A Machine Learning Model for detecting Covid-19 Misinformation in Swahili Language

Filbert Mlawa
School of Computational and Communication Sciences and Engineering, The Nelson Mandela African Institution of Science and Technology, Tanzania
mlawaf@nm-aist.ac.tz (corresponding author)

Elizabeth Mkoba
School of Computational and Communication Sciences and Engineering, The Nelson Mandela African Institution of Science and Technology, Tanzania
elizabeth.mkoba@nm-aist.ac.tz

Neema Mduma
School of Computational and Communication Sciences and Engineering, The Nelson Mandela African Institution of Science and Technology, Tanzania
neema.mduma@nm-aist.ac.tz

Received: 29 December 2022 | Revised: 2 February 2023 | Accepted: 5 February 2023

 DOI: https://doi.org/10.48084/etasr.5636

ABSTRACT

The recorded cases of corona virus (COVID-19) pandemic disease are millions and its mortality rate was maximized during the period from April 2020 to January 2022. Misinformation arose regarding this threat, which spread through social media platforms, and especially Twitter, often spreading confusion, social turmoil, and panic to the public. To identify such misinformation, a machine learning model is needed to detect whether the given information is true (true information) or not (misinformation). The aim of this paper is to present a machine-learning model for detecting COVID-19 misinformation in the Swahili language in tweets. The five machine learning algorithms that were trained for detecting Swahili language misinformation related to COVID-19 are Logistic Regression (LR), Support Vector Machine (SVM), Bagging Ensemble (BE), Multinomial Naïve Bayes (MNB), and Random Forest (RF). The study used the qualitative research method because non-numerical data, i.e. text, were used. Python programming language was used for data analysis due to its powerful libraries such as pandas and numpy. Four metrics were used to evaluate the model performance. The results revealed that SVM achieved the highest accuracy of 83.67% followed by LR with 82.47%. MNB achieved the best precision of 92.00% and in terms of recall and F1-score, RF, and SVM achieved the best results with 84.82% and 81.45%, respectively. This study will enable the public to easily identify Swahili language misinformation related to COVID-19 that is circulated on Twitter social media platform.

Keywords-COVID-19 pandemic; misinformation; machine learning; Twitter; Swahili language

I. INTRODUCTION

Social media platforms such as Instagram, Twitter, Facebook, WhatsApp, and TikTok are exceptional sources of information. They are strong tools of communication due to their accessibility, speed, and affordability [1]. These platforms are interactive communication tools and they enable smooth sharing of information, views, and thoughts. They also facilitate sharing of knowledge and experiences. They are unavoidable in our contemporary lives [2]. They gained popularity during the pandemic due to the social isolation and quarantine [3]. In spite of their advantages, social media platforms have been intentionally and unintentionally used for spreading misinformation regarding the COVID-19 pandemic [4-6]. Misinformation regarding COVID-19 can harm people and lead them into wrong directions [4]. Approximately 90% of misinformation still remain online [7]. Several studies have been conducted to provide solutions for misinformation detection related to COVID-19. However, no machine learning model has been developed for COVID-19 misinformation detection in Swahili language despite the fact that Swahili is an East African language widely used in Twitter and other platforms. Furthermore, no Swahili language datasets have been presented for detecting COVID-19 misinformation. This study was conducted to address this gap.

Since 2019, several studies were conducted to overcome the challenge of misinformation related to COVID-19. For instance, the authors in [10] presented a dataset in Luganda-English mixed code for detecting misinformation related to pandemic from social media. Several machine learning classifiers were applied including Discriminative Multinomial Naïve Bayes (DMNB), Support Vector Machine (SVM), and Bagging Ensemble (BE). The study revealed that DMNB outperformed the other classifiers with an accuracy of 78.19% and an F1-score of 77.90%. However, the study presented the model for detecting COVID-19 misinformation in Luganda-English code-mixed. Similarly, authors in [11] addressed the challenge by developing a machine learning model for detecting misinformation regarding the corona virus disease by relying on the information from UN, WHO, and UNICEF. Ten algorithms were employed, and Decision Tree (DT), Logistic Regression (LR), and Neural Networks (NNs) achieved the best results. However, the developed model was limited to misinformation in English language. Likewise, the authors in [9] conducted a study to address the same challenge of misinformation from Twitter. A large Arabic dataset was constructed. Eight traditional and deep learning classifiers were applied along with features such as word embedding and word frequencies. The best accuracy results were achieved by the Extreme Gradient Boosting (XGBoost) classifier. The constructed dataset was limited to Arabic content.

Authors in [2] conducted a similar study and three different misinformation detection models were proposed, including LSTM, KNN, and Multichannel CNN. The proposed models were found to achieve promising results. The model was focused on detecting misinformation in English from Twitter. Authors in [12] presented a simple Natural Language Processing (NLP) methodology for detecting misinformation related to the pandemic from YouTube. State-of-the-art pre-trained transfer learning methods such as BERT base, RoBERTa base, and XLNet base were employed. The highest accuracy achieved was 89.4%. Authors in [13] proposed a simple approach that used BERT embedding and Shallow Neural Networks (SNNs) for detecting misinformation related to COVID-19 where the BERT embedding transformer was found to perform better than SNNs. The presented approach was limited to misinformation in English language. Authors in [15] conducted a similar study using fuzzy clustering with deep learning models, focused on misinformation related to pandemic in English and Chinese. Authors in [14] conducted a study for detecting fake news related to COVID-19 using feature extraction methods. Several algorithms including Random Forest (RF), AdaBoost, KNN, and DT were applied. The highest accuracy achieved was 83.50% by RF. The dataset used had approximately 1,100 records and Swahili tweets were not part of it. Authors in [3] presented two misinformation datasets related to COVID-19 on Twitter and the system known as Checkvoid was developed based on machine learning and NLP technique for automatic detection of misinformation related to COVID-19. This study was focused on English language.

II. RESEARCH METHODOLOGY

The developed model follows several steps starting with collecting tweets using snscrape and ending with the evaluation of the performance of the model. Figure 1 shows a conceptual framework of this study that depicts all the employed steps. Jupyter notebook was used for writing the code in Python.

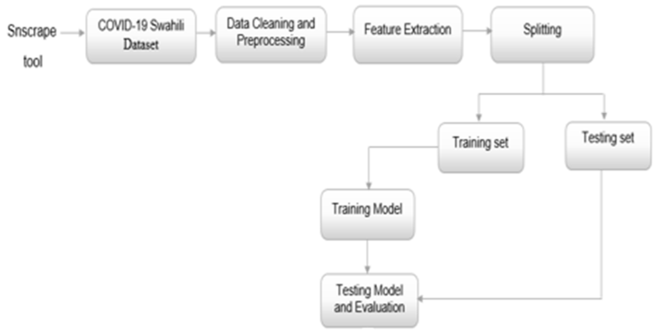

Fig. 1. Conceptual framework of the current study.

A. Data Collection

Data were collected from Twitter using snscrape. About 19,903 tweets were scraped from December 01, 2019 to January 31, 2022 based on common Swahili hashtags related to the pandemic including #covid-19, #uviko19, #chanjo_korona, #barakoa_inaokoa, #ujanjakuchanja, and #homa_kaliyamapafu and #korona_Tanzania.

B. Data Cleaning and Preprocessing

Data cleaning was conducted to remove duplicates, noise data, and data that were written in other language rather than Swahili. Data preprocessing was conducted to remove URLs, hashtags, white spaces, punctuation marks, and other special characters such as & and %. After annotation and labelling, 682 tweets equivalent to 54.47% were found to be true information and 570 tweets equivalent to 45.53% were found to be misinformation. Figure 2 and Table I shows a sample snapshot and the statistics of the dataset used in this study respectively.

```
df.head()
```

	num	tweet	label
818	0	chanjo ya uviko19 inakuweka kwenye hatari ya k...	Misinformation
37	0	CCM mkikutana na kukusanya watu korona inakimbia	Misinformation
633	1	Kila mmoja wetu anapaswa kuchukua tahadhari dh...	True information
552	1	Watu 102,093 wameripotiwa kufariki kutokana na...	True information
1020	0	Hakuna ugonjwa mbaya zaidi kwa sasa hapa dunia...	Misinformation

Fig. 2. Sample of the dataset.

TABLE I. DATASET STATISTICS

Data	True information	Misinformation	Total
Number	682	570	1,252
%	54.47	45.53	100

C. Data Visualization

Various tools and methods can be used for visualizing data. This study used the word cloud method. Figures 3 and 4 show the word clouds for true information and misinformation respectively.

Fig. 3. True information word cloud.

Fig. 4. Misinformation word cloud.

D. Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) was used for the reduction of the dimensionality of the data by converting text features into word vector representation. It involved two aspects, i.e. TF and IDF. It was computed as TF*IDF, where:

$$\mathrm{TF}(t) = \frac{Ci,j}{\sum_k Ci,j} \quad (1)$$

$$\mathrm{IDF} = \log\left(\frac{N}{dft}\right) \quad (2)$$

where Ci,j is the number of term t in the document, $\sum_k Ci,j$ is the number of all the words in the document, N is the considered document, and dft is the total documents that have the term t.

E. Data Splitting

The dataset was split into training and testing sets. 80% of the data were used for training of the model and 20% for testing.

F. Model Development

Five algorithms were trained for detecting misinformation in Swahili language related to pandemic, namely LR, SVM, BE, RF, and MNB.

1) Logistic Regression

LR is an algorithm for predicting discrete or categorical values. It is a supervised machine learning algorithm for solving classification problems by using the logistic function known as the sigmoid curve which is an s-shaped curve and is represented by:

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

It can classify data in binary or binomial, ordinal or ordered multiclass, and multinomial or multiclass classification.

2) Support Vector Machine

SVM is a popular supervised machine learning algorithm used for both classification and regression problems. It aims at finding a hyperplane in an N-dimensional space for classifying data points into a number of different classes [17, 20]. Equation (4) represents the general form of the hyperplane for classifying the datapoints.

$$Wx + b = 0 \quad (4)$$

where W is the vector normal to hyperplane (weight vector), x is the input feature, and b is the bias.

3) Random Forest

RF is a supervised learning method based on DT. It is used for both classification and regression problems [18]. It randomly selects samples of the dataset, builds a DT for each selected sample, and makes predictions from each DT. The prediction results are combined and the final classification results are obtained by the voting method.

4) Multinomial Naïve Bayes

MNB is supervised machine learning method based on simple and probabilistic family. The classifier assumes that all features in the given class are conditionally independent.

5) Bagging Ensemble

Bagging stands for Bootstrap Aggregation (Bootstrap AGGregatING). It is an Ensemble Learning (EL) algorithm in which every predictor or model is independent of the others. It involves two key ingredients such as bootstrap and aggregation.

G. Model Evaluation

The performance of the models was evaluated by four metrics derived from the confusion matrix. These are: accuracy, precision, recall, and f1-score [10, 19].

1) Accuracy

It is expressed as the ratio of the correctly classified observations to the total observations:

$$\mathrm{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+FP+TN}} \quad (5)$$

2) Precision

It is expressed as the ratio of the correctly classified positive observations to the total observations that were classified as positive:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (6)$$

3) Recall

It is expressed as the ratio of the correctly classified positive observations to the observations that were classified correctly:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (7)$$

4) F1-Score

Is the harmonic mean of the precision and recall, and is mathematically expressed as:

$$\mathrm{F1-Score} = \frac{2*\mathrm{Precision}*\mathrm{Recall}}{\mathrm{Precision+Recall}} \quad (8)$$

In (5)-(8), TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative.

III. RESULTS AND DISCUSSION

A. Results

The results presented in Figure 5 show that the SVM achieved the highest accuracy of 83.67% followed by LR with accuracy of 82.47%. MNB and RF obtained the same accuracy of 80.48% while the BE had the lowest accuracy of 76.01%. As for precision measure, MNB achieved the best results of 92.00%. In terms of recall and F1-score, RF and SVM achieved the best results with 84.82% and 81.45%, respectively.

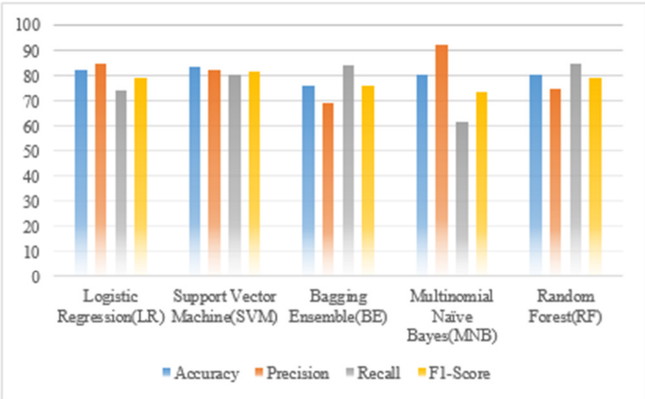

Fig. 5. Evaluation.

B. Discussion

The dataset used in this study contains only 1,252 data entries. LR and SVM were found to perform much better in predicting COVID-19 misinformation in Swahili language. LR was then chosen as the final algorithm to be considered.

IV. CONCLUSION

The threat of the COVID-19 pandemic resulted to the generation of a huge amount of information related to it on social media platforms such as Twitter. Some of these information are true and constructive while some are misinformation and have negative consequences to the public [6]. Several studies have been conducted to address the challenge of misinformation circulating on twitter and other social platforms, however, most of those studies have focused on languages other than Swahili [2, 9-13]. This study was conducted to fill this gap.

This study also presented a well labelled Swahili language misinformation detection dataset which will be used as a starting point by practitioners and researchers for further research regarding the COVID-19 pandemic. This study was limited to the Twitter social media platform. Future research can be extended to other social media platforms as well, such as Facebook, Instagram, and Whatsapp.

REFERENCES

[1] A. Glazkova, "A Comparison of Text Representation Methods for Predicting Political Views of Social Media Users," in *Proceedings of the International Scientific and Practical Conference "Information Technologies and Intelligent Decision Making Systems,"* Moscow, Russia, Jan. 2021, Art. no. 41.

[2] M. N. Alenezi and Z. M. Alqenaei, "Machine Learning in Detecting COVID-19 Misinformation on Twitter," *Future Internet*, vol. 13, no. 10, Oct. 2021, Art. no. 244, https://doi.org/10.3390/fi13100244.

[3] S. Dadgar and M. Ghatee, "Checkovid: A COVID-19 misinformation detection system on Twitter using network and content mining perspectives." arXiv, Jul. 20, 2021, https://doi.org/10.48550/arXiv.2107.09768.

[4] B. Al-Ahmad, A. M. Al-Zoubi, R. Abu Khurma, and I. Aljarah, "An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information," *Symmetry*, vol. 13, no. 6, Jun. 2021, Art. no. 1091, https://doi.org/10.3390/sym13061091.

[5] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating Deep Learning Approaches for Covid19 Fake News Detection," in *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, 2021, pp. 153–163, https://doi.org/10.1007/978-3-030-73696-5_15.

[6] A. B. Shams *et al.*, "Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic," *Healthcare*, vol. 9, no. 2, Feb. 2021, Art. no. 156, https://doi.org/10.3390/healthcare9020156.

[7] M. Bahremani, D. Berezovski, R. Perencsik, and Y. Zhang, "COVID-19 Fake News Detector," Sep. 2022, [Online]. Available: https://ssc.ca/sites/default/files/imce/second_prize.pdf.

[8] *Module 8: 'False News', Misinformation and Propaganda*. Media Defence, 2020.

[9] S. Alqurashi, B. Hamoui, A. Alashaikh, A. Alhindi, and E. Alanazi, "Eating Garlic Prevents COVID-19 Infection: Detecting Misinformation on the Arabic Content of Twitter." arXiv, Jan. 09, 2021, https://doi.org/10.48550/arXiv.2101.05626.

[10] P. Nabende, D. Kabiito, C. Babirye, H. Tusiime, and J. Nakatumba-Nabende, "Misinformation detection in Luganda-English code-mixed social media text." arXiv, Apr. 03, 2021, https://doi.org/10.48550/arXiv.2104.00124.

[11] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting Misleading Information on COVID-19," *IEEE Access*, vol. 8, pp. 165201–165215, 2020, https://doi.org/10.1109/ACCESS.2020.3022867.

[12] J. C. Medina Serrano, O. Papakyriakopoulos, and S. Hegelich, "NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube," in *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Apr. 2020.

[13] G. S. Cheema, S. Hakimov, and R. Ewerth, "TIB's Visual Analytics Group at MediaEval '20: Detecting Fake News on Corona Virus and 5G Conspiracy." arXiv, Jan. 10, 2021, https://doi.org/10.48550/arXiv.2101.03529.

[14] S. Khan, S. Hakak, N. Deepa, B. Prabadevi, K. Dev, and S. Trelova, "Detecting COVID-19-Related Fake News Using Feature Extraction," *Frontiers in Public Health*, vol. 9, 2022, https://doi.org/10.3389/fpubh.2021.788074.

[15] M.-Y. Chen and Y.-W. Lai, "Using Fuzzy Clustering with Deep Learning Models for Detection of COVID-19 Disinformation," *ACM Transactions on Asian and Low-Resource Language Information Processing*, Apr. 2022, https://doi.org/10.1145/3548458.

[16] H. Alhakami, W. Alhakami, A. Baz, M. Faizan, M. W. Khan, and A. Agrawal, "Evaluating Intelligent Methods for Detecting COVID-19 Fake News on Social Media Platforms," *Electronics*, vol. 11, no. 15, Jan. 2022, Art. no. 2417, https://doi.org/10.3390/electronics11152417.

[17] S. Nuanmeesri, "A Hybrid Deep Learning and Optimized Machine Learning Approach for Rose Leaf Disease Classification," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7678–7683, Oct. 2021, https://doi.org/10.48084/etasr.4455.

[18] M. A. Alsuwaiket, "Feature Extraction of EEG Signals for Seizure Detection Using Machine Learning Algorithims," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9247–9251, Oct. 2022, https://doi.org/10.48084/etasr.5208.

[19] W. M. S. Yafooz, E. A. Hizam, and W. A. Alromema, "Arabic Sentiment Analysis on Chewing Khat Leaves using Machine Learning and Ensemble Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 6845–6848, Apr. 2021, https://doi.org/10.48084/etasr.4026.

[20] M. Anwer, S. M. Khan, M. U. Farooq, and Waseemullah, "Attack Detection in IoT using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7273–7278, Jun. 2021, https://doi.org/10.48084/etasr.4202.