

While the COVID-19 pandemic has led to unprecedented damage, fake news has also become an issue. Citizens might mistake fake news for real news. Human lives have been lost due to fake information about COVID-19. Our goal is to use machine learning to find remedies for the “Infodemic.”

Table 1: Examples of fake news and real news from the dataset.

Text	Label
Florida Governor Ron DeSantis Botches COVID-19 Response - By banning Corona beer in order to flatten pandemic curve.	Fake
Masks can help prevent the spread of when they are widely used in public. When you wear a mask you can help protect those around you. When others wear one they can help protect people around them incl. you.	Real
Americans With Coronavirus Symptoms Are Being Asked To Cough Directly Onto President Trump	Fake
The main mode of transmission of is through droplets and it is possible that infected smokers may blow droplets carrying the virus when they exhale. Regardless of you should steer clear of second-hand smoke as it may cause various health problems.	Real
Justin Trudeau Resigns Amidst Coronavirus Pandemic	Fake
The positive rate has fallen a lot since early April. Back then it was ~20%. Now it's more like 4-5-6%. A lot of that change has been driven by the rising tests and plummeting positive rates in the northeast.	Real
Experts Call Out Claims That Cow Dung/Urine, Yoga, AYUSH Can Prevent Or Treat COVID-19	Fake
Georgia is particularly worrisome. The state had not seen a large rise in reported deaths despite rising infections and a steep hospitalization curve. Today the state reported its second-highest deaths since the beginning of the pandemic and the highest number since April 7.	Real

Table 2: Examples of hashtag cleaning.

Original Hashtag	Processed Hashtag
COVID19	covid19
CoronaVirus	coronavirus
CoronaVirusFacts	coronavirusfacts
COVID2019	covid2019
NYCLockdown	nyclockdown
FakeNews	fakenews

Table 3: Accuracies of our BERT model and previous approaches.

Model	Accuracy
Our DistilBERT	75.37%
Decision Tree (Patwa et al., 2021)	85.23%
Gradient Boost (Patwa et al., 2021)	86.82%
Our LSTM	87.21%
Logistic Regression (Patwa et al., 2021)	92.76%
Support Vector Machine (Patwa et al., 2021)	93.46%
Our BERT	95.61%

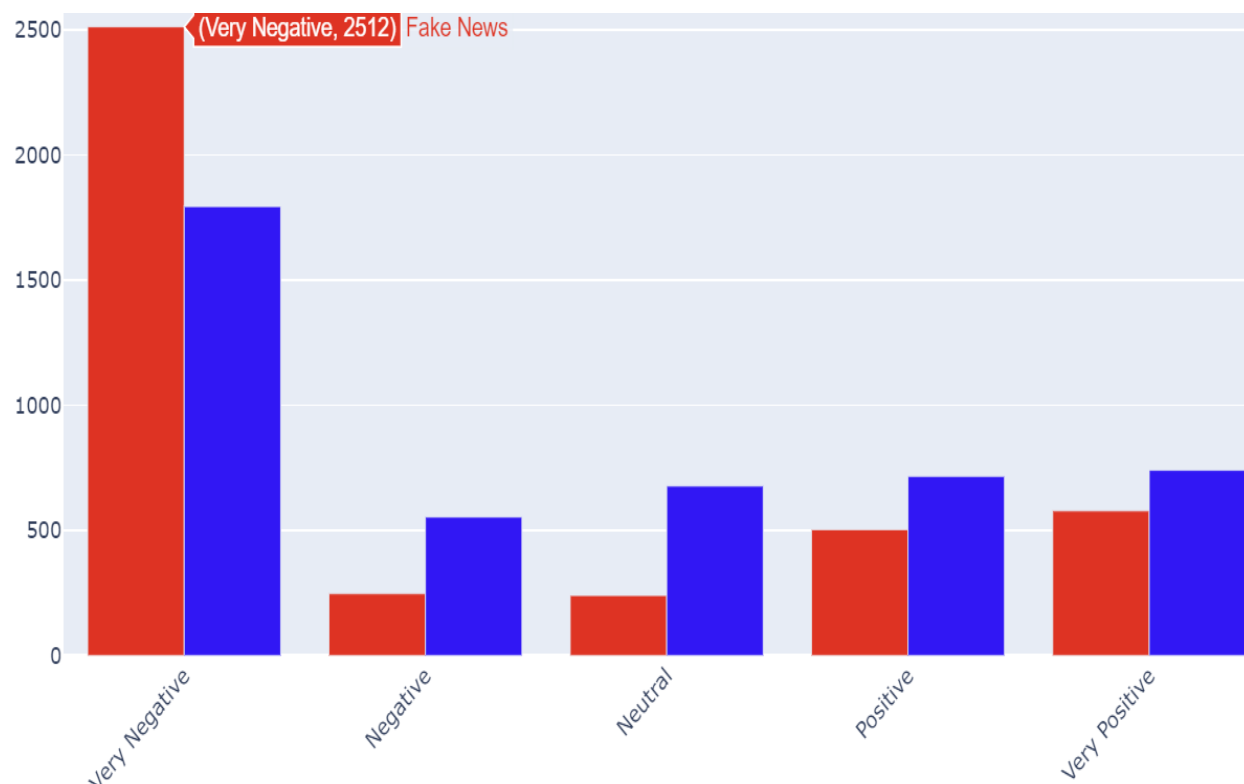


Figure 1: Sentiment Distribution of the dataset.

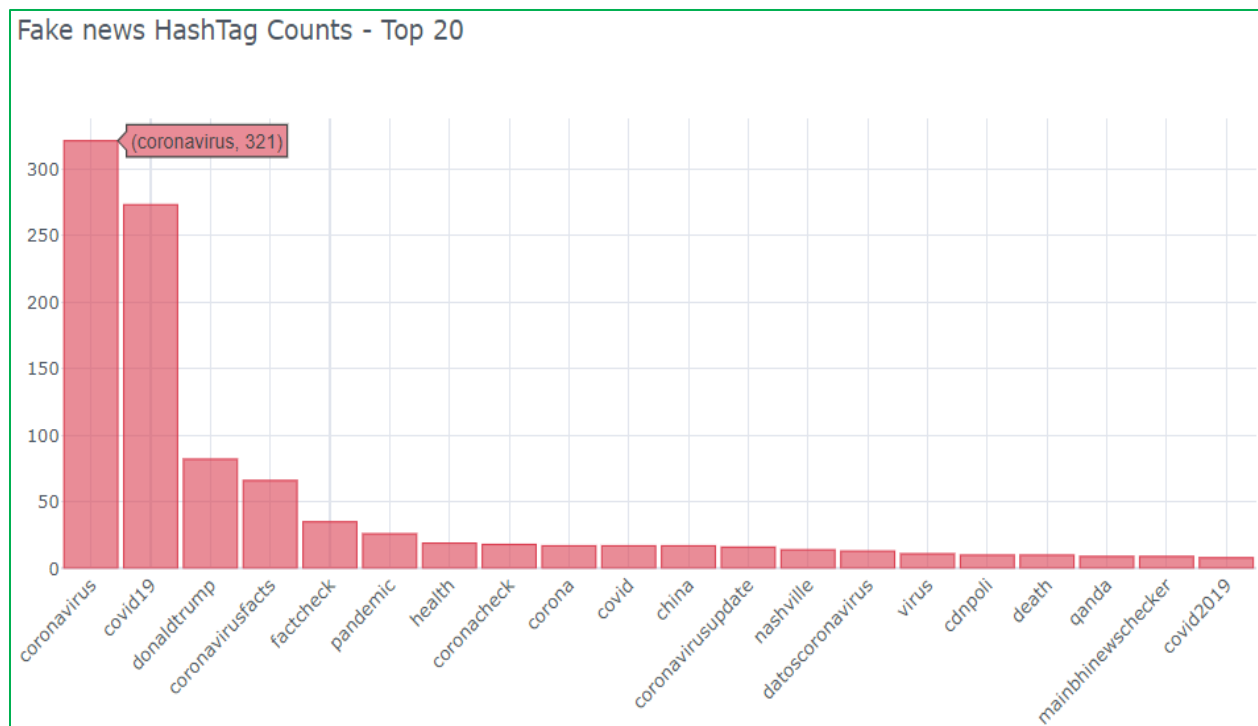


Figure 2: Top hashtags in fake news.

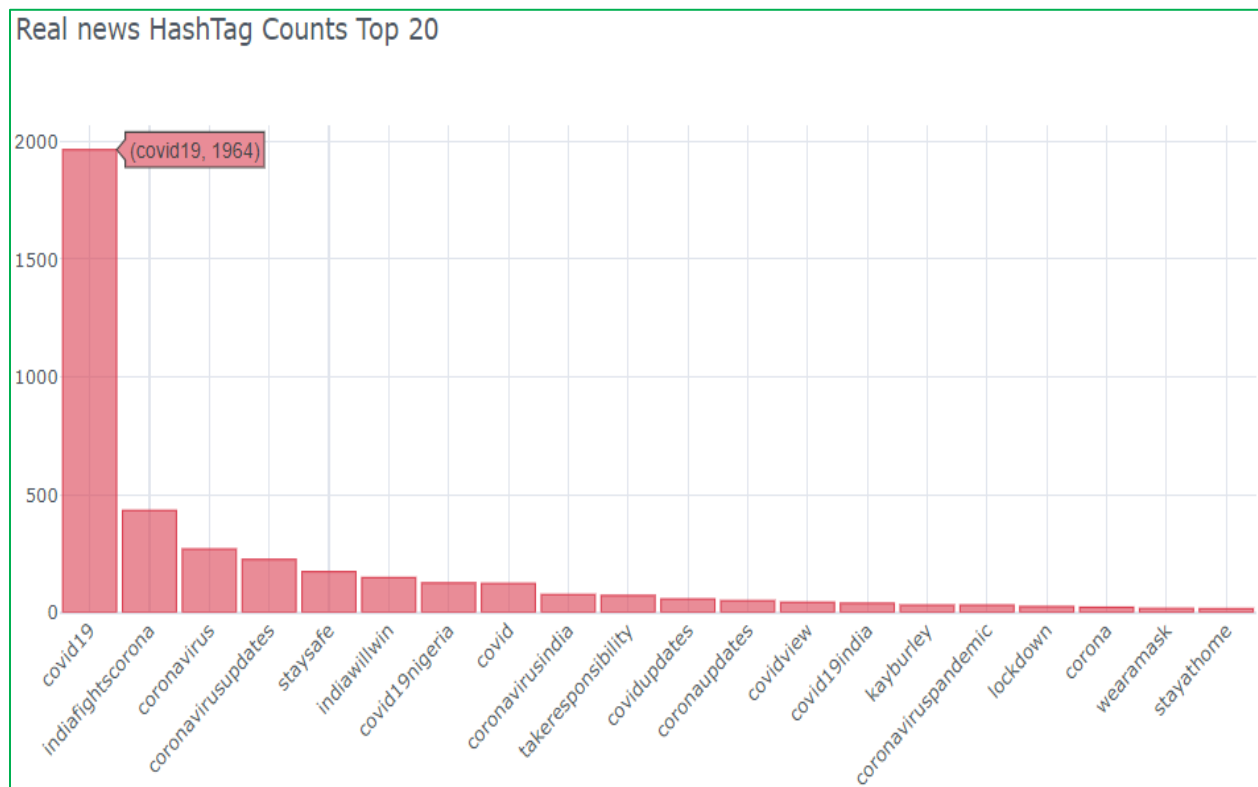


Figure 3: Top hashtags in real news.

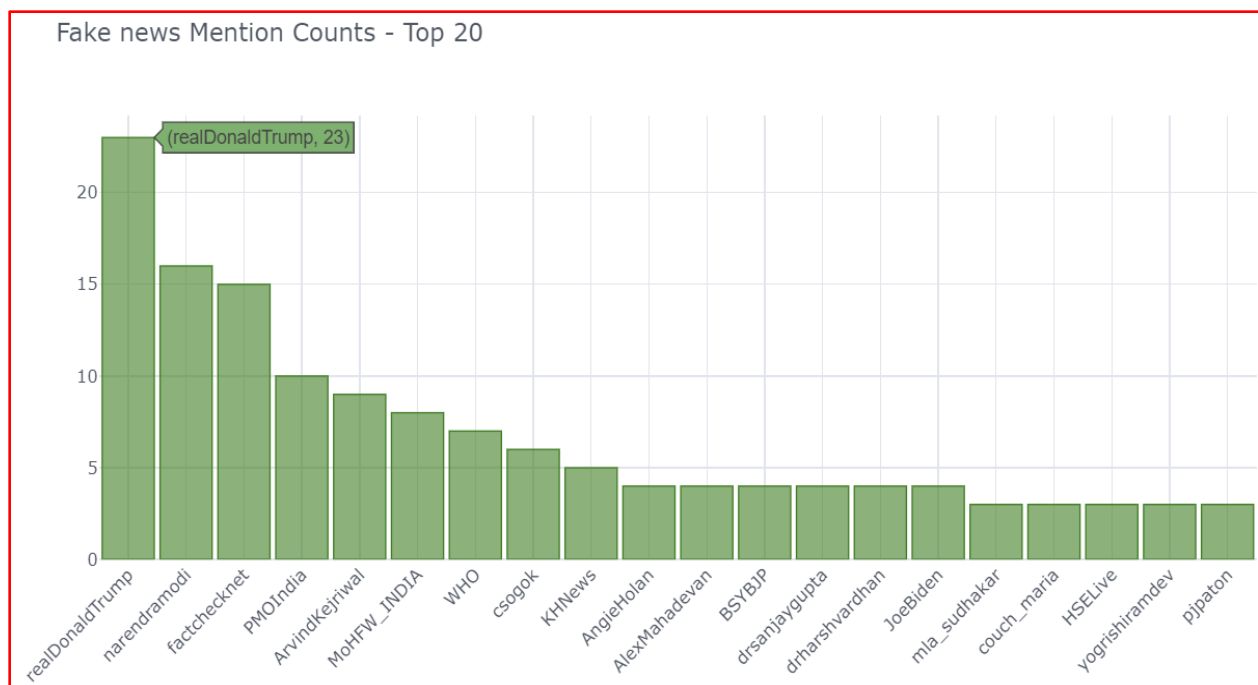


Figure 4: Top mentions in fake news.

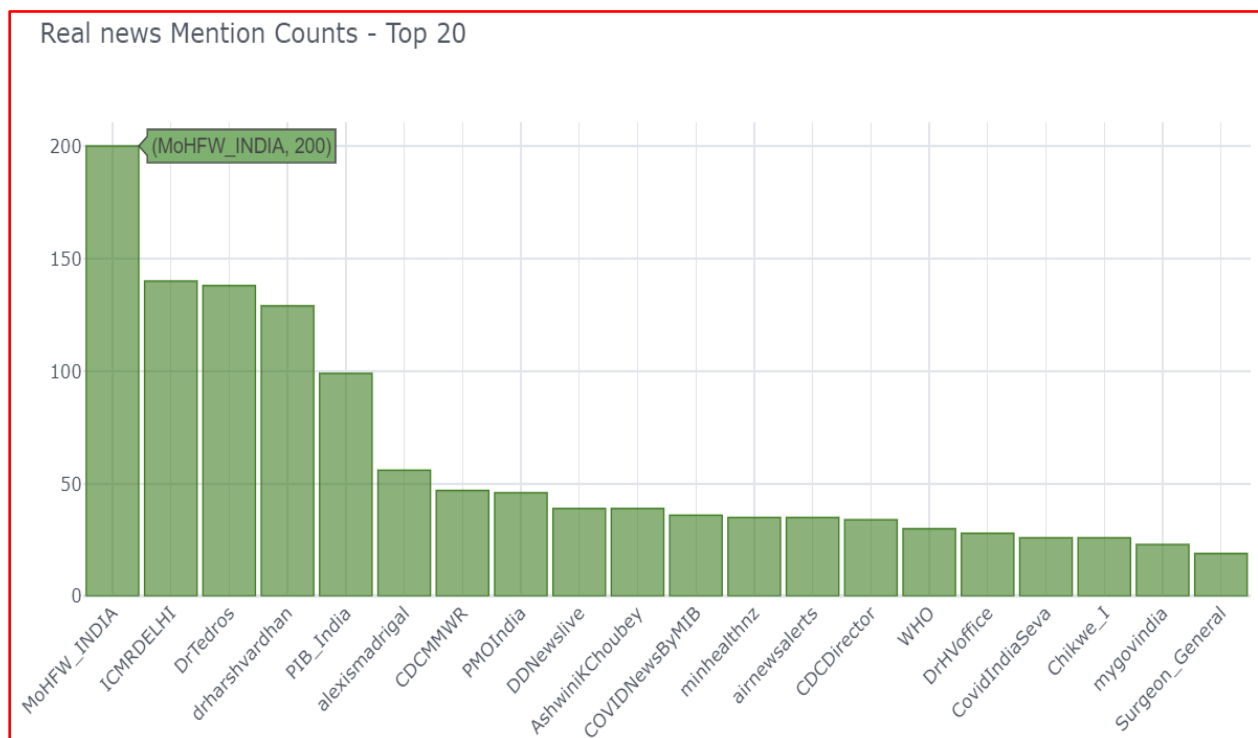


Figure 5: Top mentions in real news.

Table 4: BERT Model Accuracy on DS3 with feature elimination.

Text with Hashtag and Mention	Remove Hashtag	Remove Mention	Remove both
95.61%	95.03%	95.41%	94.81%

- Our BERT model achieved state-of-the-art results compared with previous studies. Deep Learning models with feature elimination show differences between detection models' robustness.
- Real news posts are on average 40% longer than fake news. This implies that to recognize fake news, length can provide a hint.
- The concern index of fake news is greater than that of real news by 10%, which is statistically significant.
- Fake news and real new prefer different hashtags. Hashtags in real news include inspiring and admonishing messages.
- In fake news, top mentions are the handles of politicians and fact checking sites, while in real news, top mentions are the handles of public health experts and institutes.
- We are working on an online platform and a mobile app where users can detect fake news.