Product Name: Covangers
Team Name: tokoDAMI
Owner: **Aulia Muthia** **Denise Sonia Rahmadina** **Amanda Athayasar** **Muhamad Ihsan**

Context

Who Are We

Covangers is a B2G (Business to Government) platform for presenting data & analyzing results of public sentiment on COVID-19 vaccination data in DKI Jakarta from Twitter using NLP (Natural Language Processing). This technology is projected to **help in the acceleration of COVID-19 vaccinations by mapping location and community sentiment using a psychological and technological approach at the sub-district level in DKI Jakarta.** All in all, this platform will help to understand the public reaction and aid the policymakers to project the vaccination campaign as well as health and safety measures in the ongoing global health crisis.

Almost a year and a half from a pandemic, Indonesia is now seeing its worst COVID-19 outbreak. As of early July, more than 2.3 million COVID-19 cases and 63,000 deaths have been confirmed. In response, the country is stepping up public health measures while aiming to accelerate vaccinations and has set a goal of inoculating two million people per day by August. [source]

Indonesia started its vaccination program in January 2021. Based on data from kawalcovid. id, as of April 27, 2021, the positivity rate in Indonesia is still at 16.1%, far above the WHO standard which sets 5% as a normal number. Therefore, vaccination must be carried out in an effort to fight Covid-19 in Indonesia.

Early reports of COVID-19 vaccination may not accurately convey community-level concern about the pandemic during early stages, particularly in Indonesia where vaccination capacity was initially limited. Social media interaction may elucidate public reaction and communication dynamics about COVID-19 vaccination in this critical period, during which communities may have formulated initial conceptions about the perceived severity of the pandemic.

Why do We need It

1. From the perspective of behavioral science, one of the human weaknesses is sometimes they get disconnected from their future selves. Dynamical communication in Twitter will affect human decisions to get vaccinated or not. We need help from the government to garner trust will be essential to their success, and to the emergence of more resilient societies after the crisis.
2. Public sentiments are different among locations. By knowing where to start counseling, we can get a whole new perspective of its neighborhood and how to create strategies to accelerate vaccination rates in that specific location.

How We Help

We propose solutions to improve the communication strategy of health organizations and build a location-based community of engaged influencers that support the dissemination of scientific insights, including issues related to vaccines and their safety. All in all, this technology will help the government to win its society so that they want to be vaccinated.

Benefit For Stakeholders

a. **For Government**
   - A well-informed public sentiment of COVID-19 vaccination will help them in **strategy prioritization to accelerate the COVID-19 vaccination rate** in DKI Jakarta. This platform will help to understand the public reaction and aid the policymakers to project the vaccination campaign as well as health and safety measures in the ongoing global health crisis.
   - As the first entity to know how the public reaction towards government policy related to vaccines. Therefore, the government could **mitigate the side effect of public societies' negative sentiment**.

b. **For Public Society**
   - By strengthening the government's prioritization strategy of vaccination, **public society will fastly get vaccinated**. Other than that, it will **help the distribution of vaccination to a specific location that needs to be prioritized** within DKI Jakarta.
   - Covangers in the short term will act as a **platform to capture public complaints towards vaccination** in DKI Jakarta. But in the long term, having a reliable designated Twitter account for users to complain and a platform to map their sentiment will increase trustworthiness towards the government.

Proposed Solution

a. **Sentiment Analysis Modeling**
   Tweets were collected from the Twitter public API stream filtered for keywords related to COVID-19 vaccination. We performed sentiment analysis to have an overview of people's opinions regarding the COVID-19 vaccination.

b. **Location Based Sentiment Modeling**
   Location-based sentiment analysis is the use of natural language processing or machine learning algorithms to extract, identify, or characterize the sentiment content of a 'text unit', according to the location of origin of the text unit. In our product, we use the cases of DKI Jakarta as our origin location, narrow down by the city level (East Jakarta, North Jakarta, Central Jakarta, West Jakarta, and South Jakarta)

c. **Visualization**
   We provide a real-time solution map visualization for the government for the easiest to maintain the public sentiment towards COVID-19 vaccination. This dashboard includes location-based sentiment and other beneficial information for the government.

Technology Concept

a. Lexicon Based Sentiment Analysis

The lexicon-based approach involves calculating orientation for a document from the semantic orientation of words or phrases in the document (Turney 2002). Generally speaking, in lexicon-based approaches, a piece of the text message is represented as a bag of words. Following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message.

The lexicon approach means that this algorithm constructed a dictionary that contains a comprehensive list of sentiment features. This lexical dictionary does not only contain words, but also phrases (such as “bad ass” and “the bomb”), emoticons (such as “:-)”) and sentiment-laden acronyms (such as “ROFL” and “WTF”). All the lexical features were rated for the polarity and intensity on a scale from “-4: Extremely Negative” to “+4 Extremely Positive” by 10 independent human raters. The average score is then used as the sentiment indicator for each lexical feature in the dictionary. For example, in Vader, the word “okay” has a positive rating of 0.9, “good” is 1.9 and “great” is 3.1, whereas “horrible” is -2.5, the frowning emoticon “:(“ is -2.2, and “sucks” is -1.5. Vader’s lexicon dictionary contains around 7,500 sentiment features in total and any word not listed in the dictionary will be scored as “0: Neutral”.

Input	neg	neu	pos	compound
"This computer is a good deal."	0	0.58	0.42	0.44
"This computer is a very good deal."	0	0.61	0.39	0.49
"This computer is a very good deal!!"	0	0.57	0.43	0.58
This computer is a very good deal!! :-)"	0	0.44	0.56	0.74
This computer is a VERY good deal!! :-)"	0	0.393	0.61	0.82

To calculate the sentimental score of the entire text, Vader scans the text for known sentimental features, modified the intensity and polarity according to the rules, summed up the scores of features found within the text and normalized the final score to (-1, 1) using function:

$$\frac{x}{\sqrt{x^2 + \alpha}}$$

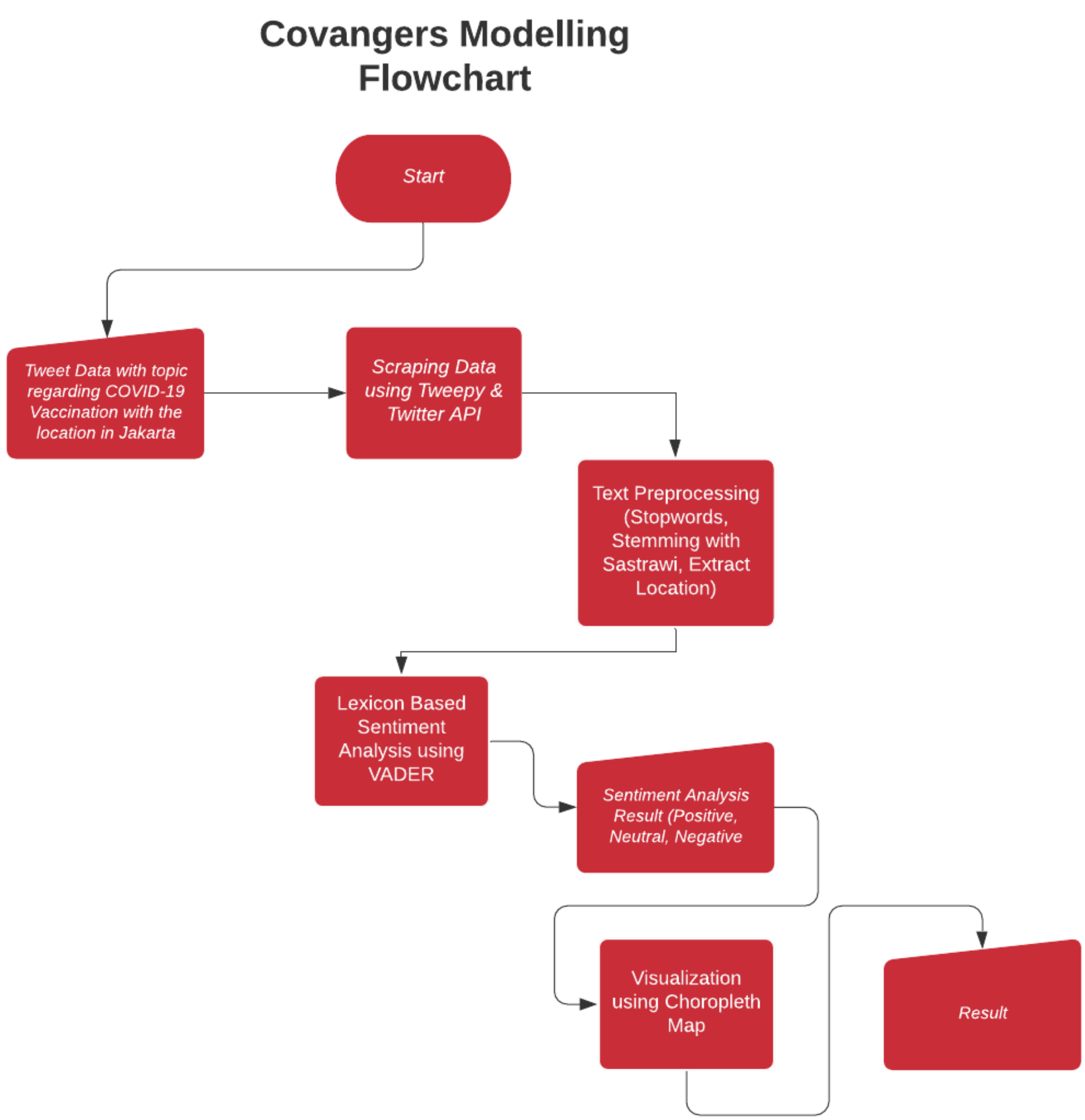
In Vader, alpha is set to be 15 which approximates the maximum expected value of x.In addition to the compound score of the sentence, Vader also returns the percentage of positive, negative and neutral sentiment features, as shown in the previous example.

Success Measurement

We will measure our success with the following metrics in order of importance.

Metrics	Description
COVID-19 Full Vaccination Rate (%)	Rate of DKI Jakarta citizens who already got fully COVID-19 vaccination

Flowchart Modelling



Result

The Lexicon Based Model using VADER successfully identified the sentiment of each tweet in the dataset with an example result of below table.

```
In [61]: df.to_csv("data_tweet_vaksinasi_sentiment_v2.csv", index=False)
```

```
In [62]: df
```

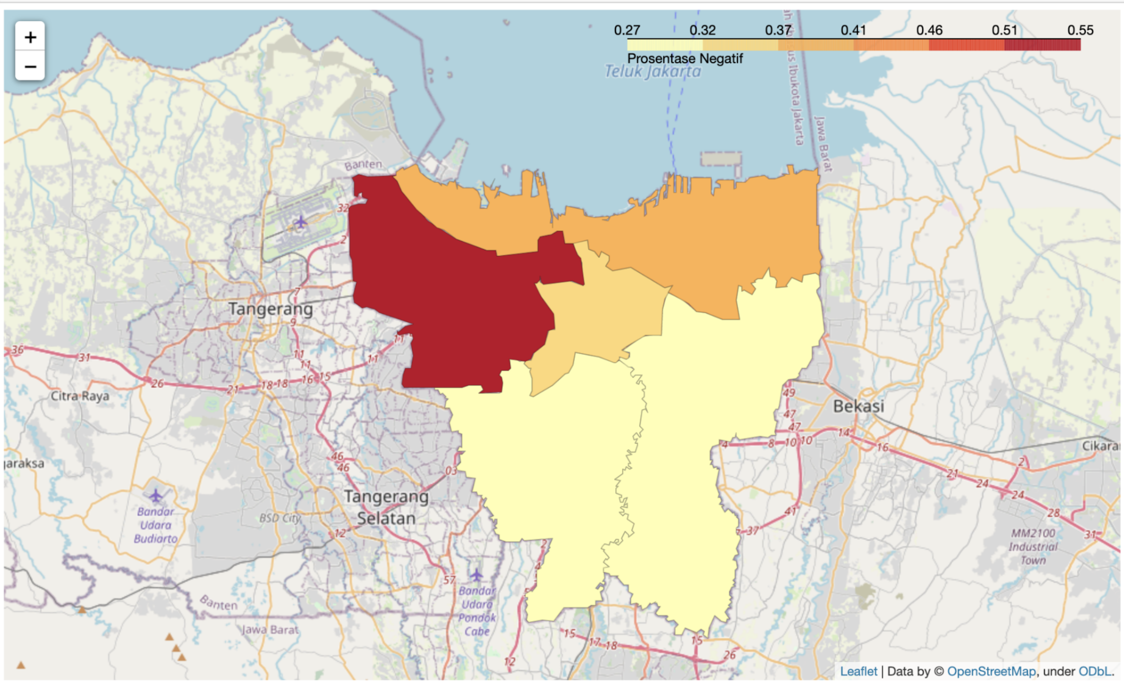
Out[62]:

created_at	user	location	coordinates	text	retweet_count	favorite_count	id	nama_kota	avg_word	word_count	sentiment	analysis
2021-08-14 14:53:43	restrotangkot	Jl. Daan Mogot No.5, Tangerang	NaN	jajar polsek ciledug polres metro tangerang pi...	0	0	ID3172	JAKARTA TIMUR	6.125000	21	2	Netral
2021-08-14 14:53:42	dryxanne	Jakarta	NaN	habis tu bangun lsg brasa lrga hanya mimpi seg...	0	0	ID3172	JAKARTA TIMUR	5.000000	26	-4	Negatif
2021-08-14 14:53:24	bukanlucinta	Jakarta, Indonesia.	NaN	bertanyarl lho syukur bisa astra arti sehat ba...	0	0	ID3172	JAKARTA TIMUR	5.461538	24	19	Positif

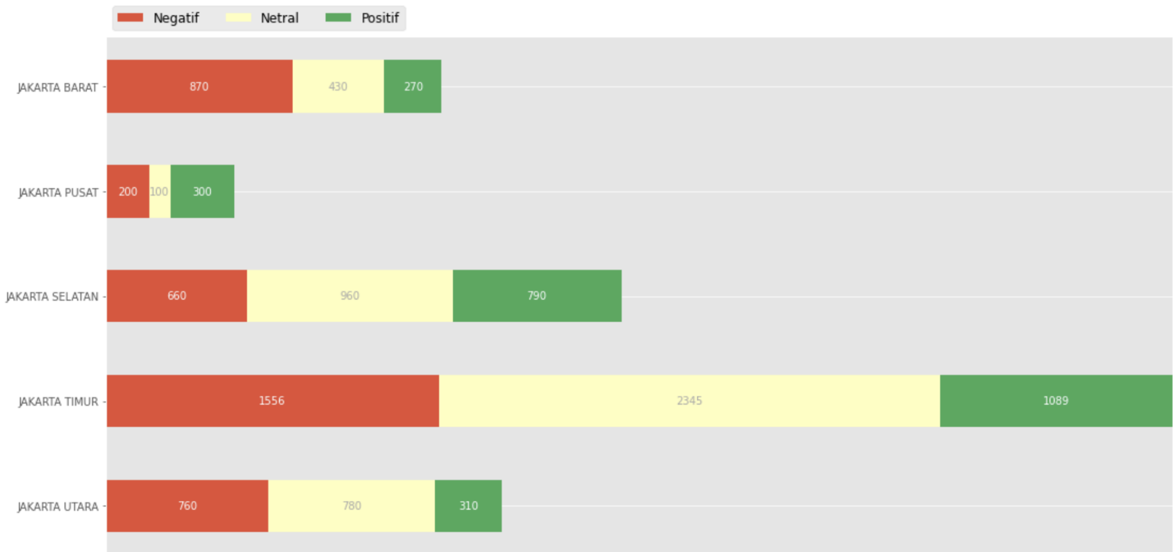
To make it easier, we visualize the modelling result using choropleth map and bar chart. The results below show an example of a negative sentiment in each district of Jakarta.

```
In [55]: m_neg
```

Out[55]:



The map above shows that Jakarta Barat has the highest Negative tweets rate with the topic of Covid-19 Vaccinations, while Jakarta Selatan and Jakarta Timur have the lowest Negative tweets rate out of all districts. We also provided the detailed number of all sentiments (Positive, Neutrals, Negative) in a bar chart below.



Summary

As a summary, we can start to approach people and check the vaccination process in Jakarta Barat as they have the highest negative sentiment rate towards covid-19 and vaccination. Then moving to Jakarta Utara as second highest. Government can utilize this result as an early warning and supporting data to help vaccine distribution better.

Room For Improvement

- By knowing the public behavior/sentiment towards COVID-19 vaccination based on location in DKI Jakarta, we, as a government, can build a strategy for which location that should be prioritized. This location is targeting a pioneer that might affect its neighborhood sentiment.
- Live stream tweet data to get more real-time information
- Directly hit API related to COVID-19
- For the next analysis, we can get a more specific location by its administrative base, not only in DKI Jakarta but also in Indonesia. We can apply the same method to see public sentiment in Indonesia.