

# A Decision Support System for Detecting Age and Gender from Twitter Feeds based on a Comparative Experiments

Roobaea Alroobaea<sup>1</sup>, Sali Alafif<sup>2</sup>, Shomookh Alhomidi<sup>3</sup>  
Ahad Aldahass<sup>4</sup>, Reem Hamed<sup>5</sup>, Rehab Mulla<sup>6</sup>, Bedour Alotaibi<sup>7</sup>

Department of Computer Science, College of Computers and Information Technology  
Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia<sup>1</sup>

Department of Computer Science, College of Computers and Information Technology, Taif University, Saudi Arabia

**Abstract**—Author profiling aims to correlate writing style with author demographics. This paper presents an approach used to build a Decision Support System (DSS) for detecting age and gender from Twitter feeds. The system is implemented based on Deep Learning (DL) algorithms and Machine Learning (ML) algorithms to distinguish between classes of age and gender. The results show that every algorithm has different results of age and gender based on the model architecture and power points of each algorithm. Our decision support system is more accurate in predicting the age and the gender of author profiling from his/her written tweets. It adopts the deep learning model using CNN and LSTM methods. Our results outperform those obtained in the competitive conference s CLEF 2019.

**Keywords**—Decision support system; age detection; gender detection; author profiling; deep learning; machine learning

## I. INTRODUCTION

Recently, many studies have focused on the field of Information Retrieval (IR). The primary target of IR is to extract critical information of a field by predicting the behavior, preference, and person's characteristics. Subsequently, filtering these massive amounts of data to convert it to useful information. This information will be used in the decision support systems. These systems can be used for creating marketing strategies [1], giving viewing suggestions to the user [2], filtering and translate languages [3]. Furthermore, support investigations by studying the sensitive text for national security that detects the source of a threat and helps police in detecting characteristics of the criminal from his/her linguistics [4] [5].

In addition, a decision support system (DSS) plays a great deal nowadays. For bidding websites, there is a needed to know age and gender of the user in order to propose to him/her the service or the clothes. It is becoming a challenging problem. For this reason, the decision support system is a needed by using author profiling.

The aim of this work is to solve the problem of author profiling (AP) by proposing a decision support system (DSS) for detecting age and gender from Twitter Feeds. Author profiling means determining the author's gender and the author's age group of a text by examining his or her writing and identifying the stylistic features. English tweets were

taken as input from the PAN-AP-2019 dataset, then generate age and gender based on extraction features distinguish between gender and age class. To achieve the best result, multiple algorithms were implemented such as Deep Learning (like 'Convolutional Neural Network (CNN)', 'Deep Neural Network (DNN)', and 'Long Short-Term Memory (LSTM)' and Machine Learning (like, 'Naïve Bayes (NB)', 'k-Nearest Neighbors (KNN)', 'Decision trees (DT)', 'Support Vector Machines (SVM)', 'Neural Network', and 'Random Forest (RF)'. The structure of the paper is organized as follows; Section 2 presents a brief review on author profile prediction methods. Section 3 introduces our method used to discriminate between authors. Section 4 presents the experiments as well as the evaluations. Finally, the summarization with the conclusion will be mentioned, following by future work.

## II. LITERATURE REVIEW

Text classification is depending on statistical procedures and text mining that provide outputs from calculations of extracted terms regularity. Author profiling distinguishes between classes of authors studying their socialist aspect, that is, how people share language. This helps in identifying profiling aspects such as gender, age, native language, or personality type. Author profiling is a problem of increasing significance in applications in forensics, security, and marketing. Table I shows up-to-date researches about detecting age and gender based on DL and ML algorithms.

In addition, there is a need to cover some works that had done by pioneer scientists. The following are some of them:

The author(s) in [14] discovered the possibility of the author automatically classified, which depending on the writer's gender or age using the AP method. AP is predicting features linked to the text author. It includes many dimensions for example gender, age, native language, personality, education level, etc. As stated by [14], men who use more determiners to classify things (this/ this/the, an/an, etc.) and quantifiers (more, little, two, etc.). However, women are more concerned in relations and, thus, using personal pronouns (I, me, her, you, etc.) more than latter. Advanced, [14] modeled the author profiling which is a prediction system. He proposed a technique in his name called the Koppel algorithm. It

contains quantifying the duplication of 467 English keywords (such as too, a, too, their, yourself, etc.) in a text to compute the gender of its writer. All writing styles were analyzed from textbooks, fiction, and tests. Then, the program was able to deliver four of five correct answers.

TABLE I. MODERN RESEARCH ON AGE AND GENDER DETECTION

Work	Method used	Publishing Year	Text languages	Age Result	Gender Result
Alroobaea [6]	NB, Neural Network	2020	English	0.63	0.96
Ashraf and Nie [7]	Bi-LSTM	2020	English Roman-Urdu	0.73	0.88
Rosso and Rangel [8]	RF Logistic regression	2020	Arabic	0.56	0.81
Gishamer [9]	SVM	2019	English	NON	0.89
			Spanish		0.84
Kapočiūtė-Dzikicnė and Damaševiči u [10]	LSTM CNN	2018	Lithuania n	0,316	0.609
Dias and Paraboni [11]	LSTM char n-gram	2018	English	NON	0.66
			Spanish		0.67
			Arabic		0.68
Bayot and Gonçalves [12]	LSTMs	2018	English	NON	0.772
			Spanish		0.687
			Arabic		0.676
Veenhoven et al, [13]	Bi-LSTM CNN	2018	English	NON	79.3
			Spanish		80.4
			Arabic		74.9
Sboev et al [33]	CNN LSTM	2016	Russian	NON	0.86
Mechti et al [3]	DT J48	2013	English	0.58	0.58
			Spanish	0.51	0.556
Gaustad et al [15]	DT J48, RF, IBk, JRip, SMO, Bagging, AdaBoostM1	2007	Arabic	0.7210	0.8115
Tayfun et al [16]	NB Term-based, k-NN	2006	Spanish	0.51	0.556
	NB Style-based, k-NN				0819
Koppel et al [14]	winnow	2003.	English	NON	0.80

The author in [15] worked on automatic classification of messages in the Arabic language and English languages. They used datasets for three demographic and four psychometric. Authors trait from the linguistic features were trained by the

WEKA toolkit and machine learning which includes: 1) decision trees J48, 2) Random Forest, 3) lazy learners IBk, 4) rulebased learners JRip, 5) Support Vector Machines AMO, 6) ensemble/meta-learners Bagging, 7) AdaBoostM1. Bagging looks to income from feature choice, where the SVM based SMO algorithm does not add progress when joint with feature choice. As a result, SMO and Bagging give the best performing for all traits examined. The research got a percentage of 81.5% of well-classified documents about the sex and of 72% about age.

The author in [16] considers the first five problem as a text classification problem. Second, the influence of stylistic features (e.g., word lengths) on predicting the gender. They used a dataset from a chat server (Heaven BBS), when users have peer-to-peer communication via textual messages. The dataset gathering of messages logs storing the users' outgoing messages. They used two different techniques according to gender classes, which are style-based classification and term-based classification. In the Term-Based Classification, they used supervised learning and employed four algorithms (k-NN, naive Bayesian, covering rules, and backpropagation). Every experiment separately was repeated five times. In the Style-Based Classification, they used many stylistic features from the message dataset. The chat messages include emotion-carriers called smileys and emoticons, and they research the word and phrase lengths as a computer-mediated text.

In regarding to the decision support systems (DSS), many systems have been proposed or improved for some areas such as diagnosis of periodontal disease, sales, business ideas competitions, etc. These systems are a computer program used to support determinations, judgments, and ways of action in an organization or a commercial. It collects and analyzes a huge of unstructured data to predict information that can help to get right step in the future or solve problems [36][37]. As mentioned above, the proposed decision support system will help to solve the problem of author profiling. In the next section, our data and methodology will be explained in detailed.

### III. CORPUS AND METHODOLOGY

#### A. Corpus

The data was collected from the CLEF (Cross-Language Evaluation Forum) initiative. It is a self-organized entity whose key goal is to foster science, creativity and the advancement of structures for access to knowledge, with a focus on multilingual and multimodal information at diverse aspects of structure. Since 2010, CLEF has housed evaluation labs for PAN (Plagiarism, Authorship and Near Duplicate Content), offering excellent locations at all times.

Therefore, our corpus is taken from the PAN-AP-2019 dataset. The feeds are taken from Twitter in English language. The dataset is split into two subsets: "training" and "test". The considered dataset is labeled with classes of ages and genders [6]. For age classification, there are four classes (i.e., "18 -24", "25 - 34", "35 - 49" and "50 and more". For gender classification, there are two classes ("Female" and "Male"). There are '14166' tweets that are unbalanced distributed as shown in Table II.

TABLE II. AGE AND GENDER DISTRIBUTION

Class	Number of Tweets
<b>Gender Classifications</b>	
Male	6986
Female	7180
<b>Age Classifications</b>	
18-24	5588
25-34	5624
35-49	1865
50-XX	1089

B. Architecture of the System

The proposed system contains the next steps (see Fig. 1).

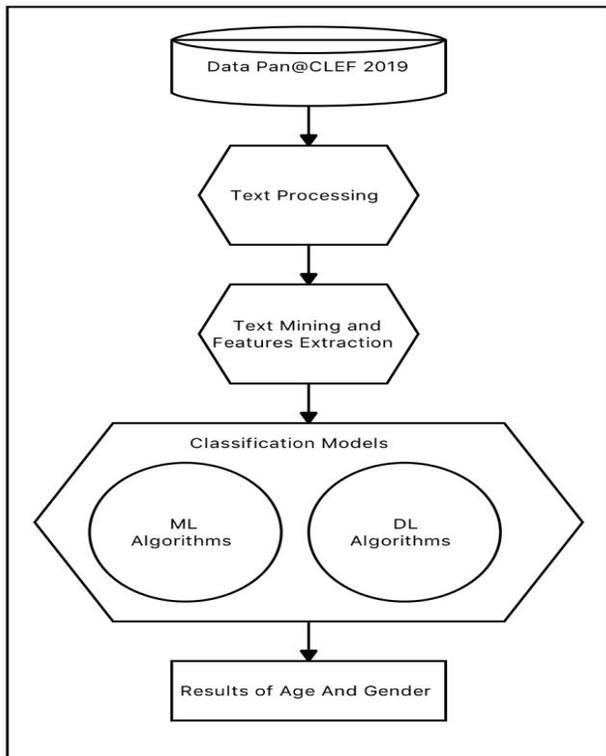


Fig. 1. Construction of the Applied System.

- Text Processing: There are two main problems with the Corpus. First, the model cannot take the tweets directly as input. Secondly, the text could be messy (raw data). As a result, there is a need to converting the raw data into a clean data set by removing noisy data, such as HTML tags, white spaces, inflection words, symbols that would reduce the accuracy of detecting [17].

- Text Mining and Features Extraction: In this step, information that has patterns and trends from the text was extracted by implementing linguistic and statistical processing that study word frequency distributions occurred in the corpus. Then, it sorted these words by frequency [17]. After that gathering these terms into thematic classes to distinguish between age group and gender classes depending on Stylistic features and those based on content [18]. To determine the most word frequency in the corpus and to reduce redundant attributes and to computational capacity to process and to enhance classification performance, TF \* IDF were calculated for each class and iterative this step until getting the most discriminating attributes [17], as shown below in Eq.(1) [19]:

$$w_d = f_{wd} \times \log(|D|/f_{wD}) \tag{1}$$

With: d: documents, W: the word, D: one document  $\in D$ ,  $f_{w,d}$ : number of occurrences of w in d, |D|: the size of the corpora  $f_{w,D}$ : number of document [17] [19].

- Classification Models: As mentioned above, DL and ML algorithms were used to implement the classification step to gender and age files. Next is explanation for these methods.

1) Deep learning (DL): It is primarily used as a statistical tool for categorizing patterns based on sample data and applying multi-layer neural networks. A group of input units, such as words or pixels, are present in the input layer. The hidden layer includes hidden units (the deeper a network is said to be, the more such layers), and ultimately the output units. With ties running between those nodes. After a while, a back-propagation algorithm allows a process called gradient descent to set the links between units that used a process, so that the same output is generated by any given input [20]. Deep learning methods are used include the following:

- Deep Neural Network (ANN): An artificial neural network technology simulates human brain neural networks activity. It contains an input layer, an output layer, and multiple hidden layers that finds the correct mathematical manipulation to turn the input into the output and flows without looping back [21].
- Convolutional Neural Network (CNN): It is a technology used in image recognition. It consists of fully connected networks that mean each neuron in one layer linking to all neurons in the next layer to analyzing it by using unsupervised learning and adding weights to the loss function [22].

- Long Short-Term Memory (LSTM): It is a class of artificial recurrent neural network (RNN) that address the problem of training RNNs with the technique of back-propagation through time (BPTT). It can process long sequence of dependencies such as speech or video. So, it is considered as the solution to short-term memory in RNN by using distinct units called memory blocks in the recurrent hidden layer [23]. The memory blocks contain [24]:
  - Memory Cells: remembering the historical state of the network. Also, it has weights input, output, and the internal state as following:
    - Input Weight: Weight input for the present time.
    - Output Weight: Weight output from last time.
    - Internal Weight: Internal state applied to calculate the output.
  - Three control gates: in charge of controlling the information flow:
    - Input Gate: Controls the flow of input activations into the memory cell.
    - Output Gate: Controls the output flow of cell activations into the rest of the network.
    - Forget Gate: Decides what information to delete from memory cells.

Input gate and forget gate using to updating of internal state. The output gate is the last filter for output. All these gates with consistent data flow called constant error carousel (CEC) that help to save cells neither exploding gradient (values of the model's weights quickly become very large during training) or vanishing gradient.

Most of LSTM advantages are achieved by using back-propagation through time (BPTT) supervised learning algorithm that is used to modify weight to minimize the error of output compared to the expected output in the response of corresponding input [24].

2) Machine learning (ML): It is a research field of artificial intelligence that extract knowledge from data by analyzing them to predict and make the design. The applications of machine learning models become in everyday life from automatic recommendation movies to what food to order or recognizing people in photos and many other applications that contain ML models [25]. The following are ML techniques used in our approach:

- K-Nearest Neighbors (K-NN): It can be defined as a classification algorithm that can predict a new data point by discovering the nearby training set relying on fixed number 'k' to that point and allocate it to the training set. It is the simplest algorithm to using and understand. It offers best performance without any significant adjustment. Nevertheless, it is lazy and cannot handle several features [25].

- Naïve Bayes Classifiers (NB): They are types of classification algorithm based on the statistical theory of Bayes. They assumed that there is a specific feature in a class that is not related to the presence of any other feature. Their advantages are predicting fast and dealing with large datasets [25].
- Decision Trees (DT): They are types of classification and regression algorithms. They performed classification without requiring much computation. In addition, they provide a clear indication of which fields are most important for prediction or classification. The target of this model is predicting the value of a target variable based on several input variables. They are unlike linear models; they are adaptive to solve any type of problems (classification or regression) [25].
- Random Forest (RF): It is defined as a classification algorithm. It has several decision trees, where each tree is a little different from the others. It solves the problem of over-fitting in training data. The over-fitting can be reduced by being an average of their results. [25].
- Support Vector Machine (SVM): An efficient supervised ML algorithm delivers better performance for both classification and regression exercises. It is accurate and works on a small amount of revised data, and it does not work on a large amount of data because it requires more training time. SVM adapts training data containing features and its preferred class, it contains two stages:
  - Learning Stages: where SVM detects closest data points to the decision boundary known as Support Vectors (SVs), which forms the best separation among the classes [26].
  - Prediction Stages: These SVs are applied to predict the class of test records [26].

### C. Evaluation of the System

By using the confusion matrix shown in Table III, precision, accuracy, recall, f-score can be calculated with the equations (2), (3),(4), and (5) that evaluate the performance for our unbalanced dataset on ML algorithms [27]. The accuracy for DL algorithms were calculated [28].

TABLE III. CONFUSION MATRIX

		Predicted class	
		Class A	Class B
True Class	Class A	True Positives (TP)	False Negatives (FN)
	Class B	False Positives (FP)	True Negatives (TN)

Where TP (True Positives) and TN (True Negatives) positive and negative correct labeled predictions. In Opposition, False Positives (FP) and False Negatives (FN) are positive and negative incorrect labeled predictions.

- Accuracy: Measure how the system is correctly predicted [34].

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (2)$$

- Precision: The number of true positives divided by all (positive) predictions [34].

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

- Recall: Percentage of positive predictions in the system that correctly identified [34].

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

- F\_score: Measure the harmonic between recall and precision [34].

$$F\_score = 2 \times \frac{Precision \times Recall}{precision+Recall} \quad (5)$$

#### IV. EXPERIMENTATION AND RESULTS

In this research, DL and ML algorithms were implemented to detect the age and gender of an author. DL algorithms results are more efficient for our model rather than ML algorithms. In the first time of implementing our approach, we did not get as expected results from CNN age’s code, also LSTM age and gender codes.

Notice that using either "sigmoid" or "softmax" as an activation function will not affect gender because we have only two classes called "Binary Classification" [31]. On the other hand, we used “softmax” for age’s code “Multi Class Classification”, but we had the same accuracy and loss in each step in both models. According to this problem, the solution is mentioned at [32]. We had changed the activation function to “sigmoid”, but this provides loss results in negative. We decide to change the loss function from "binary\_crossentropy" to “mean\_squared\_error”. This provides good accuracy as shown below in (Table IV). As a result, the searched was done for the reasons that causing these problems and solved them as shown below:

##### A. LSTM Gender’s Code Problem and Solution

By using 50000 words as top frequent words in LSTM. In addition to 10 hidden cells in the “Dense” layer. Had not provided good accuracy as in CNN algorithm. As a result, the number of hidden cells was increase to 20 “Dense” layer and number of frequent words was decreased into 5000 as shown in this problem [29]. Our approach had increased the accuracy from 0.5068 to 0.9906 as shown below in (Table V).

##### B. Age’s Code Problem and Solution

Convolutional Neural Network and Long Short-Term Memory algorithms had provided great results for gender code according to the balance of gender tweets as shown in Table VI. Table IV shows how we implemented "binary\_crossentropy" loss function to the model, that is used when there are only two label classes (Male and Female) as this was mentioned at [30]. In addition to using “sigmoid” as "activation" function.

Several classification models were tried as shown in (Table IV) and (Table V); the best result is achieved by implementing Convolutional Neural Network CNN as the first place and LSTM algorithms as the second place, in both age and gender detection. According to our search, we believe that if our model had a larger dataset size; the LSTM algorithm will be more powerful rather than the CNN algorithm, as mentioned in [35]. In Table VI, our approach is compared with the best accuracy result achieved of gender detection of competition “PAN at CLEF 2019” implemented by Valencia et al team for English tweets. It is obvious that our model system obtains more accurate results than others in predicting gender from author profiling.

TABLE IV. AGE DETECTION RESULTS

Deep Learning Algorithms					
Algorithm	Epoch Numbers	Accuracy	Loss		
CNN	30	0.7489	0.5210		
LSTM	30	0.7265	0.5434		
Machine Learning Algorithms					
Algorithm		Precision	Recall	Accuracy	F-score
K-NN	0	0.61	0.61	0.50	0.61
	1	0.66	0.40		0.50
	2	0.38	0.50		0.43
	3	0.16	0.39		0.23
NB	0	0.61	0.89	0.62	0.72
	1	0.63	0.67		0.65
	2	1.00	0.04		0.07
	3	0.00	0.00		0.00
DT	0	0.62	0.68	0.58	0.65
	1	0.59	0.61		0.60
	2	0.49	0.38		0.43
	3	0.34	0.25		0.29
SVM	0	0.66	0.83	0.66	0.74
	1	0.64	0.71		0.68
	2	0.82	0.32		0.47
	3	0.84	0.17		0.28
Neural Network	0	0.69	0.69	0.63	0.69
	1	0.63	0.72		0.67
	2	0.54	0.41		0.46
	3	0.35	0.24		0.28
RF	0	0.63	0.88	0.64	0.73
	1	0.63	0.65		0.64
	2	0.94	0.26		0.41
	3	1.00	0.06		0.11

TABLE V. GENDER DETECTION RESULTS

Deep Learning Algorithms					
Algorithm	Epoch Numbers		Accuracy	Loss	
CNN	30		0.9965	0.0023	
LSTM	30		0.9906	0.0211	
Machine Learning Algorithms					
K-NN	0	0.59	0.66	0.61	0.62
	1	0.63	0.56		0.59
NB	0	0.75	0.56	0.69	0.64
	1	0.66	0.82		0.73
DT	0	0.61	0.59	0.61	0.59
	1	0.61	0.63		0.62
SVM	0	0.70	0.66	0.69	0.68
	1	0.69	0.72		0.70
Neural Network	0	0.69	0.68	0.69	0.69
	1	0.69	0.70		0.70
RF	0	0.71	0.62	0.69	0.66
	1	0.67	0.75		0.71

TABLE VI. COMPARED BEST RESULTS

Teams	Method Used	Best Gender Accuracy Result
Our Team	CNN	0.9965
	LSTM	0.9906
Valencia et al [36]	SVM Logistic Regression Multi-Layer Perceptron	0.8432

## V. CONCLUSION

In this research, DL and ML algorithms were used to solve the PA problem from English tweets that was taken from the PAN-AP-2019 dataset. To achieve the best result, multiple algorithms of Deep Learning were implemented, which are DNN, CNN and LSTM. In addition, Machine Learning algorithms were implemented too, which are “KNN, NB, DT, SVM, Neural Network, and RF”. The results showed that every algorithm has different results of age and gender based on the model architecture and power points of each algorithm.

Moreover, our decision support system and model are achieved good result than others, who participated in the competitions CLEF 2019 (Table VI). It is based on deep learning model using CNN and LSTM methods. It is more accurate in predicting the age and the gender of author profiling from his/her written tweets. Then, it will be great if it is adopted in any bidding websites that need to know the age and the gender of their users in order to propose to them the service such as book or clothes, etc.

## REFERENCES

[1] E. Lundeqvist, and M. Svensson. Author profiling: A machine learning approach towards detecting gender, age and native language of users in social media, 2017.

[2] PranavDar/SEPTEMBER11,2018/analyticsvidhya/february,29,2020/http s://www.a nalyticsvidhya.com/blog/2018/09/facebook-rosetta-process-text-billions-images/. Acecc [22/2/2020].

[3] S. Mechti., M. Jaoua , L. H. Belguith, & R. Faiz.. Author profiling using style-based features. Notebook Papers of CLEF2, 2013.

[4] S.S. Yatam, T.R. Reddy. Author profiling: Predicting gender and age from blogs, reviews & social media. International Journal of Engineering Research & Technology (IJERT), ISSN, pp.2278-0181. 2014.

[5] Tom Buchanan, A. John Johnson, and R. Lewis Goldberg. Implementing a Five-Factor personality inventory for use on the internet. European Journal of Psychological Assessment, 21:115–127, 2005.

[6] R. Alroobaea. An Empirical combination of Machine Learning models to Enhance author profiling performance. International Journal,9(2), 2020.

[7] M.A. Ashraf, R.M.A Nawab, and F. Nie, Author profiling on bi-lingual tweets. Journal of Intelligent & Fuzzy Systems, (Preprint), pp.1-11.2019.

[8] P. Rosso, F. Rangel. Author Profiling Tracks at FIRE. SN Computer Science, 1(2), pp.1-11, 2020.

[9] F. Gishamer. Using Hashtags and POS-Tags for Author Profiling Notebook for PAN at CLEF , 2019.

[10] J.Kapočiūtė-Dzikicnė and, R. Damaševičius. Lithuanian author profiling with the deep learning. In 2018 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 169-172). IEEE. 2018.

[11] R.F.S. Dias, and I. Paraboni. Author Profiling using Word Embeddings with Subword Information: Notebook for PAN at CLEF 2018. In CLEF (Working Notes), 2018.

[12] R.K. Bayot, and T. Gonçalves. Multilingual author profiling using lstms. In Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018) , 2018.

[13] R. Veenhoven, S. Snijders, D. van der Hall and R van Noord. Using translated data to improve deep learning author profiling models. In Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018) (Vol. 2125), 2018.

[14] M. S. Koppel and A. Shimoni, Automatically categorizing written texts by author gender, Literary and Linguistic Computing, pages 401-412, 2003.

[15] T. Gaustad, D. Estival and B. Hutchinson. TAT: an author-profiling tool with application to Arabic emails. Proceedings of the Australasian Language Technology Workshop, pages 21-30, Melbourne, Australia, 2007.

[16] B. Tayfun Kucukyilmaz, B. Barla Cambazoglu, Cevdet Aykanat, and Fazli Can. "Chat mining for gender prediction." In International conference on advances in information systems, pp. 274-283. Springer, Berlin, Heidelberg, 2006.

[17] R. Alroobaea, A.H. Almulih, F. S. Alharithi, S., Mechti, M .Krichen. and L.H. Belguith., A Deep learning Model to predict gender, age and occupation of the celebrities based on tweets followers. 2020.

[18] J. Pennebaker . The secret life of pronouns: What our words say about us. New York,USA, 2011.

[19] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval, information processing and management, pages 513--523, 1988.

[20] G. Marcus. Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631. 2018.

[21] M.J. Willis, C. Di Massimo, G.A. Montague, M.T. Tham and A.J. Morris. May. Artificial neural networks in process engineering. In IEE Proceedings D (Control Theory and Applications) (Vol. 138, No. 3, pp. 256-266). IET Digital Library. 1991.

[22] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun and R. Fergus. Regularization of neural networks using dropout. In International conference on machine learning (pp. 1058-1066). 2013.

[23] H. Sak, A. Senior and F. Beaufays. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128. , 2014

- [24] J. Brownlee. Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning. Machine Learning Mastery, 2017.
- [25] A.C. Müller and S. Guido. Introduction to machine learning with Python: a guide for data scientists. "O'Reilly Media, Inc.".2016.
- [26] A. Kulkarni, Y. Pino, & T. Mohsenin., SVM-based real-time hardware Trojan detection for many-core platform. In 2016 17th International Symposium on Quality Electronic Design (ISQED) (pp. 362-367). IEEE. 2016.
- [27] A. Alsufyani, R. Alrobaea & K.A. Ahmed. Detection of single-trial EEG of the neural correlates of familiar faces recognition using machine-learning algorithms. International Journal of Advanced Trends in Computer Science and Engineering. [Online] 8 (6). Available from: doi:10.30534/ijatcse/2019/28862019. 2019.
- [28] M. Aykanat, and P. Bendevis. Keras Accuracy Does Not Change. [online] Stack Overflow. Available at: <<https://stackoverflow.com/questions/37213388/keras-accuracy-does-not-change>> [Accessed 4 November 2020].
- [29] TensorFlow. Tf.Keras.Losses.Binarycrossentropy | Tensorflow Core V2.3.0. [online] Available at: <[https://www.tensorflow.org/api\\_docs/python/tf/keras/losses/BinaryCrossentropy](https://www.tensorflow.org/api_docs/python/tf/keras/losses/BinaryCrossentropy)>.
- [30] b. chandra, P. Nayak and n. swami. How Does Sigmoid Activation Work In Multi-Class Classification Problems. [online] Data Science Stack Exchange. Available at: <<https://datascience.stackexchange.com/a/72731>> [Accessed 4 November 2020].
- [31] M. Aykanat, and P. Bendevis. Keras Accuracy Does Not Change. [online] Stack Overflow. Available at: <<https://stackoverflow.com/questions/37213388/keras-accuracy-does-not-change>> [Accessed 4 November 2020].
- [32] A., Sboev, T., Litvinova, D., Gudovskikh, R., Rybka, & I. Moloshnikov. Machine learning models of text categorization by author gender using topic-independent features. Procedia Computer Science, 101, 135-142, 2016.
- [33] W. Ahmed and M., Bahador. The accuracy of the lstm model for predicting the s&p 500 index and the difference between prediction and backtesting. 2018.
- [34] S. Mechti, M. Jaoua, R. Faiz, H. Bouhamed, L.H. Belguith. Author Profiling: Age Prediction Based on Advanced Bayesian Networks. Research in Computing Science; 110: 129–137.2016
- [35] A.I.V. Valencia, H.G. Adorno, C.S. Rhodes, and G.F. Pineda. Bots and Gender Identification Based on Stylometry of Tweet Minimal Structure and n-grams Model. 2019.
- [36] M., Farhadian , P. Shokouhi and P. Torkezaban. A decision support system based on support vector machine for diagnosis of periodontal disease. BMC Research Notes, 13(1), pp.1-6. 2020.
- [37] D., Martins, R., Assis, R. Coelho and F. Almeida. Decision Support System for Business Ideas Competitions. Journal of Information Systems Engineering and Management,4(3), 2019.