United International Journal for Research & Technology

Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

Lexicon-Stance Based Amharic Fake News Detection

Ibrahim N. Awol[1] and Sosina M. Gashaw[2]

[1,2]Addis Ababa Institute of Technology (AAiT)

Abstract— Due to the noisy nature of social media content, and the rapid propagation of fake news, the identification and detection of false information become a challenging problem. In recent years, several studies propose to use content based text representation approaches to automatically detect fake news as early as possible before being propagated into social Medias. However, fake news has a different stylistic nature of writing, and attempting to capture its various unique features may help us improve detection rather than focusing solely on text representation. In this study, we propose to hybrid stance-based features (Page score, headline to article similarity and headline to headline similarities) with the previous text representation lexicon based detection. To build the detection model, we used three machine learning algorithms: Logistic regression, Passive Aggressive and Decision tree. The proposed approach is evaluated using a newly collected Amharic fake news dataset from Facebook. Our experiment results show that the hybrid features (lexicon-stance) are capable of improving the previous lexicon based detection results up to 4.1% accuracy, 3% precision, 4% recall and 4% F1-score using passive aggressive algorithm. In addition, the hybrid feature improves the area under curve from 0.982 to 0.995 by reducing the false positive rate by 4% and improved the true positive rate by 4.4%. Furthermore, we found that Page score, out of the proposed stance features included, has contributed the most to the improvement of lexicon-based fake news detection.

Keywords— Content-based detection, Stance-based detection, Lexicon-based detection, text representation approach, Fake news.

I. INTRODUCTION

Obtaining news information has become considerably easier and more comfortable thanks to the Internet and social media. Because of the tremendous growth of information available online, distinguishing between real and misleading information is becoming increasingly difficult, culminating in the problem of fake news. Detecting fake news on social media can be a tedious task mainly for two reasons. First, it is extremely time consuming to validate suspicious news and to look for verified evidences. The second reason is the content of the posts from social media are normally very noisy without enough information to easily verify their credibility. As the issue is a worldwide and global challenge, various researches are being conducted. The majority of automatic fake news detection research, however, is focused on resourceful languages such as English; nonetheless, detecting false news in non-resourceful languages such as Amharic is still in its early stages.

Many researchers have worked on a fake news detection system that automatically determines whether any content stated in the article contains false information. In a broad sense, the forms of their research are carried out using a method that connects the linguistic pattern of news to deception and verifies deception by utilizing external knowledge. Previous studies that applied a content-based approach for fake news detection mostly focused on text representation (data representation) analysis, in which deception is identified by analyzing individual and semantic n-gram word frequency [9] [10] [2]. Lexicon-based detection is a text representation technique in which news articles are stored as a vector by counting the words in the news article in order to analyze which words are more frequently repeated on the false news article and which words are more frequently repeated on the authentic news article. This method is simple and quick for detecting fake news, but it does not take into account the news's various unique writing style. Fake news is written by the majority of people who are not journalists- that being said, the style of writing can vary [11]. It has a distinct writing style than actual news, thus it's up to the researchers to discover and study the differences. Several researchers looked into and extract the news's different stylistic and hand-crafted features using feature engineering mechanisms. Stance analysis is also another form of extracting news features that examines the news's similarity and difference or stance towards other news or parts of the news [12]. In this study, stance refers to how similar one news article is to another in terms of word frequency, or how similar one news headline is to its news body, or what the author's position is on fake news topics.

In this study we developed a hybrid lexicon-stance based approach, which combines the lexicon-based and stance-based analysis to better comprehend and analyze the stance feature effect on the identification of false news, as an extension to the prior text representation (lexicon based) approach, which ignored some unique news writing characteristics. We used Author (the person who posted the news), Headline (the title of the news), and Article (body of the news) as news attributes. To improve performance, the ’Article’ attribute is extracted from text representation analysis (lexicon based detection) and combined with the proposed stance-based features extracted from the ’Author’ and ’Headline’ attributes of the news. We proposed a hypothesis for generating a weight (score) for each author who posted the news for the author attribute. The idea is that people who have spread inaccurate information in the past are more likely to spread false information in the future, likewise users who have spread real information in the past are more likely to post authentic news in the future and we construct a grading system to account for this. The generated score for each authors shows the position or stance of the author towards fake news topics. The headline attribute of the news is extracted via known stance analysis, which examines the similarity of various news headlines and headline-article similarity of the news. Furthermore, we investigated the effect of incorporating Amharic stop words at the preprocessing stage on the performance of fake news detection, as researches have shown that stop words can have a positive impact on fake news detection [13] for resourceful languages such as English. As far as we are aware, there is no publicly available Amharic fake news dataset. So, in order to complete this research and encourage other researchers to study and investigate this topic further, we have compiled a new set of labeled Amharic news datasets.

II. RELATEDWORK

Despite the fact that there are various definitions and interpretations for the phrase ”fake news,” the most widely accepted meaning is fake news refers to a specific type of disinformation: It is false, it is intended to deceive people, and it does so by trying to look like real news [11] [15]. Two major findings can be noted using the aforementioned definition. First, there is the element of news that contains misleading information, and then there is the part of news that is structured to deceive consumers. According to researches fake news can be in different forms and types, the most common include [16]: satire, parody, fabrication, image manipulation, advertising, and propaganda. Satire: which is the style of writing that exposes real-world individuals or organizations in a humorous style usually by the treatment of irony. Parody: focuses on the ludicrousness of an affair and highlights them by producing untrue news stories instead of stating comments in a humorous oriented style. Fabrication: in this case the author or producer of an item is often intentionally trying to misinform the interested individuals. Image Manipulation: refers to the manipulation of an image either on smaller or greater scale to confuse and mislead people. Advertising: false information is formed in order to characterize or promote advertising materials and it is usually for financial gain. Propaganda: it is a way of deliberately affecting public opinion and consciousness for the government advantage or particular organization.

Fake news detection evaluates the truth value of a news piece, it tries to identifies whether an opinion claim in the article contain fake content or not. Even if there are many expert based and crowd sourced manual fact-checking websites nevertheless, manual fact-checking does not scale well with the volume of newly created information, especially on social media. As a result the research got an attention from different researchers on automatically identifying fake news and various works have been done. Fake news detection require two important procedures which are feature extraction and model construction. In feature extraction, we capture the differentiable characteristics of news pieces to construct effective representations; based on these representations, we can construct various models to learn and transform the features into a predicted label. When we deal with capturing and analyzing the various characteristics of news and construct its representation, currently fake news detection approach can be divided into three main categories. Content (linguistic cue), social context based, and propagation based approaches.

Content-based approach: which are used in the majority of works on fake news detection, rely on linguistic (lexical and syntactical) features that can capture deceptive cues or writing styles [8]. In this approach the content of deceptive messages is extracted and analyzed to associate language patterns with deception [4] [5] [17] [6]. These features can be analyzed based on text representation, stance analysis, syntax, discourse analysis and other stylistic hand crafted features. The advantage of this approach is not requiring external data or knowledge, dependent on the dataset content only and to detect bogus news at an early stage before it is spread on social media. The drawback of this method is the performance of each used linguistic feature in terms of detection differs from language to language.

Network-based (social context) approach: Unlike content based approach network based are not rely on deceptive language and leakage cues to predict deception. This method used network properties and behavior to analyze news misinformation. crowdsourcing was used to identify the fake news which network information, such as message metadata or structured external knowledge network queries can be harnessed to provide aggregate deception measures [7] [14]. This method is crucial for the future of fact-checking approaches. With the help of outside sources, news items can be fact-checked by assigning a"truth value" based on the context. However the approach requires an existing body of collective human knowledge to assess the truth of new statements.

Propagation-based approach: This method aims to detect fake news by exploring how news propagates on social networks. Researchers believed that fake news has a different propagation style on the social media compared to real news and analyzing this behavior could lead to detecting misinformation on the social network. The disadvantage of this approach is that it detects fake news after it has spread on social media, and without previous information, the method does not work at all.

In the content-based fake news detection approach, features can be extracted based on text representation techniques, syntax, discourse analysis, stance analysis or hand-crafted features analysis.

A. Data (text) Representations techniques

Deception is usually recognized based on data representation (lexicon) analysis in content-based approaches, according to research. The frequency of individual and "n-grams" words is used to detect fraud in data format. Researchers employ a variety of text representation methods to represent and extract data. The term frequency-inverse document frequency, bag of words, and word embedding are the three most used text representation techniques. The simplest yet efficient enough feature extraction method in text categorization is the bag of words (BOW), which just considers whether or not a recognized word appears in a document [19] [4] [1]. The word frequency can be done using an N-gram, which is a contiguous sequence of n items from a given sample of text or voice in order to grasp the semantic aspect of the text [17] [20]. Tf-idf stands for term frequency-inverse document frequency is one of the most prevalent term-weighting methods today and it is a numerical statistic that is designed to indicate how relevant a word is to a document in a collection or corpus. The lexicon features used in this research are derived from bag of words and tf-idf text representation model. These text representation models can be used for further analysis or for extracting hand crafted features [10] [25] [26]. The main drawback of lexicon features is lack of representing various unique fake news writing characteristics. Fake news is written by the majority of people who are not journalists- that being said, the style of writing can vary [11]. It has a distinct writing style than actual news, thus it's up to the researchers to discover and study the differences. The third text representation model is word embedding data representation which allows words with similar meanings to have comparable representations. Word embedding is a type of feature learning that seeks to map words from a real-number vocabulary into a low-dimensional space. It is a distributional representation of words with comparable meanings that machine learning models can understand [21].

Daraje Gurmesa develops Afaan Oromo fake news detection model by using lexicon TF-IDF data representation model [10]. The author collected 752 datasets and used passive aggressive as a classifier. TF-IDF with uni-gram words show better accuracy results of 97.2%. A deep neural network model for detecting fake news in Amharic was built using Amharic fastext word embedding as a data representation [1] [3]. A total of 12000 Amharic news articles were collected, with a detection accuracy of 93.92% found using a Convolutional Neural Network (CNN) [1]. Fantahun Gereme et al. [3], on the other hand, used an edited fastext word embedding representation to find 99.36% accuracy in a dataset of 6834 news articles. Another study employed TF-IDF as a data representation in conjunction with a passive aggressive classifier to detect Amharic fake news [2]. The authors gathered 961 real news articles and 457 false news articles and discovered a 96% accuracy rate. In this research our proposed approach (hybrid lexicon-stance features) is compared with a state-of-the-art text representation lexicon-based approaches. For comparison different works [10], [17], [2], [24], [22], [26] are chosen. The authors used lexicon (text representation) model to detect and identify fake news.

B. Stance based fake news detection

Researchers believe that liars and truth tellers communicate in distinct ways, so they looked at how rumors are written to see whether they could discern the difference. In this type of detection mechanism, the main focus is on feature engineering. They will fine-tune the features or add new ones in order to increase detection accuracy. Stance analysis is one form of extracting news features that examines the news's similarity and difference or stance towards other news or parts of the news [12][23]. The stance features for this research, are derived from three assumptions. The first one is, if the newspaper headline is unconnected to their body, there is a strong likelihood that the news is phony. The other is, fake news have similar headline writing style and same for that of real news. If one news headline has similar writing style with various fake news headline stories in terms of word frequency, then there is strong likelihood that the news is fake. The third one is based on our proposed hypothesis, which state that people who have spread inaccurate information in the past are more likely to post false information in the future, likewise users who have spread real information in the past are more likely to post authentic news in the future and we construct a grading system to account for this. The generated score for each authors shows the position or stance of the author towards fake news topics.

In this study, stance refers to how similar one news headline is to another in terms of word frequency, or how similar one news headline is to its news body, or what the author's position is on fake news topics. To extract stance of headline towards its news body or other news headlines, cosine similarity score was constructed between the headline and article feature vectors in order to determine their resemblance. Tf-idf and count vectorizer feature extraction techniques were used to retrieve the vector. After that cosine similarity score is applied between the vectors to find the stance features. Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space.

C. Machine learning classifiers

The selection of appropriate machine learning algorithms for recognizing and classifying the collected features is another key aspect of false news identification. Fake news detection is often a binary classification problem, and researchers have employed a variety of classification algorithms to find the best match for categorization. The most extensively used techniques for detecting fake news are supervised learning and deep neural networks. Different researches employed classifiers such as Nave Bayes, Decision tree, Passive-Aggressive, Logistic regression, Random forest, XGBoost and support vector machine for feature engineering and flexibility purpose in detecting bogus news on social media [27] [28] [29]. Because fake news is an online-based problem, various researchers chose the passive-aggressive classifier, which is an online-learning algorithm in which the input data comes in sequential order and the machine learning model is updated step by step [10] [30] [31] as opposed to batch learning, where the entire training dataset is used at once. For semantic and context data representation, the most generally used models in text classification and generation issues are the LSTM network, which can efficiently capture sequential information, and CNN networks, which are effective in categorizing short and long texts in a variety of applications. [1] [32] [18] [33]. However, for small to medium number of datasets, for feature engineering and handcrafted feature works, most writers prefer to employ supervised learning classifiers rather than deep neural networks.

III. METHODOLOGY

From content-based approaches for Amharic fake news detection, the proposed methodology combines lexicon-based (text representation techniques) and stance-based features. Fig.1 depicts the steps taken to implement the proposed technique.

A. Data collection and Preparation

To the best of our knowledge, there is no publicly available fake news dataset for the Amharic language, so we gathered and compiled our own, with each news item containing an Author (the person who posted the news), a Headline (the news title) (optional), an Article (the news body), and a label of the news (1 for real and 0 for fake news). To acquire the dataset for Amharic false news identification, we targeted the most popular and mainstream trustworthy news Facebook sites for real news and pages that propagate satire and misleading context for fake news. Despite the fact that these real pages are assumed to give true news, each piece of information was fact checked by a professional journalist. We gathered fake news piece by piece from the ground up and had it fact-checked again by journalists.

After collecting the dataset, we have preprocessed the data which is a crucial phase in data analysis that includes a variety of techniques, including data transformation, cleaning, and reduction. This phase assists in cleaning up the acquired data and making it suitable for analysis. We have removed URL links, numbers, and punctuation, symbols, emoji's, entries with no value and non-Amharic terms, from the data

Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

collection. Following that, we employed tokenization, normalization and removal of stop words on the cleaned dataset. Stop words are typically deleted during the preprocessing stage in many NLP studies. However, when n-gram words are considered at the same time, some researchers find that stop words have a positive effect [13]. Stop words are often used terms that appear repeatedly in a text but provide relatively little information on their own. Amharic language has its own Stop words, we have analyzed whether incorporating Amharic stop words affects the model performance or not.

B. Fake news detection

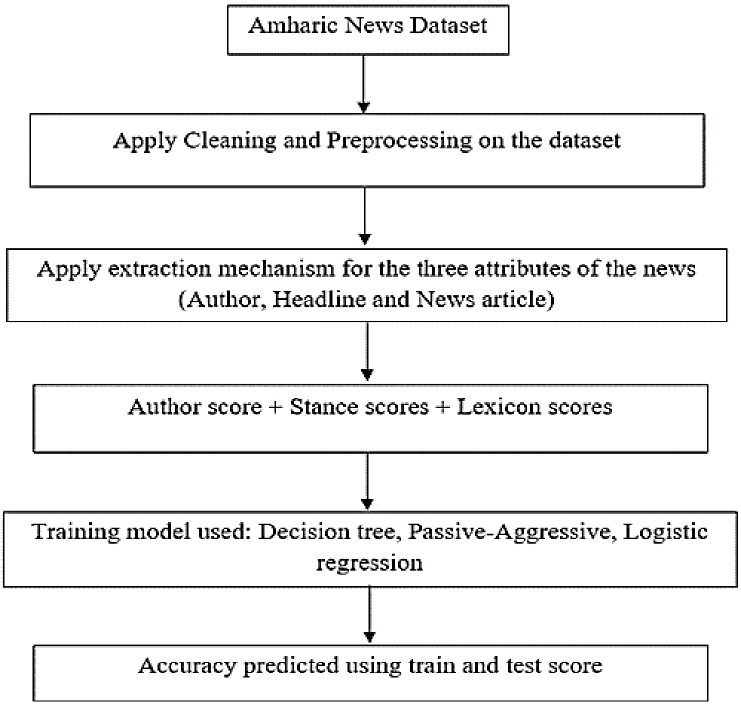

Figure 1: *General steps for the Proposed Method.*

We used countvectorizer (count of terms in vector/text), term frequency-inverse document frequency (Tf-idf) vectorizer, and N-gram (series of N-gram words taken from a given text) as feature extraction strategies to convert texts to weighted vectors. These text representation (lexicon-based) techniques are used for extracting the news attributes ’Article’ and ’Headline’. For the ‘Headline’ attribute, additional extraction is used to find stance features. The stance features extracted from ‘headline’ attribute are derived from two assumptions. First, if the title of the news is not similar to its news article, then there is strong likelihood that the news is fake. Second, if the news title is more similar to most of fake news headlines in terms of word frequency, then there is strong likelihood that the news is fake. So, in order to implement these assumptions cosine similarity score between the vector (various news headlines) and the cosine similarity score between the headline and article of the same news were used. For the ‘Author’ attribute, we generate weight based on our proposed assumption which state that ‘people who have spread false information are more likely to post fake content in the future likewise users who have spread real information in the past are more likely to post true news in the future’ hence we developed a grading system to account for this. The generated score for each author shows the stance of each author towards fake news topics. If the score is high, then that author has high probability of posting fake news in the future. If the score is low, then that author has low probability of posting fake news in the future. After extracting each news attributes and preparing both the lexicon and stance features, then we merged the feature vectors in order to improve the detection performance. Finally training model was built using three classification algorithms, namely Decision tree, Logistic regression, and Passive Aggressive (PA) and performance is evaluated and compared using evaluation performance metrics (accuracy, precision, recall and F1 score). Fig.2 shows the proposed lexicon-stance based features extraction techniques.

UIJRT ISSN: 2582-6832
United International Journal for Research & Technology
Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

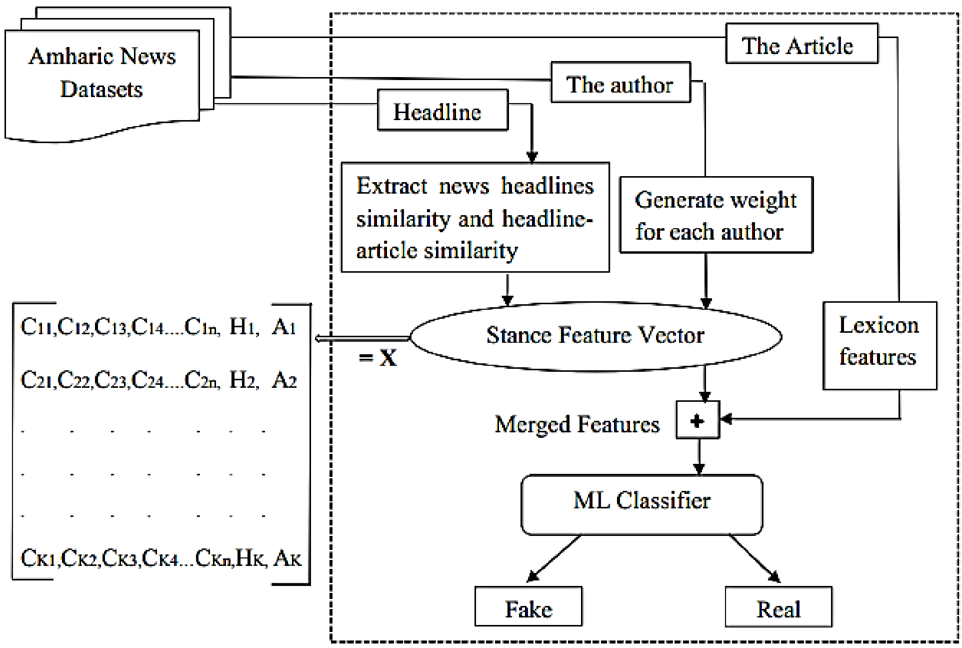

Figure 2: *Proposed lexicon-stance based features extraction techniques.*

Lexicon-based features

The text (data) representation models provide the lexical features (coming from known feature extraction techniques). We utilized Term Frequency-Inverse document frequency and CountVectorizer provided by scikit-learn library in Python. CountVectorizer is used to convert a collection of text documents to a vector of term/token counts while Tf-idf is used to determine word importance in a given article in the entire news dataset (1). The frequency of the words is rescaled by considering how frequently the words occur in all the news dataset. We used our dataset to compare these two feature extraction techniques and chose the best one for our model. The 'Article' attribute of the news is extracted based on this feature extraction technique and directly applied to the merged features without further analysis.

$$w_{ik} = tf_{ik} \times \log(N/n)$$

Where,

w_{ik} = Weight of term t_k in statement s_i

tf_{ik} = Frequency of term t_k in statement s_i

N = Number of statements in the news dataset

n = Number of statements where term t_k occurs

We took N-gram of words such as bi-gram and tri-gram when the article attribute are extracted in order to grasp the semantic part. Fig 3 shows word cloud representation of Amharic news article that give greater prominence to words that appear more frequently in the document.

Words appear mostly in real news article *Words appear mostly in fake news article*

Figure 3: *Word cloud representation of Amharic news article*

Stance-based features

The major goal of this study is to see how introducing stance-based elements affects the prior lexicon analysis. We used the 'author' and 'headline' attributes to examine the news's stance nature.

Page score: from our proposed hypothesis which state that 'people who have spread false information are more likely to post fake content in the future likewise users who have spread real information in the past are more likely to post true news in the future' hence we developed a grading system to account for this. The generated score for each author indicates the average stance of each author towards various fake news stories. This hypothesis is mathematically modeled using simple scoring mechanism: we can assign a probability score to each author based solely on the dataset (without any outside knowledge of the author) (2). For each author, we counted the number of fake news posted by him/her and counted the total number of news posted by him/her and calculated the probability of fake news posted to each author (taking ratio). Because our dataset is well-balanced, it is fair to use this ratio as the weight for the 'author' attribute. If the score is high, then that author has high probability of posting fake news in the future. If the score is low, then that author has low probability of posting fake news in the future. We verified our hypothesis by incorporating the score and observed improved fake news detection accuracy. Table 1 shows authors with their scores.

$$\text{Author i Score} = \frac{F_i}{R_i + F_i}$$

Where,

Fi = Count of fake news author i posted.

Ri = Count of real news author i posted.

Table 1: *Page scores for sample Authentic News and Fake news*

Pages which primarily posts real news	Weight (score)	Pages which primarily posts fake news	Weight (score)
page1	0	page2	0.666
page3	0.177	page4	0.751
page5	0	page6	1
page7	0	page8	0.647
page9	0.167	page10	0.833
page11	0	page12	0.769
page13	0.012	page14	0.928
page15	0.384	page16	0.916
page17	0.253	page18	0.568
page19	0.025	page20	0.72

Stance feature: In this instance, the cosine similarity score will be crucial. The attribute 'Headline' is extracted based on feature extraction techniques (CountVectorizer and Tf-idf) then further cosine similarity is applied between the vectors
(Between different news headlines and headline-article of the same news).

Mathematically, cosine similarity measures the cosine of the angle between two vectors projected in a multi-dimensional space. The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance (due to the size of the document), chances are they may still be oriented closer together. The smaller the angle, higher the cosine similarity. If the cosine score is 1, the two vectors are similar and if the cosine score is 0, then the two vectors are completely different.

Cosine similarity among news headlines: In order to grasp the degree of similarity across distinct news, a cosine similarity score is produced between the vectors of headline Hi and Hj (3). Each headline's vector similarity with others calculated and stored as another feature vector (each vector element contain cosine similarity score with others news).

$$(H_i, H_j) = \cos(\Theta) = \frac{\sum_{k=1}^{c} W_{ik} W_{jk}}{\| H_i \| \| H_j \|}$$

Where,

Hi and Hj are the corresponding weighted term vectors from headline attribute,

C is the count of terms in headline statement, and Wik of ith is defined as (1).

Cosine similarity between headline and article of the same news: We calculated the cosine similarity between headline and article of the same news to analyze the position of the headline towards the body of the news.

This feature is actually a column vector and we combined with the above stance feature vector.

United International Journal for Research & Technology
Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

IV. EXPERIMENTAL SETUP

A. Datasets

To acquire the dataset for Amharic false news identification, we targeted the most popular and mainstream trustworthy news Facebook sites for real news and pages that propagate satire and misleading context for fake news. Despite the fact that these real pages are assumed to give true news, each piece of information was fact checked by a professional journalist. We gathered fake news piece by piece from the ground up and had it fact-checked again by journalists. To gather legitimate news, Facepager, a Facebook Graph API data scraping application, was used. We have collected 3000 news (1500 actual news and 1500 fraudulent news) that each have an author (the person who wrote the news), a headline (the title of the news)(optional), an article (body of the news) and label of the news (1 for real news and 0 for fake news). To split the dataset to training and testing, we have used 80/20 (80 percent for the training and 20 percent for testing). Table 2 shows the dataset characteristics used for this research.

Table 2: *Dataset characteristics*

News group	Total number of articles	Total number of headlines	Extracted Source	Total number of Authors	Collection Period
Real news	1500	1350	Facebook		11/10/2020-4/9/21
Fake news	1500	1100		50	

B. Preprocessing

We have removed URL links, numerals, and punctuation, as well as entries with no value and non-Amharic terms, from the data collection. We employed normalization (different characters with the same sound but written in various ways) and stop word removal after cleaning the datasets. Amharic language has its own stop words, as an additional for this research, we have analyzed whether incorporating Amharic stop words affects the model performance or not.

C. Machine learning classifer

We compute the classification of the labeled dataset using a series of machine learning algorithms: logistic regression, Decision tree and Passive Aggressive classifier. Since we used hand-crafted features, there was no need to include a neural network model in the comparison since it would only associate weights with the features, rather than find new ones.

V. RESULT AND ANALYSIS

This section presents the details of each experiment result. In order to make the results clear and easy to follow, we present each set of experiments with its associated research question.

RQ1. What effect does adding stance-based features have on Amharic lexicon-based fake news detection?

Using our prepared dataset, we first performed an experiment with lexicon-based features on a variety of supervised learning algorithms, primarily logistic regression, Passive Aggressive, and Decision Tree. Table 3 shows the result of the experiment.

Table 3: *Received result for lexicon features*

Feature extraction methods	N-Gram Words	Accuracy			Recall			Precision			F1-score		
		LR	PA	DT	LR	PA	DT	LR	PA	DT	LR	PA	DT
Count Vectorizer	Uni-gram	90.8	90.3	86.8	86	87	83	96	94	91	91	90	87
	Bi-gram	87.8	88.2	83.5	83	82	75	93	95	93	88	88	83
	Tri-gram	78.1	77.2	73.8	61	60	53	96	95	96	75	74	68
	Uni-Bi gram	92	91	86.3	86	88	83	97	94	90	92	91	86
TF-IDF Vectorizer	Uni-gram	93.2	93	85	92	91	83	95	95	88	93	93	85
	Bi-gram	89.3	90.2	81.6	87	89	73	92	92	90	90	90	81
	Tri-gram	80	70	73	65	93	51	95	65	96	77	76	66
	Uni-Bi gram	93	93.4	86	92	93	83	94	95	89	93	94	86

As we can see from the results on table 5.3, Tf-idf Vectorizer outperforms Count Vectorizer in most of the performance metrics with taking n-gram parameter for the three classifiers algorithms. Taking both Uni and Bi gram words at a time has also better performance result than taking a uni gram, bi gram or tri-gram words at a time. The Passive Aggressive classifier outperforms the other on all metrics in lexicon-based features except on ROC curve in figure 4 which has equal AUC (area under curve) of 0.982 with Logistic regression. The Passive

Aggressive classifier has an accuracy of 93.4%, 93% recall, 95% precision and 94% f1-score results taking Tf-idf and uni-bi gram words feature at a time. Decision tree classifier has low performance compared to others on all metrics using Tf-idf and uni-bi gram words vectorizer. It has an accuracy of 86%, recall of 83%, precision of 89% and 86% f1-score.

Table 4 shows different metrics result for each stance features for the three classifiers. As we can see that the Page score has better performance on overall metrics than the other stance features. 92% of accuracy and F1-score is found when the Page score as a single feature is given to the classifiers and highest recall of 94% using Passive aggressive classifier is observed. Looking at the precision metrics decision tree gives a better result of 95%. From the proposed stance features the headline-article cosine similarity features gives lowest metrics result compared to the other stance features. The headline-headline cosine similarity feature of different news gives better result compared to headline-article feature. 87.2% of accuracy using logistic regression, 91% of recall using passive aggressive, 90% of precision using decision tree and 87% of F1-score using logistic regression is observed in headline-headline feature.

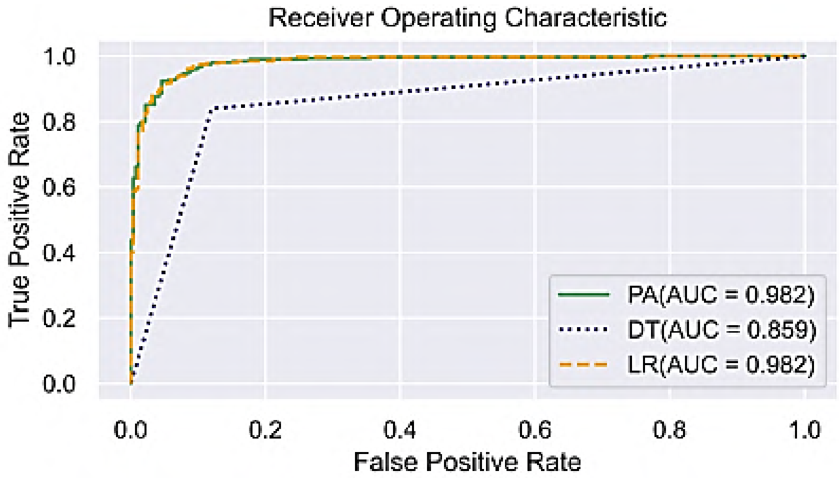

Figure 4: *ROC (Received operating characteristics) curve for three algorithms using lexicon based features only.*

Table 4: *Received result for stance features*

	Features Generated	Accuracy LR	Accuracy PA	Accuracy DT	Recall LR	Recall PA	Recall DT	Precision LR	Precision PA	Precision DT	F1-score LR	F1-score PA	F1-score DT
Stance Features	Headline -Article	79.6	83.3	80.6	72	88	77	87	82	85	79	85	81
	Headline-Headline	87.2	65	85.3	85	91	81	90	62	90	87	73	85
	Author Score	92	92	92	92	94	89	92	91	95	92	92	92
	Merged Stance	93	95	94.3	92	95	94	94	95	95	93	95	95

The proposed hybrid lexicon-stance based feature performance metrics result displays in table 5. Received result shows that there is an improvement of performance from the previous lexicon based approach. As we can see that the Passive aggressive classifier gives 97.5% accuracy (A), 98% of precision (P), 97% of recall (R) and 98% of F1-score result using lexicon-stance based features.

Table 5: *Received result of lexicon-stance features*

Lexicon-Stance Based Features	ML Classifiers	A	P	R	F1 score	AUC
	LR	97.3	97	97	97	0.994
	PA	97.5	98	97	98	0.995
	DT	95	96	94	95	0.951

Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

Looking at the AUC metrics in ROC curve of figure 5, the Passive aggressive classifier has got the highest area under curve of 0.995 which is better result compared to previous lexicon based feature result of 0.982. Overall performance is improved in the proposed lexicon-stance based feature for the classifiers.

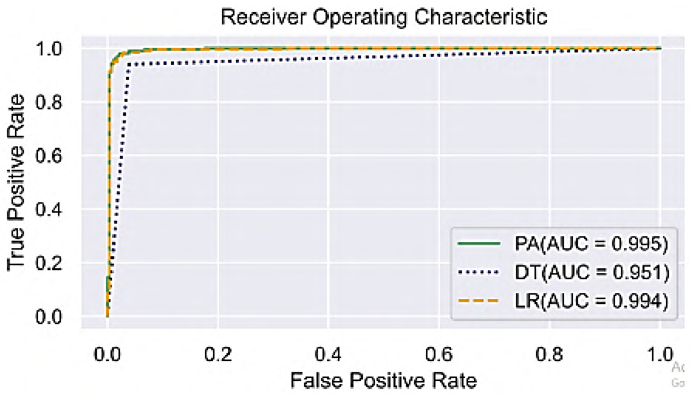

***Figure 5:** ROC curve of the three algorithms using lexicon-stance based features*

RQ2 Can the generated Page score (based on hypothesis) be used to improve Amharic fake news detection performance?

From the previous result we have observed that the proposed hypothesis for the author attribute which mathematically modeled as equation (2) and given to the classifiers as a single feature yields 92% of accuracy. The generated score shows good performance result in detecting fake news. To see the effect of this single feature on lexicon based Amharic fake news detection and to answer this research question, we have merged the Page score feature with lexicon based feature. Table 6 shows the performance metrics result of three classifiers using combined feature of Page score and lexicon feature. It can be observed that there is improved performance change in all metrics when including the Page score as a feature. 96.6% of accuracy result found using logistic regression classifier which shows improved performance by 3.6% from the previous lexicon based feature. The passive aggressive classifier improves the accuracy performance from 93.4% to 96% using the merged features. Highest AUC of 0.997 is found using merged Lexicon-Page score features on passive aggressive classifier compared to the other combined features.

***Table 6:** Result found using lexicon-Page score*

Lexicon-Author score features	ML classifiers	A	P	R	F1 score	AUC
	LR	96.6	97	97	97	0.996
	PA	96	97	97	97	0.997
	DT	93.3	94	94	94	0.933

RQ3. Does the use of stop words affect the performance of Amharic fake news detection?

In section 3.4 we have seen that there are researches that shows incorporating stop words at the preprocessing improves the fake news detection performance.

Amharic has its own stop words and we have done an experiment to see the effect of this stop words on the Amharic fake news detection performance.

Table 7 and 8 shows the experimental output when stop words are removed and stop words are included respectively. As can be seen from the results, there isn't much of a difference between the two.

However, there is a slight improvement of 0.5% accuracy in the passive aggressive algorithm. Since we used Tf-idf vectorization for text conversion, many stop words were already removed during the preprocessing stage, so there was no significant difference in the results.

We tried to figure out which stop words could account for some of the performance difference and discovered that words like ብለዋል, አይደለም, ገልፀዋል, አስታውቀዋል, ግን are the most common.

United International Journal for Research & Technology
Volume 04, Issue 05, 2023 | Open Access | ISSN: 2582-6832

Table 7: *Result found for lexicon-stance when stop words are removed.*

Lexicon-Stance Based Features	ML Classifiers	A	P	R	F1 score	AUC
	LR	97.3	97	97	97	0.994
	PA	97.5	98	97	98	0.995
	DT	95	96	94	95	0.951

Table 8: *Result found for lexicon-stance when stop words are included.*

Lexicon-Stance Based Features	ML classifiers	A	P	R	F1 score	AUC
	LR	97.3	97	98	97	0.994
	PA	98	98	97	98	0.995
	DT	94	95	95	95	0.943

Finally, we evaluated our system to determine how dependent it is on the author attribute by injecting untrained authors into the previous testing dataset and assigning fake news to pages that primarily post real news and real news to pages that primarily post fake news. Table 9 shows that the passive aggressive classifier has decreased by 2% and the logistic regression algorithm has decreased by 0.8%. When compared to other algorithms, decision trees have a higher reduction in change and are more dependent on the author characteristic. However, we must keep in mind that real-world datasets are not the same as the one we utilized for this experiment. We purposefully changed the author or the news for the sole purpose of testing.

Table 9: *Received result of our system using untrained author datasets.*

Testing our system using untrained Author	ML Algorithms	A	P	R	F1 score
	LR	96.5	97	95	96
	PA	96	97	94	96
	DT	92.1	93	91	92

VI. CONCLUSIONS

Fake news is a global problem that is not limited to a single language or location, hence various efforts should be made to identify and detect it as much as possible. In this study, we attempted to detect Amharic fake news using content-based approach without using external source or knowledge about the news. We propose to incorporate important news stance features into a state-of-the-art text representation lexicon based approach to enhance the performance of detection. To accomplish our study, we used a newly collected dataset as there has been no publicly available resource regarding the area that we want to explore and we present a new labeled Amharic fake news dataset. The stance features extracted from the news attributes are headline to article similarity, headline to headline similarity and author generated score. To assess the impact of the stance features on lexicon based detection, we conducted an experiment using models built with three commonly used machine learning algorithms: Passive-Aggressive, Logistic Regression and Decision tree. The experiment compares the performance of the lexicon based detection with the hybrid one containing both lexicon and stance features. Another experiment was done to see the generated Page score effect on lexicon based detection and to analyze the performance this single feature as individual. Finally, experiment conducted on incorporating Amharic stop words effect on the fake news detection performance.

The results show that the hybrid lexicon-stance detection improves the performance of lexicon based detection up to 4.1% accuracy, 3% precision, 4% recall and 4% F1-score using Passive-Aggressive algorithm. In addition, the hybrid feature detection improves the AUC from 0.982 to 0.995 by reducing the false positive rate by 4% and improved the true positive rate by 4.4%. Individually each stance features have good performance result and we observed that the generated Page score has better detection performance of 92% accuracy compared to the other two stance features. Lastly, integrating Amharic stop words at the preprocessing stage has no significance difference on the proposed fake news detection performance however there are some Amharic stop words which accounts for minor improvement on the detection.

ACKNOWLEDGMENT

I want to thank Dr. Sosina Mengistu for her consistent guidance and supervision. I would also like to express my gratitude to journalist Henok Dessalegn for his assistance and support in fact-checking the news and annotating the dataset that I used on this research.

REFERENCE

[1] N. Ermias, “Fake news detection for amharic language using deep learning,” Ph.D. dissertation, ASTU, 2021.

[2] T. Tewodros and R. Raghavendra, “Building a dataset for detecting fake news in amharic language,” International journal of advanced research in science, communication and technologyl, pp. 76–83, 2021.

[3] F. Gereme, W. Zhu, T. Ayall, and D. Alemu, “Combating fake news in “lowresource” languages: Amharic fake news detection accompanied by resource crafting,” Information, vol. 12, no. 1, p. 20, 2021.

[4] J. C. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised learning for fake news detection,” IEEE Intelligent Systems, vol. 34, no. 2, pp. 76–81, 2019.

[5] O. Ngada and B. Haskins, “Fake news detection using content-based features and machine learning,” in 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE, 2020, pp. 1–6.

[6] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar, “Fake news detection using sentiment analysis,” in 2019 twelfth international conference on contemporary computing (IC3). IEEE, 2019, pp. 1–5.

[7] K.-H. Kim and C.-S. Jeong, “Fake news detection system using article abstraction,” in 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2019, pp. 209–212.

[8] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, “Fake news detection on social media using geometric deep learning,” arXiv preprint arXiv:1902.06673, 2019.

[9] S. A. Alameri and M. Mohd, “Comparison of fake news detection using machine learning and deep learning techniques,” in 2021 3rd International Cyber Resilience Conference (CRC). IEEE, 2021, pp. 1–6.

[10] D. Gurmessa, “Afaan oromo fake news detection using natural language processing and passive-aggressive,” 2020.

[11] S. B. Parikh and P. K. Atrey, “Media-rich fake news detection: A survey,” in 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, 2018, pp. 436–441.

[12] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, “Fake news stance detection using deep learning architecture (cnn-lstm),” IEEE Access, vol. 8, pp. 156 695–156 706, 2020.

[13] I. Wijers, “Don’t believe everything you read: Identifying fake news using an author attribution framework.

[14] G. Jardaneh, H. Abdelhaq, M. Buzz, and D. Johnson, “Classifying arabic tweets based on credibility using content and user features,” in 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). IEEE, 2019, pp. 596–601.

[15] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD explorations newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[16] V. Tsarapatsanis, “Fake news detection,” 2020.

[17] H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer, 2017, pp. 127–138.

[18] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, “Banfakenews: A dataset for detecting fake news in bangla,” arXiv preprint arXiv:2004.08789, 2020.

[19] M. Granik and V. Mesyura, “Fake news detection using naive bayes classifier,” in 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON). IEEE, 2017, pp. 900–903.

[20] P. Bharadwaj and Z. Shao, “Fake news detection with semantic features and text mining,” International Journal on Natural Language Computing (IJNLC) Vol, vol. 8, 2019.

[21] S. H. Kong, L. M. Tan, K. H. Gan, and N. H. Samsudin, “Fake news detection using deep learning,” in 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, 2020, pp. 102–107.

[22] B. Al Asaad and M. Erascu, “A tool for fake news detection,” in 2018 20th International Symposium

on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). IEEE, 2018, pp. 379–386.

[23] X. Chai, "Fake news stance detection.".

[24] M. S. Mokhtar, Y. Y. Jusoh, N. Admodisastro, N. Pa, and A. Y. Amruddin, "Fakebuster: Fake news detection system using logistic regression technique in machine learning," International Journal of Engineering and Advanced Technology, vol. 9, no. 1, pp. 2407–2410, 2019.

[25] S. R. K. Indarapu, J. Komalla, D. R. Inugala, G. R. Kota, and A. Sanam, "Comparative analysis of machine learning algorithms to detect fake news," in 2021 3rd International Conference on Signal Processing and Communication (ICPSC). IEEE, 2021, pp. 591–594.

[26] N. F. Baarir and A. Djeffal, "Fake news detection using machine learning," in 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH). IEEE, 2021, pp. 125–130.

[27] K. [Stahl, "Fake news detection in social media," pp. 4–15, 2018.

[28] G. Bharath et al., "Detecting fake news using machine learning algorithms," in 2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021, pp. 1–5.

[29] Awan. et al., "Fake news data exploration and analytics," Electronics, vol. 10, no. 19, p. 2326, 2021.

[30] R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica, and A. Krishna, "Identification of fake news using machine learning," in 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE, 2020, pp. 1–6.

[31] J. Shaikh and R. Patil, "Fake news detection using machine learning," in 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC). IEEE, 2020, pp. 1–5.

[32] E. Qawasmeh, M. Tawalbeh, and M. Abdullah, "Automatic identification of fake news using deep learning," in 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2019, pp. 383–388.

[33] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," in 2018 13th International Conference on Computer Engineering and Systems (ICCES). IEEE, 2018, pp. 93–97.