A Minimal Approach to Fake News Detection

Daniel Markusson

Abstract

The need for efficient categorization of fake and real media increases as the ubiquity of generative AI and motivated bad actors make producing fake news ever easier. Researchers have estimated that in 2021, $2.6 billion dollars of ad revenue can be attributed to misinformation publishing sites (Skibinski, 2021), providing ample motivation for the aforementioned bad actors to fabricate stories. This paper seeks to create an effective machine learning solution that gives readers the ability to classify articles they want to read as fake or real, enabling the consumption of solely accurate news. As users tend to prefer simple solutions, we provide a parsimonious model consisting of only 5 features, yet still able to achieve 71% testing accuracy. Among the most effective predictors of a real article is that of “perceived effort” - predicated by an article’s length, number of authors, and readability.

INTRODUCTION

Establishing a Problem

Fake news is, by its nature, very deceptive, making classification a difficult problem. An experiment conducted by Kumar et al. (2016) demonstrates this, presenting evidence that fake news writers succeed at their cozenage, and that humans are poor classifiers of false media. Subjects were shown one fake and one real Wikipedia article, and were asked to discern the fake one. They were able to successfully pick out the fake article only 66% of the time (Kumar et al., 2016). This demonstrates the difficulties one faces when attempting to classify fake and real media. How can a human consistently distinguish between media that was inherently designed to fool humans? Indeed, research finds that a majority of Americans admit they get at least some of their news from social media (Gottfried & Shearer, 2017), making it crucial to ensure that these sources are accurate and limit bias. Unfortunately, the motives of these social media news sources rarely align with maximizing accuracy and limiting bias, and instead align with producing sensational, or even completely fabricated pieces of news in order to drive readership and to grow profit (Banic & Smith, 2016).

Difficulty of the Problem

Further, the increased ability to mass-produce convincing fake media on account of advanced AI accelerates the issue. Long, convincing-sounding articles can be written by Large Language Models (LLMs), making their use tempting for any writer. Even if one is not trying to deceive their audience, the simple fact that an LLM was used can introduce fictitious details called hallucinations. In the context of LLMs, hallucinations describe the creation of fake facts in response to a prompt. In a study by Stanford University, LLMs were tasked with a series of legal queries. Their responses were then examined, finding that, on average, LLMs hallucinated facts about 75% of the time (Dahl et al., 2024). Other types of AI have also been used to fabricate stories, more specifically: Deep Fakes.

Deep Fakes are a form of AI trained on real images, video, or audio that are then able to subsequently generate fake images, video, or audio according to a prompt. Already, Deep Fakes have seen use in the spread of misinformation. Millions of people listened to a Deep Fake of the United Kingdom Labour Leader, Sir Keir Starmer, berating his employees (Sky News, 2023). It is unclear what proportion of listeners knew it was fake. Moreover, when defamatory campaigns are backed by nefarious nation states or political campaigns, the consequences can be catastrophic, especially in a “scenario where a false and damaging video or audio recording lands right before election day, without any opportunity to debunk it” (Butcher, 2024). An example of this is seen right before Slovakia’s general election on September 30, 2023 when an audio recording of Michal Šimečka, the Progressive Party leader, began circulating online, depicting him discussing plans to rig the election (Conradi, 2023). The audio was a Deep Fake. His opponent ended up winning the election.

How this problem is addressed

This report addresses the proliferation of fake news in two steps. First, by identifying key features that differentiate fake articles from real ones. This is done by visualizing the dataset using Empirical Distribution Functions (ECDFs). We can then create a machine learning model using the key features to categorize fake articles from real articles. As found previously, humans

are poor classifiers of fake news. This report proposes a simple tool that uses the basic characteristics of a news article (number of shares, number of authors, sentiment score, length and readability score) to identify fake articles from real ones. This gives users an accurate way to judge their internet consumption.

BACKGROUND

Definitions

Two types of individuals have been observed to proliferate fake news: complicit spreaders and non-complicit spreaders. Complicit spreaders are individuals who knowingly share fake media to achieve some goal like increasing a product's rating on an e-commerce platform or pushing a political agenda. Non-complicit spreaders are individuals who unknowingly share pieces of fake media, usually because they themselves were duped by the fictitious piece of media. Complicit spreaders commonly utilize bot accounts as puppets to spread false media en masse (Kumar et al., 2018). These bot accounts are either 1) created maliciously by a single bad actor or 2) authentic accounts that were somehow compromised.

Similarly, false information has also been placed into two broad categories: misinformation and disinformation. Misinformation is described as information that is “created without the intent to mislead” (Kumar & Shah, 2018, p. 2), and disinformation is described as information that is “created with the intent of misleading and deceiving the reader” (Kumar & Shah, 2018, p. 2). Kumar and Shah (2018) further divide these categories into opinion-based media and fact-based media. Opinion-based media consists of fake opinions. This includes product reviews on Amazon, where fictitious personal anecdotes are passed off as fact, leaving the reader with a false judgment. Fact-based media consists of lies about actual events. This includes the creation of fake political stories for the sole purpose of driving clicks and inciting emotion within the reader.

Platforms

The rise of “micro-blogging” (DeVoe, 2009) on sites like X (formerly Twitter), Tumblr, Plurk, and Threads has let people share opinions, thoughts, and ideas at an unprecedented rate (Kaplan & Haenlein, 2011). The term “micro-blogging” comes from a shared quirk of the aforementioned platforms, which, among other things, limit posts’ length to around 200 characters. Posts on these sites can be anything, ranging from heated political discourse to movie reviews. This leads many to dub these platforms ‘digital town squares,’ comparing them to a marketplace for exchanging ideas (Burgess, 2022).

Consequences of Micro-Blogging

Both good and bad have come from the proliferation of these sites, the root cause of both being the increased ability for the propagation of opinion. An example of a positive phenomenon described as “ambient journalism” has been observed (Hermida, 2010). According to Hermida (2010), “Traditional journalism defines fact as information and quotes from official sources, which have been identified as forming the vast majority of news and information content” (p. 297). Ambient journalism describes the shift away from traditional journalism, as described by Hermida, towards the interpretation of public interactions observed on these micro-blogging sites, giving journalists “early warnings about trends, people, and news” (Hermida, 2010, p.

302). On the other hand, polarization and hate speech are common trends within these sites, caused in large part due to fake news and disinformation (Vasist et al., 2023). Polarization characterizes the division and fragmentation of a population along a contrasting set of beliefs. These 'digital town squares' enable the discussion of such topics that cause polarization, leading to hate speech as discourse devolves into tribalism.

Propagation of Fake News on Social Media Sites

Bad actors have two main incentives for the creation of disinformation: generating revenue and influencing the public (Vasist et al., 2023). Clicks on fake news articles translate to tangible profit, demonstrated by the aforementioned $2.6 billion USD of ad revenue generated by misinformation publishing sites in 2021 (Skibinski, 2021). This creates a major incentive for bad actors to develop more and more sensational headlines that drive susceptible users to their articles. In the same vein, political campaigns are heavily incentivized to spread defamatory lies about their opponents in an attempt to influence voters (Allcott & Gentzkow, 2017).

As previously mentioned, propagation of fake news has been shown to cause ample amounts of polarization within social communities. Research has found that if only 15% of users within a network believe fake news to be true, polarization *consistently* arises (Azzimonti & Fernandes, 2023). Further, polarization begets polarization; consuming polarizing content is shown to cause one to trust the other side less and view them more negatively, creating what is known as an echo-chamber (Levendusky, 2013; Gao et al., 2023). Echo chambers have been observed on sites other than micro-blogging platforms, more specifically, short-form video content. Content on such sites has been observed to homogenize as algorithms on platforms such as TikTok, Instagram Reels, and Snapchat Reels optimize for what keeps users on their platform the longest. This effect spirals and has negative social and societal impacts, contributing to the spreading of lies, rumors, and misleading information.

What Makes Users Susceptible to Fake Media

Two traits that consistently predict one's susceptibility to fake media are confirmation bias and naive realism (Kumar et al., 2018). Naive realism occurs when one believes that their perception of reality is correct and that any challenges to their belief are objectively wrong. Confirmation bias describes the tendency for one to pursue media that confirms their preexisting beliefs. The amalgamation of the two traits causes consumers to find meaning in places that there may not be, for the simple fact that the piece of media conforms to their preexisting notions. For example, social conservatives may sympathize with posts demonstrating conservative viewpoints, and irrationally disagree with liberal ones.

RELATED WORK

Methods

The identification of fake opinion-based and fact-based media has been done by analyzing four features: text, users, graphs, and time.

Textual Analysis

Researchers have identified many ways to analyze the text of fake opinion-based and fact-based media to discern which ones are fake. In the context of opinion-based media like

product reviews, malicious actors may reuse the same review over and over to minimize effort (Kumar et al., 2016). Jindal et al. identified that 6% of reviewers have at least one review identical to someone else's, presenting evidence that many reviewers copy other reviews. Moreover, Sandulescu et al. identified that while some reviews may not be identical, they are semantically saying the same thing. For example, one review may say, "The service was horrible," while another may say, "The service was terrible."

In the context of fact-based media like news articles, researchers have identified that malicious actors try to pack as much information in the titles of fake articles while using simpler and, more often than not, capitalized words. This is done for three reasons: it generates emphasis and emotion, catching the user's attention; it allows users to skip reading the entire article and form conclusions by just reading the title, which may be very different from the rest of the article; and it makes them easier to read, using simple words to lower the barrier to entry. Pérez-Rosas et al. identified that fake articles typically have more social words and are more focused on present and future issues.

User Analysis
Popular sites online that use a review system for their products (e.g., Amazon, eBay, Yelp) typically use a five-point scale, one being the worst possible score and five being the best possible score. Shah et al. have found that fake reviewers online create reviews that skew heavily to either one or five - very positive or very negative - atypical of normal reviews. Further, research done by Kumar et al. found that fake reviews create reviews that are very different from a product's average review; they create positive reviews for products that have otherwise received negative reviews and negative reviews for products that have otherwise received positive reviews. Users who create fake fact-based media have been found by Kumar et al. to have young accounts with a lower number of written articles, suggesting that these are 'throw-away' accounts - temporary online profiles used for one-time purposes, often for privacy or anonymity.

Graph Analysis
Unreliable information gets referenced less than reliable sources do. Using network graphs, Kumar et al. measured the connectedness of fake and real Wikipedia articles along with calculating the average clustering coefficient of each subnetwork created by the local reference network of each article. They find that fake articles have both fewer references and a smaller clustering coefficient when compared to real articles, revealing that malicious authors add hyperlinks to other articles in their articles to appear genuine, while genuine authors add hyperlinks to other articles by necessity. A network graph can also represent user-user follower interactions on social media sites. Subrahmanian et al. identified that bot accounts on *X* (*Twitter*) that spread false information are closer together and appear in groups on such a network graph, suggesting that they have similar followers, and follow similar people.

Temporal Analysis
Researchers have found that many fake reviews are created at the same or similar times, as a large group of reviews are made in bulk and propagated all at once (Kumar & Shah, 2018). This can be explained by the use of scripts for the creation of fake news. Research has found that measuring interarrival times (IATs) - the time between each subsequent review - is effective for

the detection of fake review spammers. Unlike regular users, who typically have large gaps between each review, spammers have very short IATs, typically less than 10 minutes (Bryan et al., 2015).

Fact-based media has been observed to be propagated very early in its lifecycle - before it is debunked. Zent et al. (2016) found that tweets containing unverified information generated more retweets than verified information. During a crisis, all users, even those affiliated with reputable news organizations, are found to propagate unverified information in its first few minutes. After the piece is fact-checked, however, the amount of interaction and sharing drops significantly.

DATASET

Source of Dataset & Analysis

The dataset used for this analysis is the “FakeNewsNet” dataset from Kaggle.com (Shu et al., 2018). The dataset provides four forms of information: article information, user-user interactions, and user-article interactions. Article information includes the title, body text, URL, images/video, author(s), publish dates, and, of course, whether or not the article is fake. User-user interactions consist of following interactions; for example, User A follows User B. User-article interactions describe the sharing behavior of users, and the proliferation of different articles.

Basic Dataset Counts

Table 1. Article counts

Fake Articles	Real Articles	Total
182	211	393

Table 2. Author counts

Authors that publish both fake and real articles	Authors that exclusively publish fake articles (Fake Authors)	Authors that exclusively publish real articles (Real Authors)	Total
11	77	158	246

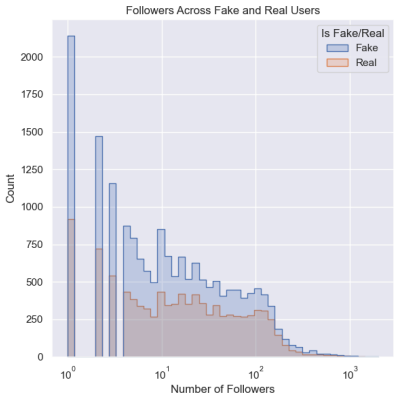
Table 3. User counts

Users that share both fake and real articles	Users that exclusively share fake articles (Fake Users)	Users that exclusively share real articles (Real Users)	Total
7043	18085	10193	35321

As can be seen in Tables 1 through 3, there are 29 more real articles than fake articles, which could potentially hurt an ML model. There are many more authors that publish exclusively real articles (hereafter called real authors) than there are authors who publish exclusively fake ones (hereafter called fake authors), while the fewest are authors who publish both. This may serve as a potential feature in a ML model, where fake and real authors serve as a potential feature. Finally, one can see in Table 3 that the largest subset of users are users who exclusively share fake articles (hereafter called fake users).

User-User Analysis

A potential reason for the large number of fake users is highlighted in Figures 1 & 2, which show that fake users have less followers than real ones, and that over 2000 fake users have zero followers. This could be an indicator of bot accounts that spread disinformation.



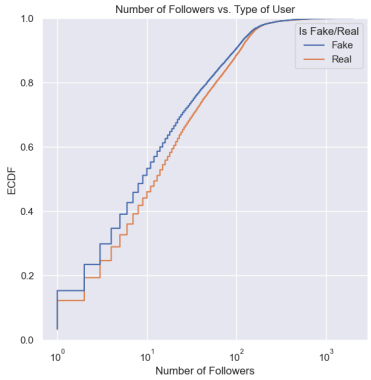


Figure 1 (left). Histogram of followers for fake and real users.
Figure 2 (right). Empirical cumulative density function of followers for both fake and real users.

Propagational Analysis

As demonstrated in Figure 3, fake articles are propagated a lot more than real articles are. On average, fake articles received 179.88 shares while real articles only received 98.02 shares. This 87.86 share difference is explained by the sensational nature of these articles; they entice users to share the emotional messages contained within the articles to people they know, whether it is out of fear, anger, or some other extreme emotion that the article attempts to forge. Moreover, this concurs with the evidence of bot accounts seen in Figures 1 and 2; the difference in the number of shares between fake and real articles could be at least partially accounted for by the bots' actions.

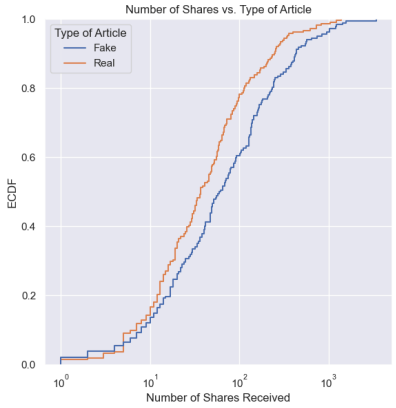


Figure 3. Empirical cumulative density function of the number of shares for each article, categorized by the validity of each article.

Author Analysis

The number of authors an article has is a useful feature that helps differentiate fake and real articles, seen in Figure 4. The number of authors was measured based on the number of *displayed* authors within an article, thus the majority of articles had zero authors, for the simple fact that most articles do not display their authors. However, of the articles that do, the majority of them are real, making this a helpful factor when determining the validity of an article. In fact, Figure 4 demonstrates that having simply one author displayed within the article is a strong indicator of validity.

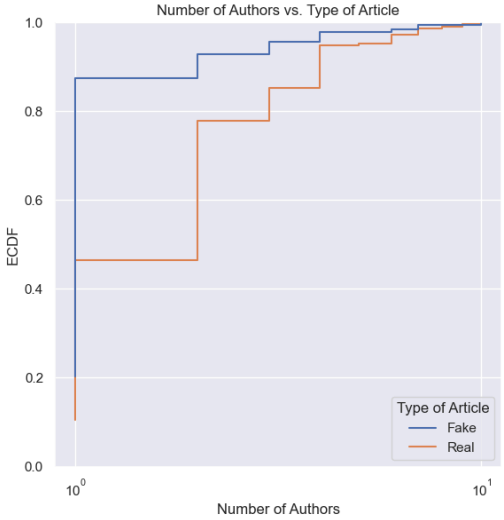
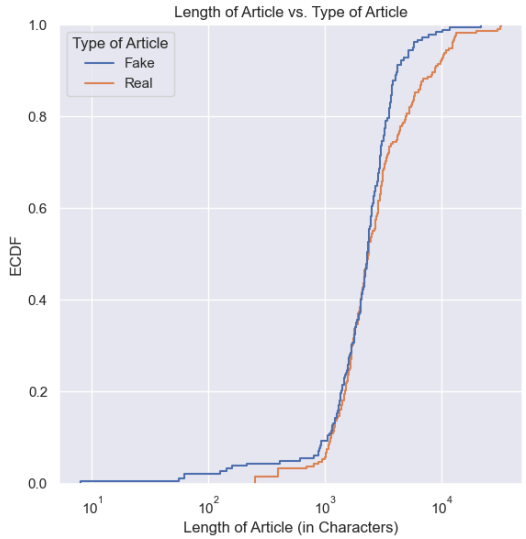


Figure 4. Empirical cumulative density function of the number of authors articles have, categorized by validity of each article.

Textual Analysis

The creation and message of fake and real articles are vastly different, substantiated by two factors: article length and textual sentiment. Although a simple metric, the difference in the typical lengths of fake and real articles can be used as a feature in a machine learning model, as demonstrated in Figure 4. Extremely short articles - less than about 350 characters - are very

likely to be fake. Medium length articles - greater than 350 characters but less than 4,000 characters - have an approximately equal chance to be fake or real. Long articles - greater than 4,000 characters - are likely to be real. This perpetuates an overarching theme: the more demonstrated effort put into an article, whether it be the number of authors or sheer length, the more likely it is to be a valid article.



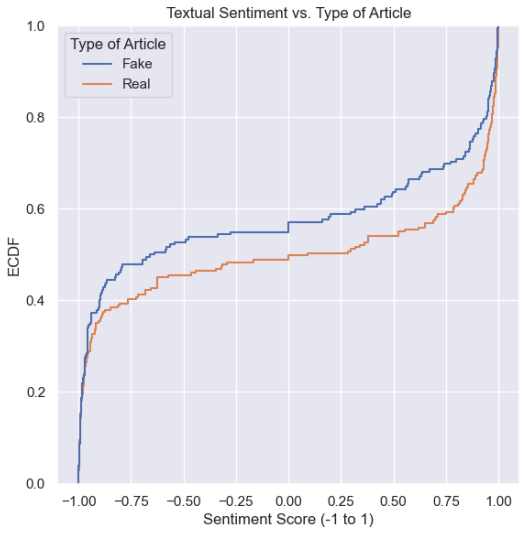


Figure 5 (left). Empirical cumulative density function of the lengths of articles, categorized by the validity of each article.
Figure 6 (right). Empirical cumulative density function of the sentiments of articles, categorized by the validity of each article.

Furthermore, the textual sentiment of articles can, rather simply, be analyzed, providing a useful insight on the validity of articles. Textual sentiment is an objective rating of the emotion demonstrated within a string of text. It is done through a dictionary, where each key value pair is a word and a corresponding score pertaining to the word's emotional connotation; lower scores for more negative words and vice versa. The function first pre-processes the text by tokenizing, removing stop words, and lemmatizing, and then it iterates through the array of tokens, summing each word's individual score. Punctuation is then taken into account, adding or subtracting a constant from the overall score. Finally, the score is normalized to be between -1 and 1. For example, the phrase, "I had such a great time today!" is first pre-processed into: "great time today !". The words are then iterated through, with 'great', 'time', 'today', and '!' having sentiment scores of 0.6249, 0, 0, and 0, respectively. Finally, taking into account the exclamation mark and normalization, the final score of the overall phrase is 0.6588. As demonstrated in Figure 5, fake articles seemingly receive a lower sentiment score than real articles, indicating that fake articles tend to surround more negative topics, potentially attempting to tug on the emotions of readers. Negative topics may strike fear into readers, leading them to believe that reading an article is a more crucial matter than it actually may be.

The final textual feature that is useful in the differentiation of fake and real articles is the readability of the article. There are various methods one may use to determine a body of text's readability, and each saw discernible differences between fake and real articles; however, the

Dale-Chall Readability Index (DCRI) was found to produce the best results. The DCRI calculates readability according to what a reasonable student would be able to read, seen in Table 4 (Chall & Dale, 1995).

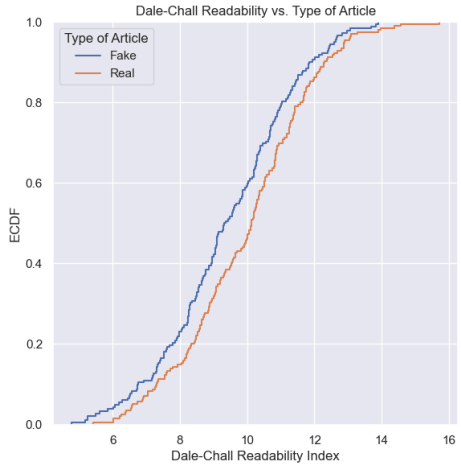


Figure 7. Empirical cumulative density function of Dale-Chall Readability Index of fake and real articles.

Table 4. Dale-Chall Readability Score interpretation table (Dale & Chall, 1948).

DCRI Score	A student at this grade could easily read this piece of text
4.9 or less	≤4
5.0 - 5.9	5 - 6
6.0 - 6.9	7 - 8
7.0 - 7.9	9 - 10
8.0 - 8.9	11 - 12
9.0 or greater	>12

This is done using the following formula, where difficult words are considered words not found in a set of 3,000 familiar words:

$$0.1579\left(\frac{difficult\ words}{words} \times 100\right) + 0.0496\left(\frac{words}{sentences}\right)^{1}$$

As seen in Figure 7, fake articles tend to be a lot more readable than real articles, further perpetuating the ideal that demonstrated effort is a powerful indicator of validity.

[1] If the proportion of difficult words is greater than 0.05, add 3.6365.

APPROACH

Features

The machine learning model used the five features found to have the highest ability to differentiate between fake and real articles, while also being simple and easy to procure: the number of shares, the number of displayed authors, the article's textual sentiment score, the article's readability, and the length of the article in characters. These features are all ones that can quantify the effort the writer put into the article - demonstrated effort - which has been seen to be a consistent theme surrounding what differentiates fake and real articles.

Model

The features were then used to train an XGBoost (eXtreme Gradient Boosting) machine learning model, outlined in Figure 8.

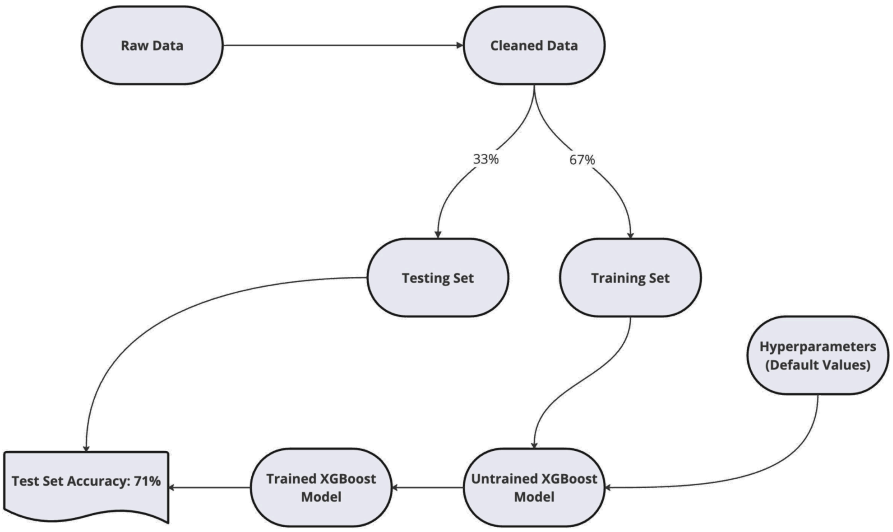


Figure 8. Outline of training process of the XGBoost machine learning model, along with the model's test set accuracy.

XGBoost is a popular gradient boosting model released in 2014. Gradient boosting is a machine learning method that involves training a series of submodels based on the residuals of the preceding model, trying to get the residuals to as close to zero as possible, with the first, primary submodel being trained on a random sample of the training data. Each iteration of training naturally involves a separate random sample of training data, giving each submodel the opportunity to generalize. The number of submodels can be tuned to reduce overfitting within the entire machine learning model. In layman's terms, gradient boosting involves an initial estimate based solely on the input data, with a series of submodels who receive the guess, and then estimate how far away they are based on their trained biases. Therefore, the output of the overarching model must be a sum of each estimate by each submodel. This is further illustrated in Figure 9.

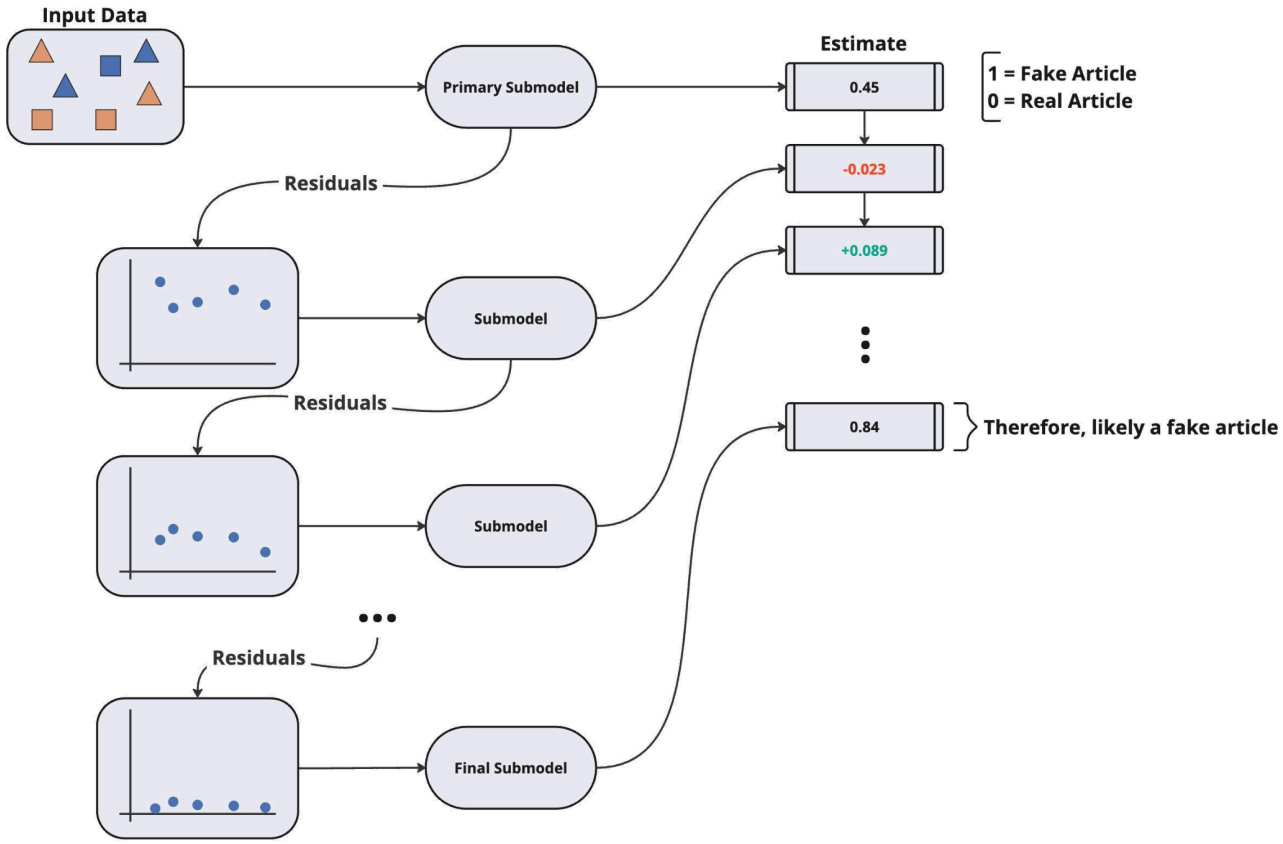


Figure 9. A gradient boosting machine learning model.

Evaluation

The XGBoost model proved to be a somewhat effective classifier of fake and real news. With a test set accuracy of 71%, the ML model proves to be more effective than the human baseline of about 66%.

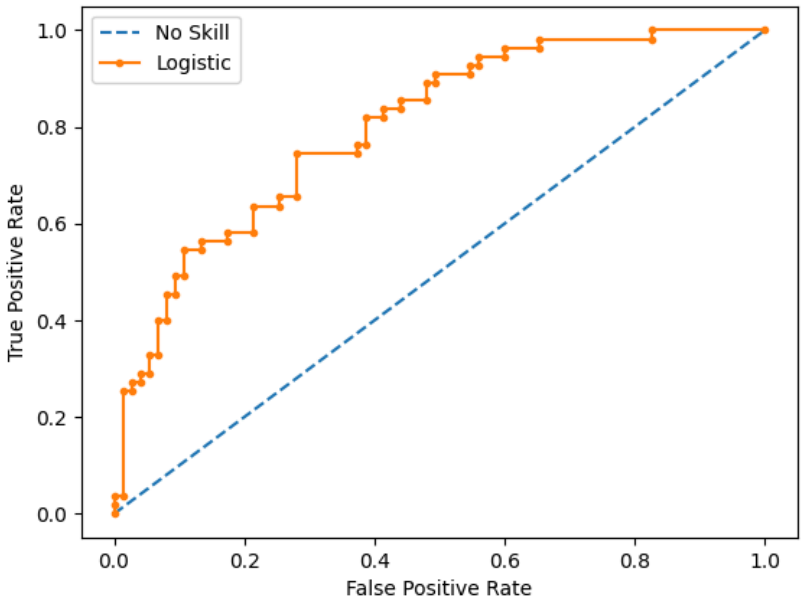


Figure 10. Receiver operating characteristic curve of the XGBoost machine learning model.

A receiver operating characteristic curve (ROC curve) is used to assess how well a classifier performs by plotting its true positive rate (sensitivity) against its false positive rate (1 - specificity) at various decision thresholds. Using python's scikit-learn library this can be done trivially, giving a deeper understanding of the model's performance. An ROC curve closer to the top left indicates a better performance by the classifier, as that is where the true positive rate is highest, and the false positive rate is lowest. When the decision threshold is varied, changes in the true positive and false positive rates can be observed; for example, if the decision threshold is lowered, the true positive rate increases while also increasing the false positive rate, and vice versa. How these graphs should be interpreted is using the Area Under the Curve (AUC) logistic, and for this ML model, the AUC logistic was 0.800, suggesting a strong ability to distinguish between the two classes.

Another way of analyzing the performance of a machine learning model is by using a confusion matrix. A confusion matrix is a square matrix where rows typically indicate true classes and columns typically indicate predicted classes. Each cell (i, j) within the matrix reveals how many articles from class j were predicted as i. Figure 11 is a confusion matrix for the XGBoost model created based on the aforementioned features. Within the context of fake news classification, there are consequences for both false positives and false negatives, but it can be argued that false negatives are more consequential than false positives as one would rather not believe a real article than believe a fake article. Thus the model's relative tendency to do more false positives can be better than the alternative.

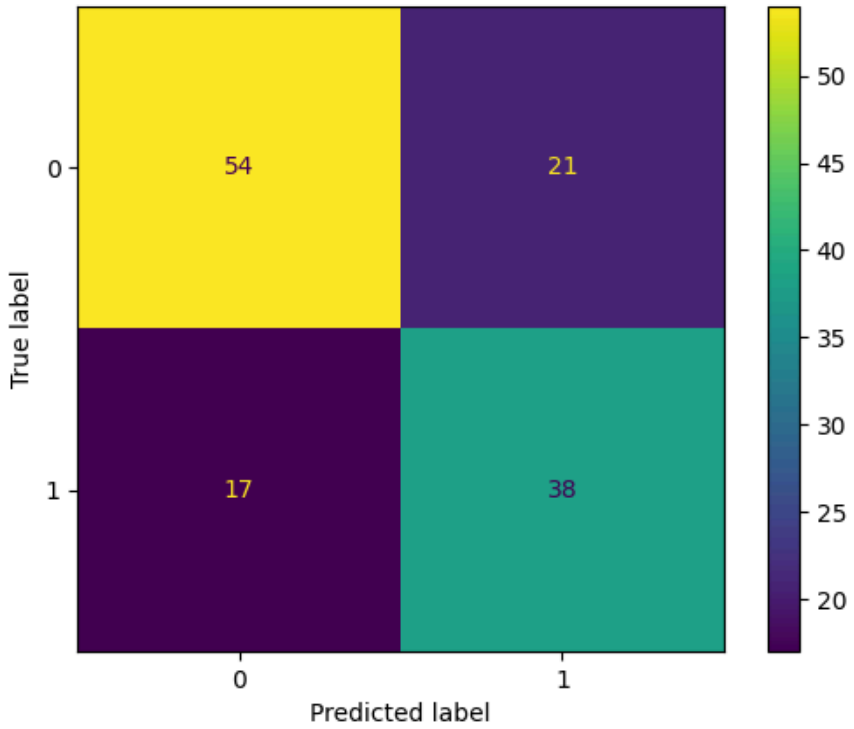


Figure 11. Confusion matrix of the XGBoost machine learning model. On the axes, 1 indicates a fake article and 0 indicates a true article.

The model seems to fail on articles that exhibit characteristics of the opposite class. For example, within the test set, there is a fake article with 220 shares, two authors, a sentiment score of approximately -0.5848, a length of 1714, and a readability score of approximately 10.2449. This article was incorrectly labeled as real by the model. The following is the article in its entirety:

> *FBI Issues Warrant For Obama's Arrest After Confirming Illegal Trump Tower Wiretap – Americas Last Line of Defense*
>
> *Former president and breaker of laws, Barack Obama, will either surrender himself or be picked up by the FBI sometime today to be booked and charged with unlawful use of authority, wire fraud and conspiracy to interfere with free elections after it was confirmed that he ordered the tapping of the phones at Trump Tower during the presidential election.*
>
> *The order, which isn't something even a president can do without the signature of a federal judge, was to listen in on Trump and his children to try to find a connection to Russia. Nothing came of it since President Trump nor any of his campaign staff have ever been to, spoken with or had anything to do with Russia or its agents.*
>
> *Trump first disclosed the bombshell in an early morning tweetstorm, knowing already that the FBI was preparing charges and asking a judge to sign a warrant for Obama's arrest. Todd McMartin, a spokesman for the FBI, told Fox News:*
>
> *"The proof is undeniable. Obama basically confessed in a private call to one of Hillary Clinton's aides that he had the Trump Tower tapped and we can't find any federal order legally authorized by a judge to do so."*
>
> *The call, between Obama and Huma Abedin, was intercepted by the FBI after President Trump ordered Obama's phones tapped to catch him in a lie over the Russia scandal. That tap was authorized by executive order for national security reasons. If convicted, Obama could face up to 40 years in prison, and no President will be pardoning him anytime soon.*[2]

For one who follows American politics, it is relatively obvious that this article is fake because it possesses many false facts. However, without looking at the lies, it is well written and has all the characteristics of a genuine article. This is likely why the machine learning model finds an article like this difficult to categorize. Textually and semantically, there is ample reason for the model to call this a real article.

[2] https://thelastlineofdefense.org/ has since been taken down; however, this article is still viewable through the Wayback Machine using this link: https://web.archive.org/web/20170520025842/http://thelastlineofdefense.org/breaking-fbi-issues-warrant-for-obamas-arrest-after-confirming-illegal-trump-tower-wiretap/

FUTURE WORK

Limitations

The current approach to differentiating fake and real news is somewhat effective, but false short of being a fool-proof method. As seen previously with the obviously fake article, the model takes no consideration of fact, and blindly follows the textual characteristics of the article, along with the shares and number of authors. This has the benefit of making this an easy-to-use and intuitive approach, but also has the detriment of failing to classify the aforementioned article despite being clearly fake. The solution to this would be to expand the scope of the features of the current model, and include some insight into what the actual article is trying to say, and then base that off of some factual database. The current approach is useful in its minimalism, but the training data could have been larger. While close to 400 articles were used, this is simply a miniscule drop in the bucket in comparison to all articles written on the internet, so getting a larger dataset to train the model on would doubtlessly improve its performance.

Future Work

Deep fakes are becoming an increasingly larger problem on social media platforms. Machine learning approaches to differentiating deep fakes from real images could potentially be vastly applicable on social media sites like instagram, where image sharing is such an integral part of their platform. Other generative AI like ChatGPT has also seen use within the classroom, raising concern among teachers. Machine learning could also help teachers differentiate student's work from AI, helping teachers protect their classroom.

CONCLUSION

We attempted to solve the growing issue of differentiating fake and real news using lightweight and minimal characteristics about the article. Features that attempt to describe the effort that the author put into making the article - demonstrated effort - seemed to be the greatest differentiator; features like length, number of authors, and readability seemed to have the greatest differences between fake and real articles, while also being easily obtained by a simple web scraper. Using the number of shares, the number of displayed authors, the article's textual sentiment score, the article's readability, and the length of the article in characters gave a 71% accuracy. The major limitation with this approach is that it has no way to fact-check what the article is saying, and has to merely rely on patterns observed within the aforementioned features. Despite this limitation, lightweight machine learning models have many uses, especially in areas where computing power is limited. Therefore, minimal approaches to fake news detection can aid many in their internet navigation.

REFERENCES:

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Azzimonti, M., & Fernandes, M. (2023). Social media networks, fake news, and polarization. European Journal of Political Economy, 76, 102256. https://doi.org/10.1016/j.ejpoleco.2022.102256

Banic, V. & Smith, A. (2016). Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies. NBC News. Retrieved from https://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451

Burgess, J. (2022). The 'digital town square'? What does it mean when billionaires own the online spaces where we gather? The Conversation. https://theconversation.com/the-digital-town-square-what-does-it-mean-when-billionaires-own-the-online-spaces-where-we-gather-18204

Butcher, S. (2024). 2024 may be the year online disinformation finally gets the better of us. Politico. Retrieved from https://www.politico.eu/article/eu-elections-online-disinformation-politics/

Chall, J. S., & Dale, E. (1995). Readability revisited. Brookline Books.

Conradi, P. (2023). Was Slovakia election the first swung by deepfakes? The Times. Retrieved from https://www.thetimes.com/world/russia-ukraine-war/article/was-slovakia-election-the-first-swung-by-deepfakes-7t8dbfl9b

Dale E; Chall J (1948). "A Formula for Predicting Readability". Educational Research Bulletin. 27: 11–20+28.

David, A. (2024, June 18). Misinformation might sway elections — but not in the way that you think. Nature. https://www.nature.com/articles/d41586-024-01696-z

Dawber, A. & Tomlinson H. (2023). Deepfakes of Donald Trump 'arrest' spread on social media. The Times. Retrieved from https://www.thetimes.com/business-money/technology/article/donald-trump-deepfakes-ai-twitter-g50n7vnbm

DeVoe, K. M. (2009). Bursts of Information: Microblogging. The Reference Librarian, 50(2), 212–214. https://doi.org/10.1080/02763870902762086

Editors at Sky News. (2023, October 9). Deepfake audio of Sir Keir Starmer released on first day of Labour conference. Sky News.

https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181

Gao, Y., Liu, F., & Gao, L. (2023). Echo chamber effects on short video platforms. Scientific Reports, 13(1), 6282. https://doi.org/10.1038/s41598-023-33370-1

Gottfried, J. & Shearer, E. (2017). News Use Across Social Media Platforms 2017. Pew Research Center. Retrieved from https://www.pewresearch.org/journalism/2017/09/07/news-use-across-social-media-platforms-2017/

Hermida, A. (2010). TWITTERING THE NEWS: The emergence of ambient journalism. Journalism Practice, 4(3), 297–308. https://doi.org/10.1080/17512781003640703

Hooi, B., Shah, N., Beutel, A., Gunnemann, S., Akoglu, L., Kumar, M., Makhija, D., & Faloutsos, C. (2015). BIRDNEST: Bayesian Inference for Ratings-Fraud Detection. https://doi.org/10.48550/arXiv.1511.06030

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, 219–230. https://doi.org/10.1145/1341531.1341560

Kaplan, A. & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. Business Horizons, 54, 105-113. https://doi.org/10.1016/j.bushor.2010.09.004

Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. In Proceedings of the 25th International Conference on World Wide Web, 591–602. https://doi.org/10.1145/2872427.2883085

Kumar, S., & Shah, N. (2018). False Information on Web and Social Media: A Survey.

Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., & Subrahmanian, V. (2018). REV2: Fraudulent User Prediction in Rating Platforms. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 333–341). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3159652.315972
9

Levendusky, M. (2013). Partisan Media Exposure and Attitudes Toward the Opposition. Political Communication, 30(4), 565–581. https://doi.org/10.1080/10584609.2012.737435

Matthew Dahl, Varun Magesh, Mirac Suzgun, & Daniel E. Ho. (2024). Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models. https://doi.org/10.48550/arXiv.2401.01301

Pérez-Rosas V., Kleinberg B., Lefevre A., & Mihalcea R. (2017). Automatic Detection of Fake News. https://doi.org/10.48550/arXiv.1708.07104

Sandulescu, V., & Ester, M. (2015). Detecting Singleton Review Spammers Using Semantic Similarity. In Proceedings of the 24th International Conference on World Wide Web. ACM. https://doi.org/10.48550/arXiv.1609.02727

Shah, N., Beutel, A., Hooi, B., Akoglu, L., Gunnemann, S., Makhija, D., Kumar, M., & Faloutsos, C. (2015). EdgeCentric: Anomaly Detection in Edge-Attributed Networks. https://doi.org/10.48550/arXiv.1510.05544

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv Preprint arXiv:1809. 01286. https://doi.org/10.48550/arXiv.1809.01286

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.48550/arXiv.1708.01967

Shu, K., Wang, S., & Liu, H. (2017). Exploiting Tri-Relationship for Fake News Detection. arXiv Preprint arXiv:1712. 07709. https://doi.org/10.48550/arXiv.1712.07709

Skibinski, M. (2021). Special Report: Top brands are sending $2.6 billion to misinformation websites each year. NewsGuard. Retrieved from https://www.newsguardtech.com/special-reports/brands-send-billions-to-misinformation-websites-newsguard-comscore-report/

Subrahmanian, V., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., & Menczer, F. (2016). The DARPA Twitter Bot Challenge. Computer, 49(6), 38–46. https://doi.org/10.48550/arXiv.1601.05140

Vasist, P. N., Chatterjee, D., & Krishnan, S. (2023). The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configural Narrative. Information Systems Frontiers. Advance online publication. https://doi.org/10.1007/s10796-023-10390-w