Improving the Process of Identifying Internally Displaced Persons Using Big Data Technologies

HakimaFathi Mahamoud, Raja Rajeswari Ponnusamy, Ho Ming Kang, Jacob Sow Tian You

Abstract: ***This data-driven project is systematically contributing on enhancing the conflict-violence or disaster-related displacement within an internationally recognized state border, namely internal displacement. With the availability of a training set with pre-defined categories, the project tackles document classification and information retrieval applications through supervised machine learning. This research can be divided into three core objectives. Firstly to eradicate non-relevant documents by filtrating documents not in English and not providing information on human mobility related to internal displacement. Secondly,*** *to tag* ***documents relatively to the themes Internal Displacement Monitoring Centre (IDMC) used to monitor the causes behind internal displacement, notably conflict/violence or disasters. Thirdly,*** *to extract* ***vital displacement information reported in online sources, such as location, displacement figures, etc. Documents are further analysed by training them using Support Vector Machine for tagging and Multinomial Naïve Bayes for information extraction, added to the pre-processing operations such as mainly working on natural language processing annotators, since the training set is mainly composed of textual documents. Finally, after having adjusted the parameters and learning, the performance of each of the resulting functions, notably Support Vector Machine and Multinomial Naïve Bayes on the training set, were measured on two different test sets, one for tagging and the other for information retrieval. By evaluating the provided dataset, the results were good with a result of 95.83% for classification and 81% for information retrieval.***

Keywords: document classification, information retrieval, Support Vector Machine, Multinomial Naïve Bayes

I. INTRODUCTION

Humans' movement from one locality to another has long back existed over time. While some of these migrations happened voluntarily, a significant percentage of migrations happened forcibly or beyond the will of oneself due to conflicts, natural disasters, famine, development projects and many more. In this paper, the internal displaced persons (IDPs) are the subject of the research.

Revised Manuscript Received on December 08, 2018.

HakimaFathiMahamoud School of Computing, Asia Pacific University of Technology and Innovation, Malaysia

Raja RajeswariPonnusamy School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia

Ho Ming Kang School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia

Jacob Sow Tian You School of Media, Art and Design, Asia Pacific University of Technology and Innovation, Malaysia

According to General Assembly (2016), IDPs are defined as persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border. Internal displacement of these IDPs has been one of the major concerns and issues by both governments and non-government organizations around the world. However, most of the time they will only be notified when an obvious humanitarian crisis is present, which is reported by those reliable sources such as media press or international observation teams. As the technology advances to a new height, data and posts on social media may be another source of information to be considered for such a monitoring action. Gigantic volume of data (both relevant ad irrelevant) and possibly with certain degree of biasness is available in the virtual world 24/7. In order to filter and process, and eventually extract useful insight out of these data quickly, support vector machine (SVM) learning and multinomial Naïve Bayes (MNB) have been proposed in this research. In this research, the aim is to develop a tool that embraces features to ease the process of identifying key information related to internal displaced people, by extracting and analysing essential facts from any online sources. This tool is then tested with the data provided IDMC. The remaining paper consists of literature review as found in Section 2.0, system implementation in Section 3.0 followed by the system evaluation. The paper ends with a conclusion and future work in Section 5.0 and a list of references in Section 6.0.

II. LITERATUREREVIEW

Based on several studies the common causes of IDPs include armed conflicts, human right violation or situations of generalized violations, sudden or slow disasters (Asplet, 2013; IDMC, 2017). In the context of IDPs, the Internal Displacement Monitoring Centre (IDMC) was established in year 1998 with the purpose of monitoring the IDPs worldwide and for the purpose of contributing on building the capability for national responses on internal displacement. It reported that conflicts in under-developed countries particularly, were commonly more complicated and drive long-term crisis due to the little interest or capability of the governments in concern to cope with them (IDMC, 2017). Besides, it also highlighted some conflicts that cause displacements might be left undetected because of their main common drivers go unaddressed and this led to resurfacing cyclically, when new violence and displacements erupt. In terms of natural and human made disasters, sudden-onset disasters such as earthquakes and volcanism phenomena, slow-onset disasters such as the insecurity of food and shelter, erosion and loss of inhabitant livelihood are some of the factors reported by Asplet (2013). Displacement does happen due to development investments according to IDMC (2017). Even though there are international standards, the procedure in guiding people who are displaced and resettling them, many of the displaced people are still left behind, in search for circumstance. Secondly, the collection on data concerning the IDP population is a challenging issue because it involved data privacy and sometimes the security of a nation. Thirdly, the availability of resource on IDP's is usually limited even if it is available due to some incidents or factors such as isolated locations, instable security circumstances and risk of contagious illness. This has somehow hindered necessary protection and assistance to be offered to these IDPs.In this research work, with the help of big data technology and the furtherance of computing technology in machine learning, it is a hope that challenges in gathering information of IDPs as mentioned by Baal and Ronkainen (2017), and European Commission (2017), and subsequently the lack of awareness of IDPs and necessary protection and responsibility towards this group of people by both governing bodies and societies, could be minimized and overcome.

A. System Implementation

In this research, a system has been developed in Python language. The datasets have been provided by IDMC, two sets of training data and two sets of testing data. These datasets are in the comma-separated values (csv) file format. In the training set, information that describes internal displaced people, such as the displacement figures, reporting term, reporting unit, and location are pre-defined (Figure 1). Two sets of testing sets are used to test and evaluate the effectiveness of the system configured in tagging online documents and extracting terms in natural language (in this case English) from online documents respectively (Figure 2 and Figure 3).

	B	C	D	E	F	G	H
1	URL	Tag	Excerpt	Displacement figure	Reporting term	Reporting unit	Specific location
2	http://www.securitycouncilreport.	Conflict and violence	United Nations agencies, funds and programmes	82000	Displaced	People	southern areas of Aby
3	http://www.independent.co.uk/ne	Disasters	Floods left a village devastated in the remote ea	60	Destroyed Housing	Households	Afghan province of N
4	http://reliefweb.int/sites/reliefwe	Disasters	Heavy rainfall on 3 February 2013 caused localiz	299	Sheltered	Households	18 districts across the
5	http://floodlist.com/asia/afghanis	Disasters	Quoting an official from the Badakhshan provinc	120	Destroyed Housing	Households	Xinhua
6	http://reliefweb.int/sites/reliefwe	Disasters	Fifteen people were injured, 91 homes damaged	6	Destroyed Housing	Households	
7	http://erccportal.jrc.ec.europa.eu/	Disasters	Since 1 February, over 9 300 households have be	1500	Destroyed Housing	Households	
8	http://www.unhcr.af/UploadDocs/	Conflict and violence	12,754 individuals displaced by conflict assesse	12754	Displaced	People	
9	http://floodlist.com/africa/thousa	Disasters	According to Angola news agency ANGOP, Luanc	800	Uninhabitable Housing	Households	
10	http://www.portalangop.co.ao/an	Disasters	The report reads that the rains of March 28 and	483	Homeless	People	Sumbe and Porto Amb
11	http://www.portalangop.co.ao/an	Disasters	The report reads that the rains of March 28 and	150	Destroyed Housing	Households	Sumbe and Porto Amb
12	http://www.argentinaindependent	Disasters	An estimated 3,000 people have had to be evacu	3000	Evacuated	People	Chaco and Corrientes
13	http://erccportal.jrc.ec.europa.eu/	Disasters	According to media, due to the recent heavy rair	14000	Evacuated	People	Formosa,Misiones, Ch
14	http://www.lanacion.com.ar/1679	Disasters	n toda la provincia hay más de 1300 evacuados.	2600	Evacuated	People	Neuquen, Rio Negro,
15	http://www.buenosairesherald.co	Disasters	Salta Governor Juan Manuel Urtubey rushed to t	15	Evacuated	Households	Salta
16	http://reliefweb.int/sites/reliefwe	Disasters	More than 4,400 people were evacuated as resu	4400	Evacuated	People	Cordoba, Santa Fe, Sa
17	http://www.internal-displacement	Conflict and violence	The majority had returned by 2005, when around	8400	Displaced	People	

Figure 1: Sample training set.

unarranged options by their own and until certain assistance is offered, it is frequently deficient to rebuild their lives. This scenario is further weakening due to the lack of the agenda of international development and displacements. It is not easy to trace and track these IDPs. Many challenges have been encountered in this effort.

According to Baal and Ronkainen (2017), the first issue was no registration systems or particular list of the overall population of the IDPs exist to permit the analysis of the current situation or

	A	B
1	url_id	url
2	1	http://reliefweb.int/sites/reliefweb.int/files/resources/UNICEF%20Mexico%20%20Earl%20Tropical%20Storm%20SitRep%20%20Aug%2008%202016.pdf
3	2	http://reliefweb.int/report/indonesia/asean-weekly-disaster-update-18-24-january-2016
4	3	http://www.dawn.com/news/print/1287094
5	4	http://www.telesurtv.net/news/Fuertes-inundaciones-deja-temporal-en-Neuquen-Argentina-20161026-0013.html
6	5	http://hiphopdx.com/news/id.43019/title.rolling-loud-music-festival-facing-cancellation
7	6	http://reliefweb.int/report/nicaragua/34-comunidades-afectadas-por-r-os
8	7	http://reliefweb.int/sites/reliefweb.int/files/resources/IB4_Italy_Earthquake_09092016.pdf
9	8	http://adinet.ahacentre.org/reports/view/805
10	9	http://reliefweb.int/report/nicaragua/familias-de-el-rama-ya-han-retornado-sus-casas
11	10	http://inciweb.nwcg.gov/incident/article/5108/34753/
12	11	http://reliefweb.int/sites/reliefweb.int/files/resources/Reporte%20complementario%20n%20081.pdf
13	12	http://reliefweb.int/sites/reliefweb.int/files/resources/Redhum_DO_Informe_de_Situacion_No_6_por_vaguada_28_5_16_11_am_COE-20160528-IA-18505.pdf
14	13	http://reliefweb.int/report/haiti/response-january-12th-2010-earthquake-displacement-tracking-matrix-dtm-haiti-round-25
15	14	https://www.humanitarianresponse.info/en/system/files/documents/files/ocha_iraq_flashupdate3_mosul_19october2016_final.pdf
16	15	http://reliefweb.int/sites/reliefweb.int/files/resources/snapshot_burundi_20160801_0.pdf
17	16	http://floodlist.com/europe/greece-floods-zakynthos-november-2016

Figure 2: Sample testing set for tagging online documents.

	A	B
1	excerpt_i	excerpt
2	1	Since 1
3	2	About 172,000 people fled western neighbourhoods of Mosul for camps and emergency sites between 19 February and 23 March as military operations to retake the western part of Mosul from ISIL moved
4	3	This brings to almost 274,000 the total number of people displaced from eastern and western Mosul as of 23 March.
5	4	More than 350,000 people were displaced between 17 October and 23 March, of whom 76,000 returned home to eastern Mosul and surrounding areas.
6	5	About 10,000 sugar cane workers fled the Kokang self-administered zone in northern Shan state after fighting erupted on 6 March in Laukkai town between Myanmar armed forces and the Myanmar Nation
7	6	Laukkai was also hosting a camp with 50 people displaced during 2015 conflict. Most residents were evacuated
8	7	About 120,000 mostly Rohingya people displaced since 2012 should be allowed to return home from rundown temporary camps in western Rakhine state, said a panel led by former UN Secretary General K
9	8	Nine hundred and twenty-one families were displaced â€œin mass displacementâ€ events (groups of more than 50 people or more than 10 families) between 1 January and 10 March by fighting among irre
10	9	about 3,500 people were displaced â€œin mass displacementâ€ events (groups of more than 50 people or more than 10 families) between 1 January and 10 March by fighting among irregular armed groups
11	10	As many as 92,000 people were displaced when more than 25,000 homes collapsed or were left uninhabitable between late November 2016 and mid-March after persistent rainfall and flooding. In early 20
12	11	More than 55,000 people were displaced from Manbij in Aleppo governorate in two major waves of displacement.
13	12	More than 35,000 were displaced northwards because of fighting between government forces and Islamic State in Iraq and the Levant (ISIL), and clashes between ISIL and the Syrian Democratic Forces (SDF
14	13	About 20,000 people were displaced from Maskana, Dayr Hafir and Khafsa to villages in the south-western countryside of Manbij and to Manbij city between 1 and 16 March.
15	14	An estimated 25,000 to 30,000 IDPs do not have adequate shelter. Most are older people, women and children, as reports indicate most young men have been recruited into the SDF or government forces,
16	15	About 17,000 people were displaced from Raqqa city to areas to the north and south of the city between November and early December 2017.
17	16	Between 18,000 and 25,000 people were displaced from areas to the north and west of Raqqa between December 2016 and early February 2017, with a large proportion returning to their communities of or
18	17	About 38,000 conflict-related displacements across 21 of Afghanistanâ€™s 34 provinces were verified between 1 January and 17 March.
19	18	About 19,000 people were registered as displaced in Ouham Pende prefecture after violence between two armed groups on 2 February in the northern town of Bocaranga.
20	19	As many as 6,600 people were registered as displaced in Bambari in Ouaka prefecture. This includes the entire population of Liwa village, who fled to Sangaris camp in Bambari in February because of viole

Figure 3: Sample testing data for extracting terms from online documents.

B. Data Cleaning and Pre-processing

As the raw data to be used in this research is taken from the Internet, the process of data cleaning is needed prior to the data analysis. Firstly, these data (as shown in Figure 2) are website links and the documents are in file type of HTML or PDF. Therefore, a web scrapping is needed to extract these reports, in respect to their individual file type. It is followed by data filtration, in which irrelevant documents in the training set are to be removed (documents that are either not written in English or not reporting and not providing relevant information on internal displacement and human mobility).Next, these pieces of textual data is to be processed with different Natural Language Processing (NLP) annotations. In this NLP process, it first tokenizes at word level (i.e. to split text into words), followed by finding root words of all the split words (lemmatization).

After that, stop words such as “the” and “is” are eliminated before the process of part-of-speech tagger is taking place to indicate the role and structure of those words in a sentence such as nouns, adjectives, verbs and etc.Thirdly, as the title and the content of an article are usually stored separately in a file structure, and title usually contents significant information about that article, title cannot be discarded but to be taken as part of the data to be processed. Thus, the title and the content have been merged and treated as one single piece of data.

C. Machine Learning

After those data have been cleaned up, they are ready to be used in the system. In this research, the approach of supervised machine learning has been chosen due to the nature of classifying online documents with tags. In this approach, a classifier needs to primarily repose its training over a pre-classified document and pre-specified class set by human experts. Once trained, the classifier will be able to apply the model it has learnt from the pre-classified documents on unseen documents, which might ultimately favor to output a prediction with high accuracy. Specifically, the SVM has been chosen because of a lighter requirement on resources (computing power and memory) compared to another infamous approach, the Neural Networks in the same approach.

Besides, SVM has also been recommended by various researchers in the context of classification (Joachims, 2012; Lee, 2010).In the training set used in this research, documents are classified into two categories, namely the “conflict and violence”, and “disaster”. A summarized flow of training the SVM used in this system configured is found as below:

1. Data transformation and scaling: A binary representation is used to present the output class labels, 0 for Disasters and 1 for conflict and violence. Next, by using the built in TfidfVectorizer function in a Python machine learning library – sklearn, the training dataset is converted into a sparse matrix of token counts and then to a matrix of TF-IDF(term frequency–inverse document frequency) features. It ranks the importance of words in a concern document and subsequently be used as weights and the output of this function is the SVM classifier’s input vectors.
2. Cross validation and Grid-Search: It is used to find the best configuration parameters for the SVM classifier.
   a. After obtaining the best parameters, the training dataset is split into 70% of data for training and 30% as the evaluation set for accuracy evaluation. This evaluation set reserved and set apart from the training dataset is used to test the accuracy of the classifier. This process of splitting into training set and evaluation set is repeated 5 times for accuracy scoring (precision and recall).
   b. If the accuracy score is less than 80 % the SVM classifier is reconfigured with a new set of parameters and tested again by using the Grid-Search CV function. The best parameter values obtained and used in this research are C=10, gamma=0.01, and a class weight of (0:1, 1:1) from a range of values as found in Figure 4.
3. Trained model: the best parameters from (2) are used to train the model before using it on the test data.
4. Learned algorithm: the algorithm produced from (3) is stored and to be used with the test data set.

```
# cross-validation
tuned_parameters = [{'kernel': ['rbf'],
                     'gamma': [.01, .03, 0.1, 0.3, 1.0, 3.0],
                     'class_weight':[{0:1,1:1},{0:1, 1:4}, {0:1, 1:5}, {0:1, 1:10}],
                     'C': [1/x for x in [0.1, 0.3, 1.0, 3.0, 10.0]]}]

scores = ['precision', 'recall']
```

Figure 4: Tuned parameters with various ranges of values.

Similarly, after those online documents have been classified with tags, useful information from each of these articles can be retrieved using the approach of supervised machine learning as performed in tagging process, with a modification on learning algorithm and test data. MNB is used to determine the probability of those terms extracted from the document $d$being classified in class c.

$$P(c|d) \propto P(c) \prod_{1 \le k \le n_d} P(P(t_k|c)|c)$$

where

$P(c|d)$is the probability of a document d being in class c.

$P(t_k|c)$is the conditional probability of term t_k of class c. It basically measures how much the term $t_k$corresponds to the correct class c.

$P(c)$is the prior probability of a document occurring in class c.

$(t_1, t_2, \ldots, t_{n_d})$are the tokens in documents d which are part of the words that are used for classification and n_d is the number of such tokens in d.

The MNB gives every processed term a probability that it belongs to a particular class, which is calculated from the occurrences of the term in the training set where the categories are pre-defined.

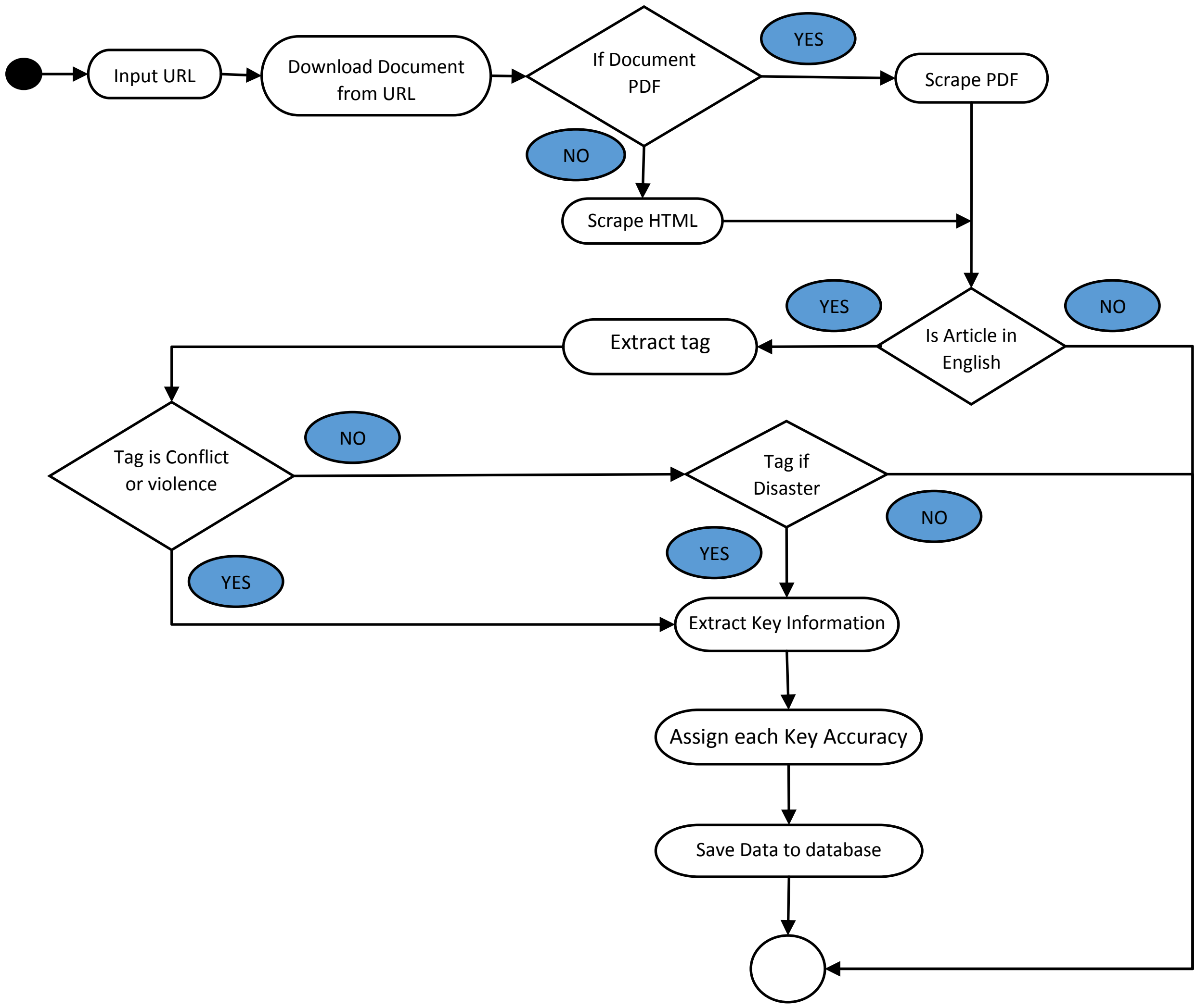
Next, in the process of information retrieval, the following details are extracted:

1. Reporting unit: from the training data set, the reporting unit used are “people” and “household”. The labelled training data is loaded into data frame and with the NLP helper function, excerpt is split into reported facts, where relevant reported facts are chosen based on whether they contain
2. reporting unit entities. Subsequently, a vectorizer of most relevant reports from each excerpt has been trained in loaded data frame against respective reporting unit. Finally, reporting unit is predicted by using grid search over MNB on the vectorizer training output.
3. Reporting terms: Reporting terms consist of descriptive words reporting on the current situation endure by the victims of the displacement event in matter. The extraction of the reporting terms is similar to the process used for extracting reporting units.
4. Location and country extraction: Country name is done by using the Spacy entity extraction. If there are more than one country identified in the fragment, the following rules are used to choose the country:
   a. Country that occurs the highest number of times.
   b. Country that appears first in the text.

5. Displacement figure: Number of people / household displaced is obtained in two ways:
   a. Using the quantity extracted in the most likely report
   b. Heuristic approach (as explained in 5(c))
6. Using a heuristic approach:
   a. When multiple countries are mentioned, the country mentioned in the title is selected. If it is not mentioned in title, the most frequently mentioned country is selected. Otherwise the first mentioned country is chosen.
   b. When multiple reporting terms are recognized, the selection priority is: people families and lastly structures.
   c. Numbers reported in brackets are treated as supplementary information and ignored.

The overall program flow of the system developed is shown in Figure 5.

Figure 5: Program flow of the developed system in tagging online documents and extracting key terms from the documents.



D. System Evaluation

After the model has been trained, it is evaluated by using the test data. There are in total 72 links in the test dataset for classification and 48 documents which have been manually reviewed and verified as relevant documents to the IDPs for information retrieval test.While accuracy has become one of the most commonly used measurement in evaluating the performance of a binary classifier (Bijalwan, 2015), it would also be good to indicate its sensitivity and specificity for false positive and false negative. Any predicted data may fall into one of these four categories when its tag is examined: False Positive (FP) if the system labels it as a positive (in this case "Conflict and Violence") while it is a negative (i.e. "Disaster"); False Negative (FN) if the system labels it as negative while it is a positive; True Positive (TP) and True Negative (TN) if the system predicts the label correctly. Meanwhile, with these FP, FN, TP, TN, the sensitivity and specificity of a learning machine are defined as a ratio as shown below (Veropoulus, Campbell &Cristianini, 1999):

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

According to the result obtained, the following confusion matrix has been built to display the performance of the system (Table1). The sensitivity, specificity and the accuracy are 100%, 92.68% and 95.83% respectively. The sensitivity rate here indicates that the system has successfully detected and classified all conflict and violence related documents correctly. If any concerned governing bodies or organizations are targeting at incidents of IDPs due to the insecurity and chaotic of society or community, this system can give an extremely accurate documents that are related to these human made factors. On the other hand, this system managed to classify 92.68% of the documents under the class of Disaster.

In other words, this system sometimes gives a false alarm in relating some of the disaster related documents with the human made factors. When the FN and FP are compared, in some of the scenarios, it may not be a significant disadvantage to have FP over FN. It is because most if not all of the disasters cannot be stopped and usually requires effort from parties from all around the world to rectify it such as global warming and waste production. While the FP, in the context of this research, the human relatedfactors which resulted a move of IDPs, should be given higher priority in minimizing such an impact by national, regional or international authorities and organizations. On average, this system has assigned correct tags to documents with a success rate of 95.83%. On the other hand, in terms of information retrieval, the accuracy of retrieving relevant documents is 81% from a test dataset of 48 documents. In other words, the developed system manage to extract essential and important terms from the documents and report those information to the users without the intervention of users.

III. RESULT

In this paper we find, the increase of the number of IDPs globally has alerted many countries and international organizations. Most of these incidents are due to factors of conflict and violence, and disaster, according to the IDMC. There are more cases which may not be recorded accurately due to major challenges such as the insecurity of the community, lack of interests from local authorities in resolving it, data privacy and others. Data gathering becomes crucial in detecting such event in order to provide necessary protection and aid to the affected IDPs. With the help of big data technology and machine learning algorithms, it is promising to have data gathered from around the globe as early as possible and analysis can be done within a shorter period of time in order to facilitate more actions to be taken. The proposed system in this paper demonstrated that it is possible to categorize online textual information into the categories suggested by IDMC with outstanding performance and to retrieve useful information such as the reporting unit, reporting term, location and countries involved according. However, as the data samples are relatively smaller, a larger sample size may be needed to further investigate and justify the performance of this system.

Table 1: Result of the proposed system.

N= 72	Classified as Conflict and Violence	Classified as Disaster
Conflict and Violence	TP= 31	FP= 3
Disaster	FN= 0	TN= 38

IV. CONCLUSION

In this paper we conclude as for the improvement, textual data from other sources such as social media can be considered as the input data to better understand and trace the case of IDPs. Of course, the data pre-processing becomes more crucial in eliminating false information and biasness from sources. Besides, the target of this system, the IDPs may be further expanded to refugees who encounter similar experience and difficulty as these IDPs. Their track of movement, impact and consequences to a new settlement area and society may be more serious and significant. Lastly, as the system currently run document tagging and information retrieval manually, automation may be another helpful move towards fully utilizing the technology of big data and machine learning.

ACKNOWLEDGEMENT

Special appreciation expressed to United Nation (UN) and IDMC for providing the datasets used in this research in conjunction with the #IDetect challenge in 2017.

REFERENCES

1. Asplet, M., (2013). *Internal Displacement:Responsibility and Action*, Inter Parliamentary Union.
2. Baal, N. & Ronkainen, L., 2017. *Obtaining representative data on IDPs: challenges and recommendations*, UNHCR the UN Refugee Agency.
3. Bijalwan, V., (2015). *Machine learning approach for text and document mining*, Uttarakhand Technical University Dehradun: Uttarakhand.
4. European Commission, (2017). *Forced displacement: refugees, aslyum-seekers and internally displaced persons (IDPs)*, s.l.: European Comission.
5. General Assembly, (2016). *Protection of and assistance to internally displaced persons*, s.l.: United Nations.
6. IDMC, 2017. *Internal Displacement Mentoring Centre: Displacement in the Context of Conflict and Violence*. [Online] Available at: http://www.internal-displacement.org/internal-displacement/context-of-displacement/conflict-and-violence/ [Accessed 15 July 2017].
7. Joachims, T., (2012). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*, Dortmund: Universitat Dortmund.
8. Lee, Y., (2010). *Support Vector Machines for Classification: A Statistical Portrait*, Ohio: The Ohio State University.
9. Veropoulos K., Campbell C. & Cristianini, N., (1999). *Controlling the Sensitivity of Support Vector Machines*, In: Proceedings of the International Joint Conference on AI, pp. 55 – 60.

AUTHORS PROFILE

1. **HakimaFathiMahamoud** is working as School of Computing, Asia Pacific University of Technology and Innovation, Malaysia

2. **Raja RajeswariPonnusamy** is working as School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia

3. **Ho Ming Kang** is working as School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia

4. **Jacob Sow Tian You** is working as School of Media, Art and Design, Asia Pacific University of Technology and Innovation, Malaysia