Outline

● Background
● Timeline
● Work Completed
● Lessons Learned
● Future Work
● References
● Acknowledgements
Background

- Client: Dr. Mohamed Farag
- Mis/disinformation spread on Twitter
- Disabling mis/disinformation sources

- Misinformation
  - False information that is spread, regardless of intent to mislead.

- Disinformation
  - Deliberately misleading or biased information; manipulated narrative or facts; propaganda.

- Deliverables: Frontend UI, machine learning model, data analysis
**TimeLine**

**Feb.**

**Milestones 1**
- Create team roles
- Meet with Client
- Created OKRs

**Mar.**

**Milestones 2**
- Hydrate Tweets
- Complete backend
- Label Tweets

**Apr.**

**Milestones 3**
- Train the model from labeled tweets
- Complete frontend
- Complete frontend testing
Work Completed (Backend)

- Data cleaning
  - Filter non-English tweets
- Tweet Hydration
  - Twarc
  - Twitter v2 API
- Label training data
  - 5,000 labeled, hydrated tweets
- Machine learning model
  - SVM
  - Multi-class
  - Scikit learn
import json
import io
import json_normalize
import warnings

file_name = 'covid.jsonl'
warnings.filterwarnings("ignore")
# clean_tweets = open("clean_tweets.txt", "w")
num_lines = 0

with io.open(file_name, 'r', encoding='utf-8') as f:
    for line in f:
        tweet = json.loads(line)
        if tweet['lang'] != 'en': continue
        # Grab the quoted tweet id, and original id
        # if tweet['is_quote_status'] == 'true':
        #     continue
        # grab original tweet id
        # if tweet['retweeted'] == 'true':
        #     continue
        id = str(tweet['id'])
        # clean_tweets.writelines(id + "\n")
        num_lines = num_lines + 1

print(num_lines)
# clean_tweets.close()
Work Completed (Backend): Tweet Hydration

Step 1: Enter
CONSUMER_KEY,
CONSUMER_SECRET
Step 2: Enter
ACCESS_TOKEN,
ACCESS_TOKEN_SECRET
<table>
<thead>
<tr>
<th>Work Completed (Backend): Label training data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5</strong></td>
</tr>
<tr>
<td>In case the scientific ‘experts’ haven’t realised it yet, the establishment, rightly or wrongly are about to throw them under the bus! Special Report: Johnson listened to his scientists about coronavirus - but they were slow to sound the alarm <a href="https://t.co/iPZeUhMVbV">https://t.co/iPZeUhMVbV</a></td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>6</strong></td>
</tr>
<tr>
<td>Media briefing on #COVID19 with @DrTedros. #coronavirus <a href="https://t.co/BOe2WSGfnK">https://t.co/BOe2WSGfnK</a></td>
</tr>
<tr>
<td>Truth</td>
</tr>
<tr>
<td><strong>7</strong></td>
</tr>
<tr>
<td>55% of Americans are smart enough to see that Trump has done a poor job of preventing the spread of the coronavirus in the United States. The other 45% of Americans think the former host of Celebrity Apprentice is doing a great job because they are total fucking morons.</td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>8</strong></td>
</tr>
<tr>
<td>Should you invest during coronavirus pandemic? <a href="https://t.co/wJdewyVSFE">https://t.co/wJdewyVSFE</a> @MorningsMaria @FoxBusiness</td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>9</strong></td>
</tr>
<tr>
<td>ICE is trying to hide the coronavirus outbreaks raging through detention centers.</td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>10</strong></td>
</tr>
<tr>
<td>Another thing to watch for? ICE may try to hide deaths by &quot;releasing&quot; people in comas and on ventilators at a hospital. Once &quot;released,&quot; the death is no longer &quot;in custody.&quot; They've done it before. <a href="https://t.co/yjSJU5UtGp">https://t.co/yjSJU5UtGp</a></td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>11</strong></td>
</tr>
<tr>
<td>Pope Francis has said the coronavirus pandemic is one of &quot;nature's responses&quot; to humans ignoring the current ecological crisis. <a href="https://t.co/mgrDyNhIEX">https://t.co/mgrDyNhIEX</a></td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>12</strong></td>
</tr>
<tr>
<td>HUGE: Brilliant Dr. Shiva, Inventor of Email, Outlines Connections Between Bill Gates, Dr. Fauci, the WHO and the CDC – Relevant to Coronavirus Pandemic <a href="https://t.co/dhZEIm6spu">https://t.co/dhZEIm6spu</a> <a href="https://t.co/A0pAutRrTT">https://t.co/A0pAutRrTT</a></td>
</tr>
<tr>
<td>General misinformation</td>
</tr>
<tr>
<td><strong>13</strong></td>
</tr>
<tr>
<td>UK rejects Trump offer to help Boris Johnson's coronavirus treatment - &quot;We're confident the prime minister is receiving the best possible care from the National Health Service,&quot; &quot;Any treatment he receives is a matter for his doctors.&quot; <a href="https://t.co/Lj7CGglMbR">https://t.co/Lj7CGglMbR</a></td>
</tr>
<tr>
<td>Truth</td>
</tr>
</tbody>
</table>
Work Completed (Frontend): Timeline

- Shows number of tweets on given day for given classifier
  - Only dates with > 20 tweets are shown
- Buttons change the tweets that are shown
- Vertical timeline (scrolling webpage) for full tweets
Work Completed (Frontend):

- Routes completed for each button
- Functional statistics page
- JSON file
- Statistics
Future Work

- Increase AI model accuracy
- Make frontend online
- Hydration optimization
## Lessons Learned

<table>
<thead>
<tr>
<th>Problems</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple records of same tweet</td>
<td>Remove duplicates through dictionary</td>
</tr>
<tr>
<td>User Interface</td>
<td>Learn React</td>
</tr>
<tr>
<td>Twitter API request limits</td>
<td>Multiple accounts</td>
</tr>
</tbody>
</table>
Acknowledgements

- Dr. Mohamed Magdy Gharib Farag
- Dr. Fox
References