IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Arabic Fake News Detection Model Using Ensemble Machine Learning Classifiers

[1] Khadega Anaam, [2] Abdullah Alhashdi

Department of Computing and Information Technology
University of Science & Technology, Sana'a, Yemen

Abstract: Information sharing on social media has grown faster, less costly and easily accessible. This useful information may contain fake news in many fields including politics, medicine and sports for various reasons such as advertising and propaganda. Ability to identify, analyze and address such information is significantly important. Many studies have been conducted to detect fake news in English, but there is a lack in Arabic language. This paper exhibits model to detect Arabic fake news on X-Platform during Ukraine-Russia conflict with the assistance of Ensemble Machine learning. In this work, an Arabic fake news dataset was collected related to the Ukraine-Russia conflict. Six different classification algorithms are trained to classify news as fake or real and are compared considering accuracy, recall, precision. The experimental results demonstrated that Random Forests (RF) using Term Frequency-Inverse Document Frequency (TF-IDF) preforming with SMOTE techniques achieved the best predictions with an accuracy of 99 % and testing accuracy with 98 %. Indeed, the results show that applying ensemble algorithms with TF-IDF and SMOTE achieved better improvement on evaluation metrics compared to the baseline classifier and other classifiers without SMOTE.

IndexTerms - Fake news, social media, Arabic corpus, X platform.

I. Introduction

In recent years, information has spread rapidly across social media platforms like X (formerly Twitter), Instagram and Facebook. It became easier and affordable for anyone to acquire the latest news from social media available 24/7 at his/her fingertips. Millions of news are generated daily on social media and shared widely without evaluating its truthfulness. This has contributed to spread the fake news concept. In fact, the term "Fake news" has been named the word of the year by the Macquarie Dictionary in 2016 (Wong et al., 2016). There has been no agreed definition of the concept of Fake news. However, one of the most adopted definition of Fake news is presented by (Allcott & Gentzkow ,2017) which is defined as "news articles that are deliberately and verifiably false and designed to mislead readers". Furthermore, Fake news has always been used as a weapon to achieve economic and political gains and to shape people's thoughts and decisions to influence their behavior (Zhou et al.,2018). Hence, more studies with various topics should be conducted on detecting fake news on social media by using the best techniques.

The effect of social media on conflict perception has drawn more attention and concern, especially when considering divisive geopolitical situation like the conflict in Gaza and the Russia-Ukraine conflict (Ghosh, C.,2024). The conflict between Ukraine-Russia was one of the most widely discussed subjects on social media in 2022 (Fung & Ji, 2022). This conflict stems from a variety of intricate issues that have surfaced after the fall of the Soviet Union, notably the geopolitical division between the eastern and western regions of Ukraine (Karácsonyi et al., 2014). The conflict intensified significantly on February 24, 2022, when Russian forces entered Ukraniain territory as part of a special military operation (Fung & Ji, 2022).

Unfortunately, at the time writing, the conflict is still ongoing. People across the world have been actively sharing their opinions and following the news on social media platforms. The nature of these social media platforms led to the rise of fake news which harms individuals and societies. This paper present Ukraine-Russia Conflict (URC) dataset and suggests a robust model based on ensemble learning methods for detecting Arabic fake news related to the Ukraine-Russia conflict on X platform. This model is built on a three an ensemble learning methods (XGBoost, Random Forests (RF), and Adaboost), alongside three traditional machine learning methods (Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT)), used as baselines to evaluate and compare performance outcomes.

II. METHODOLOGY

This paper proceeds by delineating a series of steps aimed at establishing the framework for detecting Arabic fake news. The experiment environment used of this study is Google Collab. The Google Collab environment serves as an outstanding tool for gaining proficiency in machine learning and data science. Python programming language is well-suited for implementing machine learning models due to its extensive of libraries.

X-Platform has been considered one of the most widespread social media platforms for spreading fake news worldwide. Thus, this study proposed a methodology framework in order to detect Arabic fake news on X-Platform related to Ukraine-Russia Conflict. In our proposed methodology, several phases such as data collection, data preprocessing, data annotation, feature selection and model building were involved. For getting better results, the state-of-the-art ensemble machine learning models were explored. Besides, we used Microsoft Excel to gather the dataset and Google Collab cloud platform to preprocess and analyze the dataset. The details of our proposed methodology framework described in **Fig.1**

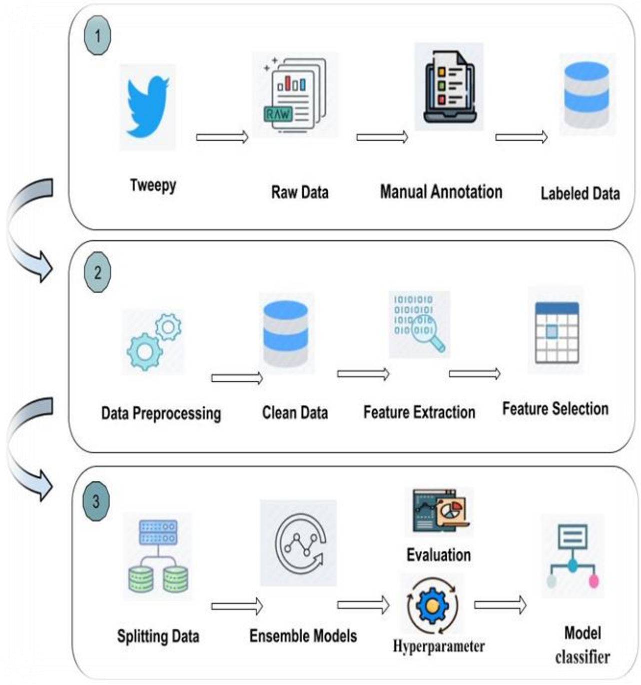

Figure 1. The Framework of the Proposed Arabic Fake News Detection Model

A. Data Collection

This study gathered more than 50k Arabic tweets about Ukraine-Russia conflict from the X-Platform. The Arabic tweets were extracted by using the Tweepy library and through the X-Platform developer account from the beginning of the conflict on February 24, 2022. The search process of tweets relies on the trend's hashtags that appears during the conflict such as Russian_Ukrainian_War, Nato and Kiev, as shown in the **Table 1.**

Table 1. List of hashtags used to construct the dataset

#	Hashtag	English Translation
1	#الحرب_الروسية_الاوكرانية	#Russian_Ukrainian_War
2	#أوكرانيا	#Ukraine
3	#روسيا	#Russia
4	#الناتو	#Nato
5	# الحرب_العالمية_الثالثة	#World_War_III

B. Data Preprocessing

To prepare the gathered tweets for next step, we performed some text preprocessing techniques. We eliminated the noise data from all tweets such as mentions, hashtags, images, videos, URLs, numbers, punctuation, stop words, emojis, punctuations, non- Arabic words, Arabic diacritics. To convert the large text of tweets to small separated words, we have implemented normalization and tokenization techniques.

A number of rules were used to normalize the text, such as replacing each Alif with different forms (أ, إ, آ) with a bare Alif character "ا." In addition, characters that were repeated more than twice and Tatweel "_" were eliminated, as were English symbols, extra spaces, and Tashkeel (fattha and wasla).

Stop words are a group of regularly used, less-meaningful terms that appear frequently in natural language. A wide list of Arabic stop words may be found in the Arabic-reshaper library of the Natural Languages Toolkit. Stop words in Arabic are incorporated and saved in a file. Stop words would occupy important processing time and space in our dataset, therefore they were eliminated.

Pre-processing was done using Python NLTK library which is an open-source. **Table 2 and Table 3** shows the tweets before and after preprocessing.

Table 2. Tweets before Preprocessing

Tweet	Label
الكويت ترحب بتوقيع #روسيا و #أوكرانيا اتفاقاً باستئناف تصدير # الحبوب # خطوة ستسهم في تعزيز #الأمن_الغذائي والحد من ارتفاع أسعار - الحبوبhttps://t.co/4sBoShEja5	Real
??وزارة الدفاع الروسية: العملية العسكرية توقف عمل 5 مختبرات بيولوجية في العاصمة الاوكرانية #كييف التعليق:مؤشر خطير جداً وجود 5 معامل بيولوجية في مدينة واحدة !! #روسيا بهذا الإنجاز خدمت البشرية وتستحق جائزة عالمية للمحافظة على الكرة الأرضية وسكانهاhttps://t.co/rvgvt6AVdG	Fake
مع استمرار #الحرب_الروسية_الأوكرانية، حثت (7) دول أوروبية مواطنيها على الامتناع عن الانضمام إلى المقاومة العسكرية الأوكرانية ضدّ #روسيا. https://t.co/RF1NbH3IOj	Real

Table 3. Tweets after Preprocessing

Tweet	Label
الكويت ترحب بتوقيع روسيا أوكرانيا اتفاقا باستئناف تصدير الحبوب خطوة ستسهم تعزيز الأمن الغذائي والحد ارتفاع أسعار الحبوب	Real
وزارة الدفاع الروسية العملية العسكرية توقف عمل مختبرات بيولوجية العاصمة الاوكرانية كييف التعليق مؤشر خطير جدا وجود معامل بيولوجية مدينة واحدة روسيا بهذا الإنجاز خدمت البشرية وتستحق جائزة عالمية للمحافظة الكرة الأرضية وسكانها	Fake
استمرار الحرب الروسية الأوكرانية حثت دول أوروبية مواطنيها الامتناع الانضمام المقاومة العسكرية الأوكرانية ضد روسيا	Real

C. Data Annotation

I. Manual Annotation

After gathering the relevant tweets and filtering of repeated tweets, the study applied manual annotation by collected a set of topics associated with the most circulated fake news during the conflict and verify it by two popular Arabic fact-checking platforms: Fatabyyano and Misbar, as shown in bellow **Table 4.** The collected topics were used for the manual annotation to classify each tweet as “Real” or “Fake”. In the annotation process, about 600 were labelled as fake tweets and 9,600 as real tweets.

Table 4. List of verified rumors and misinformation topics

#	Keyword	English Translation
1	نهاية العالم	The End of the World
2	حرب عالمية	World War
3	حرب نووية	Nuclear War
5	علامات الساعة	Signs of the hour
7	انهيار الاقتصاد الروسي	Collapse of the Russian Economy
8	دروع بشرية	Human Shield
10	انهيار العملة الروسية	Collapse of the Russian Currency
12	جيش بوتن يخسر	Putin’s army loses
15	القوات الروسية	Russian Forces
17	زعيم الشيشان	Leader of Chechya
19	يغتصبون النساء	Violate the women
21	الجنود الروس	Russian Soldiers
24	أزمة مكدونالدز	McDonald Crisis
25	عقوبات اقتصادية	Economic Sanctions
28	أوكراني قصف	Ukraine Bombing
29	وفاة رئيس أوكرانيا	Death of President of Ukraine
30	القوميون الأوكرانيون	Ukrainian Nationalists
31	قتال حزب الله	Hezbollah Fighting

33	قصف أمريكي	American Bombing
34	الممثل الأمريكي ستيفن سيغال	Steven Segal American Actor
36	ازمة نفطية حادة	Acute Oil Crisis
37	مواد كيميائية سامة	Toxic Chemicals
38	إرسال بوتين إلى الفضاء	Sending Putin into Space
39	شبح كييف	Ghost of Kiev

2. SMOTE Techniques

Once the data were labeled manually, the data showed highly imbalance where the clear majority label is Real, while the minority of label is Fake. According to the **Fig.2** below, around 6.1 % of the tweets was found to be fake news, while 93.9 % was real data, which indicates that the dataset is not balanced. We integrated SMOTE (synthetic minority oversampling technique for nominal and continuous) on the training data to balance an imbalanced dataset.

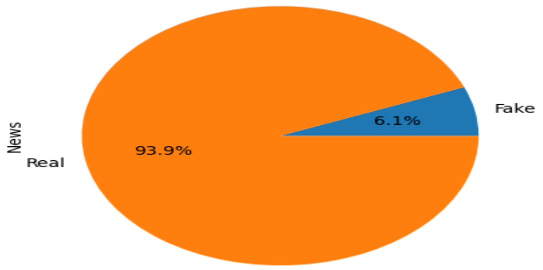

Figure 2. Dataset Before SMOTE

To handle highly imbalanced dataset, we used resampling (SMOTE) techniques. It consists of adding more examples from the minority label. The SMOTE (Synthetic Minority Oversampling Technique) consists of synthesizing elements for the minority label, based on those that already exist. **Fig.3** below demonstrates dataset after executed resampling techniques with 50% Fake tweets and 50% Real tweets.

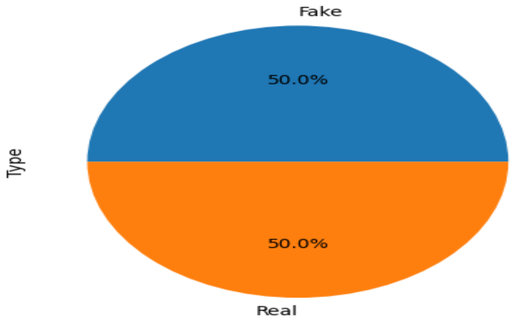

Figure 3. Dataset After SMOTE

3. Feature Extraction

The text tweet has to be converted into vectors expressions in order to perform mathematical operations and before giving it into the classifier. In this study, two types of vectorization methods are used to represent words in tweet column as numerical vectors which include a count-vectors and TF-IDF. Moreover, we replace categorical values to numerical values in label column with label =1 indicating that the tweet contains fake news and label =0 indicating that the tweet does not contain fake news.

4. Feature Selection

Feature selection is useful for selecting only those features that have an impact on the determination of whether a piece of news is fake or not. This study applied both on stylometric and textual features to reduce the dimensions of the dataset. To improve model accuracy, we have removed unwanted columns as we will not need them any more in classification analyses and kept only with two class “label” and “tweet”.

III. EXPERIMENTAL SETUP

After extracting and selecting features, we divided the whole dataset into a training set of 80% and a test set of 20% for each experiment. We trained the model by resampling it using the training set, setting hyperparameters, and evaluating its performance. In this paper, two classification experiments were investigated and evaluated on our dataset that we have gathered, applying with and without SMOTE techniques. We have implemented and used six traditional and ensemble algorithms such as Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), XGBoost, Random Forests (RF), and Adaboost for the classification purpose. The collected dataset has not many dimensional therefore, DL algorithms eliminated in this study.

The traditional algorithms have used as a baseline for comparing the performances with the ensemble algorithms. TF-IDF technique is used to generate vectors and all the listed algorithms applied to investigate the best model for identifying fake news.

Improving the performance score based on data patterns and observed evidence is one of the main goals and problems of the machine learning process. Nearly all machine learning algorithms are designed to accomplish this goal by estimating a set of parameters from the dataset that will optimize the performance score.

We made advantage of Scikit-learn's GridSearchCV function to enable an automated and repeatable method for hyperparameter tuning. Each classifier uses the hyper-parameters listed below:

- Naive Bayes: alpha=0.5
- Logistic Regression: with default values
- Random Forests: with default values
- XGBoost: with default values

IV. EXPERIMENTAL RESULTS

The accuracy show that traditional and ensemble models have without applying SMOTE techniques similar results and not achieved 95 % of accuracy score. Xgboost (XGB) classifier achieves the best training accuracy with 95.30 %. Decision Tree (DT) classifier and Random Forests (RF) classifier achieved the same accuracy score with 95.01 %. In addition, Adaboost (Ada) classifier outperforms 94.82%, while Logistic Regression (LR) classifier achieved 94.14%. Naive Bayes (NB) classifier takes the smallest accuracy score with 94.09 % as depicts in **Table 5.**

Table 5. Results of the Six algorithms without SMOTE

CLASSIFIERS	Results
0 NB	94.09
1 DT	95.01
2 LR	94.14
3 XGB	**95.30**
4 RF	95.01
5 ADA	94.82

From the **Table 6.**, it can be observed that the performance of the six models after applying SMOTE has been improved compared with the performance of models without applying SMOTE **Fig.4.** The RF classifier achieves the highest training accuracy with 99 % and testing accuracy with 98 %, Adaboost achieve the lowest train accuracy with 88 % and testing accuracy with 88%. The training accuracies of all the rest models are more than 97%, while the testing accuracies outperforms more than 94%. The execution time of the models is the period it takes to be executed. Naive Bayes (NB) takes the smallest time execution of 0.02 second while Decision Tree (DT) takes the longest time with 4.22 minutes.

The six algorithms are performed using SMOTE on balanced dataset. This experiment shows better results as compared to the first one. The performance of the classifiers with balanced dataset is more robust and performed high accuracies without any overfitting.

Table 6. Accuracies of the six algorithms with SMOTE

Model name	Train accuracy	Test accuracy	Exec time
Random Forest	**0.999742**	**0.989943**	**30.74**
XGBoost Classifier	0.978917	0.963125	**21.97**
AdaBoost	0.885429	0.886282	**24.85**
Multinomial NB	0.973436	0.944817	**0.02**
Logistic Regression	0.988459	0.977824	**0.94**
Decision Tree	0.999742	0.952037	**4.22**

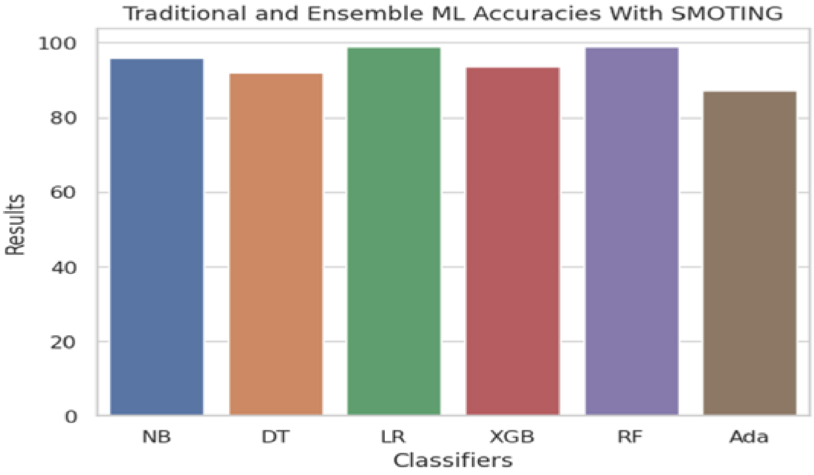

Figure 4. Accuracy Results of the Six Algorithms with SMOTE

V. DISCUSSION

The main goal of this study was to create a large dataset for fake news in Arabic related to the conflict between Ukraine and Russia. To this purpose, we present a new Arabic-language fake news corpus that was gathered from X platform. The experimental results clearly show that the manually annotated dataset can be used as a baseline for future research in the field of fake news and misinformation. As there is currently no benchmark dataset for fake news detection in Arabic related to the conflict between Ukraine and Russia, this corpus will be useful to the research community once it is made publicly available.

We trained various machine learning models on the proposed corpus using three ensemble machine learning classifiers and three traditional machine learning classifiers. The best model was chosen to predict fake news classes of balance data (roughly 10,000 tweets) using SMOTE Techniques, which have the advantage of reducing overfitting and imbalanced class issues during the training process. The proposed corpus was manually annotated by two annotators to ensure the quality and usefulness of the developed corpus. The experimental results of this study proved that the proposed ensemble classifiers have outperformed the methods used in the related work, with achieving high accuracy.

VI. CONCLUSION

Over the years, X-Platform has become one of the most popular social media platforms in the world. Currently, many users share significant amounts of information and news, in the form of tweets, with friends and family members without knowing whether they are real or fake tweets. In fact, fake news spread is a global problem as mentioned in the introduction of this study. Automatic Fake detection in Arabic tweets is more difficult than other languages due to changes in the structural and morphological nature of the Arabic language. To address this issue, this study introduced a best model on Arabic dataset that constructed from X platform and related to Ukraine-Russia conflict news.

The study used SMOTE techniques to improve the results of the model by balancing the collected dataset. The results showed lower accuracy values in the imbalance dataset. While the balance dataset showed improved results by using SMOTE Techniques. The best model was achieved by Random Forests (RF) with accuracy up to 99 % and testing accuracy with 98 %.

VII. REFERENCES

[1] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-36.

[2] Ballantyne, N., Wong, Y. C., & Morgan, G. (2017). Human services and the fourth industrial revolution: From husITa 1987 to husITa 2016. Journal of Technology in Human Services, 35(1), 1-7.

[3] Karácsonyi, D., Kocsis, K., Kovály, K., Molnár, J., & Póti, L. (2014). East-West dichotomy and political conflict in Ukraine-Was Huntington right?. Hungarian Geographical Bulletin, 63(2), 99-134.

[4] Fung, Y. R., & Ji, H. (2022). A weibo dataset for the 2022 russo-ukrainian crisis. arXiv preprint arXiv:2203.05967.

[5] Ghosh, C. The Impact of Social Media on Conflict Perception: Case Studies of Russia-Ukraine and Gaza Conflicts.

[6] Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research. Detection Methods, and Opportunities. arXiv preprint arXiv, 2492706.De Prado, M. L. Advances in Financial Machine Learning, 2018.