Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers

Mohammed Boukabous, Mostafa Azizi
MATSI Research Lab, Ecole Supérieure de Technologie Oujda (ESTO), Mohammed First University, Oujda, Morocco

Article Info

ABSTRACT

Sentiment analysis (SA) is widely used today in many areas such as crime detection (security intelligence) to detect potential security threats in real-time using social media platforms such as Twitter. The most promising techniques in sentiment analysis are those of deep learning (DL), particularly bidirectional encoder representations from transformers (BERT) in the field of natural language processing (NLP). However, employing the BERT algorithm to detect crimes requires a crime dataset labeled by the lexicon-based approach. In this paper, we used a hybrid approach that combines both lexicon-based and deep learning, with BERT as the DL model. We employed the lexicon-based approach to label our Twitter dataset with a set of normal and crime-related lexicons; then, we used the obtained labeled dataset to train our BERT model. The experimental results show that our hybrid technique outperforms existing approaches in several metrics, with 94.91% and 94.92% in accuracy and F1-score respectively.

Corresponding Author:

Mohammed Boukabous
MATSI Research Lab, Ecole Supérieure de Technologie Oujda (ESTO), Mohammed First University
BP 473 complexe Universitaire Al Qods, Oujda 60000, Morocco
Email: m.boukabous@ump.ac.ma

1. INTRODUCTION

Security intelligence is nowadays one of the primary concerns of any country. It is an orientation of security techniques based on artificial intelligence that aims to collect and organize all data related to threats in cyberspace [1] to detect possible attacks in real-time and focus on identifying criminal characteristics [2]. This requires deeply identifying and analyzing both patterns and trends of crime [3]. Such profiling allows relevant protection and anticipating the various incidents by earlier detection of major attacks. In today's world, individuals, including cybercriminals, have become technically sophisticated and repeatedly expressing their emotions on the web, especially on popular social media websites [4]. The internet phenomenal growth has led more users to express their opinions online whether in daily chat (status messages about what the user is doing), conversations (tweeting to a user or group of users within a community), information sharing (posting links to web pages), or news reporting (updates on the current case).

Transfer learning has become one of the most used image classification approaches for reusing architectures and weights learned on huge datasets to enhance small and particular classification tasks [5]. A similar outcome may be achieved in natural language processing (NLP) by reusing and transferring a language model [6]. The bidirectional encoder representations from transformers (BERT) algorithm, in particular, has been demonstrated to perform well on a variety of English text categorization tasks, such as sentiment analysis [7]. Sentiment analysis using NLP and deep learning (DL) methods has recently been a

hot study subject. Emotions are one of the most important components of human existence that may aid in the identification of trends and user necessities.

DL is a subfield of machine learning that allows machines to automatically link processes together, by allowing several algorithms to be used progressively while passing from step to step, it can solve complex problems in almost the same way the human brain does. BERT is a model for learning linguistic representations that use the attention transformers method. A DL model designed to handle sequential data, to learn the word-to-word contextual relation in a text [8].

The rest of this paper is structured as follows. In section 2, we discuss relevant related works. Our methodology is presented in section 3. Experiments and results are described in section 4. Finally, section 5 concludes the study and suggests some areas for future research.

2. RELATED WORK

Ventirozos *et al.* [9] applied a lexicon-based approach to detect aggressive, antisocial, or inappropriate behavior, within the context of the discussion. The authors employ SA at the message level, but study the whole communication thread using a set of n-gram, which is used for classifying the whole thread as aggressive or neutral. The lexicon-based approach has the advantage to not require prior training (on a dataset) to mine the data but the final result can differ according to the context in which the lexicons were created.

Siriaraya *et al.* [10] applied the machine learning (ML) approach to create a crime-solving tool that gives context for criminal events. The authors utilize data from Twitter by providing contextual information about crime incidents occurring in a specific area (San Francisco) using machine learning classification models (logistic regression and support vector machine). The machine learning approach can adapt and create trained models for specific purposes and contexts, but it has low applicability on new data because it necessitates the availability of labeled data.

Pereira-Kohatsu *et al.* [11] applied a deep learning approach to develop a Twitter-based intelligence system for detecting and analyzing hate speech called HaterNet. The authors utilized a multilayer perceptron neural network that accepts as input the tweet's word, emoji, and expression embeddings tokens enhanced by the tf-idf and output the area under the curve (AUC).

BERT is one of the best DL algorithms in SA, as shown in [12]. Yadav *et al.* [13] used the BERT algorithm to identify cyberbullying on social media platforms, utilizing the BERT model as a classifier with a single linear neural network layer, trained and evaluated on two social media datasets, one small and one fairly big.

Weir *et al.* [14] proposed a system that combines both machine learning and lexicon-based approaches to detect terrorist web pages, gauge the strength of their content, and categorize data collected on terrorism and extremism networks. The authors developed an in-depth frequency study of the syntax using the Posit textual analysis toolkit, which included multi-word units and their related parts of speech. After that, utilizing knowledge extraction techniques, the findings are used in a knowledge extraction process (decision tree, random forest).

The above-mentioned related works can give a decent sentiment analysis prediction rate in security intelligence context. However, works that use a hybrid approach of both lexicon-based and deep learning approaches are rare, especially in this context. As a result, our research contributes to the above-mentioned works as an essential experimental expansion using a hybrid approach that combines both lexicon-based and deep learning approaches.

3. METHOD

Sentiment analysis has been practiced on a variety of topics like movie reviews, service and product reviews, news, blogs, chatbot, and security intelligence [7]. During an extreme incident, most of the messages shared on Twitter are commentary-related, such as expressing thoughts, reporting them, and attempting to understand what happened and why. Sentiment analysis research has mainly focused either on identifying whether a given textual entity is objective or subjective or on identifying the polarity of subjective texts [15].

Our case studies are examples of severe occurrences brought on by individual choices that had a physical impact on individuals, and therefore these events would have multiple social media communications. Sentiment analysis (or opinion mining) is employed on Twitter by the means of one of the following techniques:
− Lexicon-based (rule-based) approach: it is used to characterize the polarity (negative, positive, and neutral) of textual material (words or phrases that convey the sentiment of the entire text). This method

can be divided into i) corpus-based (using corpus data in either a statistical or semantic manner) and ii) dictionary-based (using dictionary data).

−   Machine learning approach: this technique allows machines to learn tasks without being particularly programmed to perform them. By training ML algorithms with examples of labeled emotions in texts (dataset), machines automatically learn how to predict sentiment without human intervention. ML models can be trained in SA tasks to read beyond simple definitions, and understand more complex things like context, metaphor, sarcasm, and misapplied words. The most promising algorithms are DL ones.

−   Hybrid approach: it is the combination of ML methods and lexicon-based approaches into one system to enhance the model's performance as well as sentiment scoring. This gives us a machine learning model that has been trained on a labeled corpus.

Our proposed method is to use the hybrid approach to build a BERT model on a tweet's dataset, labeled using the lexicon-based approach by following the next steps as shown in Figure 1.

i)   First Step: Data collection. In this stage, data for our analysis is crawled from Twitter using Twitter application programming interface (API). We remind that the data from this social media platform is a valuable record of the intersubjective understanding of an extreme event by individuals and groups due to the navigational ease, anonymity (the ability to post any material without disclosing names), and the publication system's flaws (users only need to have each a valid account).

ii)   Second Step: Enrichment. We used the Spark natural language toolkit (NLTK) library to break the sentences into words and tagging them. After doing the parts of speech (POS) tagging which stands for Part of Speech tagging on the words of tweets, we also tagged the different words based on five dictionaries that we fed into our system (positive, negative, incremental, decremental, and inverse words) each one with a different weight.

iii) Third Step: Data classification. Using the above dictionary (tags), the model calculates the sentiment score of each sentence and gives the overall sentiment score. A positive score is taken as a normal tweet and a negative one is taken as a crime-related tweet.

iv)   Fourth Step: Pre-processing. Before we start building our BERT model, the acquired data is cleaned to make it ready for feeding it into the classifier.

v)   Fifth Step: Building the model. We load the pre-trained BERT Sequence Classifier and Tokenizer. Then, we used the Sequence Classifier and BERT's Tokenizer, to build our model and tokenizer. Finally, we used this model on the Twitter dataset that we already labeled to fine-tune the classifier.

vi)   Sixth Step: Evaluating the model. We assess our model using the test dataset by predicting sentiments after it has been built and generated.
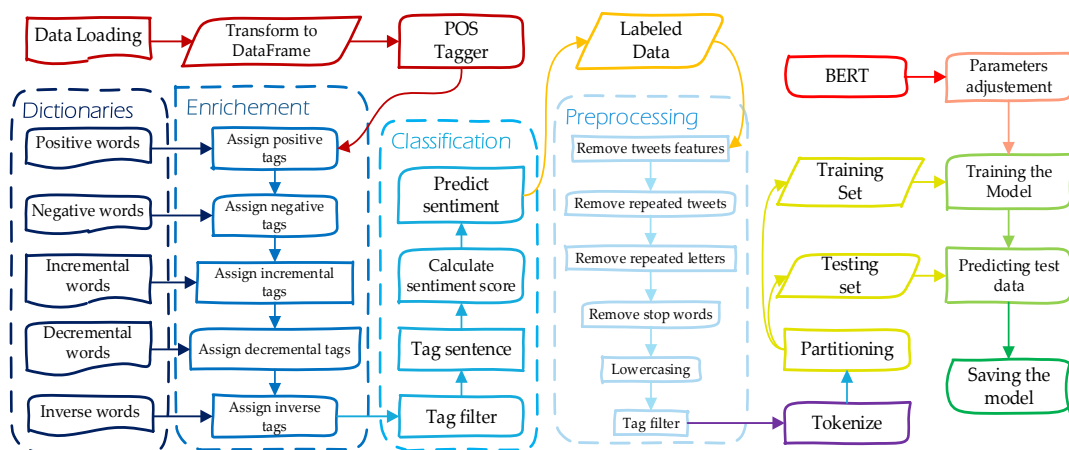

Figure 1. Proposed method

3.1. Labeling the data

The Twitter data collected for text analysis contain 70,000 tweets (27,000 crime tweets, 43,000 normal texts). The corpus for Christchurch Mosque attacks contains 37,000 tweets, El Paso shooting contains 15,000 and Hanau attacks contain 18,000. The streaming Twitter API was used to gather these tweets. We were unable to acquire historical data using Twitter since it only permits us to collect data from the last seven days, and so we begin to collect tweets data these last three years. The data we used was taken over three

days for each of the incidents, one day before, one day after, and the day in question, all using the city's name as a keyword.
- − Christchurch Mosque attacks are a series of far-right terrorist attacks committed on March 15, 2019, by Brenton Tarrant against two mosques in the city of Christchurch, New Zealand, which left 51 dead and 49 injured [16]. It is the deadliest massacre to have occurred in peacetime in New Zealand since that of Boyd in 1809 (66 to 70 dead).
- − El Paso shooting took place on August 3, 2019, in El Paso, Texas [17]. Twenty-three people are killed, at least twenty-three others are injured. The killer was arrested after surrendering to the police, says he wanted to kill as many Mexicans as possible. Racist hate crime is the hypothesis favored by a survey. The Mexican government considers it an anti-Mexican attack, and the UN speaks of a terrorist act against the Latin American community in the United States.
- − Hanau attacks took place on February 19, 2020, 43-year-old Tobias Rathjen from Hanau, hereinafter referred to as "R.," shot and killed nine Hanau citizens with a migration background in front of a shisha bar [18]. He then shot himself and his mother in his parents' apartment.

After collecting the dataset, we tokenize it and perform part of speech tagging (POS) for each word using the Spark NLTK library to break the sentences into words and tag them. Then, we enrich it using five dictionaries that we fed into our system (positive, negative, incremental, decremental, and inverse words) each one with a different weight (+1, -1, *2, /2, *-1, respectively). The positive and negative dictionaries are based on Liu and Hu opinion lexicon [19] that we have completed with other security intelligence lexicon from different public security sources like the FBI NIBRS data [20]. We have also included in our negative lexicon the Islamophobia (terrorism and extremism against Muslims) most used words, resulting from incidents and events that affect Muslim communities in general [21] as shown in Figure 2. Subsequently, we get labeled words which we used to calculate sentences sentiments in general, whether normal or criminal. And so on, we were able to label our dataset.

Figure 2. Word cloud of Islamophobia

3.2. Preprocessing

After the Twitter data has been gathered and converted to text format, it must be cleaned and pre-processed before being utilized in sentiment analysis. Twitter data is notorious for being highly loud, with a lot of banal chatter and linguistic inconsistencies [22]. To make the Twitter data as clean and easy as feasible for our sentiment analysis algorithm, the following preparation procedures were done.
- − Removal of tweets features: Tweets are frequently clogged with data that isn't immediately relevant or useful for sentiment analysis. The hashtag sign (#), URLs/ hyperlinks, numbers, references to other users, retweet symbol (RT), and emoticons are all included in this data. While not necessarily lacking in intrinsic worth, sentiment analysis algorithms do not identify this data correctly, thus it must be eliminated first.
- − Removal of repeated letters and tweets: In tweets, words are frequently written in extended or exaggerated forms (for example, "no" and "nooo!!"). To accommodate for this, the correctly spelled variant of each lengthy version of a word has been substituted. Additionally, while extracting tweets, the API occasionally provides identical tweets, which we removed to prevent giving a single tweet excessive weight.
- − Removal of stop words: we filtered the stop words using Python NLTK stop words as they are

considered neutral polarity and are not useful for polarity decisions.
- Lowercasing: as sentiment analysis algorithms include case sensitivity, all capital characters must be converted to lowercase. Words with the distinct case would otherwise be considered as independent words and processed accordingly.

Finally, we filter our tags to have only adjectives, adverbs, and nouns as they are the most relevant POS in SA for any language [23].

3.3. Building our model

Now that we have cleaned and prepared our data, we divide it into data X (texts) and label Y (sentiments), and then into a random training subset (80%) and testing subset (20%). We built our SA model using a hybrid approach with first the lexicon-based to generate labeled output. Then, the output was fed into our BERT model for training. This model used the basic BERT model (BERT-base that consists of 12 transformer layers) and built our sentiment classifier on top of it.

For training the BERT model, we need to do some additional preprocessing:
- Add special tokens to separate sentences and do classification:
  - [SEP]: ending of a sentence.
  - [CLS]: start of each sentence.
  - [PAD]: special token for padding.
  - [UNK]: everything else (unknown token).
- Pass the sequences with a constant length (padding), that we limit to 512 tokens.
- Create an array of padded tokens (0 s) and real ones (1 s) called attention mask.

We trained our model on three epochs (BERT models are already pre-trained, and a delicate fine-tuning usually gives better results) with a batch size of six (the total number of training samples in a batch) and a learning rate of 2e-5. To get the predicted probabilities from our trained model, we applied the softmax function to the outputs. On the prediction step, we used a forward pass to compute logits and softmax to calculate probabilities.

4.   EXPERIMENTS AND RESULTS
4.1. Technical resources

The following hardware specs were used in our studies on the MARWAN high-performance computing (HPC) infrastructure:
- CPU: 2 Intel Xeon Gold 6148,
- RAM: 192 GB,
- GPU: 2x NVIDIA Tesla P100 (12 GB).

We utilized Keras (2.4.0) [24], an open-source python DL framework that operates on top of Google's open-source data flow software, and TensorFlow-GPU [25] as the backend engine in our experiment.

4.2. Evaluation metrics

To evaluate our model, we used the following metrics: accuracy, loss, precision, recall, F1-score, and the confusion matrix.
- Accuracy: is the percentage of correct predictions among all predictions as shown in (1).
- Loss: is the difference between the model's predicted value and the actual value. Cross-entropy is the most widely used loss function in DL as shown in (2), where p(x) is the real distribution and q(x) is the estimated distribution, both specified over the discrete variable x [26].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$H(p,q) = -\sum_{\forall x} p(x)\log(q(x)) \tag{2}$$

Where:
- True positive (TP): is the number of successfully classified positive class records.
- True negative (TN): is the number of successfully classified negative class records.
- False positive (FP): is the number of incorrectly classified negative class records.
- False negative (FN): is the number of incorrectly classified positive class records.
- Precision: is the percentage of all positive results that were accurately identified as shown in (3).
- Recall: is the proportion of accurately identified positive results among the total number of existing positive class as shown in (4).

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

– F1-score: commonly known as the F-score, is the harmonic mean of precision and recall [27], with a value β that emphasize one or the other as shown in (5). F-score has a maximum value of 1 and a minimum value of 0, and is always between precision and recal,

$$F_\beta = \left(1+\beta^2\right)\cdot\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} \tag{5}$$

where $\beta = 1$, we have the standard F-measure or balanced F_1-score as shown in (6).

$$F_1\text{-score} = 2\cdot\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} \tag{6}$$

– Confusion matrix: is a table that visualizes the model's performance by putting predicted class instances in the rows and real class instances in the columns.

4.3. Evaluating our model

This study consists of predicting crimes using data from social media (Twitter) with a hybrid approach (lexicon-based and BERT) and has provided very promising results. In the following figures (Figures 3 to 5) generated by Tensorboard, we present the training accuracy, validation accuracy, training loss, and validation loss and for each class, we present the Recall and Precision, along with the F1 score which was trained over three epochs. As shown in Figure 3, our model in these three epochs (7 hours 10 minutes of training) the accuracy reached 97.24% and the loss attained 8.41% in training, and for the validation, it reached 94.91% in the accuracy and 16.26% in the loss as shown in Figure 4. Regarding the Recall, Precision, and F1-score, we obtained more precise results for each of our two classes (crime or not). We see that for the crime-related class, we got a precision of 92%, a recall of 94%, and an F1-score of 93%, and for the normal class (normal text chat), we obtained a precision of 96%, a recall of 95%, and an F1-score of 95% as shown in Figure 5.
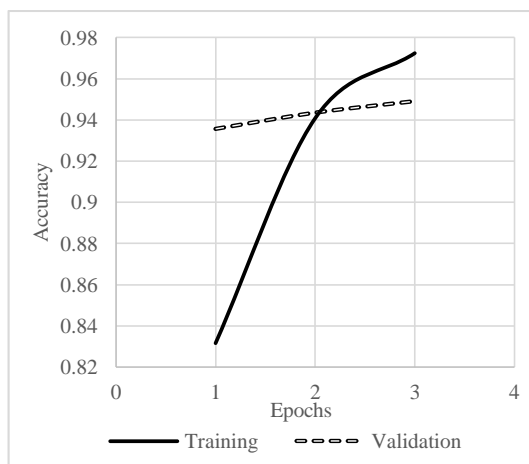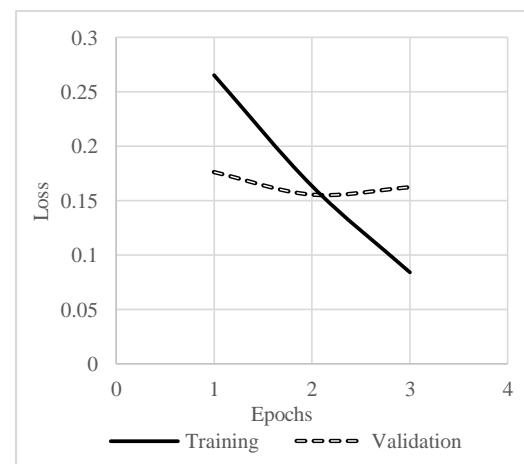

Figure 3. Training and validation accuracy

Figure 4. Training and validation loss

To verify the performance of our model, we compared the results obtained (accuracy and F1-score) with other papers' results that also used Twitter as the dataset and dealt with the security intelligence field as shown in Table 1. We can see that the hybrid approach using the NLP's model BERT gives better results than the lexicon-based and machine learning approaches in both accuracy and F1-score metrics, as BERT outperform the other NLP's model in context understanding part, especially in the "context heavy" one, for which it is so important for analytics purposes.
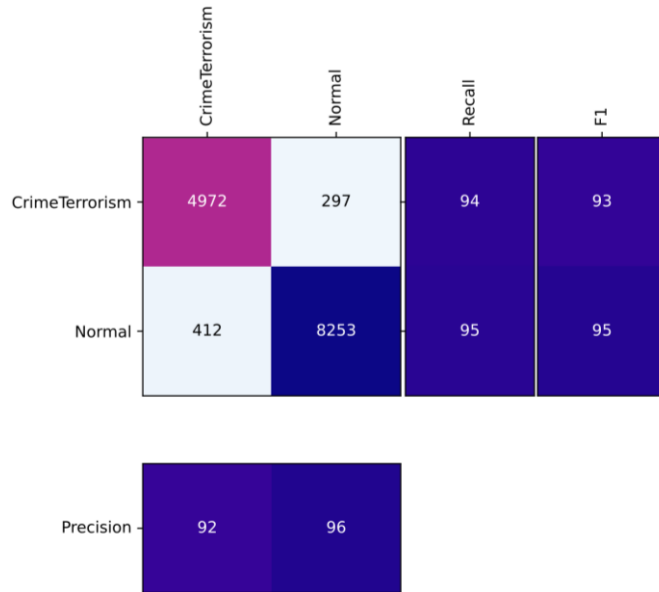
Figure 5. Model confusion matrix

Table 1. Comparison of our results with those of other papers

Paper	Approach	Accuracy	F1-score
Our approach	Hybrid (lexicon-based + BERT)	94.91%	94.92%
[28]	Machine Learning-SVM	91.55%	-
	Machine Learning - CNB	88.17%	-
	Machine Learning-DT	82.46%	-
	Machine Learning-KNN	78.06%	-
[29]	Lexicon-based	-	55%
[30]	Machine Learning-MNB	-	83%
	Machine Learning-SVM	-	89%
	Machine Learning-RF	-	87%
[31]	Lexicon-based	93%	-

The issue of analyzing and processing English metaphorical, sarcastic, and encrypted expressions in the context of sentiment analysis produces a degradation in the performance of the model when it comes to these types of sentences. Although substantial progress has been achieved in these areas of study, present computational models and methods are still unable to handle these sorts of expressions.

A metaphor is a figure of speech that provides a language technique for conveying thoughts and notions that are not the same as they appear on the surface, it directly refers to one thing by mentioning another, and to efficiently identify its hidden sentiment, the model must be capable of reading between the lines. In a sarcastic message, people use optimistic phrases to express their bad feelings. Because of this, sarcasm may readily mislead SA models unless they are expressly constructed to account for it. It can be difficult to determine its sentiment without a thorough awareness of the situation, the subject, and the surrounding environment. Cryptography focus on protecting messages (ensuring confidentiality, authenticity, and integrity) by often using secret keys. It makes a message supposedly unintelligible to anyone other than those who are entitled to it, and so it is no longer possible to understand its content without more knowledge.

5. CONCLUSION

This paper presents a crime detection and classification method based on a lexicon-based approach combined with the BERT deep learning algorithm. The classification accuracy of the method used in this article on the Twitter dataset reached 94.91%, along with 16.26% loss, 94.94% precision, 94.91% recall, and 94.92 F1-score. It can be seen from the experimental results that the method used in this article can be effective for crime detection. Furthermore, it can be observed that the hybrid method based on BERT outperforms existing works in crime detection. As future works, we will detect crime using other media such as audio, images, and videos based on this model.

ACKNOWLEDGEMENTS

REFERENCES

[1] I. Idrissi, M. Boukabous, M. Azizi, O. Moussaoui, and H. El Fadili, "Toward a deep learning-based intrusion detection system for IoT against botnet attacks," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 110-120, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp110-120.

[2] M. Berrahal and M. Azizi, "Augmented binary multi-labeled CNN for practical facial attribute classification," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 2, pp. 973-979, 2021, doi: 10.11591/ijeecs.v23.i2.pp973-979.

[3] E. A. Kirillova, R. A. Kurbanov, N. V. Svechnikova, T. E. Zul'fugarzade, and S. S. Zenin, "Problems of fighting crimes on the internet," *Journal of Advanced Research in Law and Economics*, vol. 8, no. 3, pp. 849–856, Jun. 2017.

[4] M. Salter, *Crime, justice and social media*, England, U.K.: Routledge, Jan. 2016.

[5] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, edited by E. Soria, J. Martin, R. Magdalena, M. Martinez, and A. Serrano, IGI global, 2009, pp. 242–264.

[6] I. Idrissi, M. Azizi, and O. Moussaoui, "Accelerating the update of a DL-based IDS for IoT using deep transfer learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 2, pp. 1059-1067, 2021, doi: 10.11591/ijeecs.v23.i2.pp1059-1067.

[7] M. Boukabous and M. Azizi, "Review of Learning-Based Techniques of Sentiment Analysis for Security Purposes," *Innovations in Smart Cities Applications Volume 4. SCA 2020. Lecture Notes in Networks and Systems*, vol. 183, pp. 96–109, doi: 10.1007/978-3-030-66840-2_8.

[8] A. Vaswani *et al.*, "Attention is all you need," *arXiv*. 2017.

[9] F. K. Ventirozos, I. Varlamis, and G. Tsatsaronis, "Detecting aggressive behavior in discussion threads using text mining," In: Gelbukh A. (eds) *Computational Linguistics and Intelligent Text Processing. CICLing 2017. Lecture Notes in Computer Science*, vol. 10762, 2018, doi: 10.1007/978-3-319-77116-8_31.

[10] P. Siriaraya *et al.*, "Witnessing crime through tweets: A crime investigation tool based on social media," *SIGSPATIAL '19: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, pp. 568–571, doi: 10.1145/3347146.3359082.

[11] J. C. Pereira-Kohatsu, L. Quijano-Sánchez, F. Liberatore, and M. Camacho-Collados, "Detecting and monitoring hate speech in twitter," *Sensors (Switzerland)*, vol. 19, no. 21, 2019, doi: 10.3390/s19214654.

[12] M. Boukabous and M. Azizi, "A comparative study of DL-based language representation learning models," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 2, pp. 1032–1040, 2021, doi: 10.11591/ijeecs.v22.i2.pp1032-1040.

[13] J. Yadav, D. Kumar, and D. Chauhan, "Cyberbullying Detection using Pre-Trained BERT Model," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 1096-1100, doi: 10.1109/ICESC48915.2020.9155700.

[14] G. R. S. Weir, E. Dos Santos, B. Cartwright, and R. Frank, "Positing the problem: Enhancing classification of extremist web content through textual analysis," *2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)*, 2016, pp. 1-3, doi: 10.1109/ICCCF.2016.7740431.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000001.

[16] K. Gelineau and J. Gambrell, "New Zealand Mosque shooter is a white nationalist who hates immigrants, documents and video reveal," *Chicago Tribune*, 2019. Accessed: Apr. 09, 2011. [Online]. Available: https://www.chicagotribune.com/nation-world/ct-mosque-killer-white-supremacy-20190315-story.html

[17] T. Law and J. Bates, "El Paso Shooting Suspect Told Police He Was Targeting 'Mexicans.' Here's What to Know About the Case," *Time*, 2019. Accessed: Apr. 09, 2011. [Online]. Available: https://time.com/5643110/el-paso-texas-mall-shooting/

[18] F. Gardner, "Germany shooting: 'Far-right extremist' carried out shisha bars attacks - BBC News," *BBC News*, 2020. Accessed: Apr. 09, 2011. [Online]. Available: https://www.bbc.com/news/world-europe-51567971

[19] M. Hu and B. Liu, "Mining and summarizing customer reviews," *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177, doi: 10.1145/1014052.1014073.

[20] FBI, "Federal Bureau of investigation, crime data explorer," *FBI Uniform Crime Reporting Program*, 2021. Accessed: Apr. 08, 2021. [Online]. Available: https://crime-data-explorer.fr.cloud.gov/downloads-and-docs

[21] Caleb Elfenbein, "Data – Mapping Islamophobia," *Mapping Islamophobia*, 2021. Accessed: Jun. 10, 2021. [Online]. Available: https://mappingislamophobia.org/data/

[22] P. Burnap and M. L. Williams, "Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making," *Policy and Internet*, vol. 7, no. 2, pp. 223-242, 2015, doi: 10.1002/poi3.85.

[23] M. Sokolova and G. Lapalme, "Classification of opinions with non-affective adverbs and adjectives," *International Conference RANLP 2009 - Borovets, Bulgaria*, 2009, pp. 421-427.

[24] "Keras: the Python deep learning API," *Keras*. Accessed: Aug. 18, 2020. [Online]. Available: https://keras.io/

[25] "TensorFlow." Accessed Aug. 18, 2020. [Online]. Available: https://www.tensorflow.org/?hl=fr

[26] Z. Zhang and M. R. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels," *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 8778–8788.

[27] L. Derczynski, "Complementarity, F-score, and NLP evaluation," *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, pp. 261-266.

[28] H. AL-Saif and H. Al-Dossari, "Detecting and classifying crimes from arabic twitter posts using text mining techniques," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 10, pp. 377–387, 2018, doi: 10.14569/IJACSA.2018.091046.

[29]   S. Aghababaei and M. Makrehchi, "Mining Twitter data for crime trend prediction," *Intelligent Data Analysis*, vol. 22, no. 1, pp. 117-141, 2018, doi: 10.3233/IDA-163183.

[30]   H. El Hannach and M. Benkhalifa, "WordNet based implicit aspect sentiment analysis for crime identification from Twitter," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, pp. 150-159, 2018, doi: 10.14569/IJACSA.2018.091222.

[31]   J. Hao and H. Dai, "Social media content and sentiment analysis on consumer security breaches," *Journal of Financial Crime*, vol. 23, no. 4, 2016, doi: 10.1108/JFC-01-2016-0001.

BIOGRAPHIES OF AUTHORS

Mohammed Boukabous 🆔 ⑧ SC Ⓟ is a Ph.D. candidate in Computer Engineering at Mohammed First University in Oujda, Morocco, where he is conducting research in security intelligence using deep learning algorithms in exchanged messages. He holds a M.Sc. degree in internet of things from Sidi Mohamed Ben Abdellah University in Fez, Morocco (2019), as well as a B.Sc. degree in Computer Engineering from Mohammed First University (2016). Furthermore, he holds several certifications in natural language processing, artificial intelligence, security intelligence, big data, and cybersecurity. Additionally, he served as a reviewer for various international conferences. He is currently employed at Mohammed First University as an administrative. He can be contacted at email: m.boukabous@ump.ac.ma.

Prof. Dr. Mostafa Azizi 🆔 ⑧ SC Ⓟ received a State Engineer degree in Automation and Industrial Computing from the Engineering School EMI of Rabat, Morocco in 1993, then a Master degree in Automation and Industrial Computing from the Faculty of Sciences of Oujda, Morocco in 1995, and a Ph.D. degree in Computer Science from the University of Montreal, Canada in 2001. He earned also tens of online certifications in Programming, Networking, AI, Computer Security. He is currently a Professor at the ESTO, University Mohammed First of Oujda. His research interests include Security and Networking, AI, Software Engineering, IoT, and Embedded Systems. His research findings with his team are published in over 100 peer-reviewed communications and papers. He also served as PC member and reviewer in several international conferences and journals. He can be contacted at email: azizi.mos@ump.ac.ma.