ADDIS ABABA UNIVERSITY

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Lexicon-Stance Based Amharic Fake News Detection

A thesis submitted in fulfillment of the requirements for the degree of Master of Science in Computer Engineering

Author:

Ibrahim Neji

Supervisor:

Dr. Sosina Mengistu

May 16, 2022

Addis Ababa, Ethiopia

Lexicon-Stance Based Amharic Fake News Detection

By:Ibrahim Neji

Thesis Advisor	Examiner
Dr. Sosina Mengistu	Dr Surafel Lemma

Chairman of Department	Examiner
Dr. Bisrat Derebssa	Dr Bisrat Derebssa

Submitted in Partial Fulfillment of the Requirements for Masters of Science in Computer Engineering

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

May 16, 2022

Declaration

I the undersigned declared that this thesis work is my original work. The work has not been presented elsewhere for assessment and all sources of materials and literary works referred for the thesis have been fully acknowledged.

Ibrahim Neji

Date of Submission: May 16, 2022

Place: Addis Ababa,Ethiopia

Statement of Authorship

I hereby declare that I am the sole author of this master thesis and I have not used any sources other than those listed in the bibliography that is identified as reference. I further declare that I have not submitted this thesis at any other institution to obtain a degree.

Signed——————

Date———————-

Certificate

This is to certify that this thesis work entitled ”Lexicon-Stance Based Amharic Fake News Detection” which is submitted by Ibrahim Neji is carried out under my supervision and guidance for fulfilling the nature and standard required for partial fulfillment of the requirements of masters of Science in Computer Engineering. The work in this thesis has not been submitted elsewhere for a degree.

Supervisor: Dr. Sosina Mengistu
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering
Date————

Acknowledgment

First and foremost, I thank the almighty GOD for his blessing me with good health and well being which made me successful in my all academic pursuits. Next, I want to thank my advisor 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 in my research. Finally, I must express my very profound gratitude to my brothers for providing me with unfailing support and continuous encouragement throughout my years of study, especially I want to forward my sincere thanks to Mulugeta Neji and Tadesse Neji for the priceless contribution they made to me in all aspects of my life. This accomplishment would not have been possible without them may Allah grant you long and healthy life.

May God repay all of you for all the good things you did for me.

Abstract

Due to the noisy nature of social media content, and the rapid propagation of false information, the identification, and detection of fake news become a challenging problem. In recent years, several studies propose to use text representation techniques from content-based approaches to automatically detect fake news on the social media. However, fake news has a distinct writing pattern, and attempting to capture its distinguishing features may help us improve detection rather than focusing solely on text representation. In this study, we propose to combine the stance-based features (page score, headline to article similarity, and headline to headline similarities) with lexicon-based features from text representation techniques to enhance the detection performance. 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 by 4.1% accuracy, 3% precision, 4% recall, and 4% F1-score. 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 techniques, Fake news

Contents

List of Figures

List of Tables

Acronyms

ACC .... Accuracy

API .... Application Programming Interface

AUC .... Area under the ROC Curve

BOW .... Bag of Word

CBOW .... Continuous Bag-of-Words

CNN .... Convolutional neural network

DT .... Decision Tree

FN .... False Negative

FP .... False positive

GloVe .... Global Vectors for Word Representation

LR .... Logistic Regression

LSTM .... Long Short-Term Memory

NLTK .... Natural Language Toolkit

NLP .... Natural Language Processing

PA .... Passive Aggressive

PRE .... Precision

PCFG .... Probability Context Free Grammar

POS .... Parts of Speech

REC .... Recall

ROC .... Receiver Operator Characteristics

RQ .... Research Question

RST .... Rhetorical Structure Theory

TF-IDF .... Term frequency Inverse document frequency

TP .... True positive

TN .... True Negative

VSM .... Vector Space Model

XGBoost .... Extensive Gradient Boosting

Chapter 1

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. Fake news is increasingly becoming a threat to individuals, governments, and freedom of speech, news systems, and society as a whole. It disturbs the authenticity balance of the news system, creating real-life fears in the world's societies. There are numerous social science studies that have been done on the impact of fake news and how humans react to it. Various news organizations and websites have also published articles on how to combat and identify fake news. Appendix A depicts one of these websites, which demonstrates the news literacy project's common strategies for detecting 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 is focused on resourceful languages like English. However, detecting false news in non-resourceful languages like Amharic is still in its early stages.

Fake news and deception have grown in Ethiopia's media landscape, particularly online. This had caused major, tragic real-world consequences, heightened pre-existing tensions, and fueled violence and war. It has been putting groups at extremes of their ethnic, political, and religious identity, resulting in instability across the country's regions. Despite the fact that there are self-motivated journalists who are attempting to provide fake news alerts to newsrooms and their broader audiences via social media in Amharic language, little has been done from the government or academician side to detect Amharic fake news [1] [2] [3] [4].

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. Detecting fake news approaches can be categorized broadly in-to two [5], content-based and network-based approaches. Content-based approaches: the content of deceptive messages is extracted and analyzed to associate language patterns with deception [6] [7] [8] [9] [1]. Network Approaches: crowd sourcing 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 [10] [11]. Researchers recently attempted to use propagation-based approach which is more similar to network approach, to identify real and fake information based on their respective diffusion patterns on social media [12] [13].

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 [14] [15] [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 [16]. It has a distinct writing style than actual news, thus it's up to the researchers to discover and study the differences. Several researchers examine and identify the news's various unique features that can differentiate fake news from real news. Stance analysis is one of the approaches for extracting news features that examine the news's similarity and difference or stance towards other news or parts of the news [17]. In this study, stance refers to how similar one news article is to another in terms of word frequency, how similar one news headline is to its news body in terms of word frequency, or how much the page's position is towards fake news.

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 lexicon based from text representation techniques, which ignored from representing some of unique pattern news writing style. We used Source (the page 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 'Source' and 'Headline' attributes of the news. In this study, Source attribute represents the Facebook page that posts the news, and we used the terms interchangeably. We proposed a hypothesis for generating a weight (score) for each page for the Source attribute. The idea is that people who have spread inaccurate information in the past are more likely to post false information on their next post, likewise users who have spread real information in the past are more likely to post authentic news and we construct a grading system to account for this. The generated score for each page shows the position or stance of the page 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. The assumptions are that if a news headline is not similar to its article in terms of word frequency, the news is likely to be false, and if one news headline is more similar to various fake news headlines in terms of word frequency, the news is also likely to be fake. 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 [18] 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.

1.1 Problem Statement

Researchers have done various works on automatic detection of fake news using lexicon-based approach (text representation technique) in which deception is detected by analyzing individual and semantic n-gram word frequency in the news article [19], [2], [20], [14] [15]. Although this method is simple and quick for detecting fake news, it has the disadvantage of not taking into account unique writing pattern of the news. The question is, what kind of unique news feature can we extract solely from the content of the news which can differentiate fake news from real news? In this research we incorporated news stance detection analysis which analyze how much stance or position, source of the news and title of the news have towards news articles.

We employed the source (pages that post the news) and headline (title of the news) as additional news attributes. For extracting the source attribute we hypothesize that pages that have spread more false news in the past are usually highly susceptible of posting fake information, likewise pages that have spread real information in the past are more likely to post true news on the basis of empirical observations, and we therefore develop a scoring procedure. The generated score for each page indicate that what the pages's stance or position towards fake news. The headline of a news attribute, on the other hand, is extracted using a similarity approach, which calculates how similar one news headline is to its news body and how similar different news headlines are.

After extracting stance features using the proposed approach, we combined with the lexicon-based features to improve the fake news detection. Furthermore, we see from some researches that Stop words have a significant impact on changing the performance of fake news detection, particularly when the N-gram of words is taken into account. We do not know if Amharic stop words will have an impact on the detection of fake news. Generally,this research paper will in particular aim to answer the following **research questions**:

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

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

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

1.2 Objective

1.2.1 General Objective

The general objective of this research is to examine the impact of stance based features towards lexicon based Amharic fake news detection. The effect of incorporating stop words in the dataset on the detection of Amharic fake news.

1.2.2 Specific Objective

This research paper has the following specific objectives.

- Scraping and collecting public Facebook news posts and build a dataset.
- Formulating the criteria for labeling the dataset as fake and real news.
- Proposing a relationship between the news's page and false information
- Analyzing the effect of adding Stance based feature on lexicon-based Amharic fake news detection.
- Comparing the proposed hybrid approach with previously used lexicon-based techniques.
- Analyzing the effect of incorporating Amharic stop words on the fake news detection.

1.3 Research Methodology

In order to accomplish the objectives of the research, the following procedure will be followed

i) **Literature Review**: the first task was doing literature survey on the state-of-the-art fake news detection and related works.

ii) **Data Collection and Preparation**: 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. Following that, each piece of information is fact-checked by a professional journalist. In addition, preprocessing activities will be done on the collected dataset to make the data suitable for analysis.

iii) **Feature extraction and generate weight**: Each attributes of the news (headline and article of the news) will be extracted and score will be generated for source attribute based on the proposed hypothesis.

iv) **News classification**: the merged features will be given to selected classifiers to train and classify.

v) **Evaluation:** Experiment will be conducted to test the effect of the proposed stance-based features and use of stop words at the preprocessing stage for different classifier. The performance of the system will be evaluated in terms of precision, recall, accuracy, AUC, and F score.

1.4 Scope

This research includes the design and development of fake news detection for Amharic language using the newly collected Amharic news dataset from Facebook. The research incorporate the stance detection techniques to the previous text representation lexicon-based approach by analyzing and extracting source and headline of the news as an additional attributes. The limitation of this research arises from the collected dataset. Content based approach requires large amount of quality dataset with various extracted features to detect fake news accurately and the other is news posted by pseudo-pages (pages with the same name as an actual registered pages) are excluded from the dataset.

1.5 Contribution

The main contributions of this thesis to the research field of fake news detection are as follows:

- Show the effect of incorporating stance based features on lexicon-based Amharic fake news detection model.

- Generate a hypothesis and develop weighting mechanism for the news's source (Page) attribute.
- Show the effect of incorporating Amharic stop words on the fake news detection.

1.6 Thesis Organization

The rest of this document is organized as follows. The second chapter discusses definitions, types, theoretical backgrounds related to fake news and its context in Ethiopia, the detection mechanism, features extraction techniques and news classification algorithms used.

The third Chapter discusses previous research works related to fake new detection and state of the art approaches for fake news detection. Our proposed approach for fake news detection is elaborated in Chapter Four. The experimental setups, procedures, evaluation metrics, results, and discussion are discussed in Chapter Five. The final chapter discusses the conclusion and future research direction on Amharic fake news detection.

Chapter 2

Theoretical Background

This chapter is dedicated to provide theoretical backgrounds related to fake news and its context in Ethiopia, state of the art fake news detection approaches, features extraction techniques and news classification algorithms used.

2.1 Fake News

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 [16] [21]. 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 [22]: 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. Satire is a literary genre in which humor is used to comment on people or activities and their perceived vices, flaws, or faults. Humor is used in satire to emphasize a viewpoint or point about a topic or event. The majority of the time, satirists utilize wit to criticize or attack something they do not agree with.
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. Between parody and satire, there are many parallels. In both circumstances, comedy is the most important aspect in capturing the audience's attention. The difference is that while a parody targets and mimics the original work to make a point, a satire uses the original work to criticize something else entirely.

Fabrication: in this case the author or producer of an item is often intentionally trying to misinform the interested individuals. Fabrication items depend on actual affairs but often with political bias and involves news stories containing 100% false content created to deceive and perform harm. When it comes to news fabrication, there are two aspects. The author's financial incentive is the first dimension. To put it another way, when the amount of clicks on news increases, marketers get more interested. As a result, monetary incentives might be offered. The creation of bots that flood the news is the second part of news fabrication. As a result, the news item in question gains a lot of traction. Furthermore, the fabricated fake news has a similar or even identical substance and format to the real news.

Image Manipulation: This category refers to the manipulation of an image either on smaller or greater scale. More specifically, a simple violation of the photo could be the color alteration or removing minor parts. On the other hand, more significant adjustments could be the deletion or insertion of an individual into an image. Thus, manipulated images can be shared and confuse people or even worse mislead them.

Advertising and "Clickbait": False information is formed in order to characterize or promote advertising materials and it is usually for financial gain. "Clickbait" is a current phenomenon that is spreading over the internet at an increasing rate, with the goal of making money. It's the act of persuading a person to click on a link that leads to a non-relevant website page. As a result, the user is transferred to a different environment, which is frequently a commercial site or an unrelated web source. As a result, it is classified as fake news because it deceives people.

Propaganda: It is a way of deliberately affecting public opinion and consciousness for the government advantage or particular organization. It is more relates to the political scene and often may have some plausible truth to it; however, the information is paired with strong political bias intended to persuade those who read and, or, see the presented information.

Recently, videos have been edited in a new technical approach known as Deepfake, which uses deep learning algorithms to manipulate images and videos to make it appear as if something happened that did not. Creating fake videos used to be an expensive process that took a great deal of skill, time, and money. Today, all you need is a gaming laptop, an Internet connection, and a basic understanding of neural networks to make bogus videos. There are also apps that allow you to change faces in films with a simple click [23].

Fake News context in Ethiopia Prior to the popularity of Facebook in Ethiopia, people preferred to listen to radios and watch TV broadcasts from outside the country, such as Voice of America and Duetchevelle (broadcast in Amharic from Germany) radio stations. Today, many in Ethiopia are observed to keep up with public matters in their country and other world problems, tucked inside their communication instruments, engrossed in online. Social media has a usage for the purpose of trading, entertainment, education, and maintaining acquaintanceships. Beside that social media facilitated effective and instant information sharing possible in Ethiopia's politics and Facebook has become a preferred media outlet for presenting political views from government supporters, activists and opposition groups. Due to that, traditional media also began to prioritize its online presence and started to use the new distribution channels, mainly social networks.

There are plenty of information distributed on the social media and it is difficult to know the source where that information come from. As a result, most information that people access on Facebook is usually unproven and frequently assumed as real. These days misinformation generators are working hard to gain a better reputation in the minds of their followers in order to monetize their content through splashy headlines and captions aimed at being shared virally [1]. Especially, in a country where individuals have fought and killed one another due to political differences, ethnic rivalries, and/or religious extremism, it become easy to gain many followers and fueled the violence and conflict using fake content. Unfortunately, from the different types of social media activists in Ethiopia who have large number of followers are mostly unity centered activists, Ethno nationalism activists and religion based activists.

Researches on context of fake news in Ethiopia have assessed and found that Ethiopia's media ecosystem's weaknesses have made it vulnerable to fake news and misinformation. The driving factors were undoubtedly historical, including the weak state of private media in Ethiopia, the critical role of the Ethiopian diaspora in media ownership, and the proliferation and wild rise in popularity of entertainment-news page services Facebook and Twitter. Even if we all know that fake news has its own damaging effect in the country and its people, the government of Ethiopia response to the spread of fake news and misinformation is very little. In November 2019, the Ministers Council, under the Computer Crime proclamation was approved a bill drafted by the attorney-general to combat hate propaganda and misinformation being circulating. However, the law enforcing was not effective as researchers expected because of news literacy is not provided for the citizens at an early stage and the essence of strong professional journalism is not entertained well.

There are few fact checking pages and self-motivated journalists working on tackling and identifying fake news in Amharic language (officially working language of the country and spoken by more than half of the population). However, peoples have recently criticized the neutrality of the pages for some reasons like the pages are created for the government propaganda and others selects the type of news they are checking. Academically, few works have been done recently in detecting fake news technologically for Amharic and Afaan Oromo languages [15] [1] [2] [3] [4]. However, researches have limitations and beside that fake news detection is ongoing research which needs further analyzation for better detection. To the best of our knowledge still there is no publicly available annotated Amharic fake news dataset. Our objective is to prepare the datasets and fill the gap on the state of the art fake news detection researches.

2.2 Fake News Detection

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.

2.2.1 Content-based approach

Content-based approaches, 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 [11]. In this approach the content of deceptive messages is extracted and analyzed to associate language patterns with deception [6] [7] [19] [9]. The theory behind this approach is rumors or misinformation generators have a typical writing style of news which intentionally written to mislead the readers compared to real news. Various researches used different data representation model and hand crafted features to extract and investigate the news article. This approach tried to extract linguistic features based on the news dataset content only without knowing any external knowledge or social context about the news. The linguistic features can be analyzed based on data representation, stance analysis, syntax, discourse analysis and other stylistic hand crafted features. We have explained each feature analysis on part 3.2.

Fake news detection using content based (linguistic cue) method requires large and quality annotated dataset with many dimension of feature analyzation in order to accurately detect the false news. 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. It is one of the reason for many researchers to use this method to detect fake news. The disadvantage of this method is that it is language dependent. The performance of each used linguistic feature in terms of detection differs from language to language. The other drawback is in the detection performance, as fake news on social media is frequently written in a very noisy manner in order to deceive users. As a result, detecting disinformation only based on content analysis can be challenging, and experts urge that social context and news behavior be used as additional features.

2.2.2 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. The approach requires an existing body of collective human knowledge to assess the truth of new statements. 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 [10] [24]. It's 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. The two main methods that are being used under the Network Analysis approach are Linked Data and Social Network behavior.

Linked data: Searching existing knowledge networks, or publicly available structured data, is the basis for the method. So, external knowledge or link data are employed to judge the authenticity of the news. As a result of this, false "factual assertions" can be deceptive for particular types of data because they can be extracted and compared with other statements about the known world (this relates to facts proven to be true and or statements that are widely accepted). Social Network Behavior: On social media, identity verification is critical to the concept of trust and investigating the page or author to determine the veracity of the news is the first step in detecting fake news. In this technique external knowledge about the news generator is collected and analyzed. User features are extracted based on the profile like their amount of followers and followings, likes, posts, their time since creation as well as their activity rate. Other factors that can be considered include domain popularity, Internet site ages of the publishers, Web site registration behavior of the publishers, Domain rankings, positioning of message sources in the network, reputation of cites, trustworthiness, credibility, expertise, and the tendency to spread rumors.

2.2.3 Propagation-based approach

Propagation-based fake news detection 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. A new article goes through three basic stages: creation, publication on a news outlet(s), and dissemination on social media. Propagation based method rely on third stage to investigate the news. This approach has recently gained popularity, and it is more similar to the network approach in that it only classifies real and fake information based on their respective diffusion patterns on social media [12] [11]. It tries to explore relationships among entities such as news articles, publishers, users (spreaders) and user posts to construct propagation graph or tree using different nodes and edges. Generally represent posts (or users) as nodes of a graph and social connections (follower, following) or influence paths (repost, mention, comments, etc.) as edges. 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. Each method has its own set of pros and downsides. Researchers attempted to combine methodologies and combine features in order to effectively combat and detect fake news [25] [24].

2.3 Data (text) Representations techniques

The prepared text should be converted to numeric once it has been cleaned and preprocessed in order for the machine learning model to understand and train on it. The tf-idf, bag of words, and word embedding are the three most used text representation techniques. According to researches, deception is usually recognized based on text representation (lexicon-based) analysis in content-based approaches. In lexicon based analysis, the frequency of individual and ”n-grams” words are used to detect fraud in data format.

2.3.1 Bag of Words Model (BOW)

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 [26] [6] [1]. The other element is document word counting, which may 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 [19] [27]. Unigram, Bigram, Trigram and N-gram are different approaches for creating a vocabulary. Uni-gram is taking each word as a token and N-gram is taking N pairs of words as a token, N can be 1, 2, 3 or any other number. To implement bag of words with n-gram, CountVectorizer of sklearn is used in this research.

2.3.2 Term Frequency- Inverse Document Frequency (TF-IDF)

TF-IDF stands for term frequency-inverse document frequency, and it is one of the most prevalent term-weighting methods today. It is a numerical statistic that is designed to indicate how relevant a word is to a document in a collection or corpus and it is calculated as equation 2.1. Although the performance of data representation models varies and is dependent on the specific research, for most research works, the TF-IDF model outperforms the bag of words model.

$$\text{TF-IDF} = \frac{\text{Number of term occurrence}}{\text{Terms in text}} \times \log \frac{\text{Number of text in collection}}{\text{Number of texts where term occurs}} \tag{2.1}$$

2.3.3 Word embedding data representation

The most well-known and extensively used word representation strategy in NLP applications is word embedding, 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 a distributional representation of words with comparable meanings that machine learning models can understand [28]. The majority of studies use these strategies to capture semantic and syntactic information about words. However, it is not recommended for flexibility, extracting handcrafted features, or further text analysis.

Mainly, there are three well known word embedding algorithms: Glove, Fastext and Word2Vec. Glove: Global vector for word representation is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space. FastText is a library written in c++ for efficient representations of words and sentences classification in both supervised and unsupervised learning. FastText supports both continuous bag of words (CBOW) and Skip-gram models. Word2Vec: The word2vec algorithm learns word connections from a huge corpus of text using a neural network model. Once trained, a model like this can detect synonyms and recommend new words for a sentence. As the name suggests, word2vec associates each different word with a specific set of numbers known as a vector. There are two word2vec architectures for learning word embedding namely, Skip-Gram and a Continuous Bag of Words (CBOW). Skip-gram and CBOW are similar except CBOW predicts a context of single target word whereas Skip-gram (SG) predicts the context of a sentence using neighboring words.

After extracting news attributes using the above data representation models, some researchers analyze the extracted vectors further to prepare for their specific objectives. One of mathematical technique is cosine similarity calculation. Various researches utilizes cosine similarity as a technique in order to find relationship between two documents. For this research work, we have used for finding similarity between different news headlines and headline-article similarity of the same news.
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. In this case the two vectors are consists of elements which found on counting the words in each of the two document. The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance because of the size they could still have a smaller angle between them. Especially it is best choice for our research in which the news headline and article can have different word size. Smaller the angle, higher the similarity or the higher cosine value, higher similarity. 1 means the two vectors are similar and 0 means the two vectors or documents are completely different. Figure 2.1 indicates two vectors in a multidimensional space and equation 2.2 shows how cosine similarity of two vectors calculated. Cosine similarity between two vectors corresponds to their dot product divided by the product of their magnitudes.

$$\text{Cosine Similarity} = \cos(\Theta) = \frac{H \cdot B}{\parallel H \parallel \parallel B \parallel} = \frac{\sum_{i=1}^{n} H_i B_i}{\sqrt{\sum_{i=1}^{n} H_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2.2}$$

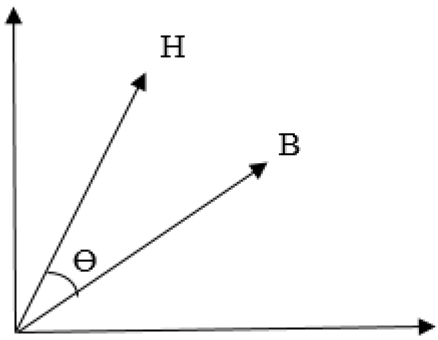

Figure 2.1: Two vectors H and B in multidimensional space

2.4 Different Machine Learning Algorithms

2.4.1 Passive-Aggressive

Passive Aggressive algorithm is a margin-based online learning algorithm for binary classification. It is one of the few online-learning algorithms (get a training example, update the classifier, and then throw away the example). In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. On each round, the algorithm investigates an instance and predicts the label to be either +1 or -1. After the completion of the prediction, the error is calculated, and the algorithm adjusts it in order to learn about the weight vector and improve its performance. Passive-Aggressive algorithms are somewhat similar to a Perceptron model, in the sense that they do not require a learning rate. However, they do include a regularization parameter. Passive aggressive classifiers have been also utilized in batch learning, because the technique is well-suited to binary classification issues and small batches of data can be classified in the initial phase.

To get a real algorithm, we need to figure out how to establish the weight vector w1 and what the update rule is for changing the weight vector at the conclusion of each round. In general, there are three types, each with its own updating rule, and we will concentrate on the simplest to explain the idea. The vector W1 is initialized to (0,...0) which, on round t, assigns the new weight vector Wt+1 as equation 2.3 the solution to the constrained optimization problem below.

$$W_{t+1} = \text{argmin}\frac{1}{2} \parallel w - w_t \parallel^2 \text{ s.t l}(w; (x_t, y_t)) = 0.\text{where w}\epsilon R^n \tag{2.3}$$

$Wt + 1$ is set to be the projection of Wt onto the half-space of vectors which attain a hinge-loss of zero. When the hingeloss is zero, the resulting algorithm is passive; that is, $Wt + 1 = Wt$ whenever $lt = 0$. On those rounds where the loss is positive, the algorithm aggressively forces Wt+1 to satisfy the constraint $l(Wt + 1; (Xt, Yt)) = 0$ regardless of the step-size required.

- Passive: If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
- Aggressive: If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it. Basically there are three main important parameters on Passive-Aggressive classifiers.
- C : This is the regularization parameter, and denotes the penalization the model will make on an incorrect prediction.
- Max-iter: The maximum number of iterations the model makes over the training data.
- Tol: The stopping criterion. If it is set to none, the model will stop when (loss > previous loss – tol). By default, it is set to 1e-3.

2.4.2 Logistic Regression

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. The method enables an algorithm in a machine learning application to classify incoming data based on previous data. The goal of logistic regression is to estimate event probabilities, which includes establishing a link between variables and the likelihood of specific outcomes. It is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary) and uses a logistic function to model a binary dependent variable.

Major Assumptions is taken in using binary logistic regression

- The dependent variable should be dichotomous in nature (yes or no).
- No appearance of extreme values in the data (There should be no outliers in the data).
- There should be no high correlations (multicollinearity) among the predictors.

The task of predicting the log odds of an occurrence lies at the heart of the logistic regression study. Logistic regression calculates a multiple linear regression function that is defined as equation 2.4:

$$\ln(\frac{P}{1-P}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{2.4}$$

Where P is defined as the probability that Y=1 and X is the independent input variables. Logistic regression is a popular method for binary classification problems in such research.

2.4.3 Decision Tree

Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. It is one of the most popular and practical supervised learning algorithms and created using an algorithm that determines multiple ways to segment a data set based on certain factors. Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. Different algorithms use different metrics for measuring best. These generally measure the homogeneity of the target variable within the subsets. Gini impurity, Information gain (based on the concept of entropy and information content from information theory), Variance reduction, and Measure of ”goodness” are some of them.

These metrics are applied to each candidate subset, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split. Each decision tree internal node gives a condition or ”test” on an attribute, and branching is based on the test conditions and results. Finally, the leaf node has a class label that is determined when all attributes have been computed. Predictive models that use supervised learning methods frequently use tree-based learning algorithms to achieve high accuracy. The biggest advantage of decision trees is that they make it very easy to interpret and visualize nonlinear data patterns. They also work very fast, especially for exploratory data analysis. The popular metrics which applied for each entity is information gain Gain(S, A) in equation 2.5: expected reduction in entropy due to sorting S on attribute A. Information gain is used to decide which feature to split on at each step in building the tree.

$$\text{Gain(S,A)} = \text{Entropy(S)} - \sum_{v \epsilon D_a} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \text{Where Entropy(S)} = -P^{+} \log 2^{P^{+}} - P^{-} \log 2^{P^{-}} \tag{2.5}$$

Generally, decision trees are used in various studies because they are easy to understand and interpret, require little data preparation, perform well with large datasets, have built-in feature selection (additional irrelevant features will be used less in subsequent runs), and decision trees can approximate any Boolean function.

Chapter 3

Literature Review

We will discuss literatures that use content (linguistic) approaches to detect fake news in this Chapter. Content-based methods are further classified based on how they extract features to detect fake news. This Chapter is divided into three sections. The first section focuses on content based approaches. The second section discusses the various categories of content-based approaches. The third section discusses the related works with the proposed approach and the final section discusses the news classification.

3.1 Content (Linguistic cue) Based Approaches

As fake news pieces are designed deliberately, rather than to report objective claims, for financial or political purposes, they often contain opinions and inflammatory words. Fake news is used to influence the consumer, and in order to do that, they often use a specific language in order to attract the readers. On the other hand, non-fake news will mostly stick to a different language register, being more formal [29]. To avoid getting detected, most liars use words strategically. Despite their best efforts, language "leakage" happens with certain linguistic features that are difficult to manage, such as pronoun, conjunction, and negative emotion word usage frequencies and patterns [13]. In linguistic approaches the content of deceptive messages is extracted and analyzed to associate language patterns with deception. Empirically focuses on analyzing and investigating the news articles' linguistic characteristics in content structure and style as a foundation for news credibility inference [30]. In this approach different linguistic features from the content of the dataset are extracted to assess the truthfulness of the news. Mostly content based approaches are chosen by researchers to detect fake news. The main reason is that in content based approach external data or knowledge is not required to identify the fake news and only analyzing the text pattern for deception detection is required.

Various researchers tried this approach by analyzing different features to reveal fake news. They utilize Known feature extraction techniques and hand crafted features to identify deceptive texts [7] [9] [30] [22]. Features extracted from content of the news can be lexicon based in which individual words or "n-grams" (multiword) frequencies are aggregated and analyzed to reveal cues of deception [15] [1] [31]. Deep Syntax-based is a content based approach where deeper language structures (syntax) is analyzed to predict instances of deception [32]. Rhetorical Structure and Discourse Analysis based: a description of discourse can be achieved through the Rhetorical Structure Theory (RST) analytic framework that identifies instances of rhetoric relations between linguistic elements [33].

Other stylistic and hand crafted features such as the subjectivity of the news, sentiment of the news, stance of the news towards the other, the number of causal and emotional words can be extracted and it is up to the researcher to identify unique features of fake news by capturing the manipulators in the news content writing style [6] [34] [9] [20].

3.2 Categories of Content Based Approaches

3.2.1 Text representation (lexicon) based techniques

According to research, deception is usually recognized based on text representation, lexicon analysis in content-based approaches. Typical lexicon features include character and word-level signals such as amount of unique and n-gram words with their frequency in the text. Researchers employ a variety of text representation (data representation) methods to represent and extract data, including word embedding [14] [28], which is a distributional representation of words with comparable meanings that machine learning models can understand. This data representation approach has been used in a variety of studies to examine the context and semantic aspects of words.The other is TF-IDF, or term frequency-inverse document frequency, which is one of the most widely used word-weighting approaches today. It is a numerical metric for determining how relevant a word is to a document in a corpus or collection. This data representation model can be used for further analysis or for extracting hand crafted features [15] [35] [36]. The bag of words (BOW), which simply considers whether or not a recognized word appears in a document, is the simplest yet effective feature extraction method in text categorization [26] [6] [1]. The other element is document word counting, which may 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 [19] [27]. TF-IDF and BOW are popular data representation techniques due to their flexibility, simplicity, and ability to accurately represent small to medium-sized datasets. Daraje Gurmessa Develops Afaan Oromo fake news detection model by using lexicon TF-IDF data representation model [15]. 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%.

To the best of our knowledge, we recently discovered four studies on the Amharic language [1] [2] [3] [4]. A deep neural network model for detecting fake news in Amharic was built using Amharic fastext word embedding as a data representation technique [1] [4]. Ermias collected a total of 12000 Amharic news articles and found 93.92% accuracy using text representation techniques and a Convolutional Neural Network (CNN) as a model classifier. [1]. Fantahun Gereme et al. [4], 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. Menbere hailu and Michael Melese used hybrid content and network analysis to detect Amharic fake news. They discovered that combining social context features with content-based features improves detection performance [3].

3.2.2 Syntax analysis based

In this approach the syntax structure of the news text is analyzed. The arrangement and grammatical structure of the news will be examined. Parts of speech tags (POS) which investigates the frequency of different tags (Adjective, Noun, Verb, Demonstrative, Adverb, Pronoun, Conjunction, Particle, Quantifier, Postposition), punctuations used, Deep syntax analysis based on Probability Context Free Grammar (PCFG) that obtain the rewrite rule of sentence in the news article are used for further extraction of the text [6] [25] [37]. Researchers believe that understanding syntax pattern of the text is beneficial in helping the classification problem.

Muhammed Zobaer Hossain et al. [25] develop a benchmark system with state of the art text representation techniques to detect Bangla fake news. The author explore different linguistic features and observed that higher presence of some punctuation symbols like '!' in the fake news. Normalized frequency of different POS tags was also used as a feature set for each document. Based on the research the evaluation of linear classifiers and neural network based models suggest that linear classifiers with traditional linguistic features can perform better than the neural network based models. It is also found that the use of punctuations in fake news is more frequent than authentic news, and most of the time fake news is found on the least popular sites. The research shows that linear classifiers with traditional linguistic features can perform better than the neural network based models.

3.2.3 Rhetorical Structure and Discourse Analysis based

Rhetorical Structure Theory is a theory of text organization which provides a framework for an analysis of text. The theory is based on the understanding that a text is not merely a string of clauses, but consists instead of hierarchically organized groups of clauses that stand in various relations to one another. Systematic differences between deceptive and truthful messages in terms of their coherence and structure has been combined with a Vector Space Model (VSM) that assesses each message's position in multi-dimensional RST space with respect to its distance to truth and deceptive centers [33] [38]. The author [38] shows that a discourse structure analysis as a significant method for automated deception detection and develop a novel discourse-based tools to alert information users to potential deception in computer-mediated texts. The research found that RST-VSM analysis produced greater levels of deception in truthful and deceptive stories than did the human judges.

3.2.4 Stance based and other hand crafted feature analysis

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. The previous approaches were unconcerned with the news's writing style, and instead of relying on ready-made extraction methods, a deeper research of the news's writing style could be beneficial for improving false news detection. Researchers employ a variety of fake news unique feature analyses, including the following:

Sentiment of the news: the sentiment of the news was evaluated [34] [9] [39] to improve detection performance by extracting the text's positivity and negativity polarity. Researchers believed that the classification of fake news may be determined by whether the writer's attitude toward a particular topic, such as a product, book, or leader, is positive, negative, or neutral. If the sentiment of the statement is negative. It is an indication that the news is most probably fake. Bhutani bhavika et al. [9] develop fake news detection model by incorporating sentiment of the news as a column feature and found performance enhancement on the fake news detection. Ajao hypothesize that there exists a relationship between a fake message or rumors and the sentiment of the texts posted online [34]. The author computed the emotional ratio of negative to positive words and included as an additional feature on the dataset. 3% improvement was found from the previous text only features.

News subjectivity: Researchers believed that typically, documents that aim at sharing factual and impartial information, such as trustful journalistic articles and scientific papers, tend to use a more objective language that does not rely much on presuppositions or sentimental and argumentative expressions. By contrast, documents aiming at convincing or persuading tend to use a more subjective language [40]. Fabricio Murai et al. [6] combine news subjectivity with additional features in a feature column. The subjectivity can be argumentation: where its dimension represents words and expressions that are related to a more argumentative discourse, presupposition: where its dimension encompasses terms that are related to a previous assumption of something, valuation: where its dimension expresses words related to the amount or intensification of something.

Stance detection. Various researchers have examined news stance towards other news or headline of the news stance towards the article. According to researchers, if the newspaper headline is unconnected to their body, there is a strong likelihood that the news is phony. Different authors developed model predicts headlines and bodies based on the class whether the news belongs to the unrelated, agree, disagree or discuss class. As features, they employed n-gram co-occurrence between the titles and articles, word embedding, hand-crafted features such as: word overlapping between the title and the article and the existence of highly polarized words from a lexicon, similarity features between headline and article [17] [41] [20] [42] [43]. Razan Masood proposed cosine similarity metrics which is used in this study as a technique to indicate the similarity between the body and headline of the news [41]. Specifically, the headline and body are converted to sparse vectors H and B, respectively, to indicate the frequency of each word. The cosine of the angle between these two vectors H and B in the high dimensional space is calculated as an indication of similarity. [9] Compute and incorporate cosine similarity of different news statement as an additional feature vector.

Various handcrafted unique features can be imagined, and it is up to the researchers to examine news behavior and implement feature engineering mechanisms in order to improve the fake news detection. By using network analysis approach, we can extract different hand crafted features from the social media like Credibility and Trustworthiness of the page, Engagement: consider number of likes, shares, and comments from Facebook users [3], Readability: the interaction of client towards the page, Domain analysis: investigating the domain or page and Popularity. Our research is limited to a content-based approach, and this work will be left to future researchers.

Hybrid features on content based approach

Many researchers who utilized a content-based approach to detect fake news used combined feature analysis to extract as many different news features as possible. [18] Proposed lexicon-syntax based extraction approach. The author include parts of speech tags in addition to data representation models. They found that longer articles, longer sentences, more ‘hear’ words, more ‘question marks’ and a more positive tone were indicative of real news according to their models. On the other hand, longer paragraphs, more stopwords, more question marks and exclamation marks and more URLs were indicative of fake news.

Bilal Ghanem et al. [43] proposed lexicon-stance combined features to detect fake news. They incorporated word or character n-grams overlapping score, bag-of words (BOW), word embedding, and latent semantic analysis features. Deception detection in Russian language based on lexicon-discourse analysis is done by Dina Pisarevskaya [44] which consists of 174 truthful and deceptive news stories in Russian. Sentiment analysis as a feature was incorporated in some researches [34] [9] [39] in combination with other features. Oluwaseun Ajao et al. [34] proposed syntax, sentiment and lexicon based feature fake news detection. They calculated the emotional ratio of tweets by dividing the number of negative emotional terms by the number of positive emotional words, as well as including syntax aspects (counts of uppercase words, exclamation marks, hashtags and quotation).

Xinyi Zhou et al. [37] proposed combination of lexicon-level, syntax-level, style-level, and discourse-level content fake news detection. They have investigated news behavior at different level and their proposed approach outperform the state-of-the-art. To capture writing style of the news, they have used sentiment of the news, subjectivity, headline-article stance detection, readability, linguistic enquiry word count (LIWC) which provides dimensions to evaluate informality of language like swear words, fillers, assents, nonfluencies and netspeak.

There have been studies on combining features from external network analysis approach and a content-based approach. Investigating news external characteristics in conjunction with its linguistic behavior can lead to improved detection performance [3] [45] [46] [47] [48]. In addition to examining the content of the news, researchers tried to analyze the page or website background such as the registration age, the number of followers, publisher detail, postdate, user location and engagement of other users with the page.

Sahil Gaonkar et al. [48] proposed a hybrid network with content features in order to maximize the accuracy of fake news detection. They included four news attributes: source, publisher, headline, and news article. The stance detection between the headline and article of the news is extracted from the content-based view, while the source and publisher attributes are extracted based on network analysis using external knowledge about the source to give the weight.

Effect of Stop words on Fake news detection

Pre-processing is a critical step in natural language processing, information retrieval, artificial intelligence, text classification, and other related tasks for a given document. The main step in the preprocessing task is to identify and remove stop words from a given document. Stop words are general words that have little meaning when used alone, are frequently occurring words in the document, and are useful for the structure of the language [49]. 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 [18].
Most researches use built in stop word list package to remove stop words from the document at the preprocessing stage. While other researches tried to filter and generate their own stop words instead of using built in package for their particular research [19]. Eliza Shoemaker [50] shows the effect of using different stop word list package on the fake news detection. The author use built in NLTK Stop words list package and spaCy Stop words list and found that accuracy difference between them. On the other hand there are researches [37] that uses count of stop word in the news and incorporate as a single feature in the model which believed to be fake news has fewer stop words than real news.

Iris Wijers. [18] Analyses the effect of incorporating stop words at the preprocessing stage. The experiment was carried out by removing and including NLTK stop words in order to determine the effect on detection performance. The preliminary results show that incorporating stop words improves accuracy. The author discovered that using support vector machine to train the model yielded 81.3% accuracy when stop words were removed and 92.2% accuracy when stop words were included. Without conducting an experiment, we will not know whether Amharic stop words have a positive or negative effect on the performance of fake news detection. Amharic language has its own stop words, Sileshi Girmaw [49] generated generalized Amharic stop words. We used stop words gathered by Sileshi Girmaw in combination with our Amharic stop word list from the dataset for this study.

3.3 News Classification

We have described how each news attribute can be represented at different levels of language using computational features. 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. Deep neural networks models are mostly utilized for large amount of dataset to classify texts extracted from known text representation techniques such as word embedding representation while supervised learning algorithms are mostly used for small to medium dataset and to classify texts extracted from text representation techniques like TF-IDF and CountVectorizer.

As classifiers perform best for machine learning settings they were initially designed for (i.e., no free lunch theorem), it is illogical to determine algorithms that perform best for fake news detection [51]. 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 [52] [53] [54]. Even though the passive-aggressive classifier is recommended for online-learning tasks since it is a margin-based algorithm and a suitable fit for binary classification problems, some researchers picked it for batch learning, where the complete training dataset is used at once [15] [55] [56] . 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] [57] [25] [58]. 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 [6][15][36][37].

Chapter 4

Proposed Methodology

From content-based approaches for Amharic fake news detection, the proposed methodology combines lexicon-based (text representation techniques) and stance-based features to achieve objectives discussed in the paper. Figure 4.1 shows the proposed general step followed for Amharic fake news detection.

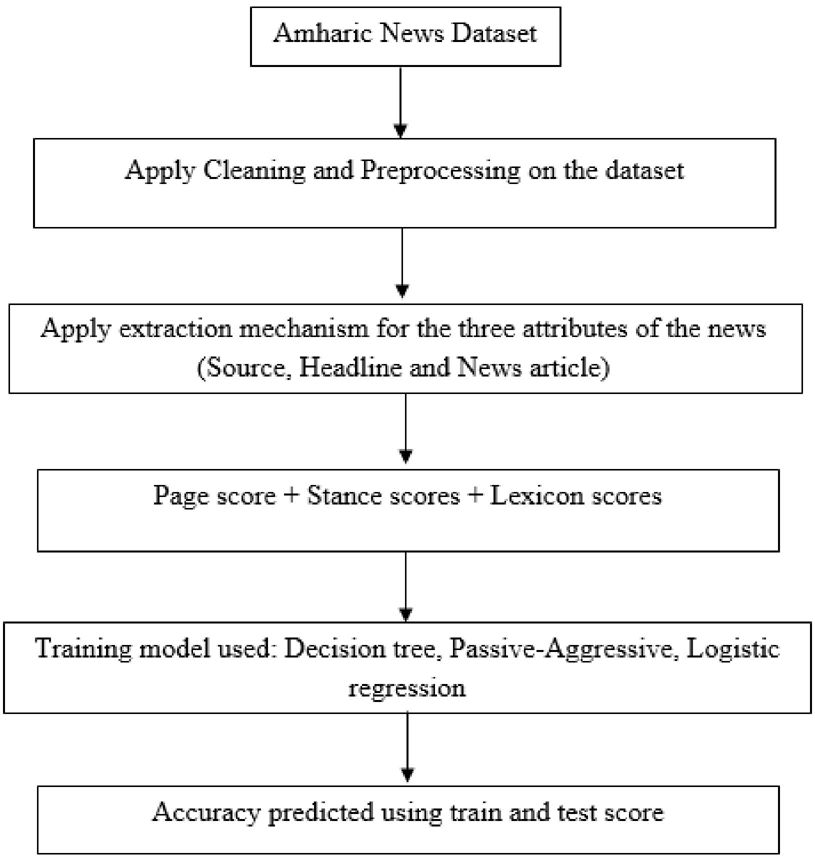

Figure 4.1: General Steps of the proposed approach

4.1 Data collection

Despite the fact that Amharic is Ethiopia's official working language and contains an enormous amount of literature in both digital and paper forms, it is still a low-resource language for NLP research [1]. 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 a source (the Facebook page that post the news), a Headline (the news title), an Article (the news body), and a label of the news (1 for real and 0 for fake news). For those of news with no headlines, since the vector element of their word count is null at the feature extraction stage, we just treat them as zeros when computing similarity.

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 is fact-checked by a professional journalist. We have collected propaganda, clickbait, fabricated and satire form of fake news. To gather legitimate news, Facepager, a Facebook Graph API data scraping application, was used. Facebook Graph API is an HTTP-based API used to programmatically query data and performs a wide variety of other tasks.
Generally based on the annotated dataset, we have formulated the criteria for labeling the news as fake and real as follows.
Real news if:

- A news is posted by several mainstream trustworthy news Facebook pages and is verified as real by the expert (journalist).
- A news is reported as real news by well-reputed fact-checkers.
- A news has additional videos and images to prove it as a real.

Fake news if:

- A news is reported as fake news by well-reputed fact-checkers and journalist.
- A news has a form of satire or click bait and the redirected link is fake.

- After some time, a news story aimed at a specific person or group was revealed to be false.

4.2 Data Preparation (Preprocessing)

Data preparation 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. After that, tokenization rule is applied to split sentences into words using a word tokenizer from NLTK. We employed normalization (different characters with the same sound but written in various ways should be changed to the same form). Both word and character level normalization are used in this thesis. Finally stop words are removed for uni gram words (single word at a time from the text) and incorporated for bi-gram (taking two words at a time from the text) and tri-gram words to see the effect on the model , as this is one of our research questions. For this study, 80 percent of the data was used for training and 20 percent of the data for testing the model. Figure 4.2 shows the preprocessing steps applied on the data.

Stop Word Removal

Stop words are typically deleted during the preprocessing stage in many NLP studies. However, when n-gram words are considered, some researchers find that stop words have a positive effect [18]. 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.

4.3 Fake news Detection

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

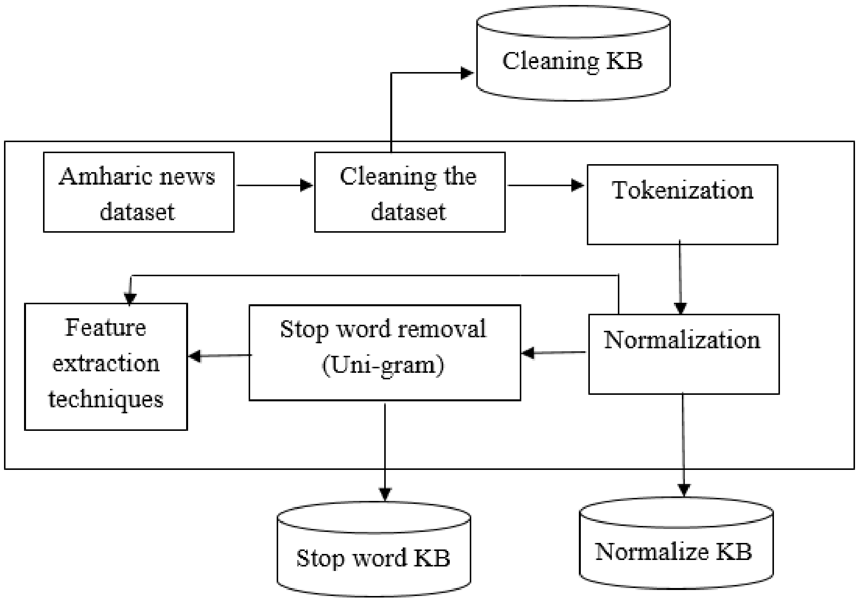

Figure 4.2: Preprocessing Steps of the proposed approach

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 detection features. The 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 to further investigate the 'Headline' attributes. For the 'Source' attribute, for this research source (which posted the news) are Facebook pages. We generate weight based on our hypothesis which state that 'people who have spread false information are more likely to post fake content likewise users who have spread real information in the past are more likely to post true news' hence we developed a grading system to account for this. The generated score for each page shows the stance of each page towards fake news topics. After extracting each news attributes and preparing features, then we merged the feature vectors in order to improve the detection performance. Figure 4.3 shows how the proposed lexicon-stance based features extracted.

Text representation (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

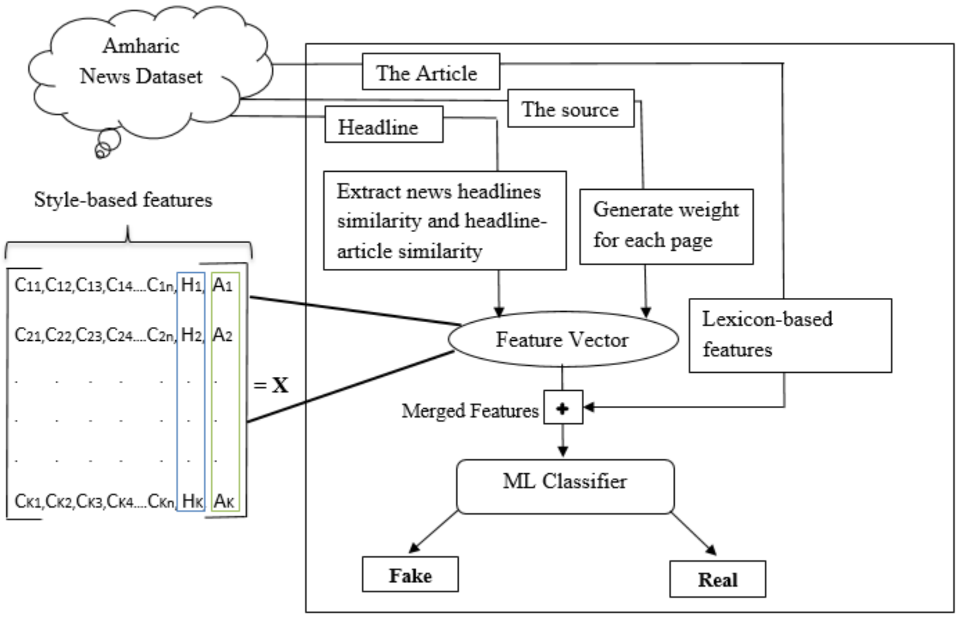

Figure 4.3: proposed lexicon-stance based features extraction techniques

Ckn= cosine similarity score of news headline k to news headline n
H= headline-article cosine similarity on the news
A= Generated Page score

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. The frequency of the words is rescaled by considering how frequently the words occur in all 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 the feature extraction techniques and directly applied to the merged features without further analysis.

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 'source' and 'headline' attributes to examine the news's stance nature.

Page score: from our proposed hypothesis which state that 'pages that have spread false information are more likely to post fake content likewise pages that have spread real information in the past are more likely to post true news on their next post' hence we

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

Table 4.1: a sample of pages and their scores

developed a grading system to account for this. The generated score for each page indicates the average stance or position of each page towards past fake news. This hypothesis is mathematically modeled using simple scoring mechanism: we can assign a probability score to each page based solely on the dataset (without any outside knowledge of the page) (4.1). For each page, we counted the number of fake news posted and counted the total number of news posted by the page and calculated the probability of fake news posted to each page (taking ratio). The ratio is calculated based on the dataset collected from the time period shown in Table 5.1. Since our dataset is well-balanced in the amount news article each page posted, it is fair to use this ratio as the weight for the 'source' attribute. If the score is high, then that page has high probability of posting fake news in the next post. If the score is low, then that page has low probability of posting fake news in the next post. We verified our hypothesis by incorporating the score and observed improved fake news detection accuracy. Table 4.1 shows pages with their scores. For this table, we purposefully rename Facebook pages to avoid exposing the real page in negative or positive manner.

$$\text{Page i Score} = \frac{F_i}{R_i + F_i} \tag{4.1}$$

Where,

Fi= Count of fake news page i posted.

Ri= Count of real news page i posted

Stance feature: In this instance, the cosine similarity metrics is used. The attribute

'Headline' is extracted based on known 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 news).

Cosine similarity between each news headline to the other: In order to grasp the degree of similarity across distinct news in the dataset, a cosine similarity score is produced between the vectors of headline Hi and Hj (4.2). Each headline's vector similarity with others calculated and stored as another feature vector (each vector element contain cosine similarity score between each news headline from the dataset). The assumption is if one news headline is more similar in terms of word frequency to various fake news headlines, then there is strong likelihood that the news is fake.

$$(H_i, H_j) = \cos(\Theta) = \frac{\sum_{k=1}^{c} W_{ik} W_{jk}}{\parallel H_i \parallel \parallel H_j \parallel} \tag{4.2}$$

Where,
Hi and Hj are the corresponding weighted term vectors for Headline i and Headline j.
Wik:The weighted term of news Headline i extracted from word count technique Tf-idf.
Wjk: The weighted term of news Headline j extracted from word count technique Tf-idf.
c is the count of terms in headline statement.

Cosine similarity between headline and article of the news: We calculated the cosine similarity between headline and article of the same news instead of examining similarity between different news to analyze the position of the headline towards the body of the news. The assumption is if one news headline is not similar to its news article in terms of word frequency, then strong likelihood that the news is fake. This feature is actually a column vector and we combined with the above stance feature vector. The combination is done horizontally merged with the other stance features.

News classification

We compute the classification of the labeled dataset using a series of machine learning algorithms: logistic regression, Decision tree and Passive Aggressive classifier. These algorithms were chosen for comparison with state-of-the-art approaches and are well-suited for binary classification tasks. 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.

Chapter 5

Experiment

This Chapter discusses the datasets used in this study, tools, experimental setup, configuration, and the evaluation mechanisms used to measure the performance of our approach, as well as the results obtained. The experiments are generally classified into three categories in order to answer our research questions which are defined in section 1.2.

5.1 Dataset Description

To the best of our knowledge there is no publicly available Amharic fake news dataset as the topic is a new research area for our country. To answer the research questions, we have collected 3000 Amharic news (1500 real news and 1500 fake news) by targeting the most popular and mainstream trustworthy news Facebook sites for real news and pages that propagate satire and misleading context for fake news. Obtaining well-balanced real and fake Amharic news is not an easy task; in particular, obtaining the fake articles was time-consuming. Every piece of news we collected is fact-checked by a professional (journalist), particularly the bogus news, which we gathered piece by piece. Each news has a source (pages): which posted the news, headline: title of the news, article or body of the news and finally label of the news either 1 for real news or 0 for fake news. Politics, sports, technology, economics, history, education, and religion are just a few of the topics we have covered. The majority of the news we gathered, however, is political in nature, and it can be found in both fake and real news formats. Concerning the topology or forms of news concerned, we have gathered mostly satire and fabrication news, as well as some propaganda news.

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

Table 5.1: Dataset characteristics

Table 5.1 provide information about the dataset and include the total number of news, the collection period, total number of pages and sample news sources.

5.2 Experiment Setup

All modules in the developed Amharic fake news detection system are implemented using Python version 3.7. Python is a high-level programming language which has different features and libraries for model development. Compared to other programming language python has short development time with interactive and portable characteristics[1]. Throughout the experiment different libraries and tools were used. Pandas: which is an open-source library that provides high performance, easy to use data structure and data analysis tools for python programming language. Numpy: a python library that adding support for large, multidimensional arrays and matrices along with a capability to operate on these data structures. Facepager: an open-source tool for collecting public data from social media platforms like Facebook. Scikit-learn: a library for python machine learning which is simple and efficient tools for data mining and data analysis algorithms for both supervised and unsupervised problems. Natural Language Toolkit (NLTK): a platform that is used for building Python programs to work with human language data. The hardware and software specification of the machine used to run all experiments is given in Table 5.2.

Manufacturer	Lenovo
Model	Lenovo 2356F82
Processor	Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz 2.90 GHz
Memory	16GB
Operating System	64 bit Windows 10

Table 5.2: Hardware and Software Specifcation of the machine

5.3 Experimental Scenarios

In order to evaluate the fake news detection approach, three groups of experiments are conducted using the collected datasets. The experiments are described as follows:

- Experiment 1: experiments conducted to evaluate the proposed lexicon-stance based analysis
- Experiment 2: experiments conducted to evaluate the generated Page score based on the proposed hypothesis between the news's page and false information.
- Experiment 3: aimed at evaluating the use of Amharic stop words on the fake news detection performance.

The proposed approach is compared with a state-of-the-art text representation lexicon-based approaches. For comparison different works [15], [19], [2], [20], [31], [36] are chosen. The authors used lexicon-based from text representation techniques to detect and identify fake news. In all of the above three experimental scenarios, the proposed fake news detection system is compared to previous lexicon-based works that only represent news texts in terms of word frequency without considering other news writing pattern.

5.3.1 Experiment I [The Lexicon-Stance based Detection]

Experiments under this category aimed at evaluating ability of the proposed hybrid (lexicon-stance) approach to accurately detect Amharic fake news. To fill the gap left by the previous lexicon-based approach, we have included stance features and will test whether the proposed combined features improve fake news detection. The stance features are consists of page scores and similarity features (headlines-article and headline-headline) of the news. Essentially, the entire goal of this study is to see if evaluating and examining these news stance features may help to improve the detection mechanism. We attempted to investigate what happens when the stance features are included and removed from the preceding text representation features.

5.3.2 Experiment II [The Page Score]

The purpose of this experiment was to assess the page score effect on the previous lexicon-based approach and to determine whether or not the proposed hypothesis would work. We have generated score for each page source and included as a single features. To see the effect of this score on the detection performance we have done an experiment by removing and incorporating this feature from lexicon-based approach. Additionally to analyze the hypothesis, the page score as a single feature was given to the machine learning classifiers and investigate the fake news detection performance.

5.3.3 Experiment III [Amharic Stop Words]

This final experiment aimed at evaluating the effect of Amharic stop words on the fake news detection performance. We have seen from some literature [18] that incorporating stop words at the preprocessing stage enhances the fake news detection. Amharic language has its own stop words and we do not know Amharic stop words can have positive effect or not on the fake news detection. We tried to do an experiment by including and removing stop words at the preprocessing stage to the effect on the detection performance.

5.4 Evaluation Metrics

In order to measure the performance of a model, there are many evaluation metrics such as accuracy, precision, recall, F score and ROC curve. The first four listed metrics are merely dependent on the confusion matrix and the numbers inside it.

5.4.1 Confusion Matrix

A confusion matrix is a table that is frequently used to describe the model's correctness and accuracy. The confusion matrix is simple to understand, but the terminology associated with it can be confusing. The following are some basic terms:
TP (True Positive): The number of instances that are positive and are correctly predicted as positive.
FP (False Positive): The number of instances that are negative but are incorrectly predicted as positive.

FN (False Negative): The number of instances that are positive but are incorrectly predicted as negative.
TN (True Negative): The number instances that are negative and truly predicted as negative.

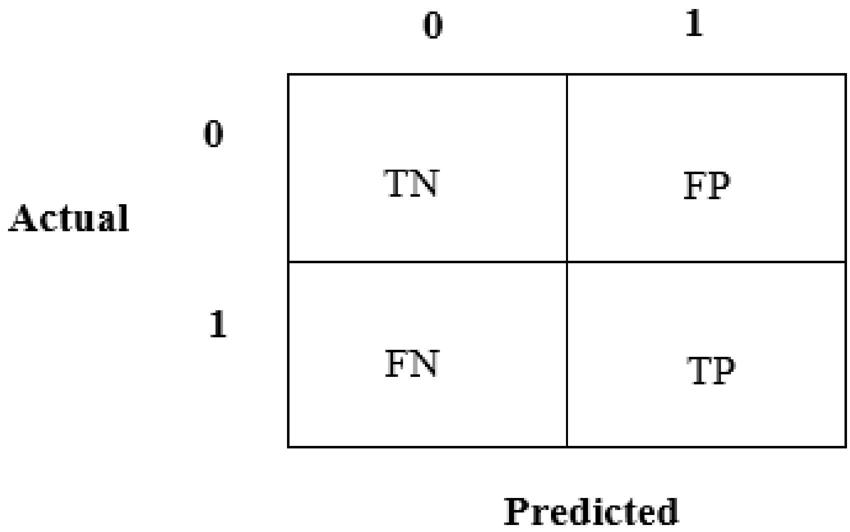

Figure 5.1: Confusion Matrix representation

Accuracy: to show how well the model detects positive and negative classes, we used the model's accuracy as an evaluation measure, because our dataset's class distribution is balanced. Accuracy measures how much accurately the model learns to classify the data. It is computed as the sum of True Positive and True Negative divided by the overall number of instances in the dataset.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Precision: Precision is a measure of the likelihood of getting correct positive class classification. It is computed as the number of True Positives divided by the number of True Positive plus False Positive.

$$PRE = \frac{TP}{TP + FP} \tag{5.2}$$

Recall: It is the measure of how sensitive our model is in identifying the positive class. A recall is computed as the number of True Positive divided by the number of True Positives plus False Negatives.

$$REC = \frac{TN}{FP + TN} \tag{5.3}$$

F1-score: It is a harmonic mean or weighted average of the model's precision and recall. F1-score is more preferable in measuring Accuracy when False Negatives and False Positives are crucial than the True Positives and True Negatives.

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5.4}$$

ROC Curve: is a probability curve that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. AUC is the area under the curve. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. This metrics mostly used for comparing the performance of different classification models at the same space.

$$FPR = \frac{FP}{FP + TN} \tag{5.5}$$

$$TPR = \frac{TP}{TP + FN} \tag{5.6}$$

5.5 Results and Discussion

5.5.1 Results

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?

To answer this research question, we must first experiment with the previous lexicon-based features using our prepared dataset, as well as determine which lexicon features (extracted from CountVectorizer or TF-IDF vectorizer) are better to proceed with the research.After we found the better lexicon features, we combined with the stance features proposed in this study to improve the performance of fake news detection. The experiment is conducted on a variety of supervised learning algorithms, primarily logistic regression, Passive Aggressive, and Decision Tree. Table 5.3 shows the result of the experiment. The rows indicate the lexicon feature extraction techniques (count vectorizer and tf-idf vectorizer) with different N-gram parameter. The columns indicate the metrics that are calculated from the confusion matrix. The results inside the Table show the three classifiers performance in the order: Logistic regression **(LR)**, Passive Aggressive **(PA)**, and Decision Tree **(DT)** respectively. 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 logistic regression and decision tree on all metrics in lexicon-based features except on ROC curve in Figure 5.2 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. We found that the Passive Aggressive and Logistic regression classifier have the same true negative and false positive value using lexicon-based tf-idf features extraction techniques. However the passive aggressive classifier has good true positive and good false negative value. When compared to the other two classifiers, decision tree has the lowest performance which has highest false positive value. When detecting fake news, the false positive value should be kept as low as possible in comparison to the other values in the confusion matrix. This is due to the fact that the system should not present false information to the user as real news and must be cautious in this regard.

Lexicon-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**

Table 5.3: Received result for lexicon features

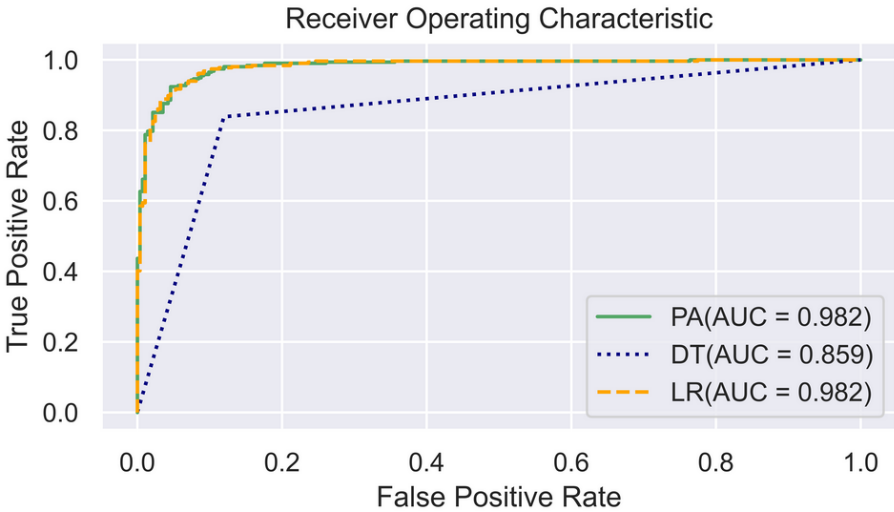

Figure 5.2: ROC curve of the three algorithms using lexicon based features

We previously experimented our model with lexicon-based features; now let's experiment it with stance features and individual performance. Finally, we will combine the lexicon and stance features to improve the detection of fake news from the state-of-the-art lexicon-based approach. Table 5.4 shows the stance detection features performance result 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 feature found from applying cosine similarity between each news headline gives better result compared to headline-article feature. 87.2% of accuracy on headline-headline feature while 79.6% accuracy on headline-article feature using logistic regression. 87% of F1-score is observed in headline-headline feature while 79% of F1-score is found on headline-article feature using logistic regression. 91% of Recall on headline-headline feature while 88% Recall is found on headline-article using passive aggressive classifiers. As we can see from the table the performance is improved when the three stance features are combined and given to the classifiers. The combined stance features gives better result of 95% accuracy, precision, recall and f1-score using passive aggressive classifier.

	Accuracy			Recall			Precision			F1-score		
Stance-features	**LR**	**PA**	**DT**	**LR**	**PA**	**DT**	**LR**	**PA**	**DT**	**LR**	**PA**	**DT**
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
Page Score	92	92	92	92	94	89	92	91	95	92	92	92
Combined Stance features	**93**	**95**	**94.3**	**92**	**95**	**94**	**94**	**95**	**95**	**93**	**95**	**95**

Table 5.4: Received result of stance features

The proposed hybrid lexicon-stance based feature performance result shows in Table 5.5. 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. Looking at the AUC metrics in ROC curve of Figure 5.3, 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. As in the previous lexicon based approach, decision tree has the lowest performance metrics result compared to logistic regression and passive aggressive classifiers. Using the hybrid lexicon-stance features, we discovered that the Passive Aggressive classifier has a lower false positive rate than the other two. The classifier wrongly predicted only 6 fake news as real from the total of 284 fake news. From the total of 316 real news logistic regression predicted 308 news as a real and 8 news wrongly predicted as a fake. In detecting fake news the false negative and false positive value should be minimum as much as possible and in this regard the passive aggressive classifier outperforms the other. As the figure 5.3 shows the area under curve (AUC) of the passive aggressive classifier is 0.995 and that of logistic regression is 0.994. Decision tree has the lowest AUC value of 0.951 compared to the other.

ML Classifiers	**Detection Techniques**	**A**	**P**	**R**	**F1 score**	**AUC**
LR(Logistic Regression)	**Proposed(Lexicon-Stance)**	**97.3**	**97**	**97**	**97**	**0.994**
	Baseline(Lexicon)	93	94	92	93	0.982
PA(Passive Aggressive)	**Proposed**	**97.5**	**98**	**97**	**98**	**0.995**
	Baseline	93.4	95	93	94	0.982
DT(Decision Tree)	**Proposed**	**95**	**96**	**94**	**95**	**0.951**
	Baseline	86	89	83	86	0.859

Table 5.5: Received result of lexicon-stance features

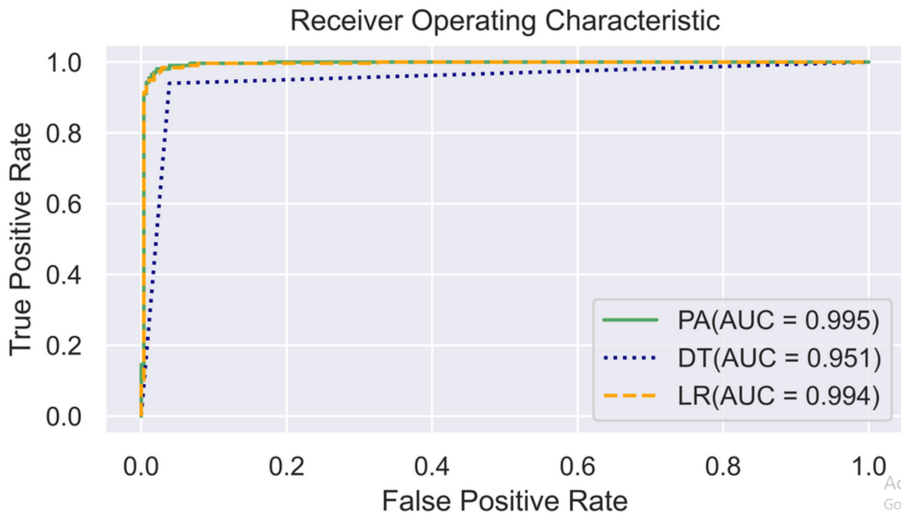

Figure 5.3: ROC curve of the three algorithms using lexicon-stance based features

Our experiment results show that the hybrid feature (lexicon-stance) are capable of improving the previous lexicon based feature results up to 4.1% accuracy, 3% precision, 4% recall and 4% F1-score. In addition the hybrid feature 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%. Furthermore, among

the three algorithms used in this research work, Passive aggressive algorithm out performs the others algorithms in most of the experiments.

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 source attribute which mathematically modeled as equation (2) and given to the classifiers as a single feature yields 92% of accuracy. To the best of our knowledge, no studies have attempted to extract the source of the news (Pages are the source of the news for this study) solely through a content-based approach. The generated page 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 5.6 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 is found using logistic regression classifier and 96% of accuracy using passive aggressive classifier when the page score is incorporated on lexicon-based features.

Table 5.7 shows the result of the other stance features (headline-article and headline-headline) combined with the previous lexicon based feature. From the table the passive aggressive classifier has better result of 94.6% accuracy compared to logistic regression and decision tree and improved performance change of 1.2% from the previous lexicon based features. The stance features are found from cosine similarity score between headline to article of the news and headline to headline of different news. From the table it can be observed that overall the lexicon-cosine similarity scores has low performance result than the lexicon-page score feature.
For this research question our experiment results show that incorporating page score feature on lexicon based approach gives improved result and enhance the detection performance. Furthermore, from the proposed three stance features (page score, headline-article and headline-headline) we found that page score has more enhanced effect on the lexicon based detection.

ML classifiers	Features	A	P	R	F1 score	AUC
LR	**Lexicon-Page score**	**96.6**	**97**	**97**	**97**	**0.996**
	Lexicon-based	93	94	92	93	0.982
PA	**Lexicon-Page score**	**96**	**97**	**97**	**97**	**0.997**
	Lexicon-based	93.4	95	93	94	0.982
DT	**Lexicon-Page score**	**93.3**	**94**	**94**	**94**	**0.933**
	Lexicon-based	86	89	83	86	0.859

Table 5.6: Result found using lexicon-page score

ML classifiers	Features	A	P	R	F1 score	AUC
LR	**Lexicon-similarity score**	**94.3**	**95**	**94**	**95**	**0.985**
	Lexicon-based	93	94	92	93	0.982
PA	**Lexicon-similarity score**	**94.6**	**96**	**94**	**95**	**0.985**
	Lexicon-based	93.4	95	93	94	0.982
DT	**Lexicon-similarity score**	**87.5**	**89**	**87**	**88**	**0.875**
	Lexicon-based	86	89	83	86	0.859

Table 5.7: Result found using lexicon-Cosine similarity score

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

In section 3.4, we saw that there are studies that show that incorporating stop words during the preprocessing stage rather than removing it improves the performance of fake news detection. Amharic has its own stop words and we have done an experiment to see the effect of stop words on the Amharic fake news detection performance. To answer this research question we have collected and prepared known Amharic stop words from the previous researches and merge with our own stop words from the dataset. We used custom function to remove or incorporate the stop words at the preprocessing stage and observe if there is any change on the fake news detection performance. Table 5.8 and 5.9 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. We tried to figure out which stop words could account for some of the performance difference using repeated trial of incorporating and removing various amharic stop words and discovered that words like the ones below are the most common.

ብለዋል, አይደለም, ገልፀዋል, አስታውቀዋል, ግን

For this research question our experiment results show that incorporating Amharic stop

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 5.8: Result found for lexicon-stance when stop words are removed

words on lexicon-stance based approach has no significant difference to enhance the detection performance. However We noticed some performance changes in the passive aggressive algorithm as a result of some Amharic stop words.

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

Table 5.9: Result found for lexicon-stance when stop words are included

Finally, we evaluated our system to determine how dependent it is on the source attribute by injecting untrained pages into the previous testing dataset (notice that for this experiment, we try to alter and use half of the test data) and assigning fake news to pages that primarily post real news and real news to pages that primarily post fake news. Table 5.10 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 page score. 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 page (source of news) or the news article for the sole purpose of testing. Furthermore, we did not use or alter the entire testing datasets, and further decrement may be seen if the entire test dataset is changed.

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

Table 5.10: Received result of our system using untrained page datasets

5.5.2 Discussion

All sets of experiments are conducted using newly collected Amharic fake news data set and three selected algorithms to show whether added stance features can have enhancement effect on the detection performance or not. By doing empirical experiments, we come up with three basic findings. The first one is, we practically showed that stance features have significant improvement effect on the fake news detection. Secondly, we experimentally examine the effect of a single hand crafted feature (page score) extracted from the source attribute of the news using pre-defined hypothesis on the lexicon based fake news detection. Finally we have experimentally identified Amharic stop words have no significant effect on the proposed lexicon-stance based fake news detection though some selective stop words have minor effect on the detection performance.

The stance features used for this researches are basically grouped in to three. The headline-article which analyzes the stance of headline towards the news article by calculating the cosine similarity between the two. The headline-headline which analyzes the stance of a news headline towards the other news headlines by calculating the cosine similarity between them and finally the page score feature generated from the stance of the page towards news based on the hypothesis. Among this features the page score shows better performance of 92% accuracy on all three algorithms. As a single feature the received result indicate that our hypothesis which state 'people who have spread false information in the past from the dataset are more likely to post fake content in the next post and people who spread real information are more likely to post real content in the next' is somehow correct. Next to page score the headline-headline feature shows better detection performance than headline-article feature. We have also observed that the stance features found from combining the three individual stance features yields 95% of accuracy using passive aggressive algorithm which is better result compared to each single feature. Our main objective was to see the effect of stance features on the lexicon based fake news detection and we have found that there is an improvement of 4.1% accuracy, 3% precision, 4% recall and 4% F1-score using passive aggressive algorithm from the previous lexicon based approach when stance features are incorporated.
In this research work, we have come up with a new hand crafted single feature page score by extracting the source attribute of the news using pre-defined hypothesis. And trying to

examine the enhancement effect of this single feature on lexicon based approach was our second research question. We observed that page score has more enhanced effect on the lexicon based detection compared to other stance features, with accuracy improvements of 3.6% and 3% using logistic regression and passive aggressive algorithm, respectively. Finally regarding Amharic stop words effect on the fake news detection performance experimental output indicate that generally incorporating stop words at the preprocessing stage does not have significant performance difference from removing it. However, we found that some Amharic stop words accounts for a minor difference in fake news detection, with a 0.5% accuracy improvement in the passive aggressive algorithm when these stop words are included on the dataset.

Chapter 6

Conclusion and Recommendation

6.1 Conclusion

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 is 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 page 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, an experiment was conducted to determine how including and removing Amharic stop words during the preprocessing stage affected the performance of fake news detection.

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. 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. From this result obtained using page score feature we can come to conclusion that mostly by only observing previous news articles posted by the page could lead to a decision on the veracity of the news on the next post. The headline to article as an individual feature got 83.3% of accuracy using Passive aggressive and headline to headline feature got 87.2% of accuracy using logistic regression. These results are good and it can be conclude that Amharic real news have mostly similar writing style of news headline in terms of word frequency and the news title stance towards the article is mostly similar. For that of Amharic fake news, mostly the title stance towards the article is different and the writing style of different fake news headlines in terms of word frequency are mostly similar. 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.

Readers should be aware that the conclusions presented above are not always correct, and that news should always be verified. One of the objectives of this study is to provide a foundation for and assistance to fact checkers. We also want to contribute to the advancement of Amharic fake news detection systems.

6.2 Recommendation

The results of our study show that stance features extracted from news attributes help to improve lexicon based fake news detection. We believe that further investigating additional stance-based features could further help to improves the performance of fake news detection. Taking this into consideration, the following research works are planned to be conducted in the future:

- This research paper is concentrated on stance detection to improve the fake news detection and in the future we recommend to included stylistic features like the number of verbs and nouns, the number of punctuations and syntax and the quantity of casual words.
- We recommend to examine the sentiment of the news, the date posted and the category of the news as important features.
- We suggest to extract hand crafted features from the news. News has different attributes and attempting to capture its various writing styles could help to improve the detection performance.
- This paper is focused on content based approach only and external source about the news is not utilized. We suggest in the future to use network analysis and examine propagation of the news on the social media as well.

Bibliography

[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 technology*, pp. 76–83, 2021.

[3] M. H. Worku and M. M. Woldeyohannis, "Amharic fake news detection on social media using feature fusion."

[4] F. Gereme, W. Zhu, T. Ayall, and D. Alemu, "Combating fake news in "low-resource" languages: Amharic fake news detection accompanied by resource crafting," *Information*, vol. 12, no. 1, p. 20, 2021.

[5] N. K. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the association for information science and technology*, vol. 52, no. 1, pp. 1–4, 2015.

[6] 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.

[7] 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.

[8] S. I. Manzoor, J. Singla *et al.*, "Fake news detection using machine learning approaches: A systematic review," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 230–234.

[9] 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.

[10] 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.

[11] 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.

[12] M. Meyers, G. Weiss, and G. Spanakis, "Fake news detection on twitter using propagation structures," in *Multidisciplinary International Symposium on Disinformation in Open Online Media*. Springer, 2020, pp. 138–158.

[13] M. A. Adamu, "Role of social media in ethiopias recent political transition," *Journal of Media and Communication Studies*, vol. 12, no. 2, pp. 13–22, 2020.

[14] 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.

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

[16] 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.

[17] 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.

[18] I. Wijers, "Don't believe everything you read: Identifying fake news using an author attribution framework."

[19] 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.

[20] 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.

[21] 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.

[22] V. Tsarapatsanis, "Fake news detection," *A thesis submitted for the degree of Master of Science (MSc) in Data Science, International Hellenic university*, 2020.

[23] J. Botha and H. Pieterse, "Fake news and deepfakes: A dangerous threat for 21st century information security," in *ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited*, 2020, p. 57.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] S. Lorent, "Master thesis: Fake news detection using machine learning," 2019.

[30] M. Mahyoob, J. Al-Garaady, and M. Alrahaili, "Linguistic-based detection of fake news in social media," *Forthcoming, International Journal of English Linguistics*, vol. 11, no. 1, 2020.

[31] 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.

[32] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2012, pp. 171–175.

[33] O. Popoola, "Using rhetorical structure theory for detection of fake online reviews," in *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, 2017, pp. 58–63.

[34] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment aware fake news detection on online social networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2507–2511.

[35] S. 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.

[36] 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.

[37] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake news early detection: A theory-driven model," *Digital Threats: Research and Practice*, vol. 1, no. 2, pp. 1–25, 2020.

[38] V. L. Rubin and T. Lukoianova, "Truth and deception at the rhetorical structure level," *Journal of the Association for Information Science and Technology*, vol. 66, no. 5, pp. 905–917, 2015.

[39] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment analysis for fake news detection," *Electronics*, vol. 10, no. 11, p. 1348, 2021.

[40] C. L. JERONIMO, L. B. MARINHO, C. E. CAMPELO, A. VELOSO, and A. S. D. C. MELO, "Characterization of fake news based on subjectivity lexicons," *Journal of Data Intelligence*, vol. 1, no. 4, pp. 419–441, 2020.

[41] R. Masood and A. Aker, "The fake news challenge: Stance detection using traditional machine learning approaches." in *KMIS*, 2018, pp. 126–133.

[42] S. Suyanto *et al.*, "Synonyms-based augmentation to improve fake news detection using bidirectional lstm," in *2020 8th International Conference on Information and Communication Technology (ICoICT).* IEEE, 2020, pp. 1–5.

[43] B. Ghanem, P. Rosso, and F. Rangel, "Stance detection in fake news a combined feature representation," in *Proceedings of the first workshop on fact extraction and VERification (FEVER)*, 2018, pp. 66–71.

[44] D. Pisarevskaya, "Deception detection in news reports in the russian language: Lexics and discourse," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 74–79.

[45] M. K. Elhadad, K. F. Li, and F. Gebali, "A novel approach for selecting hybrid features from online news textual metadata for fake news detection," in *International*

Conference on P2P, Parallel, Grid, Cloud and Internet Computing. Springer, 2019, pp. 914–925.

[46] J. Zhang, B. Dong, and S. Y. Philip, "Fakedetector: Effective fake news detection with deep diffusive neural network," in *2020 IEEE 36th International Conference on Data Engineering (ICDE).* IEEE, 2020, pp. 1826–1829.

[47] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, "State of the art models for fake news detection tasks," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT).* IEEE, 2020, pp. 519–524.

[48] S. Gaonkar, S. Itagi, R. Chalippatt, A. Gaonkar, S. Aswell, and P. Shetgaonkar, "Detection of online fake news: A survey," in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN).* IEEE, 2019, pp. 1–6.

[49] S. G. Miretie and V. Khedkar, "Automatic generation of stopwords in the amharic text," *International Journal of Computer Applications*, vol. 975, p. 8887, 2018.

[50] E. Shoemaker, "Using data science to detect fake news," *An Honors College Project Presented to the Faculty of the Undergraduate College of Intergated Science and Enginerring, James Madison University*, 2019.

[51] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

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

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

[54] M. Awan, A. Yasin, and Nobanee, "Fake news data exploration and analytics," *Electronics*, vol. 10, no. 19, p. 2326, 2021.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

Appendix A

Ten questions for fake news detection

Done by the news literacy project

Use the questions below to assess the likelihood that a piece of information is fake news. The more red color you circle, the more skeptical you should be!

1. Gauge your emotional reaction:
   Is it strong? Are you angry? Are you intensely hoping that the information turns out to be true? False? YES | NO

2. Reflect on how you encountered this. Was it promoted on a website? Did it show up in a social media feed? Was it sent to you by someone you know?

3. Consider the headline or main message:

   - Does it use excessive punctuation (!!) or ALL CAPS for emphasis? YES | NO
   - Does it make a claim about containing a secret or telling you something that "the media" doesn't want you to know? YES | NO
   - Don't stop at the headline! Keep exploring.

4. Is this information designed for easy sharing, like a meme? YES | NO

5. Consider the source of the information:

   - Is it a well-known source? YES | NO
   - Is there a byline (an author's name) attached to this piece? YES | NO

- Go to the website's "About" section: Does the site describe itself as a "fantasy news" or "satirical news" site? YES | NO
- Does the person or organization that produced the information have any editorial standards? YES | NO
- Does the "contact us" section include an email address that matches the domain (not a Gmail or Yahoo email address)? YES | NO
- Does a quick search for the name of the website raise any suspicions? YES | NO

6. Does the example you're evaluating have a current date on it? YES | NO
7. Does the example cite a variety of sources, including official and expert sources? Does the information this example provides appear in reports from (other) news outlets? YES | NO
8. Does the example hyperlink to other quality sources? In other words, they haven't been altered or taken from another context? YES | NO
9. Can you confirm, using a reverse image search, that any images in your example are authentic (in other words, sources that haven't been altered or taken from another context)? YES | NO
10. If you searched for this example on a fact-checking site such as Snopes.com, FactCheck.org or PolitiFact.com, is there a fact-check that labels it as less than true? YES | NO

REMEMBER:

- It is easy to clone an existing website and create fake tweets to fool people.
- Bots are extremely active on social media and are designed to dominate conversations and spread propaganda.
- Fake news and other misinformation often use a real image from an unrelated event.
- Debunk examples of misinformation whenever you see them. It's good for democracy!