

GOOGLE TRENDS

A. APA ITU GOOGLE TRENDS

Google trends adalah sebuah fitur yang diberikan oleh Google untuk menyediakan informasi terkait penelusuran topik tertentu. *Google trends* dapat memberikan informasi penelusuran topik tertentu dari level global (dunia) sampai dengan level kota. *Google trends* bersifat anonim, terkategori, dan agregat.

Google trends disajikan dalam bentuk indeks normalisasi dari sejumlah besar sampel penelusuran kata kunci di Google. Google menggunakan sampel untuk menghitung indeks *google trends* karena ada milyaran orang yang melakukan penelusuran dengan Google. Oleh karena itu, proses penghitungan indeks akan sangat berat dan lama jika menggunakan keseluruhan data penelusuran. Namun demikian, Google mengklaim bahwa indeks ini representatif untuk menggambarkan minat penelusuran masyarakat di mesin pencari Google. *Google* melakukan normalisasi penelusuran kata kunci untuk mempermudah perbandingan antar kata kunci.

Indeks normalisasi *google trends* berada pada rentang 0-100. Nilai 100 menunjukkan penelusuran tertinggi sebuah kata kunci tertentu pada rentang waktu tertentu. Nilai indeks 50 menunjukkan tingkat penelusuran sebuah kata kunci adalah separuh dari jumlah tertinggi penelusuran kata tersebut di rentang waktu tertentu. Nilai 0 menunjukkan tidak tersedia cukup data/tidak ada penelusuran kata kunci tersebut. Secara matematis, indeks normalisasi *google trends* dapat ditulis sebagai berikut:

$$Gt_{ip} = \frac{KW_{ip}}{\max(KW)_p} \times 100\%$$

Gt_i adalah indeks *google trends* pada titik waktu i pada interval waktu p ,

KW_i adalah jumlah penelusuran kata kunci tertentu pada titik waktu i pada interval waktu p . $\max(KW)_p$ adalah jumlah penelusuran tertinggi kata kunci tertentu pada interval waktu p .

Dalam penghitungan indeksnya, Google melakukan beberapa *pre-processing* data. Beberapa tahapan pre-processing yang dilakukan adalah:

1. Google hanya menghitung penelusuran populer. penelusuran yang dilakukan oleh sedikit orang akan muncul sebagai "0"
2. penelusuran yang dilakukan secara berulang oleh orang yang sama dalam rentang waktu yang singkat hanya akan dihitung sebagai satu kali penelusuran
3. Karakter tertentu, seperti tanda petik, tanda tanya, dll dalam penelusuran akan dieliminasi.

Ada dua jenis indeks yang disediakan *google trends*, yaitu

- Indeks *real time* yang menyediakan informasi data penelusuran pada 7 hari terakhir.
- Indeks *non-realtime* menyediakan informasi data penelusuran dari tahun 2004 sampai dengan 36 jam sebelum penelusuran.

Google trend juga menyediakan fitur penelusuran populer. Pada fitur ini, tren penelusuran harian atau tren penelusuran *realtime*. Tren penelusuran harian menyoroti penelusuran yang melonjak secara signifikan di antara semua penelusuran selama 24 jam terakhir dan diperbarui setiap jam. Pengguna dapat menggunakan tren penelusuran harian untuk melihat apa yang paling diminati orang pada waktu tertentu dan bagaimana peringkat penelusuran dibandingkan satu sama lain. Sedangkan tren penelusuran *realtime* menyoroti penelusuran yang melonjak secara signifikan di antara semua penelusuran terbaru. Tren penelusuran *realtime* adalah kumpulan tgrafik indeks *google trends*, kata kunci penelusuran populer, dan artikel berita dari Google yang

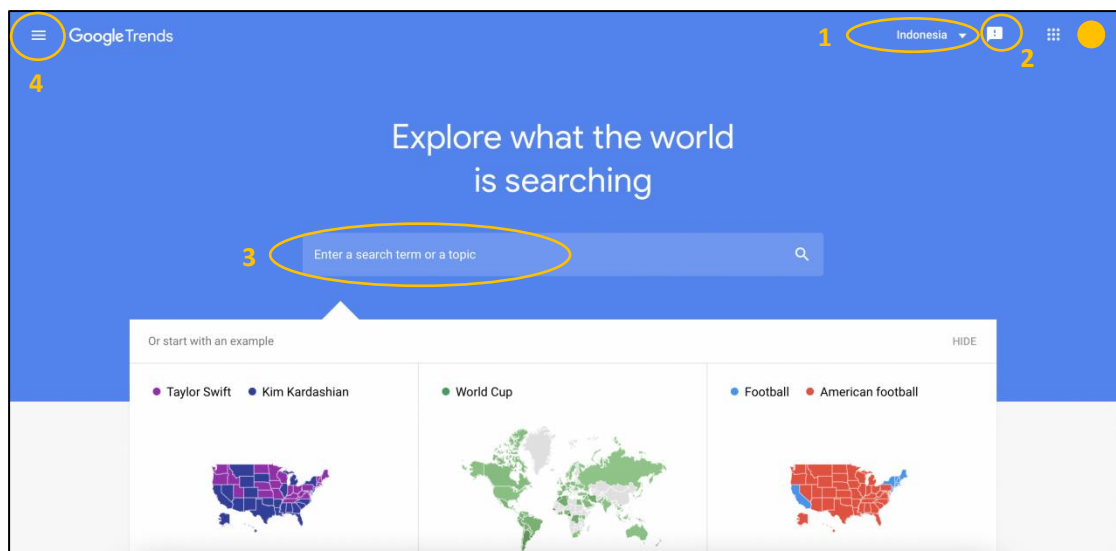
bersesuaian dengan kata kunci tertentu yang terdeteksi oleh algoritma Google. Algoritma untuk tren penelusuran *realtime* mengelompokkan topik yang sedang tren pada saat yang sama baik di Google Berita maupun Penelusuran, kemudian membuat peringkat berita berdasarkan lonjakan penelusuran. Sayangnya, tidak semua negara dapat dilakukan tren penelusuran *realtime*, termasuk Indonesia.

B. CARA MENGAKSES *GOOGLE TRENDS*

Google Trends dapat diakses melalui tautan :

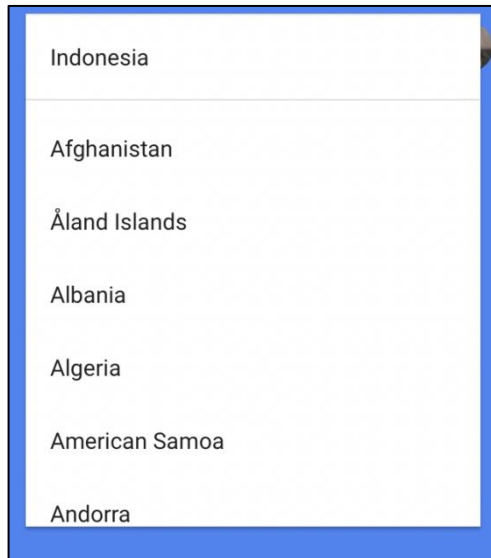
trends.google.com

Halaman pertama pada web *google trends* adalah sebagai berikut



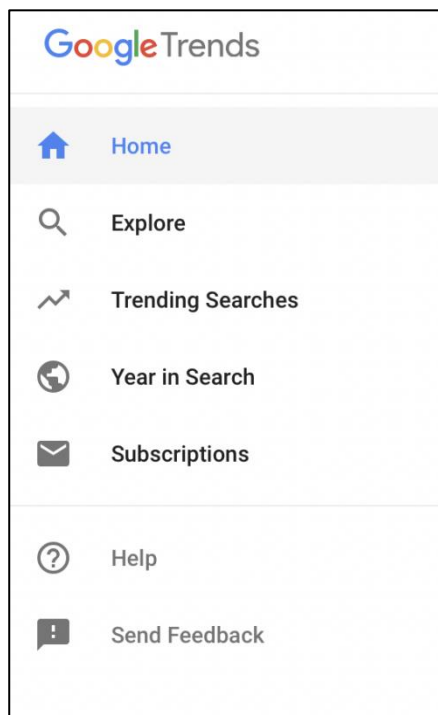
Gambar 1. Tampilan Awal web Google Trends

Ada beberapa fitur pada halaman awal *google trends*. Di sisi kanan atas terdapat pilihan untuk mengubah wilayah penelusuran (Kode 1 pada **Gambar 1.**) Pengguna dapat mengklik tombol segitiga untuk memilih wilayah penelusuran, seperti berikut diperlihatkan pada **Gambar 2.**



Gambar 2. Fitur pemilihan wilayah penelusuran

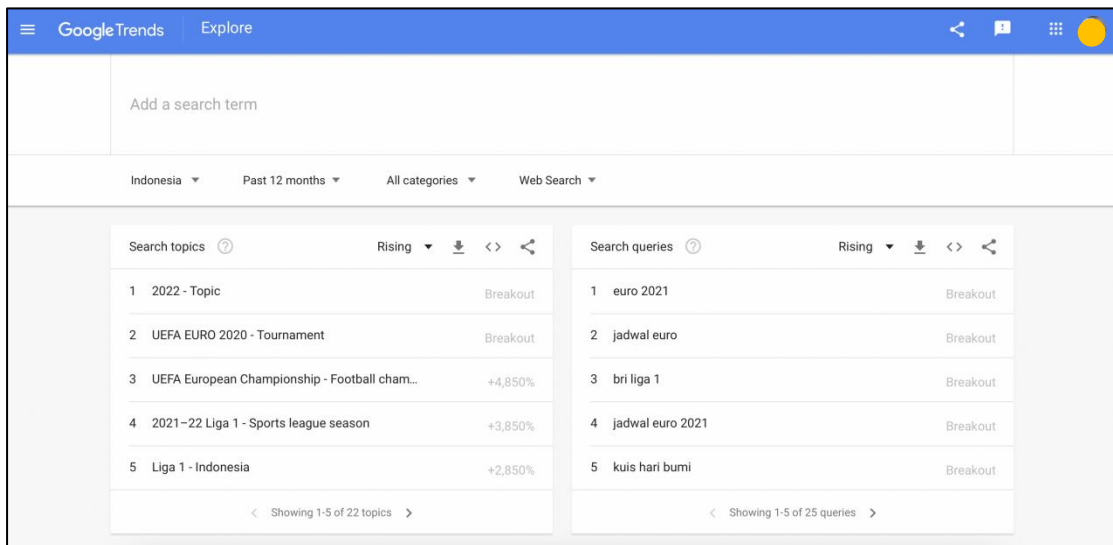
Di sebelah fitur Kode 1 terdapat fitur Kode 2 (**Gambar 1.**) yang berfungsi bagi pengguna untuk memberikan timbal balik atas layanan yang diberikan *google trends*. Fitur Kode 3 adalah fitur bagi pengguna untuk mengeksplorasi kata kunci tertentu pada mesin pencari Google. Kode 4 adalah fitur menu. Beberapa menu yang disediakan oleh google trends ditunjukkan pada **Gambar 3.**



Gambar 3. Fitur menu pada web google trends.

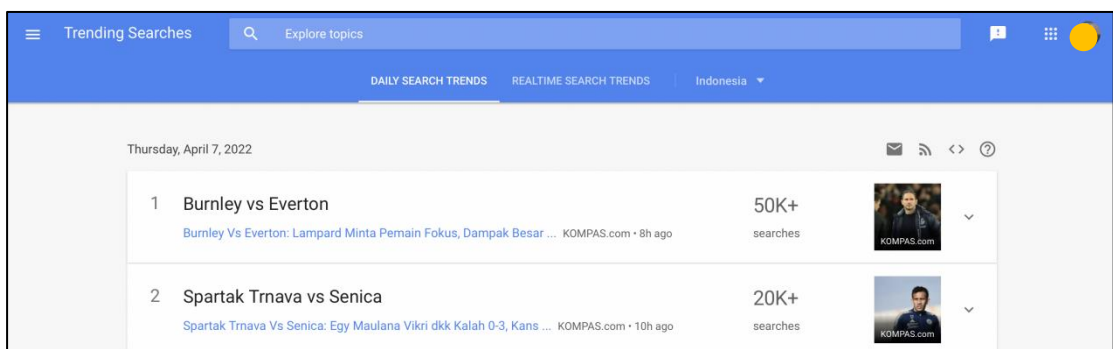
Penjelasan dari fitur menu adalah sebagai berikut:

1. *Home* adalah halaman awal web *Google Trends*.
2. *Explore* adalah halaman bagi pengguna untuk melakukan eksplorasi penelusuran kata kunci di mesin pencari google. Menu ini akan dijelaskan lebih rinci pada bagian C. Tampilan menu *Explore* adalah sebagai berikut.



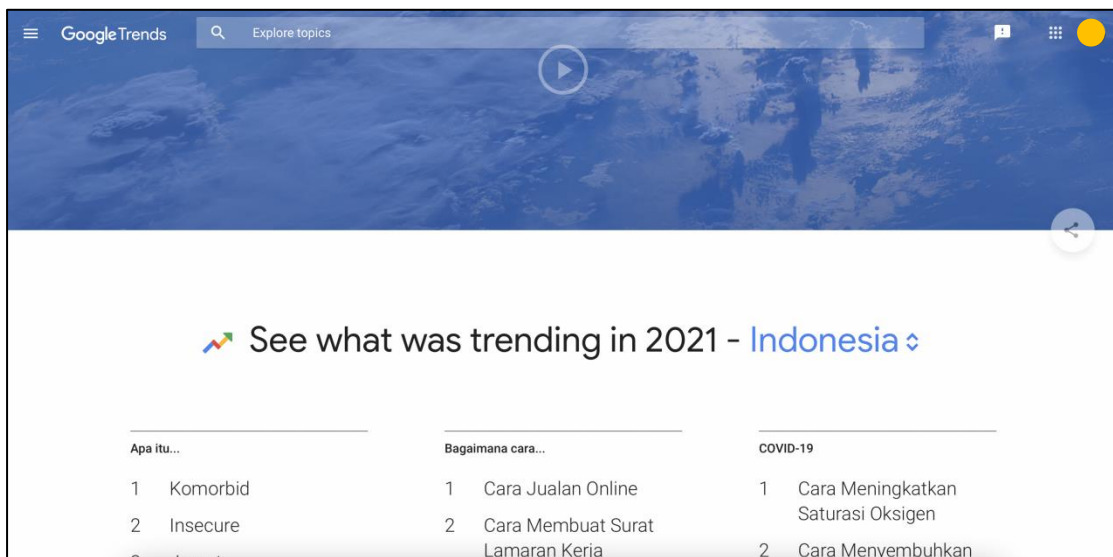
Gambar 4. Fitur menu pada web google trends.

3. *Trending Searches* (penelusuran populer) adalah halaman yang memberikan informasi pencarian yang sedang populer. Pada menu ini disediakan *daily search trends* (tren penelusuran harian) dan *realtime search trends* (tren penelusuran realtime). Tampilan menu *Trending Search* adalah sebagai berikut.



Gambar 5. Fitur *Trending Searches*.

4. *Year in search* adalah menu untuk melihat apa saja yang menjadi penelusuran populer pada tahun tertentu.

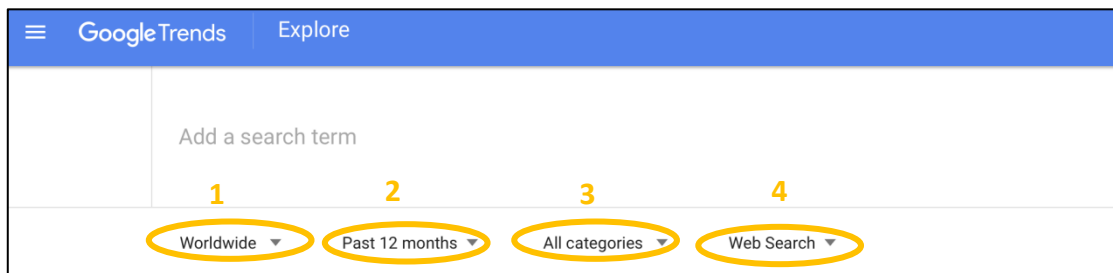


Gambar 6. Fitur *Year in Search*.

5. *Subscriptions* adalah menu untuk berlangganan
6. *Help* adalah menu yang berisi informasi-informasi penting terkait *google trends*.
7. *Feedback* adalah menu untuk memberikan timbal balik atas layanan yang diberikan *google trends*. Menu ini sama dengan menu Kode 2 pada **Gambar 1**.

C. EKSPLORASI PENELUSURAN KATA KUNCI

Indeks *google trends* yang biasa digunakan untuk analisis diperlihatkan pada menu ini. Pada Bagian B telah disajikan tampilan awal menu *Explore*. Pada bagian ini akan dijelaskan lebih rinci apa saja yang dapat dimanfaatkan pada menu tersebut untuk analisis lebih lanjut.



Gambar 7. Menu *explore*

Pada menu *explore*, pengguna dapat mengetikkan kata kunci yang ingin diketahui indeks *google trends* nya. Pengguna juga dapat melakukan mengubahsuaikan penelusuran kata kunci tersebut berdasarkan lokasi (kode 1), interval waktu (kode 2), kategori (kode 3), dan tipe pencarian (kode 4). Kategori (Kode 3) adalah filter yang digunakan untuk membantu pengguna melakukan pencarian yang spesifik sesuai kebutuhan. Sebagai contoh, kata kunci “jaguar” dapat diartikan sebagai salah satu jenis hewan (kategori hewan) atau salah satu jenis kendaraan (kategori otomotif)

Brikut ini adalah contoh penelusuran kata kunci covid (**Gambar 8.**).



Gambar 8. Penelusuran dari waktu ke waktu untuk kata kunci “covid”

Google trends menunjukkan bahwa kata kunci “covid” di Indonesia pada interval

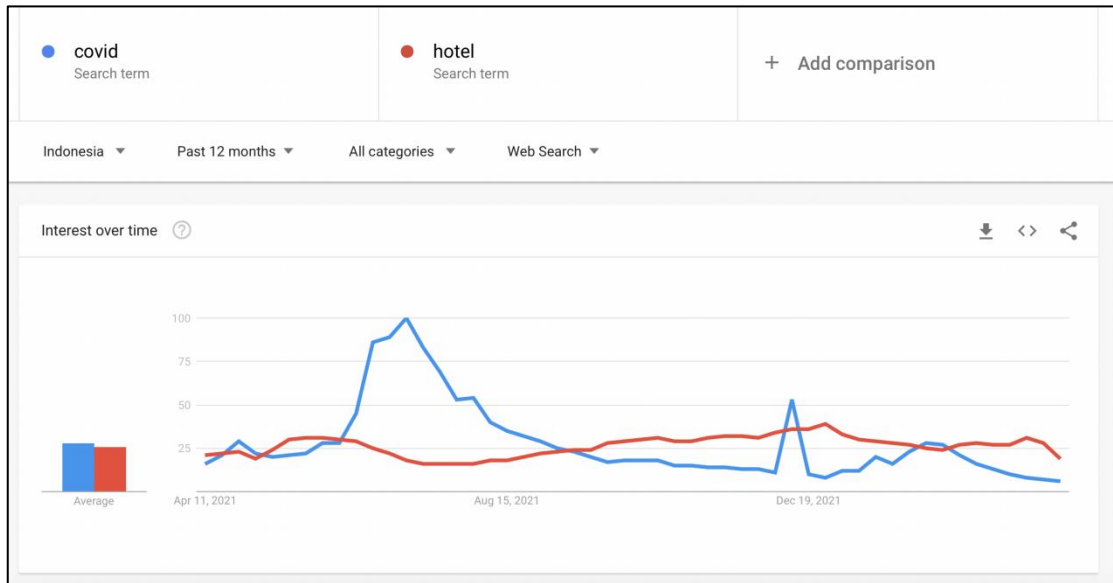
waktu 12 bulan ke belakang paling banyak ditelusuri/dicari pada tanggal 4 s.d 10 Juli 2021, sedangkan pada tanggal 12 s.d 18 Desember 2021 penelusuran kata kunci "covid" adalah sebanyak 53 persen dari penelusuran pada tanggal 4 s.d 10 Juli 2021.

Beberapa interval waktu dan periode indeks yang dapat dipilih adalah:

Tabel 1. Interval waktu dan periode indeks google trends

Interval waktu	Periode indeks
1 jam terakhir	menit
4 jam terakhir	menit
1 hari terakhir	8 menit
7 hari terakhir	jam
30 hari terakhir	harian
90 hari terakhir	harian
12 bulan terakhir	mingguan
5 tahun terakhir	mingguan
2004-saat ini	bulanan
Custom time range	variatif

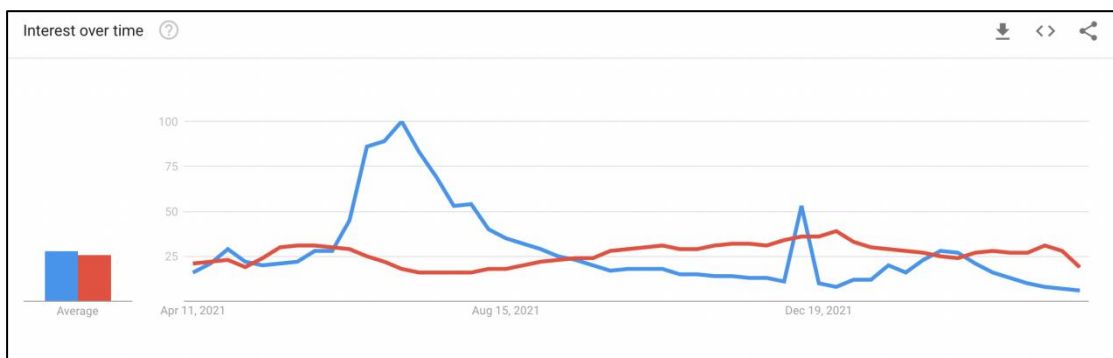
Pengguna dapat melakukan komparasi beberapa kata kunci penelusuran di google trends dengan maksimal 5 kata kunci.



Gambar 9. Penelusuran dari waktu ke waktu untuk kata kunci “covid” dan “hotel”

Beberapa informasi yang dapat dimanfaatkan dari google trends adalah:

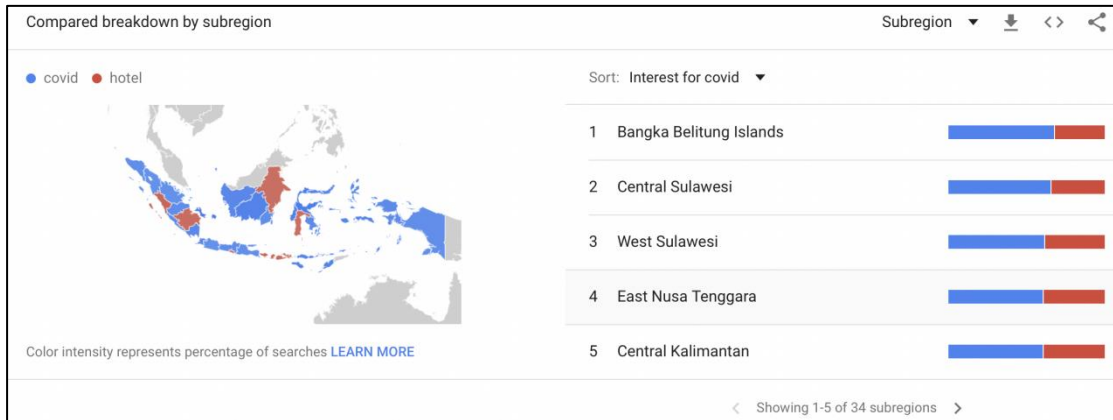
1. *Interest over time*



Gambar 10. *Interest over time*

Interest over time menggambarkan indeks *google trends* dari waktu ke waktu.

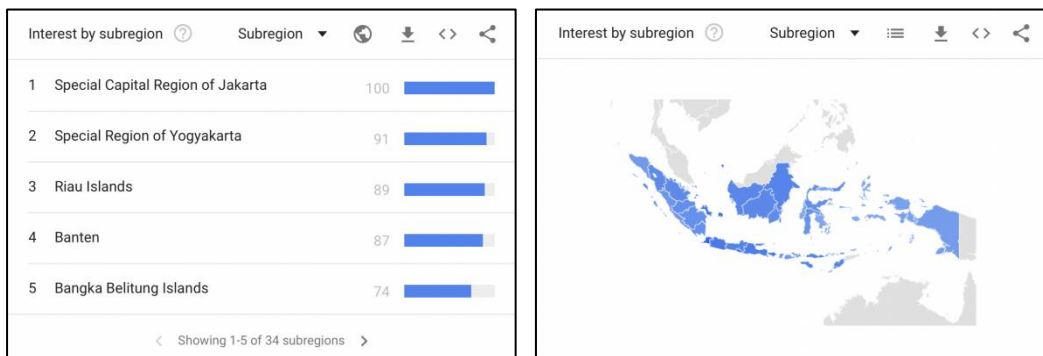
2. *Compared breakdown by subregion* (jika ada lebih dari 1 kata kunci)



Gambar 11. Compared breakdown by subregion

Compared breakdown by subregion menunjukkan persentase masing-masing penelusuran kata kunci terhadap total penelusuran kata kunci yang dibandingkan di suatu wilayah.








3. *Interest by subregion*



Gambar 12. Interest by subregion

Interest by subregion menunjukkan nilai indeks google trends menurut wilayah. Jika suatu wilayah menunjukkan nilai indeks 100 berarti pada wilayah tersebut paling banyak ditemukan penelusuran kata kunci tertentu.

4. *Related queries*

Related queries 		Rising 			
1	lokasi vaksin covid terdekat				Breakout
2	covid omicron				Breakout
3	vaksin covid terdekat				Breakout
4	covid varian delta				Breakout
5	covid vaccine near me				Breakout
 Showing 1-5 of 25 queries 					


Gambar 13. *Related queries*

Related queries memberikan daftar pencarian yang terkait dengan kata kunci yang diamati.

D. DATA COLLECTION

Ada beberapa cara untuk mendapatkan data *google trends*, beberapa diantaranya yaitu: 1) mengunduh nilai *google trends* dari web *google trends*, 2) dengan melakukan *query* menggunakan *python*, 3) dengan melakukan *query* menggunakan R.

1. Mengunduh dari web *google trends*

Cara ini merupakan cara yang paling sederhana. Cara ini dapat dilakukan oleh siapapun tanpa harus memiliki pemahaman terkait pemrograman. Kekurangan dari cara ini adalah masalah efisiensi waktu. Untuk mengunduh nilai *google trends*, pengguna hanya perlu menekan tombol unduh/download () pada setiap indikator yang disajikan maka

pengguna akan mendapatkan file berformat comma delimited (.csv). Setelah mendapatkan file tersebut, pengguna dapat melakukan analisis lebih lanjut sesuai dengan kebutuhan.

2. Query dengan menggunakan *python*

Untuk mendapatkan nilai *google trends* dengan *python*, pertama-tama pengguna perlu memiliki *python* versi 3.3 atau lebih. Indeks *google trends* didapatkan melalui akses API ke *google trends*. Oleh karena itu, pengguna perlu melakukan instalasi package yang digunakan untuk mengakses API *google trends*. Package yang dimaksud adalah **pytrends**. Untuk menginstal package ini, pengguna perlu menjalankan perintah berikut di aplikasi *python*:

```
pip install pytrends
```

Setelah itu, pengguna perlu melakukan permintaan untuk untuk dapat mengakses *google*, dengan perintah:

```
from pytrends.request import TrendReq  
pytrend = TrendReq()
```

Kemudian pengguna dapat membuat payload untuk mendapatkan akses mengambil data sesuai kata kunci yang ingin diamati. Sebagai contoh untuk kata kunci "covid" dan "hotel", perintah yang dijalankan adalah:

```
pytrend.build_payload(kw_list=['covid', 'hotel'])
```

Setelah perintah tersebut dijalankan, pengguna dapat melakukan query untuk mendapatkan data *interest over time*, *interest by region*, dan *related*

query dengan menjalankan perintah berikut:

```
# Interest Over Time
interest_over_time_df = pytrend.interest_over_time()
print(interest_over_time_df.head())

# Interest by Region
interest_by_region_df = pytrend.interest_by_region()
print(interest_by_region_df.head())
```

3. *Query* dengan menggunakan R

Mendapatkan nilai google trends dengan menggunakan R secara umum memiliki algoritma yang sama dengan menggunakan python. Untuk dapat menggunakan bahasa pemrograman R, pertama-tama pengguna perlu mengunduh dan menginstall software R versi terbaru dan Rstudio. Rstudio adalah sebuah tools yang digunakan untuk mengoperasikan bahasa pemrograman R. Kedua software ini dapat diunduh di :

R : <https://cran.r-project.org>

Rstudio : <https://www.rstudio.com/products/rstudio/download/>

Seperti python, pengguna perlu melakukan instalasi *package* untuk dapat mengakses API *google trends*. *Package* yang di maksud adalah **gtrendsR**. Untuk menginstall package tersebut, pengguna perlu menjalankan perintah berikut:

```
install.packages("gtrendsR")
```

Setelah *package* terinstall, lakukan *load package* dengan perintah:

```
library(gtrendsR)
```

Setelah *package* ter-load, pengguna dapat melakukan query untuk kata kunci yang diinginkan dengan menjalankan perintah:

```
gtrends(.....)
```

Ada beberapa argumen yang dapat digunakan di dalam fungsi *gtrends*, yaitu:

Tabel 2. Argumen pada fungsi *gtrends*

Argument	Deskripsi
keyword	Character yang menunjukkan kata kunci <i>google trends</i> yang ingin diamati
geo	Vektor karakter yang menunjukkan wilayah geografis untuk kueri, nilai default merujuk untuk query global.
time	String untuk menspesifikasi rentang waktu untuk query. Nilai yang mungkin digunakan: "now 1-H" : 1 jam terakhir "now 4-H" : 4 jam terakhir "now 1-d" : hari terakhir "now 7-d" : 7 hari terakhir "today 1-m" : 30 hari terakhir "today 3-m" : 90 hari terakhir

	<p>"today 12-m" : 12 bulan terakhir</p> <p>"today+5-y" : 5 tahun terakhir</p> <p>"all" : sejak awal <i>google trend</i> (2004)</p> <p>"Y-m-d Y-m-d" : rentang waktu tertentu, co:</p> <p>"2010-01-01 2010-04-01"</p>
hl	<p>String berisi kode <i>ISO language</i> untuk mengidentifikasi bahasa yang digunakan, co : jika ingin menggunakan bahasa Indonesia, kodenya "id-ID"</p>

Sebagai contoh untuk mengambil query untuk kata kunci "covid" dan "hotel" di wilayah Indonesia menggunakan bahasa Indonesia, maka perintah yang dijalankan adalah:

```
IDN_data <- gtrends(keyword = c("covid","hotel"), geo ="ID", hl = "id-ID")
```

Untuk mendapatkan query menurut time series, provinsi, kota, dan query terkait dapat menjalankan perintah berikut:

```
# Interest Over Time
IDN_timeseries<-as_tibble(IDN_data$interest_over_time)

#interest by region
IDN_region<-as_tibble(IDN_data$interest_by_region)

#interest by city
```

E. ANALISIS

Dari dataframe hasil *query google trend* dapat dilakukan analisis lebih lanjut. Beberapa analisis yang akan dibahas di modul ini adalah analisis deskriptif dan peramalan. *Tool* yang digunakan untuk melakukan analisis ini adalah R.

Untuk melakukan analisis di R diperlukan beberapa packages sebagai berikut:

Tabel 3. Packages untuk analisis google trends

packages	deskripsi
tidyverse	Package untuk eksplorasi dan manipulasi data
sf	Package untuk mengkodekan data vektor spasial
plotly	Package untuk membuat grafik
corrplot	Package untuk visualisasi korelasi
lubridate	Package untuk mengolah data tanggal
prophet	Package untuk peramalan

Semua *package* tersebut perlu diinstall dan di-load dengan menggunakan perintah sebagai berikut

```
install.packages("gtrendsR")
install.packages("tidyverse")
install.packages("sf")
install.packages("corrplot")
install.packages("lubridate")
install.packages("prophet")
install.packages("plotly")
```

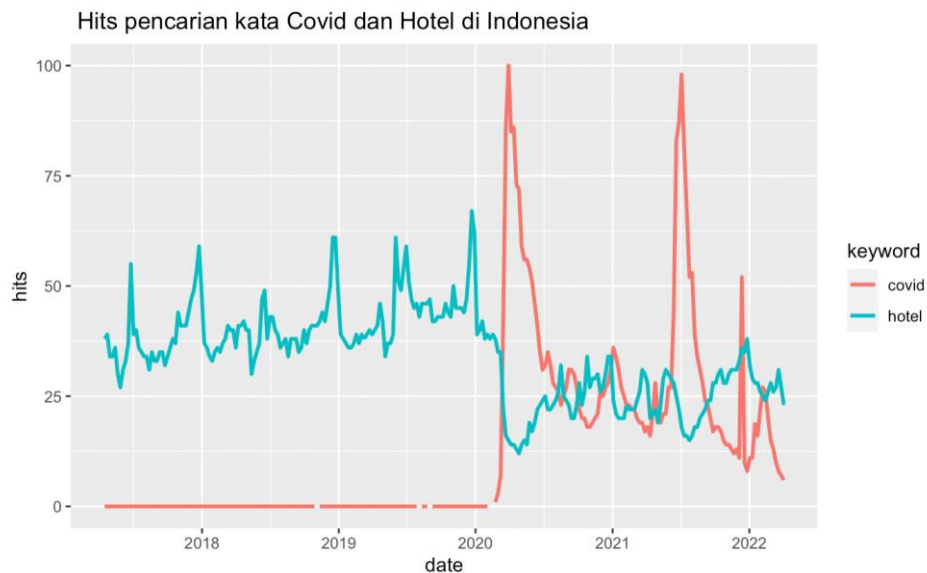
```
gtrends(.....)
```



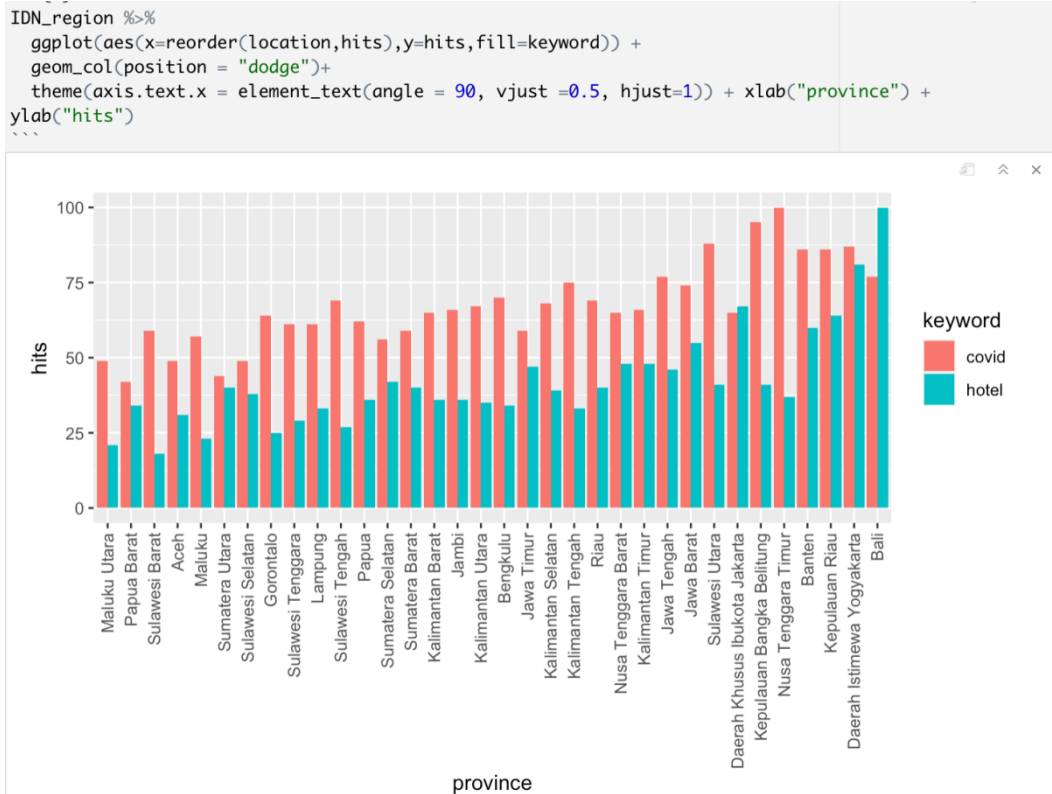
```
library(gtrendsR)
library(tidyverse)
library(sf)
library(plotly)
library(corrplot)
library(lubridate)
library(prophet)
```

Untuk menampilkan grafik time series, pengguna dapat menjalankan perintah berikut:

```
IDN_timeseries %>%
  ggplot(aes(x= date, y = hits, color=keyword)) +
  geom_line(size=1) +
  labs(title=" Hits pencarian kata Covid dan Hotel di Indonesia")
  ...
```



Untuk analisis antar wilayah, dapat dibuat visualisasi dengan menjalankan perintah



Untuk visualisasi dalam bentuk peta, perlu dilakukan pembacaan file peta (shape file) kemudian menggabungkan file tersebut dengan dataframe google trend seperti berikut

```

shp_idn <- sf::read_sf(file.choose())
str(shp_idn)

```

```

shp_idn <- shp_idn %>%
  left_join(y=IDN_region_compile,by = c("nmprov" = "location"),keep=T)

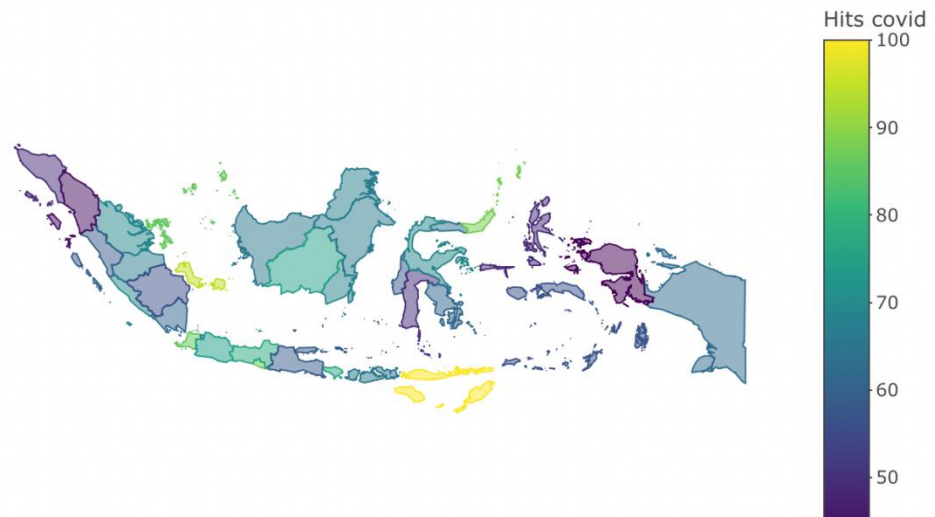
```

Perintah untuk visualisasi adalah sebagai berikut

```
plot_ly(shp_idn, color = ~covid, split = ~nmprov, span = I(1),
  text = ~paste("<b>", nmprov, "</b><br>",
    "<b>Hits covid</b>:", covid,"<br>",
    "<b>Hits hotel</b>:", hotel,"<br>"),
  hoveron = "fills", hoverinfo = "text", showlegend = FALSE) %>%
  colorbar(title = "Hits covid")
` ``
```

R Console

plotly
htmlwidget

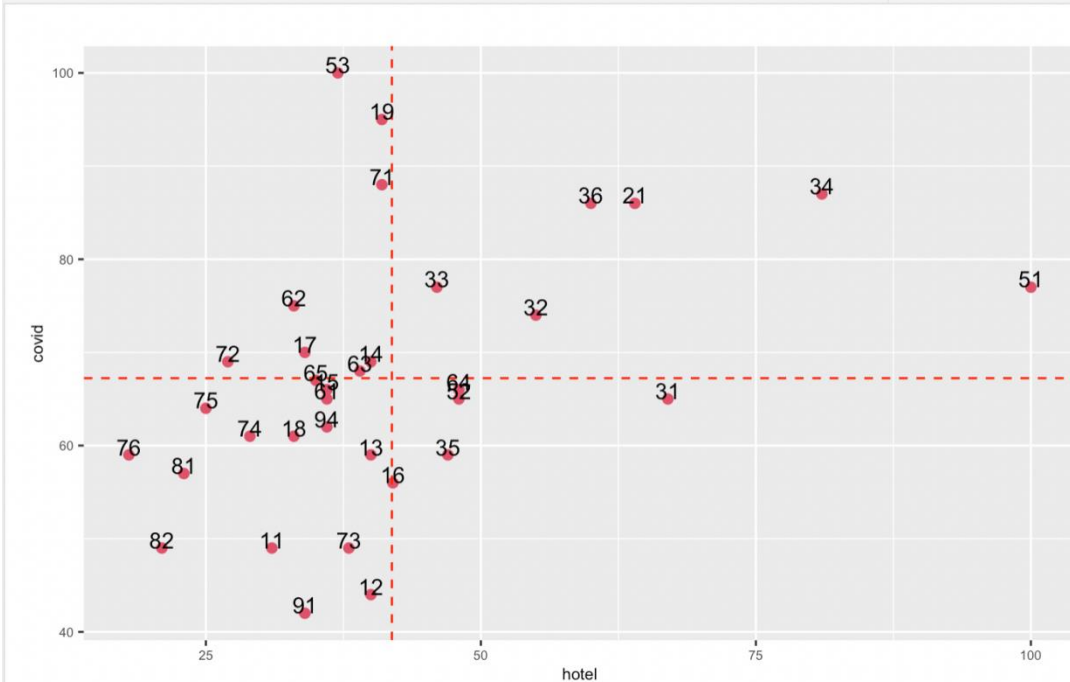


Untuk mengetahui korelasi antar kata kunci yang diamati dapat menggunakan perintah berikut

```

covid_mean = mean(shp_idn$covid)
hotel_mean = mean(shp_idn$hotel)
shp_idn %>%
  ggplot(aes(y=covid,x=hotel, label = kdprov)) +
  geom_point(col=2, size=2, na.rm = TRUE) +
  geom_hline(yintercept= covid_mean, linetype="dashed", color = "red") +
  geom_vline(xintercept= hotel_mean, linetype="dashed", color = "red") +
  geom_text(hjust=0.5, vjust=0) +
  theme(text = element_text(size = 8))
` ``

```



Untuk melakukan peramalan, pertama perlu dibuat model peramalan dengan prophet. Setelah itu menentukan berapa periode yang akan diramal (penentuan periode dalam satuan hari). Langkah terakhir adalah melakukan peramalan untuk periode yang sudah ditentukan

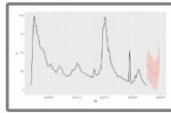
```

IDN_covid_m <- prophet::prophet(IDN_timeseries_covid)
IDN_covid_present_future <- prophet::make_future_dataframe(IDN_covid_m, periods=84)
IDN_covid_predict_data <- as_tibble(predict(IDN_covid_m, IDN_covid_present_future))
IDN_covid_predict_data

```

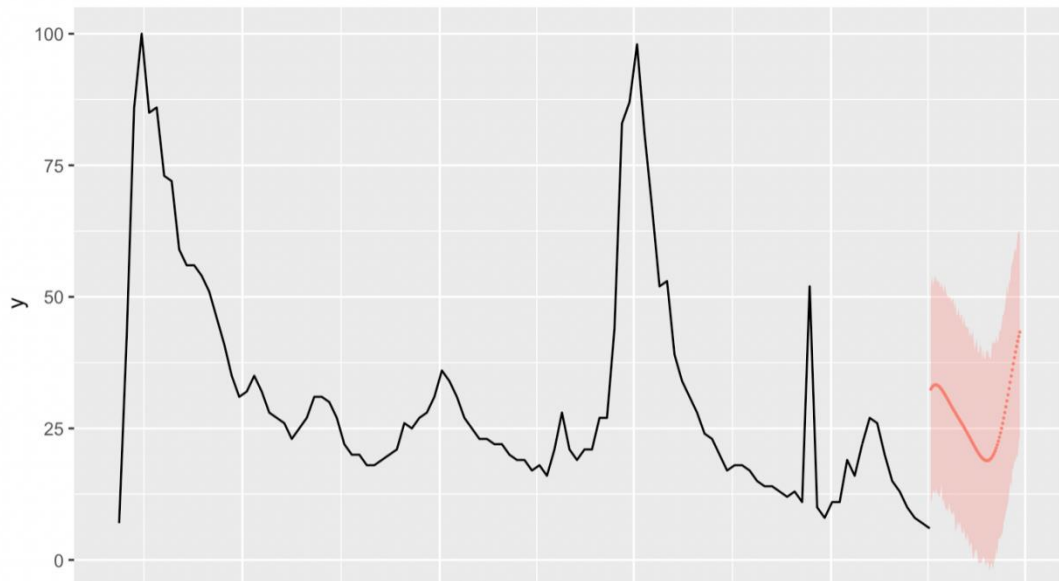
Hasil peramalan dapat divisualisasikan seperti berikut

```
IDN_covid_forecast %>%
  ggplot() +
  geom_line(aes(ds,y))+
  geom_point(data = subset(IDN_covid_forecast,category == "forecast"),
            aes(ds,yhat),color="salmon",size=0.1) +
  geom_ribbon(data=subset(IDN_covid_forecast,category=="forecast"),
            aes(ds,ymin=yhat_lower,ymax=yhat_upper),fill="salmon", alpha=0.3)
  ...
```



Warning: Replaced NA value(s) with missing warning values
(spike point(s))

R Console



ANALISIS SENTIMEN

F. PENGERTIAN ANALISIS SENTIMEN

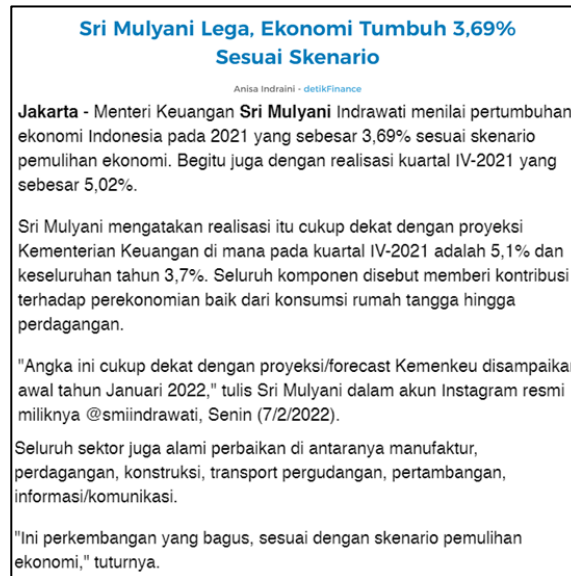
Di zaman modern ini orang-orang di seluruh dunia dapat mengekspresikan pendapat mereka secara bebas tentang berbagai topik di media sosial *online* seperti Twitter, Facebook, dan lain-lain. Pendapat-pendapat tersebut berupa data tekstual yang menyimpan '*hidden knowledge*' atau suatu pengetahuan yang tersembunyi yang sebenarnya dapat dimanfaatkan untuk satu dan lain hal. Agar data tekstual ini dapat dimanfaatkan, diperlukan suatu teknik lanjutan untuk mengolah data tekstual ini menjadi suatu informasi. Teknik ini sering disebut dengan *text mining*, yaitu teknik untuk mengekstrak informasi yang berguna dari data teks yang tidak terstruktur. Analisis sentimen merupakan salah satu turunan teknik dari *text mining*.

Secara teknis analisis sentimen atau *opinion mining*, adalah sub-bidang *Natural Language Processing* (NLP) yang mencoba mengidentifikasi dan mengekstrak opini dalam teks tertentu. Tujuan dari analisis sentimen adalah untuk mengukur sikap, sentimen, evaluasi, sikap, dan emosi seorang pembicara/penulis berdasarkan perlakuan komputasi subjektivitas dalam sebuah teks. Menurut Wahid dan Azhari (2016), analisis sentimen merupakan salah satu teknik dalam *text summarization* atau peringkasan teks, dimana dengan analisis sentimen ini kita dapat meringkas suatu teks untuk mendapatkan sentimen atau opini yang diekspresikan dalam suatu teks.

G. PENGGUNAAN ANALISIS SENTIMEN

Sesuai dengan definisi dari analisis sentimen, teknik ini dapat digunakan untuk mengetahui sentimen atau emosi yang terdapat dalam suatu data berbentuk tekstual. Sebagai contoh berikut terdapat potongan berita mengenai

pertumbuhan ekonomi Indonesia pada kuartal empat tahun 2021 dari salah satu media berita *online* di Indonesia.



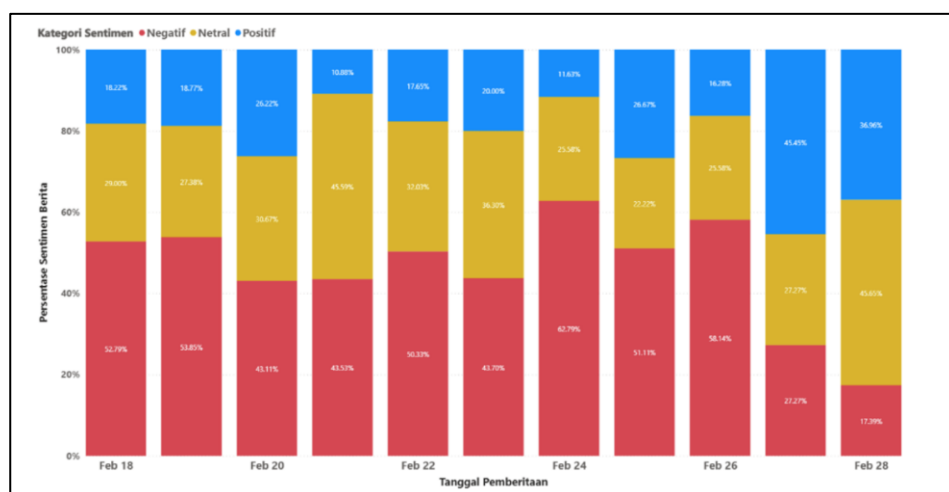
Dari potongan berita di atas dengan sekilas membaca saja kita dapat melakukan analisis sentimen secara manual dari berita tersebut. Dapat terlihat bahwa banyak kata-kata bersentimen positif yang terdapat pada potongan berita di atas, dengan demikian kita dapat menyimpulkan bahwa sentimen atau emosi yang diekspresikan dari berita mengenai pertumbuhan ekonomi Indonesia di kuartal 4 tahun 2021 adalah sentimen yang positif.

Contoh lainnya adalah data dari media sosial, dalam hal ini media sosial Twitter. Gambar berikut berisikan beberapa cuitan di media sosial Twitter yang berhubungan dengan kenaikan harga minyak goreng Maret lalu.



Apabila kita melakukan analisis sentimen secara manual untuk setiap cuitan dari sekumpulan cuitan di atas, kita dapat menyimpulkan bahwa hampir semua cuitan tersebut bersentimen atau memiliki ekspresi emosi yang negatif. Dari sini kita dapat mendapatkan suatu *insight* bahwa sentimen publik terhadap fenomena kenaikan harga minyak goreng adalah negatif. *Insight* atau temuan inilah yang nantinya dapat dimanfaatkan untuk banyak hal, misalnya dalam membantu pengambilan keputusan dan lainnya.

Dapat dilihat bahwa hanya dari beberapa potongan data tekstual saja kita dapat memperoleh suatu *insight* yang berguna, bayangkan apabila kita memiliki data tekstual dengan volume yang besar, tentu *insight* yang kita dapatkan akan lebih banyak juga. Sebagai contoh berikut adalah hasil analisis sentimen dari berita online yang berkaitan dengan peraturan Jaminan Hari Tua (JHT) di Indonesia. Pada gambar di bawah ditunjukkan bahwa kategori sentimen berita mengalami perubahan sejak pemberitaan mengenai peraturan baru JHT. Perubahan yang signifikan terjadi pada kategori sentimen negatif, sebesar 25 persen sentimen negatif pemberitaan JHT menurun pada akhir Februari 2022. Hal ini dapat disebabkan karena adanya revisi terhadap PERMENAKER No. 2/2022 terkait JHT. Perubahan kategori sentimen terkait pemberitaan JHT ini menggambarkan respons publik mulai mereda ketika Presiden Joko Widodo memerintahkan kepada KEMNAKER untuk merevisi peraturan terkait JHT.



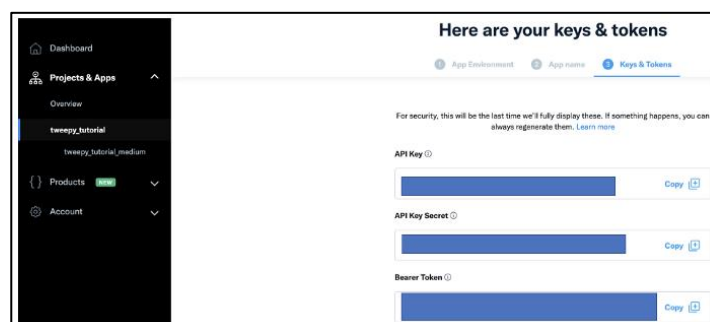
Tentu dengan data bervolume besar maka teknik analisis sentimen yang sebelumnya dapat kita lakukan secara manual (ke data berukuran kecil), harus berganti ke metode analisis sentimen khusus dengan bantuan algoritma komputasi ke data bervolume besar. Secara umum, terdapat 2 pendekatan untuk melakukan analisis sentimen, yaitu pendekatan menggunakan metode *Learning-Based* dengan teknik *machine learning* dan pendekatan menggunakan metode *Knowledge-Based*.

H. PENGUMPULAN DATA TEKS

Pengumpulan data tekstual dapat berasal dari beberapa sumber data seperti media sosial *online* dan media berita *online*. Media sosial *online* seperti Twitter dan Facebook banyak mengandung data tekstual yang di-*posting* oleh penggunanya. Pengumpulan data dari Twitter sendiri sudah difasilitasi oleh pihak pengembangnya. Kita dapat mendaftar untuk mendapat kode atau token unik yang nantinya dapat kita gunakan untuk mengumpulkan data dari Twitter. Berikut tautan yang dapat digunakan untuk mendapat token unik untuk mengumpulkan data Twitter:

<https://developer.twitter.com/en/portal/petition/essential/basic-info>

Kemudian isi *form* yang tersedia sesuai dengan tujuan dan kebutuhan kita dalam mengumpulkan data Twitter. Setelah berhasil maka kita akan mendapatkan *API keys*, *API key secret*, dan *bearer token* untuk nantinya kita pakai dalam pengumpulan data Twitter.



Setelah mendapatkan ketiga kode tersebut kemudian kita dapat memasukkan kode-kode tersebut ke sintak pengumpulan data yang banyak disediakan di internet, dalam kasus ini kita akan menggunakan Tweepy dengan bahasa pemrograman Python. Berikut adalah contoh *source code* yang dapat digunakan:

```
# Import Package

# Input API Key and Access Token
consumer_key = ""
consumer_secret = ""
access_token = ""
access_secret = ""

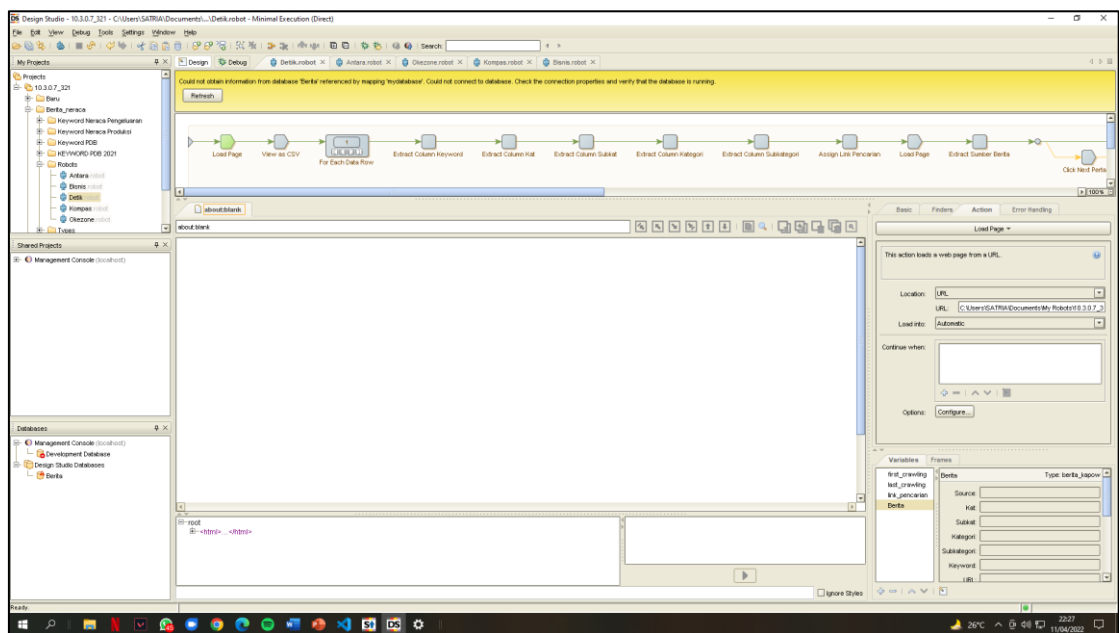
# Set authentication variable
auth = OAuthHandler (consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth)

# Set for Indonesia Area only
GEOBOX_WORLD = [-180,-90,180,90]
indo_box = [94.74,-11.11,141.04,5.93]
indonesia = indo_box

# Start streaming Twitter
class MyListener(StreamListener):
    def on_data(self, data):
        try:
            all_data = json.loads(data)
            saveFile = open('twitter.txt', 'a')
            json.dump(all_data,saveFile)
            saveFile.write('\n')
        except BaseException as e:
            print ('failed on data,',str(e))
            time.sleep(5)
    def on_error(self, status):
        print (status)

twitterStream = Stream (auth, MyListener(), verify = False)
twitterStream.filter(track=['bensin', 'pertamax', 'pertalite', 'bahan bakar'], locations=indonesia)
```

Selain dari media sosial online kita juga dapat mengumpulkan data tekstual dari media berita online seperti Detik, Antara News, Kompas, CNBC, dan banyak lainnya. Untuk pengumpulan data dari media berita online ini kita dapat menggunakan metode *web scraping*, yaitu pengumpulan data dari suatu *website* atau situs dengan memanfaatkan *tag* HTML dari situs tersebut. Cara paling mudah yang dapat digunakan untuk melakukan *web scraping* dari situs media berita online adalah dengan menggunakan aplikasi Kofax Kapow. Dalam modul ini tidak akan dijelaskan mendalam tentang bagaimana cara penggunaan aplikasi Kofax Kapow, intinya dengan aplikasi ini kita dapat membuat robot yang secara otomatis akan mengumpulkan data berita dari suatu situs media berita online yang telah ditentukan. Berikut gambar potongan dari aplikasi Kofax Kapow yang bertugas untuk mengumpulkan data berita dari situs media berita online Detik.



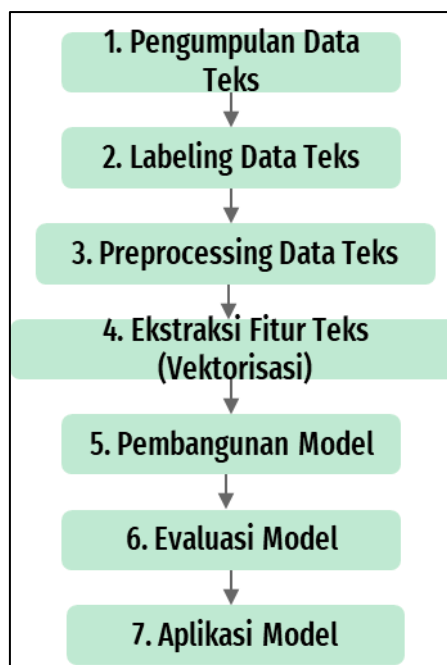
I. METODE ANALISIS SENTIMEN

Seperti yang telah disinggung pada bagian sebelumnya bahwa metode analisis sentimen secara umum dapat dilakukan menggunakan dua pendekatan yaitu dengan metode *Learning-Based* dan *Knowledge-Based*.

Metode Learning-Based

Learning-based merupakan salah satu metode klasifikasi sentiment dari hasil pembelajaran mesin (*machine learning*). Mudah-mudahan pada metode ini kita akan mengajari mesin untuk bisa membedakan mana kalimat yang bersentimen positif, negatif, atau netral. *Machine learning* sendiri telah banyak digunakan dalam kehidupan kita sehari-hari, produk rekomendasi di marketplace misalnya. Seringkali kita diberikan rekomendasi produk yang berkaitan dengan hasil pencarian produk di marketplace, ini merupakan salah satu pengaplikasian *machine learning* dimana *machine* mempelajari kebiasaan kita di marketplace.

Menurut Cao et al (2014), metode *learning-based* memerlukan dataset untuk digunakan sebagai data *training*. Oleh karena itu, dibutuhkan satu tahapan untuk mengumpulkan dan melakukan klasifikasi pada sampel dataset tersebut, selain itu proses *training* juga membutuhkan waktu. Berikut adalah tahapan atau algoritma yang paling sederhana dalam analisis sentimen menggunakan metode *learning-based*.



1. Pengumpulan Data Teks

Pengumpulan data teks dapat dikumpulkan dari beberapa sumber seperti data teks berita dan media sosial sebagaimana telah dijelaskan pada tahapan sebelumnya. Berikut adalah contoh potongan dataset yang digunakan untuk melakukan analisis sentimen.

id	date	username	tweet
1423073123960717300	2021-08-05T00:00:00.000Z	lightcanopus	@convomf Bahas perekonomian Indonesia coba
1423076063341543400	2021-08-05T00:00:00.000Z	suaradotcom	Edy meyakini pergerakan perekonomian dari desa akan mampu meningkatkan Pendapatan Asli Daerah (PAD) Kabupaten Labusel...
1423076859714015200	2021-08-05T00:00:00.000Z	bang_gar_lagi	@Kita_AMLTF Wa'alaikum salam wr wb, selamat pagi salam untuk segera sehat seluruh bangsaku agar perekonomian kembali p...
1423078264399011800	2021-08-05T00:00:00.000Z	yuuaana	Insentif ini diberikan pemerintah untuk menjaga keberlangsungan usaha sektor perdagangan eceran masa pandemi Covid-19 ...
1423078560760172500	2021-08-05T00:00:00.000Z	bisniscom	Jusindo Prediksi Pertumbuhan Ekonomi Berkisar 1,7 Hingga 2,5 Persen https://t.co/ZXkoq6dNXx
1423080164217352200	2021-08-05T00:00:00.000Z	awanakbaik	@HAECHANFTUL semangat menjunjung perekonomian
14230801455446417400	2021-08-05T00:00:00.000Z	idntimes	BPS umumkan pertumbuhan ekonomi kuartal II-2021 siang ini https://t.co/8pJ4xq5Zbu
1423082218444918800	2021-08-05T00:00:00.000Z	ndrews110611	2M Cat Pesawat terbang itu sama seperti lo sebagai kepala keluarga lagi melakukan "ikat pinggang" krna perekonomian ke...
1423082332647411700	2021-08-05T00:00:00.000Z	andriamuis	PCR Rp.900 ribu paling murah Rp. 750 ribu untuk di Palembang.. tapi toko2 masih buka.. dengan masih di bukanya toko2 t...
1423083163438309400	2021-08-05T00:00:00.000Z	meyradyana	Presiden @Jokowi menyatakan prokes merupakan kunci utama agar kesehatan dan perekonomian dpt berjalan seiringan dlm pe...
1423084126811443200	2021-08-05T00:00:00.000Z	metropublikcom	DPRD Medan Ragukan Kemampuan Pemko Tingkatkan Pertumbuhan Ekonomi https://t.co/nZ9zRiSxt4
1423084876761165800	2021-08-05T00:00:00.000Z	okezonenews	IHSG Diprediksi Menguat Jelang Data Pertumbuhan Ekonomi, Cermati Saham ANTM hingga BBRI #LengkapCepatBeritanya #Keuang...
1423085418862370800	2021-08-05T00:00:00.000Z	ndrews110611	Atau.... lo sebagai kepala keluarga lagi melakukan "ikat pinggang" krna perekonomian keluarga terkena efek C19, uang y...
1423085470552973300	2021-08-05T00:00:00.000Z	gelo_tojeng	di tengah pandemi yang memengaruhi perlambatan sektor bisnis, Pertamina tetap mampu menjadi salah satu motor penggerak...
1423086350488637400	2021-08-05T00:00:00.000Z	do_what_is_wise	Manfaat Otsus Papua sudah dirasakan oleh masyarakat, dimana masyarakat sudah dapat menumbuhkan perekonomian dalam kelu...

Sebagai contoh, semisal kita memiliki dataset berisi 5000 data hasil *web scraping* Twitter bertemakan pertumbuhan ekonomi. Kemudian akan kita ambil 1000 data dari dataset tersebut untuk digunakan sebagai data *training* dalam membuat model *machine learning*.

2. Labeling Data Teks

Labeling data dilakukan dengan memberikan label sentimen pada data teks yang telah dikumpulkan. Labeling ini dilakukan secara subjektif, semakin banyak baris yang dilabeli maka semakin banyak data *training* yang akan terbentuk. Semakin banyak data *training* yang dibentuk maka akan semakin banyak bahan yang dipelajari oleh mesin, sehingga diharapkan hasil prediksi sentimennya juga bisa lebih baik. Berikut contoh labeling sentimen dari 1000 data yang telah kita ambil sebelumnya.

id	date	username	tweet	sentiment_label
1423073123960717300	2021-08-05T00:00:00.000Z	lightcanopus	@convomf Bahas perekonomian Indonesia coba	Netral
1423076063341543400	2021-08-05T00:00:00.000Z	suaradotcom	Edy meyakini pergerakan perekonomian dari desa akan mampu meningkatkan P...	Netral
1423076859714015200	2021-08-05T00:00:00.000Z	bang_gar_lagi	@Kita_AMLTF Wa'alaikum salam wr wb, selamat pagi salam untuk segera seha...	Netral
1423078264399011800	2021-08-05T00:00:00.000Z	yuuaana	Insentif ini diberikan pemerintah untuk menjaga keberlangsungan usaha se...	Netral
1423078560760172500	2021-08-05T00:00:00.000Z	bisniscom	Jusindo Prediksi Pertumbuhan Ekonomi Berkisar 1,7 Hingga 2,5 Persen http...	Netral
1423080164217352200	2021-08-05T00:00:00.000Z	awanakbaik	@HAECHANFTUL semangat menjunjung perekonomian	Positif
1423081455446417400	2021-08-05T00:00:00.000Z	idntimes	BPS umumkan pertumbuhan ekonomi kuartal II-2021 siang ini https://t.co/B...	Netral
1423082218444918800	2021-08-05T00:00:00.000Z	ndrews110611	2M Cat Pesawat terbang itu sama seperti lo sebagai kepala keluarga lagi ...	Netral
1423082332647411700	2021-08-05T00:00:00.000Z	andriamuis	PCR Rp.900 ribu paling murah Rp. 750 ribu untuk di Palembang.. tapi toko...	Netral
1423083163438309400	2021-08-05T00:00:00.000Z	meyradiyana	Presiden @Jokowi menyatakan prokes merupakan kunci utama agar kesehatan ...	Netral
1423084126811443200	2021-08-05T00:00:00.000Z	metropublikcom	DPRD Medan Ragukan Kemampuan Pemko Tingkatkan Pertumbuhan Ekonomi https:...	Netral
1423084876761165800	2021-08-05T00:00:00.000Z	okezonenews	IHSG Diprediksi Menguat Jelang Data Pertumbuhan Ekonomi, Cermati Saham A...	Netral
1423085418862370800	2021-08-05T00:00:00.000Z	ndrews110611	Atau.... lo sebagai kepala keluarga lagi melakukan "ikat pinggang" krna ...	Netral
1423085470552973300	2021-08-05T00:00:00.000Z	gelo_tojeng	di tengah pandemi yang memengaruhi perlambatan sektor bisnis, Pertamina ...	Netral
1423086350488637400	2021-08-05T00:00:00.000Z	do_what_is_wise	Manfaat Otsus Papua sudah dirasakan oleh masyarakat, dimana masyarakat s...	Netral

3. Preprocessing Data Teks

Data teks kemudian di-*preprocessing* atau dilakukan *cleaning* (pembersihan) untuk menghilangkan elemen-elemen yang tidak dibutuhkan pada teks. Berikut tahapan pembersihan yang dilakukan:

- Contoh teks:

Selain Partai Gerindra, PDIP juga mengadakan kaderisasi menjelang pemilu di Jakarta 23/02/2019.

- *Case folding* -> menyeragamkan teks menjadi huruf kecil (bukan kapital):

selain partai gerindra, pdip juga mengadakan kaderisasi menjelang pemilu di jakarta 23/02/2019.

- Menghapus tanda baca dan angka pada teks:

selain partai gerindra pdip juga mengadakan kaderisasi menjelang pemilu di Jakarta

- Menghapus *hashtag* dan *mention* (untuk data twitter): **#Merdeka** atau **@kemarin_sore**
- Menghapus *stopwords* -> kata basic yang biasanya tidak memiliki nilai:

selain partai gerindra pdip mengadakan kaderisasi menjelang pemilu Jakarta

- *Stemming* -> Mengubah kata berimbuhan menjadi kata dasarnya:

lain partai gerindra pdip ada kader jelang pemilu jakarta

4. Ekstraksi Fitur Teks (Vektorisasi)

Salah satu masalah saat mengerjakan data berbentuk teks adalah mesin tidak dapat langsung mengolah teks mentah pada data (mesin tidak dapat membaca teks seperti yang dilakukan manusia saat membaca teks). Jadi, kita perlu melakukan ekstraksi fitur yaitu untuk mengubah teks menjadi matriks numerik (atau vektor). Kita juga perlu mengubah label sentimen menjadi bentuk numerik, misal positif menjadi angka '1', netral menjadi angka '2', dan negatif menjadi angka '3'.

tweet_clean	sentiment_label
👇	👇
kredit perban peran ekonomi nasional tumbuh persen triwulan ii	Netral
ekonomi indonesia alami tumbuh positif era pandemi covid badan pusat statistik bps tumbuh ekonomi kuartal ii year on year yoy	Netral
bangun data desakelurahan	Netral
badan pusat statistik bps sebut alas masyarakat sulawesi tenggara sultra enggan vaksin covid	Netral

Semisal kita memiliki data teks berisi 4 kalimat yang sudah di-*preprocessing* dan dilabeli sentimennya seperti pada gambar di atas. Data ini kemudian akan diubah menjadi matriks fitur numerik dengan memecah keempat kalimat tersebut menjadi kata per kata dan menjadikannya sebagai kolom matriks. Matriks fitur akan memiliki jumlah baris sesuai dengan jumlah datanya, dan jumlah kolom sesuai dengan jumlah kata dalam semua datanya. Isian baris berisi '1' bila kata itu muncul dalam data baris tersebut dan '0' bila kata itu tidak muncul. Selanjutnya label sentimen juga akan diubah menjadi bentuk numerik.

Berikut contoh matriks fitur yang terbentuk:

	alami	alas	badan	bangun	bps	covid	data	desakelurahan	ekonomi	enggan	...	sebut	statistik	sulawesi	sultra	tenggara	triwulan	tumbuh	vaksin	year	yoy	
0	0	0	0	0	0	0	0	0	1	0	_	0	0	0	0	0	1	1	0	0	0	
1	1	0	1	0	1	1	0	0	2	0	_	0	1	0	0	0	0	2	0	2	1	
2	0	0	0	1	0	0	1	1	0	0	_	0	0	0	0	0	0	0	0	0	0	
3	0	1	1	0	1	1	0	0	0	1	_	1	1	1	1	1	0	0	1	0	0	

4 rows × 34 columns

=

sentiment_label
2
2
2
2

5. Pembangunan Model

Setelah data teks diubah menjadi bentuk matriks fitur numerik barulah data tersebut dapat dikerjakan oleh mesin. Beberapa algoritma *machine learning* antara lain seperti Gaussian Naive Bayes, Bernoulli Naive Bayes, SVM, Logistic Linier, dll. Setiap algoritma di atas memiliki "perhitungan matematik"-nya masing-masing. Untuk menentukan model atau algoritma mana yang terbaik perlu dilakukan evaluasi model.

Pembangunan model dilakukan dengan membagi dataset yang telah dilabeli dan diubah menjadi matriks fitur numerik menjadi data *training* dan data *testing*, pembagiannya biasanya menjadi 7:3. Data *training* digunakan untuk melatih algoritma, sedangkan data *testing* dipakai untuk mengetahui performa atau evaluasi dari algoritma yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat sebelumnya. Sebagai contoh dataset yang telah dilabeli berjumlah 1000 baris data, kemudian kita bagi data tersebut dengan proporsi 700 baris data *training* (70%) dan 300 baris data *testing* (30%).

6. Evaluasi Model

Salah satu nilai paling sederhana yang dapat dijadikan patokan dalam evaluasi model adalah dengan menghitung tingkat akurasi dari model tersebut. Penghitungan akurasi dapat dilakukan dengan:

$$\text{Akurasi} = \frac{\text{Jumlah prediksi data testing yang benar}}{\text{Jumlah data testing yang diprediksi}}$$

Selain tingkat akurasi, evaluasi model juga dapat dilakukan dengan membandingkan tingkat presisi, tingkat *recall*, dan *confusion matrix*-nya.

7. Aplikasi Model

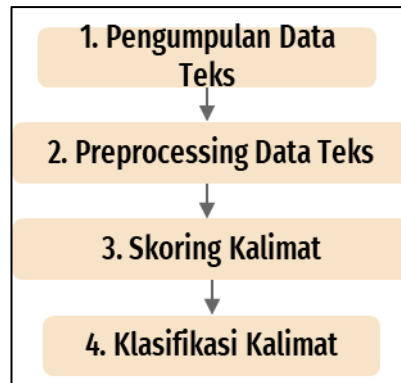
Setelah mendapatkan model prediksi sentimen kemudian model tersebut diaplikasikan ke sisa data yang belum diklasifikasikan sentimennya. Pada kasus ini maka model yang telah dibangun dari 1000 data *training* akan diaplikasikan ke sisa 4000 data lainnya.

Metode Knowledge-Based

Knowledge-based atau biasa disebut *Lexicon-based* menggunakan *dictionary* atau kamus *lexicon* untuk melakukan penilaian terhadap kata-kata pada suatu teks. Mudah-mudahan pada teknik setiap kata pada suatu kalimat akan diberikan skor positif atau negatif. Sentimen kemudian diklasifikasikan dari skor-skor tersebut. Menurut Kundi dan Asghar (2014), Klasifikasi sentimen dengan pendekatan kamus *lexicon* memiliki performa klasifikasi yang baik pada kasus lintas domain, dan isi dari kamus tersebut dapat ditambahkan kapan saja. Kekurangannya terkadang makna kata dapat berubah bergantung pada konteks dari kalimat, oleh karena itu metode ini terkadang tidak bisa menangkap makna sebenarnya dari kata yang diproses.

Banyak kamus *lexicon* yang sudah tersedia dan dapat digunakan secara gratis di internet. Beberapa contohnya seperti Bing Liu and Minqing Hu Sentiment *Lexicon*, SentiWordNet (Included in NLTK), VADER Sentiment *Lexicon*, SenticNet, SentiStrength, dll. Namun baru beberapa kamus *lexicon* saja yang mendukung

bahasa Indonesia, seperti SentiStrength_ID dan sastrawi. Pada kasus ini kita akan menggunakan kamus *lexicon* SentiStrength_ID untuk mengelompokkan sentimen teks. Berikut adalah tahapan atau algoritma yang paling sederhana dalam analisis sentimen menggunakan metode *lexicon-based*.



1. Pengumpulan Data Teks

Sama seperti metode *Learning-Based*, pengumpulan data teks pada metode *Lexicon-Based* juga dapat dikumpulkan dari beberapa sumber seperti data teks berita dan media sosial. Berikut adalah contoh potongan dataset yang digunakan untuk melakukan analisis sentimen.

id	date	username	tweet
1423073123960717300	2021-08-05T00:00:00.000Z	lightcanopus	@convomf Bahas perekonomian Indonesia coba
1423076063341543400	2021-08-05T00:00:00.000Z	suaradotcom	Edy meyakini pergerakan perekonomian dari desa akan mampu meningkatkan Pendapatan Asli Daerah (PAD) Kabupaten Labusel...
1423076859714015200	2021-08-05T00:00:00.000Z	bang_gar_lagi	@Kita_AMLTF Wa'alaikum salam wr wb, selamat pagi salam untuk segera sehat seluruh bangsaku agar perekonomian kembali p...
1423078264399011800	2021-08-05T00:00:00.000Z	yuuaana	Insentif ini diberikan pemerintah untuk menjaga keberlangsungan usaha sektor perdagangan eceran masa pandemi Covid-19 ...
1423078560760172500	2021-08-05T00:00:00.000Z	bisniscom	Jusindo Prediksi Pertumbuhan Ekonomi Berkisar 1,7 Hingga 2,5 Persen https://t.co/ZXKoq6dIXx
1423080164217352200	2021-08-05T00:00:00.000Z	awanakbaik	@HAECHANFTUL semangat menjunjung perekonomian
1423081455446417400	2021-08-05T00:00:00.000Z	idntimes	BPS umumkan pertumbuhan ekonomi kuartal II-2021 siang ini https://t.co/BpJ4xq5Zbu
1423082218444918800	2021-08-05T00:00:00.000Z	ndrews110611	2M Cat Pesawat terbang itu sama seperti lo sebagai kepala keluarga lagi melakukan "ikat pinggang" krna perekonomian ke...
1423082332647411700	2021-08-05T00:00:00.000Z	andriamuis	PCR Rp.900 ribu paling murah Rp. 750 ribu untuk di Palembang.. tapi toko2 masih buka.. dengan masih di bukanya toko2 t...
1423083163438309400	2021-08-05T00:00:00.000Z	meyradyana	Presiden @Jokowi menyatakan prokes merupakan kunci utama agar kesehatan dan perekonomian dpt berjalan seiringan dlm pe...
1423084126811443200	2021-08-05T00:00:00.000Z	metropublikcom	DPDR Medan Ragukan Kemampuan Pemko Tingkatkan Pertumbuhan Ekonomi https://t.co/nZ9zRiSXt4
1423084876761165800	2021-08-05T00:00:00.000Z	okezonenews	IHSG Diprediksi Menguat Jelang Data Pertumbuhan Ekonomi, Cermati Saham ANTM hingga BBRI #LengkapCepatBeritanya #Keuang...
1423085418862370800	2021-08-05T00:00:00.000Z	ndrews110611	Atau.... lo sebagai kepala keluarga lagi melakukan "ikat pinggang" krna perekonomian keluarga terkena efek C19, uang y...
1423085470552973300	2021-08-05T00:00:00.000Z	gelo_tojeng	di tengah pandemi yang memengaruhi perlambatan sektor bisnis, Pertamina tetap mampu menjadi salah satu motor penggerak...
1423086350488637400	2021-08-05T00:00:00.000Z	do_what_is_wise	Manfaat Otsus Papua sudah dirasakan oleh masyarakat, dimana masyarakat sudah dapat menumbuhkan perekonomian dalam kelu...

Sebagai contoh, sama seperti sebelumnya semisal kita memiliki dataset berisi 5000 data hasil *web scraping* Twitter bertemakan pertumbuhan ekonomi.

2. Preprocessing Data Teks

Berbeda dengan metode *Learning-based*, di beberapa kamus *lexicon* pada metode *Lexicon-based* kita justru tidak perlu melakukan *preprocessing* data teks seperti menghilangkan tanda baca, menyeragamkan jenis tulisan (menghilangkan kapital), menghilangkan *stopwords* atau kata-kata tidak bermakna, dan melakukan *stemming* (mengubah suatu kata menjadi kata dasarnya). Hal ini dikarenakan pada beberapa kamus *lexicon*, hal-hal tadi bisa saja menjadi pertimbangan dalam proses skoring suatu kalimat.

3. Skoring Kalimat

Proses skoring kalimat pada setiap kamus *lexicon* bisa jadi berbeda perhitungannya. Berikut adalah beberapa hal yang dipertimbangkan dalam perhitungan skor kata pada kamus *lexicon* SentiStrength_ID:

- Skor dasar kata

Skoring kata dilakukan dengan membandingkan kata dengan kamus *lexicon*. Berikut contoh penghitungan skor kata dasar suatu kalimat:

Aku tidak bahagia sebelum menang.				
0	0	4	0	4

- Kata negasi

Pendeteksian kalimat negasi dilakukan dengan melihat apakah terdapat kata negasi pada kata sebelumnya. Kata-kata yang termasuk kata negasi antara lain seperti 'belum', 'bukan', 'tak', 'tanpa', 'tidak', 'pantang', 'jangan', 'bukanlah', 'sok', dan 'tidak pernah'. Apabila terdapat kata negasi sebelum kata yang sedang dikerjakan maka skor kata tersebut menjadi kebalikannya (dikali -1).

Aku tidak bahagia sebelum menang.

0 N 4 0 4

-4

- Kata *booster*

Pendeteksian kata *booster* juga dilakukan dengan melihat apakah terdapat kata *booster* pada kata sebelum atau sesudah dari kata yang sedang dinilai. Penilaian kata *booster* antara lain:

('abis', 2),	('lebih', 1),
('agak', -1),	('mending', 1),
('amat', 2),	('mungkin', -1),
('banget', 1),	('paling', 1),
('beberapa', -1),	('pasti', 1),
('benar-benar', 1),	('sangat', 2),
('betul-betul', 1),	('seakan-akan', -1),
('hanya', -1),	('sedikit', -2),
('jauh', 2),	('sekali', 1),
('kadang-kadang', -1),	('semakin', 1),
('kali', 1),	('sempurna', 1),
('kira-kira', -1),	('sungguh', 1),
('kurang', -2),	

Kata *booster* akan menambah atau mengurangi skor dari kata yang dinilai. Contoh pada kalimat **"Aku sangat bahagia"** . Kata **'Sangat'** akan menambah skor kata 'bahagia', dari 4 menjadi 6 (ditambah 2 poin).

- Tanda baca

Pendeteksian tanda baca juga dilakukan dengan melihat apakah terdapat tanda baca tertentu di sesudah kata yang sedang dinilai. Tanda baca akan menambah atau mengurangi skor dari kata yang dinilai. Contoh pada kalimat **"Aku bahagia!"**, tanda seru setelah kata 'bahagia' akan menambah skor kata 'bahagia', dari 4 menjadi 5 (ditambah 1 poin).

- Idiom

Pendeteksian idiom juga dilakukan dengan melihat apakah kata yang sedang dinilai ketika digabung dengan kata sebelum atau sesudahnya menghasilkan suatu idiom. Beberapa penilaian idiom antara lain:

('acungi jempol', 5),	('buah bibir', -4),
('anak alay', -4),	('buah hati', 1),
('anak emas', 4),	('bunga tidur', -1),
('banting tulang', 2),	('bunuh diri', -5),
('banyak akal', 4),	('campur tangan', -4),
('batu kubur', -4),	('cari perhatian', -4),
('berbunga-bunga', 4),	('celana pendek', 1),
('berdarah indonesia', 1),	('cinta mati', 5),
('bersilat lidah', -4),	('cuci otak', -4),
('berteman baik', 1),	('darah biru', -2),
('besar kepala', -4),	('dicuci otak', -4),
('besar mulut', -4),	('diva dunia', 5),
	('empat mata', 1),

Dengan pendeteksian idiom, kata-kata yang sebelumnya tidak memiliki skor menjadi kumpulan kata yang memiliki skor. Contoh kata 'jempol' pada kamus *lexicon* akan menghasilkan skor 0, namun apabila terdapat kata 'acungi' di sebelumnya maka mendapat skor 5.

- Emoticon

Pendeteksian *emoticon* dilakukan dengan membandingkan kata yang sedang dinilai dengan kamus *emoticon* yang telah ada. Beberapa penilaian *emoticon* antara lain:

^^v	4	(o:	3
^_	4	;))	4
xp	-4):-:	-3
^^v	4)":	-5
xd	2	*_*	-2
xo	2)::	-3
d:	-4	--"	-4
o.o	-3	--	-3
o_o	-3	(^ ^)	4
x(-4	: (-5
x-(-4	.)	3
^^	4);)	4
}:)	-4	: 'Q	-3
;))	4	:#	-2
%-)	3	:o)	3
%-(-3	: (-3
(:	3	: ')	3
(-:	3	:(((-4
(o;	3	:((-4
		:))	4

Berikut contoh penilaian skor kata pada kalimat yang mengandung

emoticon di dalamnya:

Aku tidak bahagia sebelum menang :(
0	N	4	0	4	-3

4. Klasifikasi Kalimat

Setelah penilaian semua kata pada kalimat tersebut selesai kemudian akan didapat 2 nilai, yaitu skor bernilai positif terbesar dan skor bernilai negatif terkecil. Kedua nilai ini kemudian dibandingkan untuk menentukan klasifikasi sentimen dari kalimat tersebut.

- **Sentimen Positif**

Skor Positif Terbesar > Skor Negatif Terkecil

- **Sentimen Negatif**

Skor Positif Terbesar < Skor Negatif Terkecil

- **Sentimen Netral**

Skor Positif Terbesar = Skor Negatif Terkecil

Berikut beberapa contoh pengelompokan sentimen kalimat dalam beberapa kondisi:

```
1 teks = 'Aku tidak bahagia sebelum menang.'
2
3 sentimen = sentiment(teks)[0]
4 skor = sentiment(teks)[1]
5 nilai_positif = sentiment(teks)[2]
6 nilai_negatif = sentiment(teks)[3]
7
8 print(f'Hasil klasifikasi: {sentimen}')
9 print(f'Nilai positif terbesar: {nilai_positif}')
10 print(f'Nilai negatif terkecil: {nilai_negatif}')
11 skor
✓ 0.2s
Hasil klasifikasi: Netral
Nilai positif terbesar: 4
Nilai negatif terkecil: -4
['aku tidak bahagia [-4] sebelum menang [4]']
```

Kalimat dengan Negasi

```
1 teks = 'Kamu baik tapi suka bersilat lidah'
2
3 sentimen = sentiment(teks)[0]
4 skor = sentiment(teks)[1]
5 nilai_positif = sentiment(teks)[2]
6 nilai_negatif = sentiment(teks)[3]
7
8 print(f'Hasil klasifikasi: {sentimen}')
9 print(f'Nilai positif terbesar: {nilai_positif}')
10 print(f'Nilai negatif terkecil: {nilai_negatif}')
11 skor
✓ 0.1s
Hasil klasifikasi: Netral
Nilai positif terbesar: 4
Nilai negatif terkecil: -4
['kamu baik [4] tapi suka [4] bersilat lidah [-4]']
```

Kalimat dengan Idiom

```
1 teks = 'Aku sangat mencintai dirimu'
2
3 sentimen = sentiment(teks)[0]
4 skor = sentiment(teks)[1]
5 nilai_positif = sentiment(teks)[2]
6 nilai_negatif = sentiment(teks)[3]
7
8 print(f'Hasil klasifikasi: {sentimen}')
9 print(f'Nilai positif terbesar: {nilai_positif}')
10 print(f'Nilai negatif terkecil: {nilai_negatif}')
11 skor
✓ 0.1s
Hasil klasifikasi: Positif
Nilai positif terbesar: 7
Nilai negatif terkecil: -1
['aku sangat mencintai [7] dirimu']
```

Kalimat dengan Booster