need for improvement. The major in existing methods is the faulty detection of non-neutral tweets as neural tweets and low accurate analysis. This makes the existing approaches invalid and demands a new approach for improved sentiment analysis.

## 3. Proposed work methodology

In this section, the proposed Improved TwitterSentiment and Improved SentiStrength are described in detail. TwitterSentiment and SentiStrength [22] are two popular and state-of-the-art sentiment analysis tools.

Our proposed work resolved the issues addressed in these two tools with the overall objective of improving sentiment analysis and tracking. In Latent Dirichlet Allocation (LDA) based model interpretation of sentiment variation [6], TwitterSentiment and SentiStrength tool were combined and used to assign sentiment label for each tweet of the interested target. The major limitations of SentiStrength have been resolved by using machine learning technique with NLP and the issues of TwitterSentiment is resolved by using and Semantic Role Labeler and Kullback-Leibler divergence. Thus, in the proposed work, the tweets sentiments are labelled based on Improved

TwitterSentiment and Improved SentiStrength tool. Fig. 1 shows the general model of the proposed work methodology.

## 3.1 Pre-processing

Initially, the whole dataset is pre-processed by one of the most popular pre-processing technique known as n-gram. n-grams is defined as series of items from some sequence. In the semantic point of view, n grams could be described as a sequence of characters or words. Usually, an n gram as a sequence of words is the most common. If a sequence of words contains two words then it is known as bigram. When a sequence contains more than three words then it is known as n gram. The unigrams and bigrams are feature extractors with Part of Speech (POS) tags which are used to develop features for sentiment analysis. While the unigram model helps in selecting individual words from the data and using bigram model helps in extracting pair of words from the data.   The unigram model is defined as the proportional context word calculated as the summarized representation as follows:

$$P(wd_i|wd_0 \dots wd_{i-1}) \approx P(wd_i) \qquad (1)$$

This equation helps in determining the unigram property of the tweet by analysing each word. The bigram model adds one more word of context and it is defined as follows:

$$P(wd_i|wd_0 \dots wd_{i-1}) \approx P(wd_i|wd_{i-1}) \qquad (2)$$

P is the part of speech of unigram/bigram model representation, $wd$ is the context words with i=0, 1, 3,…,N. In Eqs. (1) and (2),wd represents the words in collected tweets. This representation helps in identifying the POS features of the tweets for effective classification.

## 3.2 Sentiment analysis by improved TwitterSentiment

The TwitterSentiment sometimes misjudges a non-neutral statement as neutral, which degrades the performance of sentiment analysis. The Semantic Role Labeling (SRL) is used to improve the performance of TwitterSentiment. The task of a SRL usually involves taking a pre-processed data as input and identifying arguments with their semantic roles for each predicate. This task is meaningful since it is utilized in fine-grained information extraction, sentiment analysis and summarization of tweets.

Initially the pre-processed data is given as input to the SRL, which determines semantic arguments and roles with regard to a predicate in a sentence. SRL defines a sentence with different syntactic form, but with the same semantic roles. For example, "Sara met John" can be written as "Sara and John met", "Sara met with John", "Sara and John had a meeting". These sentences have a different syntactic form but their semantic roles continue to be the same. From the above explanations, the proportions sentences can be defined as the following:

*Proposition: meet (Sara, John)*

The main objective of the SRL is to capture the semantic roles in the sentences. A tree is constructed based on the arguments defined by SRL tool. A Frameset ID based on the Frameset tool is then added, which defines the argument structure of the lexical terms in terms of case roles along with the role names known as frame elements which are local to particular conceptual frames. From the SRL, the part-of-speech can be defined which describes whether a word is a noun, pronoun, adjective verb, adverb, adjective, preposition, conjunction or interjection.

Based on the results of SRL the weight vector of the Maximum entropy is assigned. In this proposed work, the weight vectors for verb, adverb, adjective, preposition, conjunction and interjection are assigned as 0.15, 0.37, 0.46, 0.01, 0.02, 0.02 respectively. The maximum entropy is a feature based model which has a two class scenario that is the same as using logistic regression to find distribution over the classes. This model makes no independence assumptions for its features and this mean that features such as bigrams and unigrams are added to the maximum entropy model, regardless of the feature overlapping. Maximum entropy can be defined as follows:

$$P_{ME}(c|t,\lambda) = \frac{exp[\sum_i \lambda_i f_i(c, tt)]}{\sum_{c'} exp[\sum_i \lambda_i f_i(c, tt)]} \qquad (3)$$

In Eq. (3), $c$ represents the class of the tweet, tt a tweet, the class vector which is assigned based on the results $\lambda$ of SRL. Eq. (3) enumerates the process of classifying the tweets based on the entropy of each tweet. Based on this, they are labelled.In order to label the tweets, Kullback-Liebler Divergence (KLD) has been used based on the calculation of probability distribution of maximum entropy model. The modified KLD can be expressed as follows:

$$KLD\big(c_i, tt_j\big)$$
$$= \sum_{k \in V} \left\{ \begin{array}{c} \big(P_{ME}(tm_k, c_i, \lambda) - P_{ME}\big(tm_k, tt_j, \lambda\big)\big) \\ \times log\left(\dfrac{P_{ME}(tm_k, c_i, \lambda)}{P\big(tm_k, tt_j, \lambda\big)}\right) \end{array} \right\} \quad (4)$$

In the Eq. (4), $V$ denotes the set of terms, $c_i$ denotes a class from a set of classes, $tm_k$ denotes a term, $\lambda$ is the weight factor and $tt_j$ denotes a tweet. The class computation helps in determining the labels of the tweets in order to improve the accuracy of classification computation. This computation involves four cases as follows:

1. $tm_k \in tt_j \wedge tm_k \in c_i$, i.e., the term $tm_k$ appears in the tweet $tt_j$ and in the category $c_i$;
2. $tm_k \in tt_j \wedge tm_k \notin c_i$, i.e., the term $tm_k$ appears in the tweet $tt_j$ but not in the category $c_i$;
3. $tm_k \notin tt_j \wedge tm_k \in c_i$, i.e., the term $tm_k$ appears in the category $c_i$ but not in the tweet $tt_j$;
4. $tm_k \notin tt_j \wedge tm_k \notin c_i$ i.e., the term $tm_k$ does not appear in the tweet $tt_j$ and not in the category $c_i$.

By using the modified KLD method, depending on its sentiment each tweet is labelled as positive, negative or neutral.

| Algorithm 1:  Improved TwitterSentiment |
|---|
| **Input**: Target Tweets |
| **Output**: Labelled sentiments |
| 1: Apply $uni_{model}$ |
| 2: Apply $bi_{model}$ |
| 3: $Preposition = FrameNet\ (Sentences)$ |
| 4: Capture the $Sem_{roles}$//$Sem_{roles}$ represents the semantic roles in sentences |
| 5:  Define arg |
| 6:  $structure = tree(arg)$ |
| 7:  $weight\ vector = type\ (argument)$ |
| 8: Find the distribution $P_{ME}(c\|t, \lambda)$ over classes |
| 9: Find the label of a tweet by using $KLD\ \big(c_i, tt_j\big)$ |

## 3.3 Sentiment analysis by improved SentiStrength

In the Improved SentiStrength tool, NLP and machine learning techniques are used to detect poor grammars features in tweets text. This would improve the sentiment label assignment. Each word in the tweet is categorized as Part of speech (POS). Maximum entropy is used to predict the POS tag for any given word. Based on the best probability distribution of maximum entropy the tweet words are tagged as a noun, a verb or an adjective. It can be defined as follows:

$$H(p) = - \sum_{t,h} \tilde{p}(h)p(t|h)\, log\, p(t|h) \qquad (5)$$

In Eq. (5), $\tilde{p}(h)$ is the probability of context $h$ occurring in the tweet, $p(t|h)$ the probability of the tag t occurring given context $h$ as predicted by the model. Thus, the model $p$ which maximizes entropy is the preferred model for tagging the tweets words.

The constraints are all known cases and it is defined as follows:

$$f(h_i, t_i) == \begin{cases} 1, & if\, t_{i-1} = Y\ and\ t_i = X \\ 0, & otherwise \end{cases} \qquad (6)$$

In Eq. (6), $f(h_i, t_i)$ is referred to as feature of the maximum entropy model where $h_i$ refers to the context or tag or history, $X$ and $Y$ denote arbitrary tags such as a noun or a verb.

The maximum entropy is applied to parse tweet sentences which are used to define the features for the maximum entropy classifier. Initially, the training data is needed to build a MaxEnt parser, which consists of the actions taken to build parser tree. Each action is assigned a probability score and a series of actions creates a parse tree. The probability action required to build the parse tree defines the resulting score of the parse tree. Based on tag, chunk, build and check procedures of parsing are used to define a feature that considers the surrounding context of the tags as well. The features are extracted using the following equation:

$$f(f(h, a) = \begin{cases} 1, if\ condition\ satisifes \\ 0, \qquad otherwise \end{cases} \qquad (7)$$

Where the satisfying condition is $previousTag(h) = adjective$; $CurretTag =$ and $a = StartNounPhrase$.

In Eq. (7), $f$ defines the feature and is active when action $a$ is to start a new noun chunk when the previous tag in the context of his an adjective and its following tag is a noun. Input to the build and check procedures is a forest of subtrees, built around different head words. The following is a feature that may be learned during the build and check procedures:

$$f(h, a)$$
$$= \begin{cases} 1, if\, previouschunk(h) = nounchunk \\ \quad and\, previousHead = 'that' \\ \quad and\, a = StartNounPhrase \\ 0, otherwise \end{cases} \qquad (8)$$

In Eq. (8), feature $f(h, a)$ is active when action $a$ is to start a new noun phrase constituent, and when the context $h$ contains a preceding noun chunk and a previous subtree with head word "that".

Based on Eq. (4), the features apply constraints to the maximum entropy model. These features are provided as input to the machine learning algorithm known as maximum entropy classifiers. For classification purposes, training data is used to set constraints on conditional distribution. Each constraint expresses characteristics of the training data which should be present in the learned distribution. The maximum entropy classifier is defined as follows:

$$P_{ME}(c|t,\delta) = \frac{exp[\Sigma_i \delta_i f(h,a)]}{\Sigma_{c'} exp[\Sigma_i \delta_i f(h,a)]} \qquad (9)$$

In Eq. (9), $c$ refers to the class, $t$ to the term and $\delta$ to the weight vector. Suffix of the words are those added to provide grammatical appeal to the sentences which can be removed during processing. Maximum entropy classifier learns the information about the poor grammar text in tweets from Eq. (9). This information is used in SentiStrength tool which assigns the sentiment score based on the sentiment lexicon. It chooses the maximum positive score and maximum negative score among all the individual words in the tweets text. All the maximum positive score and maximum negative score are summate and denoted as FinalScore which in turn indicates whether a tweet is positive or negative or neutral.

Thus, the Improved SentiStrength and improved Twitter Sentiment tools analyze the collected tweets and assign sentiment label for each tweet. Sentiments labels of the tweets are signed by combining the two improved tools and final decision of the label class is decided using the similar strategies as proposed by [6] in the following steps:

1. If Improved TwitterSentiment and Improved SentiStrength tool makes the same judgment adopt the judgment
2. If the judgment of one tool is neutral and another tool is non-neutral then adopt the non-neutral judgment
3. If the judgment of two tools are conflict each other, then adopt the judgment of Improved SentiStrength if the FinalScore is greater than 1 otherwise adopt the judgment of Improved TwitterSentiment tool.

---

**Algorithm 2:** Improved SentiStrength

Input: Target Tweets
Output: Labelled sentiments
1: Apply $uni_{model}$
2: Apply $bi_{model}$
3: Tag each word using POS tag based on the best probability distribution of maximum entropy $H(p)$
4: $Structure = MaxEntParsing(sentences)$
5: A feature can be learned from the parsing $f(h,a)$
6: Get the poor grammar information $P_{ME}(c|t,\delta)$
7: Feed the poor grammar information to the SentiStrength
8: for *(i=0; i<n; i++)*//n is the number of words in the tweet
9: if $(pos_{score}(i) > pos_{score}(i + 1))$
10: $max_{pos} = pos_{score}(i)$
11: else
12*:* $max_{pos} = pos_{score}(i + 1)$
13: end if
14: end for
15: for *(i=0; i<n; i++)*//n is the number of words in the tweet
16: if $(neg_{score}(i) > neg_{score}(i + 1))$
17: $max_{neg} = neg_{score}(i)$
18: else
19: $max_{neg} = neg_{score}(i + 1)$
20: end if
21: end for
22: Finalscore = $max_{pos} + max_{neg}$

---

## 4. Experimental results

This section presents the experimental study to investigate the effectiveness of the proposed work. For experimental purposes Twitter dataset has been considered for sentiment analysis process. Dataset are collected automatically using Twitter API and it span around 4000 tweets collected from 11.09.2016 to 20.04.2017. In this work, two key targets to test the proposed methods were selected, which are "Narendra Modi" and "Apple". This dataset has been pre-processed using unigram and bigram models in addition to removal of URLs and the non-English tweets. The experiments have been conducted using four methods for performing sentiment analysis and have been compared in terms of accuracy, precision, recall and F-measure for in order to evaluate the performance of each method. Table 1, shows the accuracy comparison of the literature methods.

Table1. Accuracy and false rate analysis

| Reference | Method | Accuracy (%) |
|-----------|--------|--------------|
| [8] | SentiView | 79.8 |
| [9] | TASC | 76.4 |
| [10] | SACI | 81 |
| [12] | K-means & Cuckoo search | 81.64 |
| [13] | SentiCircles | 81 |
| [15] | PoliTwi system | 83.7 |
| [17] | LDA | 82.5 |
| [19] | SentiFul | 81 |
| [20] | Multiclass SVM | 83.2 |
| [21] | SVM-GA | 84 |
| [22] | SentiStrength | 86.77 |

From the above table, it can be found that though all approaches have optimal accuracy between 70% to 90%, it can be found that the highest accuracy is obtained in [22] with 86.77%. Due to the presence of several drawbacks, even SentiStrength can be improved by hybridizing or modification. The proposed model utilizes both options for improving sentiment analysis.

## 4.1 Accuracy

Accuracy is the measure of correctly labelled sentiments in all instances. A measure of how often a sentiment rating was correct. For tweets with tonality, accuracy tracks how many of those that were rated to have tonality were rated correctly. A positive or negative sentiment word can have the opposite connotation depending on context. Finding the connotation can improve accuracy. It can be calculated as follows:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (10)$$

Fig. 2 shows the comparison of accuracy of tweets labelling among combined Twitter Sentiment (TS) and SentiStrength (SS), TS with Semantic Role Labeling (SRL) and SS, Improved TS and SS and improved TS and improved SS. While the X axis represents methods Y axis represents the accuracy value. Citing the comparison in table 1, the proposed approach actually outperforms the existing methods in literature significantly. The reason for the improvement is the use of NLP and Maximum entropy technique. For 4000 tweets as said above data, the proposed model has 0.87 percent of accuracy. Fig. 2 reveals that the improved TS and improved SS provide a higher rate of accuracy, these outperforming other methods.
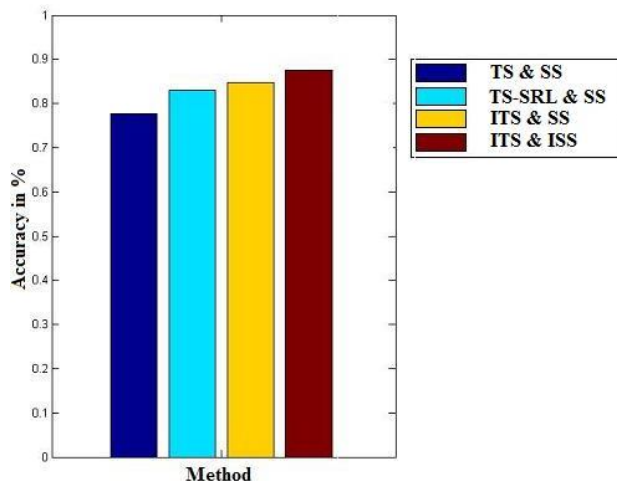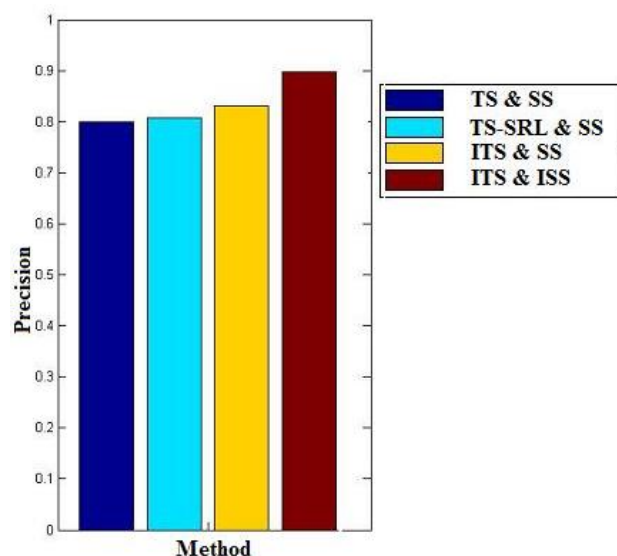


Figure.2 Accuracy comparison



Figure.3 Precision comparison

## 4.2 Precision

Precision value is evaluated according to the relevant information at true positive prediction, false positive. Precision measures the exactness of a sentiment analyzer. A higher precision means less false positives, while a lower precision means more false positives. Thus if the classification of twitter opinions is classified efficiently, it can be found by less false positives i.e. higher precision.

$$Precision = \frac{TP}{(TP + FP)} \quad (11)$$

Fig. 3 shows the comparison of precision amongst a combined of Twitter Sentiment (TS) and SentiStrength (SS), TS with Semantic Role Labeling (SRL) and SS, Improved TS and SS and improved TS and improved SS. While the X axis represents methods and Y axis represents the precision value in

percentage ranging from 0.79 to 0.89. The proposed method provides precision of 0.89 for 4000 tweet data which is outperformed the existing methods. The justification for this improvement is through the reduction of false positives due to accurate detection of the tweet arguments. From the Fig. 3, it can be inferred that the improved TS and improved SS has a higher precision value and also seem to outperform other methods.

## 4.3 Recall

The Recall value is evaluated according to true positive prediction, false negative. Recall measures the completeness, or sensitivity, of aanalyzer/classifier. A measure of how many tweets with sentiment were rated as sentimental. This could be seen as how accurately the system determines neutrality. Higher recall means less false negatives, while lower recall means more false negatives. Improving recall can often decrease precision because it gets increasingly harder to be precise as the sample space increases. In proposed model, the higher recall is needed to be obtained to justify the performance of the sentiment analysis model.

$$Recall = \frac{TP}{TP + FN} \qquad (12)$$

Fig. 4 compares recall between a combined of Twitter Sentiment (TS) and SentiStrength (SS), TS with Semantic Role Labeling (SRL) and SS, Improved TS and SS and improved TS and improved SS. The X axis represents methods and the Y axis represents the recall value in percentage ranging from 0.80 to 0.86. The improvement in recall is due to the reduction of false negatives which is achieved using the SRL and KLD methods for overcoming the shortcomings of Twitter Sentiment. From Fig. 4, it can be inferred that the improved TS and SS has a higher recall value of 0.86 that outperform other methods.

## 4.4 F-measure

F-measure is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute the score. It is a combination of precision and recall. The score is in a range of 0.0 - 1.0, where 1.0 would be perfect. The F-measure is very helpful, as it gives us a single metric that rates a system by both precision and recall. It is defined as follows:

$$F - measur = 2.\frac{Precision \cdot Recall}{Precision + Recall} \qquad (13)$$
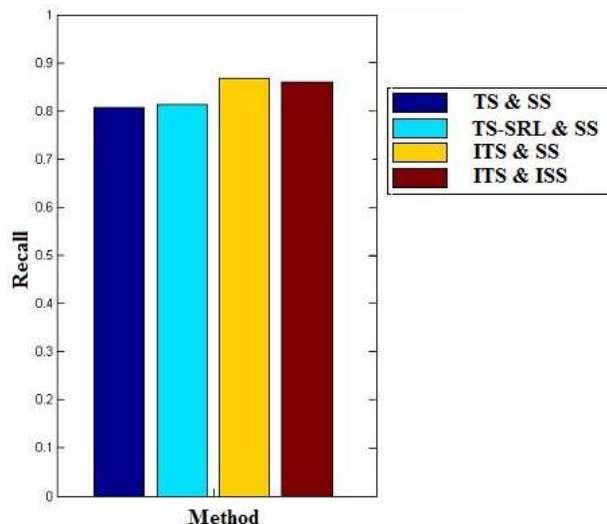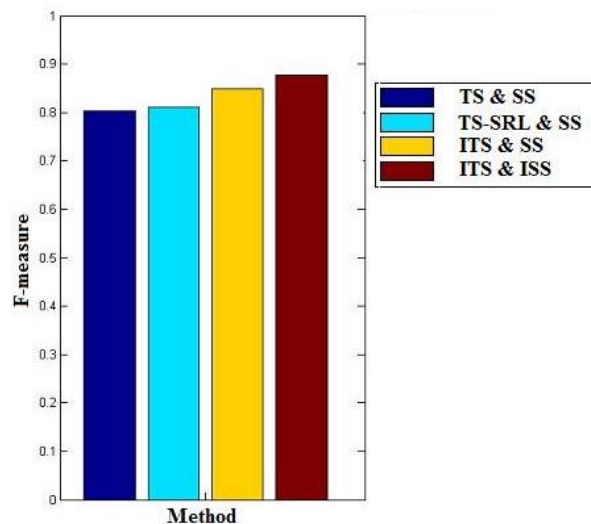


Figure.4 Recall comparison



Figure.5 F-measure comparison

Fig. 5 shows the comparison of F-measure among combined of Twitter Sentiment (TS) and SentiStrength (SS), TS with Semantic Role Labeling (SRL) and SS, Improved TS and SS and improved TS and improved SS. The X axis represents methods and the Y axis represents the F-measure value percentage ranging from 0.80 to 0.87. For 4000 tweets of data, the F-measure value is 0.87 which is almost similar to the accuracy. From Fig. 5, it is clear that the improved TS and improved SS have a higher F-measure value, thus outperforming other methods.

From the experimental results, it can be found that the proposed approach outperforms the other existing sentiment analysis methods. The novelty of this work lies in the following concepts; generally the political tweets are complex to utilize for sentiment analysis as most tweets have opposite tonality, sarcasm and indirect messages which reduce the accuracy. Here 4000 political tweets have

been collected and the use of improved TS and improved SS is a genuine effort. Instead of utilizing older and complex methods, this paper focused on developing user friendly approach that will be down-to-earth operation complexity while greater performance results. The novelty lies in the usage of tricky methods like KLD and Maximum entropy to the benefit of improving the sentiment analysis tools.

## 5 Conclusion

This paper has introduced the sentiment analysis of twitter data using a combination of Improved TwitterSentiment and SentiStrength tools. It is evident from the study that the performance of the sentiment analysis seems to improve by the use of combined of both Improved TwitterSentiment and the Improved SentiStrength. The datasets have been pre-processed using both unigram and the bigram models, which are feature extractors with Part of Speech (POS) tags. The performance of TwitterSentiment has been improved through the SRL and KLD methods while SentiStrength improved by NLP with machine learning technique. The experiment results have been conducted on Twitter dataset which prove that the improved sentiment analysis methods have outperformed existing methods in terms of accuracy, precision, recall, and F-measure. The performance of the proposed approach can be attributed to the inclusion of novel and simpler methods. This work extends the research area of social media sentiment analysis.

From this work, it can be inferred that the sentiment analysis tools can be improved by using novel ideas with the sole aim of accurate performance which will not be affected by the size of data. Though a major breakthrough is achieved through this work, still there is scope for improvement. This work demonstrates the analysis of Twitter dataset while it can also be utilized for Facebook and similar social datasets. The performance on other social media datasets namely Tumblr, Flicker, etc. has to be studied. Other than this, a noteworthy limitation of this approach is that the threshold to determine the polarity is pre-set. In future, it can be set automatically based on the statistical parameters.

## References

[1] B. Liu, "Sentiment analysis and opinion mining", *Synthesis lectures on human language technologies*, Vol. 5, No. 1, pp.161-167, 2012.

[2] S.Kolkur, G. Dantal, and G. Mahe, "Study of Different Levels for Sentiment Analysis", *International Journal of Current Engineering and Technology*, Vol. 5, No. 2, pp.768-770, 2015.

[3] R. Xia, F. Xu, J. Yu, Y. Qi, and E. Cambria, "Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis", *Information Processing & Management*, Vol. 52, No. 1, pp. 36-45, 2016.

[4] D. Tang, B. Qin, F. Wei, L. Dong, T. Liu, and M. Zhou, "A joint segmentation and classification framework for sentence level sentiment classification", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No. 11, pp. 1750-1761, 2015.

[5] X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen, "Aspect-level opinion mining of online customer reviews", *China Communications*, Vol.10, No. 3, pp. 25-41, 2013.

[6] S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. Bu, C. Chen, and X. He, "Interpreting the public sentiment variations on twitter", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 5, pp. 1158-70, 2014.

[7] M. Ponti, J. Kittler, M. Riva, T. Campos, and C. Zor, "A decision cognizant Kullback–Leibler divergence", *Pattern Recognition*, Vol. 61, pp.470-478, 2017.

[8] C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang, "SentiView: Sentiment analysis and visualization for internet popular topics", *IEEE Transactions on Human-Machine Systems*, Vol. 43, No. 6, pp. 620-630, 2013.

[9] S. Liu, X. Cheng, F. Li, and F. Li, "TASC: "topic-adaptive sentiment classification on dynamic tweets", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 6, pp. 1696-1709, 2015.

[10] L. Rocha, F. Mourão, T. Silveira, R. Chaves, G. Sa, F. Teixeira, R. Vieira, and R. Ferreira, "SACI: Sentiment analysis by collective inspection on social media content", *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 34, No. 27, pp. 27-39, 2015.

[11] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification", In: *Proc. of 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.151-160, 2011.

[12] A.C Pandey, D. S Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method", *Information Processing*

& *Management*, Vol. 53, No.4, pp. 764-779, 2017.

[13] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter", *Information Processing & Management*, Vol. 52, No. 1, pp. 5-19, 2016.

[14] N.V. Kolekar, G. Rao, S. Dey, M. Mane, V. Jadhav, and S. Patil, "Sentiment Analysis And Classification Using Lexicon-Based Approach And Addressing Polarity Shift Problem", *Journal of Theoretical and Applied Information Technology*, Vol. 90, No. 1, pp. 118-125, 2016.

[15] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, "Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis", *Knowledge-Based Systems*, Vol. 69, pp. 24-33, 2014.

[16] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network", *Expert Systems with applications*, Vol. 40, No. 16, pp. 6266-6282, 2013.

[17] Y. Li, X. Zhou, Y. Sun, and H. Zhang, "Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing", *China Communications*, Vol. 131, pp. 91-105, 2016.

[18] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence compression for aspect-based sentiment analysis", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No.12, pp. 2111-2124, 2015.

[19] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "SentiFul: A lexicon for sentiment analysis", *IEEE Transactions on Affective Computing*, Vol. 2, No. 1, pp. 22-36, 2011.

[20] Y. Liu, J. W. Bi, and Z. P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm", *Information Sciences*, Vol. 394, pp. 38-52, 2017.

[21] M. Wahyudi and D. A. Putri, "Algorithm Application Support Vector Machine with Genetic Algorithm Optimization Technique for Selection Features for The Analysis of Sentiment on Twitter", *Journal of Theoretical and Applied Information Technology*, Vol. 84, No. 3, pp. 321-331, 2016.

[22] E. Martínez-Cámara, M.T, Martín-Valdivia, L.A, Urena-López, and A.R. Montejo-Ráez, "Sentimentanalysis in Twitter", *Natural Language Engineering*, Vol. 20, No. 1, pp.1-28, 2014.