Project Proposal – Draft

LeTicia Cancel

DATA 698

September 9, 2022

Research Question

For this project I chose to analyze news datasets to identify true versus fake information, or as it is sometimes described on social media, real news vs fake news. People spend most of their time on the internet so we are more likely to get our news from online articles instead of television. Information is spread quickly and easily through social media but how can we tell if the information we are reading is accurate? What are the consequences of an article being misrepresented as true? Are most of these fake articles centered on negative news and what impact does this have on the reader?

Facebook has been accused of creating an algorithm that prioritizes negative posts to a user's feed since people are more likely to interact with content that sparks a strong emotional reaction. [1] Using the pandemic as an example, this was such a scary time for all and the spread of misinformation about a new virus was dangerous and potentially deadly. Facebook updated their system to compare information against a fact-checker and flag posts as false. [2] Twitter has also attempted to stop the spread of misinformation by asking users to flag posts that "seem misleading". [3] The existence of fake news is not new and is also not unique to the pandemic information that has been shared on social media in the past two years. What makes this so important today is just how easily information is shared to a large group of people. A system is needed to accurately identify misinformation as quickly as this information is spread and is needed across the web, not just on social media platforms.

I am glad to see these social media companies attempting to identify and stop or slow the spread of misinformation by using fact-checkers and flagging by the community. I would like to learn how the fact-checker and other methods of identifying misinformation work. What is the common thread between these "fake" articles and how accurate is the algorithm used to catch the misinformation? I plan to build a model that can categorize the information as real or fake. This model will be used with public article datasets found on Kaggle that are assumed to be "real news" to determine accuracy.

Research

My main data source will come from Kaggle's [Fake and Real News Dataset](about:blank) [4], a dataset that contains two files already categorized as real and fake. This will be used to identify predictors and train the model to categorize new datasets. I will also use the [All the News](about:blank) [5] dataset which contains articles that we expect to be classified as true to test the accuracy of the model.

My goal is to use as many datasets as possible to train the model. A fact-checker would be most accurate in identifying real vs fake new but this is not something we have access to. Instead we will use packages in R to find patterns in the news articles so we can accurately categorize them.

This academic paper about the Development of Fake News Model Using Machine Learning through Natural Language Processing is another resource for this project. [6] This paper breaks down fake news into 6 categories from False Connection which is described as "When headlines, visuals or captions don't support the content", to Fabricated Content which is described as "New content that is 100% false, designed to deceive and do harm." I will use this article as a guide when creating my own models.

Research Methods

I have found some resources that breakdown using machine learning to detect fake news using Python. Data Flair gives an example of a machine learning model built using a Passive Aggressive Classifier [7]. Data Flair describes this algorithm as remaining "passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting."

I will also use the LIAR Dataset from Activeloop [8]. It contains 12,800 classified short phrases and will be used to build the model for fake news detection. The Kaggle data only has two discrete categories, True or False. The LIAR dataset uses six categories - pants-fire, false, barelytrue, half-true, mostly-true, and true. It will be interesting to see how the accurate the model will be with 6 categories instead of just two. If we think about the impact of more categories, I think it makes more sense to have more nuanced categories instead of a Boolean category.

Sources

1. Chakradhar, S. (2021, October 26). More internal documents show how Facebook's algorithm prioritized anger and posts that triggered it. Nieman Lab. Retrieved September 10, 2022, from [https://www.niemanlab.org/2021/10/more-internal-documents-show-how-facebooks-algorithm-prioritized-anger-and-posts-that-triggered-it/](about:blank)
2. Mukul, P. (2021, May 28). Explained: How new facebook feature flags misinformation. The Indian Express. Retrieved September 10, 2022, from [https://indianexpress.com/article/explained/facebook-misinformation-fake-news-tool-7332659/](about:blank)
3. Spangler, T. (2021, August 17). Twitter is asking users to flag misinformation, including about COVID and elections. Variety. Retrieved September 10, 2022, from https://variety.com/2021/digital/news/twitter-users-flag-misinformation-covid-elections-1235043215/
4. Kaggle dataset: [https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?select=Fake.csv](about:blank)
5. Kaggle dataset: [https://www.kaggle.com/datasets/snapcrack/all-the-news?select=articles1.csv](about:blank)
6. Ahmed, S., Hinkelmann, K., &amp; Corradini, F. (2022, January 19). Development of fake news model using machine learning through Natural Language Processing. arXiv.org. Retrieved September 10, 2022, from [https://arxiv.org/ftp/arxiv/papers/2201/2201.07489.pdf](about:blank)
7. Team, D. F. (2021, March 31). Detecting fake news with Python and Machine Learning. DataFlair. Retrieved September 10, 2022, from [https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/](about:blank)
8. Liar dataset. Activeloop. (n.d.). Retrieved September 10, 2022, from [https://docs.activeloop.ai/datasets/liar-dataset](about:blank)

GENERAL COMMENTS

LeTicia,

You’ve clearly picked a fascinating topic (can one viably automate the differentiation between real and fake news?).

My main concern about your proposed is its (expansive) scope. For years, researchers in this field have stressed manual intervention (researching other sources and exercising “good judgment” – in addition to NLP models) – so the prospects of finding a reliable way to completely automate this process seems overly ambitious.

In other words, while I agree that “A system is needed to accurately identify misinformation as quickly as this information is spread and is needed across the web, not just on social media platforms,” I don’t think that’s viable, based on training models on the Kaggle Fake and Real News and All the News datasets. News or information by definition is highly situational and context-driven (unless you can convince me otherwise).

You’ve clearly picked a fascinating topic; my advice is to do a more in-depth literature review and give more thought as to defining the scope of what specifically you plan to do (within the semester timeframe allowed), with what models and datasets, to achieve what types of results (i.e. how you measure success or accuracy).

Please see my comments on the attached draft, which I’ve also posted in the “Draft Proposals with Track Changes” folder in the “Feedback on Drafts” section on the course site.

And please let me know if you’d like to discuss.

A