Bert Model for Social Media Bot Detection

Maryam Heidari
George Mason University

James H Jr Jones
George Mason University

***Abstract*—Millions of online posts about different topics and products are shared on popular social media platforms. One use of this content is to provide crowd-sourced information about a specific topic, event, or product. However, this use raises an important question: what percentage of the information available through these services is trustworthy? In particular, might some of this information be generated by a machine, i.e., a "bot" instead of a human? Bots can be, and often are, purposely designed to generate enough volume to skew an apparent trend or position on a topic, yet the consumer of such content cannot easily distinguish a bot post from a human post. This paper introduces a new model that uses Bidirectional Encoder Representations from Transformers (Google Bert) for sentiment classification of tweets to identify topic-independent features for the social media bot detection model. Using a Natural Language Processing approach to derive topic-independent features for the new bot detection model distinguishes this work from previous bot detection models. We achieve 94% accuracy classifying the contents of data set Cresci [1]as generated by a bot or a human, where the most accurate prior work achieved an accuracy of 92%.**

***Index Terms*—Bot detection, Natural language processing, Neural Network. Social media**

I. Introduction

Social media is a largely unregulated place to exchange information, so it is of great importance to differentiate between human posts and posts generated by a computer program, i.e., a bot. Bots intentionally spread information on these platforms to create a specific trend that may affect the public opinion [2]. Social media platforms are a rich source of user content and user's personal information can be disclosed in these platforms [3] and can be misused by social bots or Online fake identities to pose serious threats to financial organizations. Bots spread low-quality information, disinformation, and even misleading news, which can be hard to detect based on content. Further, bots can generate large volumes of content in online blogs from apparently multiple sources, thereby influencing quantitative sentiment and makes it hard for topic detection in online communities [**?**].

The study conducted by Cresci et al. [1], shows that Social media bots can target various audiences. They identified multiple types of bots, including promoter bots, URL spam bots, and fake followers. Promoter bots spend several months promoting specific hashtags to create fake trends and promote a specific event or product [1]. For example, they can promote a product on Amazon or help a political campaign win an election [4]. URL spam bots spread scam URL links by embedding these malicious links in retweets of legitimate users [5]. Based on Howard's study [4], URL sharing bots are used for constant tweet duplication of legitimate users in a specific time to spread malicious URLs. Their results suggest that about 23 % of accounts that use URL shortening services are bots. Other types of bots include fake followers on social media and fake reviewers of specific products [1]. The prevalence and influence of bots in recent years has led to multiple studies on detecting them. One popular bot detection service is "Botometer," a supervised learning approach to detect social bots [6]. Botometer uses metadata related to each twitter account, such as network features, user features, and temporal features, to feed a Random Forest classifier algorithm. Network features show how information diffusion happens among multiple groups of users. User features include a user name, screen name, the creation time of account and geographic location, and temporal features show patterns in a tweet's timeline. The networked nature of Twitter has led to graph-based approaches for bot detection, e.g., using fully-connected nodes among a group of Twitter accounts to detect bots [7]. One study shows that Benford's Law can detect online social bot behaviours [8], and it can be used to track bot activities on social media platforms.

Misinformation purpose is mostly to destroy a country's trust in their government, or to tamper with its economy, or to gain benefits from selling a product that gets more expensive, as these happened during the corona pandemic [**?**], [9]–[20].

Machine learning models have different applications in health, cyber security, business and social computing research [10], [14], [21]–[32] [33]–[47]. Dickerson [48] has studied the importance of sentiment features in improving prediction accuracy for bot detection models. SentiBot, their Sentiment-aware architecture on the 2014 Indian election data set, suggests that the sentiment features can play a significant role in the bot detection model's prediction accuracy. SentiBot defines sentiment features based on selecting the tweets related to the topic of interests(TOI) using Latent Dirichlet Allocation(LDA topic modeling). However, the importance of sentiment features in bot detection models is not the point of the work presented in this paper. This work's main idea is to use BERT for feature extraction and improve the prediction accuracy of the bot detection model in social media platforms.

TABLE I
CRESCI 2017 DATASET DESCRIPTION

user type	tweet count	user count
genuine	2839361	3474
social spam bot #1	1610034	991
social spam bot #2	428542	3457
social spam bot #3	1418557	464
traditional spam bot	145094	999
fake followers	196027	3351

In this paper, we use Bidirectional Encoder Representations from Transformers(BERT) [49] for sentiment classification of tweets into two categories of positive and negative. Before applying BERT on data, we do not use LDA topic modeling to categorize tweets into specific topics or filter them. As a result, our new method and extracted features are topic-independent, and the final classifier is applied to all tweets regardless of the tweet's topic. In this work, the bidirectional training of the transformer BERT offers a popular attention model to classify online posts into negative and positive posts. Then topic-independent sentiment features extracted from the tweet's text using BERT will be an input to feed the Neural network in the new bot detection model's final step to predict bots on the social media platform. This work examines several supervised learning models' accuracy, including SVM, Random Forest, Logistic regression, and the Neural network model, to choose the final classifier for the new bot detection model final phase. So the new proposed model has a neural network architecture in its final step since it provides the best prediction accuracy. The last section of this work evaluates the proposed new bot detection model compared with existing state-of-the-art bot technology techniques in the same training and test data set of Cresci [50].

II. DATASET

The Cresci 2017 data set [1], a labeled data set of bots and human accounts on Twitter, is used for this research. There is one category of human users ("genuine"), and five categories of bots: Social Spam bot #1, #2, #3, fake followers, and traditional spambots. Social Spam bot #1 are automated accounts that are from a 2014 election in Rome. Social spam bot #2 are promoter bots that spend several months on promoting specific hashtags on Twitter. Social spam bot #3 are Amazon accounts that share spam URLs on Amazon pointing to their products. Traditional spam bots are Twitter spammer accounts from the Cresci 2013 data set [51]. Cresci authors bought Fake followers of Twitter accounts from different websites [1]. Table I shows the statistics of the Cresci 2017 data set. This data set includes the following attributes: follower-count, friends-count, retweet-count, reply-count, number of hashtags, number of shared URL, tweet's text, screen name, and user ID. Figure 1 shows the distribution of bots in this data set.

III. METHOD

For sentiment classification on the Cresci data set, one approach can be using a feature-based solution such as ELMO

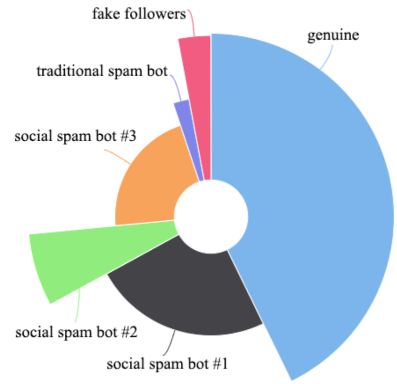

Fig. 1. Cresci dataset [1]

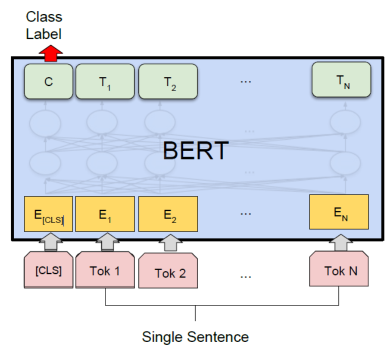

Fig. 2. single sentence classification task SST-2 [49]

[52], which is a deep contextual word representation. Another solution is a fine-tuning approach, such as Generative pre-trained transformer(GPT) [53], but the results for sentiment classification of tweets are not satisfactory in comparison with using BERT [49]. The main reason is that ELMO and GPT both useful in creating language representation, but BERT is a pre-training of a deep bidirectional transformer, a powerful model for language understanding. This paper uses a BERT model in Figure 2. We use BERT for single sentence classification task based on SST-2 [54]. SST-2 is Stanford Sentiment Treebank of binary classification of single sentences based on movie reviews.

The main advantage of introducing this model for tweet sentiment classification in this paper is that BERT has an ability to train unlabeled data based on fine-tuning on specific task. This characteristic plays a significant role in social media platforms, which generate a high volume of unlabeled tweets' texts every second, and it is a labor-intensive task to use human annotators to label real-time data stream.

A. New bot detection model

Figure 3 shows the architecture of the new proposed bot detection model by this research. The first component of the model is the Tweet sentiment classification. In this work, at the first phase of a model, BERT classifies tweets into positive and negative. At first, we trained BERT based on the 50,000 movie reviews [55]. In this paper, the BERT model

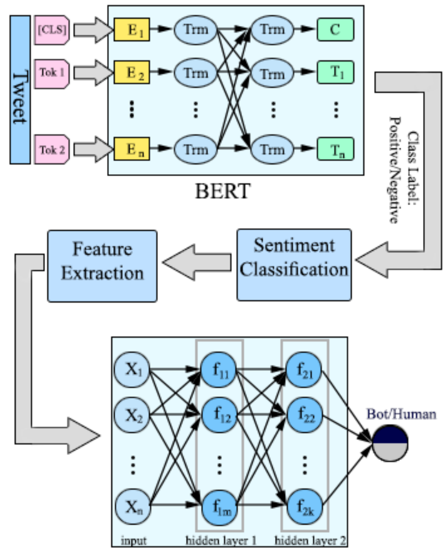

Fig. 3. New bot detection model

TABLE II
TOPIC-INDEPENDENT SENTIMENT FEATURES

New Features	Description
Count-Neutral	Count of neutral tweets for a user
Count-Positive	Count of positive tweets for a user
Count-Negative	Count of negative tweets for a user
Sum-Positive	Sum polarity scores of positive tweets which are posted by a user
Sum-Negative	Sum polarity scores of negative tweets which are posted by a user
Average-Positive	$Sum - Positive/Count - Positive$
Average-Negative	$Sum - Negative/Count - Negative$

uses the Stanford Sentiment Treebank $SST2$ [49]. SST2 is a single-sentence classification task based on sentences extracted from movie reviews labeled by human annotators. Stanford Sentiment Treebank includes 215,154 phrases with sentiment labels in the parse trees of 11,855 sentences based on movie reviews. The raw data in SST-2 is in the form of tree structure which is created to train recursive neural tensor network and each phrase has a recursive tree structure [54]. A deep neural network is trained on the tree structure of each sentence and each phrase sentiment is based on cumulative sentiments of all sentences in the specific phrase. The flow of information in the BERT model is as follows: each word is indicated as a embedding representation in the embedding layer which is the first layer in the figure 2. To create intermediate representation of each layer, the next layer apply multi-head attention computation of each word on the previous layer. So as can be seen in the first component of our new model in BERT, E_1 is the embedding representation, Trm is the intermediate representation of each token in each layer and T_1 is the final output. The final hidden state of [CLS] token is the representation of the input sequence. Using movie reviews in the first phase of the new model in this work is a new approach to classify tweets into positive and negative for bot detection tasks. Based on the sentiment classification phase, sentiment features, shown in Table II, are extracted for all users based on their tweet's text without considering the tweet topic. These features are completely topic-independent, and this work does not use LDA topic modeling to define topics on the tweets or select tweets with specific topics before the sentiment classification phase.

Then the fine-tuned BERT is used to extract sentiment features from a tweet's text. Tweets are not selected based on specific topics, and the features are extracted for all tweets related to a specific user regardless of the tweet's topic. It means that if one user tweets about politics or posts comment about specific beauty products, the BERT model in this work extracts the tweet's sentiment for all user's online posts regardless of tweets' topics, and all sentiment features are topic-independent. Using Topic-independent sentiment features in this work is one of the differences between previous bot detection models and our new proposed social media bot detection model. Using BERT, we assign a polarity score, which classifies each tweet's text into three different categories: positive, negative, and neutral. The polarity assigns X score for each tweet's text such that $-0.01 < X < 0.01$, the tweet is a neutral comment. If $0.01 < X$ means a positive comment, and $X < -0.01$ means a negative comment.After sentiment classification, sentiment features are extracted from tweets' text for each user's account based on BERT's polarity score. Table II shows a new selected set of sentiment features.

TABLE III
ALGORITHM PERFORMANCE

algorithm	accuracy	f1 score	mcc
FFNN	0.949	0.947	0.912
Random Forest	0.923	0.912	0.887
SVM	0.899	0.930	0.860
Logistic Regression	0.874	0.888	0.685

Figure 4 shows BERT results for Sentiment distribution of bots and human comments; It seems that bot and human account tweet's sentiment is similar. However, the frequency of posting positive or negative tweets is different between bot and human users. Removing network features in the new bot detection model saves significant time and memory space compared to previous bot detection models. Also, removing network features for each user enables the new bot detection model to detect bot accounts based on limited information for each online user.

In Figure 3, in the last phase of the new model, the final classifier detects bot accounts just based on the features in Table II and the tweet's text. We use GloVe [56] for the word embedding of all tweet's text. This research uses Glove(Global Vectors) Also, the statistics of the global English corpus for each token in a tweet's text can be represented by GloVe in the word embedding phase. The final classifier should provide the best prediction accuracy based on BERT's extracted features in the previous phase. Table III shows the performance of Random Forest, SVM, Logistic regression, and Neural network for bot detection based on Table II feature set. In addition to the accuracy, we also measure the Matthews correlation coefficient(MCC) since it provides a more precise explanation of the relationship between TP, TN, FP, and FN to evaluate algorithm performance. MCC is calculated as follow:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1)$$

This work uses a grid search to define the best initial setting for different types of hyperparameters for each classifier, such as the maximum depth of a tree, Kernel type, number of samples per split. The grid search used ten-fold cross-validation to fit the model. Then the set of hyperparameters is selected with the maximum recall as a setting for the model.

Table III shows prediction accuracy of each classifier. Feed-forward neural network(FFNN) outperformed other classifiers by achieving an accuracy of 94%. However, the Random forest ranked second by providing 92% in prediction accuracy in bot detection. Random Forest and SVM provide very close results in bot detection, But Random Forest outperformed SVM. Logistic regression performance is low in prediction accuracy, F1 score, and MCC compared with bot detection classifiers. This work, based on Table III results, for the final phase of the new bot detection model in Figure 3, choose the Feedforward neural network mode.

IV. RESULTS AND ALGORITHM EVALUATION

This section evaluates a new model for social media bot detection in this work compared with previously well-known bot detection techniques. To examine the new model's performance, we test the new model based on the same training and test data set used by the previous bot detection techniques. Test set # 1which is the mix of human and bot accounts in political campaign, and test set #2, a combination of human and spambots accounts in Amazon. Table IV shows a comparative analysis with previous bot detection techniques. In the test set, #1, which is related to the political campaign bots, the Cresci bot detection technique ranked first in the f-measure between all bot detection techniques. Also, for political campaign bots, Cresci, and our new bot detection model provide very close accuracy prediction; however, our model outperformed the previous bot detection techniques in MCC and prediction accuracy. As can be seen, our new proposed model outperformed previous techniques by achieving more than 94% prediction accuracy in test set #1 and test set #2. In this work, a new proposed model for social media bot detection extracts topic-independent sentiment features using a natural language processing approach for language understanding and sentiment classification of tweets. The new model detects any anomaly on the tweets posted by a single user in an early stage since the model does not use many tweets about a single topic or network features related to the online user. The new model considers the tweets' overall sentiment behavior related to the online user based on limited tweets. Using Movie reviews for social media bot detection in our new model helps us better understand social bots' online behavior in posting tweets with specific sentiments.

The new introduced model performance can be affected if the online user's tweets include slang related to a specific online community or if the tweet's text consists of minimal words. Consequently, algorithm performance can be affected since the embedding phase can not generate accurate results for the input space of the neural network model.

V. CONCLUSION

The main contributions of this work can be summarized in three major points. First, Using BERT for sentiment classification of Cresci data set into positive, negative,and neutral tweets. This research leads to creating a new public data-set of more than four million tweets based on BERT sentiment classification, which is complementary to the previous Cresie 2017 data set. The new public data set can be used in social media research, specifically in social media bot detection research. The second contribution is identifying a set of features that have a major role in bot detection accuracy, and more importantly, they can be used without extra information and network features for each online user. The new method can detect bots based on a limited number of tweets and features for each online user. Third, providing the new model for bot detection on Twitter with high prediction accuracy. The new proposed model outperformed previous bot detection models by extracting topic-independent sentiment features based on

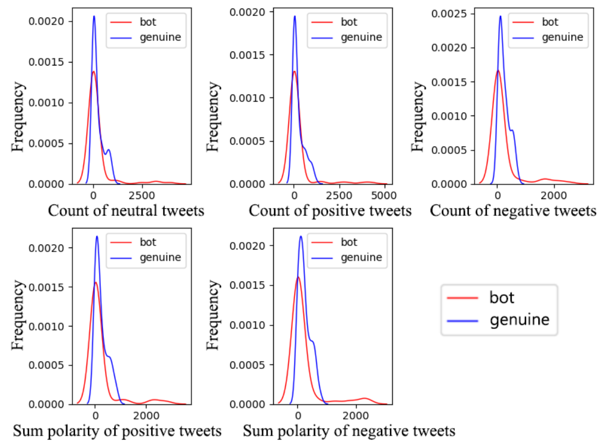

Fig. 4. Sentiment distribution of tweets

TABLE IV
PERFORMANCE COMPARISON AMONG PREVIOUS BOT DETECTION TECHNIQUES AND THE NEW BOT DETECTION MODEL [50]

technique	**type**	Accuracy	F-Measure	MCC
test set #1				
Davis et al. [6]	supervised	0.734	0.288	0.174
C. Yang et al. [51]	supervised	0.506	0.261	0.043
Miller et al. [57]	unsupervised	0.526	0.435	0.059
Ahmed et al. [58]	unsupervised	0.943	0.944	0.886
Cresci et al. [50]	unsupervised	0.976	**0.977**	0.952
Our new model	supervised	**0.979**	0.976	**0.962**
test set #2				
Davis et al. [6]	supervised	0.922	0.761	0.738
C. Yang et al. [51]	supervised	0.629	0.524	0.287
Miller et al. [57]	unsupervised	0.481	0.370	-0.043
Ahmed et al. [58]	unsupervised	0.923	0.923	0.847
Cresci et al. [50]	unsupervised	0.929	0.923	0.867
Our new model	supervised	**0.948**	**0.942**	**0.891**

the BERT sentiment classification and applying new features to the Neural network model to detect social bots. The new proposed model takes advantage of BERT's transfer learning approach and uses labeled data of Movie reviews to detect bot in social media; It uses a new way to use labeled data from other online platforms to detect bots in a specific platform of Twitter.

REFERENCES

[1] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*, pp. 963–972, 2017.

[2] M. Forelle, P. N. Howard, A. Monroy-Hernández, and S. Savage, "Political bots and the manipulation of public opinion in venezuela," *CoRR*, vol. abs/1507.07109, 2015.

[3] A. Mosallanezhad, G. Beigi, and H. Liu, "Deep reinforcement learning-based text anonymization against private-attribute inference," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, (Hong Kong, China), pp. 2360–2369, Association for Computational Linguistics, Nov. 2019.

[4] P. N. Howard, S. Woolley, and R. Calo, "Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration," *Journal of Information Technology & Politics*, 2018.

[5] Z. Chen and D. Subramanian, "An unsupervised approach to detect spam campaigns that use botnets on twitter," *CoRR*, vol. abs/1804.05232, 2018.

[6] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume*, pp. 273–274, 2016.

[7] W. M. Campbell, C. K. Dagli, and C. J. Weinstein, "Social network analysis with content and graphs," *Lincoln Laboratory Journal*, 2013.

[8] L. Madahali and M. Hall, "Application of the benford's law to social bots and information operations activities," in *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, pp. 1–8, 2020.

[9] 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)*, pp. 0466–0472, 2021.

[10] S. Zad, J. Jimenez, and M. Finlayson, "Hell hath no fury? correcting bias in the nrc emotion lexicon," in *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pp. 102–113, 2021.

[11] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting misleading information on COVID-19," vol. 8, pp. 165201–165215, 2020.

[12] 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)*, pp. 480–487, IEEE, 2020.

[13] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop*

on Semantic and Social Media Adaptation and Personalization (SMA, pp. 1–6, IEEE, 2020.

[14] S. Zad and M. Finlayson, “Systematic evaluation of a framework for unsupervised emotion recognition for narrative text,” in *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pp. 26–37, Association for Computational Linguistics, 2020.

[15] M. Heidari and S. Rafatirad, “Semantic convolutional neural network model for safe business investment by using bert,” in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1–6, IEEE, 2020.

[16] 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)*, pp. 1–6, IEEE, 2021.

[17] 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)*, pp. 0103–0109, 2021.

[18] 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)*, pp. 0285–0291, IEEE, 2021.

[19] S. Zad, J. Jimenez, and M. A. Finlayson, “The abbe corpus: Animate beings being emotional,” *ArXiv*, vol. abs/2201.10618, 2022.

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

[21] M. Heidari, H. James Jr, 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)*, pp. 1–5, IEEE, 2021.

[22] R. Abdolazimi, S. Jin, and R. Zafarani, “Noise-enhanced community detection,” in *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pp. 271–280, 2020.

[23] 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)*, pp. 1–5, IEEE, 2021.

[24] 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)*, pp. 322–329, IEEE, 2020.

[25] K. O’Haire, S. Soleymani, B. Chaudhary, P. Aghdaie, J. Dawson, and N. M. Nasrabadi, “Adversarially perturbed wavelet-based morphed face generation,” in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pp. 01–05, IEEE, 2021.

[26] A. Esmaeilzadeh, “A test driven approach to develop web-based machine learning applications,” UNLV Theses, Dissertations, Professional Papers, and Capstones, 2017.

[27] P. Aghdaie, B. Chaudhary, S. Soleymani, J. Dawson, and N. M. Nasrabadi, “Morph detection enhanced by structured group sparsity,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 311–320, 2022.

[28] 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)*, IEEE, 2021.

[29] 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)*, IEEE, 2021.

[30] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, “Review of graph neural network in text classification,” in *2021 12th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, IEEE, 2021.

[31] M. Torkjazi, L. K. Farrokhvar, and B. Kamali, “Main Contributing Factors and the Heuristic Approach for Assessing Risk at Mass Gatherings,” *Operations Research Forum*, vol. 3, p. 13, Feb. 2022.

[32] P. Hajibabaee, “Transit agencies performance assessment and implications,” 2020.

[33] 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*, pp. 175–189, Springer, 2021.

[34] P. Aghdaie, B. Chaudhary, S. Soleymani, J. Dawson, and N. M. Nasrabadi, “Attention aware wavelet-based detection of morphed face images,” in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2021.

[35] 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)*, pp. 0274–0280, 2022.

[36] 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)*, pp. 0112–0118, 2022.

[37] 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)*, pp. 0080–0086, 2022.

[38] Y. L. Hsu, C. C. Jeng, P. S. Murali, M. Torkjazi, J. West, M. Zuber, and V. Sokolov, “Bayesian learning: A selective overview,” *arXiv preprint arXiv:2112.12722*, 2021.

[39] M. Torkjazi and H. Fazlollahtabar, “Applying a utility based fuzzy probabilistic α-cut method to optimize a constrained multi objective model,” *Creative Mathematics and Informatics*, vol. 24, no. 2, pp. 221–230, 2015.

[40] P. Hajibabaee, M. Malekzadeh, M. Ahmadi, M. Heidari, A. Esmaeilzadeh, R. Abdolazimi, and J. H. J. Jones, “Offensive language detection on social media based on text classification,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0092–0098, 2022.

[41] P. Aghdaie, B. Chaudhary, S. Soleymani, J. Dawson, and N. M. Nasrabadi, “Detection of morphed face images using discriminative wavelet sub-bands,” in *2021 IEEE International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, IEEE, 2021.

[42] 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.

[43] 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.

[44] 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.

[45] R. Abdolazimi and R. Zafarani, “Noise-enhanced unsupervised link prediction,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 472–487, Springer, 2021.

[46] 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*, pp. 513–521, Springer, 2022.

[47] 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.

[48] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, “Using sentiment to detect bots on twitter: Are humans more opinionated than bots?,” in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2014, Beijing, China, August 17-20, 2014* (X. Wu, M. Ester, and G. Xu, eds.), pp. 620–627, IEEE Computer Society, 2014.

[49] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)* (J. Burstein, C. Doran, and T. Solorio, eds.), pp. 4171–4186, Association for Computational Linguistics, 2019.

[50] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, “Dna-inspired online behavioral modeling and its application to spambot detection,” *IEEE Intell. Syst.*, vol. 31, no. 5, pp. 58–64, 2016.

[51] C. Yang, R. C. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," *IEEE Trans. Information Forensics and Security*, vol. 8, no. 8, pp. 1280–1293, 2013.
[52] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," 2018.
[53] A. Radford, "Improving language understanding by generative pre-training," 2018.
[54] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1631–1642, ACL, 2013.
[55] "Imdb dataset of 50k movie reviews."
[56] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation.," in *EMNLP*, vol. 14, pp. 1532–1543, 2014.
[57] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, "Twitter spammer detection using data stream clustering," *Inf. Sci.*, vol. 260, pp. 64–73, 2014.
[58] F. Ahmed and M. Abulaish, "A generic statistical approach for spam detection in online social networks," *Comput. Commun.*, vol. 36, no. 10-11, pp. 1120–1129, 2013.