Less Sparse Feature Set with Meta Heuristic Weighted Classifier for Tweet Sentiment Classification

Ravinder Singh, Rajdeep Kaur

Abstract*: Twitter using Machine Leaning Techniques has been done. While consideration Bigram, Unigram,. SVM and naïve Bayes classifier which hybrid with PSO and ACO for effective feature weight. In Fig. 4.9 compare all experiment by on graph which shows that SVM_ACO and SVM_PSO better perform than SVM. NB_ACO and NB_PSO perform better than NB but if compare between hybrid approaches then SVM_PSO show 81.80% accuracy,85% precision and 80% recall. IN case of naïve Bayes NB_PSO 76.93% accuracy,76.24 precision and 82.55% recall, so experiments conclude that Naive Bayes improve recall and SVM improve precision and accuracy when use as hybrid approach.

Keywords: Sentiment Analysis, Twitter sentiment analysis, Support Vector Machine

I. INTRODUCTION

In the previous years, the individuals of young age are moving forward in the direction of the online social networking like Twitter, WhatsApp, Google Plus, Facebook, and so on. The social media is also revolving with those people to get them involved by making current trending insights concepts that is trending within a second. In the ongoing years, the individuals are uncovering their social relating issues by means of online networking using hashtags, remarks, surveys, emoji's, posts, and so on which was trailed by numerous individuals and their tweets become famous shortly. Likewise, the life based on social media is additionally carrying a huge open-door platform for organizations to associate with the purchasers so effectively [2,7] [9]. Individuals lay on generally client created content such as comments or remarks, over the web for settling on the choice. For instance, in the event that anybody needs to purchase an item or settle on a choice, they at first search its audits on the web, speak about it via web-based networking media. The content that is shown for that item is for the most part taken into the point just as the dialog in the web-based life is additionally seen and these made the best approach to make business a triumph. The investigation dependent on the surveys or remarks in the web-based life by the individuals represents the concept of sentiment analysis (SA). SA is acquainted with the world to reveal to us the data is right or wrong in every situation utilizing the online networking labels. Therefore, we can think about how individuals or world is responding to each angle as of now going on the planet [2, 13].

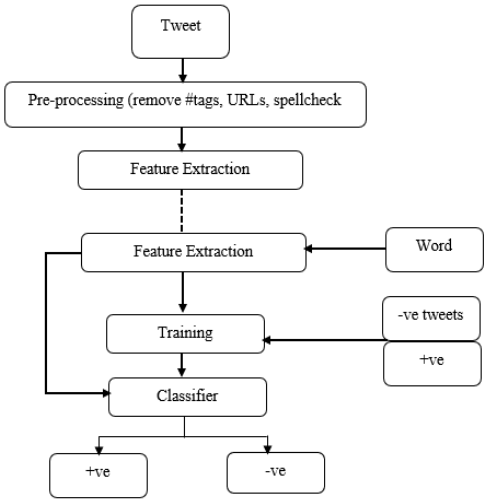

Figure 1: Architecture of the system

The framework design comprises of the parts as appeared in the figure 1, for example, extraction of Tweets from twitter, data pre-processing methodology, extraction of features. For the purpose of analysis, the training sets are usually characterized. The preparation or the training set is acquired by predefined set tweets (positive or negative) which should be possible utilizing SVM or naive Bayes and yield in positive, neutral, or negative tweets. The tweets are usually classified by means of a classifier that will characterize the tweets compatible with trained set and controls the tweet's polarity in the form of the result or the yield [2].

1.1 Sentiment Analysis

The concept of sentiment analysis can be characterized as a procedure that helps in mining of feelings, emotions, views, and opinions from content, tweets, database, and speech in an automatic way by mean of NLP i.e. Natural Language Processing. SA examination includes the classification of opinions in content into classifications like "positive" or "neutral" or "negative". It's likewise indicated as opinion-based mining, subjectivity examination, and the extraction based on judgement. The words sentiment, opinion, belief, and view are utilized in alternative manner however there occur contrasting remarks amongst them. SA examination has numerous applications in different spaces like the exposure of real-time events for example earthquakes, humanism, and the political area [2,4,5].

1.1.1 Classes of Sentiment Analysis

Sentiments can be arranged into three type of classes. for example, positive, neutral, and negative sentiments.

Revised Manuscript Received on January 05, 2020

Ravinder Singh, Computer science engineering, Chandigarh university, India.

Er. Rajdeep Kaur, Computer science engineering, Chandigarh university, India.

1. Positive Sentiments:
These represents the decent words about the objective in thought. If the sentiments based on positive feelings are expanded, it is alluded to be decent or good. While reviewing a product if the positive audits about the item are more, it is purchased by numerous clients.
2. Neutral Sentiments: These are neither great nor awful words about the objective. Henceforth it is neither favoured nor ignored [10].
3. Negative Sentiments: These are the awful words about the objective in thought. In the event that the negative assumptions are raised, it maintains a strategic distance from the preferred list. While reviewing a product if the negative audits are more, nobody cares to buy it.

1.1.2 Levels of sentiment Analysis
We can separate sentiment analysis based on certain levels discussed below:
1. Document: In this level, the task is to classify opinion for database. This database is based on an individual subject. In this manner, writings which involve relative learning can't be measured under the document level [1].
2. Aspect or Entity: It provides examination in detail. The centre undertaking of this level involves the identification of content-based aspect [Bing]. For instance, in a mobile survey if a client says," Sound is great yet the handset isn't helpful." In this audit handiness and sound represents the aspect. Here SA winds up with the undertaking based on two-level for example finding the perspectives in the content and after that characterizing in separate aspect. SA investigation on the basis of aspect-level is better than sentence and the document level examination [8, 11].

3. Sentence: At this level, the task drives to the sentences; it helps in determining that how the sentence is communicated (positive, neutral, or negative). On the off chance that a sentence expresses no conclusion implies it is an unbiased. This degree of examination is firmly identified with subjectivity arrangement [12]. The subjective kind of statement shows the element's polarity in agreed negative terms for example terms relying over good or bad analysis. Subsequently it is simple to gain sentiment from it. Yet, Objective proclamation doesn't give partition straightforwardly by positive negative terms. These are unique sentences which are based on the reality.

1.1.3 Features of sentiment analysis
Sentiment features are as follows:
1.Terms frequency and presence: These kinds of features are only individual words or n-gram words and their frequency checks. It may provide double weighting to the words or may utilize the term frequency-weights.
2. Opinion phrase and words: On their own these words usually express opinion/sentiments about the service or the product in the content. For example, hate or like, good or bad. Most of the phrases likewise express the opinion without utilizing opinion-based words.
Negations: Not good is considered as bad when there is presence of negative words which may the opinion
3. Parts of speech (POS): It set up discovering descriptive words from the content, as they are significant markers of sentiments.

1.2 Twitter
1.2.1 Definition: The term 'micro' in microblogging indicates the content limitation allowing the user to represent brief posts to other users of the service. A twitter client can make maximum 140 characters for each tweet. A tweet isn't just a straightforward instant message yet it is a mix of content and the metadata related with the tweet. These traits represent tweet features. They tell us exactly what the tweet is about and how the tweet content is expressed. Metadata can be used to discover tweet domain. These represent a few elements and spots. These substances incorporate client makes reference to, URLs, media Users, Twitter client ID, and hashtags. RT represents retweet,'@' trailed by a client identifier report the client, and'#' trailed by a word portrays a hashtag.

1.2.2 Specific features of Twitter
(a) URL: Most of the tweets share a connection alongside preface to the connections. The existence of gives its component value as 1, otherwise 0.
(b) Recency: When the question is terminated to get a tweet, it is smarter to get latest tweet about that issue. In this way, the feature of recency estimates the time of tweet in seconds after its generation process.
(c) ***Emoticons:*** These are outward appearances pictorially portrayed utilizing accentuation and letters; they express the client's state of mind
(d) Hastag: It is a word beginning with a symbol #. It alludes to a word referring to the textual content or demonstrating the subject of tweet. The value based on binary feature offers the response of whether the tweet comprises of hashtag or it does not contain any hashtag [8].
(e) Singleton: If a tweet carries no answer is a retweet (re-posting), at that point the tweet is known to be singleton.
(f) Retweet: A tweet represents only a client-based statement, or could be an answer to additional tweet. Retweets are set apart either with "RT" usually followed by "via @user id" or '@user id'. The property of Retweet is viewed as the element that makes Twitter a fresh medium of broadcasting information along with a direct communicational process [8].
(g) Mention: In a tweet when client need to allude to another client, he can compose his name beginning with a symbol '@'. It is known as Mention and it additionally presented as '@username'. If tweet encases mention the twofold component presenting to it will possess a value equal to 1, or otherwise 0.

1.3 Twitter sentiment analysis
SA is known to extract the opinions from the textual content or text assessment from the content. There are different angles, reasons, direction of extricating these feelings or emotions for examination purpose. The operation of tweet can support location as well as event-based detection. At the point when this task is realized

on twitter information, the architecture or framework to perform SA variates as indicated by what sort of result one needs to accomplish from the twitter-based analysis (or tweets). One progressively significant factor behind the changing idea of TSA is utilization of various approaches and systems. Ordinarily, specialists determine their very own structure or flow to do conclusion investigation to expand effectiveness of the outcome. Most of the regular procedural steps in the analysis of TSA and the watchwords in it are characterized underneath:

1.3.1 Pre-processing

In spite of these summed up framework of TSA examination, one can design this point into the accompanying work process. In this manner, the general steps associated with this system are as per the following: Before beginning of SA, pre-preparing of data must be done.

1. Removal of Non-English Tweets: The extraction of heat from huge datasets such as Clueweb or TREC dataset, it holds English just as the non-English kind of tweets. Subsequently, one need to run identification of language on every tweet, and need to erase from the accumulation of all tweets that are assigned with a zero-likelihood of being English.

2. Feature Selection: Lexicon Features: The features are chosen according to the following conditions:

- *Part-of-speech*: PoS features for example adjectives, adverbs, nouns and so on are labelled in each tweet.
- *Micro-blogging:* With the creation of binary features one can notice the existence of positive, neutral, and negative nearness feelings. Also, the existence of intensifiers and abbreviations, classification of tweets can be done based on the positive, neutral, and negative tweet. Slang dictionaries available online can be used for abbreviations and emotions [6].
- Lexicon Features: On the basis of word-based subjectivity, one can arrange the words into positive, neutral, and negative and nonpartisan dictionaries. We need to contrast each and every word along with predefined wordnet libraries.

2. Removal of Retweets: We need to erase any content that pursued a token of RT (just as the token of RT itself), since such text or message corresponds typically to the retweeted or cited material.

3. ASCII Conversion: A lot of tweets contain non-standard or uncommon characters, that are dangerous for the mechanism of processing down-stream. To report the problems, one need to utilize a mixture of Unidecode6 and BeautifulSoup5 to translate and change over all tweets to ASCII.

4. Removal of Empty Tweets: After finishing the majority of the other pre-preparing, one need to erase any unfilled tweet.

5. Restoration of Abbreviations: We can re-establish well known abbreviations utilized in tweets, to their relating unique structures

utilizing a vocabulary of shortened forms (for example "week-end" as "wknd"). Punctuations are generally kept since individuals frequently express sentiment by tokens, for example, ":-)"", :)", :(" These feelings can likewise be utilized for the classification of sentiments.

1.3.2 Phases for Features Extraction

1. Case Normalization: In this progression whole archive is changed over into lowercase.

2. Tokenization: it involves separating the frameworks of content into individual tokens or terms. This system can take numerous sorts, as per the phrasing being inspected. For English, viable tokenization strategy is to utilize white-space and the punctuation in the form of token delimiters [8].

3. Stemming: It represents a system of reducing significant tokens into a solitary kind of token. This strategy contains the acknowledgment and disposal of prefixes, suffixes, and inadmissible pluralization [3, 8].

4. Produce n-Grams: These are the nearby 'n' figures from a known sequence of feedback. For instance, a 3-gram of an expression 'FORM' would be presents as: 'F','_FO', 'FOR', 'ORM', 'RM', 'M'. N-grams having a 1-D are called as 'unigram', 2-D grams are called as 'bigram', 3-D grams are called as 'trigram' and the rest measurements are kwown as n-grams.

II. RELATED WORK

Sudarshan Sirat, et.al [1] utilized Naïve Bayes Algorithm to prepare the dataset based on a movie review, likewise utilizes the package named TextBlob in python to compute the twitter sentiment analysis. Alongside the assumptions of the tweets the analysts were additionally ready to extricate different qualities of the tweets for example Retweets, Preferences. The most likely used tweet and the number occasions it got retweeted. The accuracy-based classification can be improved by utilizing better models which can be prepared utilizing a bigger dataset. The procedure in this manner characterized was exploratory and it tends to be additionally improved by means of better algorithms and approaches. Shobana G, et.al [2] examined twitter sentiment analyses using tweets and fed the information to a model base on machine learning process to prepare it and after that check its precision, with the goal that analysts can utilize this model for later use as indicated by the outcomes. It involves steps like detection of sentiments, textual pre-processing, testing, and training the model. This examination subject has advanced during the most recent decade with models arriving at the proficiency of practically 85%-90%. But then, despite everything it does not have the component of decent variety in the information. Alongside this it has a ton of utilization issues with the usage of slang and the short types of words. Brinda Hegde, et.al [3] planned to concentrate and investigate tweets, group them as negative or positive tweets with the assistance of the algorithms and methods based on machine learning, lastly focus on the techniques of performance evaluation. In view of the demonetization dataset mining from Twitter utilizing Twitter-based API, the method of pre-processing was performed using Scikit-learn and NLTK that was further exposed to algorithmic

implementations, for example, Logistic Regression, Support Vector Machines, and Naive Bayes. A correlation of this implementation was measured to figure out which type of algorithm outperform for the known dataset in context to precision, accuracy, recall and F1-Score.

Mika V.Mäntylä, et.al [4] exhibited the best 20 referred papers from Scopus and Google Scholar and a scientific classification of research points. As of late, SA examination has moved from viewing the web-based audits to the textual form of social media writings from Facebook and Twitter. Numerous themes past item surveys like securities exchanges, elections, cyberbullying, disasters, and software-engineering expand the usage of sentiment-based investigation. Rahman, et.al [5] exhibited the SA of adolescents from the tweeter information. Social organization is the stage which is for the most part utilized by youngsters to associate with one another to share their emotions, feelings, and thoughts in pictures or text. This information is utilized for the investigation of items and aid to recover the nature of the items. The recursive neural network (RNN) and convolution neural system (CNN) were utilized for sentiment-based classification. The aftereffects of this methodology classified the sentiments as positive, neutral and negative assessments. Nicolas Tsapatsoulis, et.al [6] Nicolas Tsapatsoulis, et.al [6] contended that tokens utilized by people for the SA of tweets were most likely the best list of capabilities one can use for that reason. The analysts have contrasted a few naturally extricated features along with tokens or features utilized by people for classification of tweets below the framework of machine learning. The outcomes demonstrate that the physically shown in conjunction with DT i.e. Decision Tree classifier outperforms well on comparing it with other combination of feature-set classification algorithm. The physically explained dataset that was utilized in the tests was openly accessible for any individual who needs to employ it. Falguni Gupta, et.al [7] focussed over a work whose objective was to utilize a novel methodology which likewise considered sentiments based on region to provide SA another stage. Additionally, there has been a radical change in India because of economic demonetization. The primary point was to mine this information on the basis of region and check whether the classification based on region provides order can give us increasingly explicit outcomes or not. Kishori K. Pawar, et.al [8] gave a concise idea of tweets. When one needs to do tweets, sentiment-based analysis, he needs to do it in a specific part of SA-based examination. Thus, a concise information about TSA has been provided in the paper. Various procedures and strategies were talked about in a similar way. The precision/consequence of every strategy empowers us to envision the productivity of connected system in individual conditions. Adarsh M J, et.al [9] have seen the impact of Micro blogging webpage twitter on the present patterns and issues. Twitter based opinion mining causes to investigate different brands economically and furthermore to examine the practices of individuals utilizing interpersonal organizations. The examination of Twitter information was done using different viewpoints, the existence of words such as great, terrible and furthermore emojis in the tweets can be utilized to derive the supposition. The Twitter clients can be grouped into positive, neutral, and negative clients dependent on the adherents and the devotees and their practices can be overviewed dependent on the activity or tweeting and retweeting. Tweets can likewise be utilized to break down the impact factor in decisions and consequently be utilized to anticipate the outcomes as Twitter is the key device utilized by the candidates of US presidential to foresee the results. Virmani.et al [10], has examined the cooperation of SA with the extraction of opinions, summary and keep up the record of every single individual/student. In order to get collaborated and enhanced opinion about the student it adjusts the existing algorithm used in the paper. To investigate the opinion, a database relying on SA has been utilized. Initially, a score was set in general to every single assessment word in the database. Every time when the sentiment word was experienced in the sentence, it coordinates with the system database and sets the score in like manner. At that point from these scores the combined feeling were assessed. The algorithm used in the process provide a numerical incentive for the assessment. When the numerical score was high, it demonstrates the positive comment and when the numerical score was low it represented negative comments. The general execution relies on the instructor comments, the sentiment word utilized by the educator didn't coordinate with word utilized in database which influenced the general score. Medhat, et al. [11] has talked about the different applications of SA. Algorithmic ongoing upgrades were explored and presented in the paper. Numerous articles were involved in the proposed paper that assembles the user's interest in the field of SA (identification of emotions, transfer learning developing of the assets). The overviews were led on different algorithms of SA providing sophisticated classification. This work involves the use of different algorithms utilized to present the emotions and sentiments. Some of them were Chi-square, Latent Semantic Indexing (LSI), Pointwise Mutual information (PMI). The techniques of SA were fragmented into the process of machine leaning (ML), lexicon and hybrid-based methodology. V. S. Jagtap, et.al [12] talked about different strategies for the classification of SA. Web gives us a boundless source of the opinionated and the most differing content, and starting at yet just a small portion of the current spaces have been investigated. A lot of work has been executed on the item surveys short reports that include a well-characterized point. Progressively broad composition, for example, blog entries and pages, have lately getting more consideration over relative limitations. Bo. Peng, et.al [13] projected a new strategy of machine learning that applies the techniques of text-categorization to the document-based subjective sections. The separation of these segments can be actualized utilizing proficient systems for discovering least cuts in charts; this incredibly encourages cross-sentence incorporations.

III. THE PROPOSED METHOD

Proposed Algorithm

1.5.1 Classification using Naïve Bayes

Naïve Bayes classifier is the one among the family of probabilistic classifiers in machine learning, between features which is generally based on naïve independence assumption with applied Bayes' theorem. Number of features is required to be parameter linear in learning problem where Naïve Bayes classifier is highly scalable.

The probability distribution over the set of features:
$P(x) = P(c_i)P(X_1,X_2,X_3\ldots\ldots X_n/c_i)$
$P(X_1,X_2,X_3\ldots\ldots X_n/c_i) = \prod_{i=1}^{n} P(^{X_n}/_{c_i})$
$P(x) = \prod_{i=1}^{k} P(c_i)P(^{x_n^d}/_{c_i})$

Where
$X_{1,2,..n}$features values to certain class label c,
k←is the number of classes,
c_i←is the i^{th} class

$F(x_i) = sgn(\sum_{i=1}^{m} \alpha_i v_i A(u,u_i) + B)$

1.5.3 ACO-NB Algorithm

Algorithm 1: ACO-NB Module

Step 1: Initializing ants, where for each ant_n, n=1,2,3……..N.
Step 2: In ant_n, each variable$\mathbf{x_n^d}$, d=1,2,3…….D.
Step 3: Updating pheromones by choosing $\boldsymbol{\mu_i^d}$ from the pheromone table with probability in eq. (1), where i∈{1,2,3……K}.
Step 4: If minimum error is obtained, then it has higher probability.
Step 5: Generating a standard deviation$\boldsymbol{\sigma_i^d}$, if rv ≤ x_1 by eq (2) with the use of uniform distribution U (0,1), where rv is the random value lies between x_1, the predefined threshold 0 and 1.
Step 6: Generating a new value for variable$\mathbf{x_n^d}$: if rv ≤ x_2, by normal distribution N ($\boldsymbol{\mu_i^d}$,$\boldsymbol{\sigma_i^d}$).
Step 7: Else, uniform distribution generates random value, and generating random solution for$\mathbf{x_n^d}$.
Step 8: Obtained variable $\mathbf{x_n^d}$ denoting the observed attribute values to certain class label c.
Step 9: Computing probability for each class: $P(x_n^d) = \frac{P(y_i)P(y_j)}{\sum_{i=1}^{c} P(y_i)P(y_j)}$, j=1,2……..c
Where,
$\mathbf{P(y_i)}$is the $y_i$prior probability,
$\mathbf{P(y_j)}$is the conditional class probability density function.
Step 10: Calculate probability distribution over the set of features: $P(x) = \prod_{i=1}^{k} P(c_i)P(^{x_n^d}/_{c_i})$

Where
k is the number of classes,
c_i is the i^{th} class.
Step 11: Calculate accuracy, precision and recall.
Algorithm

1.5.3 ACO-SVM Algorithm

ACO-SVM Module

Step 1: Ants are initialized, where for every A_n, n=1,2,3……..N.
Step 2: In A_n, each variable$\boldsymbol{P_R^A}$, R=1,2,3…….n.
Step 3: The pheromones is updated by choosing $\boldsymbol{p_A(R,N)}$with probability
in above given eq., where i∈{1,2,3……m}.
Step 4: If the obtained error is minimum, then probability is higher.
Step 5:A standard deviation$\boldsymbol{\eta(R,U)^{\beta}}$is generated, if $\boldsymbol{N \epsilon P_R^A}$by eq (2)
with the use of uniform distribution U (0,1). where N is the random
value that has the predefined threshold 0 and 1.
Step 6: A new value for variable$\boldsymbol{\Delta\tau_A(R,N)}$ is generated.
Step 7: Else, random value is generated by uniform distributed and generating random solution is obtained for$\boldsymbol{\Delta\tau_A(R,N)}$.
Step 8: For each class, the probability is calculated :

$$p_A(R,N) = \begin{cases} \frac{\tau(R,N)^{\alpha}.\eta(R,N)^{\beta}}{\sum_{U \epsilon P_R^A} \tau(R,U)^{\alpha}.\eta(R,U)^{\beta}}, & if\ N \epsilon P_R^A \\ 0 & otherwise\ if\ Q > Q_0 \end{cases}$$

Where
In formula where,
$\boldsymbol{p_A(R,N)}$← transition probability,
$\boldsymbol{\tau_i(R,U)^{\alpha}}$←pheromone intensity among city R and city U in i^{th} group,
$\boldsymbol{\eta(R,U)^{\beta}}$←path length from city R to U,
$\boldsymbol{P_R^A}$←unvisited cities set of A^{th} ant in i^{th} group,
$\boldsymbol{\alpha\ and\ \beta}$← control parameters
Q←uniform probability [0,1]
Step 9: The training sample is
$\boldsymbol{n = \{(u_i,v_i)|i = 1,2,\ldots,m\}}$
Step 10: SVM classification model is described with optimization model $\boldsymbol{min_{\omega,\xi,B} P(\omega,\xi)}$

$$\boldsymbol{min_{\omega,\xi,B} P(\omega,\xi_i) = \frac{1}{2}\omega^t\omega + \frac{1}{2}\gamma\sum_{i=1}^{m}\xi_i^2}$$

$$\boldsymbol{v_i[\omega^t\phi(u_i) + B = 1 - \xi_i, i = 1,2,\ldots,m}$$

$$\boldsymbol{\xi = (\xi_1,\xi_2,\ldots\ldots\xi_m)}$$

Where
$\boldsymbol{\xi_i}$←Slack variable
B←Offset
$\boldsymbol{\omega}$←Support vector
$\boldsymbol{\gamma}$←Classification parameter for balancing the model complexity and fitness error.
Step 11: SVM classification model is described with optimization model $\boldsymbol{min_{\omega,\xi,B} P(\omega,\xi)}$

$$\boldsymbol{min_{\omega,\xi,B} P(\omega,\xi_i) = \frac{1}{2}\omega^t\omega + \frac{1}{2}\gamma\sum_{i=1}^{m}\xi_i^2}$$

$$\boldsymbol{v_i[\omega^t\phi(u_i) + B = 1 - \xi_i, i = 1,2,\ldots,m}$$

Step 12:Then, describing the classification decision function:

$$\boldsymbol{F(x_i) = sgn(\sum_{i=1}^{m}\alpha_i v_i A(u,u_i) + B)}$$

Step 13: Calculate accuracy, precision and recall.

1.5.4 SVM _PSO Algorithm

SVM _PSO ModuleStep 1: SVM classification model is described with optimization model $min_{\omega,\xi,B}P(\omega,\xi)$

$$min_{\omega,\xi,B}P(\omega,\xi_i)=\frac{1}{2}\omega^t\omega+\frac{1}{2}\gamma\sum_{i=1}^{m}\xi_i^2$$

$$v_i[\omega^t\phi(u_i)+B=1-\xi_i, i=1,2,\ldots..,m$$

$$\xi=(\xi_1,\xi_2,\ldots\ldots.\xi_m)$$

Where

ξ_i←Slack variable

B←Offset

ω←Support vector

γ ←Classification parameter for balancing the model complexity and fitness error.

Step2: SVM classification model is described with optimization model $min_{\omega,\xi,B}P(\omega,\xi)$

$$min_{\omega,\xi,B}P(\omega,\xi_i)=\frac{1}{2}\omega^t\omega+\frac{1}{2}\gamma\sum_{i=1}^{m}\xi_i^2$$

$$v_i[\omega^t\phi(u_i)+B=1-\xi_i, i=1,2,\ldots..,m$$

Step 3:Then, describing the classification decision function:

$$F(x_i)=sgn(\sum_{i=1}^{m}\alpha_i v_i A(u,u_i)+B)$$

Step 4: Calculate accuracy, precision and recall.

Step 5: In PSO model for each particle **i** in **S do**

Step6 : for each dimension **d** in **D do**

Step7: //initialize each particle's position and velocity

Step8: $x_{i,d}=Rnd(x_{max},x_{min})$

Step9: $v_{i,d}=Rnd(-v_{max}/3, v_{max}/3)$

Step10: end for

Step11: //initialize particle's best position and velocity

$v_i(k+1)=v_i(k)+\gamma 1_i(p_i-x_i(k))+\gamma_{2i}(G-x_i(k))$

New velocity

$x_i(k+1)=x_i(k)+v_i(k+1)$

Where

i- particle index

k- discrete time index

v_i –velocity of i^{th} particle

x_i – position of i^{th} particle

p_i- best position found by i^{th} particle(personal best)

G- best position found by swarm(global best, best of personal bests)

$G_{(1,2)i}$- random number on the interval[0,1]applied to the i^{th} particle

Step12: $pb_i=x_i$

Step13: // update global best position

Step14: if $f(pb_i)<f(gb)$

Step 15: $gb=pb_i$

Step16: end if

Step17: end for

Algorithm 4: NB_PSO Module

Step 1: Computing probability for each class: $P(x_n^d)=\frac{P(y_i)P(y_j)}{\sum_{i=1}^{c}P(y_i)P(y_j)}$, j=1,2……..c

Where,

$P(y_i)$is the $y_i$prior probability,

$P(y_i)$is the conditional class probability density function.

Step 10: Calculate probability distribution over the set of features: $P(x)=\prod_{i=1}^{k}P(c_i)P(x_n^d/c_i)$

Where

k is the number of classes,

c_i is the i^{th} class.

Step 2: Calculate accuracy, precision and recall.

Step 3: In PSO model for each particle **i** in **S do**

Step4: for each dimension **d** in **D do**

Step5: //initialize each particle's position and velocity

Step6: $x_{i,d}=Rnd(x_{max},x_{min})$

Step7: $v_{i,d}=Rnd(-v_{max}/3, v_{max}/3)$

Step8: end for

Step9: //initialize particle's best position and velocity

$v_i(k+1)=v_i(k)+\gamma 1_i(p_i-x_i(k))+\gamma_{2i}(G-x_i(k))$

New velocity

$x_i(k+1)=x_i(k)+v_i(k+1)$

Where

i- particle index

k- discrete time index

v_i –velocity of i^{th} particle

x_i – position of i^{th} particle

p_i- best position found by i^{th} particle(personal best)

G- best position found by swarm(global best, best of personal bests)

$G_{(1,2)i}$- random number on the interval[0,1]applied to the i^{th} particle

Step10: $pb_i=x_i$

Step11: // update global best position

Step12: if $f(pb_i)<f(gb)$

Step 13: $gb=pb_i$

Step14: end if

Step15: end for

3.2 Proposed methodology: Flowchart

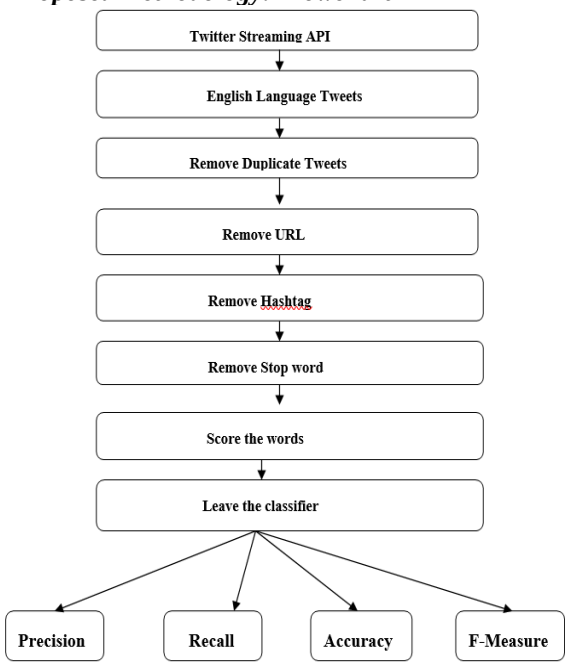

Figure 8: Proposed Flowchart

IV. RESULT ANALYSIS

4.1 Result Analysis

Table 4.1 comparison of parameters in different approaches

Classifier	Accuracy	Precision	Recall	Fscore
Naïve bayes	74.7	71.79	76.93	74.27
SVM	71.42	72.92	71.98	72.45
Naivebayes-PSO	85.48	80.24	82.55	81.38
SVM-PSO	85	89.72	90	89.45
Naivebayes-ACO	84.94	80.85	85	82.67
SVM-ACO	83.45	82.34	83.45	83

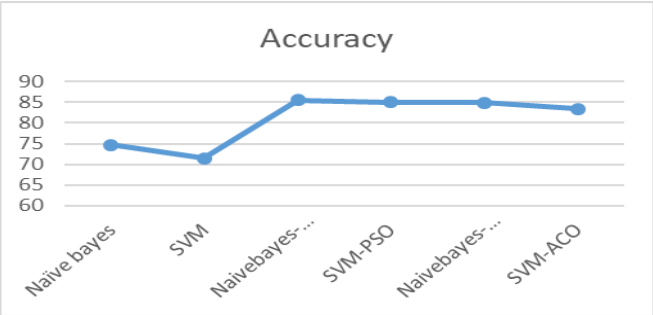

Fig 4.1 Accuracy comparison in different approaches

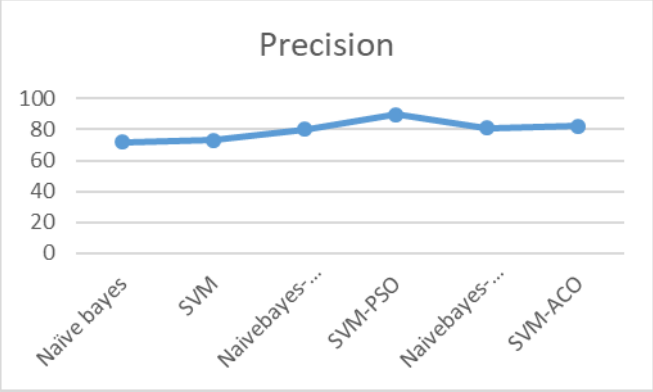

Fig 4.2 Precision comparison in different approaches

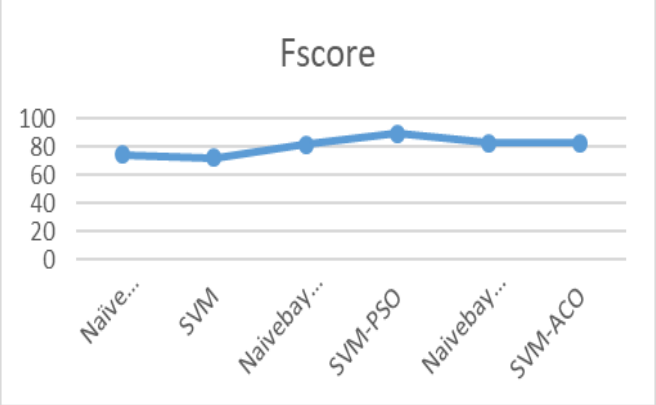

Fig 4.3 Fscore comparison in different approaches

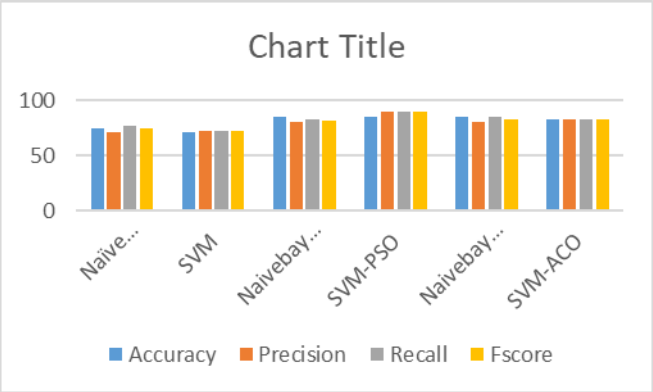

Fig 4.4 comparison in different approaches

V. CONCLUSION

Twitter using Machine Leaning Techniques has been done. While consideration Bigram, Unigram, Object-oriented features has been applied, comparative analysis of accuracy and precision between four algorithms showing the effect of features optimization. NB_ACO and NB_PSO perform better than NB but if compare between hybrid approaches then SVM_PSO show 81.80% accuracy,85% precision and 80% recall.

REFERENCES

1. Sirsat, Sudarshan, S. Rao, B. Wukkadada. "Sentiment Analysis on Twitter Data for product evaluation". IOSR Journal of Engineering (IOSRJEN), 2019, pp.22-25
2. Shobana, G., B. Vigneshwara, and A. Maniraj Sai. "Twitter Sentimental Analysis." International Journal of Recent Technology and Engineering (IJRTE)7, no.4 (2018).
3. Hegde, Brinda, Nagashree, H S. and Madhura Prakash. "Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets "International Research Journal of Engineering and Technology (IRJET), 5, no.6, (2018).
4. Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuutila. "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers." Computer Science Review 27 (2018): 16-32.
5. Rahman, Lizur, Golam Sarowar, and Sarwar Kamal. "Teenagers Sentiment Analysis from Social Network Data." In Social Networks Science: Design, Implementation, Security, and Challenges, pp. 3-23. Springer, Cham, 2018.
6. Tsapatsoulis, Nicolas, and Constantinos Djouvas. "Feature extraction for tweet classification: Do the humans perform better?." In 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), pp. 53-58. IEEE, 2017.
7. Gupta, Falguni, and Swati Singal. "Sentiment analysis of the demonitization of economy 2016 India, Regionwise." In 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, pp. 693-696. IEEE, 2017.
8. Pawar, Kishori K., Pukhraj P. Shrishrimal, and R. R. Deshmukh. "Twitter sentiment analysis: A review." International Journal of Scientific & Engineering Research 6, no. 4 (2015): 957-964.