Harnessing Social Media Data to Measuring Mental Health Statistics

Action Area D. Modernizing statistical business processes (SD3)

Modernisation is a must – six reason why?

Author:

Fitri Andri Astuti (Presenter)
Nurul Ainun Nisa’
BPS- Statistics Indonesia
Introduction

Mental Health is “a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community”. (WHO)

PROBLEM

“The mental health statistics in Indonesia using surveys is not up to date and costly”

In this time, mental health statistics in Indonesia already count in Ministry of Health and Statistics Indonesia. But some source of the data are from international agency like Mini International Neuropsychiatric Interview (MINI). And beside that, domestic agency also do interview with Self Reporting Questionnaire-20 (SRQ-20) that shared in their website and do survey that called SUSENAS to get some indicators to disaggregate the total cases of mental health.

“Advances in information technology and data storage that made social media data is an incredibly rich data source in both volume and variety.”
Methodology

Mental Health Statistics from Twitter Data

**Data Acquisition**
- Identify keywords related to mental health
- Keyword to querying in Twitter data

**Data Preprocessing**
- Module 1: Extract data tweet from 1 Jan – 30 April 2020 in Twitter Search using GetOldTweets
- Module 2: Extract data tweet from 23 – 30 April 2020 using search Twitter Developer API
- Module 3: Extract data tweet about "covid" from 23 – 30 April 2020 using search Twitter Developer API
- Database MentalHealth Local
- URL Removing Tweet Data
- Stop Word Removing Tweet Data
- Stemming Tweet Data
- Formalizing Text In Indonesia Language

**Data Classification**
- Data Training: Classifying Commercial, News
  To classify tweet contains commercial, news or not
- Classifying Related or Not
  To classify tweet related to mental health or not
- Classifying Mental Health Category
  To classify mental health disorders

**Data Dissemination**
- Database MentalHealth Local
- Data Dissemination:
  To generate query from data classification
- Stop

#apstatsweek2020
Methodology

Data Acquisition

TOP 5 KEYWORDS RELATED TO MENTAL HEALTH INDICATORS THAT USED BY PEOPLE

1. Sedih (Sad) : 38,710 tweets
2. Capek (Tired) : 14,632 tweets
3. Pusing (Headache) : 10,682 tweets
4. Gila (Crazy) : 9,935 tweets
5. Marah (Angry) : 9,222 tweets
1. Classifying Commercial, News

5,069 Data Tweets

4,350 Commercial/News

719 Non commercial/news

Accuracy: 0.939842

2. Classifying Related or Not Related

• 1719 Related to Mental Health
  • 799 related
  • 920 not related

• 7.852 Tweets

3. Classifying Mental Health Category

7.852 Tweets

3,588 Depression;
950 Stress;
726 GAD;
2,588 CMD-NOS

Accuracy: 0.906517
Modul 1

In period 1 January – 30 April 2020 we managed to collect tweet in mental health indicators.

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Tweets</td>
<td>78.236</td>
<td>66.377</td>
<td>75.265</td>
<td>249.243</td>
</tr>
<tr>
<td>Non Comercial/News Tweets</td>
<td>29.358</td>
<td>25.377</td>
<td>27.636</td>
<td>95.354</td>
</tr>
<tr>
<td>Having Relation with Mental Health</td>
<td>14.802</td>
<td>13.315</td>
<td>14.833</td>
<td>46.456</td>
</tr>
</tbody>
</table>
Modul 1

In period 1 January – 30 April 2020 we managed to collect tweet in mental health indicators.
Modul 2

In period 23 - 30 April 2020 we managed to collect 386,695 tweet in mental health indicators by using Twitter API.
Modul 3:

In period 23 – 30 April 2020, we collected tweet that contains keyword “COVID or COVID-19”.

<table>
<thead>
<tr>
<th>Total Tweets</th>
<th>Commercial/News Tweets</th>
<th>Non Commercial/News Tweets</th>
<th>Contain Mental Health Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>102,000</td>
<td>100,851</td>
<td>149</td>
<td>63</td>
</tr>
</tbody>
</table>

#apstatsweek2020
• **Findings :**

1. We developed tools and method to identify potentially relevant content in mental health indicators.

2. To increase the accuracy of the data training that code by human manually, we needs automated methods that will replicate human coded examples.

3. Third, the results of the Twitter analysis shows the nature of the human conversation about mental health disorders.

• **Recommendations :**

1. Explore other social media platform more deeply in order to measuring indicator of mental health statistics.

2. Examining the characteristics and sentiment of the classified text data.

3. Developing a robust search strategy in social media to get relevant data about mental health.
THANK YOU