InterrogatorX
BY
IZBA ATIF

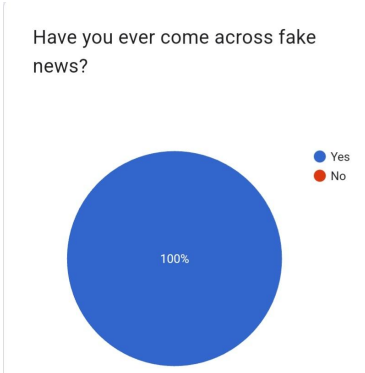
InterrogatorX is a fake news detection software. It is a simple html web page where the user can provide some information about the news article(such as title and body) as input and select the machine learning method and they will get back a webpage telling them whether the news is fake or not.

Introduction

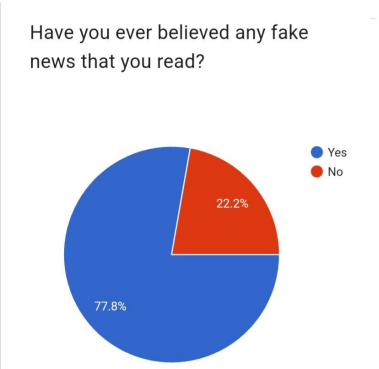
What is fake news? Fake news is misleading news information. It can be found on various social media platforms or personal blogs etc. But why is it important to identify fake news? In today's day and age, anyone can easily write a news article and publish it online and readers may believe it so this makes the detection of fake news a very important issue nowadays. Problems such as widespread panic may take place due to some fake news about certain topics such as politics. An example of a problem is when a shooting occurred due to a fake news spread earlier [1] . Fake news has also been weaponized in modern information warfare and there have been potential consequences to society and these have been discussed in [2].

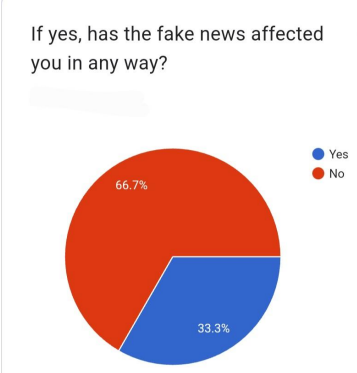
Survey Results

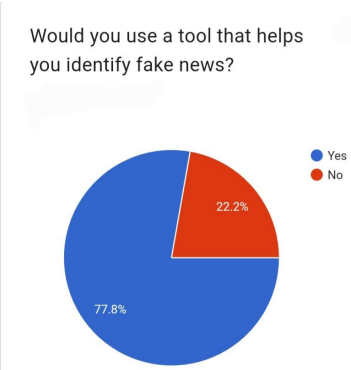
Before choosing the purpose of the tool being developed, it had to be made sure that it had a target audience and what the easiest tool to use was for the audience. So for these reasons, a survey was made and the results can be seen below..

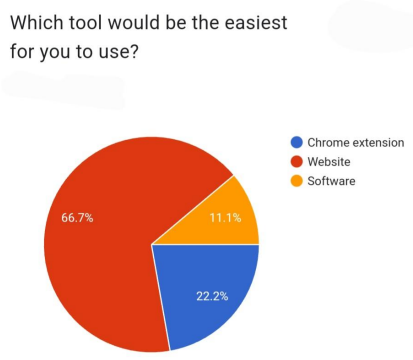


So as it can be seen all of the people who filled the survey have come across fake news and most of them have even believed it. The ratio of people who have been affected by it is not as much but exists nonetheless. And since most users wanted a website that can help them, I decided to make that.









Target Market

In 2019, a study was conducted and it showed that liking or sharing etc is one of the main causes of spreading of fake news and people who are more active on social media tend to share fake news more frequently. Also users who engage with fake news tend to be more politically polarised and less likely to fact-check information before sharing it. [3]

So this study supports the need for InterrogatorX and it can be used by several people for fact checking the news articles they read. The key amongst them may be the following:

- Media companies' work of validating an article before publishing can be made easy by using this software. Previously it was taking up time and human resources to read and verify an article but using InterrogatorX, their work could be cut short and they can get their results in a matter of seconds.
- A lot of pages on Facebook and Instagram nowadays publish any major news and people trust those news stories. So it is their ethical responsibility to check the news' authenticity before publishing and that's where InterrogatorX can be useful for them.
- Some individuals may be affected by the fake news story and take some actions which may harm them so they have the need to verify it before believing it and they can make use of our product. An example of a news story affecting individuals is dropping stock prices.

Similar Existing Products:

- The factual https://www.thefactual.com/index.html
  This is a similar chrome extension which identifies any fake news based on the author, diversity of sources in the news article etc. It also uses a machine learning algorithm. However, this does not take into account what is the data inside the news article as compared to the metadata so a fake news story coming from a trusted source may not be classified correctly.
- Check by Meedan https://meedan.com/check
  This tool is used to identify fake information in WhatsApp and Facebook messages. It also identifies fake information based on similarity between the information provided and the ones it has been trained on.
- Logically [6] https://www.logically.ai/
  This is a mobile app and browser extension which verifies both facts and images. AI is used for this verification. It can be used to check information on social media etc.

Why the need for InterrogatorX

Since tools such as these already exist in the market, why would someone use InterrogatorX instead of ones they already use and trust? InterrogatorX provides the user with the option to use different techniques to verify the news articles. Additionally the users can try all three techniques and based on their average results, come to the conclusion whether or not the news is fake.

Literature Reviews

In [4], the authors talk about how detecting fake news is a significant concern today because it can impact public opinion. This article includes a detailed survey on challenges related to detecting fake news etc. The authors have explored various approaches which are used for detecting fake news such as linguistic analysis, social network analysis, and machine learning techniques. The importance of data collection is talked about and the challenges associated with the quality and biases within the data. The limitations of the model are also discussed and future research direction.

The authors first start by talking about the impact of fake news in society and the need for detection methods. They emphasise that fake news detection is a complex task that requires insights from various fields such as natural language processing, social network analysis and data mining.

Then the researchers give and overview of linguistic analysis techniques used for fake news detection, such as lexical and syntactic analysis. They talk about how styling and grammar etc can help identify fake news. They also talk about the role of social network analysis in detecting fake news by examining some data such as user credibility, information diffusion patterns and network structure. It is also talked about how social media is a leading factor of spreading fake news as I talked about previously.

Additionally, machine learning methods are discussed in detail as being a popular method for detection of fake news. Various machine learning techniques are talked about including both supervised and unsupervised techniques. The importance of feature engineering is also talked about and how appropriate features have to be selected. Furthermore, the challenges that arise from machine learning models are also discussed. Examples of such challenges may be data bias or a dataset not being available.

The authors also talk about the need for high quality datasets for training the model. The problems associated with the datasets are discussed and how labelled datasets often suffer from the subjective nature of news categorization and the difficulty in obtaining ground truth labels for real-world news articles. They talk about how future research should focus on developing a diverse and representative dataset.

Lastly the authors talk about how fake news stories are always evolving so timely detection is crucial. They conclude the article with several research questions and future directions for fake news detection.

In [5] the authors also talk about social media being a huge cause of spread of false news and it may spread different types of emotions in people and thus the authors want to extract features based on sentiment analysis on news articles and emotional analysis on the reader's comments. The authors use a LSTM model to detect fake news and achieved an accuracy of 96.77%.

The authors start by talking about how fake news affects people's life and emotions and it has a 70% higher chance of spreading than real news. It is also discussed how the main source of news for a lot of people is just social media and they believe the information that they come across. They talk about how people's comments play a huge part in detecting whether the news is fake or not. It was found that users' replies to fake news carry the emotions of fear, disgust, and surprise, while users' replies to real news carry the emotions of anticipation, sadness, joy, and trust.

The authors then talk about the method they used in details and discussed their challenges and how they overcame those to finally be able to achieve a model with a high accuracy.

In [6] the authors explore the use of deep learning techniques for fake news detection. They talk about the growing issue of spread of fake news in this digital era and how deep learning models can play a huge part in identifying fake news articles from real ones.

The authors begin by talking about the characteristics of fake news and how it affects society, politics, and public opinion. The importance of reliable and accurate methods to address the problem of fake news detection is then talked about.The authors also talk about how deep learning is a promising solution to this problem.

The authors give an overview of various deep learning techniques that are or can be used for detection of fake news. They talk about the use of RNNs and CNNs for understanding the textual characteristics of news articles. The authors further explain how these deep learning techniques can learn sequential dependencies and get meaningful representations within text.

The authors further discuss the importance of the need of labelled datasets for training deep learning models. The challenges of obtaining large labelled datasets for fake news detection is also talked about and the authors came up with the solution to solve this which is creating a labelled dataset by combining multiple existing dataset and using crowdsourcing to gather data. Using this solution, the authors were able to obtain a more diverse dataset for training the model..

Furthermore, the authors stress on the importance of feature representation in deep learning models for fake news detection. They talk about the use of textual, social network and user based features which can be used to improve the performance of the deep learning model. They also talk about the importance of considering both, the content of news article and social context in which they are shared and consumed.

In [7], the author's research is based on fake news detection and in particular on the spread using social media. The authors talk about the challenges faced due to how quickly fake information is spread on social media and further investigates how data mining can be used to solve this problem.

The authors start with talking about the unique characteristics of fake news which is spread on social media and talk about how fake news articles often use sensationalism, emotional appeals, and misleading headlines to increase their audience and get more attention. This can make detection of fake news on social media difficult.

The authors talk about the different data mining techniques that can be used to detect fake news on social media platforms and further talk about feature-based approaches that can help identify features in the content of news articles such as writing style, metadata and linguistic patterns. Extraction of these features and comparing them to real news articles from trusted sources can help identify fake news.

Propagation pattern analysis is also talked about which basically focuses on the spread of fake news through social media. The authors discuss how examining patterns of information spread and user interactions make it possible to identify suspicious patterns related to the spread of fake news. The consider factors like the speed of spread and reach of the article etc for this approach. The involvement of bots and associated accounts also contribute to the spread of fake news.

The authors further talk about the challenges that arise with data reliability in fake news. Social media has a lot of spam, bots and fake accounts so this makes it important to ensure the quality and authenticity of data that is being used. They also discuss how to overcome these problems by using methods such as filtering of data or spam detection. They also talk about incorporation of external knowledge sources.

Another thing the authors talk about in this article is real time detection. Due to the rapid spread of fake news on social media, it is important that it is detected on time. The authors discuss the importance of algorithms and techniques which operate in real-time and adapt to the evolving fake news patterns.

After reviewing these articles, I concluded that I wanted to go with different machine learning methods to train my model but before that I should clean out my data using nlp techniques such as lemmatizing words and removing stopwords. I also plan on combining as many datasets as possible and train on that to be able to tackle the problem on limited or biassed data. If there are still problems after testing further editing of data will be done.
.

References

---

[1] Marubeni Research Institute. Real events caused by Fake News in the US. https://www.marubeni.com/en/research/potomac/backnumber/19.html
[2] P.W. Singer and Emerson T. Brooking. 2018. LikeWar: The Weaponization of Social Media. Fake News and the Limits of Information Warfare. Cambridge University Press.
[3] Guess, A., Nyhan, B., & Reifler, J. 2019. The perils of mixing survey and experimental data. Cambridge University Press
[4] Shu, K., Mahudeswaran, D., Wang, S., Lee, S., & Liu, H. 2017. A Survey of Fake News Detection: Fundamental Principles, Models, and Challenges. ACM Transactions on Intelligent Systems and Technology. 1-31.
[5] Kh Hamed, Suhaib. 2023. Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. Sensors (Basel, Switzerland) , 23 (4), p. 1748.
[6] Ruchansky, N., Seo, S., & Liu, Y. 2017. Fake News Detection: A Deep Learning Approach. ArXvix Print.
[7] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. .2017. Detecting Fake News on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 22-36.