International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

A Word Embeddings based Approach for Author Profiling: Gender and Age Prediction

Karunakar Kavuri[1], M Kavitha[2]
[1]Department of Computer Science and Engineering, Research Scholar
VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, Tamilnadu, India
karunakar.mtech@gmail.com
[2]Department of Computer Science and Engineering, Professor
VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, Tamilnadu, India
kavitha@veltech.edu.in

Abstract— Author Profiling (AP) is a method of identifying the demographic profiles such as age, gender, location, native language and personality traits of an author by processing their written texts. The AP techniques are used in multiple applications such as literary research, marketing, forensics and security. The researchers identified various differences in the authors writing styles by analysing various datasets. The differences in writing styles are represented as stylistic features. The researchers extracted several style based features like structural, content, word, character, syntactic, readability and semantic features to recognize the profiles of the authors. Traditionally, the researchers extracted various feature combinations for differentiating the profiles of authors. Several existing works are used Machine Learning (ML) methods for predicting the author characteristics of a new author. The existing works achieved good accuracies for predicting the author characteristics by considering the both stylistic features and ML algorithms combination. Recently, in advent of Deep Learning (DL) techniques the researchers are proposed approaches to author profiling by using these techniques. Few researchers identified that the deep learning techniques performance is good for author profiles prediction than the results of style based features. In this work, a word embeddings based approach is proposed for gender and age prediction. In this approach, the experiment conducted with different word embedding models such as Word2Vec, GloVe, FastText and BERT for generating word vectors for words. The documents are converted as vectors by using the document representation technique which uses the word embeddings of words. The document vectors are transferred to three different ML algorithms such as Extreme Gradient Boosting (XGBoost), Random Forest (RF) and Logistic Regression (LR) for generating the trained model. This model is used for predicating the accuracy of age and gender prediction. The XGBoost classifier with word embeddings of BERT achieved good accuracies for age and gender prediction than other word embeddings and ML algorithms. The experiment implemented on PAN 2014 competition Reviews dataset for age and gender prediction. The proposed approach attained best accuracies for predicting age and gender than the performances of various existing approaches proposed for AP.

Keywords- Author Profiling, Gender Prediction, Age Prediction, XGBoost, Word Embeddings, BERT

I. INTRODUCTION

In last decade, the textual data is tremendously growing in the web mainly through Twitter tweets, Facebook, Reviews, and Blogs etc. The crimes like creation of profiles with fake details and harassing messages also increasing in the web through textual data. Most of the authors are not specified their correct details while entering their data onto the web. The anonymous data in the web becomes a problem to the most of the businesses and government sectors. The researchers started research on this issue and developed a concept of analyzing the authorship of a document which is also named as Authorship Analysis (AA). The AA is a technique of determining the information of the authors by analyzing their texts. Mainly, AA is categorized into three varieties like Authorship Attribution (AAT), Authorship Verification (AV) and Authorship Profiling (AP) [1]. AAT is a technique for uncovering the author information of a document by examining the documents of multiple documents of known set of authors [2]. AV technique confirms that whether the suspected author written the unknown document or not by examining the suspected author's documents [3]. AP is a technique of detecting the author's characteristics of the text by inspecting the style of writing in the text [4].

The AP is exploited in different applications such as literary research, forensic analysis, marketing and security. In literary research, some researchers are claimed the content of anonymous text in the web. In this context, AP techniques are used to discover the information of the anonymous text. In marketing, the authors write reviews about products and services. The author details of reviews are very important for the companies to change their business strategies like change features of a product and service rules modification.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

Sometimes, some of the authors are not specified their details while writing the review. In this context, the methods of AP are used to discover the authorship details of the review by investigating the reviews. In forensic analysis, the property wills are processed to detect the correct author who wrote the property will. In this context, AP techniques are used to find the basic details of anonymous property wills. In security, the unauthorized sources or terrorist organizations forwarded threatening mails to the government sectors. Here, AP techniques are used to know the location (from where the mail came) and details of mails by analysing the mails textual content.

Different researchers identified writing style differences in different datasets. Koppel et al., identified [4] that more determiners and quantifiers are used by male whereas female used more pronouns and prepositions in their writings. They also observed that male used the words related to the topics of politics, women, sports and technology wherein the female concentrated on words of topics like shopping, beauty, jewellery and kitty party. Generally the authors define their own set of grammar rules and selection of words when they are writing about some topics. In another observation [5], they identified that male used more words related to politics and technology whereas female used more adverbs and adjectives and the words related to wedding styles and shopping. One researcher [6] observed that the content of a text and features of a text play a crucial role in the identification of the gender of the authors. They identified that the words like boyfriend, my husband, pink, family, negative words and emotional words are contained in more female writings and the male used more number of articles, prepositions, longer words, and statistics.

In the past, some of the researches classified the writing styles by categorizing the author into their age groups. These age groups categorized into 3 to 5 categories. In the literature, some researchers identified that the age group among 13 to 17 writes about school life and friends. The age group among 23 to 27 writes about heroes, heroines, college life, sex and premarital life. The authors in the age group of between 33 to 47 post their about kids, post marriage life, social activities etc. J.W. pennebaker et. al., recognized [7] that the use of idioms, prepositions, determiners will increase as age increase and also observed that older authors has tendency to use longer words, longer posts and more commas whereas the younger authors uses more articles, pronouns and less nouns.

The organizers of PAN included the Author Profiling task from 2013 competition onwards by changing the Applications, predicting demographic features and the dataset. In 2013 PAN competition [8], organizers concentrated on two profiles such as age and gender. The blogs dataset was given in two different languages such as Spanish and English for experimentation. In 2014 PAN competition [9], they used same profiles for prediction but increased the datasets such as social media, twitter, reviews, and blogs. 2013, 2014 competitions use same profiling characteristics to predict but the number of sub profiles varied in the context of age prediction. In 2013, they used three sub classes such as 10s (13 – 17), 20s (23 – 27), 30s (33 – 47) for age prediction whereas in 2014, five sub classes such as 18-24, 25-34, 35-49, 50-64, 65-xx for age prediction.

In this work, we used the PAN competition 2014 Reviews dataset for experimentation. Two profiles like gender and age are predicted in this work. We proposed a new word embeddings based approach for age and gender prediction. In this approach, the experiment implemented with 4 varieties of popular word embedding techniques such as BERT, FastText, GloVE and Word2Vec for converting words into word vectors. A document representation technique is defined in this approach for representing the documents as vectors. The document vectors are used to train the ML methods such as LR, RF and XGBoost for generating the ML model. This model is used for predicting the accuracy of gender and age prediction.

This work is arranged in IX sections. Section II discusses existing approaches proposed for author profiling. The dataset description is given in section III. The section IV explain machine learning techniques used in this experimentation. The proposed word embedding based approach for gender prediction is described in section V. the Section VI described the proposed word embedding based approach for age prediction. The section VII discusses different word embedding techniques that are used in the proposed approach. The section VIII presents the experimental results of this work. The conclusions and future directions of this work are specified in section IX.

II. EXISTING APPROACHES

Most of the existing works are proposed for author profiling by using machine learning algorithms and deep learning algorithms. Burger et al. [10] developed a model for gender classification of Twitter users. They experimented with various features such as character 1- to 5-grams and word unigrams and bigrams, and achieved 76% accuracy for gender prediction. They experimented with Balanced Winnow learning algorithm and observed that this algorithm shows good efficiency in terms of robustness, accuracy and speed when compared with linear SVM and Naive Bayes. Schler et al., extracted [11] 71000 blogs from blogger.com for experimentation. They experimented with a set of stylistic features such as 1000 word unigrams that are selected based on information gain scores, hyperlinks and POS N-Grams. The authors claimed that their adopted algorithm namely MCRW (Multi-Class Real Winnow) learning algorithm attained best accuracy for gender classification than the accuracy of SVM. The proposed method obtains 80% accuracy for gender classification.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

Weren et al., experimented [12] with different length based features like number of sentences, words and characters. They also used readability features like Flesch-Kincaid readability score and information retrieval features like cosine similarity in their experimentation. Marquardt et al., used [13] different style based features like number of capitalized words, emoticons, total number of posts, total grammatical and spelling mistakes, readability features, html tags, and number of capitalized letters and content-based features such as LIWC, MRC and sentiments in their experimentation.

Most of the approaches of AP follow common steps like pre-processing techniques, features extraction, features identification, vector representations of documents and ML algorithms to detect the author profiles information of a document. Rishabh Katna et al., applied [14] a machine learning approach in combination with Natural Language Processing (NLP) for author profiling. They used NLP methods such as Tokenization, lemmatization, and N-Grams of characters and words in combination with ML algorithms like Support Vector Machine (SVM), Decision Tree (DT), RF, and LR. Among four classifiers, the SVM obtained best accuracy for age and gender prediction. Ameer. Iqraa et al., presented [15] a content based method to predict the traits like age group and gender for author profiles of same genre. They used various features sets that include the character n-grams, word n-grams, POS n-grams and syntactic n-grams of POS tags in their proposed method. They identified from the results that the combination of word n-grams significantly attained good accuracies on well-known datasets.

Yaakov HaCohen-Kerner presented [16] a survey by focusing on two profiles such as gender and age, which are mostly used by several research works in author profiling. Author developed an overview of different datasets including datasets released in PAN competition and representative works in the field of author profiling with various important leaps. They also reviewed various Deep Learning (DL) techniques for predicting gender and age due to the enormous development of DL methods in recent times. Most of the gender and age datasets contain Twitter messages or blog posts that are written in Arabic, Spanish and English. There are also many datasets that are relevant to author profiling was written in Russian, Turkish, Portuguese, Italian, and Dutch. They observed that there is no uniformity and no consistency in the quality measures that are used for evaluating the results of classification, the datasets related to the type and number of their documents, the varieties of different pre-processing techniques that are applied on the dataset, the division of documents in the dataset into train, test and validation sets. An important interesting observation was the best accuracies for age group prediction was not as larger as the classification results of gender that contain only two sub-profiles. Other interesting observation was several ML classification methods were better than DL techniques for prediction of gender and age. Many classical approaches for author profiling used 3-4-5-Grams of characters and unigrams of words and bigrams of words. Various approaches also used different varieties of stylistic features. While several traditional systems do not consider any pre-processing techniques, most of the latest systems implemented various pre-processing techniques such as conversion of characters into lowercase and substitution of different strings (for example User Mentions, LF characters and URLs). They also recommended many possible future directions in gender and age profiling research.

Seifeddine Mechti et al., developed [17] an approach for author profiling. The major goal of their research work was to predict the age and gender of author. The authors implemented combination of classical ML methods and latest proposed DL methods. More specifically, they adopted the DL method of GRU (Gated Recurrent Unit) model. From their findings in their work, they identified that the proposed work results are outperformed than most of the well-known methods to AP. Piot–Perez-Abadin developed [18] a method for detecting the profile like gender. They considered both semantic and psycho-linguistic features in their experiment. The authors analyzed the detailed impact of linguistic features for the model's classification accuracy. Danique Sabel applied [19] various ML techniques such as SVM, RF, DT and KNN also for determining gender of author, where features developed by using word knowledge dataset with reaction times and word prevalence scores. Roobaea Alroobaeaet al., developed [20] an approach for constructing a DSS (Decision Support System) to predict age and gender from Twitter Tweets. The proposed system implemented with various Deep Learning (DL) methods such as DNN (Deep Neural Networks), LSTM and CNN techniques, and different ML techniques of RF, SVM, DT, KNN and NB for distinguishing profiles of gender and age. This work experimented on the dataset of PAN 2019 competition author profiling. The proposed DSS outperform results that are achieved in the research works of CLEF 2019 conference.

III. DATASET CHARACTERISTICS

The dataset was gathered from PAN competition 2014 reviews dataset. The dataset comprises of 4160 authors reviews. The Table I displays the description about the reviews dataset for Gender Prediction.

TABLE I. THE DETAILS OF REVIEWS DATASET FOR GENDER PREDICTION

Dataset	Number of Authors	Number of Reviews	Number of Male Authors	Number of Female Authors
Reviews	4160	5453	2080	2080

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

In Table I, the dataset is balanced for gender prediction which means that the number of male documents and female documents are equal. The Table II displays the description about the dataset for age prediction.

TABLE II. THE DETAILS OF REVIEWS DATASET FOR AGE PREDICTION

Gender	Age Group	Number of Authors	Number of Reviews
Male	18_24	180	228
	25_34	500	700
	35_49	500	707
	50_64	500	669
	65+	400	520
Female	18_24	180	208
	25_34	500	651
	35_49	500	659
	50_64	500	617
	65+	400	494

The effectiveness of proposed approaches in TC is evaluated by using various classification algorithms. These algorithms present the results by using various measures like recall, precision, F1 Score, support, and accuracy. Different measures considered various kinds of information for evaluating the efficiency. In this work, the results are displayed by using accuracy measure. Accuracy is the ratio between count of test documents correctly predicted and total count of documents in the test dataset. The ranges of accuracy values are 0 to 1.

IV. MACHINE LEARNING ALGORITHMS

The Machine learning algorithms doesn't follow any programming instructions to train the classification model and these algorithms develop a model by observing the dataset characteristics [21]. The machine learning techniques classified into different classes such as unsupervised, supervised, transfer learning and reinforcement [22] which are usually applied for different kind of problems and domains. In supervised learning techniques, all samples or instances of training dataset are labeled with class names and these known data are used for training the ML techniques. The goal of supervised learning is producing a function to map previously unseen samples to known class labels [23]. Regression and classification are two well-known techniques of supervised learning. Classification is a type of supervised learning which predicts the class label of an anonymous document [24]. Over-fitting is one big problem in classification process. Over-fitting is avoided by representing the documents with fewer features. Identification of fewer features from all feature sets is a complex task and this task is solved by using the concept of feature selection methods. In unsupervised learning techniques, the samples of dataset are not having any class labels. In reinforcement learning techniques, extract the sequence of actions to interact with the environment directly.

The performance of classification and learning time of classifier mainly depends on the set of features used in vector representation. In this work, various ML algorithms such as LR, RF and XGBoost are used in the experiment for training the datasets, which results in creating the model for finding the accuracy of the gender and age prediction.

A. Logistic Regression (LR)

LR is a method for classification which is originated from the statistics field [25]. LR was effectively used for resolving various real-world problems because of its less computational cost. It is a general method for solving problems of binary classification. The main aim of LR is determining the weight of every input coefficient which is similar to linear regression. LR works based on the logistic function which is also called as transformation function. The logistic function is using probabilities or rules for predicting the class label. This function converts the values in the range of 0 to 1. The LR classifier performance is improved when removing the attributes which are correlated or attributes which are not interacting with the output.

B. Random Forest (RF)

The Decision Tree (DT) classifier is one among the non-parametric learning techniques in data mining and it is implemented on categorical, numerical or combination of both data types [26]. DT constructs a tree structure based classification model for mapping instances to suitable class labels. In a DT, each internal node indicates a feature which is used for taking decision about the given instance. The arcs are used for connecting two internal nodes and internal node to leaf, which is labeled as the outcome of a feature. Each leaf in the tree denotes the class label. The class label of a previously unseen sample is recognized based on the path reached or satisfied from root to a leaf in the tree, corresponding class label of leaf is assigned to new instance. Breiman developed [21] classification method of Random Forest to average the decisions of various decision trees for reducing the variance. RF evades the general problem of over-fitting which is paced by DTs by implementing a usual method of bagging or bootstrap aggregating to the DT learners. Therefore, it is a better fit for unstructured textual data.

C. Extreme Gradient Boosting (XGBoost) classifier

In recent times, the researchers are concentrating more on tree boosting algorithms because of its high speed and high accuracy in the performance. One of the most recent tree boosting methods is XGBoost which is widely used, scalable, efficient, powerful, highly robust and can be used by downloading the XGBoost library. Tianqi Chen et al., proposed

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

[27] XGBoost algorithm, which is a supervised learning algorithm. Moreover, XGBoost has won all the competitions on machine learning in recent time which influences us to take a look on XGBoost. XGBoost works based on a technique of gradient boosting which follows the similar tree structure that different tree boosting algorithms follows. The differences which made XGBoost superior than the other tree boosting algorithms like GBM (Gradient Boosting Machine) and Adaboost, is the internal functionality and the attributes of it. In the tree structure, the head node denotes the condition and the process divides into branches based on this condition. The tree goes down into deeper based on the tree length specified in the model. The node becomes final decision node or leaf node based on where tree ends.

V. WORD EMBEDDING BASED APPROACH FOR GENDER PREDICTION

In this work, we proposed a new word embedding based approach for gender prediction. The architecture of proposed approach is presented in Figure 1.

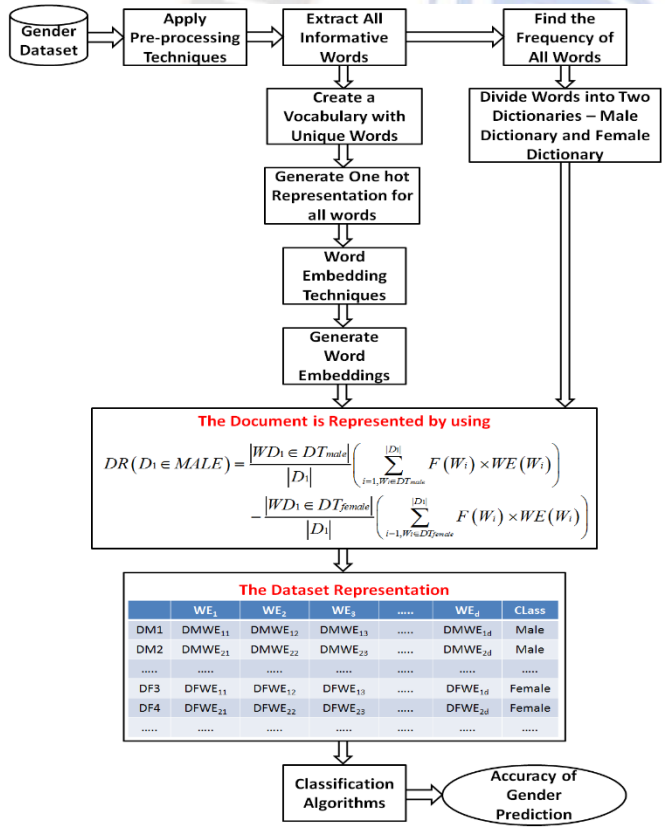

Figure 1. The Word Embedding based Approach for Gender Prediction

In this proposed approach, the first step is identification suitable dataset for experiment. Once the dataset is ready, the next step is applying suitable pre-processing techniques based on the type of word embedding technique is used. The pre-processing techniques are applied for two purposes such as one for identification of informative words and other for implementation of word embedding techniques. For informative words extraction, we implemented pre-processing techniques of stop-words elimination and lemmatization. After cleaning the datasets by using pre-processing techniques, then, collect all words from the dictionary and compute their frequencies. Based on the frequencies of words, all words are divided into two dictionaries like male dictionary and female dictionary. The male dictionary contains words which are having highest total frequency in all male documents than the frequency in female documents.

In other direction, the documents of dataset are passed to word embedding techniques to identify the word embedding of words in the dataset. In this work, four word embedding techniques such as BERT, FastText, GloVE and Word2Vec. These word embeddings techniques are used for generating word vectors. The documents are represented by using Equation (1). In Equation (1), the $\{D_1, D_2, \ldots, D_p\}$ are the set of documents in male dataset, WD_1 denotes the word in document D_1, DT_{male} and DT_{female} are the dictionaries of male and female sub-profiles respectively, $|D_1|$ is number of words in document D_1, W_i is i^{th} word, $F(W_i)$ the frequency of i^{th} word, $WE(W_i)$ the word embedding of i^{th} word.

$$DR(D_1 \in MALE) = \frac{|WD_1 \in DT_{male}|}{|D_1|}\left(\sum_{i=1,W_i \in DT_{male}}^{|D_1|} F(W_i) \times WE(W_i)\right) - \frac{|WD_1 \in DT_{female}|}{|D_1|}\left(\sum_{i=1,W_i \in DT_{female}}^{|D_1|} F(W_i) \times WE(W_i)\right) \quad (1)$$

The Equation (1) is used for representing a document that belongs to male sub-profile. In this representation, all the words in document D1 is tested for count of words that belongs to male dictionary and count of words that belongs to female dictionary. Once identifying the dictionaries of all words in document D_1, multiply each word embedding vector with the frequency of that word. Then, add all word embedding vectors that belongs to male dictionary. Then, normalize this word embedding vector by multiplying with proportion of words in document D_1 belongs to male dictionary. Likewise, add all word embedding vectors that belong to female dictionary and normalize by multiplying with proportion of words in document D_1 belongs to female dictionary. The final document representation is obtained by subtracting the word vectors of female dictionary from the word vectors of male dictionary. Similarly, all the male documents are represented by following this procedure.

$$DR(D_1 \in FEMALE) = \frac{|WD_1 \in DT_{female}|}{|D_1|}\left(\sum_{i=1,W_i \in DT_{female}}^{|D_1|} F(W_i) \times WE(W_i)\right) - \frac{|WD_1 \in DT_{male}|}{|D_1|}\left(\sum_{i=1,W_i \in DT_{male}}^{|D_1|} F(W_i) \times WE(W_i)\right) \quad (2)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

The Equation (2) is used for representing the documents that belongs to female sub-profile as vectors. Here, follows the same procedure followed in male document representation. For female document representation, subtract sum of word embeddings of words that belongs to male dictionary in document D_1 from the sum of embeddings of word embeddings of words that belongs to female dictionary in document D_1. All the document representation of female sub-profile dataset is created like this.

Once all the document vectors are ready, represent the document vectors which are suitable for training with machine learning algorithms. In dataset representation of Figure 1, {DM_1, DM_2, …, DM_p} is set of all male documents in gender dataset, {DF_1, DF_2, …, DF_q} is set of all female documents in gender dataset, $DMWE_{pd}$ is value of d^{th} dimension in word embedding vector of p^{th} document in male dataset, $DFWE_{pd}$ is value of d^{th} dimension in word embedding vector of p^{th} document in female dataset. In this work, three ML algorithms such as LR, RF and XGBoost are used for training the model. The developed model is used for predicting the gender prediction accuracy.

VI. Word Embeddings Based Approach for Age Prediction

The figure 2 shows the word embeddings based approach for age prediction. In this approach, collect the dataset from PAN competition 2014 English reviews age dataset. Apply suitable pre-processing techniques like stop-words elimination and lemmatization to obtain informative words from the dataset. Find the frequency of all words in the dataset. Create five dictionaries such as 18_24 (AG_1), 25_34 (AG_2), 35_49 (AG_3), 50_64 (AG_4), 65+ (AG_5) based on the frequency of words. The 18_24 dictionary contains words which are having highest total frequency in all 18_24 age group documents than the frequency in all other age group of documents. In other direction, forward the documents to different word embedding techniques to generate the word embeddings. The documents are represented by using Equation (3). In Equation (3), the {D_1, D_2, …., D_l} is the set of documents in 18_24 age group dataset, WD_1 denotes the word in document D_1, $DT_{18\text{-}24}$, $DT_{25\text{-}34}$, $DT_{35\text{-}49}$, $DT_{50\text{-}64}$, and DT_{65+} are the dictionaries of 18_24, 25_34, 35_49, 50_64, 65+ age sub-profiles respectively, $|D_1|$ is number of words in document D_1, W_i is i^{th} word, $F(W_i)$ the frequency of i^{th} word, $WE(W_i)$ the word embedding of i^{th} word.

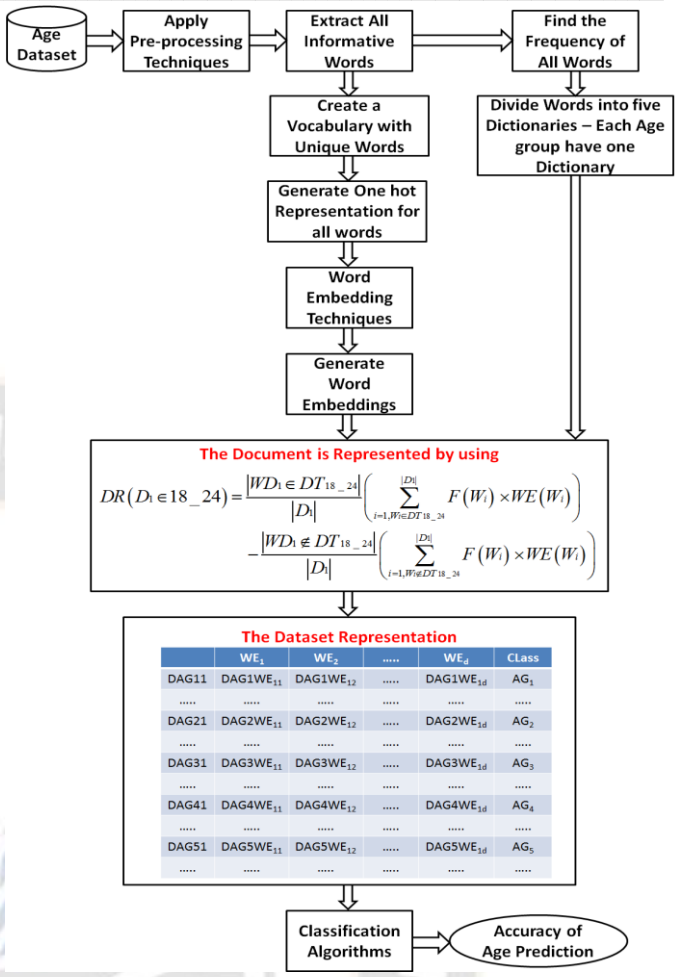

Figure 2. Word Embeddings based Approach for Age Prediction

$$DR(D_1 \in 18\_24) = \frac{|WD_1 \in DT_{18\_24}|}{|D_1|}\left(\sum_{i=1, W_i \in DT_{18\_24}}^{|D_1|} F(W_i) \times WE(W_i)\right) - \frac{|WD_1 \notin DT_{18\_24}|}{|D_1|}\left(\sum_{i=1, W_i \notin DT_{18\_24}}^{|D_1|} F(W_i) \times WE(W_i)\right) \quad (3)$$

The Equation (3) is used for representing a document that belongs to 18_24 sub-profile of age. In this representation, all the words in document D1 is tested for count of words that belongs to 18_24 dictionary and count of words that belongs to other age groups dictionary. Once identifying the dictionaries of all words in document D1, multiply each word embedding vector with the frequency of that word. Then, add all word embedding vectors that belongs to 18_24 dictionary. Then, normalize this word embedding vector by multiplying with proportion of words in document D1 belongs to 18_24 dictionary. Likewise, add all word embedding vectors that belong to other age group dictionaries and normalize by multiplying with proportion of words in document D1 belongs to other age group dictionaries. The final document representation is obtained by subtracting the word vectors of other age group dictionaries from the word vectors of 18_24

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

dictionary. Similarly, all the remaining documents of 18_24 age group are represented by following this procedure. Likewise, represent all the documents of other age group documents as vectors.

Once all the document vectors are ready, represent the document vectors which are suitable for training with machine learning algorithms. In dataset representation of Figure 2, DAG11, DAG21, DAG31, DAG41 and DAG51 are the D_1 documents of 18-24, 25_34, 35_49, 50_64, 65+ age groups respectively. $DAG1WE_{1d}$ is the value of d^{th} dimension in word embedding vector of document D_1 in 18_24 dataset, $DAG2WE_{1d}$ is the value of d^{th} dimension in word embedding vector of document D_1 in 25_34 dataset, $DAG3WE_{1d}$ is the value of d^{th} dimension in word embedding vector of document D_1 in 35_49 dataset, $DAG4WE_{1d}$ is the value of d^{th} dimension in word embedding vector of document D_1 in 50_64 dataset, $DAG5WE_{1d}$ is the value of d^{th} dimension in word embedding vector of document D_1 in 65+ dataset. In this work, three machine learning algorithms such as LR, RF and XGBoost are used for training the model. The developed model is used for predicting the age group prediction accuracy.

VII. WORD EMBEDDINGS

Bengio et al., (2003) introduced [28] the concept of word embeddings. Word embeddings can generally be seen as fixed-length word vectors, distributed, dense, and generated by using co-occurrence statistics of words. One-hot encoding is a simple method for representing words in numerical vector. This method assigns 1 to word in its corresponding index position in the vector representation and all other positions are filled with 0. This basic representation has various disadvantages. First, every word need to be represented with the size of the number of words in vocabulary, which results the vector representation of word contains most of 0's which is named as sparse representation. Secondly, the words are not arranged properly in the vocabulary so that the sematic information among the words and relationships among the words in documents are not captured in the vector representation.

The word embeddings are classified into two types such as "context-free" word embeddings and deep-contextual word embeddings. Context free word embeddings were generated by taking the advantage of words co-occurrence, there is an assumption that the semantic relationship exists rarely among words when they occurred together. The Word2Vec, fastText, and GLoVe are examples of "context-free" word embeddings and BERT is the method for generating deep-contextual word embeddings.

There are various hyper-parameters to impact the representation of the word embedding. Two hyper-parameter variables such as the dimensionality of a vector and the size of the training set are most important to enhance the performance. Word embeddings are used an algorithm for training the continuous-valued vectors and fixed-length dense vectors based on huge textual dataset. Every word is denoted as a point in the vector space and semantic relationships are preserved by moving these points around the target word. The words representation in vector space clusters the words within the space which are having similar meanings. Different word embedding techniques represent the same text into different numerical vector representations.

Google and Facebook was developed several models for understanding the actual meaning of every word present in the internet databases. Facebook was developed FastText model and Google was developed GloVe and Word2Vec models. Word2Vec was trained on approximately 100 billion numbers of words from a dataset of Google News to generate the word vectors for a vocabulary size of roughly 3 million words. Similarly, FastText and GloVe also trained on huge amount of data to generate word vectors. The embedding is the set of dimensions in which all words are represented based on the meaning of words and different contextual words surrounding to the target word.

A. Word2Vec Embeddings

The aim of Word2Vec is providing an effective implementation for both the CBOW ("Continuous Bag Of Words") and skip gram models for generating vector representations of words [29]. CBOW model detects a word based on the given context. Skip-gram model predicts the context words for a specified input word. The major idea behind Word2Vec is to train a model with context of every word, which generates similar numerical representations for similar words. In Word2Vec process, first, the sentence was divided into different words and produces the possible word pairs based on the size of window. For example, one of the word pair combinations is (X, Y), where X is the independent variable and Y is target dependent variable that we need to predict. We pass the X to the neural network through an embedding layer by initializing with random weights, and it was transferred through the softmax layer with an ultimate goal of predicting 'Y'. The optimization method like SGD (Stochastic Gradient Descent) was used to minimize the loss function ("(target word | context words)") of target word prediction for a given context words.

1) Continuous Bag Of Words (CBOW)

The steps in CBOW model when there is a single context word are:

- Both input and output layers are one hot encoding size of [1 X V], where V is the size of vocabulary.
- The network contains two varieties of weights such as one among the input layer and hidden layer and other

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

among hidden layer and output layer. The matrix size of input to hidden layer is [V X N] and the matrix size of hidden to output layer is [N X V]. Where, N is count of dimensions we decided to represent a word and also N represents neurons count in hidden layer.

- The activation function is not considered among the layers.
- Multiply the input with the weights among the input to hidden layer, which is named as hidden activation.
- The output is computed by multiplying the hidden input with the weights of hidden to output layer.
- Error among target and output is computed and back-propagate the error to weights re-adjustment.
- The weight among the hidden and output layer is considered as the vector representation of the word.

The process of CBOW model when there is k number of context words are as follows. The input layer takes input as 'k' number of one-hot encoded vectors. Input layer contains k number of [1 X V] vectors as input and the output of output layer is one [1 X V] vector. The remaining architecture is same as CBOW model with one context word. Most of the steps are similar except the process of computing the hidden activation. The hidden activation is computed as the average of input to hidden weight matrix of all k number of context words.

The major benefits of CBOW model are it shows superior performance for deterministic methods due to probabilistic nature, it consumes less memory which means that its runs on small sized RAM environments also. The main drawback of CBOW model is the context of words are not utilised properly in vector representation of words. For example, Apple is used in the context of fruit and a company, but CBOW compute the average of these contexts to generate the word vector.

2) Skip-gram

Skip-gram follows the reverse process of CBOW model. The goal of skip-gram is predicting the context words for a given word. The topology of skip-gram model was same as CBOW model. The skip-gram architecture is obtained by simply flipping the architecture of CBOW model. The input for the input layer of skip-gram model is same as the input process of one context CBOW model. Also, the computations up to activations of hidden layer are going to be same. The difference is in the computation of target variable. The computation of loss function is also similar to CBOW model. After training, the weights among the input layer and hidden layer are considered as the vector representations of words.

The one-hot encoding size of input in input layer is [1 X V], N is neurons count in hidden layer, the weight matrix size of input-hidden layer is [V X N] and the output size of each context word in output layer is [1 X V].

The process of skip-gram model when there are k number of context words

- Input layer take one word vector size of [1 X V] and output layer contains k number of context words vectors size of each vector is [1 X V].
- Hidden activation is computed by multiplying the input vector with the weight matrix of input-hidden layer. Hidden activation is computed for each target context word and the final activation is the average of all hidden activations.
- Error value is calculated by comparing the target context word with output context word. k number of errors obtained in the network when there is k number of context words. The final error of architecture is the sum of all the k number of errors.
- The error value is back-propagated for updating the weights.

The main benefits of skip-gram model are the skip-gram model generates different sematic vectors for same word based on the context the word is used, and the skip-gram model shows best performance than all methods when it is combined with negative sub-sampling.

In this work, skip-gram model is used for generating all word embeddings.

B. GloVe Embeddings

GloVe model generates word vectors by following an unsupervised learning algorithm [30]. This model is trained by using statistics of aggregated global word co-occurrence from a dataset. The initial step is to create a word co-occurrence matrix where each cell counts how many times one word appears in the context of another word. This can also be used to count how many times any word occurs in the context of a given word, and the probability of a word that will appear in the context of a given word.

For example, the "solid" word probability is closely associated to "steam" or "ice". The probabilities of these word co-occurrences in the GloVe example corpus are P(solid|ice) is more than P(solid|steam). The ratio of these two probabilities i.e P(solid|ice)/P(solid|steam) is higher than P(solid|steam)/P(solid|ice) which means that ice correlates better with the word "ice" than it does with the word "steam".

The computational complexity of this GloVe model mainly depends on the nonzero elements count in the probability matrix. We need to understand about two important methods such as global matrix factorization and local context window for understanding the functionality of GloVe. In Natural Language Processing (NLP), the global matrix factorization is a method of reducing the size of large term frequency matrices by using the matrix factorization methods from linear algebra. These matrices generally represent the absence or appearance of words in a document. Global matrix factorizations are called

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

as Latent Semantic Analysis (LSA) when they were applied on term frequency matrices. The skip-gram and CBOW models are used as methods of Local context window. The skip-gram models works effectively when training data contains small amounts of data and denotes even rare words also, whereas CBOW model has slightly good accuracies for frequent words and the CBOW have faster training times. Glove showed good efficiency when compared with other contemporaries.

C. FastText Embeddings

The Facebook handles the massive amount of textual content in the form of comments, status updates etc., on daily basis. The utilization of this textual content is more crucial for Facebook to serve the requirements of users in a better way. The computation of word representations was highly expensive and time consuming task for the Facebook by using the textual content produced by the billions of users. To overcome these problems, the research team of Facebook developed its own library named as FastText for representing words as vectors and for text classification [31].

FastText is an extension to Word2Vec model. The FastText model represents every word with character n-grams instead of learning words directly to generate the vectors. For example, consider the word "intelligence", the n value is 3, then the representation of this word in FastText model is <in, int, nte, tel, ell, lli, lig, ige, gen, enc, nce, ce>, where, the angular brackets denotes the word beginning and ending. This process is helped to capture the shorter words meaning and allows the embeddings for understanding prefixes and suffixes. Once the word is represented with n-grams of characters, then, a skip-gram model is trained on these n-grams to learn the embeddings. FastText functioning is very good with rare or infrequent words. It can split the word into different character n-grams to generate its embeddings when the word was not seen during the process of training.

FastText assumes a word is formed with character n-grams. The FastText offers several benefits over GloVe and Word2Vec by using the new word vector representation.

- It is more helpful to generate the vector representations for rare words. Rare words are divided into n-grams of characters and these n-grams are shared with common words.
- It can generate vector representations for the words which are not available in the dictionary also called as OOV (Out-Of-Vocabulary) words. Later, these words can be broken into n-grams of characters. Both GloVe and Word2Vec failed to generate vector representations for words which are not in the dictionary.
- The embeddings of character n-grams shows superior performance on smaller datasets when compared with Glove and Word2Vec.

D. BERT Embeddings

The BERT model was trained on Book Corpus which contains more than 10000 books of various genres. BERT ("Bidirectional Encoder Representations from Transformers") follows up on transformers to use them in a bidirectional manner [32]. BERT follows ELMo in that training involves a split between two tasks such as pre-training and fine-tuning. The "pre-training" step has two tasks such as MLM ("Masked Language Modelling") and NSP ("Next Sentence Prediction") to improve the performance of BERT. In MLM, a percentage of input tokens (commonly 15%) masked and those tokens will be predicted by a classification layer. In NSP, pairs of sentences will be input and a classifier will try to predict if given sentence B is the correct sentence that follows sentence A. Fine tuning is the method of considering the embeddings generated in the task of pre-training and then applying them to various NLP tasks. These tasks include but are not limited to part of speech tagging, answering tasks, and sentence prediction.

BERT is currently designed in two models based on the number of transformer layers it contains. The base BERT has 12 x transformers with 110 million parameters, and the large BERT includes 24 x transformers with 340 million parameters. The transformer layers helped to implement the Attention mechanism [33]. Each transformer layer contains a large feed-forward network. Traversing both directions within the features helps BERT to gather context and subsequently knowledge which could be fine-tuned for NLP tasks. The base BERT and large BERT contain hidden units of 768 and 1024 respectively in each transformer layer. They also contain attention heads of 12 and 16 respectively.

In this work, we experimented with base BERT. In base BERT, each word or sentence or document is represented with 768 values of word embedding. In the architecture of base BERT, each transformer layer transfers the 768 vector representation of word to next layers. The BERT environment is used for multipurpose like classification and feature extraction. In this work, we used the services of BERT for extracting features like representation of words as vectors. In both tasks, the BERT Tokenizer is used to split the document into tokens or words. In classification purpose, add the [CLS] token in first position of review sentence and [SEP] token is added as a separator in between sentences. In base BERT, the words of a sentence are passed through each layer. Self-attention is applied on every layer and forwards its outcomes through feed forward network to next layer. The output of word in each layer is the vector representation of 768 float values.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

After 12th layer, all word representations are discarded and only [CLS] token information is used for classification.

VIII. Experimental Results

In this work, the experiment conducted with a word embeddings based approach for age and gender prediction. In this approach, various word embedding models such as BERT, FastText, GloVe and Word2Vec for generating word vectors for words. The documents are converted as vectors by using the word embeddings of words in that document.

A. Experimental Results for Gender Prediction

The Table III displays the accuracies of logistic regression, Random Forest and XGBoost classifiers for gender prediction when experiment performed with different word embeddings for generating the vector representation for document.

TABLE III. The Accuracies of Logistic regression classifier for Gender prediction when different word embeddings are used to generate the document vector

Word Embedding / ML algorithms	**Gender Prediction Accuracy**		
	Logistic regression	***Random Forest***	***XGBoost classifier***
Word2Vec	0.8823	0.8956	0.9134
GloVe	0.8756	0.8889	0.9178
FastText	0.8789	0.9067	0.9267
BERT	0.8997	0.9123	0.9345

In Table III, the BERT embeddings with XGBoost classification algorithm attained best accuracy of 0.9345 for gender prediction than other word embedding techniques and classification algorithms.

B. Experimental Results for Age Prediction

Table IV represents the accuracies of Logistic regression classifier for age prediction when various word embedding techniques are used for generating the document vectors.

TABLE IV. The Accuracies of Logistic regression classifier for Age prediction when different word embeddings are used to generate the document vector

Word Embedding / ML algorithms	**Age Prediction Accuracy**		
	Logistic regression	***Random Forest***	***XGBoost classifier***
Word2Vec	0.8589	0.8656	0.8767
GloVe	0.8478	0.8712	0.8878
FastText	0.8512	0.8789	0.8823
BERT	0.8623	0.8878	0.8945

In Table IV, the BERT embeddings with XGBoost classification algorithm attained best accuracy of 0.8945 for age prediction than other word embedding techniques and classification algorithms.

The proposed word embeddings based approach attained best accuracies for gender and age prediction when compared with other proposed approaches of author profiling. The proposed approach used a novel document vector representation. The word embeddings are generated based on the global information of words. These embedding techniques are not considered the distribution information of words within a document and across documents in a dataset. In this work, we considered the importance of a word in document while generating word embedding vectors. This is the main reason, our proposed approach attained best accuracies.

IX. Conclusions and Future Work

In this work, we developed a word embedding based approach for gender and age prediction. The experiment performed with various word embedding techniques like BERT, FastText, GloVe and Word2Vec for generating the word vectors. The document vectors are represented by using a document representation technique. The document vectors are passed to three ML algorithms to identify the accuracy of gender and age prediction. The XGBoost classifier achieved best accuracies of 0.8945 and 0.9345 for age and gender prediction respectively when the word vectors are generated with the BERT model.

In future work, we are planning to implement this proposed approach on prediction of other characteristics like native language and location of authors. We are also planning to implement the combination of autoencoder and generative adversarial networks to increase the accuracy of characteristics prediction.

References

[1] Raghunadha Reddy T, Vishnu Vardhan B, Vijayapal Reddy P, "A Survey on Author Profiling Techniques", International Journal of Applied Engineering Research, March 2016, Volume-11, Issue-5, pp. 3092-3102.

[2] E. Stamatatos, "A Survey of Modern Authorship Attribution Methods", Journal of the American Society for Information Science and Technology, Vol.60, No.3, pp.538-556, 2009.

[3] M. Koppel, J. Schler, and E. Bonchek-Dokow, "Measuring differentiability: Unmasking pseudonymous authors", The Journal of Machine Learning Research, Vol.8, pp.1261–1276, 2007.

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

[5] Nerbonne, J., The secret life of pronouns. What our words say about us. 2013, ALLC.

[6] Newman, M.L., Groom, C.J., Handelman, L.D. and Pennebaker,J.W., "Gender differences in language use: An analysis of 14,000 text samples", Discourse Processes, Vol. 45, No. 3, (2008), 211-236.

[7] Pennebaker, J.W., Francis, M.E. and Booth, R.J., "Linguistic inquiry and word count: Liwc 2001", Mahway: Lawrence Erlbaum Associates, Vol. 71, No. 2001, (2001), 2001-2009.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

[8] Rangel, F., Rosso, P., Koppel, M., Stamatatos, E., Inches, G.: Overview of the author profiling task at PAN 2013. In CLEF Conference on Multilingual and Multimodal Information Access Evaluation, CELCT, pp. 352-365 (2013).

[9] Sakura Nakamura, Machine Learning in Environmental Monitoring and Pollution Control , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[10] Rangel Pardo, F., Rosso, P., Chugur, I., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., Daelemans, W.: Overview of the 2nd Author Profiling Task at PAN 2014. In: Cappellato, L., Ferro, N., Halvey, M., Kraaij, W. (eds.) CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers, 15-18 September, Sheffield, UK. CEUR Workshop Proceedings, CEUR-WS.org (Sep 2014)

[11] Burger, J.D., Henderson, J., Kim, G., Zarrella, G.: Discriminating Gender on Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1301–1309. EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA (2011).

[12] J. Schler, Moshe Koppel, S. Argamon and J. Pennebaker (2006), Effects of Age and Gender on Blogging, in Proc. of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs, March 2006. Vol. 6, (2006), 199-205.

[13] Edson RD Weren, Viviane Pereira Moreira, and Jose Palazzo M de Oliveira. Exploring information retrieval features for author profiling. In CLEF (Working Notes), pages 1164-1171, 2014.

[14] James Marquardt, Golnoosh Farnadi, Gayathri Vasudevan, Marie-Francine Moens, Sergio Davalos, Ankur Teredesai, and Martine De Cock. Age and gender identification in social media. Proceedings of CLEF 2014 Evaluation Labs, 2014.

[15] Rishabh Katna, Kashish Kalsi, Srajika Gupta, Divakar Yadav, Arun Kumar Yadav, "Machine learning based approaches for age and gender prediction from tweets", Multimedia Tools and ApplicationsVolume 81Issue 19Aug 2022 pp 27799–27817.

[16] Ameer. Iqraa, Sidorov. Grigoria, Nawab. Rao Muhammad Adeel, "Author profiling for age and gender using combinations of features of various types ", Journal of Intelligent & Fuzzy Systems, vol. 36, no. 5, pp. 4833-4843, 2019.

[17] Yaakov HaCohen-Kerner, "Survey on profiling age and gender of text authors", Expert Systems with Applications: An International Journal, Volume 199, Issue C, Aug 2022.

[18] Seifeddine Mechti, Moez Krichen, Dhouha Ben Noureddine,Lamia H. Belguith," A decision system for computational authors profiling: From machine learning to deep learning ", Concurrency and Computation Practiec and Experience, Special Issue, Wiley Online Library, 07 September 2020, https://doi.org/10.1002/cpe.5985.

[19] Piot–Perez-Abadin, P., Martin–Rodilla, P. and Parapar, J., "Experimental Analysis of the Relevance of Features and Effects on Gender Classification Models for Social Media Author Profiling", DOI: 10.5220/0010431901030113, In Proceedings of the 16th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2021), pages 103-113, ISBN: 978-989-758-508-1, 2021 by SCITEPRESS – Science and Technology Publications.

[20] Danique Sabel, "Gender Prediction based on Word Knowledge using Machine Learning Techniques", thesis submitted to Tilburg University, January 2019.

[21] Roobaea Alroobaea, Sali Alafif, Shomookh Alhomidi, "A Decision Support System for Detecting Age and Gender from Twitter Feeds based on a Comparative Experiments", International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 2020.

[22] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[23] S. J. Russell and P. Norvig, Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,, 2016.

[24] N. M. Nasrabadi, "Pattern recognition and machine learning," Journal of Electronic Imaging, vol. 16, no. 4, p. 049901, 2007.

[25] L. Olshen, C. J. Stone, et al., "Classification and regression trees," Wadsworth International Group, vol. 93, no. 99, p. 101, 1984.

[26] Prof. Prachiti Deshpande. (2016). Performance Analysis of RPL Routing Protocol for WBANs. International Journal of New Practices in Management and Engineering, 5(01), 14 - 21. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/43

[27] Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Baltic Journal of Modern Computing, 5(2), 221

[28] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.

[29] T. Chen and C. Guestrin, \Xgboost: A scalable tree boosting system", in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785-794, isbn: 978-1-4503-4232-2.

[30] Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. "A neural probabilistic language model. journal of machine learning research, Vol. 3, No." (2003): 1137-1155.

[31] Chaudhary, A. ., Sharma, A. ., & Gupta, N. . (2023). A Novel Approach to Blockchain and Deep Learning in the field of Steganography. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 104–115. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2514

[32] X. RONG, word2vec parameter learning explained, arXiv preprint arXiv:1411.2738, (2014).

[33] J. PENNINGTON, R. SOCHER, AND C. MANNING, Glove: Global vectors for word representation, in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6996
***Article Received**: 29 March 2023 **Revised**: 15 May 2023 **Accepted**: 30 May 2023*

[34] A. JOULIN, E. GRAVE, P. BOJANOWSKI, AND T. MIKOLOV, Bag of tricks for efficient text classification, arXiv preprint arXiv:1607.01759, (2016).

[35] J. DEVLIN, M.-W. CHANG, K. LEE, AND K. TOUTANOVA, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, (2018).

[36] Yathiraju, D. . (2022). Blockchain Based 5g Heterogeneous Networks Using Privacy Federated Learning with Internet of Things. Research Journal of Computer Systems and Engineering, 3(1), 21–28. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/37

[37] Karunakar. Kavuri and M. Kavitha, "A Term Weight Measure based Approach for Author Profiling," 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), 2022, pp. 275-280, doi: 10.1109/ICESIC53714.2022.9783526.

[38] Karunakar Kavuri, Kavitha, M. (2020). "A Stylistic Features Based Approach for Author Profiling". In: Sharma, H., Pundir, A., Yadav, N., Sharma, A., Das, S. (eds) Recent Trends in Communication and Intelligent Systems. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-0426-6_20

[39] Martínez, L., Milić, M., Popova, E., Smit, S., & Goldberg, R. Machine Learning Approaches for Human Activity Recognition. Kuwait Journal of Machine Learning, 1(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/146

[40] A. VASWANI, N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, Ł. KAISER, AND I. POLOSUKHIN, Attention is all you need, in Advances in neural information processing systems, 2017, pp. 5998–6008.