International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 16 (2024) pp. 623–635
© Cerebration Science Publishing

Article

Transformer Models for Authorship Profiling in Arabic Social Media Texts

Bsir Bassem [1,*], Ajith Abraham [2] and Zrigui Mounir [1]

[1] RLANTIS, University of Monastir, Hammam Sousse 4011, Tunisia; mounir.zrigui@fsm.rnu.tn
[2] School of Artificial Intelligence, Bennett University, Greater Noida 201310, UP, India; ajith.abraham@ieee.org
* Correspondence author: bsir.bassem@yahoo.fr

Received date: 23 November 2023; Accepted date: 26 June 2024; Published online: 10 July 2024

Abstract: Authorship analysis is generally performed to extract information about the author or authors of documents by examining the inherent features present within these texts. The significance of the authorship profiling task is escalating, particularly with the widespread rise in social media users and platforms. This research explores Multi-Task Learning (MTL) with transformers, focusing on refining and evaluating advanced pre-trained models for author profiling using Arabic tweets. ARBERT, MARBERT, AraBERT, and BERT base Arabic transformers undergo fine-tuning for binary classification tasks in Arabic tweet analysis. The study incorporates MTL by concurrently training models on tasks such as dialect identification, sentiment analysis, and topic classification to enhance overall performance. Parameter optimization plays a pivotal role, achieving reliable results with AraBERT demonstrating the highest F1 score on the test dataset. MTL integration showcases promising outcomes, reinforcing the transformers' efficacy in authorship profiling.

Keywords: authorship profiling; Multi-Task Learning; AraBERT; BERT; transformer-models; deep learning; information retrieval; transformers

1. Introduction

The objective of author profiling is to address inquiries such as: how can we discern variations in writing styles on social networks among different demographics like gender, age groups, locations, or psychological profiles? By tackling these questions, we aim to resolve contemporary challenges within the social network era, such as combating fake news, detecting plagiarism, and optimizing marketing strategies. Gender detection stands out as one of the prominent sub-tasks in Author Profiling (AP) [1]. However, most research efforts in AP have focused on appropriately categorizing the author's profile based on textual analysis [2].

With the robust advancement in computing power of contemporary hardware, pre-trained deep learning models, trained on vast datasets, have exhibited their superiority over conventional methods. This substantial progress primarily stems from the exceptional representational capacity of transformers and their variant architectures. Indeed, Transformers and their derivatives have showcased remarkable success as potent unsupervised or self-supervised pretraining frameworks across various natural language processing tasks. For instance, models like GPTs are trained in an autoregressive manner, predicting the next word within extensive text datasets. BERT, on the other hand, learns from data without explicit supervision, predicting masked words based on contextual cues. Devlin et al. [3] introduced a universal pre-training framework applicable to multiple downstream tasks, while Y. Liu et al. [4] proposed a robust variant of the original BERT [5].

Certainly, Transformer-based pre-trained language models have shown impressive performance in learning language representations from large amounts of unlabeled data, particularly in English and other high-resource languages. However, there is a crucial need to enhance their accuracy for Arabic gender identification tasks, particularly in the tokenization phase of data processing. Multi-Task Learning (MTL) emerges as a pivotal approach to address this challenge by concurrently training models on multiple related tasks such as dialect identification, sentiment analysis, and topic classification. This simultaneous learning process allows the models to leverage shared knowledge and linguistic patterns across tasks, thereby potentially improving their effectiveness in Arabic gender identification. Indeed, models like AraBERT, MARBERT, and other transformer variants specifically tailored or trained for Arabic have demonstrated robust capabilities in understanding and generating Arabic text. They have excelled in various NLP tasks such as sentiment analysis, named entity recognition, machine translation, and document classification [6,7]. Integrating MTL not only enhances their performance but also underscores its significance in advancing the capabilities of transformer models for complex language tasks in Arabic.

The rest of the paper is organized as follows. Section 2 discusses various studies in the domain of author profiling based on statistical and deep learning-based models. In Section 3, an author profiling method is proposed for a social media corpus with minimal manual intervention. We investigate ARBERT, MARBERT, AraBERT and BERT base Arabic transformers. Section 4 illustrates the results achieved.

2. Related Works

Natural Language Processing (NLP) operates within artificial intelligence, centering on the interface between computers and human language. Its scope encompasses various tasks like text understanding, language generation, machine translation, and sentiment analysis [8,9]. Within NLP, Authorship Profiling stands as one domain, involving the analysis of written text to unveil traits of the author, such as gender, age, writing style, or other demographics. While NLP techniques can be utilized in authorship profiling, it's important to note that they aren't interchangeable [10,11].

Authorship Profiling, a field with applications in forensic investigations, legal linguistics, and cybercrime, has seen significant advancements in recent years [12]. A range of studies have been observed in the related field of author profiling, that is, the computational task of learning author demographics from text such as gender, age, and others.

The introduction of transformers has led to a paradigm shift in NLP, with the starting point for training models moving from specific to general-purpose pretrained architectures. The use of transformers in ASR has also been explored, with the proposal of replacing positional embedding with convolutional learned input representations [13,14]. Despite their success, the implementation of transformers in real-world applications presents challenges such as computational efficiency, interpretability, and ethical considerations [15,16].

Devlin et al. [17] proved that the main challenge in NLP consists of the small quantity of the training data. To deal with this issue, authors suggested transformer-based models trained on huge unlabeled datasets (e.g., Wikipedia's dataset). Thus, the authors were able to apply the pretrained models on smaller datasets without the need for developing training models from scratch. Even though the proposed technique provided high accuracy in executing various NLP tasks, "fine-tuning" should be performed on the pretrained models before being applied on smaller datasets. As example of the pretrained models, we can mention the bidirectional encoder representations from transformers (BERT) characterized by its directionality (i.e., it considers context when words have dual valence) Henry Tsai et al., [18].

Recent research has made significant strides in author profiling using transformer models.

Vaswani et al. [19] introduced a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers.

Ai, M [20] demonstrated the effectiveness of contrastive training and BERT transformers in generating authorship embedding's and verifying authorship. There methods have shown high accuracy in distinguishing between different authors. They achieved 94% accuracy when tasked to distinguish a text piece from a set of 100 authored texts. Vaswani et al. [21] presented BERT, an innovative language representation model acronymized as Bidirectional Encoder Representations from Transformers. BERT is engineered to pre-train deep bidirectional representations from unannotated text, incorporating both right and left context across all layers. This approach yielded remarkable advancements across eleven natural language processing tasks, elevating GLUE score to 80.5, MultiNLI accuracy to 86.7%, SQuAD v1.1 Test F1 to 93.2, and SQuAD v2.0 Test F1 to 83.1%. Radford et al. [22], employed unidirectional language models for pre-training, BERT utilizes masked language models, enabling the acquisition of deep bidirectional representations. Moreover, BERT minimizes the necessity for intricate, task-specific architectures. It emerges as the pioneering fine-tuning-based representation model, exhibiting state-of-the-art performance across various sentence-level and token-level tasks, surpassing much task-specific architecture.

Manolache et al. [23] introduced a dataset characterized by distinct input data distribution. They subsequently utilized this dataset to scrutinize the domain generalization proficiency of models under low-data conditions, and the performance variations upon employing the proposed PAN-2020 splits for fine-tuning. Their findings illustrate that these splits enhance models' ability to transfer knowledge over significantly dissimilar datasets.

Schlicht et al. [24] introduced a feature transformation method to address changes in an author's writing style over time, improving the performance of author identification. These studies collectively highlight the potential of transformer-based models in author profiling tasks. Some authors provided comprehensive overviews of the methodologies and techniques used in authorship analysis, with Misini specifically focusing on stylometric features and classification methods [25].

Zhang et al. [26] delved into the practical applications of authorship analysis, with Zhang proposing a framework for tracing cyber criminals through their online messages and Tamboli discussing the use of behavioral feature extraction and classification in authorship analysis and identification.

Grivas et al. [27] further enhanced the predictive power of these features by combining them with appropriate preprocessing steps, achieving successful gender, age, and personality prediction. They suggested a systematic categorization of features paired with tailored preprocessing procedures for each category. Their approach involved two main groups: stylometric and structural features, which encompassed trigrams and counts of Twitter-specific attributes, among others. Gender and age prediction were treated as classification tasks, while personality prediction was approached as a regression problem. Support Vector Machines (SVMs) were employed for gender and age classification, and Support Vector Machine Regression was utilized for personality prediction. These algorithms were applied to documents formed by aggregating the tweets of individual users. Ray et al. [28] addressed the high dimensionality features by proposing a new approach that aggregates term weights to capture the relationship between them. They proposed approach to address the high dimensionality feature space problem by aggregating the term weights to find the weight of a document against the profile of the authors. The proposed approach was experimented on reviews domain to predict the gender and age group of the authors using accuracy as a measure.

Bassem et al. [29] utilized a neural network (NN) model with Gated Recurrent Units (GRU) to ascertain the authors' identities by analyzing their Facebook and Twitter posts. The NNP model's input underwent preprocessing and was segmented into two distinct layers: the embedding layer and the stylometric features extraction phase. Specifically, the output of the embedding layer was connected to a bidirectional GRU layer, followed by an activation layer. Meanwhile, the stylometric features were normalized and directly appended to the same activation layer employed after the GRU layer. Despite their efforts, the authors found that their results fell short compared to the superior findings reported by Basile et al. in PAN' AP (2017).

Furthermore, Estruch et al. [30] advanced an early fusion model, enhancing fusion after the

decision-level single-source classification. Achieving a remarkable 91% Gender Identification (GI) accuracy on an English dataset sourced from Foursquare, Instagram, and Twitter in Singapore. Alvarez-Carmona et al. [31] introduced a method to assess authors' genders using multi-modal information, encompassing both texts and images. This approach involved learning multimodal representations using Gated Multimodal Units (GMUs). Impressively, accuracy rates of 0.74 and 0.81 were achieved in the multi-modal scenario for English, Spanish, and Arabic datasets, respectively. Moreover, the gold standard data was translated by Veenhoven et al. [32] into the language of interest. Bi-LSTM and CNN architectures were also utilized to solve the GI problem by considering PAN-AP (2018) dataset. By considering the RNN, the highest obtained GI accuracy was equal to 79.3%, 80.4% and 74.9% for English, Spanish and Arabic languages, respectively.

A. Conneau et al. [33], in 2019, train a transformer based masked language model on one hundred languages, using more than two terabytes of filtered CommonCrawl data. They dubbed XLM-R, significantly outperforms multilingual BERT (mBERT) on a variety of cross-lingual benchmarks, including +14.6% average accuracy on XNLI, +13% average F1 score on MLQA, and +2.4% F1 score on NER. XLM-R performs particularly well on low-resource languages, improving 15.7% in XNLI accuracy for Swahili and 11.4% for Urdu over previous XLM models.

For Arabic language, Mageed et al. [34] introduced the AraT5 model, a text-to-text transformer that outperformed the multilingual T5 model and the MARBERT model in both language understanding and generation tasks. This was further improved upon by Antoun et al. [35] with the development of AraGPT2, a pre-trained transformer specifically designed for Arabic language generation. They developed the first advanced model, AraGPT2, trained from scratch on a large Arabic corpus of internet text and news articles. Our largest model, AraGPT2-mega, has 1.46 billion parameters, which makes it the largest Arabic language model available.

Bsir et al. [36] fine-tuned Ara-BERTv2-large model and they achieved higher accuracy (79.7%) on the test data set compared to XLM-RoBERTa (76.8%), their results from the conducted experiments demonstrated that the developed Ara-BERTv2-large achieved state-of-the-art performance on

Arabic datasets.

Multilingual Language Models (LMs) are a crucial asset in the realm of Natural Language Processing (NLP). Among these, mBERT stands out as the multilingual adaptation of BERT [37], renowned for its encoder model architecture incorporating bidirectional representations from Transformers. Trained with a denoising objective, mBERT has undergone extensive training on Wikipedia, covering a staggering 104 languages, including Arabic. Similarly, XLM-R, another notable player in the multilingual LM arena, follows a Transformer-based approach, serving as a masked language model trained on an extensive dataset exceeding 2TB of CommonCrawl (CC) data spanning 100 languages. This includes comprehensive coverage for Arabic, totaling 2.9 billion tokens. On the Arabic-specific front, AraBERT emerges as a dedicated pre-trained language model tailored specifically for Arabic text processing. Built upon the BERTBase architecture, AraBERT draws its strength from a substantial corpus of 24GB of Modern Standard Arabic (MSA) data.

In addition, ARBERT and MARBERT, spearheaded by Mageed et al. [38], offer specialized BERT-based models catering to Arabic language processing. ARBERT is finely tuned for MSA, leveraging a dataset of 61 GB, while MARBERT extends its scope to encompass both MSA and dialects, harnessing a corpus of 128 GB. In fact, transfer learning encompasses various approaches aimed at leveraging prior knowledge or experience to enhance learning in new domains. These approaches, often termed differently by different authors, include multitask learning, lifelong learning, knowledge transfer, knowledge consolidation, model adaptation, concept drift, and covariance shift, among others. While some researchers distinguish between these methods, considering them as distinct approaches, others view transfer learning and multitask learning as interchangeable concepts. However, from our perspective, these methods can be seen as specific implementations of transfer learning tailored to different circumstances or methodologies [39]. For instance, model adaptation suits scenarios where there's a clear disparity in data distributions between the source and target domains, whereas covariance drift addresses situations where distributional changes occur gradually. Similarly, knowledge transfer occurs when models are sequentially trained from a source to a target, contrasting with multitask learning where both source and target models are trained concurrently. Despite their differences, the overarching objective of these methods remains consistent: transferring knowledge from external sources to improve the current learning task, leading to benefits such as accelerated convergence, more robust models, and reduced sensitivity to data variations. Transfer learning can be classified into several categories based on the conditions under which they are applied. Following the taxonomy proposed by Pan and Yang, we consider data and task as two primary factors influencing transfer learning. The data condition involves the feature space (e.g., audio or text) and its distribution (e.g., financial news or scientific papers), while the task condition pertains to the label space (e.g., speech phones or speaker identity) and the model employed (e.g., probabilistic models or neural models).

BERT-Base: Velankar et al. [40] introduced bidirectional Encoder Representation from Transformers (BERT) aims to train bidirectional representations from unlabeled datasets, capitalizing on collaborative left and right context phenomena across all layers. Despite its simplicity, BERT yields powerful results across various machine learning tasks. Fine-tuning a BERT model only requires adding a single layer for each new task, making it highly adaptable. Utilizing a masked language model (MLM), BERT masks random words from input and predicts their IDs based on contextual information from both left and right contexts, enhancing the training of bidirectional models. Additionally, BERT incorporates next sentence prediction (NSP). BERT-Base, although smaller in size compared to other models, offers faster computation, affordability, and robust performance.

BERT-Large: Like BERT-Base but with a larger size, BERT-Large is more computationally intensive and suitable for handling larger datasets. It has been applied in various tasks such as processing COVID-19-related content on Twitter, offensive tweet classification, and mental illness detection. Among different approaches, BERT-Large has demonstrated competitive performance, particularly excelling in syntactic abilities and classification tasks [41].

Butt et al. [42] proposed RoBERTa-Base an enhanced version of BERT, incorporating advancements such as training on more data with larger batch sizes, eliminating next sentence prediction, and adjusting masking patterns. While performing well across experimental setups, RoBERTa exhibits a stronger linguistic bias compared to BERT. It has been utilized in tasks like feature learning [43], mental illness detection, and informative tweet classification, showing superior performance.

RoBERTa-Large: RoBERTa-Large, while offering enhanced capabilities, requires more computational resources compared to RoBERTa-Base. It has been applied in tasks such as dialog history attention-based response selection [44,45], medication detection on Twitter, Dutch language modeling, and eye-tracking prediction, demonstrating promising results in various domains.

Victor Sanh et al. [46] introduced DistilBERT as a lighter and faster version of BERT; DistilBERT offers comparable language understanding with reduced computational requirements. It has been employed in tasks like sociopolitical news classification, sentiment analysis, and answer selection mechanisms, showing significant performance improvements. Additionally, retraining DistilBERT on different datasets has enhanced downstream task performance. Joulin, A et al. [47].

ALBERT-Base-v2: Collaboratively developed by Toyota Technologies and Google Research, ALBERT is a scalable successor to BERT, emphasizing faster training speeds and lower memory consumption. It has been applied in tasks such as contextualized sarcasm detection, fake news detection, and question answering on COVID-19, showcasing improved performance and efficiency. Kim. et al. [48].

XLM-RoBERTa-Base: This model combines RoBERTa with XLM-R, offering multilingual capabilities. It has been employed in tasks like offensive language detection, hope speech detection, and context disambiguation, demonstrating superior performance across multiple languages and domains. Lectra-Small: Proposed as an alternative to BERT's Masked Language Modeling, Electra replaces tokens with alternative samples, reducing computational requirements. It has been used in tasks such as fake news profiling and multiword selection, showing competitive performance and efficiency [49].

BART-Large: Introduced by Facebook researchers, BART combines encoder-decoder architecture like BERT and GPT2, excelling in tasks like question answering and text summarization. It has been utilized in applications such as automated speech recognition, supervised topic label generation, and query suggestions, exhibiting improved understanding of noise and complex queries. Additionally, BART has been extended to tasks like visual common-sense generation and knowledge-grounded conversation, achieving promising results [50,51]. In fact, Lin, X. et al. [52] proposed a novel multi-facet paradigm, namely TransRec, to bridge the LLMs to recommendation. Specifically, TransRec employs multi-facet identifiers that incorporate ID, title, and attribute, achieving both distinctiveness and semantics. They introduced a specialized data structure for TransRec to guarantee the in-corpus identifier generation and adopt substring indexing to encourage LLMs to generate from any position. They implement TransRec on two backbone LLMs, i.e., BART-large and LLaMA-7B.

3. Method

Figure 1 illustrates the utilization of Multi-Task Learning (MTL) in our approach, where models are trained concurrently on interconnected tasks including dialect identification, sentiment analysis, and gender identification. Additionally, a multi-lingual transformer model was fine-tuned using a substantial Arabic corpus.

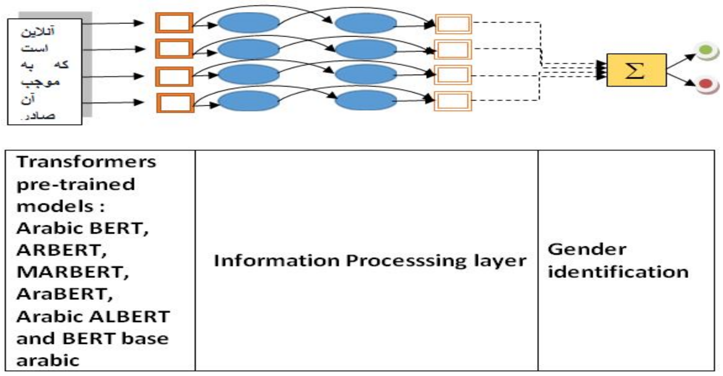

Figure 1. The architecture of Multi-Task Learning of different transformers models Approach.

By employing Multi-Task Learning (MTL), we integrated multiple pre-trained models into our approach to enhance the performance of gender identification during the fine-tuning process. Each pre-trained model was adapted and optimized through MTL, leveraging shared knowledge across tasks such as dialect identification, sentiment analysis, and gender classification. This simultaneous training on related tasks allowed the models to learn diverse linguistic features and nuances, thereby improving their accuracy and robustness in gender identification tasks specific to Arabic text.

3.1. Arbert

ARBERT, an evolution of the BERT model, underwent training on an expansive corpus of Modern Standard Arabic, encompassing 61 GB of text, equivalent to 6.5 billion tokens. This corpus, meticulously curated from a myriad of sources including literary works, news articles, web-crawled data, and Wikipedia, served as the bedrock for ARBERT's robust architecture. Mageed et al. [53] led this endeavor, propelling ARBERT into the forefront of Arabic language processing.

3.2. Marbert

MARBERT, a monumental feat in pre-trained language modeling, mirrors the foundational architecture of BERT base while achieving a groundbreaking milestone in Arabic text understanding. Enriched by a sprawling dataset comprising 128 GB of tweets from diverse Arabic dialects, MARBERT's training regimen was meticulously designed to preserve the authenticity of the original text. Minimal preprocessing was applied to maintain the essence of the tweets, ensuring a faithful representation of real-world language nuances. Mageed et al. [54] spearheaded this endeavor, propelling MARBERT as a cornerstone in Arabic natural language processing.

3.3. AraBERT

AraBERT stands as a testament to the ingenuity of BERT-based models in decoding the complexities of Modern Standard Arabic. Trained on a colossal dataset comprising 70 million sentences sourced from prominent Arabic repositories and news outlets, AraBERT underwent rigorous fine-tuning across multiple tasks including Sequence Classification, Named Entity Recognition, and Question Answering. W. Antoun et al. [55] meticulously engineered AraBERT to achieve state-of-the-art performance, extending its prowess even to Arabic dialects.

3.4. Bert Base Arabic

The inception of Bert base Arabic marks a significant advancement in Arabic language modeling, underpinned by a monumental corpus comprising 8.2 billion words drawn from diverse Arabic resources including OSCAR and recent dumps of Arabic Wikipedia. Safaya et al. [56] meticulously curated this vast dataset, employing meticulous preprocessing techniques to ensure its relevance and efficacy. Notably, this corpus encapsulates both Modern Standard Arabic and dialectical variations, thereby fortifying the model's adaptability to the dynamic landscape of Arabic language usage, particularly in social media contexts.

4. Experimental Results: Automatic Evaluation

4.1. Measures of Performance

We will measure the performance of automatic key phrase extraction methods using precision (P), recall (R), and F1-score (F) metrics. To calculate the performance of a method, i.e., precision (respectively recall), we first calculated the precision (respectively recall) for each tweet, and then took the average of these values.

4.2. Dataset

The dataset, sourced from Twitter, constitutes a pivotal component of the author profiling task within PAN@CLEF 2018. Each tweet collection comprises Arabic texts authored by 2,400 individuals, with each author contributing 100 tweets. Notably, the corpus encompasses four distinct varieties of the Arabic language: Egyptian, Gulf, Levantine, and Maghrebi.

4.3. Feature Extraction

Feature extraction methods are: Bow, TF-IDF and word embeddings.

- BOW:

The Bag-of-Words (BOW) method operates by considering the context of each word as input and aims to predict the word corresponding to that context. Each word in the vocabulary (denoted by V) is represented as a one-hot encoded vector, where only one element is set to 1, indicating the presence of that word. This one-hot encoded vector serves as the input or context word.Next, the input vector is multiplied by a weight matrix (Wvn), which has dimensions V*N, where N represents the number of neurons in the hidden layer. This weight matrix maps the input vector to the hidden layer. In the hidden layer, each neuron computes a weighted sum of the input values, followed by the application of a non-linear activation function, such as the sigmoid or ReLU function. These computations effectively transform the input data into a format that can capture complex patterns and relationships. In the final output layer, each neuron computes a weighted sum of the values from the hidden layer and applies the softmax function to produce a V-length vector. This vector contains the probabilities of each word being the target word, with higher probabilities indicating higher likelihoods.

- TF-IDF

TF-IDF is a method used to reduce the impact of frequently occurring words in a text corpus that do not provide much useful information. Tf-idf is calculated by multiplying the term frequency (frequency of a word in a particular document) and inverse document frequency (occurrence of the word in all the documents in the corpus). After completing the pre-processing steps, we fit the dataset on the tf-idf model, to create an embedding vector for each book description. These vectors are stored in a feature matrix tf-idf. When the query arrives from the user, it will be embedded in the same way as above, to generate embed query. Then it will be compared one after the other with tf-idf. Then the best three index is found from the distance matrix by compute the average of the cosine distance for each embedded sentence.

- Word embeddings

Word embeddings are dense, real-valued vectors that represent words in a continuous vector space, and are fundamental in natural language processing. Word2Vec is a widely-used technique for learning such embeddings. The Word2Vec algorithm employs a neural network model that can either predict a word given its context (continuous bag of words, CBOW) or predict the context given a word (skip-gram).

In this work, we utilize the gensim library to train the Word2Vec model. To generate word embeddings, the model requires a vocabulary list of the words to be embedded. Training the model involves running it for a set number of epochs, during which the neural network weights are adjusted to accurately predict the context of a word given its surrounding words in the CBOW mode. While CBOW is the default training mode for Word2Vec, it can be switched to the skip-gram model as needed.

4.4. Data Split

This is a crucial step in the deep learning modeling process, where the dataset is divided into two sets: the training set and the test set. The training set is used to learn the relationships between different data features, while the test set is reserved for evaluating the model's performance on unseen data. In our model, we split the dataset into two parts, allocating 20.00% for the test set and the remaining 80.00% for the training set. This division allows us to objectively evaluate the model's performance once final adjustments have been made on the training set.

4.5. Experimentation

In our exploration of author profiling, we embarked on a journey to comprehend the transformative capabilities of transformer neural networks, recognizing their significant contributions to Natural Language Processing (NLP), particularly through Multi-Task Learning (MTL). Their versatility has enabled them to tackle a wide array of projects with remarkable efficiency and effectiveness. In our quest, we delved into comparing various transformer architectures for the task of author profiling. During our experiments, we initialized the word embedding matrix W with 300-dimensional Glove word embeddings. A fully connected layer preceding the softmax activation function was configured with a dimensionality of 100. To mitigate overfitting, dropout regularization with a rate of 0.4 was applied during training. Additionally, the weight and learning rate for center loss were set to 0.001 and 0.1, respectively.

For training the models, we opted for Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01 and a momentum of 0.9. To adaptively adjust the learning rate, we implemented a decay mechanism, reducing it by a factor of 0.9 after every 10 epochs if there was no improvement observed on the development set. The batch size was set to 100, and the models underwent training for a total of 70 epochs.

4.6. Preprocessing

We apply our preprocessing function before training/testing on any dataset. We used the library farasapy. For segmentation, stemming, Part Of Speech tagging (POS tagging) and diacritization. Preprocessing is an essential step in almost any NLP tasks. It aims to eliminate the incomplete, noisy and inconsistent data. We followed these steps:

- Removing URLs: Tweets can contain links, so we need to remove them because they don't contribute to sentiment classification.
- Removing usernames: Usersnames (@user) are also removed from the tweets.
- Remove duplicated letters, as shown in Table 1: Some users, repeats some letters in a word as well as some digit in a number. For instance:" I love this phone soooooo much". The writer tries to intensify its sentiment by repeating the letter "o" in the word "so". The same behavior can be found in Arabic texts. It can lead to wrong results in the stage of processing while trying to get

the linguistic properties of the word. We replaced any letter that appears consecutively more than two times in a word by one letter.

- Remove elongation, as shown in Table 2: In Arabic writing the shape of a letter can be modified with many techniques. One type is commonly used in typing which is the "Elongation" as shown in Table 2.

Table 1. An example of an Arabic word with one duplicated letter.

Method	With Duplicated Letters	Without Duplicated Letters
Sentence Arabic	جمييييل	جميل
Transliteration	ğmyyyyyl	ğmyl
English Translation	beautiful	beautiful

- Remove punctuation and non-Arabic symbols: we also removed punctuation and other symbols that can be found in some tweets.

Also use the preprocess () function to reverse the preprocessing changes, by fixing the spacing around non alphabetical characters, and also segmenting if the model selected need pre-segmentation.

Table 2. An example of an Arabic word with elongation on the letter "يـ".

Method	with Elongation Letters	Without Elongation Letters
Sentence Arabic	جميـــــــل	جميل
Transliteration	ğmyl	ğmyl
English Translation	beautiful	beautiful

4.7. Tokenization

Tokenization serves as a cornerstone in the realm of Natural Language Processing (NLP), acting as the initial step in breaking down textual data into digestible units known as tokens. F. Rangel et al. [22, 23]. These tokens encapsulate the essence of the original text, representing either individual words or meaningful terms, thereby facilitating subsequent analysis and processing.

In our research, we delve into the intricacies of tokenization, adopting a meticulous approach to segmenting text documents, paragraphs, or sentences into coherent units. Embracing the principles of the Apriori algorithm, our tokenization method is tailored to operate at the granular level of individual sentences, enabling fine-grained analysis and comprehension of linguistic structures.

Within this experimental framework, we delineate three pivotal sub-stages:

- Precision in punctuation: The initial phase involves the meticulous removal of punctuation marks, numerical digits, symbols, and extraneous non-Arabic characters. This process ensures the preservation of linguistic integrity while eliminating noise and distractions from the text corpus.
- Strategic stop words removal, as shown in Table 3: Recognizing the ubiquity of stop words, such as prepositions, pronouns, and articles, we embark on a quest to cleanse the text of these redundant entities. By excising words like "على" ,"فوق" ,"من" ,"ال" ,"ذلك", which lack discriminatory power and contribute minimally to document distinction, we streamline the subsequent analysis, focusing solely on the salient content.
- Lemmatization: Embracing the essence of linguistic variation, we embark on the transformative journey of lemmatization. This transformative process harmonizes the myriad inflectional forms of words, distilling them to their core essence represented by a lemma or dictionary form
- . Through lemmatization, we transcend the complexities of linguistic diversity, enabling a unified and comprehensive analysis of textual data.

In essence, our approach to tokenization transcends mere segmentation, evolving into a sophisticated framework that harmonizes linguistic precision with computational efficiency. Through meticulous attention to detail and adherence to methodological rigor, we pave the way for enhanced insights and deeper understanding within the realm of Natural Language Processing.

Table 3. Arabic negation words in our list.

Negation word	Transliteration-Translation	Type
ال	lA- Not/No	preposition
لن	ln- Not	preposition
لم	lm- Not	preposition
لما	lmA- Not	preposition
ما	mA- Not	preposition
غير	Ġyra - But	Noun

4.8. Fine Tuning Models and Result

The meticulous selection of values and hyperparameters was guided by an extensive process of iterative testing, where only configurations demonstrating superior performance were retained for further refinement. Throughout the pretraining phase, the models underwent fine-tuning with a set of carefully curated specifications:

- Gated recurrent unit configurations included 256 units with a dropout rate of 0.3, 128 units with a dropout rate of 0.5, and 64 units with a dropout rate of 0.4.
- Additionally, a 1-dimensional convolutional neural network with 64 units was employed to capture intricate patterns in the data.
- To mitigate overfitting, dropout layers were strategically placed with dropout rates of 0.2 and 0.05 at key junctions in the model architecture.
- A pivotal step involved concatenating Global Average Pooling 1D and Global Maximum Pooling 1D of the previous output, enriching the model's understanding of the data's global features.
- Finally, a dense layer with a sigmoid activation function and a single unit culminated the model architecture, facilitating the classification task.

These meticulously chosen values and hyperparameters reflect the culmination of rigorous experimentation and refinement, ensuring the model's efficacy and robust performance across diverse tasks and datasets.

4.9. Discussion

In the endeavor to identify author gender utilizing the PAN Dataset, our exploration involved the deployment of various language models, augmented by Multi-Task Learning (MTL). Notably, AraBERT emerged as the frontrunner, boasting a remarkable accuracy rate of 79%, a testament to its superior performance detailed in Figure 2. Following closely, ARBERT and MARBERT achieved commendable accuracy rates of 77%.

AraBERT's triumph underscores its unparalleled ability to discern the subtle linguistic nuances crucial for gender identification within Arabic text. Enhanced by Multi-Task Learning (MTL), AraBERT leverages shared knowledge across related tasks, allowing it to excel in this domain. In contrast, while BERT base Arabic exhibits bidirectional context comprehension, it may overlook the finer linguistic intricacies necessary for accurate gender classification. This inherent limitation could result in diminished accuracy or efficacy in gender classification endeavors, despite its language-specific fine-tuning and bidirectional comprehension capabilities. Thus, AraBERT stands as a beacon of excellence, showcasing its unrivaled aptitude in navigating the complexities of Arabic linguistic structures and achieving remarkable success in author gender identification, further accentuated by the benefits of MTL.

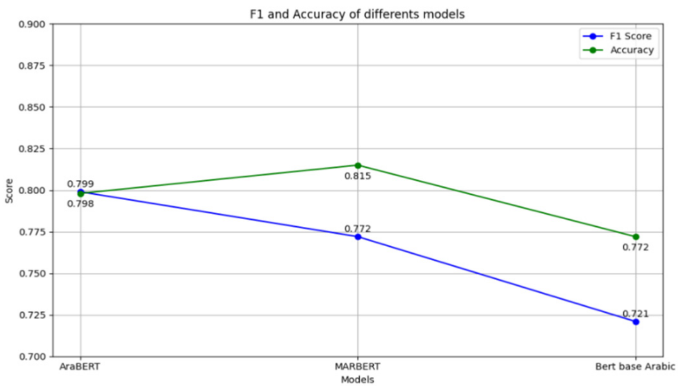

Figure 2. Performance of different Multi-Task Learning (MTL) models.

Training performance using AraBERT is depicted in Figure 2. A strategic approach unfolded, dedicating the initial 3 epochs to priming the GRU layers. During this phase, ARBERT remained immobile, its GRU layers imbued with knowledge at a learning rate of 1×10^{-4}. Subsequently, the narrative shifted as AraBERT unfurled its potential over the subsequent 12 epochs, unleashing its unfrozen prowess with a refined learning rate of 1×10^{-5}. Throughout this evolutionary process, we navigated with the aid of the Adam optimizer, steering through each epoch with a steadfast batch size of 64, and harnessing the binary cross-entropy loss function to sculpt the model's trajectory.

In the crucible of experimentation, AraBERT, with its monumental parameter count of 345 million, emerged as the beacon of excellence in gender detection within Arabic text. Its ascendancy over BERT base Arabic echoes the resonance of its enriched linguistic arsenal, meticulously crafted from a vast corpus spanning over 8.6 billion words. AraBERT's narrative is one of unparalleled efficacy, forged through the crucible of extensive linguistic exploration, ultimately converging toward superior performance outcomes.

In comparison to Bsir et al. [57], who fine-tuned the Ara-BERTv2-large model, ur study demonstrates that AraBERT performs competitively, achieving high accuracy metrics similar to those reported by Bsir et al. for the Ara-BERTv2-large model.

By leveraging Multi-Task Learning, AraBERT demonstrates competitive performance in our study, achieving an F1 score and accuracy of 0.799. This underscores AraBERT's robustness and effectiveness in authorship profiling tasks for Arabic text. MTL allows AraBERT to capitalize on shared knowledge across multiple related tasks, thereby enhancing its overall performance and efficacy in addressing diverse linguistic nuances inherent in authorship identification. These findings highlight AraBERT's suitability and competitive edge in natural language processing applications, particularly within the domain of Arabic language analysis.

5. Conclusions

The novelty of this work consists in demonstrating the effectiveness of integrating Multi-Task Learning with transformer models for Arabic author profiling, specifically focusing on gender identification of social media users. By leveraging MTL, which concurrently trains models on related tasks such as dialect identification, sentiment analysis, and topic classification, we have enhanced the performance and robustness of our approach.

The experimental results revealed that AraBERT model outperforms other transformers models. As future work, we will explore and enhance the performance of deep learning approaches in author's profiling by augmenting the size of the training set, using different tuning parameters, and employing various types of word embeddings. Regarding the data, future efforts should emphasize exploring additional augmentation and resampling strategies. Additionally, there is a need to gather a more extensive collection of harmful tweets for author profiling task. Extracting features such as account types, the number of likes, and the number of shares from tweet links available in the dataset can enhance the distinction between credible and non-credible claims.

Author Contributions

Conceptualization, B.B. and A.A.; Methodology, M.Z., B.B. and A.A.; Software, B.B.; Validation, B.B. and A.A.; Formal analysis, B.B. and A.A.; Investigation, B.B. and A.A.; Resources, B.B.; Data curation, B.B. Writing—original draft preparation, B.B.and A.A.; Writing—review and editing, B.B., A.A. and M.Z.; Visualization, B.B., A.A. and M.Z.; Supervision, A.A. and M.Z.; Project administration, B.B., A.A. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Data Availability Statement

Dataset will be made available for research purposes. Researchers can reach out to the Correspondence via email for the dataset request.

References

1. Bassem, B., & Zrigui, M. (2020). Gender identification: a comparative study of deep learning architectures. In *Intelligent Systems Design and Applications*: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 792-800). Springer International Publishing.

2. Bsir, B., & Zrigui, M. (2018). Bidirectional LSTM for author gender identification. In *Computational Collective Intelligence: 10th International Conference*, ICCCI 2018, Bristol, UK, September 5-7, 2018, Proceedings, Part I 10 (pp. 393-402). Springer International Publishing.
3. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186. 2019.
4. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," CoRR, vol. abs/1907.11692, 2019.
5. Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune BERT for text classification? In *China National Conference on Chinese Computational Linguistics*, pages 194–206. Springer, 2019.
6. BSIR, Bassem, KHOUFI, Nabil, et ZRIGUI, Mounir. Prediction of Author's Profile basing on Fine-Tuning BERT model. *Informatica*, 2024, vol. 48, no 1.
7. Abdul-Mageed, M., & Elmadany, A. (2021). ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers).2021.
8. Bsir, B., & Zrigui, M. (2018). *Enhancing deep learning gender identification with gated recurrent units architecture in social text*. Computación y Sistemas, 22(3), 757-766.2018.
9. Bassem, B., & Zrigui, M. (2020). Gender identification: a comparative study of deep learning architectures. In *Intelligent Systems Design and Applications*: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 792-800). A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, *"Improving language understanding by generative pre-training",* Computer Science, Linguistics 2018.
10. Alokla, A., Gad, W., Nazih, W., Aref, M., & Salem, A. B. (2022). *Pseudocode generation from source code using the bart model*. Mathematics, 10(21), 3967. 2022.
11. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?" In *China National Conference on Chinese Computational Linguistics*. Springer, 2019, pp. 194–206.
12. Abraham, Ajith. *Artificial Neural Networks*. Handbook of Measuring System Design (2005).
13. A. Safaya, M. Abdullatif, D. Yuret, KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media, In *Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics*, Barcelona (online), 2020, pp. 2054–2059.semeval-1.271.2020.
14. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in *Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.2017.
15. Garg, S., Vu, T., & Moschitti, A. (2020, April). Tanda: Transfer and adapt pre-trained transformer models for answer sentence selection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 05, pp. 7780-7788).2020
16. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "*BERT: Pre-training of deep bidirectional transformers for language understanding*," NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, no. Mlm, pp. 4171–4186, 2019.
17. Henry Tsai, Jason Riesa, Melvin Johnson, Naveen Arivazhagan, Xin Li, and Amelia Archer. 2019. Small and Practical BERT Models for Sequence Labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3632–3636, Hong Kong, China. Association for Computational Linguistics.2019.
18. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30. Curran Associates*, Inc., 2017, pp. 5998–6008. 2017
19. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30. Curran Associates*, Inc., 2017, pp. 5998–6008. 2017
20. A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, *"Improving language understanding by generative pre-training",* Computer Science, Linguistics 2018.
21. Manolache, A., Brad, F., Barbalau, A., Ionescu, R. T., & Popescu, M. (2022). Veridark: A large-scale benchmark for authorship verification on the dark web. Advances in Neural Information Processing Systems, 35, 15574-15588.2022.
22. Schlicht, I. B., & Magnossao de Paula, A. F. (2021, September). Unified and Multilingual Author Profiling for Detecting Haters. In *Proceedings of the Working Notes of CLEF 2021*, Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st to 24th, 2021 (pp. 1837-1845). CEUR.2021.
23. Bernard, G. *Resources to compute TF-IDF weightings on press articles and tweets* (2022). 2022
24. Zhang, Z., Wu, Y., Zhao, H., Li, Z., Zhang, S., Zhou, X., & Zhou, X. (2020, April). Semantics-aware BERT for language understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 9628-9635).2020.
25. González-Gallardo, C. E., Montes, A., Sierra, G., Núnez-Juárez, J. A., Salinas-López, A. J., & Ek, J. (2015, September). Tweets Classification using Corpus Dependent Tags, Character and POS N-grams. In CLEF (working notes). 2015.
26. Ray, A., Kumar, S., Reddy, R., Mukherjee, P., & Garg, R. (2019, October). Multi-level attention network using text, audio and video for depression prediction. In *Proceedings of the 9th international on audio/visual emotion challenge and workshop* (pp. 81-88).2019.

27. Bassem, B., & Zrigui, M. (2020). Gender identification: a comparative study of deep learning architectures. In *Intelligent Systems Design and Applications*: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 792-800). Springer International Publishing.
28. Estruch, C. P., Paredes, R. Rosso, P., Learning Multimodal Gender Profile using Neural Networks. In *Rec Adv Nat Language Process 2017*; Varna: Bulgaria, pp: 577-582. 2017.
29. Álvarez-Carmona, M. A., López-Monroy, A. P., Montes-y-Gómez, M., Villasenor-Pineda, L., & Meza, I. (2016). Evaluating topic-based representations for author profiling in social media. In *Advances in Artificial Intelligence-IBERAMIA 2016: 15th Ibero-American Conference on A*I, San José, Costa Rica, November 23-25, 2016, Proceedings 15 (pp. 151-162). 2016.
30. F. Rangel, M. Francisco, P. Rosso, M. Potthast, and B. Stein. "Overview of the 5th author profiling task at PAN 2017: gender and language variety identification in Twitter". *Working Notes Papers of the CLEF 2017 Evaluation Labs*, 2017.
31. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2020, July). Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 8440-8451).2020.
32. Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., Choi, D.: *FibVID: comprehensive fake news diffusion dataset during the COVID-19 period.* Telemat. Inform. 64, 101688 (2021).
33. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.
34. Catelli, R., Pelosi, S., & Esposito, M. (2022). *Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian*. Electronics, 11(3), 374.2022.
35. LIU, Y. *RoBERTa: A robustly optimized BERT pretraining approach*. ArXiv, 2019, vol. 1907, p. 13.2019.
36. Abdul-Mageed, M., & Elmadany, A. (2021). ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers).2021.
37. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). *Deep learning applications and challenges in big data analytics*. Journal of big data, 2, 1-21.2015.
38. Velankar, A., Patil, H., & Joshi, R. (2022). Mono vs multilingual bert for hate speech detection and text classification: A case study in marathi. In *Artificial Neural Networks in Pattern Recognition*: 10th IAPR TC3 Workshop, ANNPR 2022, Dubai, United Arab Emirates, November 24–26, 2022, Proceedings (pp. 121-128). Cham: Springer International Publishing.2002.
39. Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., Choi, D.: *FibVID: comprehensive fake news diffusion dataset during the COVID-19 period.* Telemat. Inform. 64, 101688 (2021).
40. Butt, S., Ashraf, N., Sidorov, G., & Gelbukh, A. F. (2021, September). Sexism Identification using BERT and Data Augmentation-EXIST2021. In *IberLEF@ SEPLN* (pp. 381-389). 2021.
41. Manolache, A., Brad, F., Barbalau, A., Ionescu, R. T., & Popescu, M. (2022). Veridark: A large-scale benchmark for authorship verification on the dark web. Advances in Neural Information Processing Systems, 35, 15574-15588.2022.
42. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.
43. LIU, Y. *RoBERTa: A robustly optimized BERT pretraining approach*. ArXiv, 2019, vol. 1907, p. 13.2019.
44. Sanh, V. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In *Proceedings of Thirty-third Conference on Neural Information Processing Systems (NIPS2019).*2019.
45. Joulin, A., Grave, E., & Mikolov, P. B. T. (2017). Bag of Tricks for Efficient Text Classification. In *EACL 2017*, p, 427.2019.
46. Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., Choi, D.: *FibVID: comprehensive fake news diffusion dataset during the COVID-19 period.* Telemat. Inform. 64, 101688 (2021).
47. Wang, Y., Xu, J., & Sun, Y. (2022, June). End-to-end transformer based model for image captioning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 3, pp. 2585-2594).2022.
48. Alokla, A., Gad, W., Nazih, W., Aref, M., & Salem, A. B. (2022). *Pseudocode generation from source code using the bart model*. Mathematics, 10(21), 3967. 2022.
49. Manolache, A., Brad, F., Barbalau, A., Ionescu, R. T., & Popescu, M. (2022). Veridark: A large-scale benchmark for authorship verification on the dark web. Advances in Neural Information Processing Systems, 35, 15574-15588.2022.
50. Ji, J., Li, Z., Xu, S., Hua, W., Ge, Y., Tan, J., & Zhang, Y. (2024, March). Genrec: Large language model for generative recommendation. In *European Conference on Information Retrieval* (pp. 494-502). Cham: Springer Nature Switzerland.2024.
51. Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., Choi, D.: *FibVID: comprehensive fake news diffusion dataset during the COVID-19 period.* Telemat. Inform. 64, 101688 (2021).
52. Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., Choi, D.: *FibVID: comprehensive fake news diffusion dataset during the COVID-19 period.* Telemat. Inform. 64, 101688 (2021).
53. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.
54. A. Safaya, M. Abdullatif, D. Yuret, KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media, In *Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics*, Barcelona (online), 2020, pp. 2054–2059.semeval-

1.271.2020. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.

55. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.
56. Antoun, Wissam, BALY, Fady, et HAJJ, Hazem. AraBERT: Transformer-based Model for Arabic Language Understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference* 11–16 May 2020. p. 9.2020.
57. BSIR, Bassem, KHOUFI, Nabil, et ZRIGUI, Mounir. Prediction of Author's Profile basing on Fine-Tuning BERT model. *Informatica*, 2024, vol. 48, no 1.