



# Reviews about Speech of the Canadian president about covid-19 vaccine

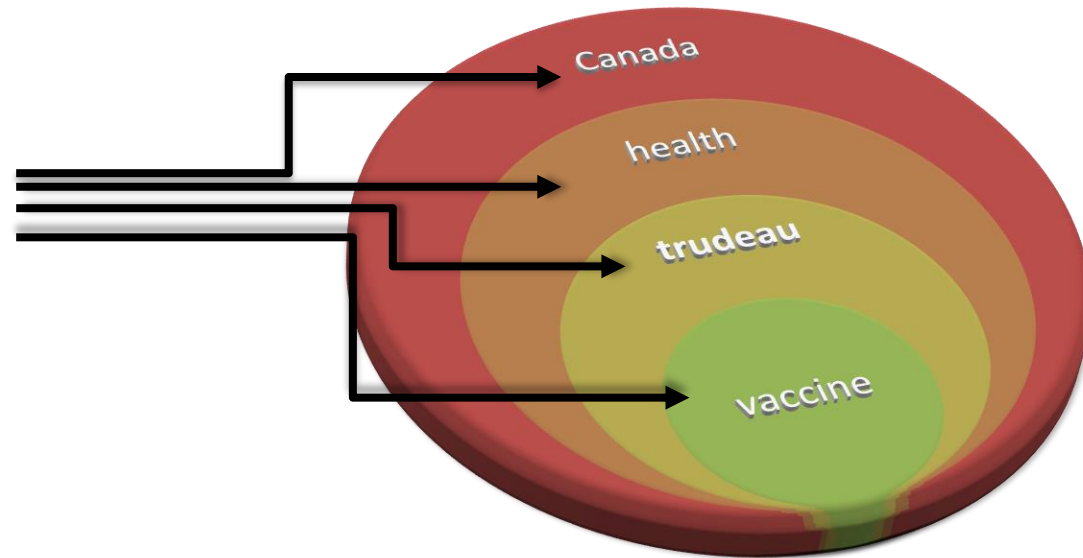
NLP Project

Shahad Abdulsalam

# Problem & Dataset

- After the speech of **Trudeau** ( president of Canada) about covid-19 vaccine plan, comments start hitting from the Canadian Citizens.

Comments:





# preprocessing

• Comment  $\longrightarrow$  Document  $\longrightarrow$  tokens

Removing:

- Stop words.
- Punctuation.

Applying:

- Lemmatization



# LDA Model

& visulaization

Start = 2, limit= 9, step=1

- Number of topics: 2
- has coherence score : 0.139
- Number of topics 3
- has coherence score: 0.285
- Number of topics: 4
- has coherence score: 0.2886
- Number of topics: 5
- has coherence score: 0.3081
- Number of topics: 6
- has coherence score: 0.303
- Number of topics: 7
- has coherence score: 0.3202
- Number of topics: 8
- has coherence score: 0.3077



# CorEx model

Anchors with strength=10

[ 'plan' , 'decide' ]

[ 'healthcare' , 'health' , 'care' , 'children' ]

[ 'government' , 'country' , 'citizen' , 'decision' ]

[ 'canada' , 'US' , 'liberal' , 'pay' ]

Covid plan:

plan,decide,low,distribution,version,citizens,falling,impact,sidelines,statistics.

Healthcare:

health,care,healthcare,children,workers,provinces,oil,jobs,approval,vaccines.

canadian\_government:

government,country,people,example,private,playing,domestically,current,ottawa,wrong.

liberal\_party:

canada,liberal,pay,vaccine,approve,foreign,production,chinese,blocked,domestic.



## Converting data into supervised data:

1. Create a dataframe for topics.
2. Create a dataframe of tokens of the documents.
3. Merge the dataframes together.
4. Return the topics columns into a dummies variables.



# Counter & Amods:

---

```
Counter(healthcare_adj).most_common(10)
```

---

```
[('possible', 28),  
 ('other', 16),  
 ('safe', 11),  
 ('good', 10),  
 ('more', 9),  
 ('first', 8),  
 ('much', 8),  
 ('long', 7),  
 ('pandemic', 7),  
 ('only', 7)]
```

---

```
Counter(canadiangovernment_noun).most_common(10)
```

---

```
[('government', 99),  
 ('vaccine', 88),  
 ('people', 37),  
 ('country', 37),  
 ('vaccines', 23),  
 ('time', 19),  
 ('plan', 14),  
 ('production', 13),  
 ('world', 12),  
 ('distribution', 12)]
```



THE END

THANK YOU