Amharic Fake News Detection on Social Media Using Feature Fusion

Menbere Hailu Worku[1(✉)] and Michael Melese Woldeyohannis[2]

[1] University of Gondar, Gondar, Ethiopia
menbere.hailu@uog.edu.et
[2] Addis Ababa University, Addis Ababa, Ethiopia
michael.melese@aau.edu.et

Abstract. These days, many people use social media as a source of information and medium of communication due to its easy to access, fast to disseminate and low-cost platform. However, it also enables the wide propagation of fake news which causes economic, political, and social crises to the society. As a result, many researchers have been working towards detecting fake news. Most of the researches concerned on linguistic analysis of news content to identify its credibility, however fake news is also written intentionally to mislead users by mimicking true news. Beside this, Amharic is one of the under-resourced language that suffer from the benefits of fake news detection. To overcome the problem of fake news using content feature and under-resourced language, this study uses a feature fusion of linguistic and social context feature of the publisher information to detect Amharic fake news. For this, a total of 4,590 instance has been collected from different Facebook pages in different domain. Each article have been annotated by professional journalists and linguist for the purposes of doing experiments. The experimental result of feature fusion-based experiment shows at least 94.13% and at most 98.7% with a high relative error reduction over the content-based approaches. The result obtained from the experiment shows that, it is promising to detect fake news using fusion feature. We are now working towards incorporating intentionally edited pictures to the news content as part of the fake news detection.

Keywords: Amharic · Amharic fake news · Content-social feature

1 Introduction

As a result of advancing technology, ease of access and a low-cost platform, many people choose social media as a source of information and medium of communication [1]. Due to this fact, the traditional news media such as TV and magazine are evolving to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats to reach the consumer in different ways [2]. Among these digital medias, social media has the potential to reach large number of audiences over the traditional news media [1]. Moreover, Social media is a powerful tool, for both business and individual to share and disseminate information in a short period than ever before when approached with appropriate content. From the very nature of social media, the information that the consumers get over the social media are not always

M. L. Berihun (Ed.): ICAST 2021, LNICST 411, pp. 468–479, 2022.
https://doi.org/10.1007/978-3-030-93709-6_31

correct and may not reach promptly. However, spreading false information on social media is for influencing others opinion while earning a money which highly affect the economy [2].

Nowadays, the issue of social media and fake news has become critical and the major issue among concerned stakeholders in country like Ethiopia [3]. Social media networks and immediate messaging applications allow misleading content to reach several people [4]. Due to the appealing nature, fake news spreads rapidly and influence people's perceptions about various subject. This causes genocides and protests that leads to economic, social, and political crises. To further detect the spread of the fake news, a number of researches attempts have been made using content-based [5–7] and social feature-based [8, 9] detection techniques despite their own drawbacks. Content-based fake news detection uses the content to detect real or not while the social features of the news.

The main problem of content-based fake news detection is that, fake news are written intentionally to mislead consumer, which makes it nontrivial to detect based on news content [10]. These difficulties are probably the reason behind the limitation of the content-based methods for fake news detection on social media. In addition to this, under-resourced Ethiopian languages like Amharic highly suffer from the lack of fake news detection. The main reason behind is the morphological complexity of the Amharic language and unavailability of linguistic resources such as stop words, Amharic stemmer and standard dictionary.

Similarly, fake news detection model using social features of news such as "like", "dislike" and "share" better determines in the presence of the social feature and tends to fail when the data did not have a social response [8, 9]. Thus, there is a need to fill the gap to overcome these problems by combining the content and social feature to detect fake news from the real on social media using state-of-the-art tools for techno-logically unsupported, morphological rich and complex Amharic language.

2 Related Work

Fake news detection is the process of discriminate fake news from genuine one news that contains false information from a certain one [11]. According to Shu and Liu [11] fake news detection methods generally focus on using news contents-based, social contexts-based and hybrid-based. News content-based approaches extract features from linguistic and visual information. Social context-based approaches incorporate features from user profiles, author analysis, and social networks [12]. Hybrid based approach is a combination of technique, method and feature from content-based method and social context-based method, using auxiliary information from different perspective [13].

An automated fake news detection system primarily introduced under the name "FakeNew-Tracker" for understanding and detecting of fake news based on the confirmed fake news and real news from fact-checking websites such as PolitiFact [14]. Then using the advanced search API of Twitter, they gathered the fake/real news related tweets that spread over the Twitter, through social engagement such as tweet replies, retweets, and favorites. To detect fake news, Auto-encoders and the doc2vec features are used for the content of the article and RNN for social context. The

experiment result shows a 71.7% accuracy. However, due to the language difference, morphological richness and complexity of the language, tweets, retweets used cannot be applied for Amharic language.

Another attempt has been made to classify the social media posts with better accuracy based on the users who "like" or "dislike" them and achieved significant accuracies even with a very small training dataset [15]. However, social interactions features cannot be used when a post has no likes, and probably perform worse when a post collects only a few social reactions. To overcome this problem, a content-based features were introduced only when the social-based methods perform poorly [16]. The model based entirely on only one type of feature at a time tested on real-world data and obtained an accuracy of 81.7%.

In the other case, content based fake news detection have been attempt and it concentrates only on political news articles and news articles from both false and truthful categories in the year 2016 [17]. The length of each article is more than 200 characters. The author utilizes the techniques of n-gram analysis and machine learning with two distinct linguistic based features extraction techniques. Term Frequency-Inverted Document Frequency (TF-IDF) and N-gram are studied and compared using six different machine classification techniques. Compared to other, TF-IDF and unigram as a feature extraction technique, and Linear Support Vector Machine (LSVM) as a classifier, with an accuracy of 92%, experimental evaluation produces the best performance. The study shows a good result but the feature extraction technique did not contain the semantics of the news content and also this study were not applicable in Amharic news.

Consequently, a content-based fake news detection used after preprocessing and feature extraction using twenty-three machine learning algorithms to build a model [7]. The performance evaluation has been performed on three readymade datasets called BuzzFeed dataset contains 1,627 news articles and achieved an accuracy score of 65% by zeroR algorithm, Random political news data contains 75 articles and achieved an accuracy score of 69.3% by sequential minimal optimization algorithm, and ISOT fake news data collection contains 44,898 article and achieved an accuracy score of 96.8% by decision tree algorithm.

In addition to this, a novel machine learning model based on the NLP technique for detection of fake news by using both content-based features and social features of news which consist of a headline and body [18]. To examine the social features of the news articles, their authenticity can be established by using Facebook Page ID, Source, and Face-book App ID. In data processing, a bag of words for feature extraction from headline and body and classification were performed using a probabilistic classifier and tested on the FakeNewsNet dataset and achieved an average accuracy of 90.62%. The performance of the experiment shows a good result but when we see the feature extraction, it works only by counting the occurrence of the word in the documents without the semantic and order of words in the news content. The above research study works either on the content or social context, however this study uses both content and social features to detect Amharic fake news article.

3 Amharic Language

Amharic is one of the Semitic languages spoken in many parts of Ethiopia and it is an official working language for the Federal Democratic Republic of Ethiopia [19]. It is also a working language of several regional states including Amhara, South Nation, and Nationalities. The language is also used for the interregional communication. According to [20], more than 25 million with up to 22 million native speakers speaks Amharic as their first language in Ethiopia and it has over 4 million second-language speakers within the country and a further 3 million around the world. The majority of Amharic speakers found in Ethiopia even though there are also speakers in a number of other countries, particularly Italy, Canada, the USA and Sweden. The language Amharic is also used in commerce, government, media, and national education. Amharic is written using a writing system called Fidel (ፈደል), adapted from the one used by the Ge'ez language [19].

3.1 Amharic Writing System

The language Amharic which is phonetic contains 34 base characters each of the characters ordered in the form of seven orders [16]. The seven orders represent syllable combinations consisting of a consonant following a vowel. In Amharic language seven vowels are used, each of them has seven distinct forms that reflect the seven vowel sounds. They are አ/a/, ኡ/u/, ኢ/i/, ኣ/a/ኤ/e/, እ/ə/, ኦ/o/. This vowel is fused to the consonant form in the form of diacritic markings. The diacritic markings are strokes attached to the base characters to change their order. In the Amharic writing system, there is no capital-lower case distinction. The 34 basic characters and their orders give 238 distinct symbols. In addition to the 238 symbols, Amharic script contains five-so called labiovelars which have five orders and 18 additional labialized consonants. These letters do not use more frequently. Table 1 presents the distribution of Amharic characters in the modern writing systems.

Table 1. Amharic writing system

	Character order						
	1	2	3	4	5	6	7
ሀ	ሀ(ha)	ሁ(hu)	ሂ(hi)	ሃ(hā)	ሄ(hé)	ህ(he/h)	ሆ(ho)
ለ	ለ(la)	ሉ(lu)	ሊ(li)	ላ(lā)	ሌ(lé)	ል (le/h)	ሎ(lo)
ሐ	ሐ(ha)	ሑ(hu)	ሒ(hi)	ሓ(hā)	ሔ(hé)	ሕ(he/h)	ሖ(ho)
መ	መ(ma)	ሙ(mu)	ሚ(mi)	ማ(mā)	ሜ(mé)	ም(me/h)	ሞ(mo)

Amharic writing scheme has some issues that are difficult to process Amharic text. One of these challenges is the redundancy of characters used in Amharic, more than one character to represent the same sound. The various forms have their meaning in Ge'ez, but there is no clear rule that shows their purpose in Amharic [19]. The problem of the same sound with various characters is not only observed with core characters but

also exhibited in the same order of characters. Those are, ሀ and ሃ; ሐ and ሓ; ኀ and ኃ; አ and ኣ; ሰ and ሠ. A word formed by using this character has the same meaning. For example, the word 'sun' could be written in a different way like ፀሀይ/S'ähay/, ፀሀይ/Ts'ähay/ጸሃይ/Tse-hi/, ፀሃይ, ጸሐይ, ፀሐይ/Ts'ähay/, ፀሓይ, ፀሓይ.

In addition to the 238 symbols, Amharic script contains five-so called labiovelars which have five orders and 18 additional labialized consonants. These letters do not use more frequently. Those characters are ሏ፣ ሟ ፣ሷ ፣ሯ ፣ሿ.

3.2 Amharic Morphology

Amharic is one of the most morphological rich and complex languages that [19]. It shows a root-pattern morphological phenomenon. With a basic lexical meaning [19]. Amharic words are categorized under five categories based on the use of morphology and position of the word in a sentence. Amharic word categories are noun, verb, adjective, adverb, and preposition. Nouns are words used to name or identify any class of things, people, places or ideas or a particular of these. Verbs are the most important part of speech because which shows the action or state, a word that tells the listener or reader what is happening in the sentence and have more to do with mental processes and perceptions. The adjective is a word that comes before a noun and adds some kind of qualification to the noun. An adverb is a word that qualifies the verb by adding extra ideas from time, place, and situations point of view. A preposition is a word that can be placed before a noun and perform adverbial operations related to place, time, cause, etc.

4 Data Collection and Preparation

In the experimental research, it was possible to collect data from social media and identify dependent and independent variable then perform experimental analysis with different experimental setting and machine learning techniques [21] to detect the news article as real or fake using both content and social feature as independent variable. This work collected the data from Facebook then prepared and annotated by experts for the purpose of the experiment. During experimental analysis, features has been extracted from content and social context of collected data, then total of 8 experiments have been performed using machine learning techniques to train and classify the dataset as real or fake news. The evaluation techniques take place after the classification task to check the performance of the proposed model.

Gereme et al. [22] states that they collect and prepare publicly available datasets for the fake news detection in local language especially for Amharic content news, however the data is not accessible. So, we collect data from the Facebook platform and make available for expert (journalist) for labeling (annotation).

The Facebook pages are selected from different categories including news media, broadcasting pages, bloggers and journalist's pages, political parties, politicians and government office pages. The pages are selected based on the frequency of news post content and contain accepted news for real and fake news through observation of the domain experts specifically the journalist. The data is collected in specific periodic manner which enables collect data with in a specific range as per suggestion of different

researchers [14]. Accordingly, the data for this study have been collected from those selected pages which contain Amharic news contents starting from April 10, 2020 up to August 25, 2020 which covers four months of data. The data collection is made by using Data-miner web scraper tool which have effective way to collect the news content and social context of the perspective news from social media platform. During the data collection process, 4,590 data with six attributes; news content, posted date, published page, created date, verification of the published Facebook pages, number of follower and like of published pages are collected from the selected pages.

After the data collection, the annotation process takes place to label the given article as real or fake, it is not easy to find well-organized true and fake Amharic news, especially getting the fake article was tedious, and however there are new pages created for promoting fake news based on the situation of the country. However, to get the clarity of the news, we checked those manually using senior journalists and linguist. There are six senior journalist who are news editor and organizer at Gondar FM and one linguist from the linguistic department of government institution. Each annotator takes 765 post that are randomly selected from the collected data's. Since, the annotation process is manual, tedious and needs detail analysis and clarity, the annotation process takes two months and annotated based on the guidelines those are, the annotator labeled as fake/false when one or more of the following conditions is met;

- The statement is fully contradict to the truth,
- The statement seems true but interpreted in the wrong context,
- The posted date event and the nature of the news is mismatch which means the news is true and it happened before and presented it as it hap-pens now,
- Check the date of event and the news.

Accordingly, from the total of 4,590 instances, 2,211 news has been labeled as Fake and 2,379 news has been labeled as Real. Figure 1 represented the distribution of fake and real news image.

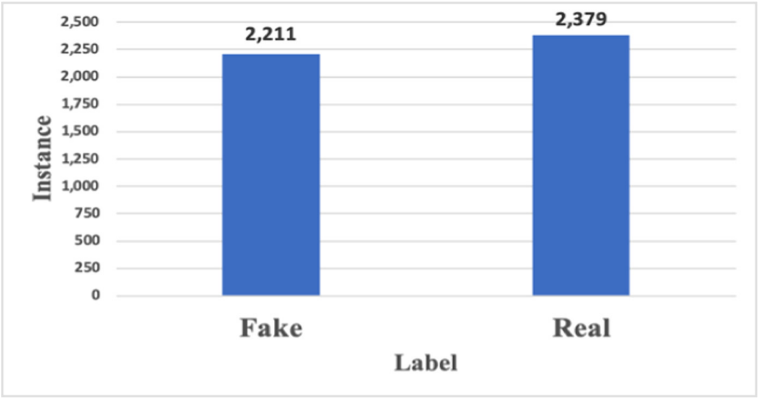

Fig. 1. Class label of distribution of the collected data

The challenge faced during the annotation was some post have written in both Amharic and Tigrinya language, in order to understand the meaning of full article it requires to translate those terms so it is the challenges for the annotator.

4.1 Data Preprocessing

In this work, two different kinds of preprocessing have been performed. One is for the content of the Amharic text and the other for the social features. Primarily, NLP based pre-processing used for the content of the article includes removing irrelevant characters. These characters include special symbol, emoji, Amharic and Arabic numbers and punctuation. In addition to this, Amharic character normalization were done to removes the Amharic character that made the same sound but different orthography. Consequently, content tokenization have been done after removing irrelevant and redundant characters. This is followed by the removal of Amharic stop words from the tokenized word and stemming using a rule-based Hornmorpho [23]. Finally, the output of content pre-processed data is used as an input for feature ex-traction process. The second one is handling categorical data that has been per-formed to convert categorical data of social-context information to numeric.

4.2 Features Extraction

In order to analyze the news content and verify the given news article, it is important to use feature extraction technique to extract features from the news content. In this study we used NLP-based feature extraction technique like TFIDF, N-gram weighted by TFIDF (i.e. unigram, bigram and trigram) and word2vec for the content. Beside content feature, for the social context features like number of like, follower, created date and credibility status of Facebook page that the news articles have been posted.

4.3 System Design

This system design for Amharic Fake News detection contain three different phase; pre-processing, feature extraction and classification. Those major phases are presented on the following figure.

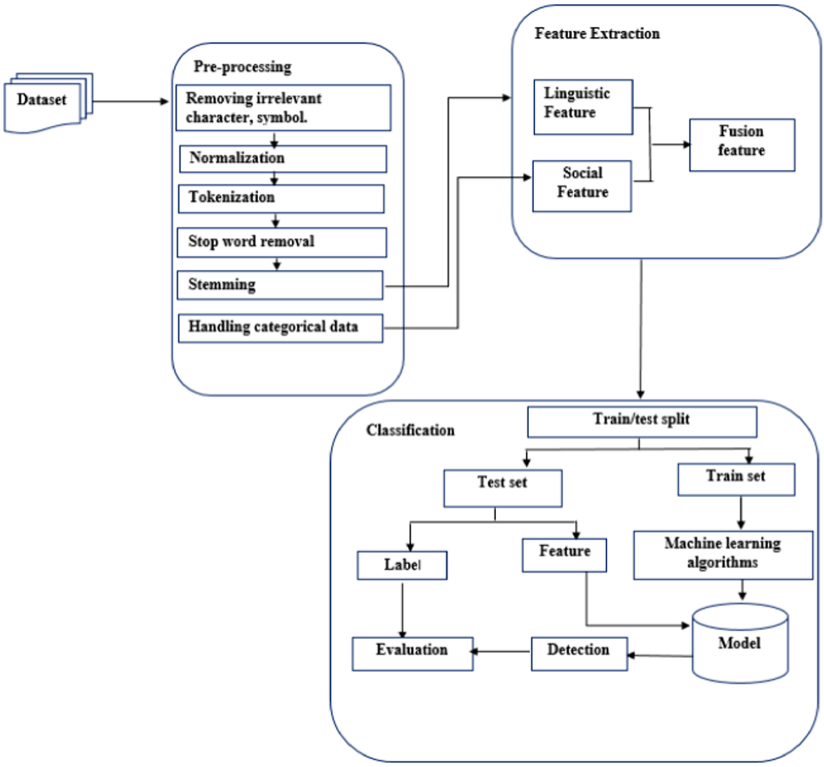

Fig. 2. The architecture of feature fusion of fake news detection

The above Fig. 2 shows the architecture of fake news which differ from other fake news detection by applying feature fusion to determine the reliability of the news article based on linguistic and social feature. Primarily, the dataset is collected from Facebook having news content and social information of news and it is annotated by expert (journalist).The news content in the dataset which have been preprocessed using Amharic data pre-processing technique including the removal of irrelevant character, normalization, tokenization, stop words removal beside stemming and social context information also preprocessed with handling categorical data technique. After all the preprocessing, feature extraction takes place to extract features from preprocessed news content using N-gram weighted by TFIDF and word2vec. The output of feature extraction of content-based and preprocessed social context information are combined and it makes an important feature vector of the dataset for training the model. After fusion of content and social feature, classification model has been built by training set and three machine learning algorithms, Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). Finally, the performance of the model on the test set has been evaluated using accuracy, recall, F-measure, and precision.

5 Experiment and Discussion

In order to perform experiments, different python libraries for machine learning including the latest one Flair, Natural language processing toolkit (NLTK), Scientific computing (SciPy), and multidimensional processing (NumPy) has been used at different stages of the machine learning. Along with Python, the Jupyter notebook is used as a tool for an integrated development environment (IDE). Because of these advantages, the researcher uses python for preprocessing as well as for building models. Natural language processing uses for preprocessing the data like for tokenization, stemming, and normalizing the news content and analyzing some other social-context feature of data. In addition to this, LR, RF, and SVM, the machine learning algorithm has been used to create an Amharic fake news detection model.

To minimize the opportunity of creating over-fitting the dataset is first divided into two different datasets randomly, 80% of the data for training and the remaining 20% for evaluating the model. Once the dataset splitted, to investigate fake news detection for Amharic language, two independent experiment were conducted using three different machine learning algorithms. The first experiment is conducted using only content features s with three machine learning algorithms LR, SVM, and RF. Similarly, the second experiment also contain four experiments which is a hybrid feature of content and social-context features with three machine learning algorithms. Beside this, both experiments are conducted using n-gram weighted by TFIDF and word2vec. Table 2 presents the experimental result with respect to the linguistic feature, algorithm and content feature extracted against weighted n-gram and word2vec.

Table 2. Experiment result of fake news detection

Feature	ML algorithm	N-gram			w2vec
		Unigram	Bigram	Trigram	
Linguistic/content feature	LR	93.35	89.76	76.57	92.48
	SVM	89.76	81.69	74.94	92.48
	RF	91.5	86.71	79.52	94.33
Fusion feature	LR	96.84	96.84	96.84	93.79
	SVM	96.4	96.4	96.4	96.4
	RF	99.56	99.34	99.34	99.67

The above Table 2 presents the experiment result in terms of accuracy. The experiment result shows that the content/linguistic feature on the unigram and bi-gram with logistic regression achieves the highest accuracy 93.33% and 89.76% respectively. The reason behind the higher performance for LR is the ability to transform the class label using the probability measures of the unigram and bigram vectors. In addition to this, unigram and bigrams are relatively context independent. However, RF is much better than SVM but lower than LR.

Unlike the unigram and bigram, trigram and word2vec achieved the highest performance in the Random Forest. Because it is a capable to selecting the highest

classification result among the selected decision tree results. In addition to that the word2vec features achieved the highest accuracy, because it uses the semantic information of the news content on the vectors than N-grams features. In other case, all fused feature achieves highest accuracy on RF.

The above Fig. 3 presents the comparison of the experimental result in terms of accuracy.

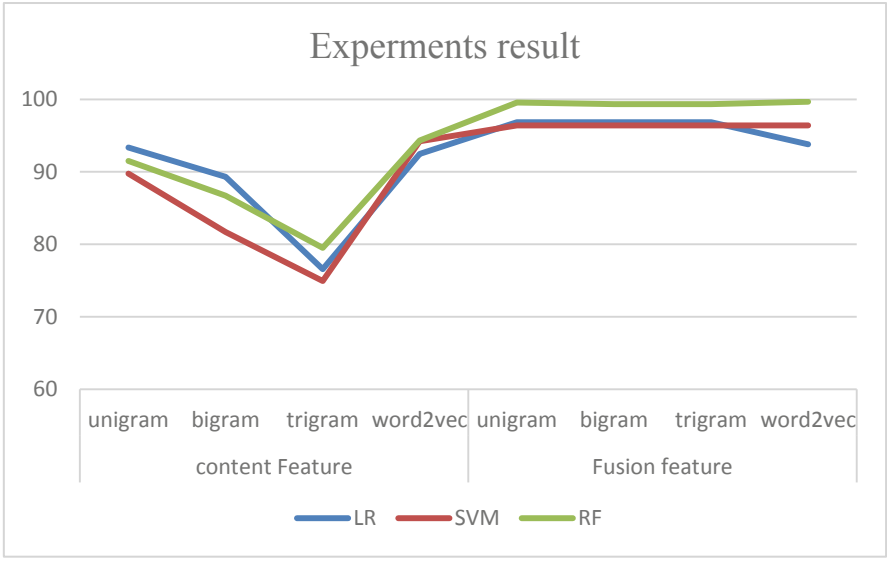

Fig. 3. Experiments result

The accuracy result has shown us which feature extraction technique, approach and model perform well than others. The result that has been observed while experimenting shows which technique should be used to build the prototype which received new data contain content news and social context. From the stated experiments, the content-based approach of the word2vec feature extraction technique with support vector machine achieves the highest accuracy 94.22% among other content-based feature extraction techniques like N-gram levels with TFIDF.

Because word2vec retains the semantic meaning of different words in a document. The context information is not lost and another great benefit of the word2vec approach is that the size of the embedding vector is very small. Each dimension in the embedding vector contains information about one aspect of the word. There is no need of huge sparse vectors, unlike the bag of words and TF-IDF approaches in addition to that trigram achieved the lowest accuracy among linguistic based analyses, because the occurrence of three consecutive words became. Whereas when we use a hybrid approach which means integrating the content-based feature with social con-text feature the accuracy is higher than content-based features. When we incorporate social context features to the content-based feature it performs higher than of con-tent-based feature.

The proposed model achieves its highest accuracy when using a fusion of word2vec and social context feature with Random Forest classifier because feature fusion get both content and social context feature other than of one of them. Among other machine learning algorithms Random Forest achieves the highest accuracy which is 99.67% because it enables to select the best classification result from random trees generated.

6 Conclusion and Future Work

This study proposed a solution for detecting fake news on social media especially Facebook by using a hybrid approach of fake news detection mechanism and ma-chine learning classification techniques for Amharic online news. 4,590 data are collected from 30 Facebook pages by using the "data_miner" tool that was posted with-in three months. When using the word2vec hybrid with social context and Random Forest classifier, the proposed model achieves 99.67% of the highest accuracy score. The first four experiment conducted on news content features (i.e. Uni-gram, Bigram, and Tri-gram word2vec). The rest of four experiments have done on incorporating of social context feature to content based feature. In addition to this, eight different experiments have been performed on three machine learning algorithms. The major contribution of this study is design a model for Amharic fake news detection using fused feature of both content and social context feature. In this research we only consider news text content and social information, it would be better for future re-searchers to incorporate also the image analysis as most of fake news distributer use unrelated and intentional edited pictures.

Acknowledgment. We would like to thank college of Informatics, University of Gondar for supporting the research work and Impact Amplifier Online Safety Project for sponsoring conference participation.

References

1. Perrin, A.: Social Media Usage: 2005–2015. Pew Research Center, Washington (2015)
2. Howell, L.: Digital wildfres in a hyperconnected world. WEF Report **3**, 15–94 (2013)
3. Abeselom, D.K.: The impacts of fake news on peace and development in the world: In the case of Ethiopia. Internal J. Curr. Res. **10**, 71356–71365 (2018)
4. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
5. Ahmed, H., Traore, I., Saad, S.: Detection of online fake news using n-gram analysis and machine learning techniques. In: Traore, I., Woungang, I., Awad, A. (eds.) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pp. 127–138. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-69155-8_9
6. Riedel, B., Augenstein, I., Spithourakis, G.P., Riedel, S.: A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv preprint arXiv:1707.03264 (2017)
7. Ozbay, F., Alatas, B.: Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A Stat. Mech. Appl. **540**, 123174 (2020)

8. Tacchini, E., Ballarin, G., Vedova, M.L.D., Moret, S., Alfaro, L.D.: Some like it hoax: automated fake news detection in social networks. In: Proceedings of the Second Workshop on Data Science for Social Good (SoGood), Skopje, Macedonia, 2017. CEUR Workshop Proceedings, vol. 1960 (2017)
9. Tlahun, A., Beshah, D.: Fake news detection model using machine learning approach: the caseof Amharic news on social media. unpublished masters thesis ,University of Gondar, Department of Information systems (2020)
10. Shu, K., Wang, S., Liu, H.: Exploiting Tri-Relationship for Fake News Detection (2017). arXiv:1712.07709 [cs]
11. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, A.H.: Fake news detection on social media: a data mining perspective. In: KDD Exploration Newsletter (2017)
12. Zhou, X., Zafarani, R., Shu, K., Liu, A.H.: Fake news: fundamental theories, detection strategies and challenges. In: WSDM (2019)
13. Mahid, Z.I., Manickam, S., Karuppayah, S.: Fake news on social media: brief review on detection techniques. In: 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACC). IEEE, Subang Jaya (2018)
14. Shu, K., Mahudeswaran, D., Liu, H.: FakeNewsTracker: a tool for fake news collection, detection, and visualization. Comput. Math. Organ. Theory **25**(1), 60–71 (2018). https://doi.org/10.1007/s10588-018-09280-3
15. Tacchini, E., Ballarin, G., Vedova, M., Della, S.M., Alfato, L.D.: Some like it hoax: automated fake news detection in social networks (2017). arXiv preprint arXiv:1704.07506
16. Vedova, D., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., de Alfaro, L.: Automatic online fake news detection combining content and social signals. In: 22nd Conference of Open Innovations Association (FRUCT), pp. 272–279 (2018)
17. Ahmed, H., Traore, I., Saad, S.: Detection of online fake news using n-gram analysis and machine learning techniques. In: Traore, I., Woungang, I., Awad, A. (eds.) ISDDC 2017. LNCS, vol. 10618, pp. 127–138. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69155-8_9
18. Shubham, B., Vijay, B., Jain, P., Chawla, M.: Natural language processing based hybrid model fordetecting fake news using content-based features and social features. Inf. Eng. Electron. Bus. **4**, 1–10 (2019)
19. Getahun, B.A., Fekade.: Event modeling from amharic news article. Unpublished Master's Thesis, Department of Computer Science, Addis Ababa University (2018)
20. ethiopiaonlinevisa (2020). https://www.ethiopiaonlinevisa.com/amharic-the-ethiopian-language/
21. Morrison, G.R., Ross, S.M.: Experimental Research Methods. Researchgate (2003)
22. Gereme, F., Zhu, W., Ayall, T., Alemu, D.: Combating fake news in "low-resource" languages: amharicfake news detection accompanied by resource crafting. Information **12**, 20 (2021)
23. Michael, G.: Hornmorpho 2.5 User's Guide. Indiana University, School of Informatics and Computing (2012)
24. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. In: KDD Exploration Newsletter (2017)
25. Potthast, M., Kiese, J., Reinart, K., Bevendorff, J., Stein, B.: A Stylometric Inquiry into Hyper partisan and Fake News (2017). arXiv preprint arXiv:1702.05638