BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Electrical Science & Engineering
https://ojs.bilpublishing.com/index.php/ese

ARTICLE

Classification and Detection of Amharic Language Fake News on Social Media Using Machine Learning Approach

Kedir Lemma Arega*

School of Technology and Informatics, Department of Information Technology, Ambo University, Ethiopia

ARTICLE INFO

Article history
Received: 16 November 2021
Revised: 31 December 2021
Accepted: 13 January 2022
Published Online: 26 January 2022

Keywords:
Amharic fake news detection
Amharic posts and comments datasets
Classification
Machine learning
Social media

ABSTRACT

The pervasive idea of web-based media stages brought about a lot of sight and sound information in interpersonal organizations. The transparency and unlimited way of sharing the data via online media stage encourages data spread across the organization paying little mind to its noteworthiness. The multiplication of misdirecting data in regular access news sources, for example, web-based media channels, news websites, and online papers has made it trying to recognize dependable news sources, in this way expanding the requirement for computational devices to give bits of knowledge into the unwavering quality of online substance. The broad spread of phony news contrarily affects people and society. Along these lines, counterfeit news identification via web-based media has as of late become arising research drawing in enormous consideration. Observing the possible damage caused by the rapid spread of fake news in various fields such as politics and finance, the use of language analysis to automatically identify fake news has attracted the attention of the research community. A social networking service is a platform for people with similar interests, activities, or backgrounds to form social networks or social relations. Participants who register on this site with its own expression (often a profile) and social links are generally offered a social network service.

1. Introduction

Counterfeit news is anything but an original idea. Strikingly, the thought has been in presence even before the development of the Web as distributers utilized bogus and deceiving data to additional their advantages [1]. The most serious issue of these days' online media papers is generally phony news stories [2]. There is an increment in the dispersion of bogus news, tricks and other misleading statements in the public eye. The distribution of phony information is going not just on a virtual world (web-based media, social network, and so forth) yet additionally from one individual to another. Counterfeit news is a major string since it can influence many individuals all throughout the planet consistently [3]. Counterfeit news is progressively turning into a threat to our general public.

*Corresponding Author:
Kedir Lemma Arega,
School of Technology and Informatics, Department of Information Technology, Ambo University, Ethiopia;
Email: kedirnaw1999@gmail.com

DOI: https://doi.org/10.30564/ese.v4i1.3885

It is normally produced for business interests to draw in watchers and gather promoting income. Notwithstanding, individuals and gatherings with conceivably pernicious plans have been known to start counterfeit news to impact occasions and approaches around the world. The use of the Internet has grown in recent years. The term "social media networks" has gained popularity as a result of the widespread use of the Internet. Everyone who uses the Internet is familiar with online access [4]. An online access is a grouping of several social networking websites. Online network is a platform that allows users to express their opinions about anything [5].

Online Social Networks are the most popular way for people all over the world to exchange information. Social networks are the focal point people use these devices for many different purposes and they provide a variety of new tools for the user community. A social network is best described as a graphical structure with nodes and edges representing users and their interactions [6]. Depending on the network structure used, the nodes and edges in a Social Network graph can be labelled or unlabelled. Because of social intelligence's high reputation, social networking sites such as Facebook, WhatsApp, Twitter, Instagram and Telegram has become the tool of choice for communicating and sharing information with a wide variety of users, including individuals and businesses [7]. Customers of online media will play an important role and will be solely responsible for the information exchanged on the networks. Users share information from websites, videos, and files that they find interesting. People share confidential information with great faith, and others have the same faith in the information shared. The surge in the reputation of online social networks, as well as the availability of a massive amount of data, allows them to be simple targets for their opponents. These goals primarily include stealing personal information from individual users without their knowledge [7].

1.1 Definition of Fake News

Counterfeit news definition is made of two sections: validness and plan. Genuineness implies that phony news content bogus data can be confirmed all things considered, which implies that paranoid fear is excluded from counterfeit information as there are hard to be refuted valid or as a rule [8]. The subsequent part, purpose, implies that the bogus data have been composed to misdirect the peruse.

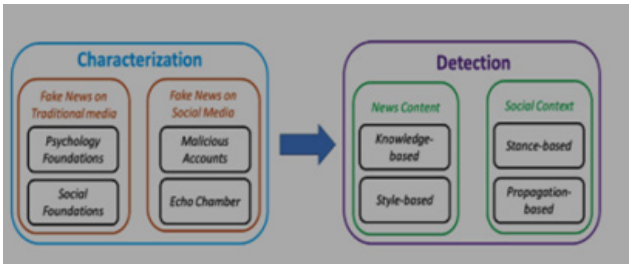

Figure 1. Fake news on social network: characterization to detection flow diagram [6]

1.2 The Amharic Linguistic

"Amharic is an Ethiopian Semitic language, which is a subgrouping inside the Semitic part of the Afroasiatic dialects. It is communicated in as a first language by the Amharas, and furthermore fills in as a most widely used language for different populaces living in significant urban areas and towns of Ethiopia

1.2.1 The Representation of Amharic Characters

Geez characters, which date back to the 4th century AD, are used in Amharic. The original Geez writing forms contained only consonants, while subsequent variants of the symbols represented consonant-vowel phoneme pairs. Amharic writing, like Geez, uses a combination of vowels and consonants. The Amharic language uses seven vowels, each in seven different forms, which correspond to seven vowel sounds. There are 33 basic characters, each of which represents a consonant and a vowel at the same time, making the Amharic script pronounced in syllables [11].

1.2.2 Punctuation Marks of Amharic Language

Recognizing accentuation marks is crucial to know word boundary for regular language handling. As indicated by Tewodros Hailemeskel (2003), the accentuation marks in Amharic are around ten however not many of them utilized in the PC composing framework. "HuletNeteb" (":") word separator and "AratNeteb" ("::") sentence separator are the significant accentuation marks. However, space is generally utilized rather than HuletNeteb (":"), especially in PC composing [5].

1.2.3 Amharic Number

Bender et al. (1976) expressed that Amharic numbers are gotten from Greek letters. Furthermore, some of them were adjusted to seem as though Amharic person [1]. They are addressed by a solitary letter, and every one of them has a level stroke above and underneath, as displayed in the accompanying table [16].

Table 1. Amharic Number System

፩	፪	፫	፬	፭	፮	፯	፰	፱	፲	፳	፴	፵	፶	፷	፸	፹	፺	፻
1	2	3	4	5	6	7	8	9	10	20	30	40	50	60	70	80	90	100	1000

1.3 Difficulties in the Amharic Penmanship Scheme

The Amharic writing plan includes a few difficulties that are difficult to cope with while dealing with Amharic text. One of these challenges is the repeating of characters in Amharic, which requires more than one person to address 19 the same sound. In Ge'ez, the various structures are significant, but in Amharic, there is no unambiguous concept that explains why they exist [5]. The issue of the same sound with different characters is evident with the central characters, as well as in similar character requests. Those are, ሀ and ሃ; ሐ and ሓ; ኻ and ኽ; አ and ኣ A word that is framed using this person has a comparable meaning.

Table 2. Amharic Penmanship Scheme Sound

Character	The other style of character	Others
ሀ (he)	ሐ and ኻ	ሃ ፡ ሓ ፧ ኽ
ሠ (se)	ሰ	
አ (a)	ዐ	ኣ ፧ ዓ
ጸ (tse)	ፀ	

2. Literature Review

A broad writing audit would be directed on programmed counterfeit news identification, language innovation, and Amharic language to know the topic exhaustively. In this paper, multiple researches were provided with diverse ways to detect false news, and they proposed a classification system for detecting fake news on social media. Following Table 3 shows works done by

Table 3. Summary of related work

Researcher	Topics	Extract feature	Accuracy and result
WorkuKelemework(2013)	Automatic Amharic text news classification: a neural networks approach	Automatic classification of Amharic news using neural network method	In order to compare the effectiveness of the Vector Quantization (VQ) algorithm, a TF*IDF weighting method was used. The accuracy for this experiment was an average of 71.96%.
Nicole O'Brien (2018)	Machine Learning for Detection of Fake News	Machine learning	95.8 %.
Samir Bajaj (2017)	Fake News Detection Using Deep Learning	Using Deep Learning	0.7
Yi-Ju Lu, Cheng-Te Li (2020)	Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media	Using GCAN	16%
Md ZobaerHossainy|, Md Ashraful Rahmany|, Md Saiful Islam|, Sudipta Kar (2020)	BanFakeNews: A Dataset for Detecting Fake News in Bangla	Using traditional Linguistic features and neural network-based methods.	90%
Ajeet Ram Pathaka, Aditee Mahajana, Keshav Singha, Aishwarya Patila, Anusha Naira (2019)	Analysis of Techniques for Rumour Detection in Social Media	Use comparative analysis of the state-of-the-art	Accuracy 82%
Josef Kapusta, Petr Hajek, Michal Munk, Lubomir Benko (2020)	Comparison of fake and real news based on morphological analysis	Using morphological analysis	to improve the existing web services for fake news classification
MontherAldwairi, Ali Alwahedi (2018)	Detecting Fake News in Social Media Networks	Use both Bayes Net and Naïve Bayes	94.4%

different researchers on these topics [2,3,7,8,12,13].

3. Methodologies

The algorithms we create to classify bogus news written in Amharic are described in this section. Traditional lang uage features as well as several neural network-based models are used in our method.

3.1 Dataset Construction

The goal of this research is to identify Amharic bogus news on social media. In addition, a new Amharic false news dataset must be created.

3.2 Traditional Linguistic Features

3.2.1 Lexical Characteristics

We extract word n-grams (n=1, 2, 3) and character n-grams (n=3, 4, 5) from news articles due to strong performance in several text categorization tasks. We use the phrase frequency-inverse document frequency as a weighting system (TFIDF).

3.2.2 Grammatical Structures

The Grammatical features of texts are frequently important for deciphering specific examples in reports that assist order issues in the end.

3.2.3 Semantic Features

By providing semantic data, dispersed depictions of word and sub-word tokens have demonstrated adequacy in text order difficulties. As a result, we experiment with several approaches to pre-prepared word inserting, such as addressing an article based on the mean and standard deviation of the vector representations of the words in it.

3.3 Punctuation and Metadata

We noticed a higher presence of some accentuation images like "!" in the phony news. Thusly, we utilize the accentuation recurrence as provisions. Moreover, we utilize some Meta data like the lengths of the feature and

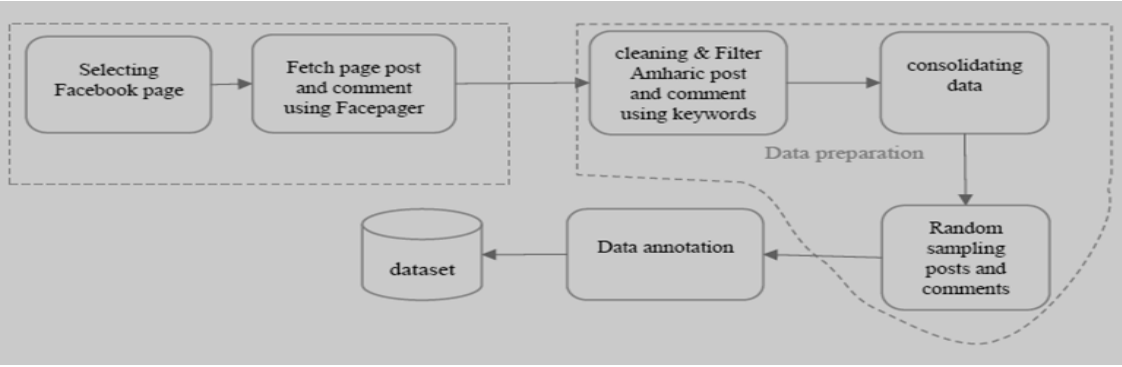

Figure 2. Technics for Building Amharic Fake News Dataset

Table 4. Some Category of Social Media Page

S.No.	Name of Category	Description for the Name
1	News media and Broadcasting pages	Conveying the news to the overall population or an objective public
2	Bloggers and Journalists pages	An individual who consistently composes material for a blog. Likewise, an individual composes for papers, magazines, or news sites or plans news to communicate.
3	Religious Group Pages and Media	Conveying strict news, strict lessons to the devotees of that religion or an objective gathering.
4	Official Pages of Political Parties, Politicians, and Government Offices	Gathering of individuals, frequently lawmakers with normal perspectives, meet up to challenge decisions and capture the power in the public authority. A gathering of legislators who have held the force of government
5	Public figure Artist and Authors pages	An individual or person of note who makes tune, painting or an essayist of a book, article, or archive.
6	Community Pages for Activists, General, or Special Interests	An individual who missions to achieve political or social change: additionally, pages advance diverse political, strict.

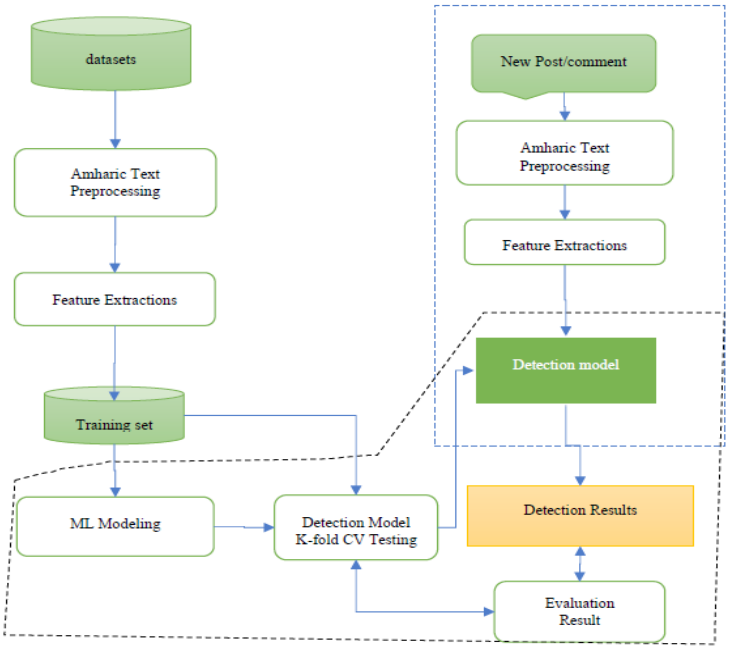

Figure 3. Amharic Fake News Detection Architecture

the assortment of news stories as provisions.

3.4 Neural Network Models

Neural organizations are exhibiting amazing execution in a wide scope of text characterization and age errands. Given a lot of preparing information, such models normally accomplish higher precision than etymological element based strategies. Consequently, we try different things with a few neural organization models that have been utilized as benchmark models in various text order errands.

3.5 Evaluation Metrics

To assess each model, different assessment measurements have been utilized. There are review, accuracy and f1-score. It is expected to utilize various measurements since they don't all record for similar qualities. For example, it is feasible to have a model with a review of 1 that act very terrible on the grounds that it just groups every one of the contributions to a similar single class.

Remind it Precision is expressed as [6].

$$Precision = \frac{TP}{TP + FP}$$

This indicates that depending on which classes are considered positive, we can have two distinct precisions. It is a ratio of appropriately categorized optimistic features to the total number of positive elements. When there are no false positives, it equals 1, but this does not mean that all of the positive elements have been correctly identified; there could be some erroneous negatives. The recall aids in the resolution of this issue.

It is express as

$$Recall = \frac{TP}{TP + FN}$$

The f1-score combines the recall and the precision. It is defined by

$$f1 - score = \frac{2 * precision * recall}{precision + recall}$$

You can likewise check out the weighted normal of these numbers. For instance, the worldwide review can be determined by duplicating the review for the two classes by the applicable class proportion [12].

4. Proposed System

A tool that can identify and delete fraudulent sites from the results presented to a user by a search engine or a social media news feed is part of the proposed system to address the issue of false news. The user can download the tool and then add it to the browser or program they use to get news feeds. Once it's up and running, the tool will use a variety of techniques, including those that look at a link's syntactic features, to determine whether it should be included in the search results [8].

5. Conclusions

Privacy concerns and the disclosure of private data

consume emerged as the greatest pressing concerns for social networking services. Users from all across the world use social networking services in various ways. The simplicity with which fake users and other dangerous users can contact these people and use the information stored in their accounts attracts them. This article proposed to foster an answer for counterfeit news via web-based media utilizing AI methods. The review attempted to create, execute and thinks about AI and text highlight extraction techniques explicitly for counterfeit news recognition for the Amharic language. The point further developed the current online media administrations for counterfeit news characterization.

References

[1] Fikremariam, Genet Mezemir, 2009. Automatic Stemming for Amharic Text: An Experiment Using Successor Variety Approach. pp. 1-79.

[2] Kelemework, Worku, 2013. Automatic Amharic text news classification: Aneural networks approach. Ethiop. J. Sci. & Technol. Vol. 6, pp. 127-137.

[3] Bajaj, Samir, 2017. Fake News Detection Using Deep Learning. Stanford University: cs 224n.

[4] Srinivas Rao Pulluri, Jayadev Gyani, Narsimha Gugulothu, 2017. A Comprehensive Model for Detecting Fake Profiles in Online Social Networks. International Journal of Advanced Research in Science and Engineering. Vol. 6, pp. 1-10.

[5] Ver´onica P´erez-Rosas, Bennett Kleinberg, Alexandra Lefevre, August 20-26, 2018. Automatic Detection of Fake News. Santa Fe, New Mexico, USA. s.n. 2018. Proceedings of the 27th International Conference on Computational Linguistics. pp. 3391-3401.

[7] Lorent, Simon. 2018-2019. Fake News Detection Using Machine. pp. 1-91.

[8] O'Brien, Nicole, 2018. Machine Learning for Detection of Fake News. pp. 1-56.

[9] Ajeet Ram Pathaka, Aditee Mahajana, Keshav Singha, Aishwarya Patila, Anusha Naira, 2019. Analysis of Techniques for Rumor Detection in Social Media. International Conference on Computational Intelligence and Data Science (ICCIDS 2019). pp. 1-11.

[10] Sachin Ingle, Satish Borade, Sagar Awasare, 2019. Detecting Fake User Accounts on. IJARIIE-ISSN(O)-2395-4396. pp. 927-931.

[11] Defar, Yonas Kenenisa. 2019. Hate Speech Detection for Amharic Language on Social Media Using Machine Learning Techniques. pp. 1-103.

[12] Lu, Y.J., Li, Ch.T., 2020. GCAN: Graph-aware Co-Attention Networks. arXiv, Vol. 1, pp. 1-10.

[13] Md Zobaer Hossainy, Md Ashraful Rahmany, Md Saiful Islam, Sudipta Kar. Bangla, BanFakeNews: A Dataset for Detecting Fake News in. 2020, University of Houston, Texas, USA. pp. 1-10.

[14] Arega, Kedir Lemma, 6, 2020. Social Media Fake Account Detection for Amharic Language. Global Scientific Journals. Vol. 8, pp. 604-614.

[15] Josef Kapusta, Petr Hajek, Michal Munk, Lubomir Benko, 2020. Comparison of fake and real news based on morphological analysis. ELSEVIER. pp. 1-9.

[16] Kai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang, Huan Liuy. Fake News Detection on Social Media:A Data Mining Perspective.