

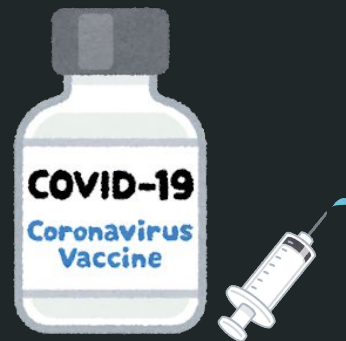
# COVID Misleading Tweets Detection

---

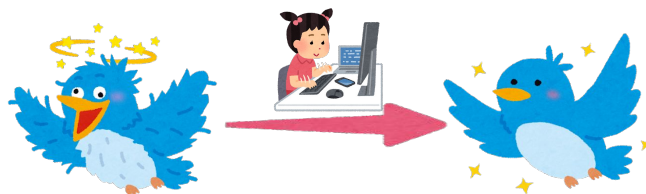
2022 JHU Data Sciences Fellowship Summer Program

Mentor: Gherman Uritskiy, PhD

Presenter: (HELEN) Ting He



# Misleading Information at Twitter



We want to broaden the range of voices in the Internet, but also don't want to have somethings which received widespread public attention by misleading information. especially for the COVID case, a pandemic spread all over the world

**Goal:** Automatically detect misleading Tweets

**My solution:** A timely web application which can run NLP model given the tweets

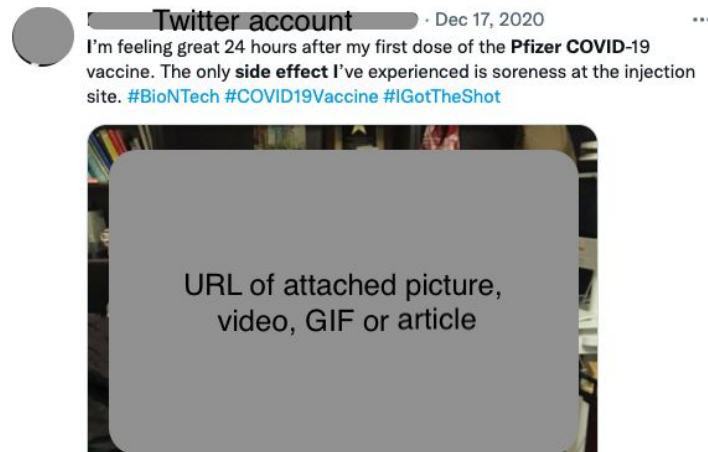
# Data Source

## Annotated Data Source:

- CoAID: COVID-19 healthcare misinformation Dataset
  - how they labeled data: Claim was annotated into fake or true by comparing the reliable sources. They classified the myth and rumor as fake claim'.
  - chosen dataset: 457 fake claim(of one or two sentences) tweets, 6342 real claim tweets which collected September 1, 2020 from through Nov 1, 2020

## Real-World Tweets:

- Scraping Tweets from website by calling API



# Experiments

## Baseline Model#1: Random Forest

- Structured features only
- F1 score at 0.66

## Experiment#2: Random Forest with TF-IDF

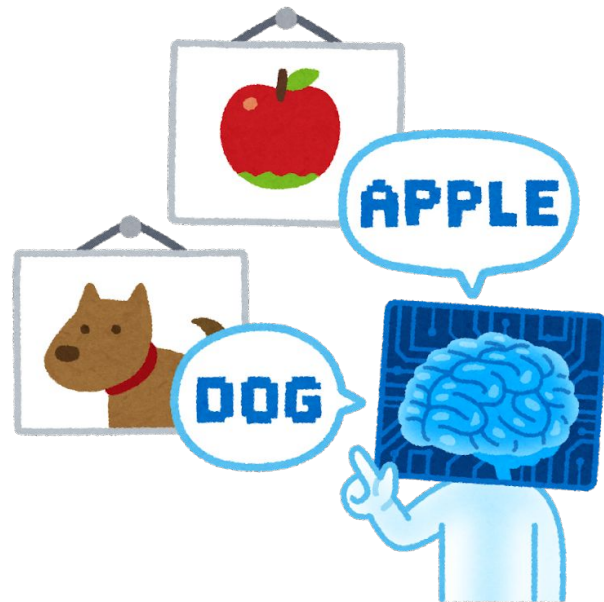
- Structured features with unstructured feature in TF-IDF encoding
- F1 score at 0.75

## Experiment#3: Random Forest with TF-IDF + NMF

- Structured features with unstructured feature in TF-IDF encoding plus with NMF to narrow feature space
- F1 score at 0.70

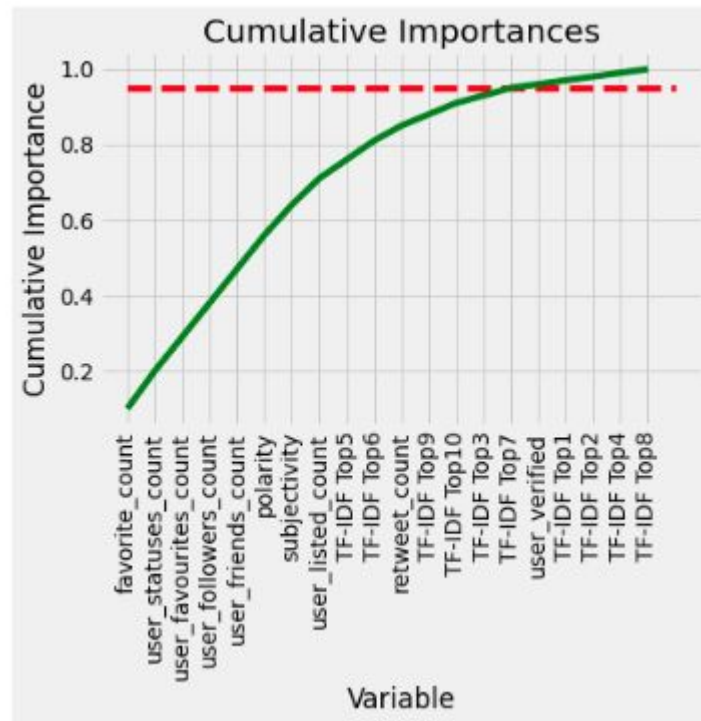
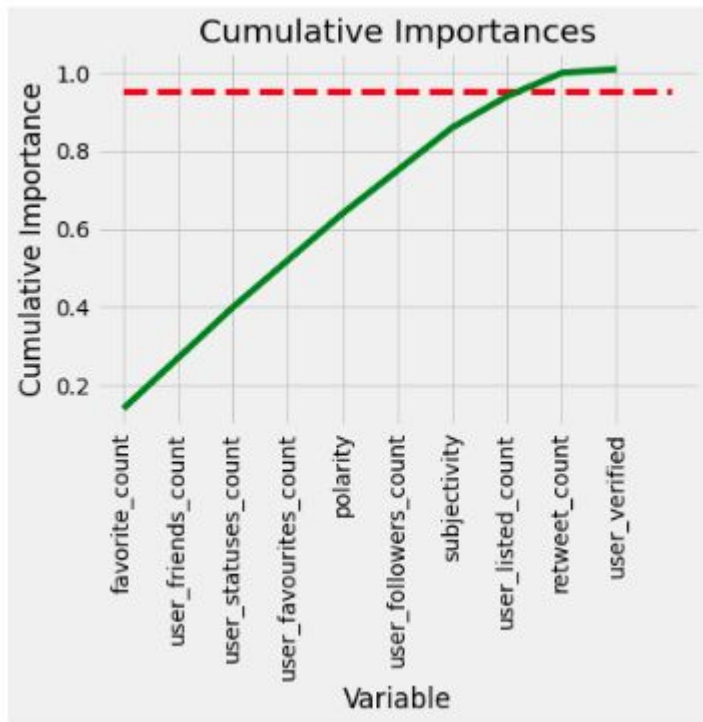
## Experiment #4: Pretrained Distill BERT Embedding

- Unstructured feature only
- F1 score at 0.96



F1 score report in validated dataset  
Tuning the hyperparameter by gridsearch

# Experiments (con't)



# Final Representation

http://localhost:8501/

## COVID Misleading Informaiton Detection 🔥

Retrive the Most Recent Tweets Given the COVID related  
Keyword

Key Word:

covid

Number of the Latest Tweets You Want to Retrieve:

10 - +

Submit & Rerun Model

	The Latest 10 Tweets on covid	Percentag
0	RT @Flyck66: "In the trial itself, it appeared that one was more likely to suffer a serious adverse event f	0.104941
1	RT @RepMaryMiller: Joe Biden is firing healthy military service members for not taking a vaccine that f	0.031431
2	RT @EliseVMike: My cousin had to return to in-person work, got COVID from her boss, and passed away	0.037403
3	RT @Atipico1996: England's COVID deaths in the first 7 weeks of summer: 2022: 6,344 2021: 2,202 2020	0.037403
4	RT @EliseVMike: My cousin had to return to in-person work, got COVID from her boss, and passed away	0.078091
5	RT @DahvidSchloss: Because I'm seeing a ton of you in infosec Twitter saying you popped positive fro	0.031997
6	RT @amazonholder1: Take a guess what the 5 cities hit the hardest for dining since Covid have in comr	0.016555
7	RT @rorycooper: Of all the dumb things the federal and NY governments have spent billions of COVID r	0.046984
8	RT @sammy4723: A reminder: * COVID is airborne. * Herd immunity did not happen. * Omicron was no	0.037403
9	RT @MeetJess: First Lady Jill Biden tests positive for COVID today. Pres Biden was masked when he sto	0.031997

