The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 22, Pages 191-198

ICBASET 2023: International Conference on Basic Sciences, Engineering and Technology

Stance Classification for Fake News Detection with Machine Learning

Maysaa ALSAFADI
Karadeniz Technical University

Abstract: The variety of resources and applications available nowadays made the growth of the news rapid; that allowed people to share their opinions, articles, news, etc.; regardless of the truth percentage they have, which caused the belief that lots of the news can be posted or published through social media and news platforms by an automatic pot or fake user for this purpose. Fake news detection (FND) is a binary classification task; that indicates if the news is right or not right, which involves predicting the probability that a certain news article is designed to be deceptive. Commonly, fake news is produced for political and financial purposes, e.g., influencing presidential elections or manipulating the stock market. Although many studies have been conducted to detect news in English as fake news, the evaluation of the credibility of news written in Arabic is still in its early stage. where FND in Arabic languages got underway to receive more interest in the last years, and many detection approaches present some ability to detect fake news on multiple datasets. Then interest in effective detection models has been growing; specifically, in the Arabic language which has lagged behind the work in other languages. In this paper, we used deep learning models and applied a convolutional neural network and long short-term memory (CNN-BiLSTM) with optimization of Stochastic gradient descent (SDG); to the Arabic accessible dataset called AFND; referring to Arabic Fake News Detection. Our experimental results based on the existing AFND dataset indicate an encouraging and good performance; as we reach an accuracy of 87.7%. We appraise the problem of detecting fake news as one of the classification problems; i.e., our target is to classify a given news as credible or not credible; where credibility is often defined in the sense of believability and quality.

Keywords: Fake news detection, Deep learning, Bidirectional long short-term memory, Convolutional neural network

Introduction

Nowadays, the internet has become an integral part of our lifestyle. Anyone can publish anything with creditable content or not that can be consumed by social networking; as the traditional information channels' roles such as newspapers and television on how we collect and consume news have become less prominent than in the past, e.g, in the "Arab Spring"; social media platforms had a vital cause to spread news and rumors, it is used as a communication tool between different sides in various revolutions across the Arab world in 2011. Due to this, many scientists and specialists have been focused on studying the fake news phenomenon; by providing solutions to detect the fake misleading information incompatible with factual reality (Gabrielle et al., 2022).

FND is a partial form of stance detection problem; as stance detection is defined as a problem related to social media analyses, information retrieval, and natural language processing, which focus on detecting the attitude of a person from their published texts', towards a specific target like an idea, concept or event, either explicitly written in the published text or implied. or implied only. Recently, the interest in effective detection techniques to identify such phenomena has been growing very fast, with many models that performed optimally across all datasets in terms of machine learning algorithms; particularly deep learning-based methods. DL is a part of machine learning which used to model and fix complex problems by applying artificial neural networks with multiple layers. It involves the use of algorithms that require large amounts of data to train effectively and can

be computationally expensive to run. with the integration of deep learning and text-based processing, we can fix fake news problems by building classifiers that can classify the news data.

Referring to the published studies and models, most of them have been conducted to identify news in the English language as credible or not credible, as we noticed the research related to the Arabic language is less, due to the lack of a labeled Arabic fake news dataset, which is still a bottleneck for advancing computational solutions to mitigate the false information spread in Arabic content. The task (Khalil et al., 2021) is a contribution to filling this gap; they proposed an Arabic fake news corpus, which consists of 606912 articles collected from 134 Arabic online news sources; then they applied different AL and ML algorithms for the detection task with the proposed dataset. Another number of related research has been published particularly since the 2016 US election. The researchers (Al-Yahya et al., 2021) provided a comparative study of neural networks and transformer-based approaches using four datasets, the authors analyzed whether current advances in deep learning models and large-scale language models for the Arabic language can be effectively applied to the task of Arabic FND. The results of this study demonstrated that transformer-based models outperform the neural network-based solutions, which led to an increase in the F1 score from 0.83 to 0.95, and it boosted the accuracy by 16%. In another research effort by (Ghaith et al., 2019) they utilized four algorithms, namely Random Forest, Decision Tree, AdaBoost, and Logistic Regression to identify fake news from Arabic tweets; and the experimental evaluation shows that the system can filter out fake news with an accuracy of 76%. The authors (Elgendy et al.,2022) evaluated transformer-based classifiers for the FND task while applying eight state-of-the-art Arabic contextualized embedding models, which were XLM-Roberta, GigaBERTv4, Arabert, Arabic-Bert, ArBert, MARBert, Araelectra, and QaribBert. Experimental evaluations lead to an accuracy exceeding 98%. (Antoun et al., 2020) the research introduced the AraBERT methodology as a pre-train BERT for the Arabic language; by following up on reaching the same success that BERT did for the English language. They evaluated AraBERT on three Arabic NLU various downstream tasks. The results showed AraBERT achieved performance on most tested Arabic NLP tasks. In the research shared by (Darwish et al.,2020) they filtered the tweets of the 5,000 most active users with 10 tweets at least; based on user-stated locations. They applied four different classifications on two training sets, namely using fastText, SVM with retweeted accounts (SVMRT), and with all words as features (SVMT EXT), and finetuned BERT embeddings with a dense neural layer and SoftMax output (BERT). As the results show, BERT provided the best results for most topics, with the highest overall averages across all scores. (Shaina et al.,2022) proposed a model called fake news detection through news content and social context (FND-NS), which fits the bidirectional and auto-regressive transformers (BART). By using NELA-GT-2019 and Fakeddit as real-world datasets, and the result was 74.89%, 72.40%, 77.68%, 70.4%, and 74.95% for accuracy, precision, recall, AUC, and F1-score respectively.

On the other hand, some projects are working on prediction of the fake news by providing a model to determine the credibility of news articles in the early stages after publishing news on social media, as the DSS model proposed by (Davoudi et al.,2022); which analyze the sentiments of replies related to the news articles by constructing the stance network and extracting various graph-based features. The proposed model is evaluated on the FakeNewsNet repository, comprising two recent well-known datasets in the field, namely PolitiFact and GossipCop. The results outperformed the state-of-the-art methods by 8.2% on PolitiFact and 3% on the GossipCop datasets. Another fake news detection system proposed by (Huang et al.,2020) which based on an ensemble learning model using deep learning techniques. They highlight the difficulty in identifying fake news due to their similarities with real news. They propose a system that preprocesses and analyzes news articles using different training models. The ensemble learning model combines four different models: embedding LSTM, depth LSTM, LIWC CNN, and N-gram CNN. To achieve higher accuracy in fake news detection, the authors optimize the weights of the ensemble learning model using the Self-Adaptive Harmony Search (SAHS) algorithm. Experimental results show that the proposed model outperforms state-of-the-art methods, achieving the highest accuracy of 99.4%. They also investigate the issue of cross-domain intractability and achieve the highest accuracy of 72.3% in that scenario.

CNN-BiLSTM is one of the most popular and best deep learning models, and several studies deployed it in experiments by adding different changes in the architecture. The study of (L. Baniata et al., 2016) presents a deep learning model for sentiment analysis of Arabic text. The research highlights that while sentiment analysis research has predominantly focused on English text, limited research has been conducted on other languages like Arabic. The proposed model combines Convolutional Neural Network (CNN) and Bi-directional Long Short-Term Memory (BiLSTM) to analyze the sentiment of Arabic text. A comparison between two architectures, CNN-BiLSTM and BiLSTM-CNN, demonstrates that CNN-BiLSTM achieves a test accuracy of 86.43% and provides a better representation of sentence features. The study concludes that the CNN-BiLSTM model is highly effective for sentiment classification in Arabic text. (Ouassil et al., 2022) proposed a fake news detection system that combines word embedding techniques and a hybrid deep learning model. The authors

highlight the prevalence of fake news in online sources and the need to detect and prevent its spread. They present a deep learning approach that combines different word embedding techniques with a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) model. The authors trained their classification model on an unbiased dataset called WELFake. They found that the best results were achieved by combining a pre-trained Word2Vec CBOW model and a Word2Vec Skip-Word model with a CNN on BiLSTM layers. This combination achieved an accuracy of up to 97%. In conclusion, the paper introduces a novel method for detecting fake news using the WELFake dataset. The approach involves representing words as numerical vectors through various word embedding techniques and training a hybrid CNN and BiLSTM model. The results demonstrate improved accuracy and precision compared to traditional machine learning algorithms and related work. Another deep learning approach was introduced by (Alharbi, 2021) for Arabic sentiment analysis by combining Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architectures with a Support Vector Machine (SVM) classifier.

The conventional deep learning architecture is modified by replacing the fully connected layer with an SVM classifier that utilizes embedded vectors extracted by CNN and BiLSTM for the polarity classification of Arabic reviews. The proposed method is evaluated on three publicly available datasets, and the results demonstrate superior performance compared to baseline algorithms of CNN and SVM on all datasets. The model outperforms one of the state-of-the-art deep learning models as well. The author concluded that their model, which combines CNN, BiLSTM, and SVM, is effective for Arabic sentiment classification. They suggest further improvements such as addressing issues like negation handling and exploring deeper layers and diverse architectures on different Arabic benchmark datasets. The task of (Abdelhady et al., 2022) presented an effective stacked ensemble deep learning framework, Stacked-CNN-BiLSTM-COVID, for sentiment analysis of Arabic COVID-19 tweets. The model combines Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) to categorize Arabic tweets. Word embedding models, namely Aravec, FastText, and ArWordVec, are used to represent the tweets as vectors and assess their impact on the model's performance. The proposed model is compared to other deep learning models such as CNN, LSTM, and BiLSTM, and experiments are conducted on three Arabic datasets related to COVID-19 and vaccines. The empirical findings demonstrate that the Stacked-CNN-BiLSTM-COVID model outperforms other models, achieving high F-measure scores of 76.76%, 87.25%, and 80.5% on the SenWave, AraCOVID19-SSD, and ArCovidVac datasets, respectively. The study highlights the importance of using stacked ensemble learning and word embedding approaches for accurate sentiment analysis of Arabic COVID-19 tweets.

In this paper, we describe the steps involved in the detection of the fake news process, including data collection, preprocessing, feature extraction, model selection, model training, and evaluation. The paper also highlights the effectiveness of stance classification-based fake news detection and its potential to be integrated into social media platforms to prevent the spread of fake news. Our contribution includes applying the deep learning techniques with the optimizer SGD which helps us to achieve an accuracy of 87.7% after many experiments. We used a large Arabic fake news corpus in our experiments; which is needed to train the ML algorithms particularly deep learning models.

Method

This section shows the datasets used in our experiments with the experimental settings, the preprocessing techniques then baseline method. Deep learning has high efficiency for stance classification with big data; it can be used to develop automated systems for fake news detection. Deep learning models can analyze large amounts of data and learn patterns that are indicative of fake news. By training those models on a dataset of labeled news articles and social media posts, it is possible to develop an effective fake news detection system.

Dataset

In our experiments, we used a large Arabic dataset called AFND; which was proposed by (A. Khalil et al., 2022) AFND was collected from newspaper and feed parser Arabic news articles. It contains 606912 public news articles that were scraped from 134 public news websites of 19 different Arab countries that include: Levant countries (Jordan, Palestine, Lebanon, and Syria), Arab Gulf countries (United Arab Emirates, Saudi Arabia, Kuwait, Qatar, Bahrain, and Oman), Arab Maghreb countries (Algeria, Morocco, Tunisia, Libya, and Mauritania), Yemen, Iraq, Egypt, and Sudan; AFND is publicly accessible. (Table 1) presents some statistical information about the corpus articles and the online sources.

Table 1. AFND statistical information

Statistics	Credible	Not Credible	Undecided
Source count	52	51	31
Articles count	207310	167233	232369
Average number of words in body text	230	217	254
Average number of words in headline	9	10	9

AFND dataset is available in JSON format; 134 news sources are stored in separate single JSON files with details including the title, text, and published date for each news article. Each news article source has a key and value. The key is the source name, and the value is a dictionary that contains the source label and two lists. The first list has the RSS website links and the second one has the website links of the local news pages. We collected those JSON files in one CSV and PKL file for the modeling process.

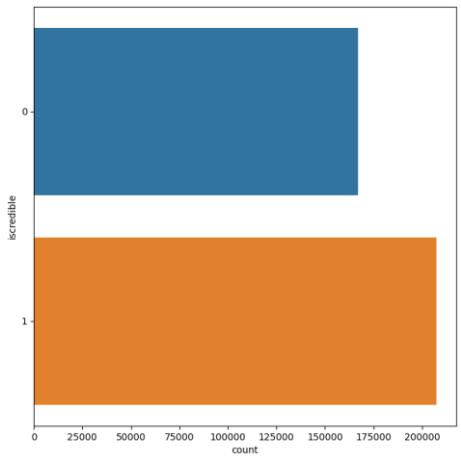

Figure 1. The number of samples per label

In this experiment, we use binary classifications; the number of articles for the training and testing are 83616 and 250850, respectively. The articles that are labeled in the dataset as "undecided" were ignored. Classification tasks used distant labeling, where each article is labeled with the same label as the corresponding source credible or not credible (Khalil et al., 2021). Figure 1. presents the number of samples per label; it indicates the imbalanced dataset that caused overfitting in the first experiment we did, as each class of data has not an equal number of observations in the dataset. We balanced the training dataset using the python Undersampling library scikit-learn. Finally, we use the balanced dataset to train a machine learning model. The reason for balancing a dataset is to prevent a machine learning model from being biased towards one class due to an unequal representation of data in the training set

Pre-preprocessing

Data preprocessing is processed to clear Arabic and non-Arabic digits, words, white spaces, websites, punctuation, and symbols. The natural language toolkit (NLTK) and the Arabic-Stopwords Python libraries were used to remove stop words. The normalization technique was applied using Tashaphyne (Taha, 2023) to extract the root for each Arabic word. Finally, a vocabulary dictionary of 196400 tokens was generated to create word sequences by assigning a unique digit for each word. Figure 2. visualize the distribution of the number of words in a text.

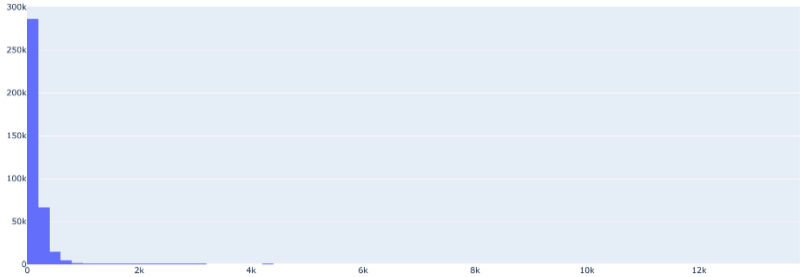

Figure 2. The distribution of number of words in a text

The preprocessing module cleans the text and minimizes the vocabulary while preserving meaning; it comprises the below five steps:

1. Word encoding using word embeddings.
2. Tokenization by splitting text into individual words.
3. Punctuation marks and Arabic stop words removal
4. Stemming and limitizer is tested and the result was the best with the limitizer function which convert the words to thier base form using a dictionary lookup.
5. Removing special characters and numbers and non- arabic characters.

Classification Model

The purpose of classification is to predict previously unseen items based on inferences derived by training on a set comprising news articles and their labels. We first describe the basic elements of our used architecture and motivate why we chose it. The first stage of the model, we use an Embedding layer to to generate inputs that are a combination of word sequences and embedded features. In general, an embedding layer is a type of layer in a neural network that is commonly used in natural language processing (NLP). Its purpose is to map input data (such as text) into a lower-dimensional space, known as an embedding space. This mapping allows the network to better understand the relationships between the inputs and make more accurate predictions. We use one-dimensional Convolutional layer (CNN) with specifying the number of filters as the output dimension, the kernel size as 3, and the activation function as 'relu' (rectified linear unit). We equalize article lengths in the preprocessing stage by zero-padding or truncation, such that the article length is constant, leading to constant channel size for all articles.

Following the Convolutional layer, we employ a type of recurrent neural network (RNN); the bidirectional Long Short-Term Memory (BiLSTM) layer because the news we predict represents word sequences, and this type of layer is specialized in learning long-term dependencies in text (Ciprian-Octavian, 2023). A further reason to employ such a layer is that it uses two simple LSTM architectures that input sequence in both directions simultaneously. This allows the model to capture both the past and future context of each word in the input sequence. We use dropout layers to deactivate a percentage of the outputs coming from the previous layer; thus, we reduce overfitting by creating artificial noise and improve the generalization capacity of the network when new, unknown data is fed for prediction. We use dense layers with linear activation as a connection between the network layers, and a final dense layer with a softmax activation function to produce the classification result, i.e., the probability of the post to be true.

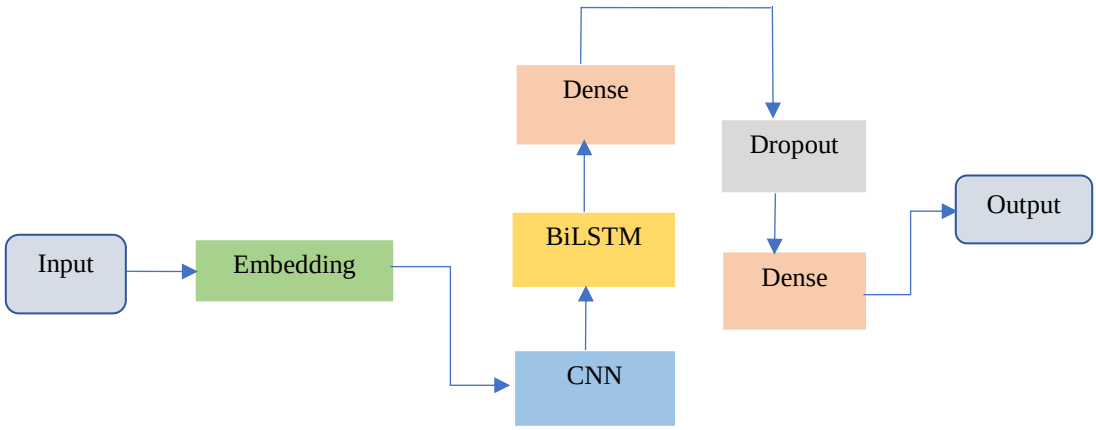

Figure 4. The model architecture

Finally, we complied the model with the optimizer Stochastic Gradient Descent (SGD); which is a type of gradient descent algorithm that optimizes a differentiable objective function by iteratively adjusting the model parameters in the direction of the steepest descent of the objective function. Its stochastic nature can introduce some noise in the training process, which can help prevent overfitting and improve generalization performance. We also update the learning rate value to be calculated based on the loss value as below equation; where β equal 0.001 and loss is the loss value of the test set.

$$\alpha = \frac{\beta}{1 + e^{-loss}}$$

To increase our model; we set up the model training with early stopping, a TensorFlow hook; which is defined as a regularization technique that stops the training process in certain conditions, for example, the validation loss reaches a particular threshold; it also optimizes the number of epochs.

Results and Discussion

This section highlights the results, lessons learned from the implementation and presents an in-depth analysis of the two modules. Figure 4 presents the accuracy of the automatic Arabic fact-check detection for the binary classification task. Compared with the reinforcement learning and traditional ML models and the other models used in (Khalil et al., 2021) The accuracy of our model is higher than the ones represented by (Khalil et al., 2021). The reason is that we improved the optimizer using SGD; which provides 87.8% accuracy.

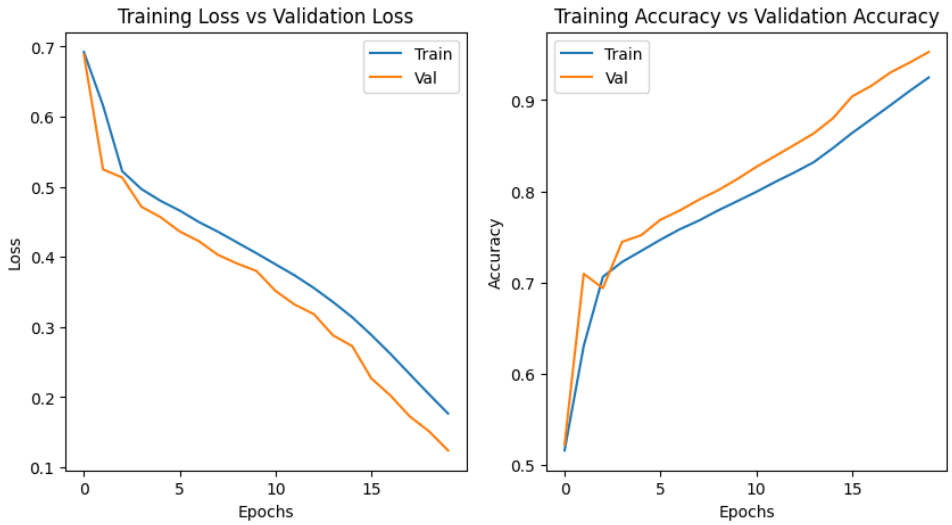

Figure 4. Accuracy of the model for each subdivision of the training and validation sets.

As presented in Table 2; we have many experiments before reaching the enhanced model, first, we work with "Adam" optimizer which provides an accuracy of 74.8% for the test set and 98% for the training set. Then we try to use SGD optimizer with two cases: 1) Using fixed learning rate with the value 0.08, but the result was worse as it reach 49.65%. 2) By changing the learning rate to be calculated based on the loss value, the accuracy increased to 87.7%.

Table 2. Accuracy of the classification task

Classifier	Accuracy
BiLSTM-CNN with SGD, dynamic LR	87.7%
BiLSTM -CNN with adam optimizer	74.8%
BiLSTM -CNN with SGD, fixed LR	49.65%

Conclusion

With the increased adaptation of the news resources such as social media and other web platforms, distinguishing verified information from fake news becomes an extremely difficult and crucial task. Especially with the Arabic language news; opposite to the other languages there are several types of writing posts and news in Arabic. In this paper, we applied a fake news detection model using convolutional and bidirectional LSTM layers. The used model is capable of accurately identifying fake news based solely on a news text. This model offers state-of-the-art performance on the tested datasets. This presents an efficient model to be used for Arabic news detection.

Recommendations

As future research directions, we will apply some improvements on the model to reach the best accuracy with the used dataset, optimizing the Arabic word processing, and finally we will use the transfer learning test to be more comprehensive.

Scientific Ethics Declaration

The author declares that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the author.

Acknowledgements or Notes

This article was presented as oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Marmaris/Turkey on April 27-30, 2023.

References

Abdelhady, N., Soliman, T. H. A., & Farghally, M. F. (2022). Stacked-CNN-BiLSTM-COVID: An effective stacked ensemble deep learning framework for sentiment analysis of arabic COVID-19 tweets. https://doi.org/10.21203/rs.3.rs-2206648/v1

Alharbi, O. (2021). A deep learning approach combining CNN and Bi-LSTM with SVM classifier for Arabic sentiment analysis. *International Journal of Advanced Computer Science and Applications, 12*(6),165-172.

Al-Yahya, M., Al-Khalifa, H., Al-Baity, H., AlSaeed, D., & Essam, A. (2021). Arabic fake news detection: comparative study of neural networks and transformer-based approaches. *Complexity, 2021*, 1-10.

Baniata, L. H., & Park, S. B. (2016). Sentence representation network for Arabic sentiment analysis. *한국정보과학회 학술발표논문집*, 470-472.

Ashwaq K., Moath J., Monther A., & Manar J. (2022). AFND: Arabic fake news dataset for the detection and classification of articles credibility. *Data in Brief,42*, 108141

Davoudi, M., Moosavi, M. R., & Sadreddini, M. H. (2022). DSS: A hybrid deep model for fake news detection using propagation tree and stance network. *Expert Systems with Applications, 198*, 116635.

Gauthier -Melancon, G., Ayala, O. M., Brin, L., Tyler, C., Branchaud- Charron, F., Marinier, J., Grande, K., & Le, D. (2022). Azimuth: Systematic error analysis for text classification. *Association for Computational Linguistics*, 298-310.

Jardaneh, G., Abdelhaq, H., Buzz, M., & Johnson, D. (2019). Classifying Arabic tweets based on credibility using content and user features. *IEEE Jordan International Joint Conference on Electricial Engineering and Information Technology (JEEIT).* 596-601.

Khalil, A., Jarrah, M., Aldwairi, M., & Jararweh, Y. (2021, November). Detecting Arabic fake news using machine learning. In *2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)* (pp. 171-177). IEEE.

Nassif,A.B, Elnagar,A., Elgendy,O., & Afadar, Y. (2022). Arabic fake news detection based on deep contextualized embedding models. *Neural Computing and Applications,34*,16019-16032.

Ouassil, M. A., Cherradi, B., Hamida, S., Errami, M., El Gannour, O., & Raihani, A. (2022). A fake news detection system based on combination of word embedded techniques and hybrid deep learning model. *International Journal of Advanced Computer Science and Applications, 13*(10), 525-534

Raza, S., & Ding.C.. (2022). Fake news detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics, 13*, 335-362.

Samih, Y., & Darwish, K. (2021, April). A few topical tweets are enough for effective user stance detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 2637-2646).

Truică, C. O., Apostol, E. S., Nicolescu, R. C., & Karras, P. (2023). MCWDST: a minimum-cost weighted directed spanning tree algorithm for real-time fake news mitigation in social media. *arXiv preprint arXiv:*2302.12190.

Zerrouki, T. (2023, April 1). *Tashaphyne, Arabic light stemmer*. https://pypi.org/project/Tashaphyne/

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.

Author Information

Maysaa Alsafadi
Karadeniz Technical University
Trabzon, Turkey
Contact e-mail: *maysaaelsafadi@gmail.com*

To cite this article:

Alsafadi, M. (2023). Stance classification for fake news detection with machine learning. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 22*, 191-198.