5-1-2019

A Hybrid Model to Detect Fake News

Indhumathi Gurunathan
Boise State University

A HYBRID MODEL TO DETECT FAKE NEWS

by

Indhumathi Gurunathan

A project

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Computer Science

Boise State University

May 2019

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Indhumathi Gurunathan

Thesis Title: A hybrid Model to Detect Fake News

Date of Final Oral Examination: 23 April 2019

The following individuals read and discussed the project submitted by student Indhumathi Gurunathan, and they evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Francesca Spezzano, Ph.D.	Chair, Supervisory Committee
Dianxiang Xu, Ph.D.	Member, Supervisory Committee
Bogdan Dit, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Francesca Spezzano, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

DEDICATION

Dedicated to my husband Ashok, my two kids Akshaya Sri and Goutham Kirthik.

ACKNOWLEDGEMENTS

I would like to address my heartfelt thanks to my advisor, Dr. Francesca Spezzano, for her support and advice she gave me during my graduate studies at Boise State University. Her guidance and knowledge have been a major factor throughout my project. I would also like to thank my committee members, Dr. Dianxiang Xu and Dr. Bogdan Dit for the guidance and support through this process.

I would also like to thank the Computer Science Administrative Department for helping me whenever needed and for providing a welcoming environment for international students. And finally, I would like to thank my parents and my amazing family for the never-ending love, care and support given to me over the years and I undoubtedly could not have done this without them.

ABSTRACT

The wide availability of user-contributed content in the online social media facilitates aggregation of people around common interests, worldviews, and narratives. But over the years, internet being the source of information also becomes the source of misinformation. As people are generally awash in information, they can sometimes have difficulty discerning misinformation propagated on web platforms from truthful information. They may also lean heavily on information providers or social media platforms to curate information even though such providers do not commonly validate sources. In this project, we primarily focus was on political news and propose a hybrid model to detect misleading news. We use different modalities including news content (headline, body, and associated image), source bias and social network of users who spread the news to detect whether the news is misleading or factual.

We study the relationship between the publisher bias and news stance and show that hyperpartisan news sources are more likely to spread misleading stories than other sources. Also, we demonstrate that it is not necessary to analyze the news content to detect misleading news, but using features such as publisher bias, user engagements, and images related to the news can achieve comparable performances (AUROC of 0.90 vs. 0.88 and average precision of 0.79 vs. 0.78).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PICTURES

LIST OF ABBREVIATIONS

RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GPU	Graphics Processing Unit
AMT	Amazon Mechanical Turk
TF-IDF	Term Frequency-Inverse Document Frequency
LIWC	Linguistic Inquiry and Word Count
SMOG	Simple Measure of Gobbledygook Index
ARI	Automatic Readability Index
LIX	Lycee International Xavier Index
CNN	Convolutional Neural Network
VGG	Visual Geometry Group
SNAP	Stanford Network Analysis Platform
MBFC	Media Bias Fact Check
BP	Belief Propagation
ROC	Receiver Operating Characteristic
AUROC	Area Under the ROC
MRF	Markov Random Field
URL	Uniform Resource Locator
TTR	Text-Type Ratio
CSI	Capture, Score, and Integrate

CHAPTER 1

INTRODUCTION

According to the Oxford dictionary, misinformation is “false or inaccurate information, especially that which is deliberately intended to deceive”. These days, the massive growth of the Web and social media has provided fertile ground to consume and quickly spread the misinformation without fact-checking. Misinformation can assume many different forms such as vandalism [6], spam [50], rumors [8], hoaxes [51], clickbait, counterfeit websites, fake product reviews [7], fake news [23], etc.

Fake news is low-quality news that is created to spread misinformation and mislead readers. The consumption of news from social media is highly increased nowadays so as spreading of fake news. According to the Pew research center [9], 64% of Americans believe that fake news causes confusion about the basic facts of current events. A recent study conducted on Twitter [10] revealed that fake news spread significantly more than real ones, in a deeper and faster manner and that the users responsible for their spread had, on average, significantly fewer followers, followed significantly fewer people, were significantly less active on Twitter. Also, human behavior contributes more to the diffusion of fake news than the real news especially when the news conforms to their preexisting attitudes and beliefs. Moreover, bots are equally responsible for spreading real and fake news, and then the considerable spread of fake news on Twitter is caused by human activity.

The volume of misleading news in social media has grown in popularity in recent years. In 2017, the Pew Research Center found that 67% of American adults (ages 18+) get news from social media, which was a 5% increase since 2016 [11]. An analysis of news leading up to the 2016 election conducted by BuzzFeed, found that there was more engagement with the leading misleading news stories than real news stories [12]. Thus, news is becoming more accessible and widespread than ever before. However, the spread of information has also contributed to the spread of misleading news which has fostered the advancement of various methods to determine the validity of news. One such method is developed upon evaluating linguistic attributes such as features determining readability and lexical information [21, 16, 30]. These methods often mimic that of what would generally be considered the most effective of all: reading through news the purpose of evaluating their accuracy. However, with the spread of misleading news it is unlikely if not impossible for everyone to spend large quantities of time reading through multiple newspapers and sources. Additionally, in a recent study, Gabielkov et al. [13] found evidence that the number of news shares is an inaccurate measure of actual readership. Thus, people are immersed in information across social media which is often shared without users reading and considering the validity of content thus leading to possible consequences of its diffusion. The impact of fake news diffusion is huge, it affects news media ecosystem, cause political damage, influences social media marketing and also impairs individuals' opinions. According to the Pew research center [9], 64% of Americans believe that fake news cause confusion about the basic facts of current events. In the same Pew survey, 23% of respondents admitted to sharing fake news, while 14% said they shared an article knowing it was fake. Even Oxford dictionary selected "post-truth" as its word of the year 2016. According to a 2016 Gallup poll [3], trust in mass media among Americans has plummeted to 32 percent, an all-time low from 72 percent in 1976.

Thus, in this project, we use machine learning techniques to develop a hybrid model to detect fake news. To the best of our knowledge, we analyze all the news data available including headline, body content, associated image, social network of the users who spread the fake news and source bias, for misleading news detection. Interestingly, our analysis highlights a correlation between publisher political bias and its credibility. In fact, by analyzing information collected from mediabiasfactcheck.com, “the most comprehensive media bias resource on the Internet”, we showed that hyperpartisan news sources are more likely to spread misleading stories than other sources. Moreover, we find out that we can avoid to “read” the news to determine its veracity, as considering publisher bias, user engagements, and images related to the news can achieve comparable performances (AUROC of 0.90 vs. 0.88 and average precision of 0.79 vs. 0.78).

The remainder of this document is organized as follows: Chapter 2 discusses about the related work that was done previously for fake news detection. Chapter 3 provides a brief description on the dataset and discuss the techniques used in the data collection process from mediabiasfactcheck.com website. Chapters 4 and 5 describe the methods to extract features from various aspects of the news and the experiments for different combination of these features to detect misleading news. Chapter 6 concludes the report discussing the efficiency of the hybrid model and suggest directions for future work.

CHAPTER 2

RELATED WORK

To detect misleading news, many works have considered news content (headline, body, image), the social network between the users and their social engagement (share, comment, and discuss given news), or a hybrid approach that considers both (see [17, 23] for a survey). The survey Shu, et al. (2017) [23], precisely gave definitions about the fake news and provided the complete review of methods to detect fake news on social media. The paper characterized the fake news by comparing different theories and properties in both traditional news media and social media. Existing algorithms to detect fake news on traditional media depends on only the news content. These methods are ineffective for the case of social media and therefore, leveraging this problem with extra social context auxiliary information helps to detect fake news more efficiently.

News content-based features include both linguistic features extracted from the text of the news, metadata-based features such as news source (author and/or publisher), headlines, etc., and visual-based features extracted from images and videos associated with the news. For instance, Seyedmehdi and Papalexakis [20] proposed a solution based on extracting latent features from news article text via tensor decomposition to categorize fake news as extreme bias, conspiracy theory, satire, junk science, hate group, or state news. Potthast et al. [24] used the writing style of the articles to identify extremely biased news from the neutral one by using the techniques called unmasking. This study used a dataset com- posed of 1,627 articles from a Buzzfeed dataset. Features such as n- grams, stop words, parts of speech and readability were considered in this study. Although there was higher accuracy in determining the mainstream articles vs. hyperpartisan (0.75 accuracy based on stylistic features and 0.71 for topic) the research was limited in deciphering between real and fake news (only a 0.55 accuracy for style and 0.52 for topic).

Horne and Adali [21] considered both news body and headline for determining the validity of news. They included three datasets: a dataset created by Buzzfeed leading to the 2016 US elections, one created by the researchers containing real, fake and satire sources, and a third dataset containing real and satire articles from a previous study. Based on textual features extracted from body and headline, they found out that the content of fake and real news is drastically different as they were able to obtain a 0.71 accuracy when considering number of nouns, lexical redundancy (TTR), word count, and number of quotes. Further, the study found that fake titles contain different sorts of words (stop words, extremely positive words, and slang among others) than titles of real news articles resulting in a 0.78 accuracy. Pérez-Rosas et al [16] collected two new datasets, the FakeNewsATM dataset covering seven different news domains (education, business, sport, politics, etc.) and the Celebrity dataset regarding news on celebrities. They analyzed the news body content only and achieved an accuracy up to 0.76 in detecting misleading content. They also tested cross domain classification obtaining poor performances by training in one dataset and testing in the other one, but better accuracies (ranging from 0.51 to 0.91) in training on all but the test domain in the FakeNewsATM dataset.

Images in news articles also play a role in misleading news detection [34, 18, 19, 25]. Fake images are used in news articles to provoke emotional responses from readers. Images are the most eye-catching type of content in news; a reader can be convinced of a claim by just looking at the title of the news and the image itself. So, it’s crucial to include image analysis in fake news detection techniques.

For instance, Jin et al. [33] used only visual and statistical features extracted from news images for microblogs news verification and obtained an accuracy of 0.83 on an image dataset collected from Sina Weibo on general news events. More recently, Wang et al. [29] proposed a deep-learning based framework to extract features from both text and image of the news that are not related to specific events to detect misleading content. Their results show an accuracy value ranging from 0.71 on a Twitter dataset to 0.82 on Sina Weibo.

Social context-based features consider (i) the profile and characteristics of users creating and spreading the news (e.g., number of followers/followees, number of posts, credibility and the reliability of the user) also averaged among all the users related to particular news, (ii) users' opinion and reactions towards social media posts (post can potentially contain fake news), (iii) various type of networks such as friendship networks, co-occurrence networks (network formed based on the number of posts the user write related to the news), or diffusion network where edges between users represent information dissemination paths among them.

Kim et al. [27] propose methods to not only detect the fake news but also to prevent the spread of fake news by making the user flag fake news and used reliable third-parties to fact check the news content. They developed an online algorithm for this purpose, so it works at the time of user spreading the fake news thus preventing it from spreading. Jin et al. [28] developed a method for detecting fake news by using the users' viewpoints to find relationships such as support or oppose and by building a credibility propagation network by using these relationships. Users on social media inclined to network with like-minded people and then they receive and share the news that promotes their interests/beliefs which will result in echo chamber effect. So, extracting these network-based features by creating different kinds of network such as stance, co-occurrence, friendship, and diffusion networks help to infer network pattern to identify the fake news.

The stance network has nodes and edges with nodes representing the posts related to the news and the edge indicating the weight of the similarity of the viewpoints. The co-occurrence network is built based on the user engagements by counting the number of posts the user-authored related to the news. The friendship network represents the network pattern of the followers and followees of the user who posts related to the news. The diffusion network tracks the information diffusion path between the users. Network metrics such as degree and clustering coefficients are used to characterize the diffusion and friendship network. Wu and Liu [31] used the way news spread through the social network to find the fake news. They used graph mining method to analyze the social network and recurrent neural networks to represent and classify propagation pathways of a message.

Finally, hybrid methods combine the two previous approaches. For instance, Ruchansky et al. [32] used temporal behavior of users and their response and the text content of the news to detect the fake news. They proposed the CSI model (Capture, Score, and Integrate) to classify the news article. Fairbanks et al. [37] show that a content-based model can identify publisher political bias while a structural analysis of web links is enough to detect whether the news is credible or not. Shu et al. [36] exploited both fake news content and the relationship among publishers, news pieces, and users to detect fake news.

Regarding clickbait detection specifically, Chakraborty et al. [41] build personalized automatic blocker for clickbait headlines by using a rich set of features that use sentence structure, word patterns, N-gram features, and clickbait language. Their browser extension ‘Stop-Clickbait’ warns users for potential click- baited headlines. Potthast et al. [39] used Twitter datasets to identify messages in social media that lead to clickbait. They gathered tweets from various publishers and constructed features based on teaser message, linked web page, and meta information. Anand et al. [38] used three variants of bidirectional RNN models (LSTM, GPU, and standard RNNs) for detecting clickbait headlines They used two different word embedding techniques such as distributed word embeddings and character-level word embeddings. Chen et al. [42] examined a hybrid approach for clickbait detection by using text-based and non-text-based click baiting cues. While textual cues use text-based semantic and syntactical analysis, non-textual cues relate to image and user behavior analysis.

CHAPTER 3

DATASETS

In this chapter, we discuss the lack of a large-scale misleading news dataset (especially in the political domain) and present the datasets used in this project, namely the FakeNewsNet dataset and a new dataset containing publisher bias and credibility details crawled from the MediaBias/FactCheck website.

3.1 Available Datasets and Limitations

There exist several datasets containing political news that have been used for fake news detection, as shown in Table 1.

Table 3.1: Available datasets for misleading news detection.

Dataset	Size	Text	Images
BuzzFeedNews [20]	1,627	✓	
Horne and Adali DS1 [11]	71	✓	
Horne and Adali DS2 [11]	225	✓	
Pérez-Rosas et al [18]	480	✓	
FakeNewsNet [24]	384	✓	✓

The BuzzFeedNews dataset contains news regarding the 2016 U.S. election published on Facebook by 9 news agencies. This dataset labels 356 news articles as left-leaning and 545 as right-leaning articles, while 1264 are mostly true, 212 are a mixture of true and false, and 87 are false. Horne and Adali used two datasets in their paper [21]. The first dataset, DS1, contains 36 real news stories and 35 fake news stories, while the second one, DS2, contains 75 real, misleading and satire news (75 for each category). The main drawback of these two datasets is that labels are assigned according to the credibility of the news source, instead of via fact-checking. However, a news source can have mixed credibility and publish both factual and misleading information. Pérez-Rosas et al [16] collected a dataset of 480 news where 240 are fact-checked real news belonging to six different domains (sports, business, politics, etc.) and 240 are fake news collected via crowdsourcing, i.e. they asked to AMT workers to write a fake news item based on one of their real news item and by mimic journalist style. FakeNewsNet [35], described in Section 3.2, is the only state-of-the-art dataset containing information beyond the news content modality and in the political domain. With the importance and relevance, this dataset was used to conduct the analysis in this project. As Table 1 shows, there is generally limited availability of large-scale benchmarks for fake news detection as collecting labels requires fact-checking, which is a time-consuming activity. As reported in [23], other datasets have been used for related tasks, but they are not suitable for our analysis as they do not contain proper news articles. For instance, LIAR [40] contains human- labeled short statements, while CREDBANK [47] contains news events, where each event is a collection of tweets. Finally, the MediaEval Verifying Multimedia Use benchmark dataset [43] used in [29] contains images and tweets instead of news articles.

3.2 FakeNewsNet Dataset

The FakeNewsNet dataset consists of details about the news content, publisher information, and social engagement information [35]. The ground truth labels are collected from journalist experts such as Buzzfeed and the fact-checking website Politifact. The dataset is divided into two networks as Buzzfeed and Politifact and the news contents are collected from Facebook web links. Dataset included all the downloaded available images related to the news in this dataset. The publishers' bias is retrieved from the dataset described in the next section. In this work, the news from both Politifact and Buzzfeed were merger to have a larger dataset to work with. After cleaning the dataset from missing news bodies or headlines, there were a total of 384 news, 175 misleading and 209 factual.

3.3 MediaBias/FactCheck Dataset

In order to exploit the partisan information of the news source, we crawled the website mediabiasfactcheck.com, whose main goal is to educate the public on media bias and deceptive news practices. This website contains a comprehensive list of news sources, their bias and their credibility of factual reporting score. Here, the publisher political bias is defined by using seven degrees of bias: extreme-right, right, right-centered, neutral, left-centered, left, and extreme-left.

The factual reporting score of all the news sources were collected under five categories: Left bias (moderately to strongly biased to- ward liberal causes), Left-center (slight to moderate liberal bias), Least (minimal bias), Right-Center (slightly to moderately conservative in bias), and Right bias (moderately to strongly biased toward conservative causes). The credibility score of these publishers falls into three categories: Very high (which means the source is always factual), High (which means the source is almost always factual) and Mixed (which means the source does not always use proper sourcing or sources to other biased/mixed sources). The publisher bias was also collected under the category Questionable Sources which contains extremely biased publishers mainly doing propaganda and/or writing misleading news. The number of publishers in each category considered is reported in Figure 1 and there is a total of 1,783 publishers. The relationship between the source bias and its credibility is analyzed in Section 4.3.

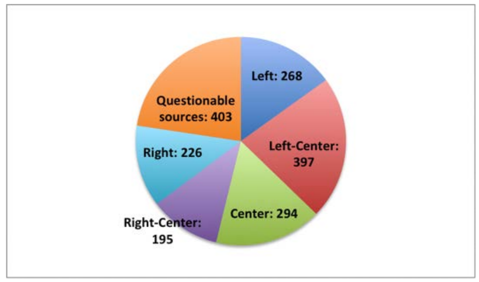

Figure 1: Number of publishers per category in the MediaBias/FactCheck dataset.

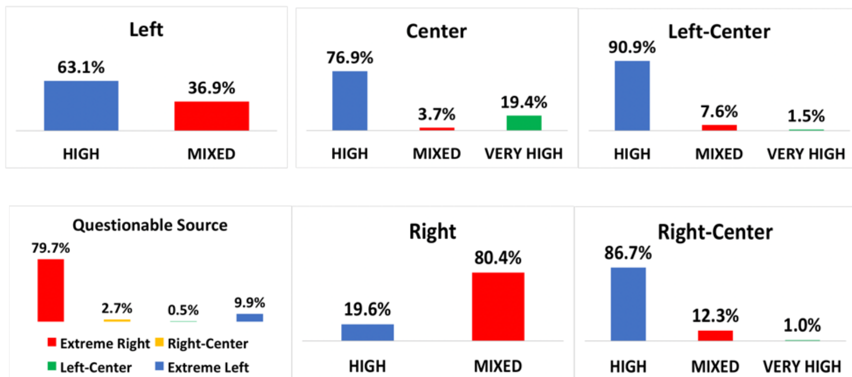

Figure 2: Publisher credibility per bias and bias distribution within questionable sources in the MediaBias/FactCheck dataset.

CHAPTER 4

METHODS TO EXTRACT FEATURES

In this chapter, we describe in detail the set of features used in the experiments to detect misleading political news. We consider five different modalities, namely news content, headline, news description, images, source bias, and social.

4.1 Textual Features

To analyze text content, we use the following groups of features, and these features are computed for both the news body content and news headline.

4.1.1 Term Frequency-Inverse Document Frequency (Tf-Idf)

In this work, a tf-idf is used to represent text, where each word is represented by its score to express the importance of the word in a corpus based on how frequently the word appears in the document and also how many other documents contain that word. We preprocessed the text by applying Stemming and punctuations and stop-word removal. A basic tf-idf scoring function is available in Eq. 4.1.

$$\text{tf-idf}_{w,di,D} = \left(\frac{C_w}{|d_i|}\right) . \log\left(\frac{|D|}{1+|d \in D :\ w \in d|}\right) \tag{4.1}$$

The first term represents the term frequency (tf) of the word w, which is the ratio of the number of occurrences of the word (C_w) to the total number of words in the document ($|d_i|$). The second term is the inverse document frequency (idf) which boosts the more informative words and diminish the impact of frequently used words likes articles, pronouns. The idf is computed by taking the logarithm of the total number of documents in the corpus (|D|) divided by the number of documents with the word offset by 1 to avoid 0 denominators (1 + |d ∈ D: w ∈ d|). The words that appear in almost all the documents will have an idf close to 0 and the words that appear in only select documents will have larger idf values, thereby increasing their tf-idf weights.

4.1.2 Linguistic Inquiry and Word Count (LIWC)

LIWC is a transparent text analysis tool that counts words in psychologically meaningful categories. LIWC 93 measures were used for analyzing the cognitive, affective, and grammatical processes in the text. To examine the difference between the factual and misleading news writing style, the LIWC features are divided into four categories: Linguistic, Punctuation, Psychological, and Summary [44].

Linguistics features refer to features that represent functionality of text such as the average number of words per sentence and the rate of misspelling. Thus, total function words as well as negations under this category were chosen.

Punctuation features are used to dramatize or sensationalize a news story which can be analyzed through types of punctuation used in the news such as Periods, Commas, Colons, Semicolons, Question marks, Exclamation marks, Dashes, Quotation marks, Apostrophes, Parentheses, and Other punctuation.

Psychological features target emotional, social process and cognitive processes. The affective processes (positive and negative emotions), social processes, cognitive processes, perceptual processes, biological processes, time orientations, relativity, personal concerns, and informal language (swear words, nonfluencies) can be used to scrutinize the emotional part of the news.

Summary features define the frequency of words that reflect the thoughts, perspective, and honesty of the writer. It consists of Analytical thinking, Clout, Authenticity, Emotional tone, Words per sentence, Words more than six letters, and Dictionary words under this category.

4.1.3 Readability

Readability measures how easily the reader can read and understand a text. Text complexity is measured by using attributes such as word lengths, sentence lengths, and syllable counts. The popular readability measures were used in the analysis: Flesh Reading Ease, Flesh Kincaid Grade Level, Coleman Liau Index, Gunning Fog Index, Simple Measure of Gobbledygook Index (SMOG), Automatic Readability Index (ARI), Lycee International Xavier Index (LIX), and Dale-Chall Score. Higher scores of Flesch reading-ease indicate that the text is easier to read and lower scores indicate difficult to read. Coleman Liau Index depends on characters of the word to measure the understandability of the text. The Gunning Fog Index, Automatic Readability Index, SMOG Index, Flesh Kincaid Grade Level are algorithmic heuristics used for estimating readability that is, how many years of education is needed to understand the text. Dale-Chall readability test use a list of words well-known for the fourth-grade students (easily readable words) to determine the difficulty of the text.

4.2 Image Features

To analyze the image associated with the news, the state-of-the-art deep-learning based technique were used to extract features from the images.

4.2.1 NeuralTalk2

NeuralTalk2 is an efficient image captioning model, coded in Torch that runs on GPU. It is similar to the original NeuralTalk, but this model implementation is batched, uses Torch, runs on a GPU, and supports CNN finetuning. All of these together result in quite a large increase in training speed for the Language Model (~100x). NeuralTalk2 model is based on a novel combination of Convolutional Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding. Then they designed a Multimodal Recurrent Neural Network architecture that uses the inferred alignments to learn to generate novel descriptions of image regions. The models can be trained using *loadcaffe* using VGGNet. But in this project, the pretrained model checkpoint were used to extract a caption describing the image content from the images associated with the news by using NeuralTalk2 [45], a pre-trained recurrent neural network that summarizes the content of an image in a one sentence description. After that a tf-idf to represent the text from these captions were computed and considered as an additional set of features in the analysis.

4.3 Source Bias

Several studies in the field of journalism have theorized a correlation between the political bias of a publisher and the trustworthiness of the news content it distributes [6, 9]. To validate this assumption, the relationship between the political bias of a news source and its credibility were examined by analyzing the information about 1,785 publishers in the MediaBias/FactCheck dataset. Figure 2 shows the distribution of the credibility score per political bias category (from Left to Right) and the bias distribution in the questionable sources. The plots show that when the news source is moderately to strongly biased (either conservative or liberal), then the source is more likely to publish misleading news than other news sources that are more moderate and declared as left-centered, right-centered, or neutral. Also, the Extreme- right (or strongly conservative) is the predominant bias among the questionable sources. Thus, the news source bias was used as another feature in the experiments.

4.4 Social Network Features

Social network features give useful information about users' social network, how fast and deep the fake news propagate through these networks. Here we use the relationship between the publishers and news, news and users, users and users to derive the veracity of each of them. Then, the credibility score for a news is computing by modeling the problem as a Markov Random Field (MRF) where we use the loopy belief propagation (BP) algorithm [49] to conduct the inference. In general, the MRF approach treats each node as a random variable and in our problem, we have three types of nodes publishers, news, and users. The random variable for each node is represented as $p_i \in \{0,1\}$, $n_j \in \{0,1\}$, and $u_k \in \{0,1\}$ where 0 being not credible and 1 being credible and the output is a marginal probability $p(p_i)$, $p(n_j)$, and $p(u_k)$ quantifying the belief that a node i belongs to class p_i, node j belongs to n_j, and node k belongs to u_k. The prior probability of each node can be assigned and represented by the function $\emptyset(p_i)$, $\emptyset(n_j)$, and $\emptyset(u_k)$ that can be obtained from our datasets. Given its bias, $\emptyset(p_i)$ gives information about whether the publisher is credible or not. For example, for the Right bias, the prior probability that there is a higher chance that the publisher is not credible as shown in Figure 2. For the left bias, even though the probability of not credible is lower compared to right bias but the chance of being not credible is higher compared to other biased publishers. The questionable sources are mostly from non-credible sources and maximum percentage of publishers are extreme right biased.

With the lack of prior knowledge and information regarding the news and based on the ground truth, the assumption was made that the news can be 50% chance of being credible and 50% not credible. A new study from MIT [46] [10] proposed that the human nature is responsible for the rapid spread of fake news than the true credible news. The research work analyzed more than 100,000 news stories on Twitter as for how many total tweets were posted and re-posted, time to reach the magnitude of engagement, and verifying the account from where it is created. They have proved that the users who spread fake news had significantly fewer followers, followed fewer people and less active on Twitter. This study was used in this research work to infer the prior probability of the users in the network by computing the in-degree, out-degree. These features representing number of followers/followings were extracted using the graph mining library, SNAP [4], based on the network between the users. These individual level features are used to infer the credibility and reliability for each user spreading news in the social network.

The function ψ_{ij} is a hyper parameter that determines the conditional probability for each node and the credibility score can be measured for the edges using edge potential function. The below tables show the choice of the affinity matrix ψ, for $\varepsilon > 0$, this choice of ψ assumes the correlation between the nodes. Table 4.1. Shows that if the publisher is not credible then there is higher probability to publish fake news and low probability to publish real news. Likewise, if the news is not credible then there is high probability that the user spread that news is also not credible and low probability to spread good news as shown in Table 4.2. Similarly, if the user is not credible then there is higher probability that the neighbor/friend users are also not credible and the same can be perceived from Table 4.3.

Table 4.1: Edge Potential functions between the node publisher and news

$\psi(p_i, nj)$	0	1
0	$1-\varepsilon$	ε
1	ε	$1-\varepsilon$

Table 4.2: Edge Potential functions between the node news and users

$\psi(n_j, u_k)$	0	1
0	$1-\varepsilon$	ε
1	ε	$1-\varepsilon$

Table 4.3: Edge Potential functions between the node users and users

$\psi(u_k, u_k)$	0	1
0	$1-\varepsilon$	ε
1	ε	$1-\varepsilon$

CHAPTER 5

EXPERIMENTS

In this work, each group of features described in the previous section was used in input to a RandomForest classifier with 5-fold cross validation to compute the performance of these features in classifying factual vs. misleading stories. Results are reported in Table 5.3 according to Area Under the ROC curve (AUROC), F1- measure, and Average Precision (or area under the precision-recall curve) and discussed in the following. The class weighting was used to deal with class imbalance. The experiments also included classification using linear SVM classifier with L2 regularization (with 5- fold cross validation) and the results are reported in Appendix. The results from both the classifiers are compared and found that RandomForest classifier performed better with these features.

5.1 News Body Content

The first modality analyzed is the news body content. Here, tf-idf features achieves the best results (0.888 AUROC, 0.811 F1-measure and 0.781 average precision). Next, the LIWC features is the second-best group of features. Among them, the psycho-linguistic features are the most important groups of features, achieving comparable performances. After LIWC, the readability features do not seem to separate well misleading news from factual ones in this dataset. The misleading news have higher frequencies of psychology related words such as personal concerns (death), relativity (motion), social (family and affiliation), and biological processes. The language used has more tentative words evoking uncertainty, more informal and more swear words. In contrast, factual news is harder to understand (higher Flesh Kincaid and Gunning Fog values), have higher risk related words, less anger, and more sad words. There are more parentheses in factual news which were used to indicate additional content providing more evidence of the news.

5.2 News Headline

Among all the features considered to analyze the news headline, it is shown that all the LIWC features combined is performing the best according to all the measures (e.g., AUROC of 0.791) and after that tf-idf with AUROC of 0.733 works better. Regarding punctuations, misleading headlines have more occurrences of the dot, exclamation mark, and semi-column (which may indicate they are packing many sentences in the news title). According to readability level, factual headlines are more complex to understand and show a higher Flesch-Kincaid score compared to misleading ones which have more tentative words evoking uncertainty, more informal, and more swear words as seen in the news body content. Overall, the analysis shows that factual political news headlines are more professionally written compared to misleading one.

5.3 News Source Bias

The news source bias is a strong predictor for news credibility as it achieves an AUROC of 0.884, 0.917 of average precision and F1-measure of 0.854. This result further confirms the correlation between source bias and the credibility of the news it distributes. It is worth noting that the publisher information is independent on the news labels as the former is collected from MediaBias/FactCheck, while the latter from Buzzfeed and Politifact.

5.4 News Image

The image associated with the news were used to determine the news validity and found that the tf-idf features of image caption from NeuralTalk2 performs better and is comparable to news headline (0.743 of F1-measure, AUROC of 0.600, and average precision is 0.725). Through a manual analysis of the images included in our datasets (see Figure 3 for examples), trends in the images used in misleading news and real news became apparent. One such trend was that real news articles included significantly more images focused on a figure speaking whereas the misleading news articles contained more images of people with only expressions on their faces. Further, the images in real news portrayed more positive impressions than misleading news. A final note from the manual inspection of our datasets was that the misleading news images were more likely to have been photoshopped by placing two images together and such images were of lower quality than the images from the real news datasets.

Picture 1: Examples of images associated with misleading (top) and factual (bottom) news.

<u>5.5 News Social Network</u>

The social network features were used to compute a credibility score for each news based on how users in the social network are sharing the news. In order to obtain the useful features from the social network, the prior probability for the news was set as 50% credible and 50% not credible because of the lack of prior knowledge about the news.

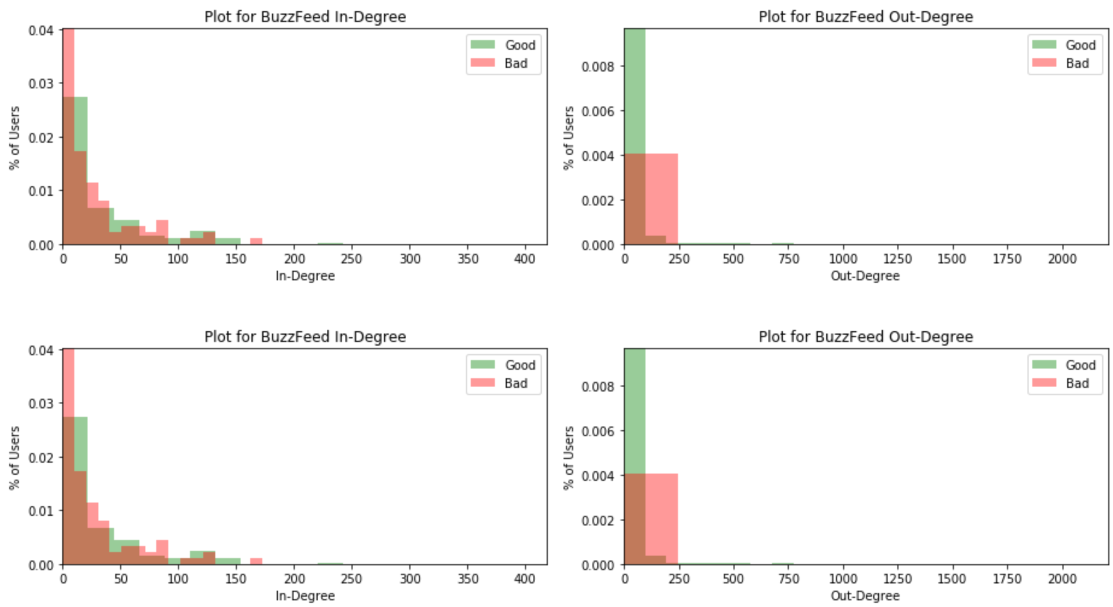

Figure 3. In-degree & Out-degree and percentage of users' distribution

The prior probability for each user was derived based on in-degree, out-degree, and combining both degrees. The indegree and outdegree against the percentage of users in each category was plotted. We plotted with all the users and also with users who spread news more than 6, 7, 8, 9, and 10 and the threshold to get the prior probability for each user was set based on those histogram plots. The Figure 3 shows the indegree, outdegree and percentage of users' distribution with the users who spread news greater or more than 8. We tried two different prior probabilities for publishers in our experiments, one with derived from source bias and another with default prior as 50% for both being credible and not credible.

The next step is to use sparse-matrix belief propagation [48], a modified form of loopy belief propagation that encodes the structure of a graph with sparse matrices to infer the credibility score for each node. In this case, the focus is on pairwise markov random fields that approximates the posterior marginals of each node. The different possibilities were tried and tested such as combining bias and in-degree, out-degree, both the degree, etc., by changing their prior probabilities to compute the credibility score using loopy belief propagation. The experiments were also conducted with all the users and also with users who spread the total news count greater than 6,7,8,9, and 10, to observe whether the users simply creating noise by just spreading only one or two news. The metrics used to compute the credibility score was AUROC and average precision and the results for the user who shared more than 6 and 7 news are shown in Table 5.1 and 5.2 for BuzzFeed and PolitiFact dataset. From this experiment, it is evident that the credibility score for users with different news count was almost same with minor difference, and both indegree and outdegree features are working for BuzzFeed dataset and indegree feature gives better results for Politifact. The credibility score improves when we use publisher bias, but to combine with other modalities we used the credibility score with the combination of total news greater than 8 and both the degree for Buzzfeed that gives better results (0.722 average precision, 0.636 AUROC), and indegree for Politifact (0.513 average precision, 0.560 AUROC) to avoid redundancy. Overall with all the modalities the results achieved are 0.880 AUROC, 0.858 F1-measure, 0.779 average-precision, that shows that exploiting social network features will definitely improves in detecting misleading news. The credibility score is also computed with 50% default prior probability for all the three nodes as shown in the table.

Table 5.1 Credibility Score for features with total news count spread by the user greater than 6 and 7 (BuzzFeed).

Features	AUROC > 8	Avg.Precision > 8	AUROC > 9	Avg.Precision > 9

Indegree	0.584	0.604	0.603	0.638
Outdegree	0.415	0.503	0.397	0.493
Both degree	0.636	0.722	0.648	0.738
All Default	0.500	0.506	0.500	0.506

Table 5.2 Credibility Score for features with total news count spread by the user greater than 6 and 7 (PolitiFact).

Features	AUROC > 8	Avg.Precision > 8	AUROC > 9	Avg.Precision > 9
Indegree	0.560	0.513	0.558	0.511
Outdegree	0.441	0.446	0.442	0.448
Both degree	0.440	0.446	0.441	0.448
All Default	0.500	0.500	0.500	0.500

Table 5.3: RandomForest Classification results with multi-modal features.

Features	F1	AUROC	Avg.Precision
News Content			
TF-IDF	0.811	0.888	0.781
Readability	0.642	0.682	0.629

Punctuation (LIWC)	0.704	0.766	0.636
Linguistic (LIWC)	0.719	0.787	0.650
Psychological (LIWC)	0.711	0.799	0.646
Summary (LIWC)	0.673	0.725	0.604
All LIWC	0.761	0.836	0.691
All News Content	0.848	0.874	0.771
News Headline			
TF-IDF	0.663	0.733	0.644
Readability	0.539	0.560	0.565
Punctuation (LIWC)	0.644	0.727	0.644
Linguistic (LIWC)	0.660	0.725	0.605
Psychological (LIWC)	0.635	0.676	0.574
Summary (LIWC)	0.657	0.705	0.600
All LIWC	0.722	0.791	0.654
All Headline	0.845	0.816	0.752
Image			
NeuralTalk2	0.743	0.600	0.725
Bias	0.854	0.884	0.917
Social Network	0.738	0.627	0.731
All	0.858	0.880	0.779

5.5 Do we need to "Read"?

To address this question, "do we need to look at the news body content?", we can refer to Table 5.4. The headline, bias, image features, and social features are combined to see if it further improves misleading news detection. Results show that specific set of features are effective for categorizing political news articles as factual or not. The feature bias plays a crucial part in detecting misleading news and the second most important feature is news headlines. Conversely, Horne and Adali [11] showed that the news headline is more informative than the body content (78% vs. 71% of accuracy). The results show that instead of "reading" the news article to figure out its validity, considering the metadata of news such as headline, bias, social network, and image can achieve comparable or even higher performances (0.90 AUROC vs. 0.88). Thus, looking at the news snippet by considering the headline characteristics, checking the publisher bias and headline keywords, and putting more attention on the associated images provides efficient tools for detecting misleading news. If these signals can be thought to humans, we can hopefully prevent people from massively spreading non-factual news through online social media.

Table 5.4: F1-measure, AUROC, and average precision results with combination of bias, headline, image, and social features.

Features	F1	AUROC	Avg.Precision
Headline + Content + Bias + Image + Social	0.858	0.880	0.779
Headline + Bias + Image + Social	0.865	0.901	0.786
Headline + Content + Bias + Image	0.860	0.879	0.777

Headline + Content + Bias + Social	0.854	0.874	0.772
Headline + Content + Image + Social	0.854	0.826	0.766
Content + Bias + Image + Social	0.864	0.898	0.785
Headline + Bias + Image	0.858	0.895	0.778
Headline + Image + Social	0.852	0.828	0.757
Headline + Bias + Social	0.867	0.896	0.790
Bias + Image + Social	0.871	0.879	0.846

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

In this project, the relative importance of different news modalities (body, headline, source bias, visual content, social network) were analyzed in detecting misleading political news. In particular, the source bias has never been analyzed before, and our findings demonstrate a strong correlation between political bias and news credibility. Moreover, it is proved that it is not necessary to analyze the news body to assess its validity (which may be time-consuming for the users), but comparable results can be achieved by looking at alternative modalities including headline features, source bias, and visual content.

One of the main limitations is for sure the size of the dataset considered, but there are no other currently available datasets containing all the information about the four considered modalities. Thus, collecting a bigger dataset will be helpful to refine our analysis as future work. Moreover, by extracting the sentiment from the news images one can achieve better performance in analyzing misleading news, as the manual inspection of the images in the dataset showed that images associated with misleading news are more emotional than the ones of factual news. Also, we would like to test the cross-domain efficiency of alternative news modalities as this has only been investigated for news body content so far [18]. From the social network, we analyzed by estimating the user credibility and network-based features such as diffusion network. It would be helpful to achieve better results if we analyze user-based features such as user profiles, user opinions and also post-based features represent users' social response in term of stance, topics, or credibility etc. Moreover, it's worth to explore effective features and models for early fake news detection, as fake news usually evolves very fast on social media.

REFERENCES

[1] [n.d.].MediaBias/FactCheckApps/Extensions. https://mediabiasfactcheck.com/ appsextensions/

[2] Pen America. 2018. Faking News: Fraudulent News and the Fight for Truth. https://pen.org/faking- news/

[3] 2016 Gallup poll - http://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx

[4] http://snap.stanford.edu/index.html

[5] https://www.factcheck.org/2016/11/how-to-spot-fake-news/.

[6] Vandalism in Wikipedia. http://en.wikipedia.org/wiki/Wikipedia:Vandalism.

[7] Anu Shrestha, Francesca Spezzano, and Maria Soledad Pera. Who is really affected by fraudulent reviews? an analysis of shilling attacks on recommender systems in real-world scenarios. In Late-Breaking Results track part of the Twelfth ACM Conference on Recommender Systems (RecSys'18), 2018.

[8] Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In COLING 2016, 26th Interna- tional Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan, pages 2438–2448, 2016.

[9] Pew Research Survey - http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

[10] Vosoughi, et al. (2018). "The spread of true and false news online." Science 359, no. 6380 (2018): 1146-1151.

[11] Shearer and Gottfried. 2017. News Use Across Social Media Platforms 2017.(2017).

[12] Craig Silverman. 2016. This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed News 16 (2016).

[13] Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on Twitter? ACM SIGMET- RICS Performance Evaluation Review 44, 1 (2016), 179–192.

[14] Joseph Kahne and Benjamin Bowyer.2017.Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. American Educational Research Journal 54, 1 (2017), 3–34.

[15] Paul Resnick,R Kelly Garrett,Travis Kriplean,Sean A Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. In Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM, 95–100.

[16] Verónica Pérez-Rosas,Bennett Kleinberg,Alexandra Lefevre,and Rada Mihalcea. 2018. Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics. 3391–3401.

[17] Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. arXiv preprint arXiv:1804.08559 (2018).

[18] Manish Gupta,Peixiang Zhao, and Jiawei Han.2012.Evaluating event credibility on twitter. In Proceedings of the 2012 SIAM International Conference on Data Mining. SIAM, 153–164.

[19] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo.2017.Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 2017 ACM on Multimedia Conference. ACM, 795–816.

[20] Seyedmehdi Hosseinimotlagh and Evangelos E. Papalexakis. Unsupervised content-based identification of fake news articles with tensor decomposition en- sembles. In MIS2: Misinformation and Misbehavior Mining on the Web Workshop held in conjunction with WSDM 2018 Feb 9, 2018 - Los Angeles, California, USA, 2018, 2018.

[21] Benjamin D Horne and Sibel Adali. 2017. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. arXiv preprint arXiv:1703.09398 (2017).

[22] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):211–36, 2017.

[23] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19, 1 (2017), 22–36.

[24] Potthast, et al. (2017). "A Stylometric Inquiry into Hyperpartisan and Fake News." arXiv preprint arXiv:1702.05638 (2017).

[25] Dongping Tian et al.2013. A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering 8, 4 (2013), 385–396.

[26] Sander Van der Linden, Anthony Leiserowitz, Seth Rosenthal, and Edward Maibach. 2017. Inoculating the public against misinformation about climate change. Global Challenges 1, 2 (2017), 1600008.

[27] Jooyeon Kim, Behzad Tabibian, Alice Oh, Bernhard Scholkopf, and Manuel Gomez-Rodriguez. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018, pages 324–332.

[28] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA., pages 2972–2978, 2016.

[29] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 849–857.

[30] MartinPotthast,JohannesKiesel,KevinReinartz,JanekBevendorff,andBenno Stein. 2018. A Stylometric Inquiry into Hyperpartisan and Fake News. In Proceed- ings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers. 231–240.

[31] Liang Wu and Huan Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018, pages 637–645, 2018.

[32] Natali Ruchansky, Sungyong Seo, and Yan Liu. CSI: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017, pages 797–806, 2017.

[33] Jin, et al. (2017). "Novel visual and statistical image features for microblogs news verification." IEEE transactions on multimedia 19, no. 3 (2017): 598-608.

[34] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2015. Vqa: Visual question answering. In Proceedings of the IEEE international conference on computer vision. 2425–2433.

[35] KaiShu, Suhang Wang, and Huan Liu.2017. Exploiting Tri-Relationship for Fake News Detection. arXiv preprint arXiv:1712.07709 (2017).

[36] Kai Shu, Suhang Wang, and Huan Liu. Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019, pages 312–320, 2019.

[37] Knauf Nathan Fairbanks James, Fitch Natalie and Briscoe Erica. Credibility assessment in the news: Do we need to read? In MIS2: Misinformation and Misbehavior Mining on the Web Workshop held in conjunction with WSDM 2018 Feb 9, 2018 - Los Angeles, California, USA, 2018, 2018.

[38] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. We used neural networks to detect clickbaits: You won’t believe what happened next! In Advances in Information Retrieval - 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings, pages 541–547, 2017.

[39] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. Clickbait detection. In Advances in Information Retrieval - 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016. Proceedings, pages 810–817, 2016.

[40] William Yang Wang.2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers. 422–426.

[41] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016, San Francisco, CA, USA, August 18-21, 2016, pages 9–16, 2016.

[42] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. Misleading online content: Recognizing clickbait as ”false news”. In Proceedings of the 2015 ACM Workshop on Multimodal Deception Detection, WMDD@ICMI 2015, Seattle, Washington, USA, November 13, 2015, pages 15–19, 2015.

[43] Symeon Papadopoulos Duc-Tien Dang-Nguyen Giulia Boato Michael Riegler Yiannis Kompatsiaris et al. Christina Boididou, Katerina Andreadou. 2015. Verifying Multimedia Use at MediaEval 2015. In MediaEval.

[44] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical Report.

[45] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2017. Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge. IEEE Trans. Pattern Anal. Mach. Intell. 39, 4 (2017), 652–663.

[46] https://techcrunch.com/2018/03/08/false-news-spreads-faster-than-truth-online-thanks-to-human-nature/

[47] Tanushree Mitra and Eric Gilbert. 2015. CREDBANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations. In Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015. 258–267.

[48] Bixler, Reid Morris. "Sparse matrix belief propagation." PhD diss., Virginia Tech, 2018.

[49] Chau, et al. (2011). "Polonium: Tera-scale graph mining and inference for malware detection." In Proceedings of the 2011 SIAM International Conference on Data Mining, pp. 131-142. Society for Industrial and Applied Mathematics, 2011.

[50] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. Detecting spammers on social networks. In ACSAC, pages 1–9, 2010.

[51] Srijan Kumar, Robert West, and Jure Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016, pages 591–602, 2016.

APPENDIX A

Title of Appendix A

Table A.1 Linear SVM Classifier results with multi-modal features

Features	F1	AUROC	Avg.Precision
News Content			
TF-IDF	0.818	0.875	0.791
Readability	0.585	0.639	0.596
Punctuation (LIWC)	0.671	0.708	0.606
Linguistic (LIWC)	0.684	0.729	0.620
Psychological (LIWC)	0.695	0.735	0.632
Summary (LIWC)	0.637	0.678	0.581
All LIWC	0.729	0.780	0.667
All News Content	0.812	0.773	0.794
News Headline			
TF-IDF	0.672	0.730	0.654
Readability	0.573	0.593	0.591
Punctuation (LIWC)	0.640	0.742	0.653
Linguistic (LIWC)	0.608	0.640	0.568
Psychological (LIWC)	0.607	0.628	0.573
Summary (LIWC)	0.551	0.555	0.529
All LIWC	0.675	0.720	0.626
All Headline	0.785	0.704	0.772
Image			
NeuralTalk2	0.721	0.670	0.761
Bias	0.843	0.878	0.890

Social Network	0.444	0.538	0.690
All	0.814	0.797	0.802

Table A.2: F1-measure, AUROC, and average precision results with the combination of bias, headline, image, and social features.

Features	F1	AUROC	Avg.Precision
Headline + Content + Bias + Image + Social	0.814	0.797	0.802
Headline + Bias + Image + Social	0.817	0.809	0.796
Headline + Content + Bias + Image	0.824	0.803	0.805
Headline + Content + Bias + Social	0.814	0.802	0.802
Headline + Content + Image + Social	0.789	0.749	0.788
Content + Bias + Image + Social	0.833	0.826	0.821
Headline + Bias + Image	0.846	0.825	0.827
Headline + Image + Social	0.780	0.720	0.815
Headline + Bias + Social	0.835	0.821	0.814
Bias + Image + Social	0.835	0.873	0.876