STEM Fellowship Big Data Challenge 2021

Infodemiology and Infoveillance of Covid19 using GPT-3

Robert Joseph[1]

[1]University of Alberta

June 1, 2021

Abstract

Fake news detection is the task of detecting forms of news consisting of deliberate disinformation or hoaxes spread via traditional news media (print and broadcast) or online social media. Fake news is especially rampant in the current COVID-19 pandemic, leading to people believing and blindly following in false and potentially harmful claims and stories. Detecting fake news quickly can alleviate the spread of panic, chaos and potential health hazards as well reducing stress and other mental health issues. Using the Generative Pre-trained Transformer 3 (GPT-3) which is an autoregressive language model that uses deep learning to produce human-like text, classify text,design, generate code and various other use cases. Using the classifications endpoint provides the ability to leverage a labeled set of examples without fine-tuning and can be used for any text-to-label task and hence by using various data sets which contain fake and real Covid19 tweets/news GPT-3 was trained on the dataset and achieved a 98% accuracy by correctly classifying fake news and real news. Apart from using GPT-3 we also used a Passive Aggressive Classifier which is an online machine learning algorithm which also achieved an accuracy of 91%. We also provide future discussions and the limitations of the Deep Learning Model (GPT-3) as well as the simple Machine Learning model (Passive Aggressive Classifier). We hope to combat the misinformation of Covid19 spread online with these two models.

Keywords

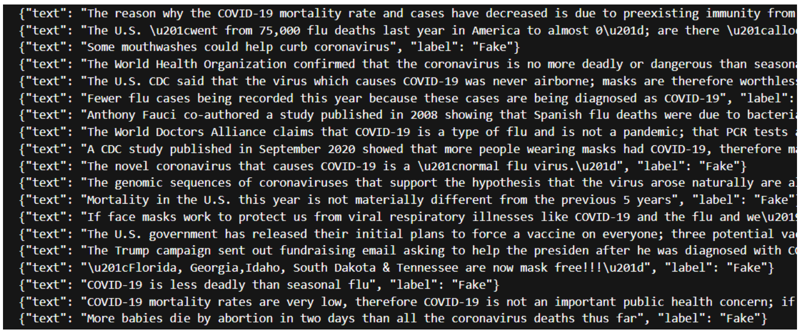
Infodemiology, Covid19, Machine Learning, GPT-3, Passive Aggresive Classifiers, NLP,News

1 Introduction

The proliferation of fake news is a significant challenge for modern democratic societies. Inaccurate information can affect the health and well-being of people, especially during the challenging times of the COVID-19 pandemic. Furthermore, disinformation erodes public trust in democratic institutions, by preventing citizens from making rational decisions based on verifiable facts. A disturbing study has shown that fake news reach more people and spread faster than actual facts, especially on social media. MIT researchers have discovered that fake news are 70% more likely to be shared on platforms like Twitter and Facebook[1]. However, people and groups with potentially malicious agendas have been known to initiate fake news in order to influence events and policies around the world. It is also believed that circulation of fake news had material impact on the outcome of the 2016 US Presidential Election.

Fake news campaigns are a form of modern information warfare, used by states and other entities to undermine the power and legitimacy of their opponents. According to EU authorities, european countries have been targeted by chinese and russian disinformation campaigns, spreading falsehoods about numerous topics, including the COVID-19 pandemic. The East StratCom Task Force has been set up to deal with that problem, by monitoring and debunking fake news about EU member states. search away![1]

As part of an effort to combat misinformation about coronavirus, I tried and collected training data and trained a ML model to detect fake news on coronavirus and present novel trends.



{"text": "The reason why the COVID-19 mortality rate and cases have decreased is due to preexisting immunity from
{"text": "The U.S. \u201cwent from 75,000 flu deaths last year in America to almost 0\u201d; are there \u201callo
{"text": "Some mouthwashes could help curb coronavirus", "label": "Fake"}
{"text": "The World Health Organization confirmed that the coronavirus is no more deadly or dangerous than season
{"text": "The U.S. CDC said that the virus which causes COVID-19 was never airborne; masks are therefore worthles
{"text": "Fewer flu cases being recorded this year because these cases are being diagnosed as COVID-19", "label":
{"text": "Anthony Fauci co-authored a study published in 2008 showing that Spanish flu deaths were due to bacteri
{"text": "The World Doctors Alliance claims that COVID-19 is a type of flu and is not a pandemic; that PCR tests
{"text": "A CDC study published in September 2020 showed that more people wearing masks had COVID-19, therefore m
{"text": "The novel coronavirus that causes COVID-19 is a \u201cnormal flu virus.\u201d", "label": "Fake"}
{"text": "The genomic sequences of coronaviruses that support the hypothesis that the virus arose naturally are a
{"text": "Mortality in the U.S. this year is not materially different from the previous 5 years", "label": "Fake"
{"text": "If face masks work to protect us from viral respiratory illnesses like COVID-19 and the flu and we\u201
{"text": "The U.S. government has released their initial plans to force a vaccine on everyone; three potential va
{"text": "The Trump campaign sent out fundraising email asking to help the presiden after he was diagnosed with C
{"text": "\u201cFlorida, Georgia,Idaho, South Dakota & Tennessee are now mask free!!!\u201d", "label": "Fake"}
{"text": "COVID-19 is less deadly than seasonal flu", "label": "Fake"}
{"text": "COVID-19 mortality rates are very low, therefore COVID-19 is not an important public health concern; if
{"text": "More babies die by abortion in two days than all the coronavirus deaths thus far", "label": "Fake"}

Figure 1: Train Dataset

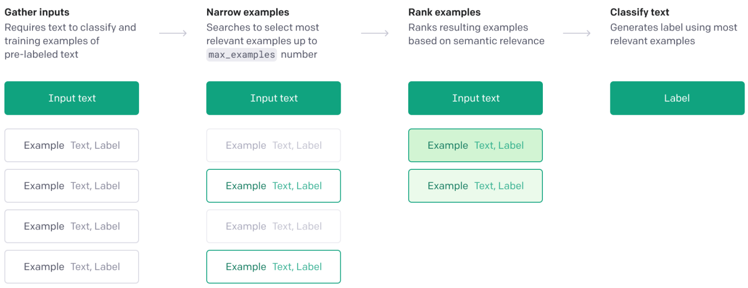


Figure 2: Classification API

2 Materials & Methods

All of the main datasets were used and transformed into JSONL files for training the pre-trained GPT-3 model.

- CoAID: COVID-19 Healthcare Misinformation Database
- FakeHealth
- COVID-19 Fake News
- COVID Fake News Dataset

All of these datasets were in mostly a CSV format which was later converted into a pandas data frame and only two major relevant columns were extracted one of them being the title of the news and the other being the label ie Fake or Real. An example of the JSONL file is shown in figure 1. This dataset was primarily used to train the GPT-3 model as required by the GPT-3 API.

2.1 GPT-3

The Classifications endpoint provides the ability to leverage a labeled set of examples without fine-tuning and can be used for any text-to-label task. By avoiding fine-tuning, it eliminates the need for hyper-parameter tuning. The endpoint serves as an "autoML" solution that is easy to configure, and adapt to changing label schema. Up to 200 labeled examples or a pre-uploaded file can be provided at query time.[2] Using this fact a dataset of around 10000 labeled examples was created and uploaded to the model. This dataset was then split into 8000 and 2000 examples for the train and the test set respectively.

2.2 Passive Aggressive Classifier

Apart from the GPT-3 model the Passive Aggressive Classifier was set up in a similar way where Passive-Aggressive algorithms are generally used for large-scale learning. It is one of the few 'online-learning algorithms'. In online machine learning algorithms, the input data comes in sequential

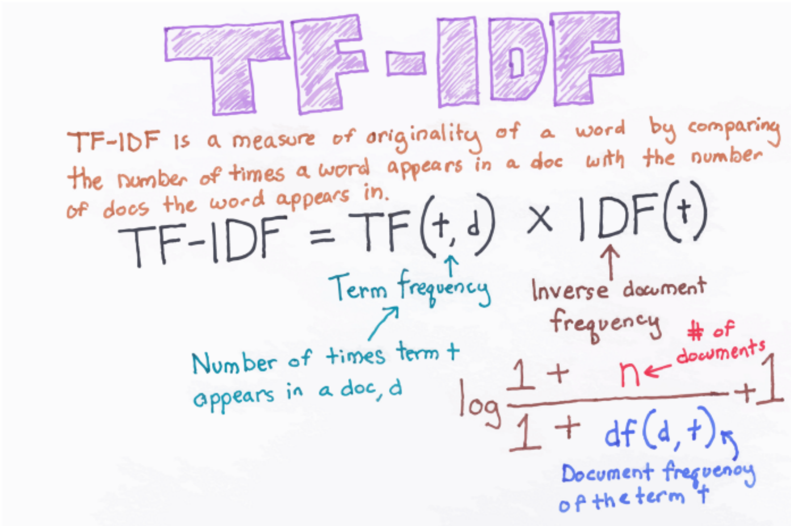


Figure 3: TF-IDF

order and the machine learning model is updated step-by-step, as opposed to batch learning, where the entire training dataset is used at once. This is very useful in situations where there is a huge amount of data and it is computationally infeasible to train the entire dataset because of the sheer size of the data. We can simply say that an online-learning algorithm will get a training example, update the classifier, and then throw away the example.[3] This is extremely useful in detecting fake news as data is continuously being streamed in every second. Similarly for GPT-3, a 7000 labeled dataset was taken and split into a 5000 and 2000 examples for the train and test set respectively.

2.3 TF-IDF

We also introduce the concept of TF-IDF vectorizer which is an abbreviation for Term Frequency Inverse Document Frequency. This is very common algorithm to transform text into a meaningful representation of numbers which is used to fit machine algorithm for prediction. This is used in our Passive Aggressive Classifier.

2.4 Exploratory Data Analysis

For the Exploratory data analysis part on the dataset a simple word Cloud was created as well as bar plots created with the frequency of the most common words in both the fake and real dataset to compare. Using regex some of the HTML tags in the title were removed and cleared out and from the NLTK library the common stop words ie (the, on, is etc) were used so that the word Cloud only counted the frequency of the rest of the words and ignored these stop words.

3 Results

3.1 GPT-3

Using the GPT-3 model query the code to get the result is as given in table 1. The GPT-3 API has 4 different engines ie [2]

- **Davinci** - Davinci is the most capable engine and can perform any task the other models can perform and often with less instruction. For applications requiring a lot of understanding of the content, like summarization for a specific audience and creative content generation, Davinci is going to produce the best results. These increased capabilities require more compute resources, so Davinci costs more per API call and is not as fast as the other engines. Another area where Davinci shines is in understanding the intent of text. Davinci is quite good at solving many kinds of logic problems and explaining the motives of characters. Davinci has been able to solve some of the most challenging AI problems involving cause and effect. **Good at: Complex intent, cause and effect, summarization for audience**

Table 1: Fake news detection accuracy of GPT-3

Search Model	Model	Number of Labelled Examples	Number of Tokens(words) in the Title	Accuracy
ada	curie	10	5	65.75
ada	curie	100	Full Length	93.4
ada	curie	200	Full Length	96.3
ada	babbage	200	Full Length	84.7
davinci	curie	10	10	83.4
davinci	curie	200	Full Length	98.2
davinci	ada	200	Full Length	91.1

Table 2: Accuracy of Passive Aggressive Classifier in detecting fake news

Regularization	Random State	Maximum Iterations	Accuracy
0	0	50	88.2
1	10	50	90.81
0.5	100	10	90.69

- **Curie** - Curie is extremely powerful, yet very fast. While Davinci is stronger when it comes to analyzing complicated text, Curie is quite capable for many nuanced tasks like sentiment classification and summarization. Curie is also quite good at answering questions and performing QA and as a general service chatbot. **Good at: Language translation, complex classification, text sentiment, summarization**

- **Babbage** - Babbage can perform straightforward tasks like simple classification. It's also quite capable when it comes to Semantic Search ranking how well documents match up with search queries. **Good at: Moderate classification, semantic search classification**

- **Ada** - Ada is usually the fastest model and can perform tasks like parsing text, address correction and certain kinds of classification tasks that don't require too much nuance. Ada's performance can often be improved by providing more context. **Good at: Parsing text, simple classification, address correction, keywords**

3.2 Passive Aggressive Classifier

Similarly for GPT-3 the accuracy of the classifier with different values in the hyperparameter's are shown in table 2.

4 Discussion

The results for both GPT-3 and the Passive Aggressive Classifier show promising results and the GPT-3 is even impressive in achieving an all time 98.2 accuracy of detecting fake news. Now there are various considerations of using 4 different engines' but as mentioned above in the result section davinci is the strongest engine compared to the rest which was used as the search engine (the engine used to search the entire dataset to compare the values of the query) and the engine used to classify is the curie engine which is used for complex classification. The interesting thing to notice in the table is the number of labelled examples used which gave varied results. For a given query, the endpoint searches over provided examples or labeled data in the provided file to select the most relevant examples for that particular query. All the label strings will be normalized to be capitalized. Semantic search is used to rank documents by relevance to the query. The relevant examples are then combined with the query to create the prompt for completion. Setting max-examples to a higher value leads to improved accuracy but with increased latency and cost. max-examples is by default set to 200. The code is as follows [2]

```
results = openai.Classification.create(
            file = "id",
            query = prompt,
```

Figure 4: Word Cloud - Fake Dataset

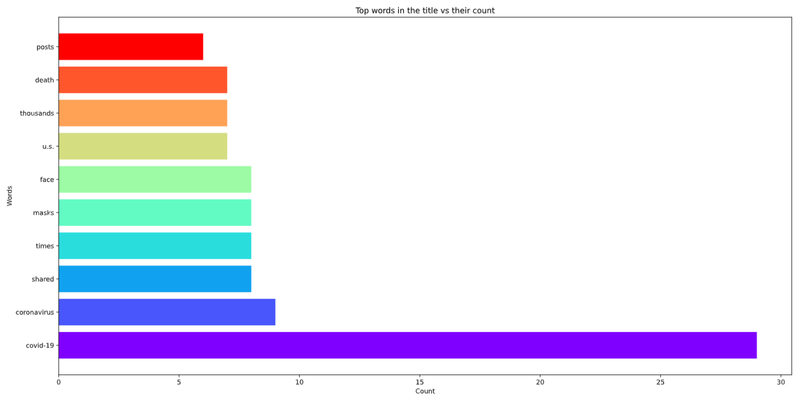


Figure 5: Bar Plot - Fake Dataset

```
        search_model="davinci",
        model="curie",
        max_examples=200)
```

I did test the accuracy of the engine's also in regard to how many words are present in the title and compare the accuracy of the model on detecting fake vs real news. A surprising result is the fact that just with 10 words in the title and not even the full length yielded an accuracy of 83.4 with just 10 labelled examples and the faster model ada resulted in 65.75 (with 5 words) accuracy. This is a huge due to the fact that the model did not need to process the whole title rather just a part of it.

Now in regards to the Passive Aggressive Classifier we get an accuracy of almost 91% and while exploring the tfidf vectorizer (code given below)

```
pac=PassiveAggressiveClassifier(C = 0.5,random_state = 10,max_iter=100)
pac.fit(tfidf_train,y_train)

#DataFlair - Predict on the test set and calculate accuracy
y_pred=pac.predict(tfidf_test)
score=accuracy_score(y_test,y_pred)
```

While doing some Exploratory data analysis on the dataset(consisting of all the fake labels only) the word Cloud generated as well as a bar plot showing the words with the most frequency present is shown in figure 4 and 5.

Similarly for the dataset(consisting of all the real labels only) the word Cloud generated as well as a bar plot showing the words with the most frequency present is shown in figure 6 and 7.

These two word clouds are have a lot of high frequency words in common such as covid, coronavirus, vaccine etc as well as the fake news word cloud seems to be having posts, death, rate, lockdown, social media and other common terms while the real news word cloud has a more positive tone with words such as may, vaccine,testing, mask and such.

Figure 6: Word Cloud - Real Dataset

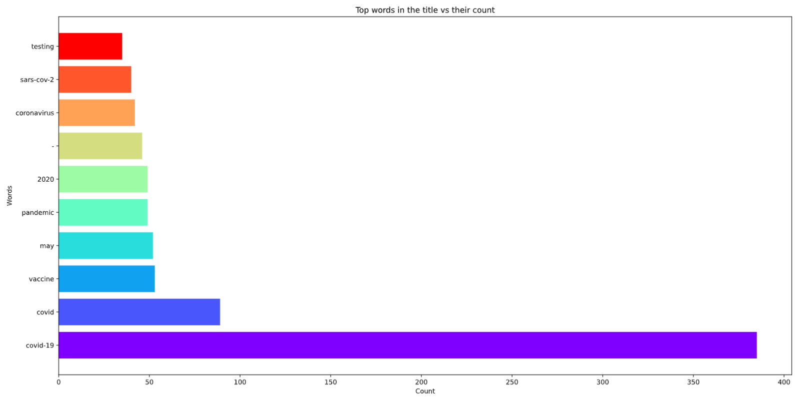


Figure 7: Bar Plot - Real Dataset

5 Conclusions

The problems of fake news and disinformation play an important role on nowadays life especially during this pandemic. This is because the advanced level of technology and communication methods we have enabled information spreading among people without any verification. This is a reason why researchers started searching for solutions to stop fake news and disinformation from spreading easily. However, it is well known that controlling the flow of information online is impossible. In this paper, we performed an attempt to verify the news articles credibility depending on their characteristics.[4]

All of the proposed models can run in near real-time with moderately inexpensive compute. The work presented here is based on the assumption that our knowledge base is accurate and timely. A certain limitation sis that this assumption might not always be true in a scenario such as COVID-19 where "facts" are changing as we learn more about the virus and its effects as well as the various variants[5].

Finally in this paper the use of GPT-3 which is the worlds largest model can definitely be used to detect fake news much faster in real time with more fine tuning and we can also use the concept of domain specific knowledge by pre-training GPT-3 so as to understand the domain and yield a better accuracy and performance. A passive aggressive Classifier was also presented in this paper which can be used and is much faster than GPT-3 (due to the fact that the GPT-3 API takes a while to get the results back) and can be instantly used online as the data streams in.

There are several ways on how this can be discussed and more research done on how to use such an advanced and powerful model to detect fake news not only for Covid19 but in general. Powerful web scrapping methods that can crawl the internet in real time and upload it to either the GPT-3 or passive aggressive classifier and label it in real time hence either removing it from the website where the news is hosted or marking it with a fake label. This way we can help fight an infodemic as well as lessen the belief in rumours which can cause significant harm.

6 Acknowledgements

I would like to thank God for protecting and blessing me during this pandemic. I would also like extend our sincerest gratitude to the Canadian STEM Fellowship for organizing the Big Data Challenge and allowing us to learn more as well as improve our technical skills and also would like to thank the various workshop facilitators for teaching us the required skills needed to present this report in the best way possible. I am grateful for my parents and friends too for supporting me all this way.

References

[1] How i created a fake news detector with python. 2021.

[2] Gpt3 - documentation. 2021.

[3] Passive aggressive classifiers. 2020.

[4] A tool for fake news detection. 2020.

[5] Two stage transformer model for covid-19 fake news detection and fact checking. 2020.