
FAKE

Problem Statement

We Have Been Accused of Allowing Fake News to Proliferate on Our Platform.

We have been accused of allowing fake news to proliferate on our social media platform.

As a result, we are facing a backlash from the public and our shareholders, who are pressuring us to respond to the charges.

Business Value

Is Fake News REALLY Proliferating on Our Platform?

Is Fake News REALLY Proliferating on Our Platform?

How do we know whether this accusation has merit?

If we can build a model that can accurately classify news as fake or not, then we can understand whether and to what extent is exists on our platform.

Business Value

Protect Our Brand Identity

And, we can protect our brand's identity from damage from the current accusation and any similar future ones.

Methodology

What is

Fake News?

In order to build this model, we first need to understand what Fake News is.

Methodology

Untrue Information

Presented As News

According to wikipedia, Fake News is defined as "Untrue information presented as news. It often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue.

Methodology

Collect News Stories from

Truthful & Fake Sources

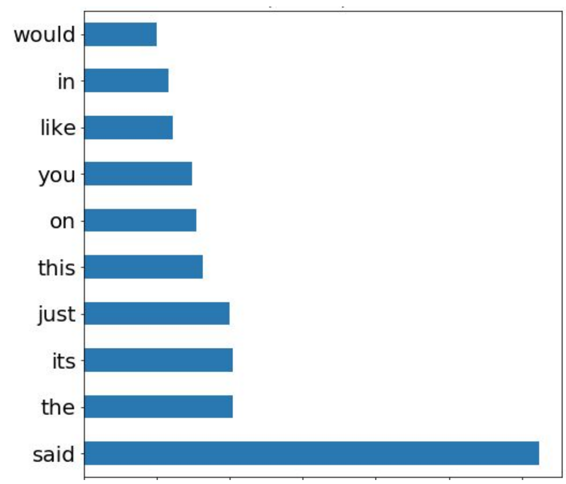
Truthful	17,447
Fake	17,447
Total	34,894

With that definition in mind, I built a dataset of about 35,000 news articles that were collected from both truthful and fake sources. The truthful articles were taken from Reuters and The Guardian, while the Fake news articles were taken from sources flagged by Politifact and Wikipedia as unreliable.

Methodology

Stopwords Only

Top 10 Most Important Words to the Model

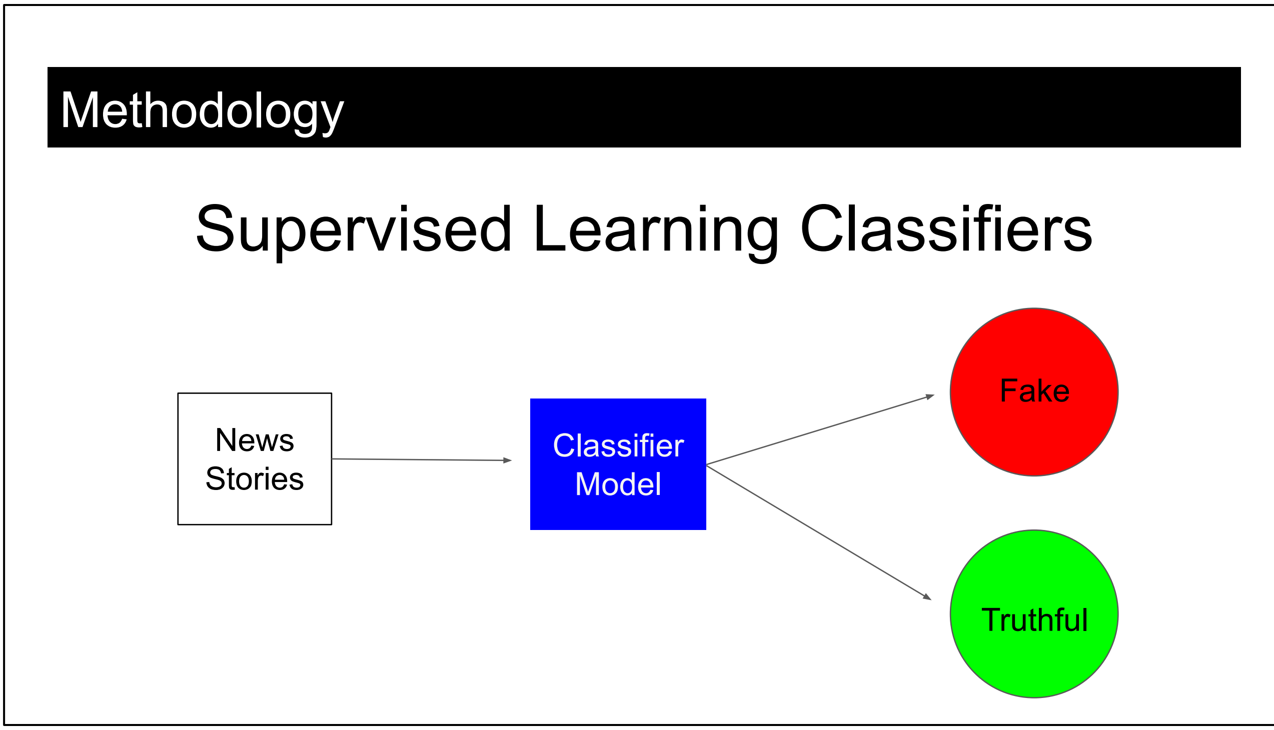


I wanted to build a model that not only could accurately classify a news story from my dataset as fake or truthful, but that could work well with news stories on people, events, organizations, or jargon not mentioned in my dataset or not mentioned in the same way. A simple way I came up with to do that was to train the model only on the stopwords in a news story in the hope that the way a fake news story is written would be different from the way a truthful story was written.

Stopwords are words that are used very frequently in a given language.

The chart on the right includes the top ten words that the model eventually found to be important when classifying a news story and their relative importances.

As you can see they are all very common words, and that the use of the word “said” was very important in determining whether to classify whether a story is truthful or fake.



The type of model I used is supervised learning classification, meaning I knew the classes (fake or truthful) of a set of news stories, and I trained a model that could accurately predict those classes with the data I had, and more importantly accurately predict the classes of news stories that the model wasn't trained on.

I used two different classifiers (Naive Bayes and Random Forest) to attempt to train a <u>useful</u> model.

Methodology

	Accuracy
Overall	89%
Identifying Truthful Stories	88%
Identifying Fake Stories	91%

Of that models I tested, a Random Forest Classifier did the best with 89% Overall Accuracy and that did almost equally well classifying Fake and Truthful news stories.

Recommendations

Audit the News

On Our Platform

Given the high accuracy of the model, my first recommendation is to:

Use the model to classify all the news stories being shared on our platform. We can use the results to determine if our platform is actually full of Fake News stories, or just has some high profile ones.

Having this information will allow our company's executives to make an informed response to the accusation.

Recommendations

Classify the News

Going Forward

Next, because fake news is a major issue, both to society and as a PR issue for our company, regardless of the results of the tests done on the news stories currently being shared, going forward all stories shared on our social network that originate at websites identified as news websites should be classified by our model during the posting process.

Keeping a log of this information will give us additional knowledge of how our platform is being used.

Recommendations

Identify Posters

And third:

Use the data collected when the news stories are classified to identify any users that are prolific posters of fake news and then consult with the legal department to determine if these actions violate our terms of service.

Future Work

Get More News

Acquire more labeled news stories to improve the model

While this dataset contained approximately 35,000 news stories that were a balanced split on classes, the origin of the news stories classified as true came from only two sources, Reuters and The Guardian, and the origin of the news stories classified as fake is unknown beyond knowing that they came from sites identified by Politifact or Wikipedia as unreliable. Having news stories from additional sources for both classes should create a more robust and generalizable model.

Collecting news stories and properly labeling them true or fake is a time consuming and labor intensive process, which is why it wasn't done already.

Recommendations

Identify the Story Origins

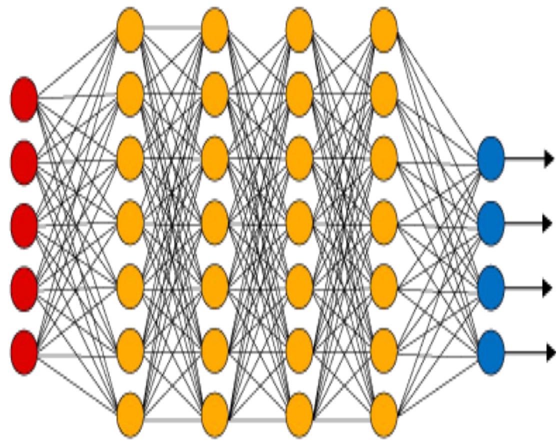
http://www.crazyfakenews.com

Identify the origin of news stories

As additional stories are collected and labeled, their URL should also be collected, and to the extent possible, determine the source URL for existing stories in our dataset. Knowing the source of the news story should be valuable data and can be incorporated into the classification process.

Recommendations

Deep Learning



Develop a more sophisticated (deep learning) model to use for Fake News detection.

While the model developed here did a good job classifying news stories, with the continual advancements in AI generated text, such as the recent GPT-3 model, detection should become harder. For example, if the GPT-3 is told to generate a news story in the style of a writer at the New York Times, it would likely be hard to detect because our model is classifying based on a currently effective, but simple way.

Developing a deep learning model to classify the news will require more time and computing power and potentially need more data.

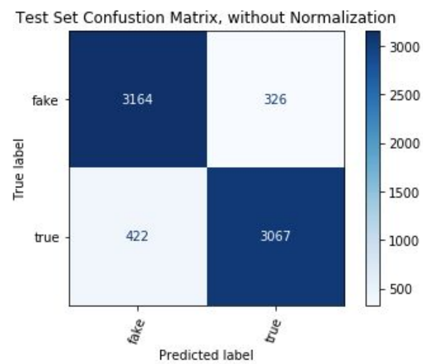
Thank You

Appendix

Performance Results of Best Model

Classification Report for Test Set

	precision	recall	f1-score	support
fake	0.88	0.91	0.89	3490
true	0.90	0.88	0.89	3489
accuracy			0.89	6979
macro avg	0.89	0.89	0.89	6979
weighted avg	0.89	0.89	0.89	6979



Appendix

Example Story From The Guardian With Only Stopwords

is willing to on according to the front page on the has that she be willing to on the of in as significant that be able to full of its while to the the that the beginning of new among the who have been and in their that when the the it will not be able to both and keep to the the similarly said the could its on that she had done so only to try to keep in the what actually did was clearly once again that the would not its four of and to allow to while to the as out in its of her she in an all from to to the after its of the were we to make an for the of with this would mean we would of the whole in the because everyone else will then want these so where did the and the come from apparently from that while the can not on the of there was for within the on the of available to who their right to said that if for example someone came to from and only for but on then i see about which we must again to me in the that the himself the he needs for himself and his in the other i am of the that we will have to further with the when this of the and potentially have the that the of for should be for further in the as to up for the whole of it is that last to on to the system with the bill from for their first five in at present are to after in for six the on is still in and it is important to was the but it would be very to that with any to on the of the and are to as they did also is that anything she says it