Offensive Language Detection on Social Media Based on Text Classification

Conference Paper · January 2022

DOI: 10.1109/CCWC54503.2022.9720804

7 authors, including:

Parisa Hajibabaee
University of Massachusetts Lowell

Masoud Malekzadeh
University of Massachusetts Amherst

Maryam Heidari
George Mason University

Armin Esmaeilzadeh
University of Nevada, Las Vegas

Offensive Language Detection on Social Media Based on Text Classification

Parisa Hajibabaee
University of Massachusetts at Lowell
parisa_hajibabaee@student.uml.edu

Masoud Malekzadeh
University of Massachusetts at Lowell
masoud_malekzadeh@student.uml.edu

Mohsen Ahmadi
Arizona State University
Pwnslinger@asu.edu

Maryam Heidari
George Mason University
mheidari@gmu.edu

Armin Esmaeilzadeh
University of Nevada Las Vegas
esmaeilz@unlv.nevada.edu

Reyhaneh Abdolazimi
Syracuse University
rabdolaz@syr.edu

James H Jr Jones
George Mason University
jjnoes@gmu.edu

***Abstract*—There is a concerning rise of offensive language on the content generated by the crowd over various social platforms. Such language might bully or hurt the feelings of an individual or a community. Recently, the research community has investigated and developed different supervised approaches and training datasets to detect or prevent offensive monologues or dialogues automatically. In this study, we propose a model for text classification consisting of modular cleaning phase and tokenizer, three embedding methods, and eight classifiers. Our experiments shows a promising result for detection of offensive language on our dataset obtained from Twitter. Considering hyperparameter optimization, three methods of AdaBoost, SVM and MLP had highest average of F1-score on popular embedding method of TF-IDF.**

***Index Terms*—offensive language detection, social media, machine learning, text mining**

I. Introduction

Deep learning have changed numerous scientific fields in the previous decade, including natural language processing [1], [2] (NLP) [3]–[6], medical imaging [7], healthcare, cyber security, social computing [8], and a variety of other topics [9]–[61]. In recent years, with the seeming growth of social media, new concerns regarding users' mental and physical safety have been introduced. Based on a report in [62], among students between 12 to 18 who reported being bullied at school, 15% were bullied online through social medias. In addition, the percentage of individuals who have experienced being victims of cyberbullying during their lifetime has more than doubled from 2017 to 2019 from 18% to 37% [63]. Offensive, hateful or threatening speech on the content exchanged by the crowd might range from minor or implicit bullying to severe and explicit violent threatening over victims with specific characteristics such as race, sex, religion, community, etc. [64] shows that the rise of public media cyberbully poses a global problem that might damage people's online lives. The state-of-the-art approaches target various contexts, domains, platforms for detecting a specific category of offensive language, e.g., hate speech with or without considering the severity. In this regard, various datasets also have been published to evaluate the correctness and precision of proposed methods [65]–[67].

In this study, we propose a modular text classification pipeline consisting of modular cleaning phase and tokenizer, three embedding methods, and eight classifiers. The experiment done in this study is based on Twitter, and a dataset was optimized effectively. Although we do not claim that our framework would perform well on *all* social media platforms, it could provide future research direction to guide academic and industry researchers. The broader impact of this paper can be related to the systematically investigation of detecting online

978-1-6654-8303-2/22/$31.00 ©2022 IEEE

2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) | 978-1-6654-8303-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCWC54503.2022.9720804

harassment on social media platforms. Moreover, due to social media platforms' inherent features, it is impossible to generalize a model for all the platforms. For example, [68] shows that training a classifier on Reddit is more challenging than Gab because the average length of posts and conversations is longer. Hence, Reddit input introduces more noise than Gab for a classification task. The following is a summary of the rest of the paper. First, we describe our experimental setup and different approaches we take in Section II. We outline our case study in Section III and present our numerical experiments and discussions in Section IV. Finally, Section V presents concluding remarks.

II. Method

This section briefly discusses the steps taken for cleaning and preparing the dataset and also conducting the experiments. Furthermore, Fig. 1 shows a graphical overview of these steps, discussed in the following.

A. Data Preparation

Data Preparation is the first step for training binary classifiers. The strategies for data preparation, which need to be carefully conducted, are described as below:

- **Basic cleaning methods**: We need to clean the data as (i) extracting the pure text from the dataset, removing duplicates, and NaNs (ii) transforming to lowercase (iii) expanding the abbreviations.
- **Slangs**: Given the micro-blogging style of Twitter, using slangs are typical. Slangs bring difficulties to text mining approaches, especially for those emerging lately and thus do not have an updated entry in any dictionaries. So, we plan to transform the text into a canonical form using the reference dictionary[1] for slangs and abbreviations.
- **Removing methods**: Using hashtags, user references, links, and emojis are typical on social media platforms. Therefore, preprocessing the data and selectively removing the typical patterns are essential to normalize the text.

[1]https://github.com/goncalopereira/twitter-moods

B. Tokenizer

To start any text analysis, we need to break down the text into smaller parts like paragraphs and sentences and then convert words into tokens. We can create our customized tokenizers on sentence-level or word-level in our framework and then pass them into embedding methods.

C. Feature Engineering

In our experiment, we utilize the conventional vectorization embedding approaches such as i) Term Frequency-Inverse Document Frequency (TF-IDF), ii) Word2Vec, iii) FastText to transform the text into numerical representation (also called embedding or vector). In the following, we shortly describe each one.

- **TF-IDF**: One way to represent words into vectors is to count the occurrence of words seen in the whole documents. One caveat of this method is the overemphasizing the frequent words in the dataset. In contrast with the word counting method, TF-IDF distributes the weight of frequent words by their relative frequency.
- **Word2Vec**: The word2vec method takes a corpus of text as input and returns word vectors as output. There are two model architectures to produce a distributed representation of words. The continuous bag-of-words (CBOW) architecture predicts the current word based on the context(window size), and the Skip-gram predicts surrounding words(defined window) given the current word.
- **FastText**: FastText represents a low-dimensional vector text that is generated by summing vectors corresponding to the words in the text. Neural Network is being used in FastText for word embedding. FastText model is often compared to other deep learning classifiers with a higher speed and accuracy for training and evaluation [69], [70].

D. Classification Algorithms

In this study, we incorporate eight classifiers for the binary classification task. Our classifiers are: i) (Gaussian) Naïve Bayes (NB), ii) Decision Tree (DT), iii) Logistic Regression (LR), iv) Random

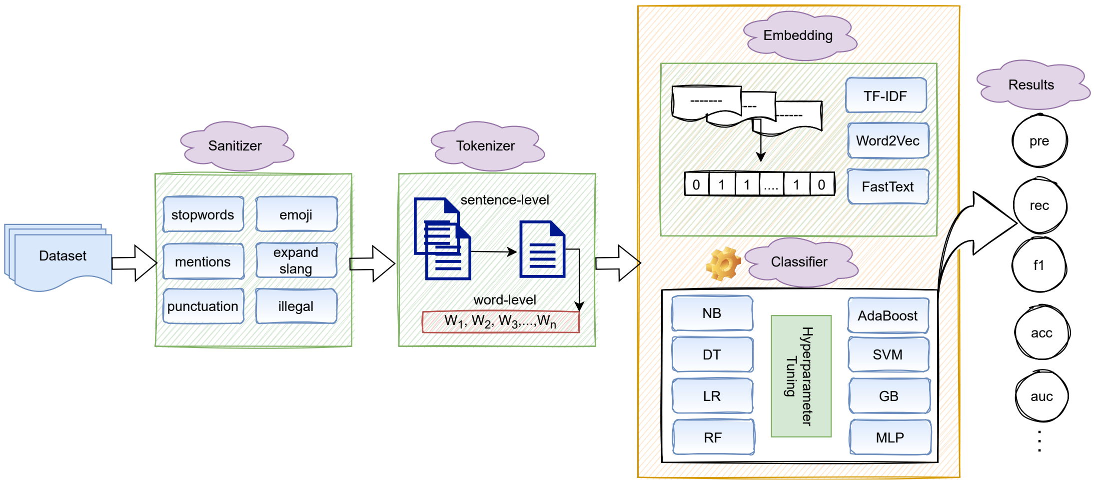

Fig. 1: The modular experimental setting with the flow of data from dataset to results.

Forest (RF), v) AdaBoost, vi) Support Vector Machine (SVM), vii) Gradient Boosting (GB), and viii) Multi-Layer Perceptron (MLP).

We also applied Bayesian optimization for the hyperparameter tuning. Bayesian optimization benefits the feedback of previous evaluations by choosing the classifier parameter combinations in an informed way. Furthermore, this approach takes fewer steps to converge to an optimal set of hyperparameter values by limiting the search space.

III. Dataset

Davidson et al.[2] has compiled a corpus containing around 24783 tweets annotated the text by crowd sourcing. This dataset represents three classes of labels as “hate speech”, “offensive language” and “neither”. They begin with a hate speech lexicon containing words and phrases identified by internet users as hate speech, compiled by Hatebase.org. Using the Twitter API they searched for tweets containing terms from the lexicon, resulting in a sample of tweets from 33,458 Twitter users. They extracted the time-line for each user, resulting in a set of 85.4 million tweets. From this corpus they then took a random sample of 25k tweets containing terms from the lexicon and had them manually coded by CrowdFlower (CF) workers. CF Workers were asked to label each tweet as one of the three categories.

Each data file contains 5 columns: Count, HateSpeech, Offensive_language, Neither, Class.

In this study, we consider ”HateSpeech” and ”Offensive_language” as offensive language. To be precise, we have 20620 tweets out of 24783 total tweets as offensive language(i.e., about 83% of tweets) which confirms that the dataset is highly imbalanced [71]–[75].

Also, Fig. 2 shows that offensive messages tends to be shorter than normal messages.

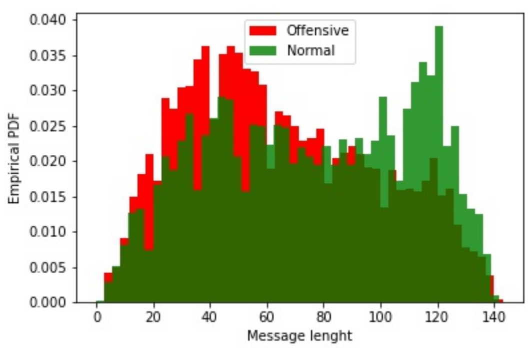

Fig. 2: Offensive/normal messages lengths

[2]https://github.com/t-davidson/hate-speech-and-offensive-language/tree/master/data

IV. Results and Discussions

This section reports the result of this study given the experimental setup described in Section II, which is also depicted in Fig 1. Table I demonstrates the results including precision, recall, F1-score, balanced accuracy, and AUC score of training binary classifiers over the dataset. It should be noted that the results are sorted based on F1-score in descending order.

TABLE I: Reporting performance metrics using eight classifiers and three embeddings.

classifier	*embedding*	*pre*	*rec*	*f1*	*acc*	*auc*
AdaBoost	TF-IDF	0.95	0.94	0.95	0.94	0.94
SVM	TF-IDF	0.95	0.95	0.95	0.95	0.93
MLP	TF-IDF	0.95	0.95	0.95	0.95	0.92
DT	TF-IDF	0.94	0.94	0.94	0.94	0.89
RF	TF-IDF	0.94	0.94	0.94	0.94	0.90
LR	TF-IDF	0.93	0.93	0.93	0.93	0.86
SVM	Word2Vec	0.92	0.93	0.93	0.93	0.85
MLP	Word2Vec	0.93	0.93	0.93	0.93	0.88
LR	Word2Vec	0.92	0.93	0.92	0.93	0.85
GB	TF-IDF	0.92	0.92	0.92	0.92	0.81
MLP	FastText	0.92	0.92	0.92	0.92	0.85
SVM	FastText	0.91	0.92	0.91	0.92	0.82
RF	Word2Vec	0.9	0.91	0.9	0.91	0.76
LR	FastText	0.9	0.91	0.9	0.91	0.79
GB	Word2Vec	0.9	0.91	0.9	0.91	0.79
AdaBoost	Word2Vec	0.89	0.9	0.89	0.9	0.78
AdaBoost	FastText	0.88	0.89	0.89	0.89	0.76
GB	FastText	0.89	0.9	0.89	0.9	0.76
RF	FastText	0.89	0.89	0.88	0.89	0.71
NB	Word2Vec	0.88	0.84	0.85	0.84	0.82
DT	Word2Vec	0.85	0.85	0.85	0.85	0.74
NB	FastText	0.86	0.84	0.85	0.84	0.78
DT	FastText	0.83	0.83	0.83	0.83	0.69
NB	TF-IDF	0.88	0.63	0.68	0.63	0.77

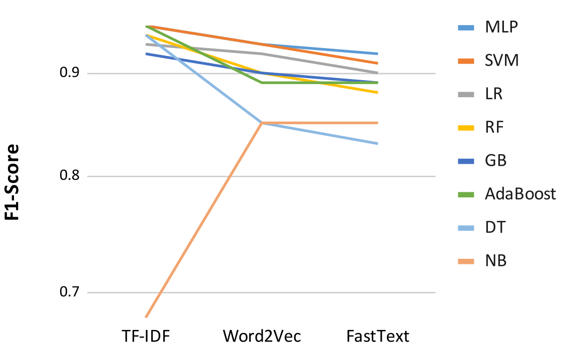

Fig. 3: F1-score trend line charts on different classifiers against each embedding method

As shown in Table I, the maximum obtained F1-score in the validation is 95% for Adaboost, SVM, and MLP classifiers on TF-IDF embedding. On the contrary, the lowest performance is 68%, which is the result of NB on TF-IDF. This suggests that while Adaboost, SVM, and MLP models perform better on TF-IDF embedding, a model like NB has a significantly lower score on this embedding approach. However, looking at Table I, NB is the worst classifier on all three embeddings. The reason might be that the NB algorithm assumes that features are conditionally independent given a class label but in practice, the independence assumption is often violated. The second worst classifier is DT on both Word2Vec and FastText embeddings although this classifier performs well on TF-IDF.

Figure 3 also shows the trend line of different classifiers against each embedding method. It clearly shows that although TF-IDF embeddings perform well for most of the classifiers, it has the lowest score when combined with NB classifier. Overall, NB and DT are the worst classifiers in this experiment. As mentioned in Section II, we used hyperparameter tuning in our experiment. For the MLP classifier, it is observed that the parameters *learning rate* and *activation* can affect the result positively if they were set to *Relu* and *Adaptive* respectively. Besides, for both DT and GB, the parameters *Minimum Sample Split* $= 2$ resulted in better results than *Minimum Sample Split* $= 5$.

V. Conclusions

In this work, we propose a modular text classification pipeline on social media datasets focusing on Twitter. Our proposed approach is to leverage a modular development that allows easy use for combining different text classification components. This paper's main contribution is that it presents a new modular text classification pipeline to facilitate benchmarking by conducting a detailed analytical study of the best-performing approaches, features, and embeddings reported by the state-of-the-art.

References

[1] N. Sabetpour, A. Kulkarni, and Q. Li, "OptSLA: an optimization-based approach for sequential label aggregation," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics,

Nov. 2020, pp. 1335–1340. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.119

[2] N. Sabetpour, A. Kulkarni, S. Xie, and Q. Li, “Truth discovery in sequence labels from crowds,” 2021.

[3] S. Khorshidi, G. Mohler, and J. G. Carter, “Assessing gan-based approaches for generative modeling of crime text reports,” in *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*. IEEE, 2020, pp. 1–6.

[4] S. Khorshidi, M. Al Hasan, G. Mohler, and M. B. Short, “The role of graphlets in viral processes on networks,” *Journal of Nonlinear Science*, vol. 30, no. 5, pp. 2309–2324, 2020.

[5] O. Jafari, P. Nagarkar, B. Thatte, and C. Ingram, “Satellitener: An effective named entity recognition model for the satellite domain,” in *Proceedings of the KMIS 2020*, vol. 3, 2020, pp. 100–107.

[6] S. Zhang, O. Jafari, and P. Nagarkar, “A survey on machine learning techniques for auto labeling of video, audio, and text data,” *arXiv preprint arXiv:2109.03784*, 2021.

[7] M. Saadati, J. Nelson, A. Curtin, L. Wang, and H. Ayaz, “Application of recurrent convolutional neural networks for mental workload assessment using functional near-infrared spectroscopy,” in *Advances in Neuroergonomics and Cognitive Engineering*, H. Ayaz, U. Asgher, and L. Paletta, Eds. Cham: Springer International Publishing, 2021, pp. 106–113.

[8] N. Kalantari, D. Liao, and V. G. Motti, “Characterizing the online discourse in twitter: Users’ reaction to misinformation around covid-19 in twitter,” in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 4371–4380.

[9] A. Esmaeilzadeh, M. Heidari, R. Abdolazimi, P. Hajibabaee, and M. Malekzadeh, “Efficient large scale nlp feature engineering with apache spark,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022.

[10] R. Abdolazimi, M. Heidari, A. Esmaeilzadeh, and H. Naderi, “Mapreduce preprocess of big graphs for rapid connected components detection,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022.

[11] M. Malekzadeh, P. Hajibabaee, M. Heidari, and B. Berlin, “Review of deep learning methods for automated sleep staging,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022.

[12] M. Heidari, J. H. J. Jones, and O. Uzuner, “An empirical study of machine learning algorithms for social media bot detection,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–5.

[13] J. Liu, M. Malekzadeh, N. Mirian, T. Song, C. Liu, and J. Dutta, “Artificial intelligence-based image enhancement in pet imaging: Noise reduction and resolution enhancement,” *PET clinics*, vol. 16, no. 4, pp. 553–576, 2021.

[14] D. Bamgboje, I. Christoulakis, I. Smanis, G. Chavan, R. Shah, M. Malekzadeh, I. Violaris, N. Giannakeas, M. Tsipouras, K. Kalafatakis *et al.*, “Continuous non-invasive glucose monitoring via contact lenses: Current approaches and future perspectives,” *Biosensors*, vol. 11, no. 6, p. 189, 2021.

[15] M. Malekzadeh, T. Song, and J. Dutta, “Pet image denoising using unsupervised domain translation,” in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2021.

[16] Y. Soofi and M. Bitaraf, “Output-only entropy-based damage detection using transmissibility function,” *Journal of Civil Structural Health Monitoring*, pp. 1–15, 2021.

[17] P. Hajibabaee, F. Pourkamali-Anaraki, and M. Hariri-Ardebili, “An empirical evaluation of the t-sne algorithm for data visualization in structural engineering,” in *2021 IEEE International Conference on Machine Learning and Applications*. IEEE, 2021.

[18] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, “Emotion detection of textual data: An interdisciplinary survey,” in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0255–0261.

[19] S. Zad, M. Heidari, J. H. Jones, and O. Uzuner, “A survey on concept-level sentiment analysis techniques of textual data,” in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0285–0291.

[20] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, “Ontology creation model based on attention mechanism for a specific business domain,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–5.

[21] M. Heidari, S. Zad, and S. Rafatirad, “Ensemble of supervised and unsupervised learning models to predict a profitable business decision,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–6.

[22] M. Heidari and S. Rafatirad, “Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment,” in *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 2020, pp. 322–329.

[23] S. H. Gheshlaghi, C. N. E. Kan, and D. H. Ye, “Breast cancer histopathological image classification with adversarial image synthesis,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 3387–3390.

[24] R. Abdolazimi, H. Naderi, and M. Sagharichian, “Connected components of big graphs in fixed mapreduce rounds,” *Cluster Computing*, vol. 20, no. 3, pp. 2563–2574, 2017.

[25] M. Torkjazi and H. Fazlollahtabar, “A fuzzy probabilistic maximum technique to optimize an unconstrained utility based multi objective model,” *Industrial Engineering and Management*, vol. 4, no. 147, pp. 2169–0316, 2015.

[26] S. H. Gheshlaghi, O. Dehzangi, A. Dabouei, A. Amireskandari, A. Rezai, and N. M. Nasrabadi, “Efficient oct image segmentation using neural architecture search,” in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 428–432.

[27] M. Heidari, J. H. Jones, and O. Uzuner, “Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter,” in *2020 International Conference on Data Mining Workshops (ICDMW)*, 2020, pp. 480–487.

[28] M. Heidari and J. H. Jones, “Using bert to extract topic-independent sentiment features for social media bot detection,” in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2020, pp. 0542–0547.

[29] S. H. Gheshlaghi, A. Ranjbar, A. A. Suratgar, M. B.

Menhaj, and F. Faraji, “A superpixel segmentation based technique for multiple sclerosis lesion detection,” *arXiv preprint arXiv:1907.03109*, 2019.
[30] A. Esmaeilzadeh and K. Taghva, “Text classification using neural network language model (nnlm) and bert: An empirical comparison,” in *Proceedings of SAI Intelligent Systems Conference*. Springer, 2021, pp. 175–189.
[31] M. Esmail Zadeh Nojoo Kambar, A. Esmaeilzadeh, Y. Kim, and K. Taghva, “A survey on mobile malware detection methods using machine learning,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (IEEE CCWC 2022)*, virtual, USA, Jan. 2022.
[32] A. Esmaeilzadeh, J. R. Fonseca Cacho, K. Taghva, M. E. Z. N. Kambar, and M. Hajiali, “Building wikipedia n-grams with apache spark,” in *Intelligent Computing*. Springer, 2022.
[33] M. Heidari, “Nlp approach for social media bot detection(fake identity detection) to increase security and trust in online platforms,” 2022.
[34] O. Dehzangi, S. H. Gheshlaghi, A. Amireskandari, N. M. Nasrabadi, and A. Rezai, “Oct image segmentation using neural architecture search and srgan,” in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 6425–6430.
[35] M. E. Z. N. Kambar, P. Nahed, J. R. F. Cacho, G. Lee, J. Cummings, and K. Taghva, “Clinical text classification of alzheimer’s drugs’ mechanism of action,” in *Proceedings of Sixth International Congress on Information and Communication Technology*. Springer, 2022, pp. 513–521.
[36] A. Esmaeilzadeh, “A test driven approach to develop web-based machine learning applications.” UNLV Theses, Dissertations, Professional Papers, and Capstones, 2017. [Online]. Available: http://dx.doi.org/10.34917/11889688
[37] M. Mostofa, S. Mohamadi, J. Dawson, and N. Nasrabadi, “Deep gan-based cross-spectral cross-resolution iris recognition,” *arXiv preprint arXiv:2108.01569*, 2021.
[38] M. TORKJAZI and H. FAZLOLLAHTABAR, “Applying a utility based fuzzy probabilistic α-cut method to optimize a constrained multi objective model.”
[39] P. Hajibabaee, M. Malekzadeh, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, “An empirical study of the graphsage and word2vec algorithms for graph multiclass classification,” in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0515–0522.
[40] M. Torkjazi, H. Fazlollahtabar, and I. Mahdavi, “Optimizing an unconstrained multi objective model using a utility based fuzzy probabilistic α-cut method,” *Journal of Intelligent & Fuzzy Systems*, vol. 26, no. 6, pp. 2927–2936, 2014.
[41] M. Heidari and J. H. J. Jones, “Misinformation detection model to prevent spread of the covid-19 virus during the pandemic,” 2022.
[42] S. Mohamadi and H. Amindavar, “Deep bayesian active learning, a brief survey on recent advances,” *arXiv preprint arXiv:2012.08044*, 2020.
[43] R. Abdolazimi and R. Zafarani, “Noise-enhanced unsupervised link prediction,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2021, pp. 472–487.
[44] M. Heidari and J. H. J. Jones, “Fraud detection to increase customer trust in online shopping experience,” 2022.
[45] S. Mohamadi, D. Adjeroh, B. Behi, and H. Amindavar, “A new framework for spatial modeling and synthesis of genomic sequences,” in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 2221–2226.
[46] M. Heidari, J. H. J. Jones, and O. Uzuner, “An empirical study of machine learning algorithms for social media bot detection,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–5.
[47] M. Ahmadi and B. Ghahfarokhi, “Preserving privacy in location based mobile coupon systems using anonymous authentication scheme,” in *2016 13th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC)*. IEEE, 2016, pp. 60–65.
[48] ——, “Preserving privacy in location based mobile coupon systems using anonymous authentication scheme,” in *2016 13th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC)*. IEEE, 2016, pp. 60–65.
[49] M. Ahmadi, “Hidden fear: Evaluating the effectiveness of messages on social media,” Ph.D. dissertation, Arizona State University, 2020.
[50] S. Zad, M. Heidari, P. Hajibabaee, and M. Malekzadeh, “A survey of deep learning methods on semantic similarity and sentence modeling,” in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0466–0472.
[51] M. Ahmadi, K. Leach, R. Dougherty, S. Forrest, and W. Weimer, “Mimosa: Reducing malware analysis overhead with coverings,” *arXiv preprint arXiv:2101.07328*, 2021.
[52] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, “Review of graph neural network in text classification,” in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2021, pp. 0084–0091.
[53] M. Ahmadi, P. Kiaei, and N. Emamdoost, “Sn4ke: Practical mutation testing at binary level,” *arXiv preprint arXiv:2102.05709*, 2021.
[54] M. Heidari, S. Zad, P. Hajibabaee, M. Malekzadeh, S. HekmatiAthar, O. Uzuner, and J. H. Jones, “Bert model for fake news detection based on social bot activities in the covid-19 pandemic,” in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2021, pp. 0103–0109.
[55] H. Ahmadian, MM.and Shahriari, M. Shafiee, and M. Ahmadi, “High survivable ransomwares detection using bayesian belief network and hybrid behavioral analysis features.”
[56] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, “Emotion detection of textual data: An interdisciplinary survey,” in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0255–0261.
[57] Y. Hsu, C. Jeng, P. Murali, M. Torkjazi, J. West, M. Zuber, and V. Sokolov, “Bayesian learning: A selective overview,” *arXiv preprint arXiv:2112.12722*, 2021.
[58] S. Hajiseyedjavadi, Y. Lin, and K. Pelechrinis, “Learning embeddings for multiplex networks using triplet loss,” *Applied Network Science*, vol. 4, no. 1, pp. 1–16, 2019.
[59] R. Abdolazimi, S. Jin, and R. Zafarani, “Noise-enhanced community detection,” in *Proceedings of the 31st ACM*

Conference on Hypertext and Social Media, 2020, pp. 271–280.
[60] M. Heidari and J. H. J. Jones, “Offensive behaviour detection on social media platforms by using natural language processing models,” 2022.
[61] O. Jafari, P. Nagarkar, and J. Montaño, “Improving locality sensitive hashing by efficiently finding projected nearest neighbors,” in *International Conference on Similarity Search and Applications*. Springer, 2020, pp. 323–337.
[62] G. Espinoza, H. Schacter, and J. Juvonen, “School and cybervictimization across high school: Normative developmental trajectories and bidirectional links with loneliness,” *Journal of Research on Adolescence*, vol. 30, no. 3, pp. 787–799, 2020.
[63] J. Patchin and S. Hinduja, “Cyberbullying among tweens in the united states: Prevalence, impact, and helping behaviors,” *The Journal of Early Adolescence*, p. 02724316211036740, 2021.
[64] H. Watanabe, M. Bouazizi, and T. Ohtsuki, “Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection,” *IEEE access*, vol. 6, pp. 13 825–13 835, 2018.
[65] Chantelmariediaz, “Predicting-cyberbulling-on-twitter.” [Online]. Available: https://github.com/chantelmariediaz/Predicting-Cyberbulling-on-Twitter
[66] R. Kumar, A. Reganti, A. Bhatia, and T. Maheshwari, “Aggression-annotated corpus of hindi-english code-mixed data,” *arXiv preprint arXiv:1803.09402*, 2018.
[67] Z. Waseem and D. Hovy, “Hateful symbols or hateful people? predictive features for hate speech detection on twitter,” in *Proceedings of the NAACL student research workshop*, 2016, pp. 88–93.
[68] J. Qian, A. Bethke, Y. Liu, E. Belding, and W. Y. Wang, “A benchmark dataset for learning to intervene in online hate speech,” *arXiv preprint arXiv:1909.04251*, 2019.
[69] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
[70] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” *arXiv preprint arXiv:1607.01759*, 2016.
[71] P. Hajibabaee, F. Pourkamali-Anaraki, and M. Hariri-Ardebili, “Kernel matrix approximation on class-imbalanced data with an application to scientific simulation,” *IEEE Access*, pp. 83 579–83 591, 2021.
[72] ——, “An empirical evaluation of the t-sne algorithm for data visualization in structural engineering,” *arXiv preprint arXiv:2109.08795*, 2021.
[73] P. Hajibabaee, “Transit agencies performance assessment and implications,” 2020.
[74] S. Enayati, Z. Yang, B. Lu, and S. Vucetic, “A visualization approach for rapid labeling of clinical notes for smoking status extraction,” in *Proceedings of the Second Workshop on Data Science with Human in the Loop: Language Advances*, 2021, pp. 24–30.
[75] H. Nazari, B. Abedin, P. Hajibabaee, and E. Fallah, “Identifying and ranking critical factors that determine meritocracy using ahp technique in automotive industry in iran,” in *International Conference on Technology and Business Management March*, vol. 24, 2014, p. 26.