Covid19-socmed.id
An effort to listen to public during the pandemic

Windy Gambetta

Center for Artificial Intelligence

Institut Teknologi Bandung

Webinar on Making History Together

The Center for Research on Society and Culture, Faculty of Humanities, University of Indonesia

10 July 2020

Speaker

- ## Windy Gambetta
  - Lecturer
    - @ Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung (ITB)
  - Researcher @ Pusat AI ITB

  - Team Leader for Indonesian Work Competency Standardization Development Team, for:
    - Artificial Intelligence/ Data Science
    - Software Development

  - Concultant for National Movement Towards 100 Smart City, Ministry of Communication and Informatics
    - Mimika Regency, Pemalang Regency, Batang Regency, Ambon City, Banjarbaru City, Gresik Regency, Bekasi City
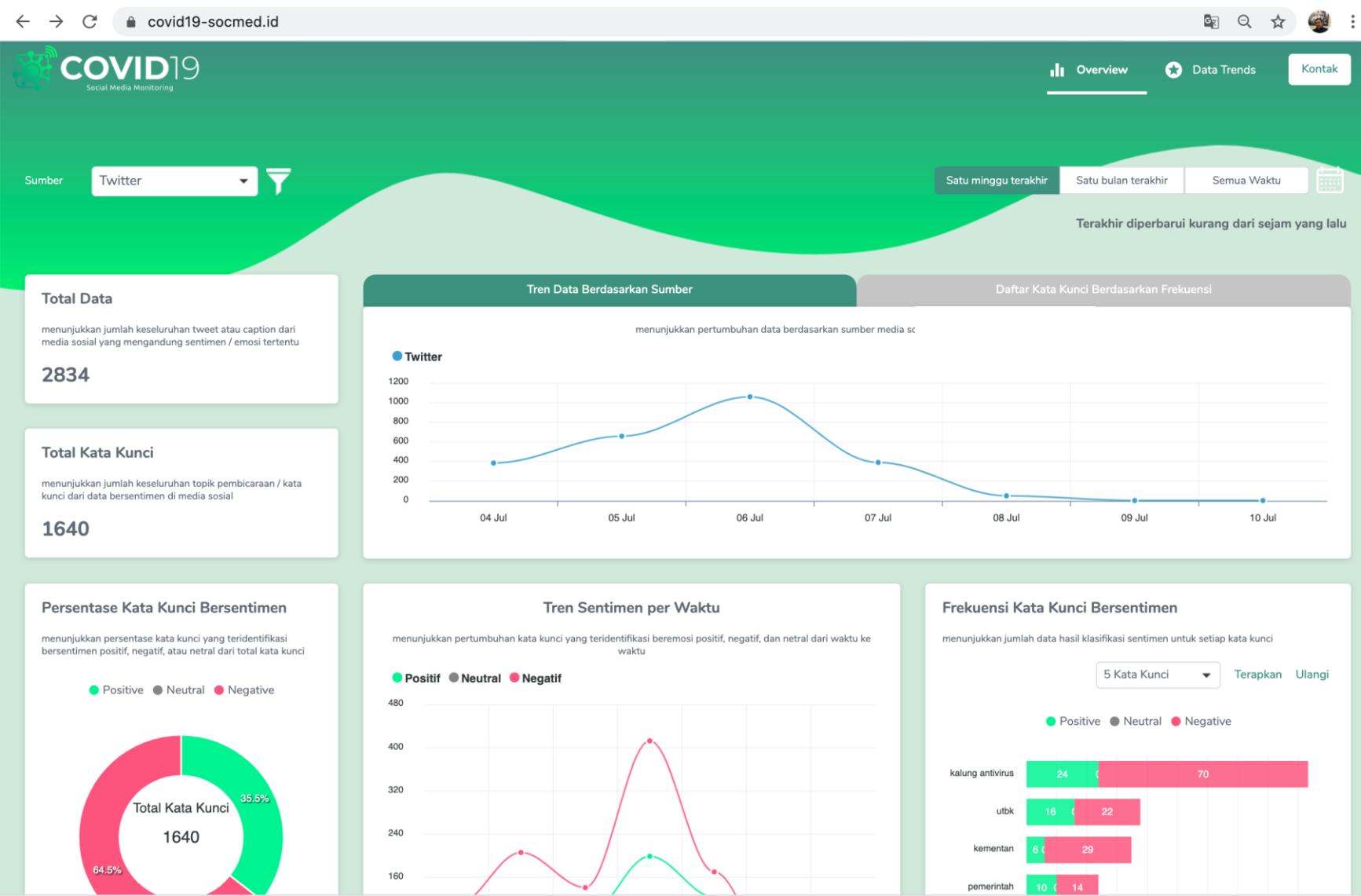
Covid19-socmed.id
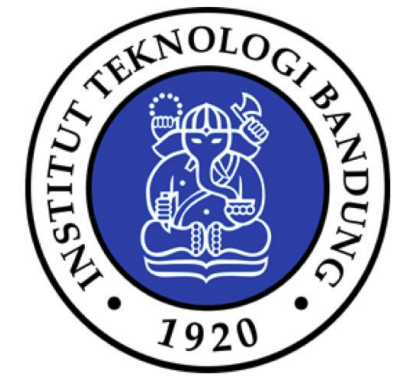
Pusat Artificial Intelligence ITB

Prosa.ai
AI for Bahasa Indonesia

Lead by Dr. Eng. Ayu Purwarianti

Covid19-socmed.id

- An internet website which monitoring various media social platforms to find out how people 'see' covid-19 and how many misinformation/ disinformation around covid-19
    - Input
        - Social media feeds (tweets, posting)
        - [News portal]
    - Output
        - Sentiment analysis
        - Hoax analysis

Sentiment Analysis

- The use of *natural language processing, text analysis, computational linguistics, and biometrics* to systematically **identify, extract, quantify, and study** **Affective States** and **Subjective Information** (Wikipedia)
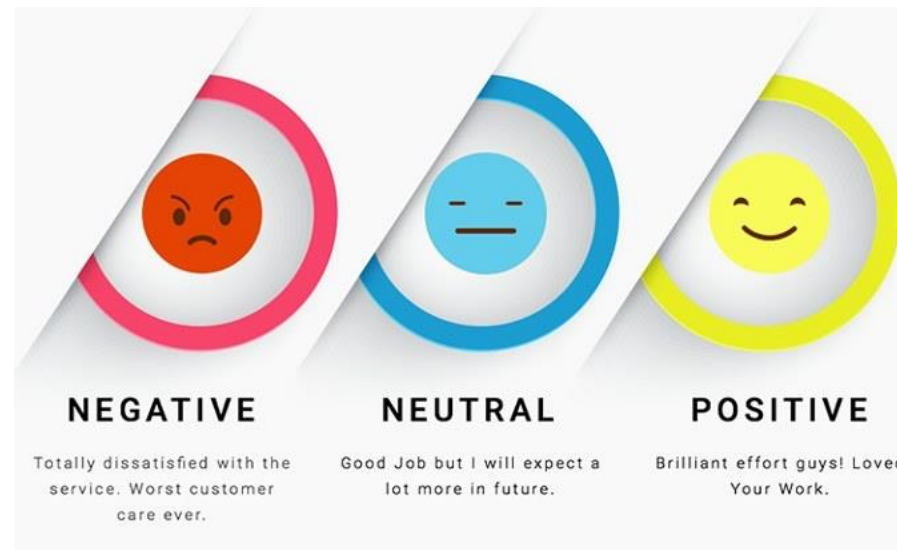
- Social Media Sentiment Analysis
  - Information analysed are from social media (and online news portals)

Hoax Analysis

- The use of *natural language processing, text analysis, computational linguistics, and biometrics* to systematically **identify, extract, quantify, and study** false information deliberately fabricated to masquerade as the truth
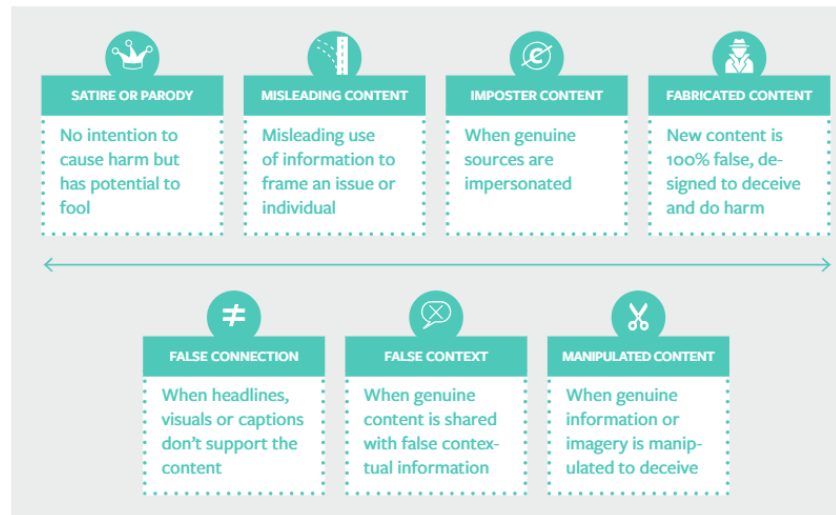
SATIRE OR PARODY	MISLEADING CONTENT	IMPOSTER CONTENT	FABRICATED CONTENT
No intention to cause harm but has potential to fool	Misleading use of information to frame an issue or individual	When genuine sources are impersonated	New content is 100% false, designed to deceive and do harm

FALSE CONNECTION	FALSE CONTEXT	MANIPULATED CONTENT
When headlines, visuals or captions don't support the content	When genuine content is shared with false contextual information	When genuine information or imagery is manipulated to deceive

Figure 4: Seven categories of 'information disorder' - *firstdraftnews.org*

- Social Media Hoax Analysis
  - Information analysed are from social media (and online news portals)

How Indonesia Fight against Hoax

How We Do It

1 Data Collecting

2 Data Cleansing

3 Analytics: Sentiment

4 Analytics: Hoax Analysis

covid19-socmed.id
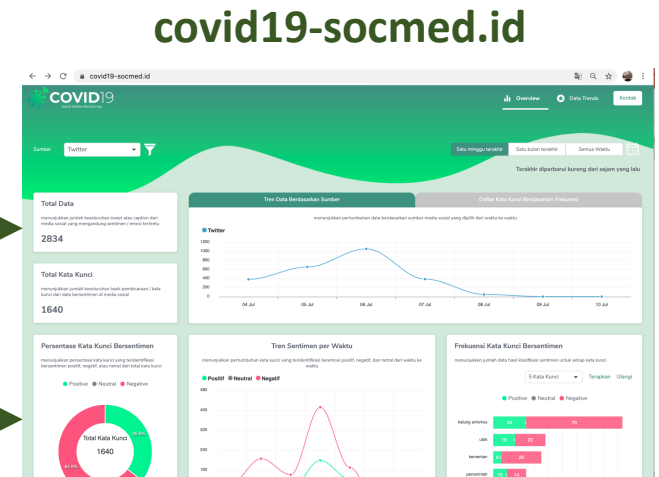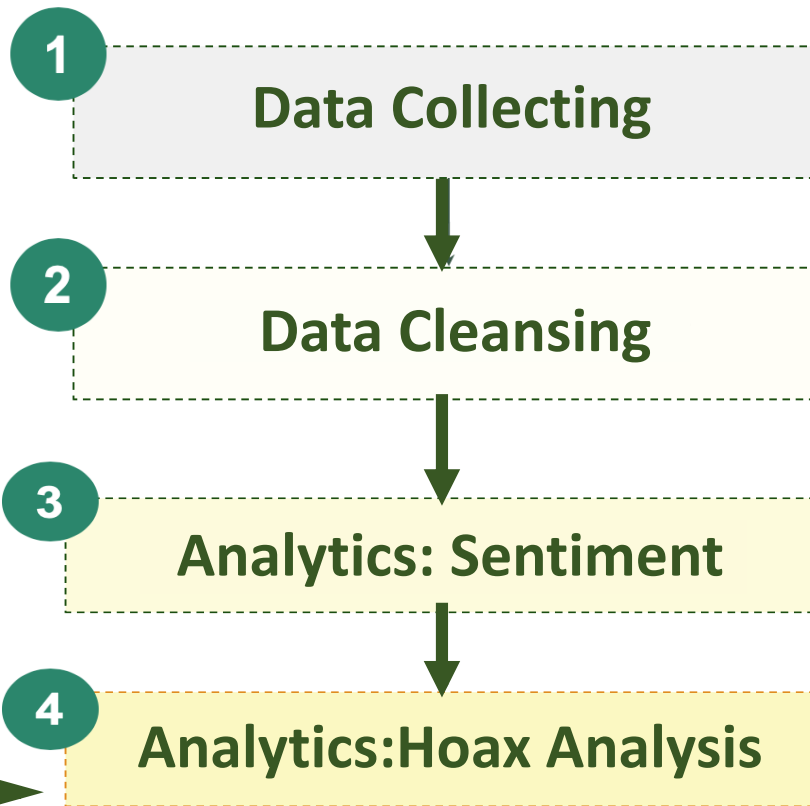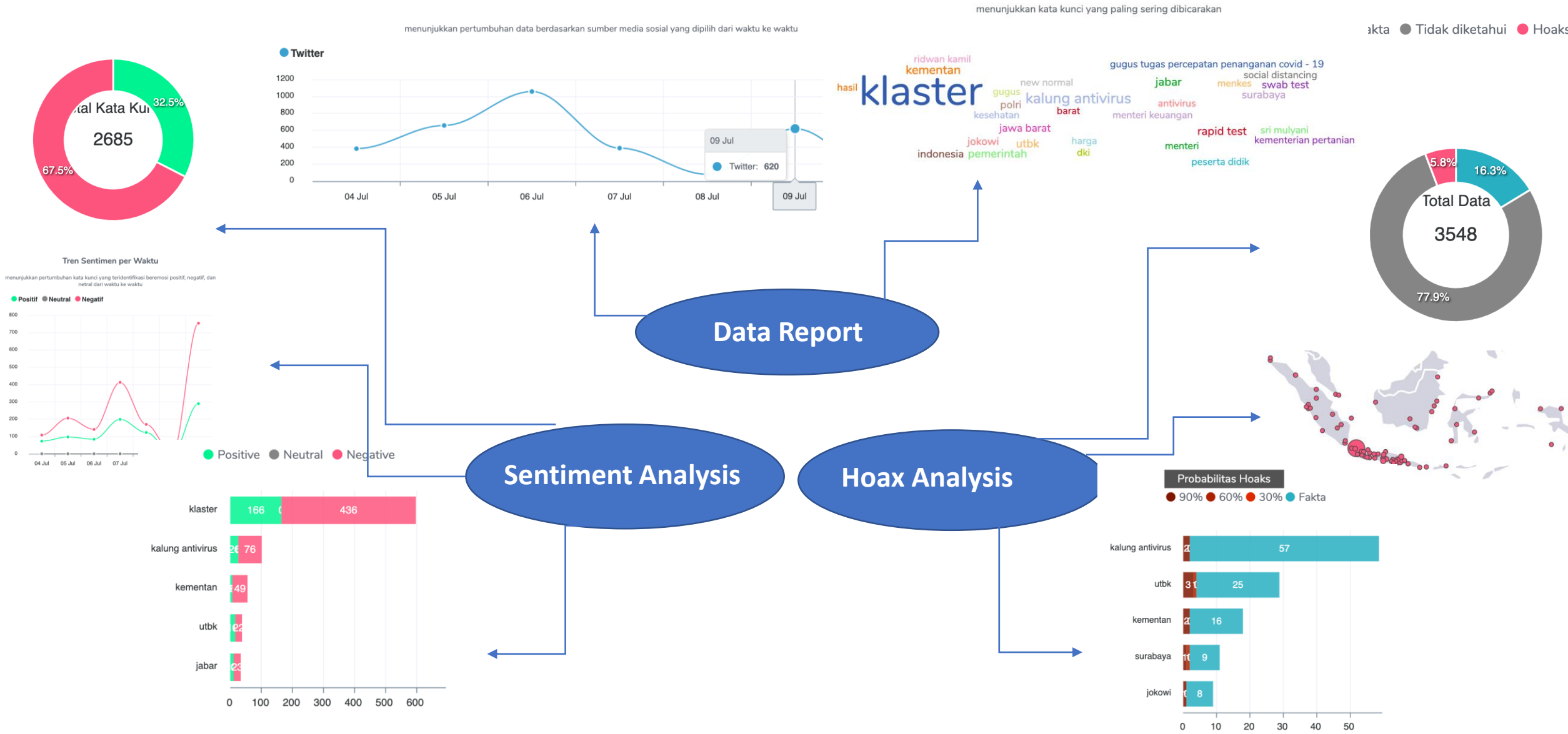
What We Implemented

What We Found so far (1)

- Background Information
  - **2 March** 2020 *First Indonesia patient* confirmed
  - **15 March** 2020 President urged to apply some forms of *social distancing* after 117 patients confirmed
  - **31 March** 2020 President signed Presidential Order No 21 Year 2020 Policy on **Large-Scale Social Restrictions (PSBB)** and declared this as a national disaster
  - **6 April** 2020 Jakarta became *the first city declared Large-Scale Social Restrictions (PSBB)*
  - Various cities and provinces declared PSBB after that
  - **Late May** 2020, **new normal** started to be applied

What We Found so far: General Data

Total Data

menunjukkan jumlah keseluruhan tweet atau caption dari media sosial yang mengandung sentimen / emosi tertentu
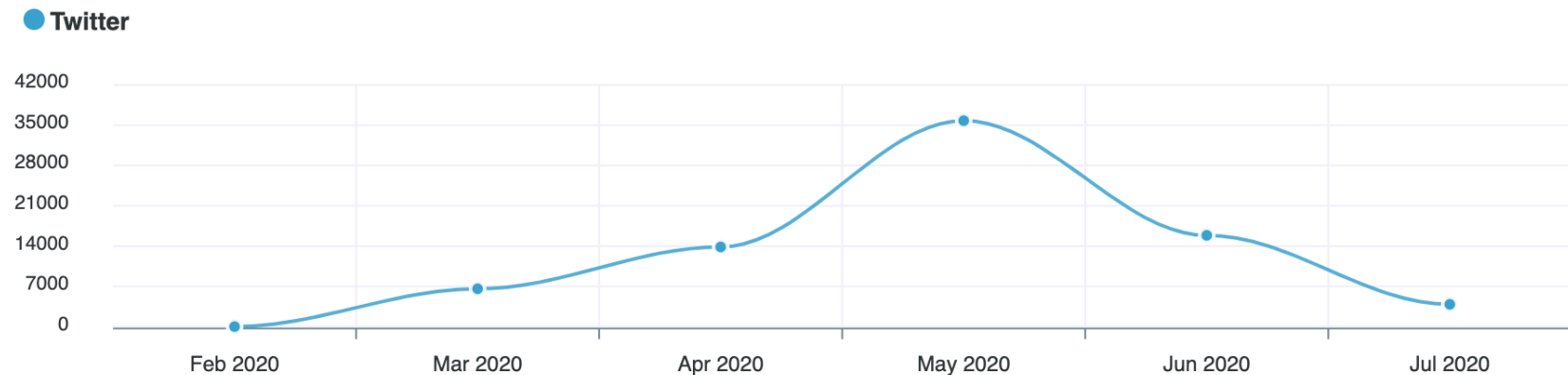
75956

Total Kata Kunci

menunjukkan jumlah keseluruhan topik pembicaraan / kata kunci dari data bersentimen di media sosial

71708

Tren Data Berdasarkan Sumber | **Daftar Kata Kunci Berdasarkan Frekuensi**

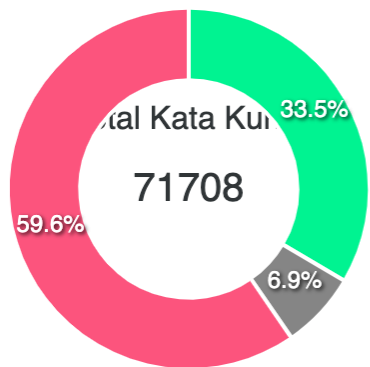menunjukkan pertumbuhan data berdasarkan sumber media sosial yang dipilih dari waktu ke waktu

What We Found so far: Sentiment Analysis

Persentase Kata Kunci Bersentimen

menunjukkan persentase kata kunci yang teridentifikasi bersentimen positif, negatif, atau netral dari total kata kunci
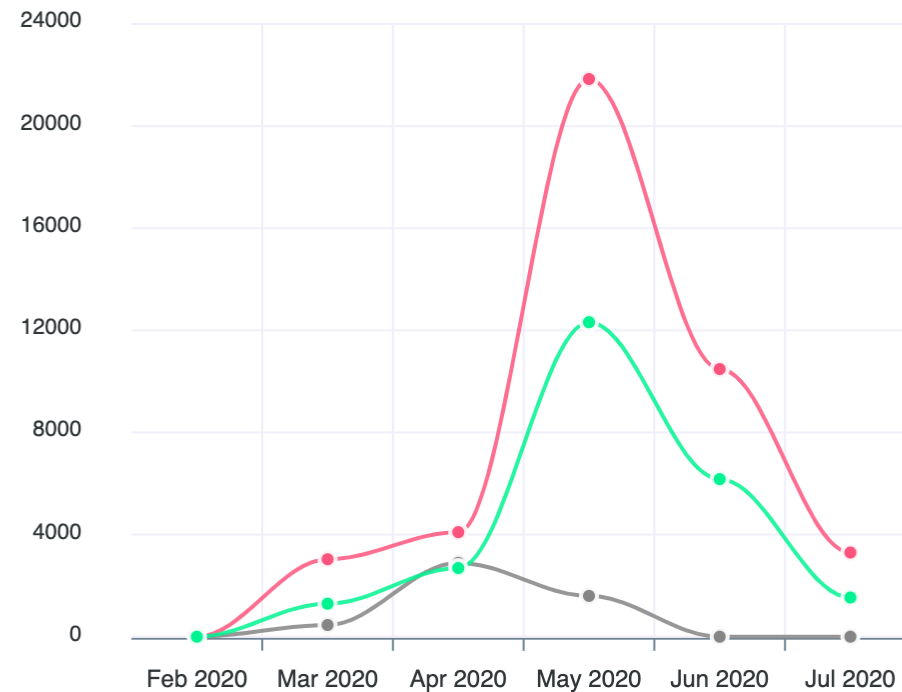
● Positive ● Neutral ● Negative

Total Kata Kunci
71708

33.5%
6.9%
59.6%

Tren Sentimen per Waktu

menunjukkan pertumbuhan kata kunci yang teridentifikasi beremosi positif, negatif, dan netral dari waktu ke waktu
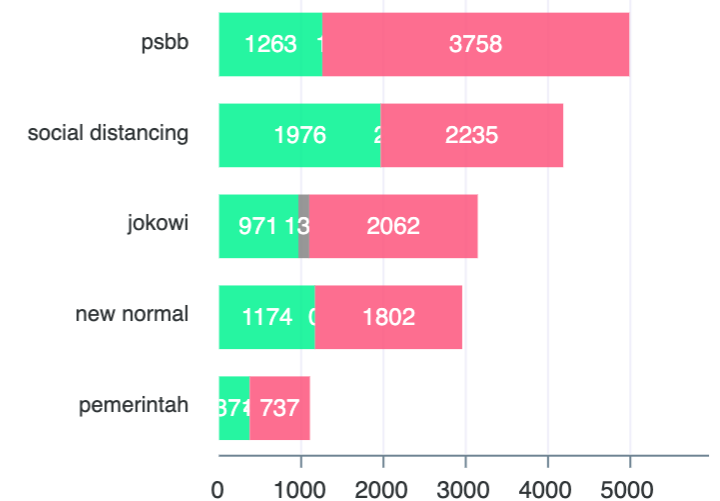
● **Positif** ● **Neutral** ● **Negatif**

24000
20000
16000
12000
8000
4000
0

Feb 2020   Mar 2020   Apr 2020   May 2020   Jun 2020   Jul 2020

Frekuensi Kata Kunci Bersentimen
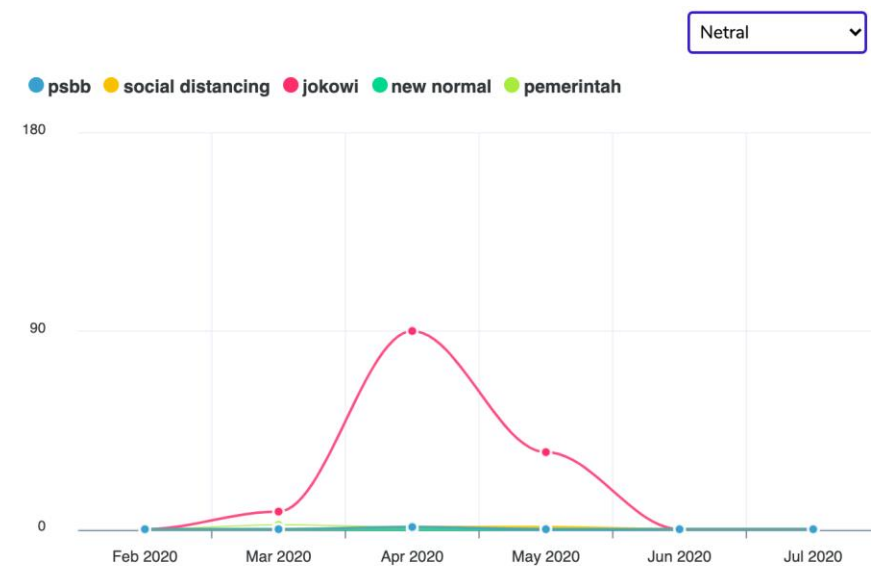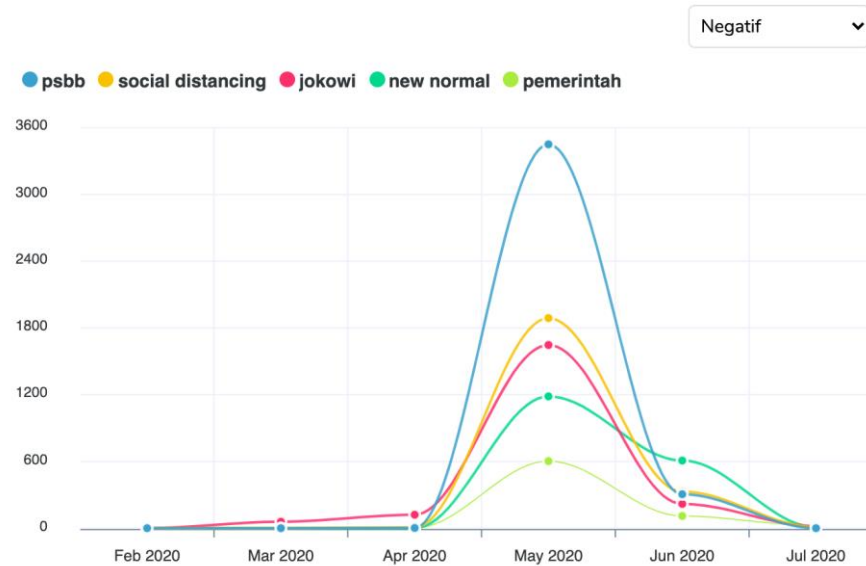
menunjukkan jumlah data hasil klasifikasi sentimen untuk setiap kata kunci

[5 Kata Kunci ▼]   Terapkan   Ulangi
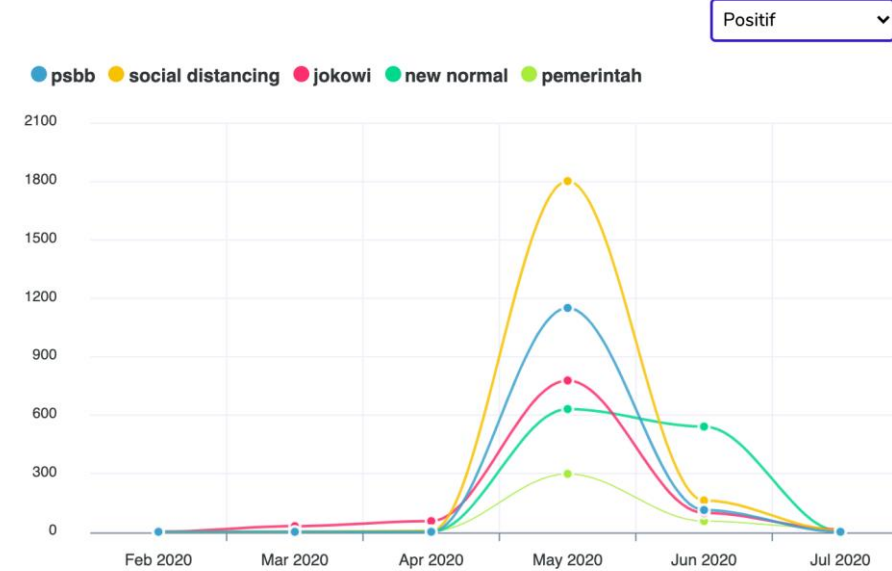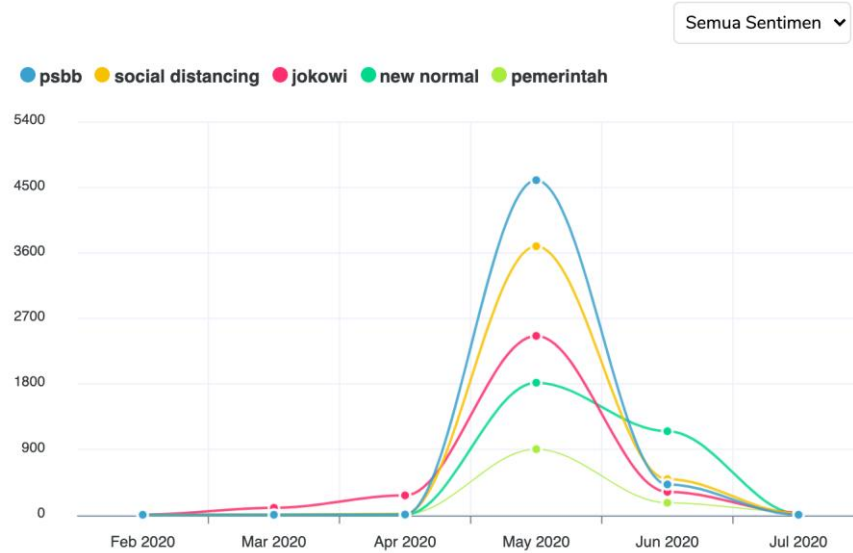
● Positive ● Neutral ● Negative

psbb — 1263 | 1 | 3758
social distancing — 1976 | 2 | 2235
jokowi — 971 | 13 | 2062
new normal — 1174 | 0 | 1802
pemerintah — 37 | 737

0   1000   2000   3000   4000   5000

What We Found so far: Sentiment Analysis

What We Found so far: Sentiment Analysis

March | April | May | June 2020

March
- menkes: 123 | 38 | 328
- jokowi: 29 | 8 | 58
- menhub: 30 | 55
- anies: 19 | 13 | 39
- presiden: 4 | 4 | 29

April
- jokowi: 55 | 90 | 121
- mudik: 50 | 28 | 124
- polri: 27 | 171 | 3
- anies: 14 | 25 | 27
- pola: 1 | 61 | 2

May
- psbb: 1151 | 3451
- social distancing: 1803 | 1889
- jokowi: 778 | 3 | 1647
- new normal: 631 | 1185
- menteri: 245 | 769

June
- new normal: 541 | 608
- peserta didik: 264 | 325
- social distancing: 161 | 329
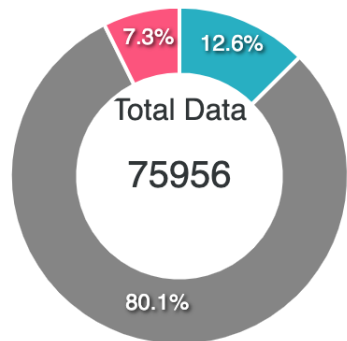- psbb: 112 | 305
- pandemi: 144 | 201
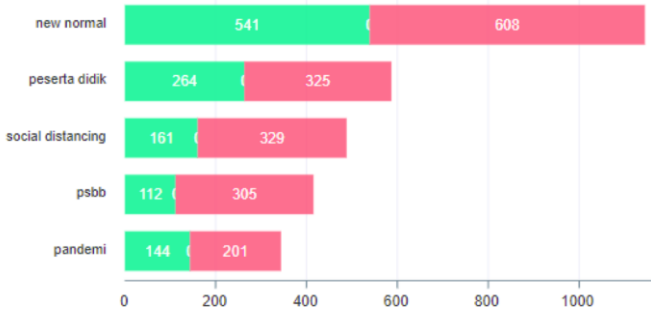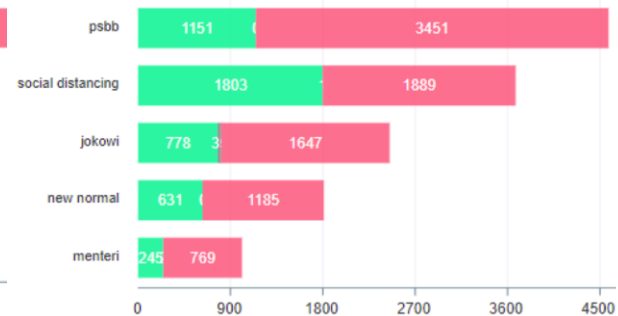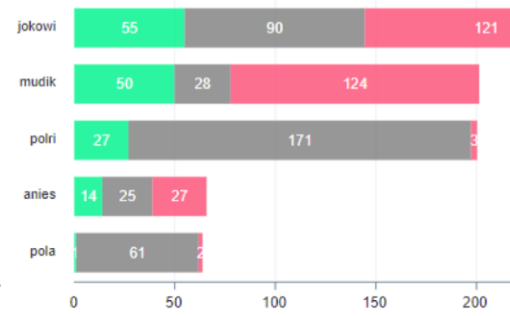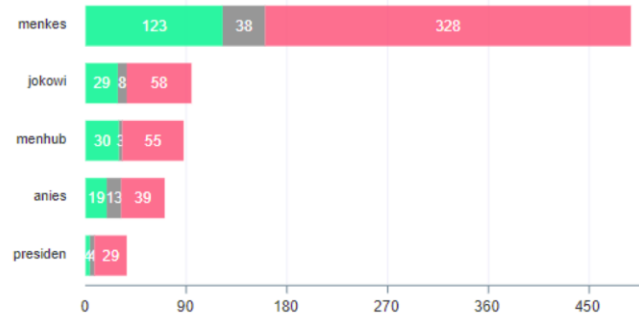
● Positive ● Neutral ● Negative

What We Found so far: Hoax Analysis

Persentase Data Hoaks

menunjukan persentase data yang teridentifikasi sebagai fakta, hoaks, atau bukan keduanya dari total data
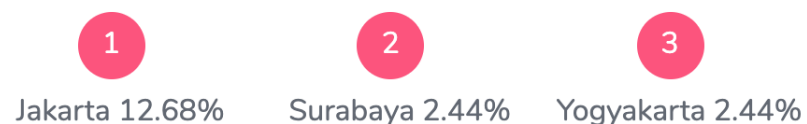
● Fakta ● Tidak diketahui ● Hoaks

7.3% 12.6%

Total Data
75956

80.1%

Distribusi Lokasi Penyebaran Hoaks

menunjukkan jumlah data yang teridentifikasi sebagai hoaks pada setiap kota di Indonesia

Kota dengan Jumlah Hoaks Tertinggi

menunjukkan tiga kota dengan persentase jumlah hoaks tertinggi

1 Jakarta 12.68%
2 Surabaya 2.44%
3 Yogyakarta 2.44%

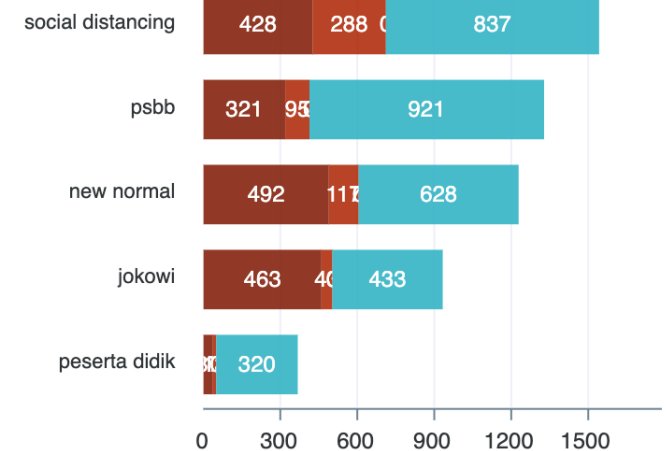## Frekuensi Probabilitas Hoaks per Kata Kunci

menunjukkan jumlah data hasil klasifikasi hoaks pada setiap kata kunci

5 Kata Kunci ▾   Terapkan   Ulangi
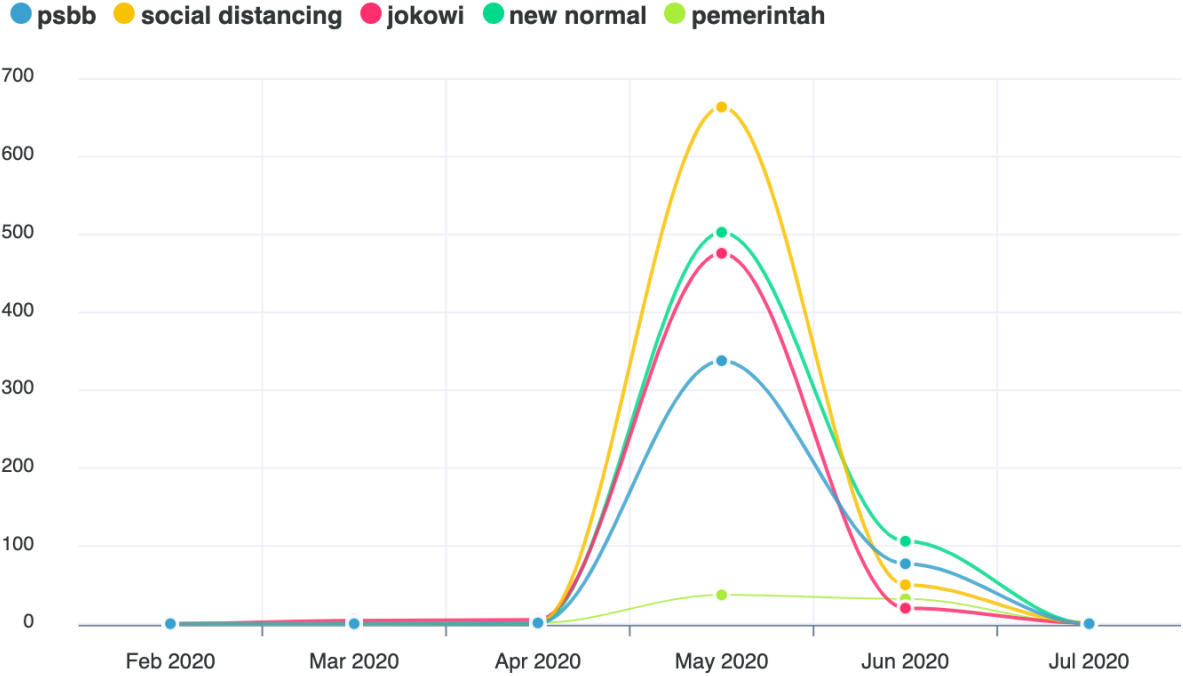
Probabilitas Hoaks
● 90% ● 60% ● 30% ● Fakta

social distancing   428   288   0   837
psbb   321   95   921
new normal   492   117   628
jokowi   463   40   433
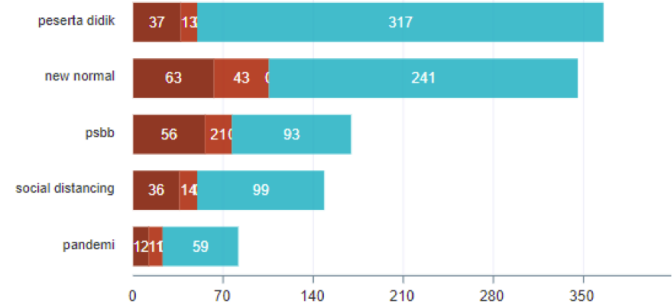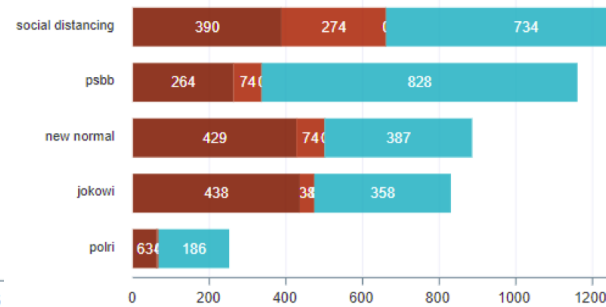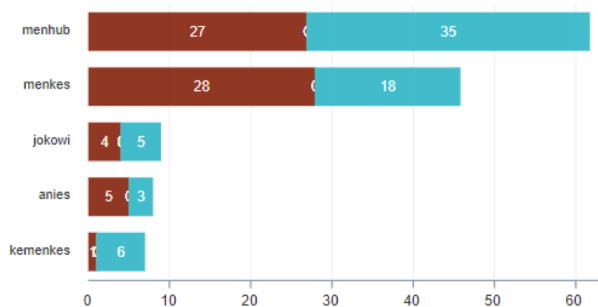peserta didik   30   320

0   300   600   900   1200   1500

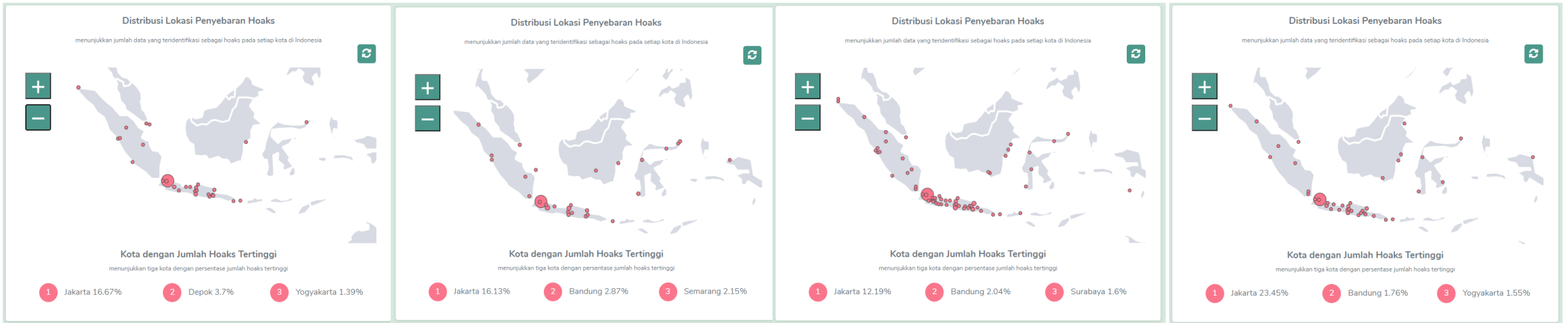# What We Found so far: Hoax Analysis

What We Found so far: Hoax Analysis

Closing

- **Social media feeds** and **News portal** can be **analysed** using **Artificial Intelligence Technology**

- Two types of analysis are conducting in covid19-socmed.id
    - Sentiment Analysis
    - Hoax analysis

- Data collected can be analysed further to find out what people's opinion during the pandemic

- It may be of beneficial when someday in the future we look back to see how we discuss the covid19 in the year of 2020.