



## Automated Detection of Bilingual Obfuscated Abusive Words on Social Media Forums: A Case of Swahili and English Texts

Geoffrey Njovangwa<sup>1,2\*</sup> and Godfrey Justo<sup>2</sup>

<sup>1</sup>Tanzania Forestry Research Institute, P. O. Box 1854, Morogoro, Tanzania,

<sup>2</sup>University of Dar es Salaam, College of Information and Communication Technologies, P. O. Box 33335 Dar es Salaam, Tanzania

\*Corresponding author, E-mail: [stargeof@gmail.com](mailto:stargeof@gmail.com)

Received 21 Jun 2021, Revised 19 Sep 2021, Accepted 25 Sep 2021, Published Oct 2021

DOI: <https://dx.doi.org/10.4314/tjs.v47i4.2>

### Abstract

The usage of social media has exponentially grown in recent years leaving the users with no limitations on misusing the platforms through abusive contents as deemed fit to them. This exacerbates abusive words exposure to innocent users, especially in social media forums, including children. In an attempt to alleviate the problem of abusive words proliferation on social media, researchers have proposed different methods to help deal with variants of the abusive words; however, obfuscated abusive words detection still poses challenges. A method that utilizes a combination of rule based approach and character percentage matching techniques is proposed to improve the detection rate for obfuscated abusive words. The evaluation results achieved F1 score percentage ratio of 0.97 and accuracy percentage ratio of 0.96 which were above the significance ratio of 0.5. Hence, the proposed approach is highly effective for obfuscated abusive words detection and prevention.

**Keywords:** Rule based approach, Character percentage matching techniques, Obfuscated abuse, Abuse detection, Abusive words, Social media.

### Introduction

It is not uncommon when engaged on online forums for words such as “as\$hole” or “m\$enge” to unfortunately surface bringing detrimental impacts on users or community psychological wellbeing and experience. In an attempt to protect against abusive words, Internet firms have set standards, policies and guidelines that users must comply with, in addition to deployment of human moderators on software platforms, that internally also potentially use regular expressions and blacklist to detect abusive words and prevent such users’ posts or comments (Nobata et al. 2016). Under political landscape, for example, the pre and post political leaders’ elections periods in Kenya, in East Africa, saw severe abusive

words exchanges amongst rival party factions, that ultimately led the Kenyan people at finding solutions to combat the usage of offensive languages (Maloba 2013). In Tanzania, the spread of abusive words used on social media platforms, led the government to come up with a legislation (the Cybercrimes Act No 14 of 2015) to combat the vice. The study by Mhagama (2016) showed that the implementation of the legislation led to reduction by 60% of offensive language cases.

In social media forums, it is uncommon to see cases of posts or comments that make use of two languages in a single thread, often in an attempt to disguise abusive words. In Tanzania for example, the social media platform, Jamii Forums, have post contributors and respondents

that use both Kiswahili and English in a single thread post or comment, some occasionally augment non-abusive words with abusive words. The abusive words may be obfuscated or non-obfuscated, which makes it harder for a common approach to detect. The choice of a bilingual context of English and Swahili is, indeed, influenced by the fact that the two languages remain as official languages in Tanzania since independence (UNICEF 2016). The prior studies by Maloba (2013) and Mugambi (2017) proposed detection methods under bilingual abusive words context, but failed to detect obfuscated abusive words. A rule based approach and character percentage match technique with features of Natural Language Processing, regression analysis and Boolean system, is proposed for obfuscated and non-obfuscated abusive words detection to address the gap.

#### **Related work**

Prior work in the area of abusive words detection had spread across many overlapping fields. This somewhat has led to confusion because different works may address specific aspects of abuse detection, notably, Cyber-bullying, Offensive language, hate speech, abusive words, etc. According to Mugambi (2017), abusive words are expressions of hostility language without any stated explanation for them. Elsewhere, Saravananaraj et al. (2016) defined abusive words as an aggressive act carried out by a group or individual, using electronic forms of contact or face to face interactions, repeatedly and over time, against a victim who cannot easily defend himself or herself. Also, Nobata et al. (2016) stated that abusive words are the terms used to refer to hurtful language and include hate speech, derogatory language and also profanity. But this paper adopts the abusive words definition by Fortuna and Nunes (2018) as words that are improperly used towards an entity or entities, often to unfairly or improperly gain benefit. Obfuscated abusive words are stated as abusive words that have been obscured of the intended meaning of

communication by making them difficult to understand, usually with confusing and ambiguities (Maloba 2013). Further, in a study of Pamungkas and Patti (2019) they defined obfuscated abusive words as abusive words written in such a way that are less clear and non-obvious for reader to comprehend. However, we define an obfuscated abusive word as an abusive word written in a twisted form to disguise the obvious meaning in order to evade detection to achieve its intention of creating hostile environment towards innocent victim(s). It is noted from Mugambi (2017) and Maloba (2013) that obfuscated abusive words are hard to detect as they tend to argument special characters and other misleading words, hence lead to poor detection performance on their proposed approaches.

Popular abusive words detection approaches make use of word classification method, including blacklist, BoW and N-gram (Nobata et al. 2016). The blacklist (dictionaries) strategy makes use of dictionaries. This approach constitutes making a list of words (the dictionary) that are searched and counted in the text. The word frequencies can be used directly as features or to compute scores. The various contexts of use for abusive words detection trading on the magnitude and computational cost requirement are studied in Dinakar et al. (2011), Dadvar et al. (2012), Liu and Forss (2015). It is noted by Maloba (2013), however, the word classification approach is not feasible for obfuscated abusive words detection because it makes use of straight forward comparison approach of context with no additional knowledge.

The bag-of-words (BoW) method relates to supervised classification approaches, in which a corpus is created based on the words that are in the training data, instead of a pre-defined set of words used in dictionaries (Burnap and Williams 2016). Subsequently, the frequency of each word in the collection is used as a feature for training a classifier (Greevy and Smeaton 2004). It is noted by Malmasi et al. (2017) that the word sequence is ignored including its

syntactic and semantic contents, thus, could lead to misclassification if the words are used in different context (Fortuna and Nunes 2018).

The N-grams method on the other hand is one of the most used techniques in abusive words automatic detection and related tasks (Malmasi et al. 2017). According to Burnap and Williams (2016), the most common n-grams approach consists of combining sequential words into lists of size N, the goal is to enumerate all the expressions of size N and count the occurrences of them. This allows improving the classifiers' performance because it incorporates at some degree the context of each word, while instead of using words; it is also possible to use n-grams with characters (Mehdad and Tetreault 2016). It is however, noted by Badjatiya et al. (2017) that this approach is not susceptible to spelling variations as when words are used.

The rule based approach has already been successfully used for word classification in some applications (Fortuna and Nunes 2018). The rule based approach is described as being enriched by linguistic knowledge and does not involve learning or any predictions, hence, it relies on pre-compiled list and dictionaries of subjectivity clues (Haralambous and Lenca 2014). The down side, however, is that it does not perform well for applications that demand predictions on an instance to occur; hence it requires a well detailed dataset (Malmasi et al. 2017). The approach can perform well in applications that have a well detailed data source, potentially also in abusive words detection, however, cannot detect obfuscated abusive words without combination with other approaches (Fortuna and Nunes 2018).

Existing studies in the multilingual contexts include that in Nobata et al. (2016) that used a supervised classification approach with NLP features in which all texts were treated as collection of unordered words while ignoring the semantic and syntactic information. Subsequently, were compared with a corpus depending on the classifier. The NLP features used are N-grams, Linguistic, Syntactic and Distributional Semantics. However, it was

remarked that better results could be obtained by experimenting with other mixed approaches of offensive language detection. On the other hand, Mugambi (2017) used a supervised classification approach using Support Vector Machines (SVMs) with features including N-grams, automatically derived blacklists, manually developed regular expressions and dependency parse features to automatically detect multilingual (English and Swahili) hate speech on Twitter. The study was limited to 140 characters texts, but also the detection of obfuscated abusive text remained pending for future research. Further, Pamungkas and Patti (2019) proposed abusive words detection method that used a hybrid approach with deep learning and a multilingual lexicon to cross-domain and cross-lingual detection of multilingual abusive words. The approach first characterized the available datasets by capturing various phenomena related to abusive language and investigated the characterization in cross-domain classification. Secondly, the approach explored the use of a domain-independent, multilingual lexicon of abusive words, called HurtLex, in both cross-domain and cross-lingual settings. Lastly, the approach took advantage of the availability of multilingual word embedding to build a joint learning approach in the cross-lingual settings. Overall, the approach performed well but struggled in the detection of obfuscated abusive words due to the fact that its major focal goal was to achieve multilingual abusive words detection.

## **Materials and Methods**

### **Proposed approach**

This work proposed a Rule-based approach in combination with Character Percentage Match technique for obfuscated abusive text detection. The rule-based approach is based on application of human-made rules to store, sort and manipulate data, thus, it mimics human intelligence. Typically, the rule-based method requires a set of facts or source data and a set of rules for manipulating the data. The rules are sometimes referred to as 'If statements' as they

tend to follow the line of ‘IF A happens THEN do B’. Hence, the rule-based approach is enriched by linguistic knowledge and does not involve learning or any predictions, as it relies on pre-compiled list and dictionaries of subjectivity clues. In particular, the approach works based on rules, which outline triggers and the actions that should follow (or are triggered). Generally, rules take the form: ‘IF’ outlines the trigger, ‘THEN’ specifies the action to complete. The following rule was formulated to govern the obfuscated abusive words detection process:

*If TOKEN = WORD in data source by 70% or higher THEN categorize it as abusive ELSE none abusive;*

The Character Percentage Match technique determines how closely two values match each other by calculating the Character Edit Distance between two String values, and also taking into account the length of the longer or shorter of the two values, by character count. In mathematical terms, the Character Percentage Match comparison used the following formula to calculate the results (Oracle 2011):

$$CP = \left[ \frac{MaxCL - CED}{MinCL} \right] \times 100,$$

where,

CP = Character Match Percentage;

CED = Character Edit Distance between two words;

MaxCL = Maximum Character Length of the two words being compared; and

MinCL = Minimum Character Length of the two words being compared.

As a proof of concept a prototype forum was developed to implement the proposed approach for obfuscated bilingual abusive words detection. The architecture of the prototype is depicted in Figure 1, which constitutes three main components: the Social Media Forum, Text Analysis and Boolean System components. The text submitted to a social media forum is used as input to the detection sub-system, which in

turn is processed for output. In particular, the Social Media component serves a user interface for text input, the Text Analysis component in turn, process the input. The latter, proceeds as follows:

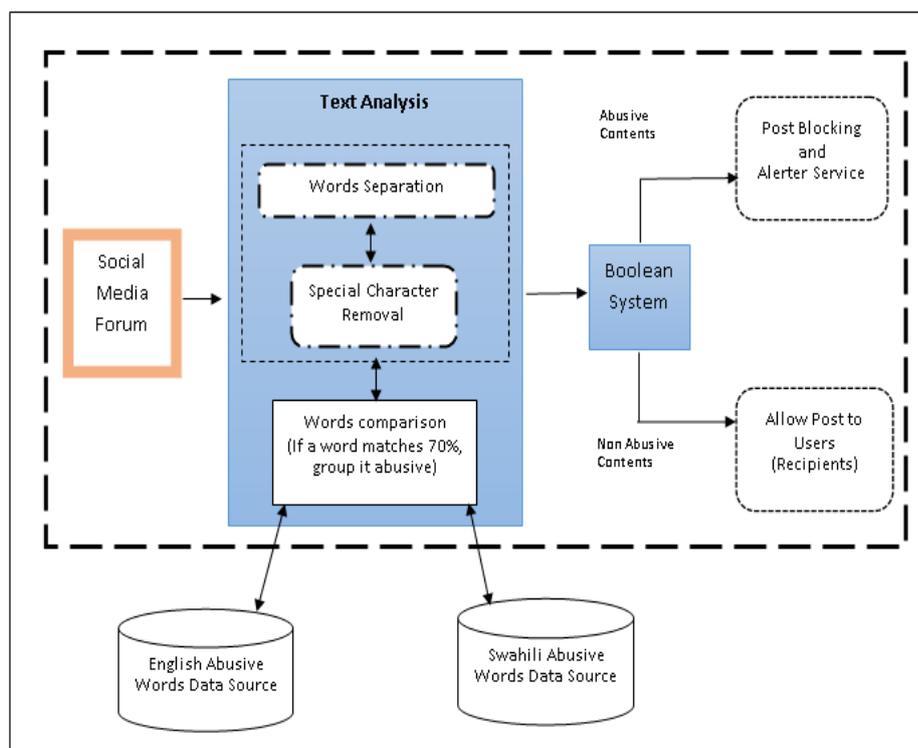
Checks on whether the input is a word or a sentence, subsequently, if the entered text is a sentence, it is tokenized with the help of *phptokenizer function "strtok ()"* which utilizes open Natural Language Processing toolkit (*openNLP*-a machine learning based toolkit for processing and tokenizing sentences) (He and Kayaalp 2006), before being sent for special character removal stage. On the other hand, if the input text is not a sentence, then it will be sent for special character removal stage. In turn, a token is checked whether it has any special character (through a *php function preg\_replace ()*-which uses regular expression to extract special character from a word (PHP 7.4.0 RC2, 2019), in which case the special character is removed, or otherwise, the token is moved for the next stage. At this stage, a token is compared to the list of abusive words in the Swahili and English data sources (corpora). This is the part where both rule-based approach and character percentage match technique are employed. The percentage match technique used a lower threshold of 70% for words similarity (adopted from Jin and Schuler 2015, where for words similarity, character percentage match of 70% or higher is recommended).

Therefore, a token with similarity to a word in the corpus by 70% or higher, is categorized as abusive, otherwise as non-abusive word. The Boolean system uses the token categorization to make decisions on what action to take accordingly. In particular, if the post has one or more abusive words, the Boolean system will block it and notify sender, otherwise the post without abusive words is displayed to public forum users.

The abusive words dataset (corpus) was initially composed from ready developed online sources (Ofcom 2016, Ndalul and Babusa 2013), where 100 abusive words were randomly selected and compiled. Expanded data set was obtained through composition by a purposely

selected trained team of five users. The team was given the initial 100 abusive words (50 English and 50 Swahili) from the ready developed online sources and each instructed to compose 20 own new abusive texts (5 English, 5 Swahili, 5 Obfuscated-English and 5 Obfuscated-Swahili). Subsequently, each was next instructed to compose 20 own new non abusive text (5 English, 5 Swahili, 5 Obfuscated-English and 5 Obfuscated-Swahili). Finally, each was instructed to generate 20 own

new obfuscated abusive words (10 English and 10 Swahili). Overall, this led to the generation of a total of 200 abusive words, 200 abusive texts and 100 non abusive texts, resulting to total data set of 200 words and 400 texts that were newly built, in addition to the 100 abusive words initially generated from online sources. The final compiled data set is archived at [https://github.com/gnjovangwa/ABUSES\\_DAT A](https://github.com/gnjovangwa/ABUSES_DAT_A).



**Figure 1:** Proposed prototype architecture for abusive words detection and prevention.

### Results and Discussions

A forum prototype based on our proposed method was implemented on TAFORI forum (<http://taforiforum.or.tz//index.php>)-a PHPBB based open source social media platform that support seamless customizations and testing flexibility. To guide evaluation for the effectiveness of the proposed approach, the prototype was subjected to purposely selected users. Quantitative data were collected

pertaining to various users' inputs and outputs, subsequently, the data were analyzed and results presented in terms of tables and figures. To aid results interpretation, percentages ratios on key performance metrics, in particular, the F1 score and accuracy (Feldman and Sanger 2006), respectively, were employed. The F1 score is used to obtain the average ratio of the precision and recall in detection of abusive

words, as calculated using the following formula (Feldman and Sanger 2006):

$$F1\text{-Score} = 2*TP / (2*TP + FP + FN)$$

The accuracy, here defined as the fraction of predictions that the detection approach got right, i.e., as the accuracy in detecting abusive words, was determined by the following formula (Feldman and Sanger 2006):

$$Accuracy = (TP + TN) / (TP + FP + TN + FN),$$

where,

TP = Abusive texts and words branded as true positive;

TN = Non abusive texts branded as true negative;

FN = Abusive texts and words that were not detected, branded as false negative;

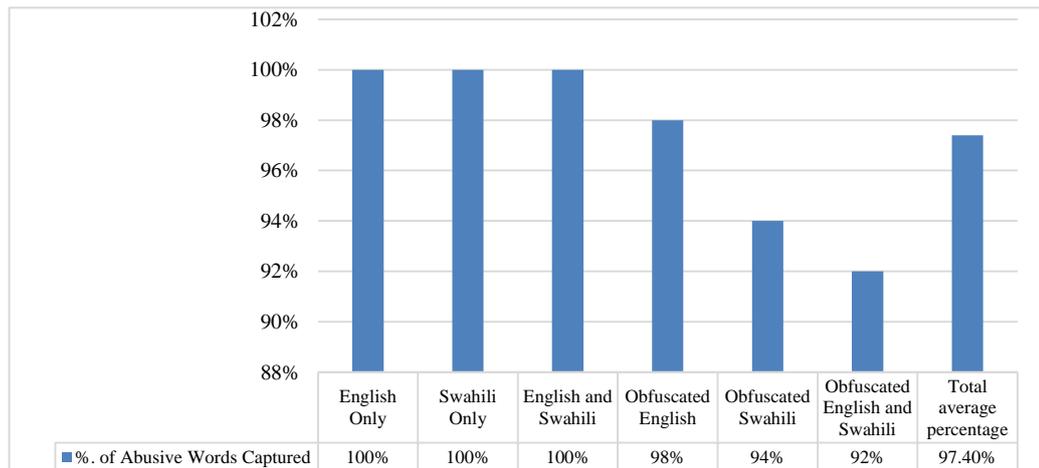
FP = Non abusive texts that were detected, branded as false positive.

A test case of 300 words was generated for experimentation where the words were divided into 6 categories, each of 50 words, in which all the words were abusive. Following the test, the results showed that 97.4% of the abusive words were captured, whereas the remaining 2.6% were not captured. The latter cases were

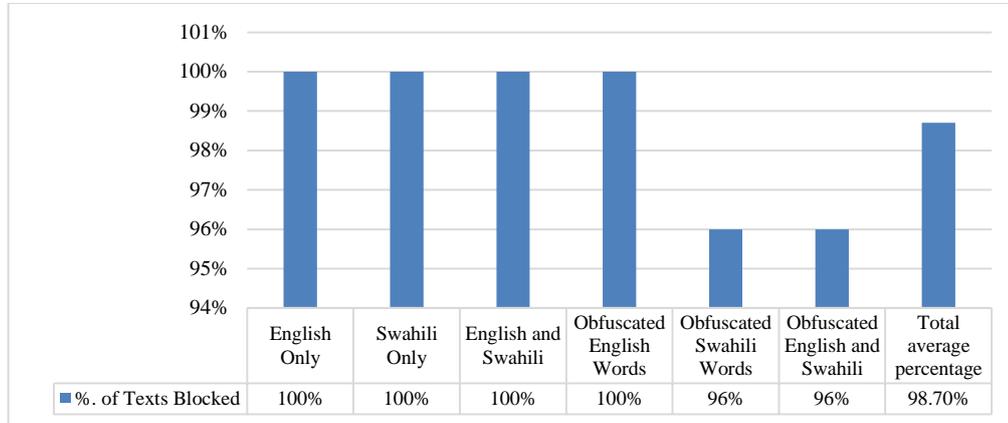
because the character similarities were less than the set threshold similarity of 70%. As a result, the 2.6% of words were seen to be not critically abusive, as depicted by Figure 2.

A second test case of 300 texts was experimented with texts divided into 6 categories whereby each category had 50 texts, in which all the texts had abusive words. Following this experiment, the results showed that 98.7% of texts that contained abusive words were blocked, whereas the rest, a 1.3% were not blocked. The undetected texts were attributed to the fact that their character similarity was less than the 70% lower threshold limit, hence, assumed to be not critically abusive texts as seen in Figure 3.

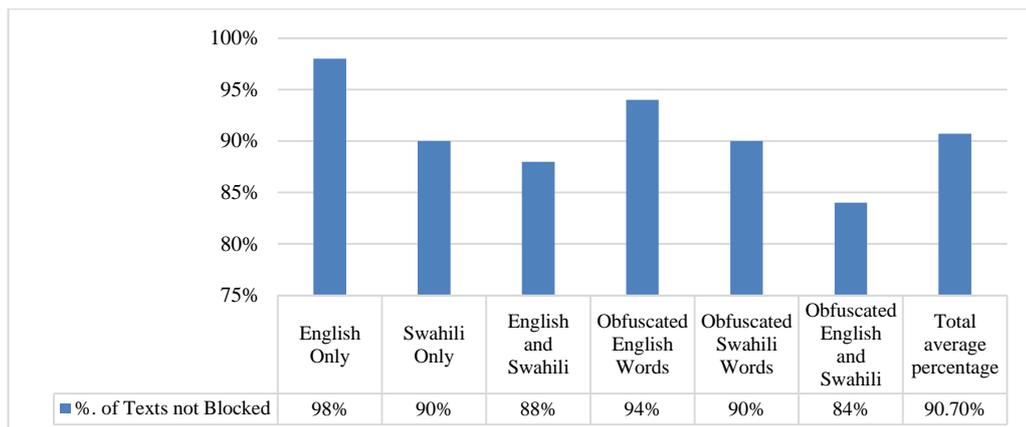
A third test case had 300 texts that were divided into 6 categories each of 50 texts, in which the texts categories had no abusive words. Following the experiment, the results showed that 9.3% of the texts were blocked whereas the rest 90.7% of the texts were not blocked. The percentage 9.3% of blocked texts were achieved due to the similarity of text entered to the abusive words in the data source in relation to the threshold of 70% similarity or high criteria hence were assumed to be abusive as depicted by Figure 4.



**Figure 2:** Percentage detection rate for the test case of 300 abusive words put under 6 categories of 50 words and corresponding total average percentage.



**Figure 3:** Percentage detection rate for a test case of 300 texts containing abusive words put under 6 categories of 50 texts and corresponding total average percentage.



**Figure 4:** Percentage detection rate for a test case of 300 texts with no abusive words put under 6 categories of 50 texts and corresponding total average percentage.

A confusion matrix was generated in order to determine the key performance metrics achievement by the proposed approach, in particular, the F1-score and accuracy for the abusive words detection. To achieve this goal, a total of 600 abusive texts and words were branded as true positive (TP), 300 non abusive texts were branded as true negative (TN), 28 abusive texts and words that were not detected were branded as false negative (FN) and 12 non abusive texts that were detected as abusive were branded as false positive (FP). The metrics evaluation resulted into F1-score of 0.97 (i.e., harmonic mean of 97% from recall

and precision) and accuracy of 0.96 (i.e., arithmetic mean of correctly detected being 96% out 100% of the abusive words). Assuming the ratio of above 0.5 (50%) as significant and the fact that most of the results obtained were 0.9 (90%) and above; the performance results implied that the proposed approach is highly effective in the detection, hence, had potential to achieve ultimate prevention of abusive words from being viewed by innocent social media forums users.

Further, a validation step was conducted to compare the performance metrics results achieved by similar previously proposed

approaches. To achieve this goal same data set was used across other approaches (data can be accessed at [https://github.com/gnjovangwa/ABUSES\\_DATA](https://github.com/gnjovangwa/ABUSES_DATA)).

The involved other approaches included:

- (i) Sentiment analysis for hate speech detection on social media: TF-IDF weighted N-Grams based approach by Mugambi (2017). The approach used support vector machine together with N-Gram for detection of bilingual abusive words and cross-domain and cross-lingual abusive language detection.
- (ii) Hybrid Approach with Deep Learning and a Multilingual Lexicon by Pamungkas and Patti (2019). The approach used a hybrid method with deep learning plus multilingual lexicon for detection of abusive content.

Figure 5 depicts the comparative performance analysis as computed outside the ratio of one. The results show that the proposed approach outperformed its predecessors on both fronts, achieving F1 score of 0.97 and accuracy of 0.96, compared to F1-score of 0.73 and accuracy of 0.71 by the approach of Pamungkas and Patti (2019) and F1-score of 0.65 and

accuracy of 0.61 by the approach of Mugambi (2017). This performance achievement by the former can be attributed to the fact the proposed approach combines the rule based approach and character percentage match technique compared to predecessor methods which relied on one approach only. In particular, the other approaches lack the right strategy to deal with bilingual obfuscated abusive words, which heavily impacted on their observed low performance.

It is noted that, although the test data used during the experimentation are different from the test data that may have been used by other researchers (as different categories have been used to separately test the effectiveness of the proposed approach), it is believed that this criterion of different test data should not have heavily affected the performance of the other approaches. This is due to the fact that there was a limited number of 156 English and Swahili non-obfuscated abusive words (Ofcom 2016, TUKI 2013), while the sample had 100 English and Swahili non-obfuscated abusive words which would bring a margin error of 5%, at confidence level of 90%. Hence, with a 5% error, it is significant to say that the results are valid.

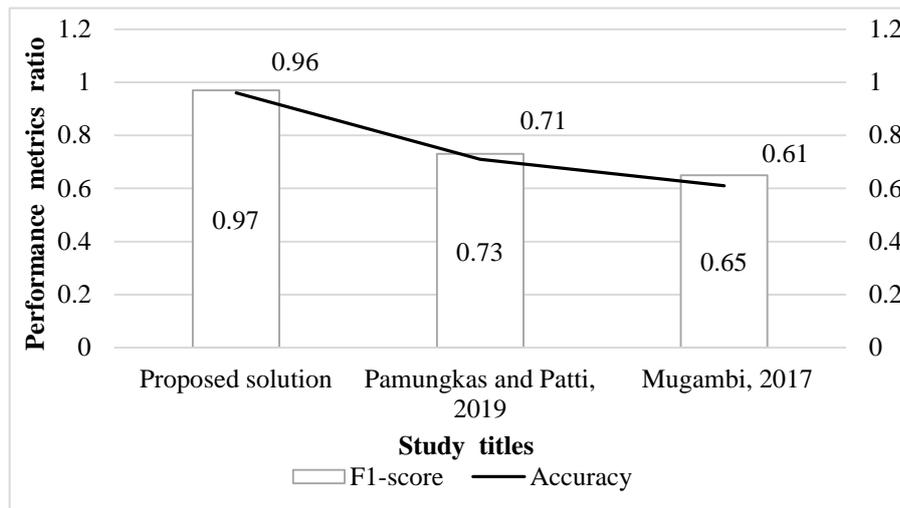


Figure 5: Performance comparison to other related approaches.

## Conclusion

The paper achieved the main goal to bridge the knowledge gap on improving the detection and prevention of obfuscated abusive words on social media forums. To achieve this goal, an improved obfuscated abusive words detection approach was proposed and its implementation prototype integrated into the TAFORI Forum for experimentation. A series of test cases were generated and applied to the tool leading to improved performance results with F1 score of 0.97 and accuracy of 0.96. Given the growing importance of social media, it is imperative that future studies to explore more improved approaches to curb the growing problem of social media abuse prone users.

## Acknowledgements

This article is an extract of the dissertation report from the research conducted under the auspices of the University of Dar es Salaam, while the corresponding author was affiliated as an M.Sc. in Computer Science student and the co-author as the supervisor. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- Badjatiya P, Gupta S, Gupta M and Varma V 2017 Deep learning for hate speech detection in tweets. In Proceedings of the 26<sup>th</sup> International Conference on World Wide Web Companion, pages 759–760. International World Wide Web Conferences Steering Committee.
- Burnap P and Williams LM 2016 Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science* 5(1): 11.
- Dadvar M, de Jong FMG, Ordelman RJF and Trieschnigg RB 2012 Improved cyberbullying detection using gender information. In: Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012) (pp. 23-25). Ghent: Ghent University. Available online on <https://research.utwente.nl/en/publications/improved-cyberbullying-detection-using-gender-information>.
- Dinakar K, Reichart R and Lieberman H 2011 Modeling the detection of textual cyberbullying. *The Social Mobile Web* 11(02). Available online on <https://ie.technion.ac.il/~roiri/papers/3841-16937-1-PB.pdf>.
- Feldman R and Sanger J 2006 *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, USA.
- Fortuna P and Nunes S 2018 A survey on automatic detection of hate speech in text. *ACM Comput. Surv.* 51: 1-30.
- Greevy E and Smeaton FA 2004 Classifying racist texts using a support vector machine. In Proceedings of the 27<sup>th</sup> annual international ACM SIGIR conference on Research and development in information retrieval, pages 468–469. Association for Computing Machinery, USA.
- Haralambous Y and Lenca P 2014 Text classification using association rules, dependency pruning and hyperonymization. ArXiv, [abs/1407.7357](https://arxiv.org/abs/1407.7357). DOI: 10.6084/m9.figshare.1189289.v1
- He Y and Kayaalp M 2006 A Comparison of 13 Tokenizers on MEDLINE (Report number: LHNCBC-TR-2006-003). The lister hill national center for biomedical communications, National Library of Medicine, Bethesda, USA. DOI: 10.13140/2.1.1133.1206
- Jin L and Schuler W 2015 A comparison of word similarity performance using explanatory and non-explanatory texts, In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 990-994. Available online at: <https://www.aclweb.org/anthology/N15-1101.pdf>.
- Liu S and Forss T 2015 New classification models for detecting hate and violence web content. 2015 7<sup>th</sup> International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 01, 487-495, IEEE.
- Maloba WJ 2013 *Use of regular expressions for multilingual detection of hate speech in*

- Kenya, Master thesis, Strathmore University. Available online on <https://suplus.strathmore.edu/handle/11071/2242>.
- Malmasi S, Evanini K, Cahill A, Tetreault J, Pugh R, Hamill C, Napolitano D and Qian Y 2017 A report on the 2017 Native Language Identification shared task. In Proceedings of the 12<sup>th</sup> Workshop on Innovative Use of NLP for Building Educational Applications, pp. 62-75. Copenhagen, Denmark: ACL. Available online at <http://www.aclweb.org/anthology/W17-5007>.
- Mehdad Y and Tetreault J 2016 Do characters abuse more than words? In Proceedings of the SIGdial 2016 Conference: The 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 299-303.
- Mhagama H 2016 Tanzania: Law Reduces Cyber Crime By 60 Per Cent. Tanzania Daily News (Dar es Salaam). Available online at <https://allafrica.com/stories/201601250502.html>.
- Mugambi SK 2017 *Sentiment analysis for hate speech detection on social Media: TF-IDF weighted N-Grams based approach*, Master's thesis, Strathmore University. Available online on <http://suplus.strathmore.edu/handle/11071/5657>.
- Ndalu A E and Babusa H 2013 Kamusi Teule ya Kiswahili. Kilele cha Lugha (Swahili Edition). East African Educational Publisher.
- Nobata C, Tetreault J, Thomas A Mehdad Y and Chang Y 2016 Abusive language detection in online user content. In: Proceedings of the 25<sup>th</sup> International Conference on World Wide Web, pages 145–153. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE.
- Ofcom 2016 Attitudes to potentially offensive language and gestures on TV and radio: Quick reference guide. The Office of Communications, UK. Available online on [https://www.ofcom.org.uk/\\_\\_data/assets/pdf\\_file/0022/91624/OfcomOffensiveLanguage.pdf](https://www.ofcom.org.uk/__data/assets/pdf_file/0022/91624/OfcomOffensiveLanguage.pdf)
- Oracle 2011 Enterprise Data Quality Help version 9.0. Oracle Corporation, USA. Available online on [https://www.oracle.com/webfolder/technetwork/data-quality/edqhelp/Content/processor\\_library/matching/comparisons/character\\_match\\_percentage.htm](https://www.oracle.com/webfolder/technetwork/data-quality/edqhelp/Content/processor_library/matching/comparisons/character_match_percentage.htm).
- Pamungkas EW and Patti V 2019 Cross-domain and cross-lingual abusive language detection: a hybrid approach with deep learning and a multilingual lexicon. In: Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 363–370. Association for Computational Linguistics, Florence, Italy.
- PHP 7.4.0 RC2 2019 PHP Manual Documentation. Available online on <https://www.php.net/manual/en/function.strtok.php>.
- UNICEF 2016 Tanzania: The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa. Available online on <https://www.unicef.org/esaro/UNICEF2016LanguageandLearningTanzania.pdf>.
- Saravananaraj A, Sheeba JI and Devaneyan SP 2016 Automatic detection of cyberbullying from Twitter. *Int. J. Comput. Sci. Inf. Technol. Secur. (IJCSITS)* 6: 26-32.
- TUKI 2013 Kamusiya Kiswahili Sanifu. Oxford University Press and Institute of Kiswahili Research. Dar es Salaam.