International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

A Data Driven Model for Identifying Potentially Radical Content on Social Media Platforms

Krishna[1], Prof. Ashish Tiwari[2]

Abstract: With the emergence of social media as a common platform for communication among different people and communities at large, the chances for malicious usage of social media platforms for malicious activities has also increased manifold. One such malicious activity is spreading radical content over social media platforms due to the ease of sharing among several individuals and groups. The challenging aspect though for social media agencies or security agencies is the screening of humongous amounts of data to detect radical content. With no clear boundary to demarcate radical and non-radical content, the classification problem becomes challenging as the data size increases.[1] The proposed work presents an artificial intelligence based technique for detection of radical content. The proposed approach uses the concept of dictionary learning to train a Bayesian Regularized artificial neural network. The performance evaluation parameters are the number of iterations, absolute time and the accuracy. It has been shown that while the proposed system attains a classification accuracy of 97% compared to 89% of previous work.

Keywords:- Radical Content, Counter-terrorism, social networks, text analysis, Bayesian Regularization, accuracy.

I.Introduction

With the power of social media invading every domain of life, its use to spread hatred and radical content has also increased [1]. Neither the international anti-terrorist coalition led by the US nor the defense Ministry of Russia officially provides the timing of combating the threat of prohibited in the Russian Federation ISIS [2]-[4]. Such fact that this young organization is highly aggressive and behave provocatively in relation to other Islamist terrorist organizations, adds the significance of prohibited in the Russian Federation ISIS in the Middle East [5].

In order to provide maximum impact on the minds of the Islamic world, including tech-savvy westernized Muslim youth, prohibited in the Russian Federation ISIS has actively used all the possibilities of media technologies and PR [6]-[8]. Modern terrorist organizations weekly, and even daily put in the Internet space professionally shot and orchestrated videos of public executions, cultural monuments destruction, fighting, interviews with their commanders, the radical preachers and activists, photos of the trophies and killed enemies. “Media Jihad” has firmly won its place in the international information space and became one of the most important activities of terrorists. Financial capability of prohibited in the Russian Federation ISIS allows to actively finance various media projects and to promote them in the network, which ultimately leads to the influx of new fighters, and also help to consolidate the power over its territory.[9]-[14].

Terrorist and radical groups of people use instant messengers and accounts on social networks to publish propaganda messages. Blocking such accounts is one of the most effective methods of countering them [15]-[18]. To do this, analysts need to read and process a huge amount of information. In the Table I we can see the statistic of the messages frequency in different social networks and messengers [19].

It is hardly in today's world to find someone who has not heard about such prohibited in the Russian Federation organization as “the Islamic state of Iraq and the Levant” (ISIL or ISIS), also known as “Islamic state” (IS) [20].The rapidity of the organization’s development and its actions magnitude, unlike any other existing terrorist organization, using new devices and techniques for ideological and informational activity allow us to say that we deal with a new phenomenon in terrorism, which has managed to take on a qualitatively different level of development [21]-[23].

II. Need for Artificial Intelligence for Radical Content Detection

The amount of data that is churned up in social network sites is staggeringly large which makes it is impossible for humans to scrutinize the amount of data [25]-[27]. To render an insight into the data size and complexity, the frequency analysis of the different social media platforms is given below:

Table.1 Frequency Analysis of Different Social Media Applications [1]

Application	Per Second	Per Day	Per Month
WhatsApp	636 (thousand)	55 (billion)	1.6 (trillion)
Telegram	175 (thousand)	15 (billion)	450 (trillion)
Facebook	2.5 (thousand)	216 (billion)	6.5 (trillion)
Twitter	5.8 (thousand)	500 (billion)	15 (trillion)
Instagram	1 (thousand)	95 (billion)	2.8 (trillion)

Since the number of messages is enormously large and complex for analysis, hence it is necessary to use artificial intelligence based techniques for analysis of social media data. To practically implement such artificial intelligence based techniques, artificial neural networks are to be designed. Various machine learning methods are most often used for the analysis of texts with radical content. Such a method as the Named Entity Recognition (NER) allows to extract structured information from unstructured or semi-structured documents and is successfully applied to short text messages. Clustering, logistic regression and Dynamic Query Expansion (DQE) are more suitable to predict terrorist acts, riots or protests. The most often methods used to identify radicalism and extremism in real-time mode are K-Nearest Neighbour, Naive Bayes classifier, Support Vector Machine (SVM) with different kernel functions, decision trees and others. However, without a clear boundary among the different classes, there arises a need for a probabilistic classifier. The solution of the thematic analysis problem is complicated by several factors. Disseminated by terrorist groups information is heterogeneous, messages in social networks are quite short, contain slang and coded words, making semantic analysis useless. The difficulty is that the communication on the forums proceeds in different languages, and also, perhaps, in their combinations (the same goes for Internet documents) [28]. Also, a simple search based on keywords or specific phrases would not help to distinguish terrorist sites from such sites as news agencies. In addition, terrorist sites are often disguised as news sites and religious forums. The number of sites is huge, which makes their manual analysis inefficient, therefore, for correct identification of these sites and forums associated with terrorist groups there are required the automatic means for effective selection and filtering [29]. More difficult problem is to determine the identity of disseminated information to one of the terrorist groups, because terrorist groups may be ideologically close, and use similar vocabulary. Another method of operating prohibited in the Russian Federation ISIS in social networks is the promotion of specific “hashtags” (special tags using the # sign in order to organize posts by subject or group). Hundreds and thousands of activists repeatedly place the necessary messages with required “hashtags” at a specific time of day. Mass mailing Islamists of messages in social networks leads to dramatic results. So, experts give the following example: during the assault by militants of Mosul was published about 40 thousand tweets in support of prohibited in the Russian Federation ISIS. It was enough to display the top need hashtags (e.g., #ISIS, #AllEyesOnISIS, #Iraqwar) and pictures by manipulating the news agenda. The approach proposed in this paper consists of several consecutive stages. First, one must clear the text from “information noise”, such as links, emoticons, words without any meaning (articles, pronouns, conjunctions, etc.). The next step is fixing typos in the keywords using Levenshtein distance metrics [30]. Then the naive Bayes classifier is trained on the test data. The classifier will allocate two groups of messages: containing the radical content and not containing such content. However, the first group may include a message dedicated to the fight against terrorism and strongly condemns radical views. In order to separate such messages from advocating extremism, it is necessary to perform texts tone analysis.

III. System Design using Regression Learning based Bayesian Regularized ANN

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Other advantages include:

1. **Adaptive learning:** An ability to learn how to do tasks based on the data given for training or initial experience.
2. **Self-Organization:** An ANN can create its own organization or representation of the information it receives during learning time.
3. **Real Time Operation:** ANN computations may be carried out in parallel, and special hardware devices are being

designed and manufactured which take advantage of this capability.

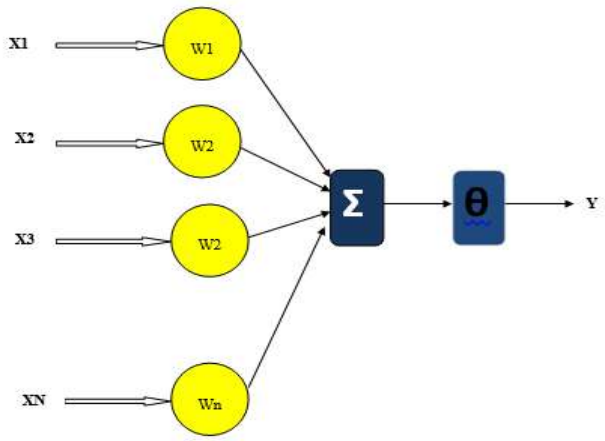

Fig.1 Mathematical Model of Neural Network

The output of the neural network is given by:

$$\sum_{i=1}^{n} XiWi + \theta \quad (1)$$

Where,
Xi represents the signals arriving through various paths,
Wi represents the weight corresponding to the various paths and
Θ is the bias. It can be seen that various signals traverssing different paths have been assigned names X and each path has been assigned a weight W. The signal traverssing a particular path gets multiplied by a corresponding weight W and finally the overall summation of the signals multiplied by the corresponding path weights reaches the neuron which reacts to it according to the bias Θ. Finally its the bias that decides the activation function that is responsiblefor the decision taken upon by the neuralnetwork. The activation function φ is used to decide upona the final output. The learning capability of the ANN structure is based on the temporal learning capability governed by the relation:

$$w(i) = f(i, e) \quad (2)$$

Here,
w (i) represents the instantaneous weights
i is the iteration
e is the prediction error

The weight changes dynamically and is given by:

$$W_k \overset{e,i}{\rightarrow} W_{k+1} \quad (3)$$

Here,
W_k is the weight of the current iteration.
W_{k+1} is the weight of the subsequent iteration.

(i) Regression Learning Model

Regression learning has found several applications in supervized learning algorithms where the regression analysis among dependednt and independent variables is eeded [31]. Different regression models differ based on the the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used. Regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a relationship between x (input) and y(output). Mathematically,

$$y = \theta_1 + \theta_2 x \quad (4)$$

Here,
x represenst the state vector of inut variables
y rperesenst the state vector of output variable or variables.
Θ1 and Θ2 are the co-efficents which try to fit the regression learning models output vector to the input vector.

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ1 and θ2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y). The cost function J is mathematically defined as:

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_1 - y_i)^2 \quad (5)$$

Here,
n is the number of samples
y is the target
pred is the actual output.

(ii) Gradient Descent in Regression Learning

To update θ1 and θ2 values in order to reduce Cost function (minimizing MSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ1 and θ2 values and then iteratively updating the values, reaching minimum cost. The main aim is to minimize the cost function J [31].

(iii) Bayesian Regularization

The Bayesian Regularization (BR) algorithm is a modified version of the LM weight updating rule with an additional advantage of usingthe Baye's theorem of conditional probability for a final classification [32].
The weight updating rule for the Bayesian Regularizationis given by:

$$w_{k+1} = w_k - (J_k J_k^T + \mu I)^{\wedge} - 1(J_k^T e_k) \quad (6)$$

Here,

w_{k+1} is weight of next iteration,

w_k is weight of present iteration

J_k is the Jacobian Matrix

J_k^T is Transpose of Jacobian Matrix

e_k is error of Present Iteration

μ is step size

I is an identity matrix.

The decision making approach of the Bayesian Classifier can be understood graphically using the graph theory approach. The approach for computing the probability among different disjoint sets can be understood using the set theory approach shown in the subsequent steps. The figures clearly depict the decision to be taken in cases of different overlapping data value categories.

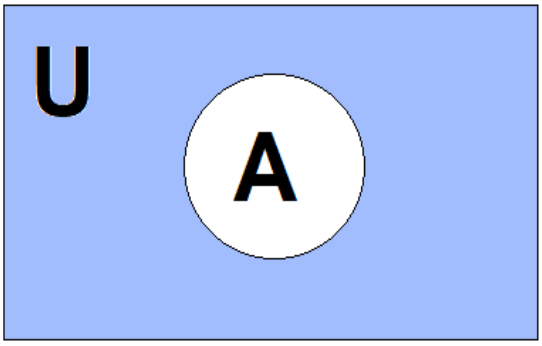

Fig.2 Universal Set Containing a Subset 'A'

Let us assume that the Bayesian Regularization algorithm needs to categorize the set A among multiple subsets in the superset U, for the time being in which only A exists exclusively.

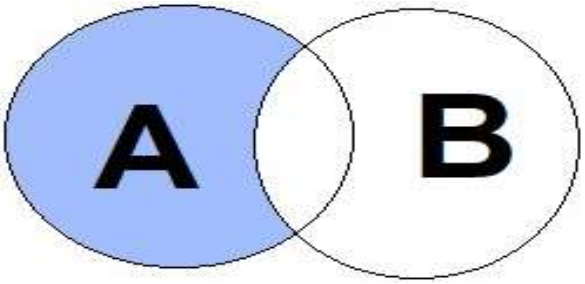

Fig.3 Probability of Exclusive Occurrence of 'A'

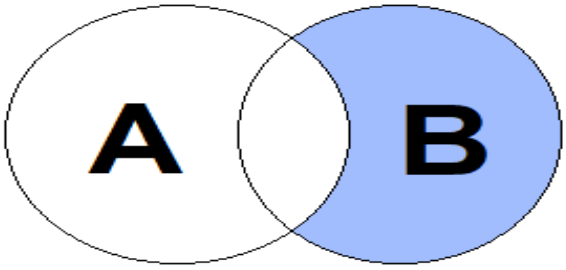

Fig.4 Probability of Exclusive Occurrence of 'B'

Figures 3 and 4 depict the probability of exclusive occurrence of events A and B respectively.

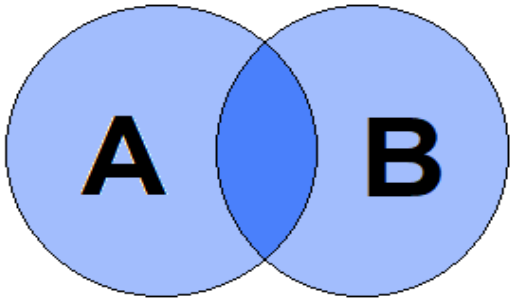

Fig.5 Probability of Union of A and B

Moreover for the predictive classification of ant data set, the Baye's Rule is followed, which is given by:

$$P\frac{A}{B} = \frac{P(A).P\frac{B}{A}}{P(B)} \quad (7)$$

Here,

$P\frac{A}{B}$ is the probability of occurrence of A given B is true.

$P\frac{B}{A}$ is the probability of occurrence of B givenA is true.

$P(B)$ is the probability of occurrence of B

$P(A)$ is the probability of occurrence of A

In the present case the, 70% of the data has been taken for training and 30% of the data has been taken for testing.

The conditional probability of the sentiment can be also seen as an overlapping event with the classification occurring with the class with maximum conditional probability. The mathematical formulation for the above mentioned probabilistic approach can be understood as follows:
Let there be 'N' classes of data sets available in the sample space 'U'.
Let the conditional probability of each of such sets be given by:

$$P(\frac{A}{U}),\ P(\frac{B}{U}),\ \ldots\ldots\ P(\frac{N}{U}). \quad (8)$$

The BR algorithm tries to find out the maximum among the probabilities:

$$P(\max) = \begin{matrix} P(\frac{A}{U}) \\ P(\frac{B}{U}) \\ \vdots \\ \vdots \\ \vdots \\ P(\frac{N}{U}) \end{matrix} \quad (9)$$

The maximum value of the probability decides the classification of a dataset into a particular category. Assuming that X attains the maximum in such a sample space:

$$P_{max} = X \quad (10)$$

Where,

$$P\left(\frac{X}{U}\right) = P\ \frac{X}{\prod_{i=1}^{i=n} U_i} \quad (11)$$

Here,
$\prod_{i=1}^{i=n} U_i$ represents the conditional probability cumulative for all possible data set classes in the sample space U
X is the maximum probability corresponding to a particular data set and n is the total number of classes of categorization.

IV. Evaluation Parameters

Since errors can be both negative and positive in polarity, therefore its immaterial to consider errors with signs which may lead to cancellation and hence inaccurate evaluation of errors. Therefore we consider mean square error and mean absolute percentage errors for evaluation. The system accuracy can be evaluated in terms of the mean square error which is mathematically defined as:

$$mse = \frac{1}{n}\sum_{I=1}^{N}(X - X')^2 \quad (12)$$

Here,
X is the predicted value and
X' is the actual value and n is the number of samples.

V. Results:

Dataset: To test algorithms for classification of text documents on the subject of radical content there was used data collected by scientists from the Artificial intelligence lab of the University of Arizona. These data represent information collected from various websites, forums, chats, blogs, social networking, etc. for designated terrorist organizations.

Data Normalization: Canonization (normalization) of the text is the process of bringing to a single format, convenient for further processing. When working with large amount of information, it is necessary to exclude from the document all non-informative parts of speech (prepositions, particles, conjunctions, etc.).


: 'A MESSAGE TO THE TRUTHFUL IN SYRIA - SHEIKH ABU MUHAMMED AL MAQDISI: http://t.co/73aFszajvr http://t
: SHEIKH FATIH AL JAWLANI "FOR THE PEOPLE OF INTEGRITY, SACRIFICE IS EASY" http://t.co/uqqzXGgVTz http://t.
: FIRST AUDIO MEETING WITH SHEIKH FATIH AL JAWLANI (HA): http://t.co/TgXT1Gd6w7 http://t.co/2sEBeisze6
: SHEIKH NASIR AL WUHAYSHI (HA), LEADER OF AQAP: "THE PROMISE OF VICTORY": http://t.co/3qg5dKIlwr http://
: AQAP: 'RESPONSE TO SHEIKH BAGHDADIS STATEMENT 'ALTHOUGH THE DISBELIEVERS DISLIKE IT.' http://t.co/2
JA'WAH SERIES BY A SOLDIER OF IN: Video Link :http://t.co/EPaPRlphSW http://t.co/4VUYsaaint
OH MURABIT! : http://t.co/huJJ9KGkG http://t.co/t9RxMtBVGK
: 'A COLLECTION OF THE WORDS OF THE U'LAMA REGARDING DAWLAH': http://t.co/AlbeyWNeDQ http://t.co/m
ew account after the previous one was suspended.@KhalidMaghrebi @seifulmasdul123 @CheerLeadUnited
: AQAP STATEMENT REGARDING THE BLESSED RAID IN FRANCE:http://t.co/qvErFO25Qj http://t.co/YtcnGMVjDI
ulmasdul123 @CheerLeadUnited @IbnNabih1
OUS ONE TAKEN DOWN AQAP-"THE FACES HAVE BEEN BRIGHTENED" -REGARDING THE BLESSED ATTACK IN FRAN
- SHEIKH ABU HASAN AL KUWAITI (HA) ADVICE OF NABI S.A.W TO THE MUJAHIDEEN: https://t.co/QcKBdeCdty ht
KUWAITI (HA): ADVICE OF NABI SAW TO THE MUJAHIDEEN: http://t.co/bJ24uNFNIT http://t.co/99nsWDNzxa
dia @Dawlat_Islam? Not translating it either.
g the new JN video? Will translate it if nobody else is @IbnNabih1 @KhalidMaghrebi_ @MuwMedia @Polder_Mu
ghrebi_ @MuwMedia @Polder_Mujahid So we will translate it then inshallah let us knw if anybody is available fo
aniyah @polder_mujaahid @Abulzzadeen2 @MuwMedia half way through, a few more days inshallah, been very
halidMaghrebi_ @IbnNabih1 @Polder_Mujahid Aslm, we have completed the translation with the exception of a
halidMaghrebi_ @IbnNabih1 @Polder_Mujahid reports and words of the Syrian people, will have that completed
halidMaghrebi_ @IbnNabih1 @Polder_Mujahid inshallah. Should we still finish it, or not?
- ENG Subs: The Path of Salvation :WATCH:http://t.co/1hiwtP3XDdhttp://t.co/jcviGUdedv http://t.co/SsDgGMvP
With English Subtitles: The Path of Salvation #Jihad: _ https://t.co/HGZPYIAsw0 http://t.co/5DUqXYFTwO

Fig.6 Sample Text data for analysis [1]

The figure renders a sample screenshot of the text data to be analysed from different social media applications for finding radical content.

Fig.7 Designed Neural Network and its training parameters

The following attributes of the designed neural network are:

1) The number of hidden layers has been limited to 2 in order to limit the complexity of the algorithm.
2) The performance evaluation parameter is the mean square error
3) The training algorithm is the Bayesian Regularization algorithm

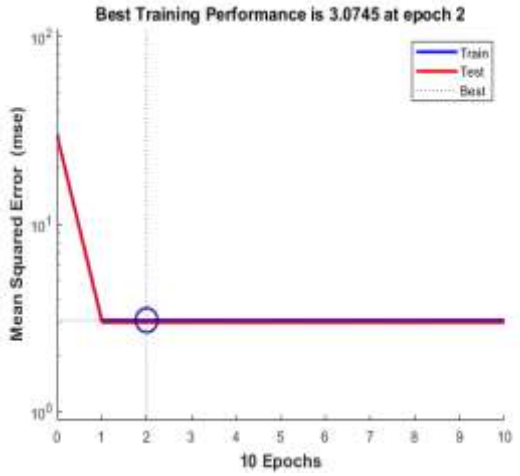

Fig.8 Training and epoch performance of the proposed system

The variation of the mean squared error as a function of the number of epochs is shown in the above figure. It can be seen that the MSE stabilizes at a value of 3.0745.

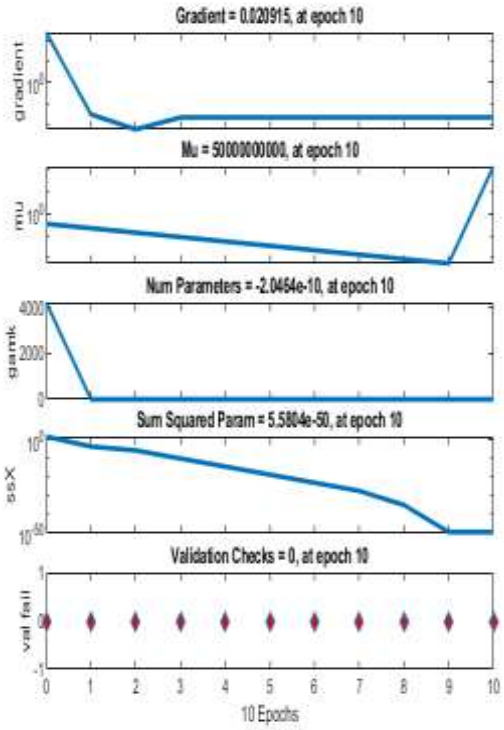

Fig.9 Training States as a function of number of epochs.

The variation in the training states such as the step size (mu), gradient, number of effective parameters and sum squared parameters has been shown in figure 9. The validation has also been shown. The gradient (g) and step size (μ) mathematically defined as:

$$g = \frac{\partial e}{\partial w} \quad \textbf{(13)}$$

Here,
e represents the error
w represents the weight

$$\mu = w_{k+1} - w_k \quad (14)$$

Here,
k represents the present iteration
k+1 represents the subsequent or next iteration

It can be clearly seen from figure 7 that the accuracy is 97% with an error metric of 3%. This is significantly higher than previously existing approaches. [1]. Here, the accuracy percentage is computed as:

$$Accuracy\ (\%) = [100 - errro]\% \quad (15)$$

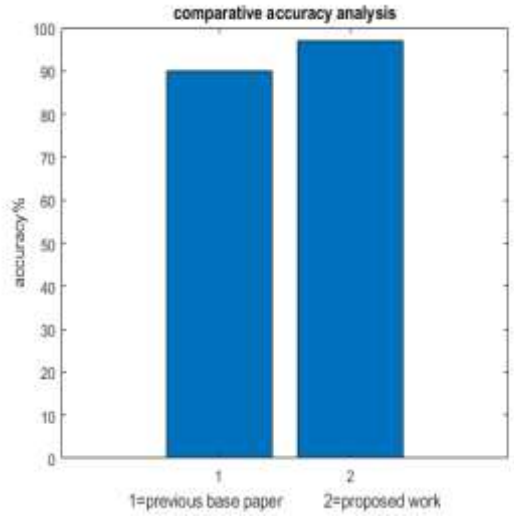

Fig.10 Comparative Accuracy Analysis w.r.t. previous work [1]

It can be clearly seen that the proposed work attains much higher accuracy of 97% compared to 89% of previous work. This can be attributed to the regression learning based BR trained ANN design which has a steep descent of error compared to the naïve Baye's classifier or the conventional Bayesian Regularization algorithm.

VI. Conclusion:

It can be concluded form the previous discussions that with the emergence of social media as a common platform for communication among different people and communities at large, the chances for malicious usage of social media platforms for malicious activities has also increased manifold. One such malicious activity is spreading radical content over social media platforms due to the ease of sharing among several individuals and groups. The challenging aspect though for social media agencies or security agencies is the screening of humongous amounts of data to detect radical content. With no clear boundary to demarcate radical and non-radical content, the classification problem becomes challenging as the data size increases. The proposed approach uses a Regression Learning based Bayesian Regularized ANN. It has been shown that the proposed work attains much higher accuracy of 97% compared to 89% of previous work

References

[1] Andrey I. Kapitanov, Ilona I. Kapitanova, Vladimir M. Troyanovskiy, Vladimir F. Shangin, Nikolay O. Krylikov, "Approach to Automatic Identification of Terrorist and Radical Content in Social Networks Message". IEEE 2023
[2] D López-Sáncez, J Revuelta, F de la Prieta, "Towards the Automatic Identification and Monitoring of Radicalization Activities in Twitter," IEEE 2023
[3] G Bobashev, M Sageman, AL Evans, "Turning Narrative Descriptions of Individual Behavior into Network Visualization and Analysis: Example of Terrorist Group Dynamics, IEEE 2022
[4] Z Li, D Sun, B Li, Z Li, A Li, "Terrorist group behavior prediction by wavelet transform-based pattern recognition", hindawi 2020.
[5] A Tundis, G Bhatia, A Jain, "Supporting the identification and the assessment of suspicious users on Twitter social media", IEEE 2019.
[6] M Al-Zewairi, G Naymat, "Spotting the Islamist Radical within: Religious Extremists Profiling in the United State", Elsevier 2018
[7] AH Johnston, GM Weiss, "Identifying sunni extremist propaganda with deep learning", IEEE 2017
[8] T Ishitaki, R Obukata, T Oda, "Application of deep recurrent neural networks for prediction of user behavior in tor networks", IEEE 2017
[9] I Lourentzou, A Morales, CX Zhai "Text-based geolocation prediction of social media users with neural networks", IEEE 2017
[10] R Lara-Cabrera, A Gonzalez-Pardo, "Extracting radicalisation behavioural patterns from social network data" IEEE 2017
[11] L Ball, "Automating social network analysis: A power tool for counter-terrorism", Springer 2016.
[12] T Ishitaki, T Oda, L Barolli, "A neural network based user identification for Tor networks: Data analysis using Friedman test", IEEE 2016

[13] T Oda, R Obukata, M Yamada, "A Neural Network Based User Identification for Tor Networks: Comparison Analysis of Different Activation Functions Using Friedman Test", IEEE 2016

[14] T Sabbah, A Selamat, MH Selamat, R Ibrahim, H Fujita, "Hybridized term-weighting method for dark web classification", Elsevier 2016

[15] R Scrivens, R Frank, "Sentiment-based Classification of Radical Text on the Web", IEEE 2016

[16] T Sabbah, A Selamat, "Hybridized Feature Set for Accurate Arabic Dark Web Pages Classification", Springer 2015

[17] T Ishitaki, T Oda, L Barolli, "Application of Neural Networks and Friedman Test for User Identification in Tor Networks" IEEE 2015

[18] R Frank, M Bouchard, G Davies, J Mei, "Spreading the message digitally: A look into extremist organizations' use of the internet", Springer 2015

[19] T Sabbah, A Selamat, "Hybridized Feature Set for Accurate Content Arabic Dark Web Pages Classification", researchgate, 2015

[20] Basant Agarwal ,Soujanya Poria,Namita Mittal,Alexander Gelbukh,Amir Hussain, "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach", Springer 2015

[21] B Akhgar, F Tabatabayi, PS Bayerl, "Investigating Radicalized Individual Profiles through Fuzzy Cognitive Maps", Elsevier 2014