Detection of Extremist Protest Content on Twitter using Feature Selection Based Classifiers Techniques

Prabha Devi D & Iniyasri S
Assistant Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamanaglam

Abstract: Online users' behaviour and activities are indicated by the opinions expressed on sites. Extremist protest content detection is criticalfor the analysis of the sentiments of users regarding certaingroups and also to deter association with theseillegal actions. Protest content is identified from Twitter tweets in this work. Classification of the Twitter content is performed through feature extraction and their classification. The objective of the Feature Selection techniqueisfor theselection of a relevant and minimal feature subset from a given data set and maintenance of its original representation. Correlation-based Feature Selection (CFS) will assess the value of an attribute subset by taking into account each feature's predictive capability together with their redundancy degree. Numerous real-world complex problems are handled by the extensive utilization of Biogeography-Based Optimization (BBO). For binary classification problems,the most efficient techniques are Support Vector Machines (SVMs). Artificial Neural Networks (ANNs) are the latest and beneficial modelswhich are employedin machine learning and problem-solving. For detecting extremist protestcontent on Twitter, this work examines the proposed FeatureSelection based on BBO.

Keywords: Correlation-based Feature Selection (CFS), Biogeography-based optimization (BBO), Support Vector Machine (SVMs) and Artificial Neural Networks (ANN)

1. Introduction

More than 37% of the global population is active on social networks [1]. Control of publications is extremely arduous as the majority of users of social networkscommunicate in diverse languages and also in a language that is non-academic and commonplace. Twitter, Facebook, and other social media networking sites enable users to share posts, videos, photos, and views and also inform each other about activities related to the real-world or online. One of the most popular microblogging sites is Twitter,whichhas massive user interactions. Users of this

network will use Twitter messages (known as tweets) for the sharing of personal opinions on specific social events, views, technical knowledge,opinions, and ideas.

Numerous researchers have tried to study the online behaviour of protestersas they have perceived the dangers of radicalization and violence and how it is turninginto a criticalglobal threat tosocieties. Based on the review of existing literary works, several existing researchintegratestechniques to finddistinctive properties which can assist in automatically detecting such users [2]. The majority of these techniques utilize a textual analysis,which is based on keywords. When this analysis type is only applied, it can exhibit many disadvantages likebeing highly dependent on the examined data, and in the production of a massive number of false positives. Protest content detection is advantageouswith regards to the classification of the affiliationof protesters or activists throughthe filter of tweets before they are transmitted, or recommended.

Feature Selection (FS)is a dimensionality reduction techniquethatintends to downsize the feature space's dimensionality. Upon application of the FS technique, there will be a selection of the most informative features whilst the uninformative and least important features are eliminated on the basis of the premise that the classification quality is not necessarily affected by the removal of these features. On the other hand, in the feature space, all the features are weighed and ranked for the selection of the most informative features.

In 2008, Simon developed the Biogeography-BasedOptimization (BBO) [3]which wasbiogeography-influencedwith regards to the species'migration between diverse habitats, and also the species'evolution and extinction. When there is an assumption of an optimization problem with certain candidate solutions, each habitat denotes a candidate solution, the habitat's feasibilitydenotes the optimization problem's fitness, and the features of the habitat denote the decision variables. As per the theory of biogeography, through migration, a superior solution will likely share more favourable information with the inferior solution, especially in cases with low immigration and high emigration, and also the other way around. As per the biogeography evolution, a mutation can also happen with some probability.

There are numerous significant aspects and restrictions related to the processing of Social network analysis like twitter and content analysis. Also, content of the social movements yield [4], which must be considered during the development of machine learning techniques for resolving these problems. Analysis of text messages isdepicted as a combination of non-text and text attributes thatdescribe the message. As a typical case, the message text is an arbitrarily sized with oneessential attribute, which is, a timestamp of the message's publication or registration. The message can also have extra attributes, for example, its author and reader (orsender and receiver), thatis utilized for identifying the topology of a network community or a user group.

Messages can vary from being huge text documents to being short,comprising of some words. There is also the issue of hashtags and web links,as they generally can comprise the whole message.Therefore, for the representation of such messages content, the content related to the resources or the referred users must be downloaded and analysed. Also, extremist protester's resources can take advantage ofjargon or slang,which is only familiarwithin a narrow user circle, and also utilizeuniquesymbols or code wordswhich take the place of keywords commonlyutilized for keyword-based search.

A good classifier is capable of predicting intricate patterns'classes and hence, producea classification with good accuracy. Classification models such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and others will have better predictability as the accuracy becomes higher. Amongst the classification tools, the SVM algorithm is favoured due to its numerous benefits like its capability to manage huge feature spaces along with the good management of high dimensionality of the feature vectors and feature redundancy [17]. SVM has also been proven to be amongst the machine learning techniqueswith the best performance in numerous fields,inclusive of classification of texts. Even though SVM is an efficient binary classifier which has been employedas a text classifier invarious existing projects, it can also be utilized for problems which require multi-label classification.

The literature review has been done in Section 2. The various techniques utilized in the investigation have been explained in Section 3. The experimental results have been described in Section 4, and finally, the conclusions of the study have been made inSection 5.

2. Literature Survey

López-Sáncez et al. [6] devised a novel platformwhich could automatically detect and monitorthe process of radicalization on Twitter. The proposed platform has two distinct tasks: (1) To identify influential users having radicalization plans, and to notify the human supervisors of the relevant profiles; (2) To monitor the confirmed radical users' interaction and to assess the radicalization risk for vulnerable users who communicate with them. The work also gave a case study on the monitoring of "Hogar Social Madrid", anextremist far-right group which functions in Spain that exhaustively utilizes social networks. The proposed platform was able to detectnumerous profiles which were engaged in the radicalization process, thatcomplemented with young individuals who had recently embraced the radical concepts and messages "Hogar Social Madrid".

Protesters use social media during civil unrest for expressing opinions of the issue, organizing events. The tweets are used by the researchers for predicting the protest activity. Idya and Geetha [7]proposed a keyword-based technique for

identifying the correlation between the tweets endorsing protest and the forthcoming protest activities. A probabilistic model to classify protest activity is presented. Experiments were conducted using Twitter dataset from #Jallikattu, #BusFareHike and #SaveFisherMen. Results demonstrated the efficacy of the proposed method in classifying the protest activity.

Fernandez & Alani [8] developeda technique for constructingthe presentation of the semantic context of the terms which are connected with radicalized rhetoric. Over 114,000 tweets which contained terms of radicalization (about 97,000 tweets posted by "general" Twitter users and 17,000 tweets posted by pro-ISIS users) was analysed by using this technique. It is disclosedin what way the contextual information varies for similarterms of radicalization within the two data sets, that denotedhow contextual semantics can aidinthe better discrimination of radical content from content which only utilizes radical terminology. The classifiers constructed to gauge this hypothesis outperformed when compared to classifiers, which ignored contextual information.

Islam et al. [9] proposed an online framework to predict future protests using tweets. In the proposed framework, tweets are filtered and classified using SVM and weights are computed and distributed among the locations. The weights are updated and based on the overall score, the status of the protest in a location is predicted. The proposed method quantified sentiment using a new keyword dictionary with keyword score The results demonstrated that the proposed framework outperforms the existing framework.

A technique for the identification of users who participate in online extremist protest communications was proposed by Wei& Singh [10]. The proposed techniqueinitially utilizes detailed feature selection for the identification of relevant posts. Later, a novel weighted network is utilized to model the information flow between the relevant posts'publishers.

Wei& Singh [11]concentrated on features and metrics which researchers have recommended as surrogates for misbehaviour on Twitter. Initially, analysis is done on the probable features of a tiny amount of manually labeled data related to ISIS supporters on Twitter. Later, these features are categorized based on the dynamics, viewpoints, and content of tweets. There was a presentation of a case study that examined the ISIS extremist group based on the discussion of diverse advanced techniques for detecting extremism and related problems. The paper described the manner in which data was collected by the surveillance system.It concluded with discussions on certain existing problems and future directions for monitoring extremism more efficiently.

Nouh et al. [2] attempted to findmethodsfor the automatic detection of radical content in social media. For classifying radical messages, various behavioural,psychological,and textual signals were also identified. The proposed contribution is threefold: (1) analysis of the extremist group's published propaganda material and the creation of a contextual text-based model for the radical content; (2) construction of a model forthe psychological characteristicsunderstood from these materials; and (3) assessment of these models on Twitter in order to ascertain the extent of the possibility of automatic detection of online radical tweets. Experimental resultsdemonstrated that radical users displayeddistinct behavioural, psychological, and textual characteristics. Experimental results also demonstrated that the utilization of textual models with vector embedding features drastically enhanced the identification over TF-IDF features. High accuracy was accomplished by the two experiments performed using the proposed techniques. The proposedoutcomes can be employed as indicators for the detection of activities related to online radicalization.

Biogeography-BasedOptimization (BBO) is an evolutionary algorithm that was influenced by the species'migration between habitats. The original BBO research was published in 2008, almosta decade ago. The BBO has been successful in resolving optimization problems in diversefields and has also arrived at a comparablestate of maturity. Taking into account the critical and growing research on BBO and its applications, the time is entirely appropriate for providingthereview of the published literature'stenth anniversary, and also to specify certainsignificantroutes for future research. The related BBO research literature from the previous ten years was summarized and organized by Ma et al. [12]. This paper started with a basic BBO's foundation, then will survey the BBO algorithm's family and then describe themodifications of the BBO, its hybridizations, and applications in mathematical theory, and engineering and science. The paper concluded with certainfascinating open issues and the direction of BBO for future research.

3. Methodology

This section describes theTwitterdata set, Term Frequency-Inverse Document Frequency (TF-IDF) Feature Extraction, Correlation-based Feature Selection, Biogeographic Optimization (BO) Algorithm, and classifiers like Support Vector Machine (SVM) and Artificial Neural Network (ANN).

3.1 Dataset

On the social media platform Twitter, hundreds of millions of tweets are shared and sent daily. 18,000 normal messages and 2,300 negative messages have been considered in this work.

3.2 Term Frequency–Inverse Document Frequency (TF-IDF) Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a simplistic and efficienttechnique for feature extraction. Being an information retrieval method, TF-IDF is employed to ascertain the relevancy of the document's terms with reference to a query [13]. Two steps constitute the TF-IDF:(1) Initially, the term frequency (TF) is evaluated;(2) Later, the inverse document frequency (IDF) is evaluated. Both of these steps also have many variations.

TF-IDF runsthroughevaluating the relative frequency of words in a particular document in comparison to that word's inverse proportion over the wholedocument's body of the text. Thisevaluation will also instinctively determine a provided word's relevancy in a specific document. Common wordswithin documents that may be a small group or be single are likely to have high TFIDF numbers compared to general words like prepositions and articles. Equation (1) is used to depict this evaluation's mathematical formula as:

$$TFIDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (1)$$

Wherein,the terms are denoted by t; each document is denoted by d; the document collection is denoted by D.

3.3 Correlation-based Feature Selection (CFS)

A problem with numerous features is managed by the long prevalent technique of Feature Selection.The best feature subset is determined with CFS utilization. This technique is generallya combination with search strategies like genetic search, best-first search, bi-directional search,backward elimination, and forward selection.

Attributes are listed by the Correlation-based feature selection (CFS) as per a heuristic evaluation function,which is correlation-based. The function will assess subsets which constitute attribute vectors, thatcorrelatewith the class label but are not dependent on each other. The assumption of the CFS [14] techniqueis that as a low correlation with the class is demonstrated by features which are insignificant, the algorithm will disregard these features. Contrastingly, there will be an examination of the excess features since they generally have strong correlations with one or more of the other attributes. CFS is evaluated as per Equation (2):

$$r_{zc} = \frac{k\overline{r}_{zi}}{k + k(k-1)\overline{r}_{ii}} \quad (2)$$

wherein the correlation between the class variable and the summed feature subsets is denoted as r_{zc} the number of subset features is denoted as k, the average of the

correlation between the class variable and the subset features is denoted as r_{zi}, and the average inter-correlation between subset features is denoted as r_{ii}.

3.4 Proposed Biogeographic Optimization (BO)Algorithm

In 2008, Dan Simon devised a novel metaheuristic algorithm for global optimization known as Biogeography-BasedOptimization (BBO). The theory of biogeography is the study of the biological organisms' geographical distribution. BBO's fundamental concept is influenced by this theory. Migration and mutation are the two major mechanisms in the BBO algorithm. The inspiration for these mechanisms comes from the species'immigration and emigration between islands to find habitats which are friendlier [15]. Every solution is referred toas a "habitat" with a habitat suitability index (HSI). An –real dimension vector is used for the solutions' representation. Suitability index variables (SIVs) are referred to asthe individual's variables, whichdescribe thehabitatcapability. There is a random generation of the habitat vectors of an initial individual. Habitats with a high HSI are treated as good solutions,whereas, habitats with low HSI are treated asweaksolutions. The features of high HSIsolutions are shared withlow HSI solutions. Many novel features from the high HSI solutions are accepted by the low HSI solutions [3].A habitat in BBO is a vector (SIV) thatarrives at the optimal solution by following the steps of migration and mutation. Migration and mutation operators are utilized for the generation of a new candidate habitat from all solutions in the population. In BBO, the strategy of migration is akin to the strategy of evolution where a single offspring can be shared by numerous parents. The migration model is employedforchangingthe existing solution and for the existing island's modification. Migration is an operator which will adjust the probability of a habitat H_i. The modification of the probability H_i is in proportion to its immigration rate λ_i.The source of the modified probability comes from is in proportion to the emigration rate μ_j. Each individual in the BBO will have their own immigration rate λ and emigration rate μ. These are functions of the habitat's number of species and are evaluated as per Equation (3):

$$\lambda_k = I\left(1-\frac{k}{n}\right)$$
$$\mu_k = E\left(\frac{k}{n}\right) \quad (3)$$

wherein the maximum immigration rate is denoted by I, the maximum emigration rate is denoted by E, the habitat is denoted by Hi, the habitat's number of species is denoted by Si, and the maximum number of species is denoted by Smax. A better solution is indicatedby the habitat with more species in BBO. Since a better solution has a higher rate of emigration and a lower rate of immigration, there will be sharing

of favourable information between other solutions and is also less prone to be damaged by migration.

3.5 Classifiers

3.5.1 Support Vector Machine (SVM)

A classifier which will classify patterns into just two classes is referred to as the support vector machine (SVM). The SVM classifies data throughdetection of the best hyperplane,which will separate one class's data points from data points of another class. An SVM's best hyperplane refers to the class with the most significant margin between the two classes. Theslab's maximum width,which is parallel to the hyperplane,which has no interior data points, is termed the margin. The data points which are nearest to the separating hyperplane are termed the support vectors.These data points are on the slab's boundary.

The theory of statistical learning is the basis of SVM. The SVM will attempt to maximize the generalization property of the classifier model,which is generated by the algorithm. In SVM classification algorithm, a set of training instances are utilized, and new instances with two possible class labels, –1 and 1, are predicted. The hyperplane's definition is given by $w^T x+b=0$, wherein, $w \in R^n$ is orthogonal to the hyperplane and $b \in R^n$ is the constant. Providing certain training data D, a set of point of the form as per Equation (4):

$$D=\left\{(\vec{x}_i,\vec{y}_i)\,|\,\vec{x}_i \in R^m, \vec{y}_i \in \{-1,+1\}\right\}_{i=1}^{n} \tag{4}$$

whereinthe m-dimensional real vector is denoted by x_i, y_i denotes the class of input vector x_i, which is either –1 or +1. The objective of the SVM is to seek a hyperplanewhichwill maximize the margin between the two classes of samples in D with the least empirical risk [16].

$$y_i(\vec{w}^T\vec{x}+b) \geq 1 \tag{5}$$

The distance between these two hyperplanes will be maximized by SVM. $\frac{1}{\|\vec{w}\|}$ is used for evaluation of distance between these two hyperplanes. For the non-separable case, the SVM training is resolved utilizing the quadratic optimization problem,which is depicted as per Equation (6):

$$\begin{aligned} &\min imize:\ \mathrm{P}(\vec{\mathrm{w}},\mathrm{b},\xi)=\frac{1}{2}\|\vec{w}\|+C\sum_{i=1}^{n}\xi_i \\ &subjectto:\ \mathrm{y}(\vec{\mathrm{w}}.\phi(\vec{\mathrm{x}})+\mathrm{b}) \geq 1-\xi_i \quad ,\ \xi_i \geq 0 \end{aligned} \tag{6}$$

3.5.2 Artificial Neural Network (ANN)

Amachine learning (ML) model is the Artificial neural network (ANN). In terms of practicality, ANNs are comparably competitive to traditionalstatistical and regression models. For this classification, the ANN method was utilized [14]. Based on the system behaviour, the ANN is capable of understanding a very complex system. Due to their outstanding attributes of advancement in input to an output mapping,non-linearity, tolerance of faults, adaptivity, and self-learning, ANNs are commonly employedin numerical paradigms for approximation universal functions.

Artificial neurons are the ANN's fundamental building units. The inputs are weighted by the multiplication of each input witha weight at the artificial neuron's entry. A function is used to sum up,all the weighted inputs and bias at the internal neuron. On the other hand, a summation of the earlier weighted contributions and bias are transferred over an activation function at the output neuron. A K-element ANN's product is given as per Equation (7):

$$y(x) = \sum_{i=1}^{k} w_i y_i(x) \tag{7}$$

where, at layer i, y_i is the output and w_i is the weight.

ANNs are made up of an assortment of interconnected information processing units called neurons, which are generally organized into three layers: an output layer, one or many hidden layers, andan input layer. Through means of synaptic weights, the input signal is relayed to the initial hidden layer by the input layer neurons. A weighted summation of the inputs is evaluated by the hidden layer neurons. Then, these neurons will utilize an activation function to decide if the value should be relayed to the subsequent layer. As a result, the neural network's learning methodology will proceed via weight adjustments. The computation and adjustment of weights are usually performed using the backpropagation algorithm.

4. Results and Discussion

For evaluation of the techniques, 18,000 normal messages and 2,300 negative messages were considered in this work. TF-IDF was used for feature extraction, CFS and proposed BO for feature selection and SVM and ANN as classifiers. The results for classification accuracy, Hitrate,and Positive Predictive Value are shown in tables 1 to 3 and figures 1 to 3.

Table 1 Summary of Results

	CFS-SVM	CFS-ANN	BO-SVM	BO-ANN
Classification Accuracy	77.7	79.87	79.58	83.8
Hitrate for normal message	0.7642	0.7848	0.7821	0.8277
Hitrate for Hate message	0.8778	0.9078	0.903	0.9187
Positive Predictive Value for normal message	0.98	0.9852	0.9844	0.9876
Positive Predictive value for Hate Message	0.3223	0.3502	0.3462	0.4052

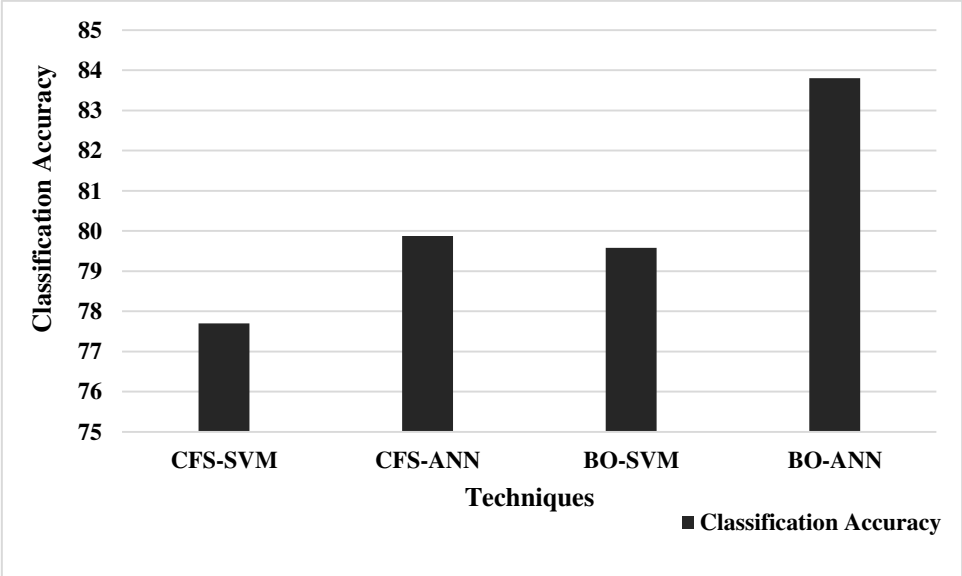

Figure 1 Classification Accuracy for BO-ANN

From Table 1 and Figure 1 shows that the Classification Accuracy for BO-ANN performs better by 7.6%, by 4.8%, and by 5.2% than CFS-SVM, CFS- ANN, and BO – SVM respectively.

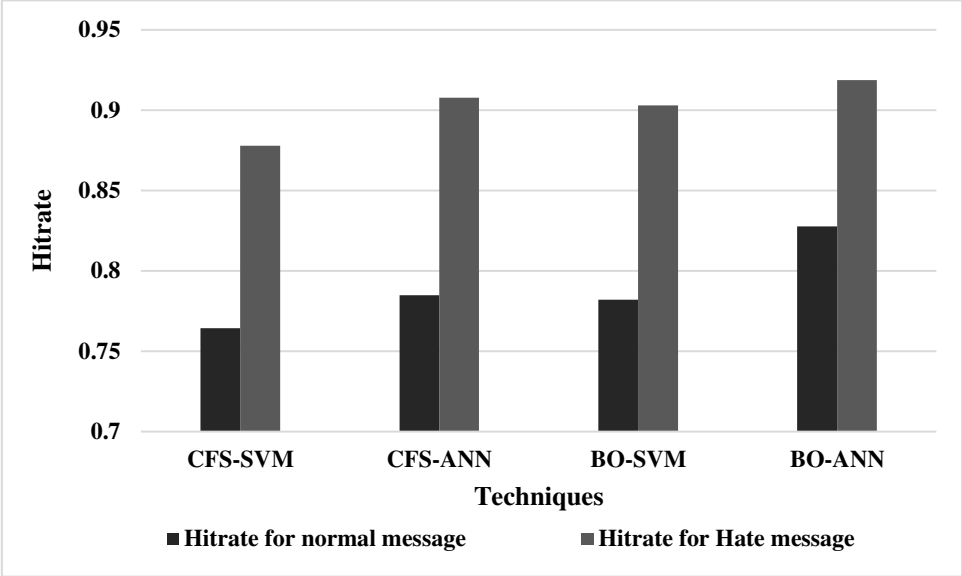

Figure 2 Hitrate for BO-ANN

From Table 1 and Figure 2 shows that the Hitratefor BO-ANN performs better by 7.98%, by 5.32%, and by 5.7% than CFS-SVM, CFS- ANN, and BO –SVM respectively for the normal message.The Hitratefor BO-ANN performs better by 4.6%, by 1.2% and by 1.7% than CFS-SVM, CFS- ANN, and BO –SVM respectively for Hate message.

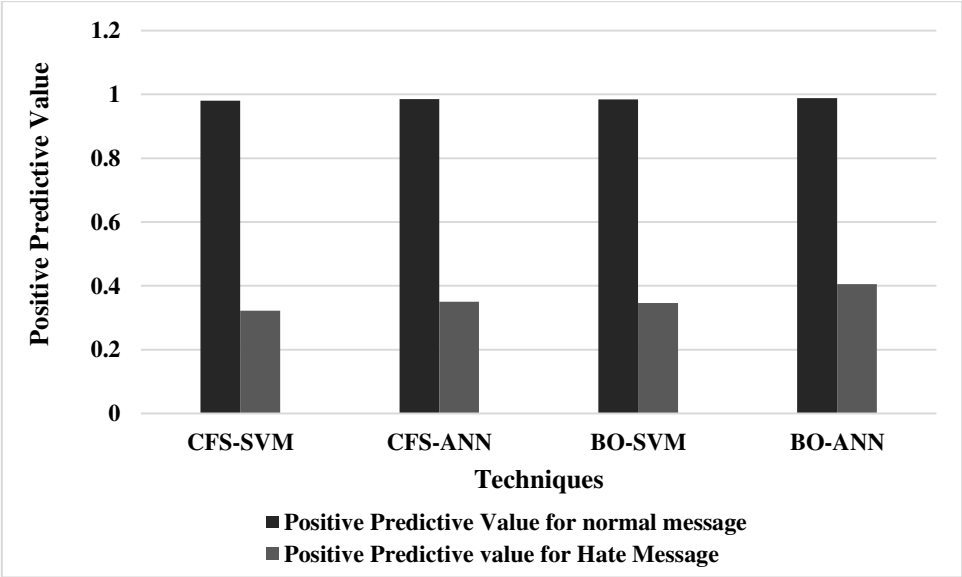

Figure 3 Positive Predictive Value for BO-ANN

From Table 1 and Figure 3 shows that the Positive Predictive Value for BO-ANN performs better by 0.8%, by 0.2% and by 0.3% than CFS-SVM, CFS- ANN, and BO – SVM respectively for the normal message. The Positive Predictive Value for BO-ANN performs better by 22.8%, by 14.6%, and by 15.7% than CFS-SVM, CFS- ANN, and BO – SVM respectively for Hate message.

Figure 4 to 6 shows the Features Selected, RMSE and RMSE versus Features Selected respectively.

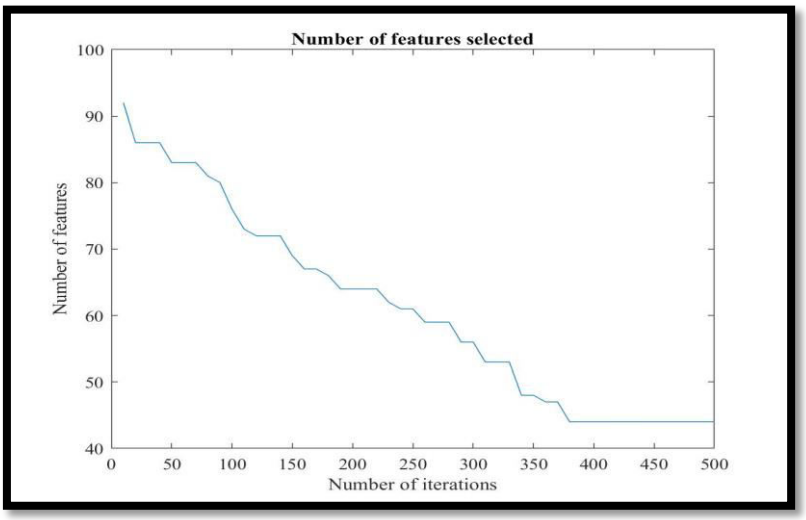

Figure 4 Features Selected

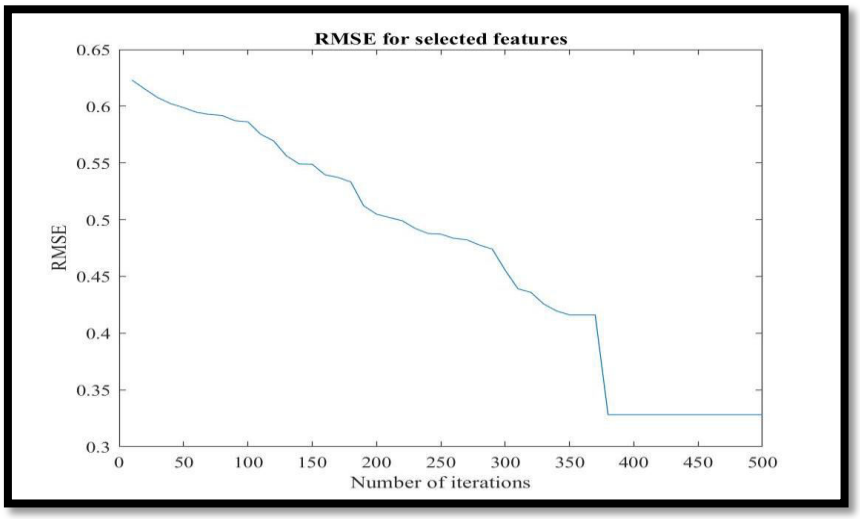

Figure 5 RMSE

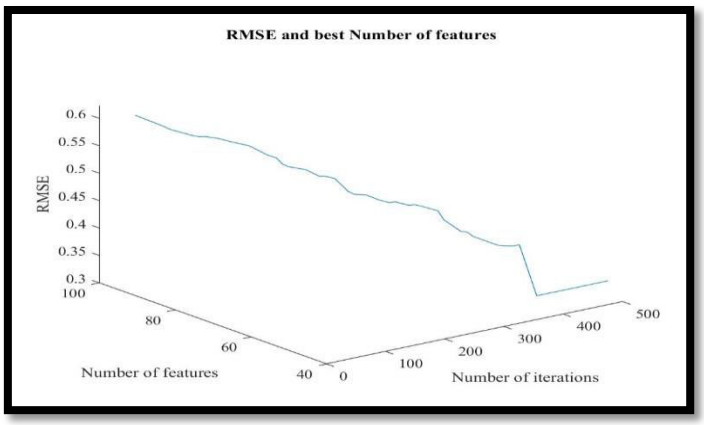

Figure 6 RMSE versus Features Selected

5. Conclusion

Similar to any other organizations and groups, extremist protest groups can be located on the Internet, inclusive of numerous groups with official websites. Extremist websites are examined to generate information on protesters relationships, activities, ideologies, communications, etc. For feature extraction's better accuracy, improvements are made on the TF-IDF method. As a unit of a habitat ecosystem, the BBO algorithm will generate a population of candidate gene subset. Then, for improvement of the classification accuracy, the BBO algorithm will apply the processes of migration and mutation over various population generations. Experimental results demonstrate that the Classification Accuracy for BO-ANN performs better by 7.6% compared to CFS-SVM, by 4.8% compared to CFS-ANN, and by 5.2% compared to BO-SVM.

References:

- Bedjou, K., Azouaou, F., & Aloui, A. (2019, July). Detection of terrorist threats on Twitter using SVM. In Proceedings of the 3rd International Conference on Future Networks and Distributed Systems (pp. 1-5).
- Nouh, M., Nurse, R. J., & Goldsmith, M. (2019, July). Understanding the radical mind: Identifying signals to detect extremist content on twitter. In 2019 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 98-103). IEEE.
- Cui, M., Li, L., & Shi, M. (2019). A Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization. Computational Intelligence and Neuroscience, 2019.
- Ogan, C., & Varol, O. (2017). What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during Gezi Park. Information, Communication & Society, 20(8), 1220-1238.
- Sabbah, T., Ayyash, M., & Ashraf, M. (2018). Hybrid support vector machine based feature selection method for text classification. Int. Arab J. Inf. Technol., 15(3A), 599-609.
- López-Sáncez, D., Revuelta, J., de la Prieta, F., & Corchado, J. M. (2018, August). Towards the automatic identification and monitoring of radicalization activities in twitter. In International Conference on Knowledge Management in Organizations (pp. 589-599). Springer, Cham.
- Iyda, J. J., & Geetha, P. (2020). Keyword-Based Approach for Detecting Civil Unrest Events from Social Media. In EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing (pp. 287-298). Springer, Cham.

- Fernandez, M., & Alani, H. (2018). Contextual semantics for radicalisation detection on Twitter.
- Islam, M. K., Ahmed, M. M., Zamli, K. Z., &Mehbub, S. (2020). An online framework for civil unrest prediction using tweet stream based on tweet weight and event diffusion. Journal of Information and Communication Technology, 19(1), 65-101.
- Wei, Y., & Singh, L. (2017, May). Using network flows to identify users sharing extremist content on social media. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 330-342). Springer, Cham.
- Wei, Y., & Singh, L. (2018). Detecting users who share extremist content on twitter. In Surveillance in Action (pp. 351-368). Springer, Cham.
- Ma, H., Simon, D., Siarry, P., Yang, Z., & Fei, M. (2017). Biogeography-based optimization: a 10-year review. IEEE Transactions on Emerging Topics in Computational Intelligence, 1(5), 391-407.
- Eklund, M. (2018). Comparing Feature Extraction Methods and Effects of Pre-Processing Methods for Multi-Label Classification of Textual Data.
- Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management, 2(2), 271-277.
- Li, X., & Yin, M. (2013). Multiobjective binary biogeography based optimization for feature selection using gene expression data. IEEE Transactions on NanoBioscience, 12(4), 343-353.
- He, Ruining, and Julian McAuley. "Ups and downs: Modeling the visual evolution of fashion trends with one class collaborative filtering." Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016.