Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

ENHANCED TWITTER SENTIMENT ANALYSIS USING HYBRID CLASSIFICATION METHODS AND RESULT ANALYSIS

Priyanka Tyagi[1*], Dinesh Javalkar[2] and Sudeshna Chakraborty[3]

[1] Research Scholar, Lingayas Vidyapeeth, Faridabad
[2] Asst. Professor Lingayas Vidyapeeth, Faridabad
[3] School of Engineering and Technology, Sharda University

ABSTRACT

The utilization of social media, in the course of recent years, has heightened hugely.Social media has framed a stage for the accessibility of plentiful information. A great many individuals express their discernments through social media. Sentiment Analysis (SA) of such perspectives and discernments is considerable to gauge public opinion on a particular/explicit topic of concern. Twitter is a microblogging webpage wherein clients can post updates (tweets) to companions (supporters).This paper proposes an instrument for extricating the suppositions from the tweets posted on Twitter. Tweets can be delegated positive, nonpartisan or negative. This model ends up being profoundly compelling and exact in the investigation of sentiments. This paper presents a crossover approach of utilizing Subterranean insect Settlement Enhancement and Molecule Multitude Improvement with classifiers. For each tweet, pre-handling will be finished by performing different cycles, for example tokenization; expulsion of stop-words, stemming etc. Besides, highlights are removed by the Uni-gram include extraction procedure. Grouping is completed by a Help Vector machine (SVM) classifier with Subterranean insect State Enhancement and Molecule Multitude Improvement to advance the arrangement execution. For solving ACSO, this work presents a hybrid classification based on ant colony optimization (ACO) and particle swarm optimization (PSO). In addition, the ACO framework includes a pheromone disruption method for dealing with pheromone stagnation. The execution showed that the new half breed slant arrangement had the option to further develop the precision execution. The productivity of the proposed framework was approved on the Creep Tweet dataset.

Keywords*:* Sentiment Mining, Naïve Bayes (NB), Support Vector Machine (SVM). k-nearest neighbor (KNN), hybrid classification method,Ant Colony Optimization (ACO),PSO , Sentiment analysis,

INTRODUCTION

Sentiment analysis is a subgroup of computational methods that naturally extract and epitomise those findings from opinions that contain a massive amount of data that the average reader can not process. [9]. Through online media sites like Facebook, Google+, Twitter, Instagram, and others, individuals everywhere on the

Article to Cite:
Priyanka Tyagi, Dinesh Javalkar & Sudeshna Chakraborty. (2021). Sentiment analysis of twitter data using hybrid classification methods and comparative analysis, Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition), 40(6). DOI 10.17605/OSF.IO/2NVJK

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

world are a piece of one another's lives in the high level Web age. With Feature Optimization, Twitter has almost 300 million active users and millions of tweets every day, making it a big social networking internet platform worldwide. Twitter has traditionally been used as "an analytical resource by numerous organisations to study public sentiment and gather critical input" due to its large user base and vast data.(# 11). Users may write their opinions on any subject or general thoughts in a tweet that is no longer than 140 characters long. Because of the short duration of tweets, people write in a rather succinct way, sometimes using slang, making sentiment analysis difficult.

Similarly, notion examination can be portrayed as the way toward arranging sentiments communicated through tweets to all the more likely comprehend the client's viewpoint on a given subject [15].

"It is beneficial for marketers to review and analyse consumer perceptions of their brand and existing/newly launched goods, as this will help them measure and enhance their results" [12] [16].

Dorigo introduced the subterranean insect settlement improvement (ACO) in 1991, some significant procedures, for example, positive criticism and covered up equal were proposed. By utilizing a positive criticism methodology, the ACO can track down the better outcome through equal pheromone trading between insects. What's more, by utilizing a secret equal technique, bouncing into the ideal arrangement can be forestalled and the ACO is likewise productive.

There has been a lot of interest in extraordinary calculation with the end goal of information order. There are enormous quantities of datasets publically accessible [24] for the model approval reason. Insect Province Enhancement (ACO), Hereditary Calculation (GA) and Molecule Multitude Improvement (PSO) are very much archived in writing for the reason for combinatorial streamlining, steering, booking and grouping and so on Grouping exactness is significantly subject to the integrity of the highlights. A portion of the highlights may corrupt the presentation too. Highlight choice is additionally an intricate interaction [25, 26, 27, 28, 29, 30]. Choices about

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

the integrity of highlights assume an essential part in the order. The issue of choosing the integrity of highlights has been demonstrated as a component determination issue.

Multitude Knowledge Improvement Techniques are utilized which depend on the activities of gatherings of bugs found in nature [32,33,34,35].. Ants' normal activity is inhibited by ACO tactics, which the HCTS use to get to their colony in the shortest possible time. When returning to the colony, they keep track of each direction by depositing pheromones, which eventually evaporate. Since it is taken more often, a shorter path would have a higher pheromone density. The same conduct has been joined into PC counterfeit subterranean insects to track down the best answers for a problem. Birds' social movement is hindered by PSO in light of the fact that they decide to live in flocks. All of the birds are endeavoring to find food in a given area [17] [18] [48,49,50].. They are, nonetheless, aware of the distance among them and the food. This conduct has been coordinated to help distinguish the best advanced arrangement by iteratively improving the competitor arrangement locally and globally. To recognize tweets as sure or negative, AI methods will be used. The utilization of a directed AI model will give somewhat better results. For this investigation, two learning strategies were thought of: NB, KNN, ACO, PSO, and SVM.

Furthermore, we will make useful recommendations for the application of various algorithms to various classes of social network data. The model was underlying Java and afterward tried against tweets, with its yield estimated utilizing four boundaries: exactness, accuracy, review, and F-score.

RELATED WORK

Assessment Investigation is the investigation of how one's perspectives and perspectives are connected to one's mind-set and demeanor as communicated in regular language because of an occasion .Recent occurrences demonstrate that sentiment analysis has advanced to a point where it can go beyond positive vs.

ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

negative and address the entire spectrum of behaviour and feelings for various groups and themes.

In the subject of slant investigation, a significant number of exploration have been finished using assorted approaches for the expectation of social perspectives.

Ankita Gupta and colleagues [1] To increase classification accuracy, SVM and KNN based hybrid models are provided in this study. While much of the work in this sector is concerned with 2-way categorization, the suggested technique isolated the tweets into good, negative, and unbiased sentiments.

The suggested model's work has gone through three stages: preprocessing, feature creation, and classifier learning. The model is evaluated analytically as far as precision and f-measure. The aftereffects of the correlations are contrasted with the SVM and KNN calculations . The model's precision and f-proportion of tweet class expectation have improved, as indicated by the examination discoveries.

Bharat Naiknawareet al.,[2] if the MAE is less than the accuracy. The findings reveal that the classifiers have the same performance. In the MAE, there is a smidgeon of a difference.

Seven datasets were used to test the classifiers' performance (Budget2017, Demonetization, GST2017, Computerized India, Kashmir, Make in India, Startup India).Nave Bayes performs best in the Budget2017 dataset . The Nave Bayes calculation performs best in the Demonetization dataset. In the GST2017, SVM has the best presentation, yet Max Entropy has the best exhibition in the Computerized India, Kashmir, Make in India, and Startup showcases. We can also see that the Mean Error is useful for estimating the Mean Absolute Error.

Taylor and Christianini.[3]The information of SVM, a machine learning method, was published and shared. The writers are able to provide a thorough grasp of the methodology as well as how to approach the SVM method in order to apply it to real-world challenges. Because research was being undertaken in all domains at the time the book was written, the approach will be theoretical.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

In 1975, John Holland described the GA based on particle natural selection for the first time [3].GA is based on the "survival of the fittest" premise proposed by Charles Darwin.

Compared to traditional methods, this method is more efficient. When inputs are altered and various sounds occur, GA does not readily break. As a result, GA offers a more advanced methodology [4].Coding, selection, crossover, and mutation are the four main GA operands.

The novel notion of PSO was proposed by Dr. James Kennedy and Dr. Eberhart (particle swarm optimization).This method aims to improve the particle solution over time in order to optimise a problem [1].The behaviour of an organism, such as schooling of fish or flocking of birds in a natural setting, inspired this algorithm. PSO is an intelligent swarm algorithm that solves optimization problems. PSO's key benefit is its ability to converge quickly [5].

Y Chen and Y. Hao proposed a method that is a hybrid artificial-ant colony method [6].In his Ph.D. thesis in 1991, Marco Dorigo introduced ant colony optimization (ACO). When compared to other optimization methods, this method's searching speed has increased thanks to the ant algorithm [7].The ants in ACO wander from their nest in search of the quickest path to the food source[36,37].

Malhar and Slam [4] utilized managed AI procedures and counterfeit neural organizations to recognize tweets, just as a contextual analysis of Official and Gathering decisions, and found that SVM outflanked any remaining classifiers.

Using the user influence factor, the authors build a mechanism for predicting election results. The authors used a combination of Principle Component Analysis and SVM to achieve the dimension reduction. Martineau and Finin [5] propose the Delta TFIDF approach, which effectively assesses word scores prior to categorization. Delta TFIDF was simple to comprehend, construct, and calculate.

The authors employed support vector machines to improve slant characterization exactness with Delta TFIDF and film audit information sets. Delta TFIDF is ideal than

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

TDFIF in that it checks word crude for all record sizes and loads for legislative identifying charge support, opinion extremity arrangement, and subjectivity location, as per the creators.

Delta TFIDF, according to the authors, is the first method for boosting and identifying the importance of selecting words utilizing the assessed solo circulation of highlights preceding grouping into two classes .Two SVM classifiers were constructed by Mohammad et al. [6,38].

One is a term-level occupation that assesses the opinion of a solitary word in a message, while the other is a message-level work that recognizes the notion of messages like SMS and tweets.

The creators partook against 44 different groups in a rivalry, and their work on tweets won in front of the rest of the competition, with a F-score of 88.93 in the term-level work and a F-score of 69.02 in the message-level undertaking.

Sentiment, semantic, and surface-form aspects were implemented by the authors.

The creators likewise made two enormous term-notion affiliations, one utilizing emojis from tweets and the other utilizing estimation term hashtags from tweets. Neri et al. [7] [39] [40].. analysed the sentiment for dynamic firm La7 and Rai, a growing Italian social broadcasting firm, based on sentiment analysis on newscast across more than 1000 Facebook posts.

The findings of the authors were compared to a study undertaken by the Osservatorio di Pavia, an Italian research centre specialising in the empirical and theoretical study of media that focuses on political communication in the media.

The authors employed the Knowledge Mining System, which is utilised by security agencies and government organisations in Italy to regulate data from Web mining and OSINT.

For recognising the polarity of English tweets, Pablo et al. delivered adaptations of Innocent Bayes classifiers.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

Benchmark (prepared to classify tweets as sure and tweets as bad, unbiased) and Parallel (utilizes an extremity vocabulary and arranges tweets as certain and tweets as negative) are two separate types of (NB) classifiers.

The properties that classifiers identified were obtained from (noun and verb and adjectives also adverbs).Valence Shifters and multiple sentences from rare resources

[9] [41,42]. Po-WeiLiang et al.Using real-time sentiment analysis on the Internet microblogging site Twitter, create a system for presidential candidates in the 2012 U.S. elections.

The typical study of elections takes time to obtain poll data, however this technique uses twitter, a microblogging site, to collect data from a larger number of individuals.

It enables societal figures like as academics, the media, and politicians to broadcast their future views on public opinion and the political process.

Finally, the authors stated that the system and methodology are general, and that they should be easily adopted and expanded over a variety of other fields.

METHODOLOGY

Figure 1 demonstrates the working of Sentiment Analysis in the extraction phase where social media acts as a data source. Data from social media like twitter is updated frequently. So, it gives the feeling of real time representation of sentiments. To obtain data on run-time an internet both known as web crawler is used. This browses through the World Wide Web in an organized manner to index web pages. In pre-processing, the extracted data is cleaned as it has large amounts of noise before sending the text for analysis. Extracted text has grammatical errors as text is of limited length. In analysis, sentiments in data, number of repetitions observed in tweets and their location is analyzed. In Knowledge Discovery phase, to find opinions of people regarding any particular occurrence, it is essential to store data

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

related to the event. Once sentiments polarity is known it generates statistical graphs and charts.

There are four groups used in sentiment analysis:

Syntactic feature: It employs word or Part Of Speech (POS) ticket, Ngrams, punctuation, or phrase patterns one among all. The authors identify that "phrase patterns like "n+aj" (positive adjective) symbolize positive sentiment direction, at the same time as "n+dj" (negative adjective) articulate negative sentiment".

Semantic feature: Semantic feature concentrate on relation between signifiers like phrases, words, and score-base method classify these stands on title figure of encompass positive or negative semantic feature.

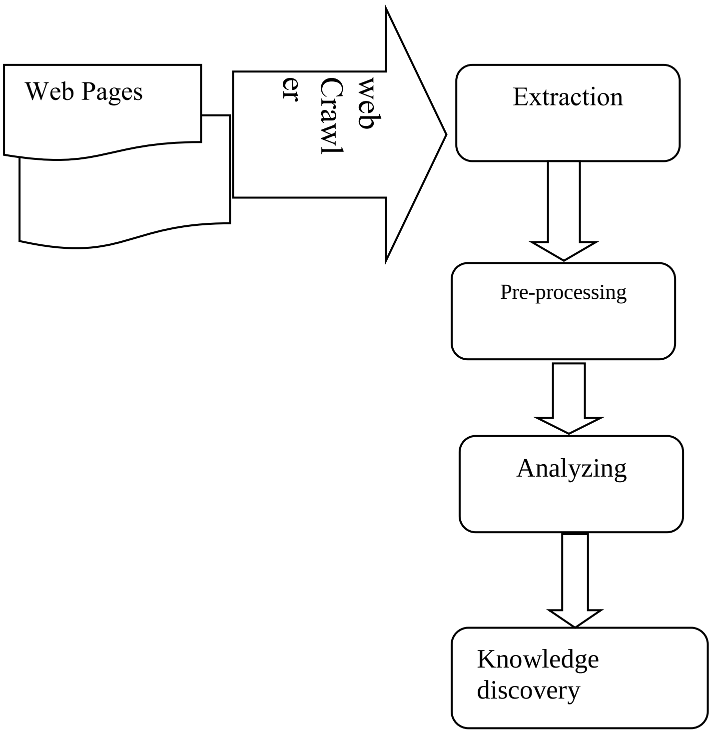


Figure 1Working of Sentiment Analysis

Link base feature: Using links present among relation and relations, link base samples are classified.

Stylistic feature: Artist's use this to go by a message to us.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

Use of symbolism: This is where a writer or an artist employs a symbol to illustrate, symbolize or differentiate a person, thing or place.

The arranged framework encases different periods of development. A dataset is designed utilizing twitter posts of film audits. As we realize that tweets hold language words and incorrect spelling. In this way, we do a sentence level slant examination on tweets. This is finished in three phases. In a first stage preprocessing is ready. At that point highlight vector is shaped through appropriate highlights. Ultimately, utilizing disparate classifiers, tweets are grouped into positive, negative and unbiased classes.

Algorithm

Input: Training and Testing Datasets

Step 1: Browse Training dataset

Step 2: Delete stopwords from training dataset.

Step 3: Tokenization of training dataset.

Step 4: Stemming of training dataset.

Step 5: Browse Testing dataset

Step 6: Delete stopwords from testing dataset.

Step 7: Tokenization of testing dataset.

Step 8: Stemming of testing dataset.

Step 9: Extract special keywords.

Step 10: Extract positive and negative keywords.

Step 11: Extract positive and negative tags.

Step 12: Form 8-features based vector

Step 13: Perform Classification

Step 14: Form BoW vector.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

Step 15: Form Hybrid Vector by fusion of vector 1 and 2.

Step 16: Measure Precision, Recall and AccuracyPerformance for existing and proposed method.So as will perform sentiment analysis, would needed with gather data from the desired source (here Twitter).

This data goes through various steps of pre-planning which makes it more machine sensible than its past structure.One of the best things that happen on machine learning is that the algorithms can memorize the data and when we need to use it for another data it has a poor performance, this behavior is called over fit. To avoid this problem, I work with test driven methodology. Each dataset is divided in three random parts and each part in three moredivisions:

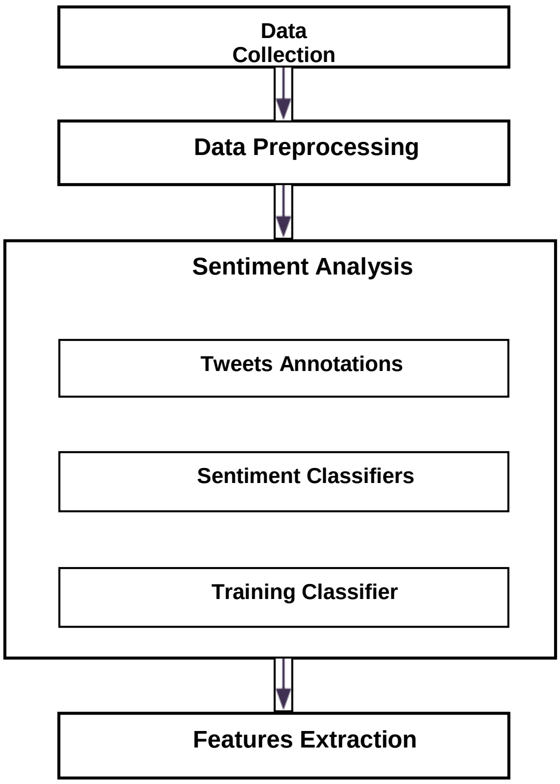


Fig. (1) : Methodology for sentiment analysis

ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

DATA COLLECTION

Is concerned with the correct acquisition of data; regardless of the various methods depending on the field, the accentuation for guaranteeing accuracy stays those same. The number one aim of any data collection attempt is to obtain quality data that may be easily translated to analyze rich data evaluation which can result in dependable and conclusive answers to questions that need been posed.

Thus, it is Tweet assortment incorporates gathering pertinent tweets about the particular space of revenue. The tweets are assembled using the Programming interface. These Programming interface assists us with social affair the information for the information. Fundamentally, it is an interface between the client and the source site from where the information tweets information could make brought. As it's far a delayed interaction, for this exploration reason the information has been gathered of different sites instead of gathering tweets from the actual Twitter.

PRE-PROCESSING

Those pre-processing of the information may be a significant step In like manner; it picks the viability of the various strides down in line. It includes grammatical amendment of the tweets as wanted.

Optimization

In this paper, ACO and PSO are used to for the enhancement cycle. This marvel is alluded to as 'combinatory improvement'. It may be multi-objective functions that need which means of searching on the preliminary values aiming to decrease the final results of our function. It is a cycle wherein the important highlights are chosen from the arrangement of the element. This interaction is finished by utilizing some calculation which is improved in advancement Subterranean insect province improvement and Molecule swarm enhancement both are based absolutely at the natural conduct of the insects and multitudes. By this procedure it tracked down the most concise way or courses of subterranean insects. Those yield for the investigation exhibit that the improve calculation not simply diminish the number for

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

ways in the ACO, yet additionally finding the briefest route at the spot of biggest way. These calculations utilize the very rule that are used by the subterranean insects and gives us optimized features

Supervised Classifiers

In this stage, order the highlights as expressed by their properties. Arrangement in this work is finished via the use of the Help Vector Machine and Credulous Bayes Classifier.

Naïve Bayes*:*

The Innocent Bayes classifier in a champion among the least complex probabilistic model meets assumptions emphatically on text order and used investigating Bayes rule with self-supporting element assortment [13,43,44,45]. meets expectations positively around quick text order and deals with Bayes rule. With Self-supporting component combination [13], it is adaptable in method of taking care of with whatever number of classes or traits. For a given tweet d, C* is a class variable which characterizes the conclusion given by

$C^* = \text{argMax}_c\ P_{NB}(C/D)$

Bayes Probability $P_{NB}(C/D)$ described as

$$PNB(CID) = \frac{(P(C)\sum_{i=1}^{m} P(f\backslash C)^{ni(d)}}{p(d)}$$

Here, f is highlight and ni(d) is include found in d, m addresses all out number of highlights and P(c) and P(f|c) are found through greatest probability gauges [14]. During order stage we discovered a word which was not found in preparing stage then we will give zero as likelihood for positive, negative and nonpartisan classes. To end this issue, we will in general make likelihood equivalent utilizing Laplacian smoothing constant*k*=1.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

$$\frac{Term\_count + k}{Tota\_Terms + \ kIcI}$$

Support vectormachines:

Backing vector machines (SVM) is a mix of a straight demonstrating Besides event based learning in a high-dimensional space. SVM might be done for the ones issues while information can't make isolated via line [21]. SVM utilize nonlinear planning – It changes the occurrence space into some another space which need higher size over thefirst.Kernel thought gave up push to help vector machines. Bit is a capacity which satisfy planning of a nonlinear information to another space.

Kernel function *K* will be an inward item Φ(*x)*• Φ (*y)* between of two points *x* and *y*:

k (x, y) = Φ(*x)*• Φ (*y*) (4.4)

where Φ (x)and Φ (y) are planning operators. The include, that piece trademark is formed as an inward item, offers a likelihood to refresh scalar item with a couple of inclinations of part [15][22] [46,47]. The issue of discovering boundaries of SVM consistent with a raised enhancement inconvenience, which implies that that close by result might be is worldwide ideal as well.In general, the categorization task usually dividing data under traineeship and experiment sets[23].Those objective of SVM may be to prepare a model (primarily based at the traineeship data).SVM for classification may be utilized to discover a linear model of the following form:

$$y(x)= W^{T}X +b(4.5)$$

wherein x is enter vector, w and b are boundaries which might be adjusted for a specific model and assessed in a test technique. In straightforward direct grouping type might be to chop down a regularized blunder work given through Equation 4.6.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

$$C\sum_{n=1}^{N}\mathcal{E}_n+\frac{1}{2}\parallel\omega\parallel^2$$

whereas $\mathcal{E}_n \geq o, \forall n = 1, \dots \dots \dots N, and$

$$y(\omega T_x x + b) \geq 1 - \mathcal{E}_n 4.6$$

Here the SVM constructs an isolating hyper plane and afterward attempts to expand the" edge" in between the two classes. With figure that edge, the SVM builds two equal hyper planes, one on each side of the underlying one.

These hyper planes are then "pushed" oppositely away from each other until they interact with the nearest models from one or the other class. These models are alluded to as SVM and are illustrated in bold in Figure 3.5.

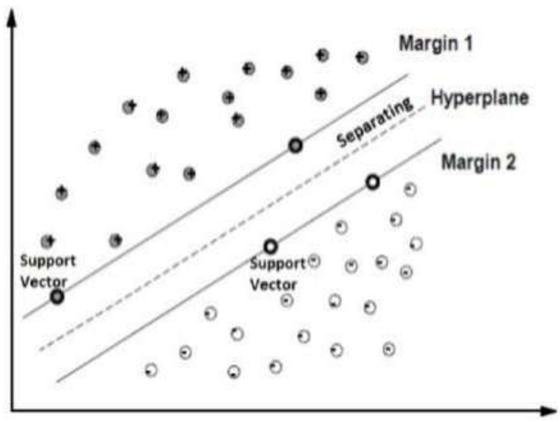


Fig.: Support Vector Machine: Classification

KernelFunctions:

Following are types of SVM piece capacities utilized for the arrangement. In research [20] the four after fundamental bits are portrayed:

Liner kernel:k(x,y)=X^TY +C

Polynomial kernel:k(x,y)= (X^TY +C)

Radial basis kernel: k(x, y) = $\exp(\frac{-IIx-yII^2}{2\sigma^2})$

Sigmoid Kernelk(X_i,Y) : tanh (Yx^Ty + C)

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

Hybrid PSO-ACO algorithm

As a rule, the normal PSO-ACO model summed up as Calculation 1.In Calculation 1, time overhead for a complicatedTSP which needs a particular number of ACO cycleswould be colossal. In this article, we attempt to reducethe calculation running time dependent on the improvementsas follows:

Algorithm 1.SVM- PSO-ACO.

Algorithm 1.SVM-PSO-ACO.Input:Hyper parameters of PSO;

Output:A sequence of Hyperparameters for ACO.

1: InitializeD-dimensional particlesPiand their velocityVi(i¼1,2,...,I,Iis the number of particles) related withDparameters of ACO;m¼0;M,Nare the maximum numbersof iterations in PSO and ACO respectively;

2:whilem<Mdo

3:ACO is invoked by PSO,i¼0;

4:whilei<Ido

5:n¼0;

6:whilen<Ndo

7:ACO deals as (1)-(5) withPi;

8:n¼nþ1;

9:Clear the pheromones;

10:The achieved result is used to evaluatePi;

11:Update the velocity of SVM,PSO,Vi, as (6);

12:Update the location ofSVM ,PSO,Pi, as (7);13:i¼iþ1;

14:m¼mþ1;

15:returnThe optimal solution of the relatedPi.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

RESULTS

Performance Evaluation:

Should figure that accuracy to classifier, we needed to gauge which exactness may an opportunity to be gained. There are two measures on which exactness might be reliant:**Accuracy, Precision, Recall, F Score.**

True Class

Prediction Class	**Positive**	**Positive**	
		True Positive (TP)	**False Positive (FP)**
	Negative	**False Negative (FN)**	**True Negative (TN)**

Table (1): Confusion Matrix

The result is produced from where of Precision, Recall, F score, and Accuracy.

Precision: It is the extent of archives ofrightly grouped under certain forecast class to all records under sure expectation class.

$$\text{Precision} = \frac{TP}{TP - FP}$$

Recall: It is the extent of archives of rightlyclassified under certain expectation class to the records that are positive in the negative forecast class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Accuracy: Exactness and accuracy are two vitalvariables significant elements to consider when bringing information assessments, we need to discover the accuracy of classifiers. Precision for any forecast model can be given as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

F-Score: (**also F-score or F-measure**) is the weighted normal of Accuracy and Review. In this way, F-Score takes both bogus positives and bogus negatives into thought. "F-score isn't generally as simple to secure as exactness, anyway it's far parcel extra helpful than precision, particularly Accepting that you have a lopsided class dispersion. Exactness works best in cases like if the bogus positives and the bogus negatives have tantamount expense.

$$\text{Accuracy} = 2 * \frac{(\text{Recall} + \text{Precision})}{(\text{Recall} + \text{Precision})}$$

Complexity analysis

In this subsection, the computational complexityof the proposed strategy is determined. In the initial step of the proposed method,the fisher score of all highlights ismeasured. The computational intricacy of the fisher score estimation is $O(ncp)$, where $n$is the quantity of the first highlights and $p$denotes the quantity of examples and $c$is the quantity of classes in the dataset. The initial step of the technique targets changing over the element space into a chart and requires $O(n2p)$time steps where $n$is the quantity of the first highlights and $p$denotes the quantity of examples. In addition, in the following stage, a local area recognition calculation is applied to discover the component groups. The intricacy of the local area detectionalgorithm is $O(n\log n)$. At that point a particular hereditary calculation based pursuit strategy is used to pick the last list of capabilities. The pursuit calculation will be rehashed for various iterative cycles (i.e., I). Accordingly, the time intricacy of this part is $O(IPkfk)$, where Pisthe number of the chromosomes in the populace, $k$is the quantity of the bunches and fkdenotes the time intricacy to compute the wellness work. The time intricacy of the KNN classifier is $O(Pn)$. Consequently, the computational intricacy of this stage is equivalent to $O(IP2nk)$.Consequently, the last computational intricacy of the proposed strategy is $O(n2p+n\log n+IP2nk)$, which are decreased to $O(n2p+p2\text{n})$

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

	OVERVIEW OF COMPLEXITY	
S.no	**Sentiment Classifiers**	**Complexity analysis**
1	**NB-SVM**	The NB-SVM hybrid model has more complexity than other hybrid models.
4	**KNN-SVM**	The KNN-SVM hybrid model has more complexity than SVM-ACO, ACO-PSO and SVM-ACO-PSO
2	**SVM-ACO**	The complexity of the SVM-ACO hybrid model is lower than that of the NB-SVM hybrid model, but it is not lower than that of the ACO-PSO, KNN-SVM, and SVM-ACO-PSO models.
3	**ACO-PSO**	ACO-PSO hybrid model complexity reduces complexity, but it is more complexity than SVM
5	**SVM-ACO-PSO**	The SVM-ACO-PSO hybrid model is shown to have reduced complexity and other model complexity is higher than in this model.

RESULTS AND DISCUSSIONS

This Paper displays and analysis the test outcomes and the assessment for our approach. To begin with, we look at consequences of changed strategies applied for the assessment examination of information got from Twitter. Second, the conversations on the impacts of different highlights are introduced. Likewise, we talked about the best gotten results (which were given by SVM, KNN, ACO,PSO and NB techniques). Correlation of NB, NB-SVM brings about the type of bar diagram having x-pivot containing exactness, review, precision and y-hub contains rate:

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

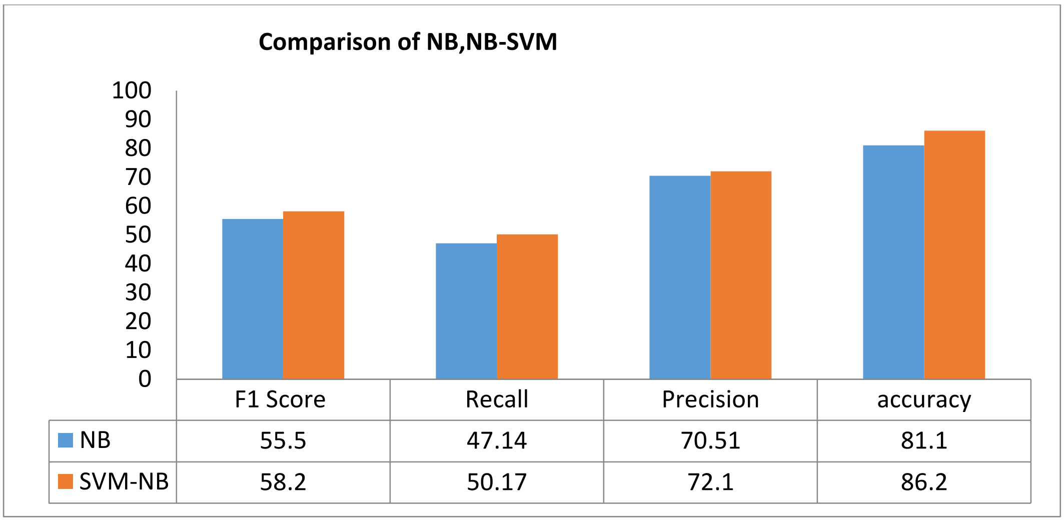


Fig. (2): Graph of results NB,NB-SVM

Correlation of SVM, SVM-ACO, SVM-PSO brings about the type of bar diagram having x-pivot containing exactness, review, precision and y-hub contains rate:

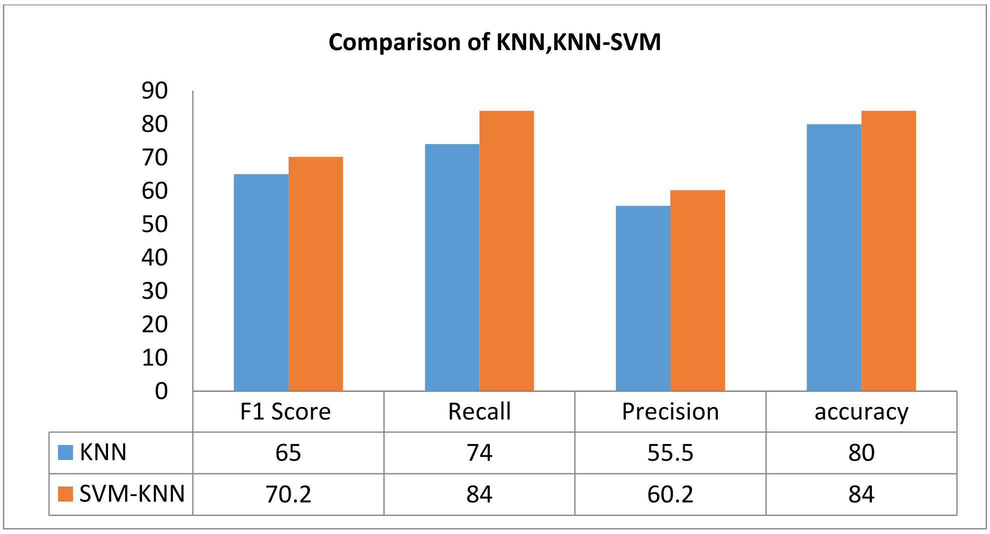


Fig. (3): Graph of results KNN,KNN-SVM

Examination of SVM, SVM-ACO, brings about the type of bar diagram having x-pivot containing exactness, review, precision and y-hub contains rate:

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

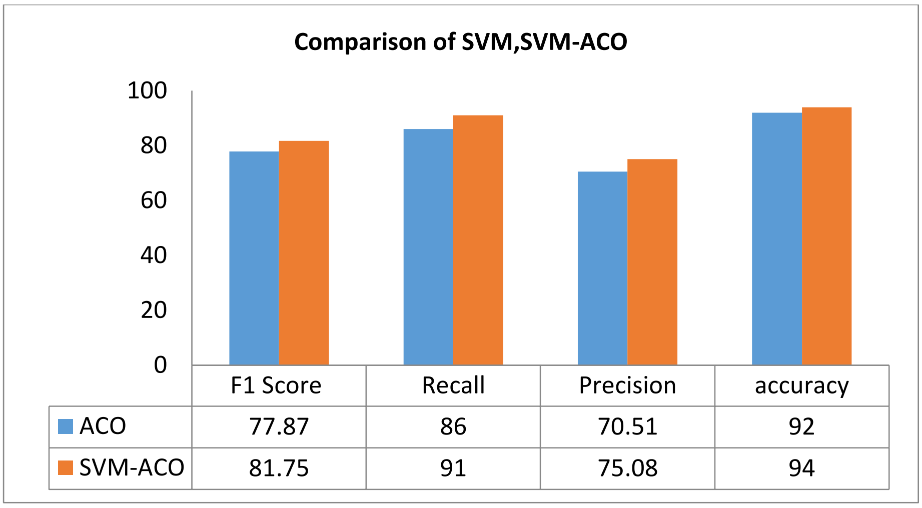


Fig. (4): Graph of results SVM,SVM-ACO

Correlation of SVM, SVM-ACO, brings about the type of bar diagram having x-hub containing exactness, review, precision and y-hub contains rate:

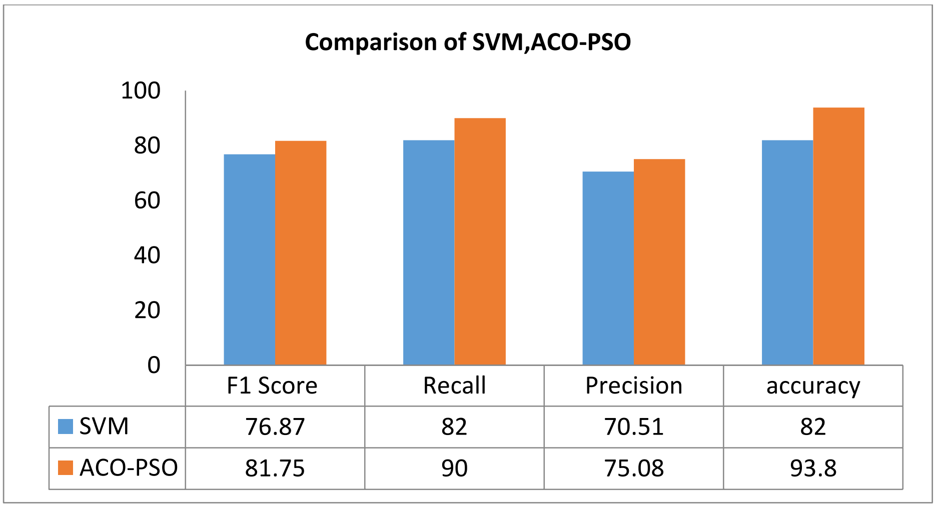
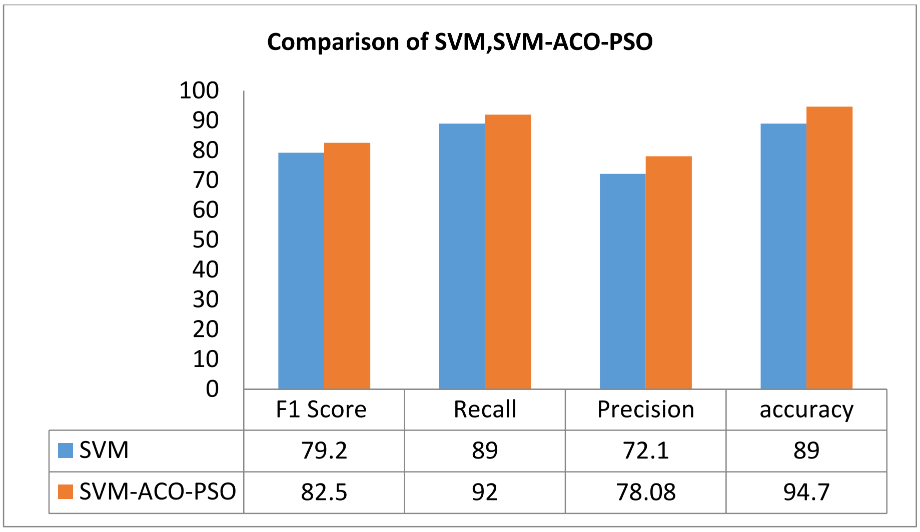


Fig. (5): Graph of results SVM,ACO-PSO

Correlation of SVM, SVM-PSO-ACO, brings about the type of bar diagram having x-hub containing exactness, review, precision and y-hub contains percentage:

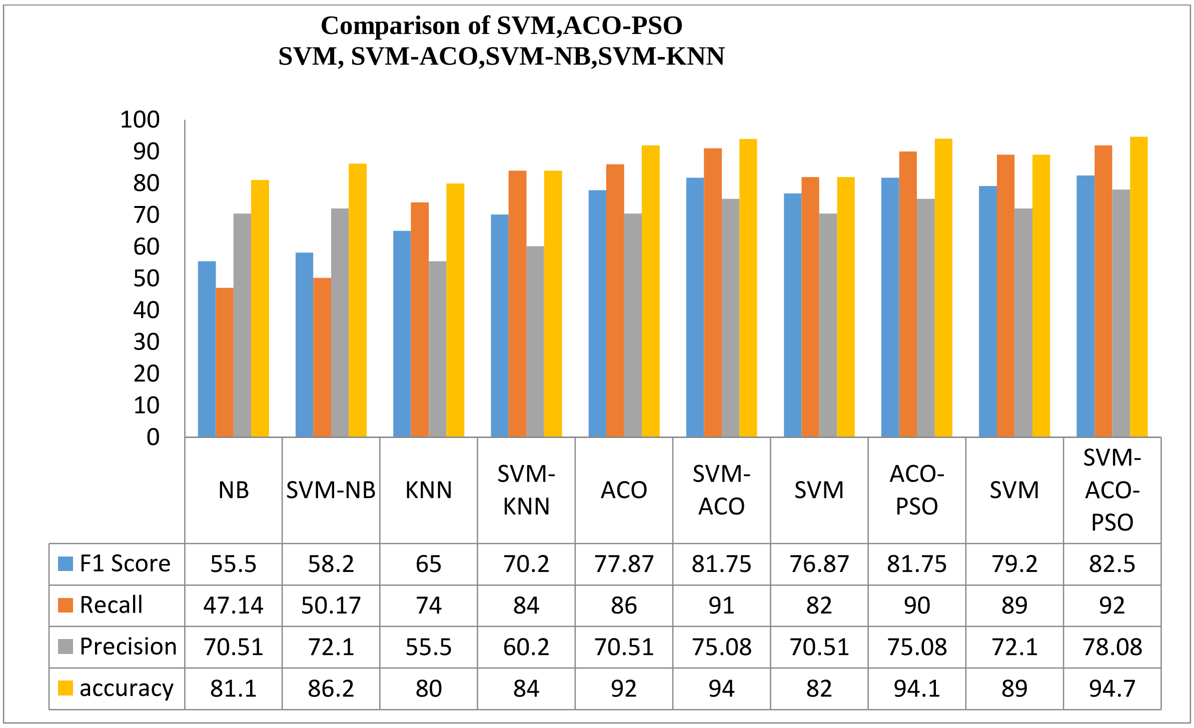
Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK



	F1 Score	Recall	Precision	accuracy
SVM	79.2	89	72.1	89
SVM-ACO-PSO	82.5	92	78.08	94.7

Fig. (6): Graph of results SVM, SVM-ACO-PSO

Examination of SVM, ACO-PSO,SVM, SVM-ACO,SVM-NB,SVM-KNN,SVM-ACO-PSO brings about the type of bar diagram having x-pivot containing exactness, review, precision and y-hub contains rate:



	NB	SVM-NB	KNN	SVM-KNN	ACO	SVM-ACO	SVM	ACO-PSO	SVM	SVM-ACO-PSO
F1 Score	55.5	58.2	65	70.2	77.87	81.75	76.87	81.75	79.2	82.5
Recall	47.14	50.17	74	84	86	91	82	90	89	92
Precision	70.51	72.1	55.5	60.2	70.51	75.08	70.51	75.08	72.1	78.08
accuracy	81.1	86.2	80	84	92	94	82	94.1	89	94.7

Fig. (7): Graph of results SVM,ACO-PSOSVM, ACO-PSO,SVM-ACO,SVM-NB,SVM-KNN

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

At long last, as far as exactness, accuracy, review, and F-measure, the SVM-ACO-PSO result beats different outcomes dependent on relative investigation with cross breed order techniques for assessment examination in this paper. Distinctive kind of mixture Model further work as there is at present an extensive proportion of opportunity to get better.

CONCLUSION

Sentiment analysis is a technique for determining people's thoughts, feelings, and emotions. Human beings can have good or negative viewpoints. The Element of speech is a function that is commonly used to extract the sentiment of written input. When determining sentiment from components, an adjective plays a crucial role. When adjectives and adverbs are used simultaneously, it might be difficult to distinguish between feeling and opinion. I designed, tested, and assessed a range of machine learning approaches for the Sentiment Analysis assignment. I learnt a lot about how to deal with machine learning issues and how to do data analysis to make the machine's learning process easier. The effective weight of KNN, NB, PSO, SVM, and ACO, as well as a discriminative classifier of SVM, are used to analyse the data in this text. Because SVM-ACO-PSO repeatedly changes the edge if the weight is diverse for different boundaries, the outcomes exhibit that SVM-NKK and SVM-PSO perform well, but not as good as SVM-ACO-PSO, which has a 94.7 percent accuracy.

For future work, we would like to make bigger the domain of our experiments and run the classifiers on multiple dataset considering number of different languages so as will have more representative inputs and thus better generalizable results.

ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

REFERENCES

[1] Go, Alec, Richa Bhayani, and Lei Huang "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford 1 (2009): 12.

[2] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREc. Vol. 10. 2010.

[3] Spencer, James, and Gulden Uchyigit. "Sentimentor: Sentiment analysis of twitter data." Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. 2012.

[4] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!" Icwsm 11 (2011): 538-541.

[5] Narr,Sascha, Michael Hulfenhaus and Sahin Albayrak "Language-independent twitter sentiment analysis" Knowledge Discovery and Machine Learning (KDML), LWA (2012): 12-14.

[6] Celikyilmaz, D. Hakkani-Tur and J. Feng "Probabilistic model-based sentiment analysis of twitter messages," in Spoken Language Technology Workshop (SLT), 2010 IEEE, pp. 79–84, IEEE, 2010.

[7] Y. Wu and F. Ren, "Learning sentimental influence in twitter," in Future Computer Sciences and Application (ICFCSA), 2011 International Conference on, pp. 119–122, IEEE, 2011.

[8] A.Pak and P. Paroubek "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of LREC, vol. 2010.

[9] R. Xia, C. Zong, and S. Li "Ensemble of feature sets and classification algorithms for sentiment classification," Information Sciences: an International Journal, vol. 181, no. 6, pp. 1138–1152, 2011.

[10] Neethu M.S., Rajasree R "Sentiment Analysis in Twitter using Machine Learning Techniques", 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India, IEEE.

[11] Shenghua Liu, Xueqi Cheng, Fuxin Li, and Fangtao Li "TASC:Topic-Adaptive Sentiment Classification on Dynamic Tweets", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, MANUSCRIPT, 2014.

[12] Rui Xia, Feng Xu, ChengqingZong, Qianmu Li, Yong Qi and Tao Li" Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE Transactions on Knowledge and Data Engineering, 2015. R. Suresh Ramanujam, R. Nancyamala, Nivedha, "SENTIMENT ANALYSIS USING BIG DATA", 2015 International Conference On Computation Of Power, Energy, Information And Communication.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

[13] Saif, Hassan, Yulan He, and Harith Alani "Semantic sentiment analysis of twitter" International Semantic Web Conference. Springer Berlin Heidelberg, 2012.

[14] Mukwazvure, Addlight, and K. P. Supreethi "A hybrid approach to sentiment analysis of news comments." Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015 4th International Conference on. IEEE, 2015.

[15] Z. Kechaou, M. Ben Ammar and A. M. Alimi, "Improving E-learning with sentiment analysis of users' opinions", *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, pp. 1032-1038, Apr. 2011.

[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 Syst. Appl.*, vol. 40, no. 16, pp. 6266-6282, Nov. 2013.

[17] I. Santos, N. Nedjah and L. de Macedo Mourelle, "Sentiment analysis using convolutional neural network with fastText embeddings", *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, pp. 1-5, Nov. 2017.

[18] Mendon, S., Dutta, P., Behl, A. *et al.* A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. *Inf Syst Front* (2021). https://doi.org/10.1007/s10796-021-10107-x

[19] Khan, Jawad and Byeong Soo Jeong "Summarizing customer review based on product feature and opinion." Machine Learning and Cybernetics (ICMLC), 2016 International Conference on. IEEE, 2016.

[20] Menaria H.K., Nagar P., Patel M. (2020) Tweet Sentiment Classification by Semantic and Frequency Base Features Using Hybrid Classifier. In: Luhach A., Kosa J., Poonia R., Gao XZ., Singh D. (eds) First International Conference on Sustainable Technologies for Computational Intelligence. Advances in Intelligent Systems and Computing, vol 1045. Springer, Singapore. https://doi.org/10.1007/978-981-15-0029-9_9.

[21] Yadav, A., Vishwakarma, D.K. Sentiment analysis using deep learning architectures: a review. *ArtifIntell Rev*53, 4335–4385 (2020). https://doi.org/10.1007/s10462-019-09794-5.

[22] Ducange P, Fazzolari M, Petrocchi M, Vecchio M (2019) Engineering applications of artificial intelligence an effective decision support system for social media listening based on cross-source sentiment analysis models. Eng Appl ArtifIntell 78:71-85. https://doi.org/10.1016/j.engappai. 2018.10.014 CrossRefGoogle Scholar.

[23] Hassan F, Usman K, Saba Q (2018) Enhanced cross-domain sentiment classification utilizing a multi-source transfer learning approach. Soft Comput https://doi.org/10.1007/s00500-018-3187-9CrossRefGoogle Scholar.

[24] Matveeva, O., Nechipurenko, Y., Rossi, L., Moore, B., Sætrom, P., Ogurtsov, A.Y., Atkins, J.F., Shabalina, S.A.: Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. Nucleic Acids Res. 35, e63 (2007) CrossRefGoogle Scholar.

[25] Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine (2007), http://www.ics.uci.edu/~mlearn/MLRepository.htmlGoogle Scholar.

[26] Nemati, S., Basiri, M.E., Ghasem-Aghaee, N., Aghdam, M.H.: A novel ACO–GA hybrid algorithm for feature selection in protein function prediction. Expert Systems with Applications 36, 12086–12094 (2009)CrossRefGoogle Scholar.

[27] Correa, S., Freitas, A.A., Johnson, C.G.: Particle Swarm and Bayesian networks applied to attribute selection for protein functional classification. In: Proc. of the GECCO 2007 Workshop on Particle Swarms, The Second Decade, pp. 2651–2658 (2007)Google Scholar

[28] Jain, C.K., Prasad, Y.: Feature selection for siRNA efficacy prediction using natural computation. In: World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), pp. 1759–1764. IEEE Press, Los Alamitos (2009)CrossRefGoogle Scholar

[29] Kušen E, Strembeck M (2018) Politics, sentiments, and misinformation: an analysis of the Twitter discussion on the 2016 Austrian presidential elections. Online Soc Netw Media 5:37–50. https://doi.org/10.1016 j.osnem.2017.12.002CrossRefGoogle Scholar

[30] Sun T, Wang J, Zhang P, Cao Y, Liu B, Wang D (2017) Predicting stock price returns using microblog sentiment for chinese stock market. In: Proceedings 2017 3rd International conference big data computer communications BigCom 2017, pp 87–96. https://doi.org/10.1109/BIGCOM.2017.59

[31] Ghiassi M, Zimbra D, Lee S (2016) Targeted twitter sentiment analysis for brands using supervised feature engineering and the dynamic architecture for artificial neural networks. J Manag Inf Syst 33:1034–1058. https://doi.org/10.1080/07421222.2016.1267526CrossRefGoogle Scholar

[32] Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA (2017) Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE 12:1-22. https://doi.org/10.1371/journal.pone.0171649CrossRefGoogle Scholar

[33] Asghar MZ, Kundi FM, Ahmad S, Khan A, Khan F (2018) T-SAF: twitter sentiment analysis framework using a hybrid classification scheme. Expert Syst 35:1-19. https://doi.org/10.1111/exsy.12233CrossRefGoogle Scholar

[34] Baccianella S, Esuli A, Sebastiani F (2010) SENTIWORDNET 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In:

ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

Proceedings of the 7th international conference on language resources and evaluation, LREC 2010. pp 2200–2204Google Scholar

[35] Hamdan H, Bellot P, Bechet F (2015) Sentiment lexicon-based features for sentiment analysis in short text. In: Conference 16th international conference on intelligent text processing and computational linguistics, pp 217–226Google Scholar

[36] Strapparava C, Valitutti A (2004) WordNet-affect: an affective extension of WordNet. In: Proceedings of the 4th international conference on language resources and evaluation, LREC 2004, pp 1083–1086Google Scholar

[37] Appel O, Chiclana F, Carter J, Fujita H (2018) Successes and challenges in developing a hybrid approach to sentiment analysis. Appl Intell 48:1176–1188. https://doi.org/10.1007/s10489-017-0966-4CrossRefGoogle Scholar

[38] Kolchyna O, Souza TTP, Treleaven P, Aste T (2015) Twitter sentiment analysis: lexicon method, machine learning method and their combinationGoogle Scholar

[39] Zainuddin N, Selamat A, Ibrahim R (2018) Hybrid sentiment classification on twitter aspect-based sentiment analysis. Appl Intell 48:1218–1232. https://doi.org/10.1007/s10489-017-1098-6CrossRefGoogle Scholar

[40] Bollegala D, Weir D, Carroll J (2011) Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In: ACL-HLT 2011—Proceedings of the 49th annual meeting of the association for computational linguistics human language technologies, vol 1, pp 132–141Google Scholar

[41] Chan WN, Thein T (2018) A comparative study of machine learning techniques for real-time multi-tier sentiment analysis. In: 1st IEEE international conference on knowledge innovation and invention, ICKII 2018, Institute of Electrical and Electronics Engineers Inc, pp 90–93. https://doi.org/10.1109/ICKII.2018.8569169

[42] Mansour R, Hady MFA, Hosam E, Amr H, Ashour A (2015) Feature selection for twitter sentiment analysis: an experimental study. In: Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and Lecture notes in bioinformatics), Springer, pp 92–103. https://doi.org/10.1007/978-3-319-18117-2_7

[43] Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert Syst Appl 40:6266–6282. https://doi.org/10.1016 /j.eswa. 2013.05.057CrossRefGoogle Scholar

[44] Liu M, Song Y, Zou H, Zhang T (2019) Reinforced training data selection for domain adaptation. In: Proceedings of the 57th annual meeting of the association for computational linguistics, pp 1957–1968Google Scholar

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication：Online Open Access
Vol: 40 Issue: 06-2021
DOI 10.17605/OSF.IO/2NVJK

[45] Ducange P, Fazzolari M, Petrocchi M, Vecchio M (2019) Engineering applications of artificial intelligence an effective decision support system for social media listening based on cross-source sentiment analysis models. Eng Appl Artif Intell 78:71–85. https:// doi.org/10. 1016/ j.engappai.2018 .10.014CrossRefGoogle Scholar

[46] Hassan F, Usman K, Saba Q (2018) Enhanced cross-domain sentiment classification utilizing a multi-source transfer learning approach. Soft Comput https://doi.org/10.1007/s00500-018-3187-9CrossRefGoogle Scholar

[47] Sanders NJ (2011) Twitter sentiment corpus. Sanders analytics. Sanders analytics LLC Web 16 Nov 2013Google Scholar.

[48] Anbananthen KSM, Elyasir AMH (2013) Evolution of opinion mining. Aust J Basic Appl Sci 7(6):359–370

[49] Appel O, Chiclana F, Carter J (2015) Main concepts, state of the art and future research questions in sentiment analysis. Acta Polytechnica Hungarica - J Appl Sci 12(3):87–108.

[50] Brenga C, Celotto A, Loia V, Senatore S (2016) Capturing digest emotions by means of fuzzy linguistic aggregation. Springer International Publishing, Cham, pp 113–139Google Scholar

[51] Dzogang F, Lesot M-J, Rifqi M, Bouchon-Meunier B (2010) Expressions of graduality for sentiments analysis – a survey IEEE international conference on fuzzy systems (FUZZ), 2010, pp 1–7