



# Latent Dirichlet Allocation Feature Extraction with Bio-Inspired Pigeon Feature Selection Technique for Twitter Sentiment Analysis

S.Kasthuri<sup>1</sup>, Dr.A.Nisha Jebaseeli<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Govt. Arts & Science College (affiliated to Bharathidasan University), Lalgudi, Trichy, Tamilnadu, India. vikramaug17081984@gmail.com

<sup>2</sup>Assistant Professor & Head, Department of Computer Science, Govt. Arts & Science College (affiliated to Bharathidasan University), Lalgudi, Trichy, Tamilnadu, India. gganapathy@gmail.com

## ABSTRACT

Feature selection act an important role in structure of machine learning. Nonlinear features in the data upset the accuracy performance of the model and rise the training period required to form the system. Feature selection (FS) is asignificant process for making sentiment analysis (SA). In this study, we propose a wrapper FS algorithm for SA. This algorithm uses the technique ofpigeon inspired optimizer (PIO) used to purpose of selection. These algorithm was assessed using three standard datasets: the STC-Twitter dataset. The proposed algorithm overcomes many FS algorithms from the most advanced related conceptssuch as, TPR, accuracy, FPR and F-score. In addition, the projected cosine likeness system for binarizing the algorithm has a quicker integration than the sigmoid scheme. Stimulation outcome display that the proposed PIOFS using the Latent Dirichlet Allocation (LDA) feature extraction technique yields better results.

**Key words:**Sentiment Analysis, Twitter data, Latent Dirichlet Allocation Feature Extraction, and Bio-Inspired Pigeon Feature Selection.

## 1. INTRODUCTION

Life of human in the world is full of feelings and ideas. Without them we cannot visualize the world. Emotions and opinions play an important role in almost all people's activities. They influence human life by manipulating the way we think, how we act and how we do. Accessing large amounts of data on the Internet and transforming them into a social network is no problem now, as every day new terabytes of information are available, available to

everyone. More importantly, it has changed the way we share facts. The recipients of the data are not only content available on the web, but actively interpret this content and create new info. Today, people not only comment on current information, bookmarks and ratings, but also share ideas, news and insights with the community at large. In this way, awhole community becomes a writer and a reader [2]. Users, such as existing blogs, wikis, forums and social networks can post information, provide comments and receive feedback on other topics, from policy, health to product reviews and travel. The growing popularity of different types of personal publishing services suggests that opinion information will become an important aspect of textual data on the web. In recent times, several investigators have concentrated on this area [3].

They attempt to obtain feedback by analyzing and summarizing the opinions automatically expressed by computers. This new area of research is commonly referred to as "Opinion Mining and SA" [4]. So far, the researchers have isolated all the techniques of problem solving. The analysis of mining concepts and current feelings is a cross-sectional study of info retrieval (IR) and (NLP) natural language processing and shares specific features with text extraction and information extraction [5]. Today, in this world, large amounts of data are created every day, sentimental analysis has become an important tool for understanding each processed data. This has allowed companies to obtain the results of the different processes that they are currently developing. This is a type of oral research or mining. In this paper, we apply statistics, NLP and ML to find, analyze and extract some essential info from tweets [6]. The main purpose is to observe a reviewer's feelings, expressions, thoughts, or judgments about a particular topic, event, or company. This type of

analysis is also called opinion extraction, which is the main objective of the extraction. The purpose of this analysis is to analyze the opinion of our audience on a particular topic from a variety of sources. [7] Repeated neural networks recognize every word founded on the consideration of earlier words. There are links with persistent loops that allow the existence of information. In a RNN, the production from the previous stage is introduced in the current stage. [8]

In traditional NN, every inputs and outputs are liberated of each other, but sometimes the next word of a sentence must be predicted because the previous words are required, so the previous words must be remembered. . To overcome the drawback, RNN solves this problem with the help of a hidden layer. SVM can be used for linear and nonlinear regression and classification. The huge array of data that contains false positives and non-essential and unwanted functions. These features not only slow finding, but also consume many computing resources. To improve accuracy, you must have a system to select the best functions. FS helps solve some common difficulties in SA by identifying appropriate functions. The relevant functions carry important information that helps a lot in the classification. An important thing to keep in mind regarding FS in SA is that it reduces processing costs, reduces memory and improves the understanding of test data. As FS is an ML idea, it is employed using several algorithms. FS is performed using systems such as statistical analysis, SVM, data mining and neural networks. In addition, FS provides a detection mechanism that can be grouped into 3 kinds: incremental, random, and reduced selection. The collection system is used to decide and pick the corresponding objects in the data set. It is interesting to note that FS can be attained using several technologies, with intelligence patterns, swarm intelligence (SI) and ANN.

Metaheuristic algorithms are mainly need for FS in intrusion detection schemes, particularly due to their great accuracy level. In this context, SI is a significant practice that is used in the process of implementation and categorization of metaheuristic algorithms. swarm intelligence and artificial intelligence is used to solve the difficult issue and also its behavior insects and swarm. There are two attitudes used in swarm exploration include PSO and ACO. In addition to the widely used swarm-based FS algorithms, other customizable algorithms use metaheuristics and SI, which were used to make various methodologies and structures. The foremost influences of this document are detailed below.

- i. Review the Sophisticated Works related to FS algorithms for SA.
- ii. Propose a novel FS algorithm for SA based on pigeon inspired optimizer.
- iii. Test the projected FS algorithm and equate the outcome with six well-known algorithms from the Sophisticated Works.
- iv. Pigeon bio inspired feature selection

## 2. LITERATURE SURVEY

Zhao Jiangqiang et al.[11] have suggested an development over traditional sentimental analysis attitudes that focus on the lexical enquiry of the emoticons, words of each unit of tweets. The projected attitude is based on unsupervised learning. Twitter Corp. considers significant hidden links of words based on contextual meaning and combined statistics between words in tweets. To create a set of tweet functions, the term “inlay” is combined with n-gram elements and punctuation characteristics of word polarity. For training and marking the classifier of sensations, a set of functions is included in deep CNN. The performance ratings of these models are compared with the base ideal in five types of Twitter datasets, in which these models work well with F-measure and accuracy.

K Lavanya et al. [12] proposed a system of classification based on a training method consistent with the subject of tweets. The system is intended to address the problem of dealing with a wide variety of topics on Twitter. In this attitude, the non-textual characteristics of tweets are used to train the algorithm for classification. The projected system can be applied to static data of totally dissimilar topics and to dynamic data for a specific timeline during transmission. Positive, neutral and negative are the three main classes in tweets. These three class tags can be further expanded to form 5 class tags: very positive, negative, neutral, positive, and very negative. This method achieved the better performance outcome of recall, accuracy value.

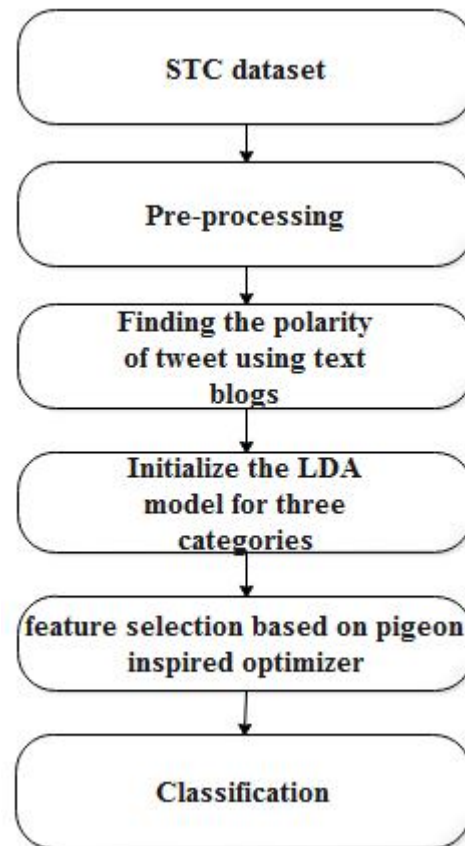
Chintan Dedhia, et al. [13] has to develop the power of the SVM classifier for sentiment classification, an add-on method was proposed, which is the most commonly used ML algorithm for SA and classification. This synchronization perfect is designed by integrating the SVM algorithm into the base classifier and the AdaBoost algorithm for set strengthening. Tweets, tags within tweets, and other essential tweet features use the projected algorithm to find information about the link between tweets. It helps to study the characteristics of social interaction. The project uses the synchronization model to categorize Twitter data into label of positive and negative. Standard SVM algorithm compares the projected approach, which performs better with esteem to accuracy, recall, and F-m

Yeqing Yan, et al. [14] has presented two synchronization models were developed using Naïve Bayes with Mallet's MaxNT and Centristrength with Pattern of Text Blob. This model overcomes the problem of properly trained classifiers when there is not enough training data; That is, to classify tweets related to a product, tweets associated to other similar products can be used to study classifiers. Both models verified to be efficient in demonstrating high correctness in the classification of feelings from twelve various data sets. Both synchronization classifiers can be implemented in equivalent and able to handling big circles of Tweet data.

Paramita Ray et al. [15] proposed an R-based structure built on a dictionary-based attitude to SA and classification of reviews about product. It make better decisions about products and facilities. Initially it involves preprocessing the input tweets, replacing the abstracts with their extended forms, describing emotions as emoticons, eliminating empty words and manipulating suggestions. The project methodology analyzes the level of the document and the level of perspective. The author's objective is to develop a hybrid work model based on ML methods for SA and classification.

### 3. PROPOSED SYSTEM

Figure 1 shows PIOFS based twitter sentiment classification. Based on the Pigeon FS and LDA, framework has provide highly effective results for sentiment classification. The proposed framework performs classification of sentiment in five major steps: data gathering using Twitter datasets, Twitter pre-processing includes filtering to filter unique Twitter attributes, and the polarity values of tweets are identified in the pre-processing phase. Extracting features, using TF-IDF features. The information of the extracted high-efficiency features is selected with the help of Pigeon FS, which provides for a more classification. The LDA classifies three groupings of tweets by launching them, and then provides them as the input to the DT for categorizing tweets.



**Figure 1:** Proposed System.

#### 3.1. Data Collection

To achieve our task of getting customer insight for telecom domain, we selected three Saudi telecom companies (STC), Mobily and Zain to collect data. Twitter search APIs have been used to retrieve 1000 tweets from each company using filtering terms that are relevant to their official and their service accounts (@STC #STC, @STCcare, @ZainHelpSA, #Zain, @Zainksa, #Mobily, @Mobily1100 and @Mobily) defined by longitude and latitude of Saudi Arabia.

#### 3.2. Preprocessing

The major part of these paper is tweet data preprocessing. The data recovered in (JavaScript Object Notation) JSON layout is first transformed to normal text data. It covers following –

- Identification of every caps letter
- Lower case
- Removal of URL
- Analysis of emotion
- Punctuations and blank spaces removal
- Word Compression

For the purpose of the algorithm, the tweet is first transformed to lowercase format, since the tweet can be uppercase or lowercase. The tweet can contain URLs, so all URLs are excluded from messages with the help of expressions or common word URLs. The

user names referred to in the recovered data are removed with the help of a regular expression or any other term with neutral polarity. Since the words in the hashtag remain unchanged, they can be used to model themes. If the word has multiple repetitions such as 'Happpppyyy', such words are converted into 'Happy' digits by eliminating the maximum number of possible repeats. Removes icons and extra white spaces, leaving only a white space in the mid of words and removing punctuation with the assistance of parsing.

#### 4. FEATURE EXTRACTION

This section applies a new method for feature extraction to extract things from text data. Counter vectorization is the theory of signifying or extracting the words from a big text. Here, a set of correlations was assigned to each word to determine the similarity between the available words. Similarly, the term frequency and reverse frequency of a document are completed to extract useful functions from text data. Frequency of a reverse file measure the frequency of occurrence of a word in a document using Equ. (1) and (2).

$$\text{Term frequency} = \frac{\text{No. of times term (t) appears in document}}{\text{Total no. of terms in a document}} \quad (1)$$

$$\text{frequency} = \log \frac{\text{Total no. of documents}}{\text{No. of documents with term (t)}} \quad (2)$$

The TF-IDF technique is used to extract useful information from tweets. In the following stages, the LDA system is launched to categorize tweets into three categories: negative, positive, and neutral.

##### 4.1. LatentDirichlet Allocation

After extracting process, the LDA approach is used to identify tweets in three classes. In general, the LDA is a probabilistic thematic model, while every document is treated as a random mix of hidden themes. Each hidden subject is marked as a fixed word distribution in the LDA, which is used to determine the underlying structure of the hidden subject based on the observed data. Normally, words are created in two steps for each document. In the first step, the distribution of topics is randomly picked for each document. In LDA, the word  $W$  is an array of  $\{1, \dots, V\}$  and  $N$  words from a dictionary index.

The LDA monitors the representation of the three layers and checks the parameters  $\pi$  and  $\mu$  when creating the case. For each document, we examine variable variables at the subject level. As a result, word variables are checked for each word in the LDA document. The co-distribution of a random variable is signified as an LDA generative process. The probability density function of the  $k$ -dimensional random differential variable is determined by using

the equation. (3) By default, the joint distribution and body potential of the thematic mixture are calculated using the Eq. (4) and (5).

$$p(\mathbf{x}|\pi) = \frac{\Gamma(\sum_{i=1}^k \pi_i)}{\prod_{i=1}^k \Gamma(\pi_i)} \mathbf{x}_1^{\pi_1-1} \dots \mathbf{x}_k^{\pi_k-1} \quad (3)$$

$$p(\mathbf{x}, \mathbf{y}|\pi, \mu) = p(\mathbf{x}|\pi) \prod_{n=1}^N p(x_n|\mathbf{x}) p(y_n|x_n, \beta) \quad (4)$$

$$p(D|\pi, \mu) = \prod_{d=1}^M \int p(\mathbf{x}_d|\pi) \times (\prod_{n=1}^{N_d} \sum_{x_{dn}} p(x_{dn}|\mathbf{x}_d) p(y_{dn}|x_{dn}, \mu)) d\mathbf{x}_d \quad (5)$$

Where  $\mathbf{x}$  Where document subject is represented as subject variables,  $\mathbf{a}$  is denoted as a Dirichlet parameter,  $M$  is signified as a document,  $N$  is denoted as sum of words,  $x$  is denoted as subject,  $x$  is denoted as a word theme, and the word is observed.

In one document, the computation of the subsequent distribution of the latent variable is an important assumption of the LDA classic. The exact inference of the subsequent distribution of the latent variable is an unresolved problem. The combination of LDA with approximate algorithms such as Gibbs sampling, Markov chain, Laplace and variable approximation is broadly used to extract the keyword. The neutral, negative, and positive classes are extracted by the individual weight value, and these values are stored in the dictionary. To reach a neutral, negative, and positive weight value, the data of the test emotions are coordinated with the dictionary of the test phase. The values of the 3 classes are 30.86 meant for Positive, 59.12 meant for Neutral and 10.013 for Negative Tweets of data gatheredby LDA. Next gaining three weight values, the decision tree algorithm is used to perform the classification process.

##### 4.2. FSusing Pigeon bio-inspired algorithm

PIO algorithms have recently been exposed to be effective in solving various optimization issues, including aerial robot trajectory planning, three-dimensional trajectory planning, an automatic landing system, and a PID development controller. In this article, we adopt the FS algorithm for IDS based on the fresh binary version of PIO. This unit offers two versions of the PIO. The first version or algorithm uses a sigmoid function to sample the speed of the doves, the second version offers an updated modified binary version of the basic PIO, which uses cosine similarity to determine the speed of the doves. Both versions use an equal fitness function, on the other hand, each version has methods that represent a pigeon or a solution. The process of comparing the optimization tasks of PIO and FS is presented in Table 1.

**Table 1:** PIO map to FS optimization problem.

PIO Concept	FS representation
Number of Pigeons $N_p$	Number of solutions
Position of Each Pigeon $X_p$	Solution (selected features)
Pigeon or solution ( $X_p$ ) length	Total number of features
Best Pigeon	The solution that has the best fitness value
The velocity of each Pigeon $V_p$	The amount of change toward the best pigeon
Fitness Function	Model Evaluation based on TPR, FPR, and number of selected features
$V_c$	Number of iterations

#### 4.2.1. Fitness Function

Fitness Function or objective or cost function is the terms of a process for evaluating the sufficiency of solutions. The fitness function evaluates the solution, which is a subset of the functions selected according to the true positive speed (TPR), the false positive speed (FPR), and the sum of functions. The sum of functions is involved in the adaptation function, so, if there is a few function that does not disturb the TPR or FPR, we want to avoid it. Eq. 6 represents the formula used to calculate the taste of a dove or solution. Here is the sum of objects chooses, the total number of objects in SF and NF is  $w_1 + w_2 + w_3 = 1$ . The weight is set as follows:  $w_1 = 0.1$ ,  $w_2 = w_3 = 0.45$ , because TPR and FPR are equal.

$$FF = w_1 * \frac{SF}{NF} + w_2 * FPR + w_3 * \frac{1}{TPR} \quad (6)$$

#### 4.2.2. Sigmoid PIO for FS

Defines a solution or a pigeon vector of length equal to the sum of key FS PIO. In the case of STC data, the length of the pigeon vector or solution is 19. Since the basic PIO procedure continuously processes the dove's position, the specific PIO solution for the FS is defined as a vector whose values of velocity and position vectors are fixed randomly among initial [0, 1]. The traditional method is used to measure the rapidity of every pigeon according to Equ. (1), and then the sigmoid function is used to translate the velocity into a binary version according to Equation 7.

For the binary files of the cluster intelligence algorithm, the location of each dove is updated based on the value of the sigmoid function and the probability of a uniform random numeral between [0, 1] according to Equation 8. The algorithm will act as an old PIO, except for updating the position of the ground operator. Additionally, the sigmoid function will be used to transfer the speed, and then the locations will be updated accordingly.

$$S(V_i(t)) = \frac{1}{1 + e^{-\frac{V_i(t)}{2}}} \quad (7)$$

$$X(t)_{(i,p)}[i] = \begin{cases} 1, & \text{if } (S(V_i(t)) > r) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

#### 4.3. Decision Tree Algorithm (DTA)

It is a data mining induction method that repeatedly divides a data set into a deep or wide range using a greedy approach until all the data elements belong to a particular class. The DT structure consists of root, internal, and leaf nodes. Tree structure is used to classify records of anonymous data. For each internal node of the tree, the best separation is determined by using impurity measures. Tree leaves are formed by class labels that classify data elements. The DT classification technique is carried out in two stages: tree assembly and tree pruning. The construction of the trees is carried out from top to bottom. It is at this stage that the tree is recursively divided into parts until all the data elements belong to the same class tag. This is a very difficult task and requires significant computational resources, since the training data set is displayed repeatedly. Tree felling is done according to the principle from the bottom up. It is used to improve forecast accuracy and algorithm classification by minimizing redundant matching. Excessive adjustment in the DTA leads to misclassification. Tree pruning is less complicated than the tree's growth phase, since the training data set is scanned only once. In the proposed system, the decision tree classification provides a better option for the end user to classify the positive and negative tweets. It is done by comparing the maximum frequent items generated by the rules in the training data have been compared with the maximum frequent items of the test data and hence the classification can be made easily.

## 5. EXPERIMENTAL RESULT AND DISCUSSION

This unit briefs the results of the test, the discussion of our technique, and details on performance measurement, experimental setup, quantitative analysis and comparative analysis. The proposed system was implemented using Python with 4GB RAM, 1TB hard disk and Intel i5 3.0 GHz processor. Presentation was analysis with other classification techniques and past research based on the Twitter dataset. To evaluate the effectiveness of the proposed system. The performance of the projected system was calculated in terms of precision (PR), accuracy (ACC), recall (RC), and F-measure (F-m).

### 5.1 Performance measure

The performance indicator is defined as the regular measurement of results and results, which allows you to obtain reliable info on the efficiency of the proposed system. In addition, an indicator of effectiveness is the process of informing, gathering and analyzing info about the results of the activities of a group or solo. The confusion metric for evaluating the classifier for binary data is displayed in Table 2 and can be understood in the following terms:

**Table 2:** Confusion Matrix

	Predicted positive	Predicted Negative
Actual positive	True Positive (TP)	False Positive (FP)
Actual Negative	False Negative (FN)	True Negative (TN)

TP refers to real positive events that are correctly classified as positive. The FP then indicates that the true positive classified is falsely negative. Similarly, TN is true negative and falsely classified as positive and FN are true negative events.

The mathematical equation of ACC, F-m, PR, and RC are denoted in the Eq. (6), (7), (8), and (9).

$$Accuracy = \frac{TN+TP}{TP+TN+FN+FP} \times 100 \quad (6)$$

$$F - measure = \frac{2TP}{(2TP+FP+FN)} \times 100 \quad (7)$$

$$Precision = \frac{TP}{(FP+TP)} \times 100 \quad (8)$$

$$Recall = \frac{TP}{(FN+TP)} \times 100 \quad (9)$$

Where, true positive is denoted as TP and TN express true negative and then FP is expressed as false positive, and FN is expressed as false negative.

## 5.2. Performance Analysis of Proposed Decision Tree

In this study, simulated Twitter data is used to compare estimates of the effectiveness of existing methodologies and the proposed approach in terms of ACC, F-m, and RC. Experiments is performed under different experimental conditions in STC data sets. The experiments have as main objective to determine the best controlled approach with the optimization of the proposed IOP compared to the random forest (RF), MLP and DT. DT is calculated for each thematic text classified in the Sanders Twitter. This classification helps the organization focus on the tweets with the greatest impact. Below is a discussion on specific topics of DT Sanders in the following sections.

### 5.2.1. Performance Analysis in terms of Precision, Recall and F-Measure

In this division, the presentation of DT with FS based PIO optimization are compared with MLP, RF and DT algorithm by both with and without implementation of PIO techniques in terms of PR, RC and F-m for three classes such as negative, positive and neutral in Table 3.

**Table 3:** Performance of DT-PIO for three classes of STC dataset

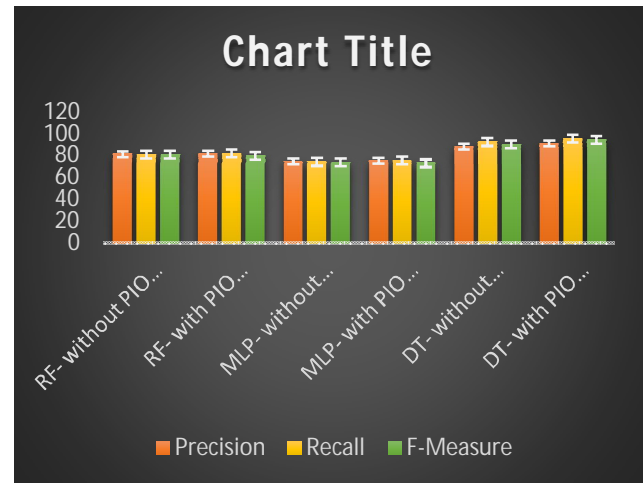
Methods	Parameters	Positive	Negative	Neutral
RF-without PIO optimization	Precision	0.56	0.27	0.66
	Recall	0.39	0.04	0.86
	F-Measure	0.46	0.07	0.74
RF-with PIO optimization	Precision	0.58	0.29	0.68
	Recall	0.41	0.05	0.87
	F-Measure	0.43	0.27	0.71
MLP-without PIO optimization	Precision	1.00	1.00	0.59
	Recall	0.06	0.02	1.00
	F-Measure	0.11	0.04	0.75
MLP- with PIO	Precision	1.11	1.11	0.61

optimization	Recall	0.08	0.04	1.11
	F-Measure	0.09	0.03	0.71
DT- without PIO optimization	Precision	0.92	0.46	0.86
	Recall	0.74	0.46	0.95
	F-Measure	0.82	0.46	0.91
DT- with PIO optimization	Precision	0.96	0.48	0.89
	Recall	0.89	0.48	0.98
	F-Measure	0.79	0.48	0.86

In table 3 display a performance of proposed DT-PIO performance is displayed. In above every techniques attain the outcome value by with PIO optimization, however after including the PIO optimization technique the performance result would much better than without PIO optimization. And then our PIO achieved 96% PR for positive class, 48% for negative and 89% for neutral, whereas the best R value for DT- PIO is 98% for neutral, 48% for negative and 89% for positive. The F-m of DT- PIO is 79% for positive, 86% neutral and 48% for negative. While compared with previous techniques by both with and without PIO optimization method, when our proposed system attain a better performance result than existing system.

**Table 4:** Whole performance of PIO in terms of PR, RC and F-m

Methods	Precision	Recall	F-Measure
RF- without PIO optimization	81.16	80.85	80.83
RF- with PIO optimization	81.88	81.98	79.55
MLP- without PIO optimization	74.74	74.20	74
MLP- with PIO optimization	75.41	75.42	73
DT- without PIO optimization	88.12	92.31	90.17
<b>DT- with PIO optimization</b>	<b>91.12</b>	<b>95.31</b>	<b>94.34</b>



**Figure 2:** Overall performance of PIO in terms of precision, recall and F-Measure

In table 4, it is clearly shows that the proposed DT-PIO achieved higher values of PR, RC and F-m when compared with existing techniques as using both with and without optimization. But every existing model by using optimization technique to attain performance better than without including optimization techniques. So optimization based FS techniques to yield the better performance. The next unit will define the performance of DT-PIO in terms of ACC.

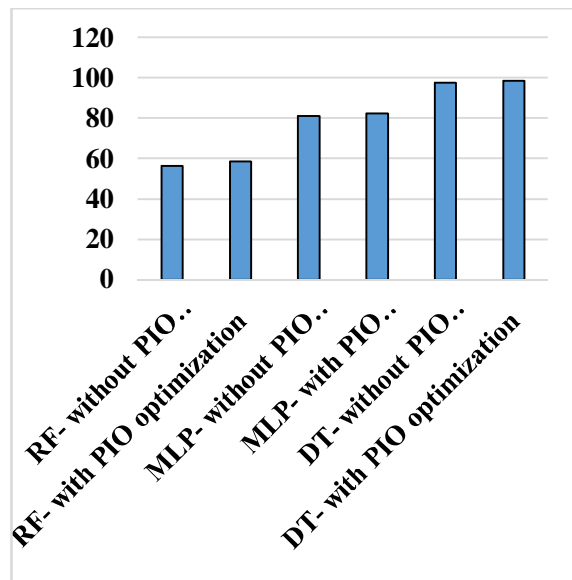
### 5.2.2. Performance Analysis of IWO in terms of ACC

In this unit, the classification ACC of proposed PIO are compared with existing techniques such as RF, MLP and DT by using both with and without optimization.

**Table 5:** Performance of ACC for proposed DT- PIO.

Methods	Accuracy
RF- without PIO optimization	56.46
RF- with PIO optimization	58.55
MLP- without PIO optimization	80.97
MLP- with PIO optimization	82.48
DT- without PIO optimization	97.51
DT- with PIO optimization	98.61





**Figure 3:** Performance of Accuracy for proposed DT- PIO.

From the investigational outcomes, the proposed DT-PIO achieved higher ACCAs 99.88%) for twitter sorting from STC datasets. The RF attained very low ACC when without optimization such as 56.46%. And the by using optimization technique RF achieved 58.55% of ACC. When compared with all other previous techniques without PIO optimization result lower than after implement of optimization. The MLP approach obtained 82.48% A by using PIO optimization. And then our proposed DT-PIO achieved the 99.88% of A. When compared with existing system, our proposed system attain the better accuracy

The results show that the proposed DT-PIO demonstrates its ability to improve the discriminatory power of terms to classify tweets. The proposed method can help reduce dimensions, analyze moods, and detect spam and many other applications that face several tweet problems. Therefore, the proposed scheme mitigates the impact of these problems on the classification task. Although our results were obtained for the data of Twitter, one of the most common platforms of social networks, we believe that this proposed method is applicable to other social networks.

## 6. CONCLUSION

In this paper, we propose a novel FS algorithm based on a PIO for emotional analysis. The proposed PIO FS aims to reduce the number of actions required to select the strongest behavior while maintaining the high rate and accuracy of detecting features with small false alarms. The projected PIO FS algorithm reduced the features in the STC data set. This

maintains TPR and high accuracy and reduces the time required to build the model. FS is a particular optimization problem. To solve this problem, a special process of continuous intelligent algorithms must be applied. A new method of discrimination is proposed equated to the traditional method used by researchers using a continuous algorithm based on the use of cosine similarity. The proposed discrimination process displays faster integration than the traditional method of using sigmoid function in the similarsum of iterations of the PIO.

## REFERENCES

- [1]. Shobana G, Vigneshwara B, ManirajSaiA. **Twitter Sentimental Analysis**.*International Journal of Recent Technology and Engineering* (IJRTE). No.78, pp. 15-78, 2018.
- [2]. Shabaz M, Kumar A. AS: **a novel sentimental analysis approach**.*International Journal of Engineering and Technology*. Vol., 23, no. 2, pp. 46-9, 2018.  
<https://doi.org/10.14419/ijet.v7i2.27.11679>
- [3]. Donepudi Babitha, Jayasankar.T, Sriram V.P, Sudhakar S, KollaBhanu Prakash,” **Speech Emotion Recognition using State-of-Art Learning Algorithms**”, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 2, pp.1340-1345, 2020.  
<https://doi.org/10.30534/ijatse/2020/67922020>
- [4]. Avudaiappan.T, Jenifer, Sisaytumsa, Subashree, T.Jayasankar, “**Twitter Sentimental Analysis Using Neural Network**”, *International Journal of Scientific & Technology Research*, Vol.9, No. 02, pp. 2573-2577, February 2020
- [5]. Aiswaryadevi. VJ, P, Rajanaren and S, Rameshkumar and S, Rajeevan, “**Using Heterogeneous Social Media As Auxiliary Information to Improve Car Recommendation Performance**”,*International Journal of Emerging Technology and Innovative Engineering*. Vol. 10, no.5, pp. 333-24, 2018.
- [6]. Heydari M. **Sentiment Analysis Challenges in Persian Language**. *arXiv preprint arXiv:vol.24, no. 23*, 2019.
- [7]. Godavarthi B, Nalajala P, Teja LR. **Wireless sensors based data acquisition system using smart mobile application. Internet of things**,”*International Journal of Advanced Trends in Computer Science and Engineering*. 5vol.1, pp. 25-9,2016.
- [8]. Li Q, Zhang Q, Si L. **Tweetsenti: Target-dependent tweet sentiment analysis**. *InThe World Wide Web Conference* vol.13, pp. 3569-3573, 2019.
- [9]. Abid, Fazeel, et al. "Sentiment analysis through recurrent variants latterly on convolutional neural network of



- Twitter."***Future Generation Computer Systems*vol.95, pp. 292-308, 2019.
- [10]. Ahmad, M., Aftab, S., Bashir, M.S., Hameed, N., Ali, I. and Nawaz, Z., 2018. **SVM Optimization for Sentiment Analysis.***Int. J. Adv. Comput. Sci. Appl*,vol.9, no. 4, pp.393-398, 2018.
  - [11]. Jianqiang Z, Xiaolin G, Xuejun Z. **Deep convolution neural networks for twitter sentiment analysis.** *IEEE Access.* vol.1, no. 6, pp.23253-60, 2018.
  - [12]. Lavanya K, Deisy C. **Twitter sentiment analysis using multi-class SVM.***In2017 International Conference on Intelligent Computing and Control (I2C2)* vol. 23, pp. 1-6, 2017.  
<https://doi.org/10.1109/I2C2.2017.8321798>
  - [13]. Dedhia C, Ramteke J. **Ensemble model for Twitter sentiment analysis.** *In2017 International Conference on Inventive Systems and Control (ICISC)*, vol. 19, pp. 1-5, 2017.
  - [14]. Yan Y, Yang H, Wang HM. **Two simple and effective ensemble classifiers for Twitter sentiment analysis.***In Computing Conference*, vol.18, pp. 1386-1393, 2017.
  - [15]. Ray, P. and Chakrabarti, A., February. **Twitter sentiment analysis for product review using lexicon method.** *International Conference on Data Management, Analytics and Innovation (ICDMAI)* vol.23, pp. 211-216, 2017.  
<https://doi.org/10.1109/ICDMAI.2017.8073512>