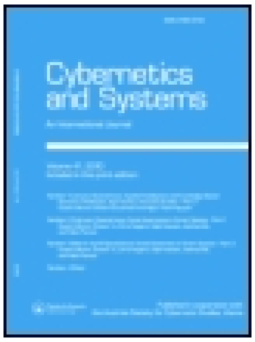

Taylor & Francis
Taylor & Francis Group

Cybernetics and Systems
An International Journal

ISSN: 0196-9722 (Print) 1087-6553 (Online) Journal homepage: http://www.tandfonline.com/loi/ucbs20

A Comparative Study and Automated Detection of Illegal Weapon Procurement over Dark Web

Jaspal Kaur Saini & Divya Bansal

CYBERNETICS AND SYSTEMS: AN INTERNATIONAL JOURNAL
https://doi.org/10.1080/01969722.2018.1553591

Taylor & Francis
Taylor & Francis Group

A Comparative Study and Automated Detection of Illegal Weapon Procurement over Dark Web

Jaspal Kaur Saini and Divya Bansal

Department of Computer Science and Engineering, PEC University of Technology, Chandigarh, India

ABSTRACT

Terrorist groups have reconnoitered smarter ways to use online discussion forums for their violent plans. They have been using their privately owned discussion forums for various illegal purposes. A comparative study of work done on various dark web forums of terroristic organizations is done in this paper. This paper proposes a novel approach to identify procurement of modern weapons over the social media forum by terrorist groups. We used data from four dark web forum websites named *"Ansar Aljihad Network", "IslamicAwakening", "Gawaher", and "IslamicNetwork"*. Multiple experts independently annotated 313 randomly selected posts as procurement (YES) or non- procurement (NO) to label the forum threads. Mutual agreement between experts is computed to find the level of significance. Furthermore, we used machine learning classification techniques (MLCT) in order to classify labeled posts. To our knowledge, our procurement model presents a first of its kind model to automatically detect procurement of modern weapons over dark web. The work done presents application of social media analytics and text mining to counter terrorism.

KEYWORDS

Dark web forums; machine learning; modern weapons; procurement model; violent extremists

Introduction

Social media analytics has become a powerful field for security and intelligence agencies to identify violent and extremist behavioral characteristics which can result in designing of effective techniques in order to counter threats of various kinds. Social media and online discussion forums have become potential tools for communication within extremist organizations. Dark web discussion forums are one such communication medium over the public infrastructure of Internet but require special configurations and customizations to gain access. These forums are discussion sites which include conversations from authenticated members of the forum. Monitoring and analyzing these dark web discussion forums of terroristic organizations can unfold useful insights to the security intelligence due to increased sophistication of technologies used by extremist organizations (Mandal and Ee Peng 2008).

CONTACT Jaspal Kaur Saini ✉ sainijassi87@gmail.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ucbs.

Pre-processing the large varying discussion forum data and applying automated machine learning methods can reduce the effort of the intelligence community to analyze the large volume of text data.

In this paper, we have developed an automated classifier to detect procurement of modern potential weapons such as drones using dark web discussion forums. Law enforcement officials use drone cameras to monitor the target of attack by terrorist organizations and then planning their massacre is quite traceable ("Obama's Drone War – The New Yorker" 2017). The extremist groups such as jihadist, etc. have been discussing and purchasing drones over the dark web. We found this a novel contribution to automatically detect procurement from thread discussions over several dark web forums of terrorist organizations. To purchase weapons like drones, terrorist organizations make use of computers and discussion forums which can be automatically detected by our proposed technique. To the best of our knowledge, such a study has not been performed till now and segues into a potential research area. Our work can help security intelligence to understand how violent extremists conduct online discussions to illegally procure weapons and potential targets in the near future by these modern weapons.

Our current work presents the use of text analytics and automatic detection of purchasing of drones by terrorist organizations over dark web. Specifically, we extract the messages communicated towards purchase of drones or discussions held regarding usage of drones for mass destruction or potential harm. All the possible threads which discuss massive destructive plans and mention usage or purchasing of drone have been considered and annotated by our experts. These messages give platform to both seller and buyer to interact with each other and settle the amount for the illegal procurement. Weapon procurement have thus been performed online with advent use of Internet and social media.

The structural layout of our paper is as follows: "*Background and Motivation*" section presents necessary background and motivation behind the work done. "*Literature Survey*" presents a comparative study of the related research work done in the field of mining patterns over various dark web forums. Problem is formally postulated in "*Problem Statement*" section. Detailed methodology is described in "*Research Methodology*" section. Machine learning classification models so developed and their corresponding experimental results have been described in "*Machine Learning Classifier and Results*" section. "*Conclusion and Future Work*" section concludes the work done.

Background and Motivation

Terrorist organizations use privately owned discussion forums to share their opinions and plans which make them more decentralized and flexible

Table 1. Four dark web discussion forum used in our study.

	Ansar 1	Gawaher	Islamic Awakening	Islamic Network
Period	12/8/2008– 1/20/2010	10/24/2004– 6/7/2012	4/28/2004– 5/22/2012	6/9/2004– 11/10/2010
Number of Posts	29,492	372,499	201,287	91,874
Number of Threads	11,244	53,235	32,879	13,995
Number of Members	382	926	3,964	2,082
Language	English	Arabic	Arabic	Arabic

in performing illegal acts. We have extracted data from Dark Web Portal Project (DWPP) (Chen 2013; McKerlich, Ives, and McGreal 2013). Under this project, data from 28 different social media discussion forums have been collected and processed primarily from Arabic sources. This dataset contains 14,297,961 messages, 1,553,122 threads originating from 362,495 authors. The discussion forums majorly belong to jihadist and general Islamic networks. We considered data from four discussion forums which are *"Ansar Aljihad Network", "IslamicAwakening", "Gawaher", and "IslamicNetwork"*. These four discussion forums are based are based on set of invitation only and are more popular with western jihadist and Islamic networks ("Dark Web and GeoPolitical Web Research | Artificial Intelligence Laboratory" 2016). The characteristics of four discussion forums used in our experiments are described in more detail below Table 1.

Each dataset consists of millions of postings written by thousands of forum members These posts are organized into threads which generally indicate the topic under discussion. Each post includes detailed metadata such as date, member name, etc. Each forum is provided as a downloadable compressed text file by University of Arizona ("Dark Web and GeoPolitical Web Research | Artificial Intelligence Laboratory" 2016) which can be used by researchers to perform study and computational analytics.

Literature Survey

We surveyed various research evidences which present findings of mining social data available from discussion forums of violent extremists. Social network analysis, content analysis, web metric analysis, and sentiment analysis are the major domains of computational analytics done on dark web forums ("Dark Web and GeoPolitical Web Research | Artificial Intelligence Laboratory" 2016). We did exhaustive study of work done on various dark web forums, a systematic representation of which has been tabulated in Table 2 with research findings.

It can be observed from Table 2 that multiple datasets have been considered for multilingual mining and topic modeling but no analysis has currently been done by integrating datasets from multiple discussion forums. No efforts, so far, have been made in developing any automated model on integrated

Table 2. Various DW forums research survey.

Dark web forum	Year of publication	Work done	Citation
Angelic Adolf, CCNU, Aryan Nation, Neo-Nazi, NSM, Smash Nazi, White knights, World Knights, Azzamy, Friends, Islamic union, Kataeb, Kataeb Qassam, Taybah, Osama Lover, Wa Islamah	2007	Measuring violent affects of multilingual messages over extremist group forums	(Abbasi 2007)
Al-Firdaws and Montada	2008	Automating sentiment polarity computation and affect analytics using machine learning technique named SVM and SVR	(Chen 2008)
Ansar 1	2010	Topic modeling using LDA and topic based key members extraction using SNA	(Huillier et al. 2010)
Alokab, Al-Boraq, Hanin Net, Attahadi Net, Ana Almuslim, Ansar Almujahideen, Al-Qimmah, Palestinian Islamic Jihad, Sheikh Hamid Bin Abdallah Al Ali, Abrar Way	2012	Topic detection from Arabic messages using Vector Space Model	(Alghamdi and Selamat 2012)
Al Ansar, Al Boraq, Al Faloja, Al Firdaws, Alokab, Alqimmah, Alsayra, Atahadi, M3f, Majahden, Medad, Montada, Muslm, Shamikh	2012	Identifying unknown forums using thirteen dark web portals by incorporating text features in sentiment classification	(Zimbra and Chen 2012)
Ansar 1	2014	Cyber recruitment detection using Naïve Bayes, logistic regression, SVM and Classification trees	(Scanlon and Gerber 2014)
Tor discussion forum	2015	Analysis of data by combination of character-level n-grams, stylometric features and timestamp features of the user posts	(Spitters et al. 2015)

datasets in the past. Furthermore, to our knowledge, no study on detecting procurement of illegal weapons by violent extremists over dark web has been done so far by researchers in the field which adds considerable novelty and applicability to reduce human efforts in analyzing large amounts of text to mine information about illegal weapon procurements. To the best of our knowledge, this is the first study of this kind which reports work on identifying illegal procurement of modern weapons within online communities over dark web. From our exploratory work, we found that application of computational analytics from Natural Language Processing to detect illegal procurement of weapons shows promising results and also highlights how drones are on hit list of modern weapon procurement for mass destruction.

Proposed Work

The main contributions of our work are as follows:

1. Highlighting the novel patterns over dark web discussion forums by extracting and combining data from multiple discussion forums.

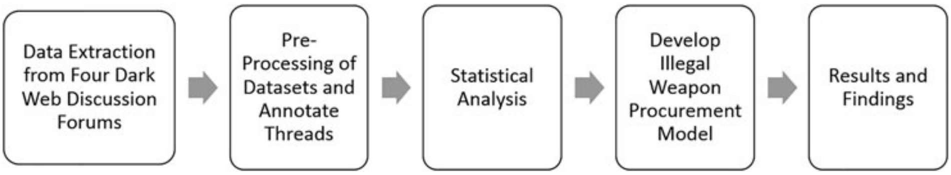
Data Extraction from Four Dark Web Discussion Forums
Pre-Processing of Datasets and Annotate Threads
Statistical Analysis
Develop Illegal Weapon Procurement Model
Results and Findings

Figure 1. Methodology followed to develop illegal weapon automated procurement model.

2. To use Machine Learning Classification Techniques (MLCT) to develop procurement models to identify illegal weapon procurement.
3. Validating proposed illegal weapon procurement model and comparing various machine learning classifiers used to train and test our model.

Research Approach

The dark web discussion forums present a big bunch of data which can be mined to find useful associations and draw hidden inferences using computational analytics. Figure 1 shows stepwise approach which is followed to develop techniques for detection of such illegal weapon procurement model.

Data Extraction from Four Dark Web Discussion Forums

To fulfill the research goals described above, we searched for datasets over dark web to model detection illegal weapon procurement. The data should come from infamous violent extremist groups. We identified the DWPP (Chen 2013; McKerlich, Ives, and McGreal 2013) as an appropriate dataset to evaluate our proposed technique of automatic detection of procurement of drones over dark web forums. The Dark Web Portal is warehouse of social data of 28 different forums or social networking sites. Most of the messages of violent extremists comes from Arabic sources but DWPP provides compilation and translation from seven discussion forums translated into English language. We considered the four most popular Jihadist and Islamic network based dark web forums named *"Ansar Aljihad Network", "IslamicAwakening", "Gawaher", and "IslamicNetwork"* in our research work. These datasets have been collected by University of Arizona under DWPP (McKerlich, Ives, and McGreal 2013) (Chen 2013). These datasets do not explicitly contain any content or posts which are related to illegal weapon procurement. So by developing an automated classifier, we detect such online illegal weapon procurement which might otherwise be hidden to us.

Pre-Processing of Datasets and Annotate Threads

All the posts of four dark web forums are extracted and recorded into initial corpus. We extracted forum posts from combined datasets and

Table 3. Sample annotations.

Sample[a]	Annotation
Russia in unprecedented Israeli drones purchase	YES
Germany to purchase Israeli killer drones (to be used by its troops in Afghanistan)	YES
CBS - U.S. Drones Have al Qaeda On the Run	YES
Air Force Report Envisions a Broader Use of Drones	YES
Houthis: Gov't air raids targeting civilian areas	NO
Amid US concerns Canada confirms Afghanistan pullout	NO
Egypt destroys 10 more Gaza's survival tunnels	NO

[a]Text sample shown is without any text pre-processing.

removed duplicates (same message ID) and empty documents (no message text) from the corpus. Most of the posts from four dark web forums were in English. However, if occasional posts were found to contain some words or slangs from Arabic, we left them intact considering that these words are supposed to be readable by an English Language Speaker. For example, "Kuffar" is a derogatory Arabic term for nonbeliever. Forum data which are extracted from four discussion forums do not provide any indication or intelligence about the content. We manually annotate the illegal weapon procurement posts within this combined data. Combined forum data from four forums were provided to two experts with following guidelines:

1. 313 posts extracted from four dark web discussion forums are being provided.
2. Each post be read carefully with intent to label discussions if it is about illegal modern weapon procurement or not. Illegal weapon procurement is defined as any discussion/message/text pertaining to purchase of modern weapon such as drones which can be used to monitor or plan massive attacks or cause damage to human lives, critical infrastructure or for surveillance purpose.
3. Annotate each post i) YES if contains any information related to drone procurement ii) NO if does not contain any such information related to procurement.

Sample of 313 posts labeled by two independent experts as either "YES" or "NO". Table 3 shows the sample of annotations from the population of annotated dataset.

Statistical Analysis

As the two experts have labeled the posts, so there is a need to compare number of agreement between two experts to validate consistency between annotations. Therefore, the Cohen's kappa coefficient (Cohen 1986) is used to check the agreement between experts which is calculated using following formula:

Table 4. Agreement matrix.

		Judge B	
	Category	Not Purchase	Purchase
Judge A	Not Purchase	0.4984	0.0287
	Purchase	0	0.4728
Kappa		0.9425	
Subjects		313	

$$k = \frac{p_a - p_c}{1 - p_c} \tag{1}$$

where p_a = proportion of observations for agreement of two experts; p_c = proportion of observations for agreement which is expected to happen by chance between two experts.

Agreement matrix of proportions for weapon purchase is shown in Table 4. Cohen' Kappa coefficient value was found to be 0.9425 at $\alpha = 0.05$ (α is probability of confidence interval for kappa statistics) which signifies an almost perfect agreement between the experts. R Programming Package "*psych*" is used to compute Cohen's kappa coefficient (Revelle 2016).

Considering significance and magnitude of kappa coefficient so computed, the annotations labeling represents the justification of process of manually labeling approach which can therefore be used in our analysis to train and test our proposed automated illegal weapon procurement model.

Development of Illegal Weapon Procurement Model

After annotating the posts and finding acceptable level of mutual agreement, we proceed to detect illegal weapon procurement automatically. We employed bag- of-words technique to build term document matrix using text mining packages of R Programming which are explained in detail in next section. Further, we used machine learning classifiers to train and test our illegal weapon procurement model.

Machine Learning Classifier and Results

To develop automated model, we have used six machine learning classification algorithms namely SVM, Boosting, Random Forest, GLMNET, Tree, and MAXENT. We have used RTextTools (Jurka et al., n.d.) package to implement and analyze the performance of our algorithm on annotated dataset. We developed a binary classifier which detects procurement of drones over four dark web discussion forum.

The probability model can be represented as:

$$\Pr[Procurement = Trueq_{\mathrm{i}}] = F[w_1(q_1), \ \ldots, \ w_n(q_n)] \qquad \ldots(2)$$

Table 5. MLCT description.

Algorithm	Description	R Package
SVM	Support vector machine, e algorithm is applied for estimating class probabilities using default package parameters and kernel method.	e1071 (Meyer and Dimitriadou 2015)
BOOSTING	R package's default settings are utilized with binomial log-likelihood as logistic loss computed as: $\sum_{i=1}^{n}\log(1+e^{-2Pro_iF(q_i)})$ where $Pro_i \in \{+1,-1\}$ is classification label and $F(q_i)$ is classification function as described in our procurement probability model in Equation (1).	caTools (Tuszynski2015)
RF	Random forest is averaging classifier which samples random features of dataset and "out of bag" error rate is monitored for each bootstrap sample in order to improve the performance of classifier.	randomForest (Breiman, Cutler and Liaw 2015)
GLMNET	Generalized linear model classifier is build using the following function: $\Pr\left[Pro_j = \pm 1 w_i(q_i)\right] = \frac{1}{1+e^{\left[-Pro_j\left(\beta_0+\sum_{k=1}^{n}\beta_k \cdot w_k(q_i)\right)\right]}}$ where $Pro_i \in \{+1,-1\}$ is procurement classification label and $\beta_0,\ldots,\beta_k$ are estimated parameters from training data.	LiblineaR (Helleputte and Gramme, n.d.)
TREE	Recursive partitioning with deviation criterion is used to select feature and build classification trees. Default package parameters are used to train our binary procurement classifier.	Tree (Brian and Ripley 2016)
MAXENT	Based on entropy provided by information theory and coding is used to estimate classifier label called maximum entropy estimate.	Maxent (Jurka 2012)

where Procurement is binary classification label, $q_i \epsilon Q$ is messages over forum and w_n is text feature function of q_i. Following are the steps undertaken to train and test our illegal weapon procurement model:

1. Create two Document Term Matrix (DTM) and Container: Various preprocessing operations are utilized from tm package (Feinerer, n.d.) of R Programming. Whitespaces, numbers and punctuations have been removed. Two DTM are generated with minimum Word Length (WL) which equals to three (WL $=$ 3 for DTM$_1$) and four (WL$=$ 4 for DTM$_2$). Our binary classifier, the illegal weapon procurement model is then developed using both DTMs to compare the performance. To implement machine learning algorithm, we need to divide the term matrix into training and testing set.
2. Train and test models: Six classification algorithms (SVM, BOOSTING, RF, GLMNET, TREE, and MAXENT) are trained and tested using different R packages to develop our binary procurement classifier (Meyer and Dimitriadou 2015; Tuszynski 2015; Hothorn et al. 2015; Breiman, Cutler, and Liaw 2015; Helleputte and Gramme, n.d.; Brian and Ripley 2016; Jurka 2012). Six classification algorithms stated below in Table 5 are trained and tested using different R packages to develop our binary procurement classifier.

Table 6. Procurement model evaluation.

	Predicted classification		
Referenced classification	$\Pr[Pro_j q_i] \geq \theta$	$\Pr[Pro_j q_i] < \theta$	$Recall\ (R)\ or\ True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN}$
$Procurement = True$	TP	FN	$Precision\ (P) = \frac{TP}{TP+FP}$
$Procurement = False$	FP	TN	$Fscore = 2 \times \frac{P \times R}{P+R}$

3. Performance Analytics: To evaluate the procurement classifier, we used performance metrics namely Recall, Precision, and F-score. The classifier algorithms are trained and tested for both term document matrix: DTM1 and DTM2 generated in step 1. Classification algorithm performance measures are described in Table 6 below with the help of confusion matrix. Table 7 shows the performance measures for all the six algorithms for both DTMs.

   It is clear from Table 7 that Random Forest (RF) classifier results into best classifier based on the Recall, Precision, and *F* score value computed though SVM, GLMNET, MAXENT are also comparable to RF with marginal difference.

   Furthermore, Classification Methods are evaluated and compared using ROC curves, which shows tradeoffs between False Positive Rate (FPR) and True Positive Rate (TPR) at different classification thresholds θ as defined in Equations (3) and (4).

$$TPR\ (\theta) = \frac{TP}{TP + FN} \qquad \text{...(3)}$$

$$FPR\ (\theta) = \frac{FP}{FP + TN} \qquad \text{...(4)}$$

4. Cross validation: It is required to execute N fold cross validation for the given dataset in order to find best estimates of accuracy for illegal weapon procurement model. We executed 10-fold cross validation and compare six illegal weapon procurement classifiers using our combined annotated dataset. Figure 2 shows mean ROC curve averaged over tenfold cross validation experiment. We have also employed area under the ROC curve (AUC) to compare the performance of each method along entire curve using single measure.

It can be seen that RF classifier performs best with an AUC of 0.82 which is more than SVM (AUC = 0.81), GLMNET (AUC = 0.81), and MAXENT (AUC = 0.81).

Therefore, it can be evidently inferred that for our illegal weapon procurement model RF, SVM, GLMNET, and MAXENT classifiers performs with reasonable accuracy.

Table 7. Classification algorithm performance measures.

Algorithm	Recall (WL = 3)	Recall (WL = 4)	Precision (WL = 3)	Precision (WL = 4)	*F* score (WL = 3)	*F* score (WL = 4)
SVM	81	80	83	81	80.5	80.5
BOOSTING	77	74.5	78	76	77	74.5
RF	82.5	80	86	85	82	80
GLMNET	81	76	85.5	77	80.5	75.5
TREE	73.5	78.5	75.5	82	73	78.5
MAXENT	81.5	80	82	82.5	81.5	79.5

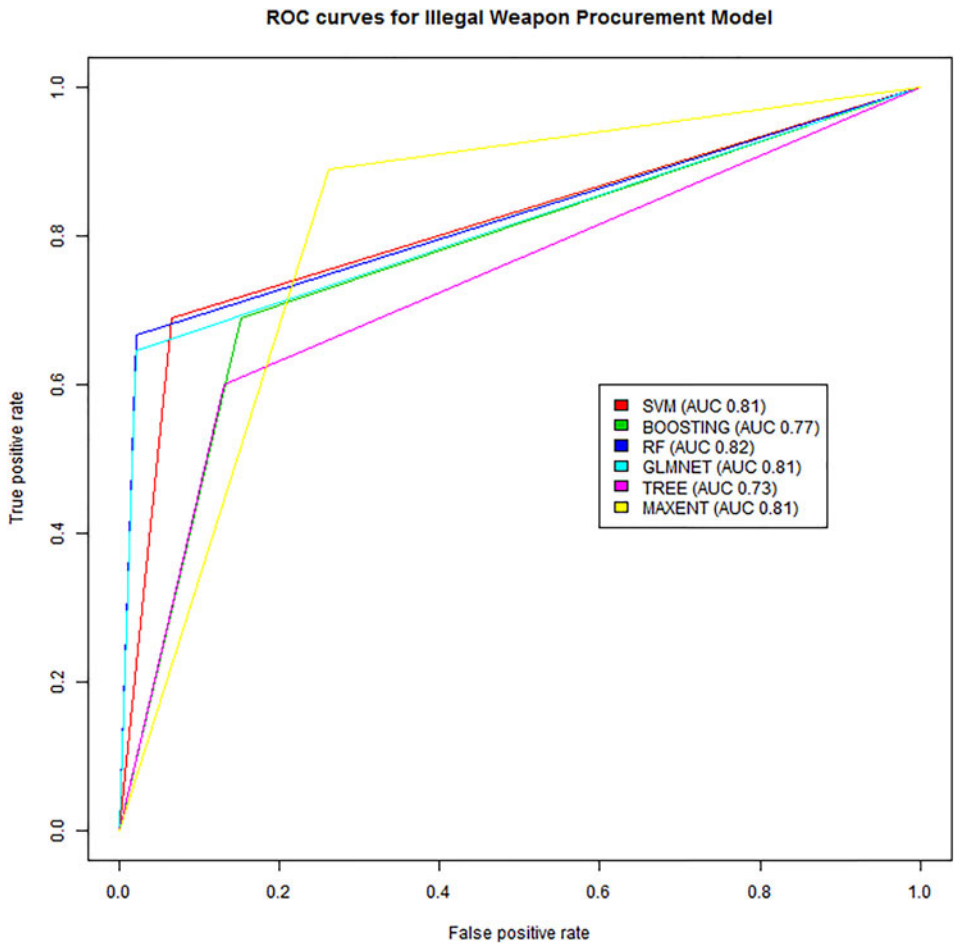

Figure 2. Comparison of illegal weapon procurement classifiers using ROC curves. Curves are averaged over 10 cross validation fold.

Conclusion and Future Work

Online activities of terrorist groups over the social media have increased exponentially in recent years which have led to increasing demand of development of automated detection models for illegal activities. Our research work is highly motivated by increased online activities of terrorist groups along with lack of automated detection models to analyze such online activities. The results are based on extraction of datasets from dark web forum by University of Arizona (Chen 2013; McKerlich, Ives, and McGreal 2013). The results show that it is possible to create automated models which can be used to detect illegal weapon procurement of modern destructive equipment such as drones which have potential to be used to record or plan massive destructive activities. After detecting procurement of such weapons, we can further trace the key individuals involved in selling and buying such equipment who are posing out to be a big threat to nation. SVM, GLMNET, and BOOSTING classifiers perform better over others for illegal weapon procurement detection. Procurement of other modern weapons through social networks may also be included and labeled to further extend this work. We have used text analytics to preprocess the messages extracted from dark web and trained the illegal weapon procurement model. As this is the first reported work on this kind of detection of procurement of drones over dark web, the results can serve as benchmark for further directions and making improvements.

In future, our illegal weapon procurement model can be enhanced to support different languages such as Arabic, etc. which are widely used by violent extremists over the dark web forums. As a future path, Social Network Analysis, Topic Modelling in order to find more hidden patterns over the dark web and detect more illegal activities. We hope that these smart solutions with effective use of computational analytics will help to counter online terroristic activities.

Acknowledgments

This work was partially supported by Cyber Security Research Centre, Punjab Engineering College (Deemed to be University), Chandigarh, India. Authors are grateful to Visvesvaraya PhD scheme for Electronics and IT for funding this research.

References

Abbasi, A. 2007. Affect intensity analysis of dark web forums. In *IEEE Intelligence and Security Informatics*, 282–88. doi:10.1109/ISI.2007.379486.

Alghamdi, H. M., and A. Selamat. 2012. Topic detections in Arabic dark websites using improved vector space model.” In *Conference on data mining and optimization*, no. September:6–12. doi:10.1109/DMO.2012.6329790.

Breiman, L., A. Cutler, and A. Liaw. 2015. randomForest: Breiman and Cutler’s random forests for classification and regression R package version 4.6-12. https://cran.r-project.org/web/packages/randomForest/index.html. https://www.stat.berkeley.edu/~breiman/RandomForests/.

Brian, A., and M. B. Ripley. 2016. Tree: Classification and regression trees R package version 1.0-37. https://cran.r-Project.org/web/packages/tree/index.html.

Chen, H. 2008. Sentiment and affect analysis of dark web forums: Measuring radicalization on the internet. *IEEE International Conference on Intelligence and Security Informatics, 2008, IEEE ISI 2008*, 104–9. doi:10.1109/ISI.2008.4565038.

Chen, H. 2013. Uncovering the DarkWeb: A case study of Jihad on theWeb. *International Review of Research in Open and Distance Learning* 14 (4):90–103. doi:10.1002/asi.

Cohen, J. 1986. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1):37–46. doi:10.1177/001316446002000104.

“Dark Web and GeoPolitical Web Research | Artificial Intelligence Laboratory.” 2016. https://ai.arizona.edu/research/dark-web-geo-web (accessed 21 December).

Feinerer, I. n.d. “Introduction to the Tm Package Text Mining in R.” https://cran.r-Project.org/web/packages/tm/index.html.” doi:10.1201/9781420068740.

Helleputte, T., and P. Gramme. 2015. Package LiblineaR: Linear predictive models based on the Liblinear C/C++ library. R Package Version 1.94-2. http://CRAN.R-Project.org/web/packages/LiblineaR

Hothorn, T., B. D. Ripley, T. Therneau, and B. Atkinson. 2015. Ipred: Improved predictive models by indirect classification and bagging for classification, regression and survival problems as well as resampling based estimators of prediction error. R package version 0.9-5. https://cran.r-Project.org/web/packages/ipred/.

Huillier, G. L., G. L. Huillier, H. Alvarez, F. Aguilera, and F. Aguilera. 2010. Topic-based social network analysis for virtual communities of interests in the dark web Topic-Based social network analysis for virtual communities interests in the dark web. In *ACM SIGKDD Workshop on Intelligence and Security Informatics*, p 9, 66–73.

Jurka, T. P. 2012. Maxent : An R package for low-memory multinomial logistic regression with support for semi-automated text classification. *The R Journal* 4 (1):56–9.

Jurka, T. P., L. Collingwood, E. B. Amber, and E. Grossman. n.d. Rtexttools: A supervised learning package for text classification 2014. https://cran.r-project.org/web/packages/RTextTools/index.html.

Mandal, S., and L. Ee Peng. 2008. Second life: Limits of creativity or cyber threat? *IEEE International Conference on Technologies for Homeland Security, HST’08*, 498–503. doi:10.1109/THS.2008.4534503.

McKerlich, R., C. Ives, and R. McGreal. 2013. A focused crawler for DarkWeb forums. *International Review of Research in Open and Distance Learning* 14 (4):90–103. doi:10.1002/asi.

Meyer, D., and E. Dimitriadou. 2015. e1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TUWien R package version 1. 6–7. https://cran.r-project.org/web/packages/e1071/index.html.

“Obama’s Drone War – The New Yorker.” 2017. http://www.newyorker.com/magazine/2014/11/24/unblinking-stare (accessed 25 January 2017).

Revelle, W. 2016. R package psych. *October*, 1–383.

Scanlon, J. R., and M. S. Gerber. 2014. Automatic detection of cyber-recruitment by violent extremists. *Security Informatics* 3 (1):1–10. doi:10.1186/s13388-014-0005-5.

Spitters, M., F. Klaver, G. Koot, and M. V Staalduinen. 2015. Authorship analysis on dark marketplace forums. *IEEE European Intelligence & Security Informatics Conference (EISIC)*. doi:10.1109/EISIC.2015.47.

Tuszynski, J. 2015. *caTools: ROC AUC tools, MovingWindow statistics (2015). R package version 1.17.1.* http://CRAN.R-project.org/package=caTools.

Zimbra, D., and H. Chen. 2012. Scalable sentiment classification across multiple dark web forums.” *ISI 2012–2012 IEEE International Conference on Intelligence and Security Informatics: Cyberspace, Border, and Immigration Securities*, 78–83. doi:10.1109/ISI.2012.6284095.