Fake News Detections Project

Cindy L Brinkmeyer

Bellevue University

DSC630

Andrew Hua

January 22, 2023

Fake News Detections Project

Milestone 3

Will I be able to answer the questions I want to answer with the data I have?

I believe that I will be able to answer the questions that I want to answer, including- Can this effort detect fake news from a list of potential news stories? It appears to be able to detect fake news. I have found several approaches to work through to detect fake news and display with what accuracy can this be achieved. Will I be able to locate additional news sources and generate similar results? There are also a large variety of news resources that I can leverage to achieve a variety of news stories available. In addition, these datasets are labelled which will aid the process. I was able to locate additional datasets, including the "LIAR", and "BS DETECTOR" datasets.

What visualizations are especially useful for explaining my data?

With additional research, I was able to determine that word clouds and network diagrams are useful visualizations for text. "Text visualization is the technique of using charts, graphs, or word clouds to showcase written data in a visual manner." (Monkey Learn Blog, n.d.)

Do I need to adjust the data and/or driving questions?

At this point, I don't need to adjust the data or driving questions, and this appears to be a solid approach to detecting fake news. This includes the original idea to add additional datasets to the initial one, it all seems achievable at this point.

Do I need to adjust my model/evaluation choices?

I haven't determined any reason to adjust my model/evaluation choices, and I have modified my Milestone 2 write-up below to reflect the comments that I received in the prior review from the professor.

Are my original expectations still reasonable?

The original expectations seem reasonable with no need to modify at this point.

Milestone 2 (Modified as per comments)

Wouldn't it be nice to detect fake news and filter it out of your news feeds to generate more trusted media? What exactly is fake news? One article, "Detecting Fake News in Social Media Networks", provides the following definition, "fictitious article deliberately fabricated to deceive readers". This project is the initial step in generating detection that could be filtered out in an online environment with a follow-on implementation and increased scope of this effort.

In this project, I would like to be able to ingest various forms of news and using a Passive Aggressive algorithm which is an online learning algorithm starting with the provided datasheet, and then, expanding to other data sources. In addition, I will explore the other models including logistic regression and random forest.

It would be interesting to see the various results for media outlets like the National enquirer stories which are sometimes expected to be sensationalized or in pop-up articles that are sometimes eye-catchers while reading other articles online.

To evaluate the results, I will provide accuracy, precision/recall and ROC.

I am hoping to learn more about media sources that might generate new that will be detected as "fake" news. Also, in expanding from the initial dataset, I will gain more experience with additional data sets and exploratory analysis.

The risks are that incorrectly labelled news stories in the next steps following the project, might be filtered out, creating bias in news sources by eliminating perhaps colorful but true articles and stories.

In the larger scope, ethical concerns at the false positives might cause bias and eliminate news stories that are not fake news. Sometimes perception alone can be an ethical concern, if the method and results are not publicly known or well understood. In addition, false negatives would allow fake news to become validated as true, and might even be more harmful and risky.

My initial data set is located at:

https://data-flair.training/blogs/advanced-python-project-detecting-fake-news/

References

(n.d.). Monkey Blog. Retrieved January 22, 2023, from HYPERLINK "https://monkeylearn.com/blog/text-visualization/"https://monkeylearn.com/blog/text-visualization/