

# ANALISIS OPINI PUBLIK

Reshuffle Kabinet Indonesia Maju

Dhea Fajriati Anas  
Sanber Code Batch 20



# OUTLINES

1

PENDAHULUAN

2

METODE PENGOLAHAN DATA

3

HASIL DAN PEMBAHASAN

4

KESIMPULAN DAN SARAN





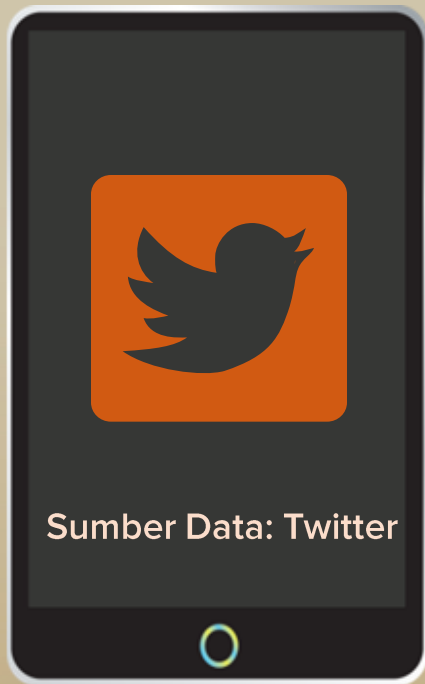
# PENDAHULUAN

## **Jokowi Umumkan Reshuffle 6 Menteri Baru di Tengah Pandemi COVID-19**

Pada tanggal 22 Desember 2020, Presiden Jokowi mengumumkan enam orang nama sebagai menteri baru di Kabinet Indonesia Maju, dan melakukan pelantikan pada keesokan harinya. Hal tersebut menimbulkan berbagai respon, baik pro dan kontra dari berbagai pihak yang mempertanyakan apakah perombakan di tengah pandemi COVID-19 adalah langkah yang tepat?. Studi kasus kali ini akan melakukan analisis sentimen dari topik tersebut dengan mengambil data dari social media, yaitu twitter.

# METODE PENGOLAHAN DATA

## TEXT MINING



### **Waktu Pengambilan Data**

22 – 29 Desember  
2020



### **Kata Kunci**

#ResuffleKabinet,  
#ReshuffleKabine,  
resuffle kabinet



### **Total Raw Data**

1818 baris  
11 kolom



### **Modul**

Tweepy :  
Consumer key  
Consumer secret  
Access token  
Access token secret

### **Format Data**

- |                   |                  |
|-------------------|------------------|
| 1. userid         | 7. follower      |
| 2. name           | 8. following     |
| 3. screenname     | 9. tweetid       |
| 4. verified       | 10. tweet        |
| 5. location       | 11. tweetcreated |
| 6. accountcreated |                  |

# METODE PENGOLAHAN DATA : PRE- PROCESSING TEXT & FEATURE SELECTION



## CASE FOLDING

Modul : re (regular expression)

Studi kasus ini melakukan pengolahan data berupa :

- Lowercase
- Menghapus \n
- Menghapus mention
- Menghapus hashtag
- Menghapus RT
- Menghapus hyperlink
- Menghapus angka
- Menghapus tanda baca
- Menghapus single character

## STOPWORDS

Modul :

- nltk.corpus
- nltk.tokenize
- itertools

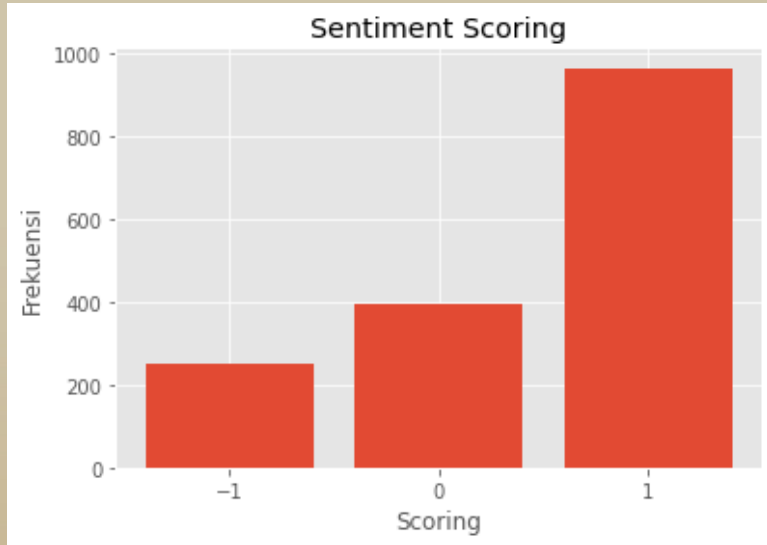
Selain menggunakan corpus bahasa indonesia, penulis juga menambahkan kata berupa :

- |       |       |
|-------|-------|
| - yg  | - dr  |
| - nya | - ya  |
| - dg  | - yaa |
| - aja | - utk |

Kemudian, dilakukan drop\_duplicate untuk menghapus data tweet yang sama

# PROSEDUR PENGOLAHAN DATA

## ANALISIS SENTIMEN



**Positif** : 965  
**Netral** : 397  
**Negatif** : 250

### Data Preparation

**Analisis sentimen** : menganalisa dan menilai suatu sentimen dalam sebuah kalimat. Studi kasus ini melakukan analisis sentimen kategori *Fined-grained Sentiment Analysis*.



#### Sentiment Scoring

<https://github.com/riochr17/Analisis-Sentimen-ID>, dan ditambahkan berdasarkan *workcloud* di halaman berikutnya. Positif (1), Netral (0), dan Negatif(-1).

Pelabelan dilakukan dengan melakukan transformasi data tweet dengan syarat sebagai berikut :

Total Nilai > 0, maka sentimen positif

Total Nilai < 0, maka sentimen negatif

Total Nilai = 0, maka sentimen netral



#### Hasil data tweet yang telah text pre-processing

Jumlah data : 1612

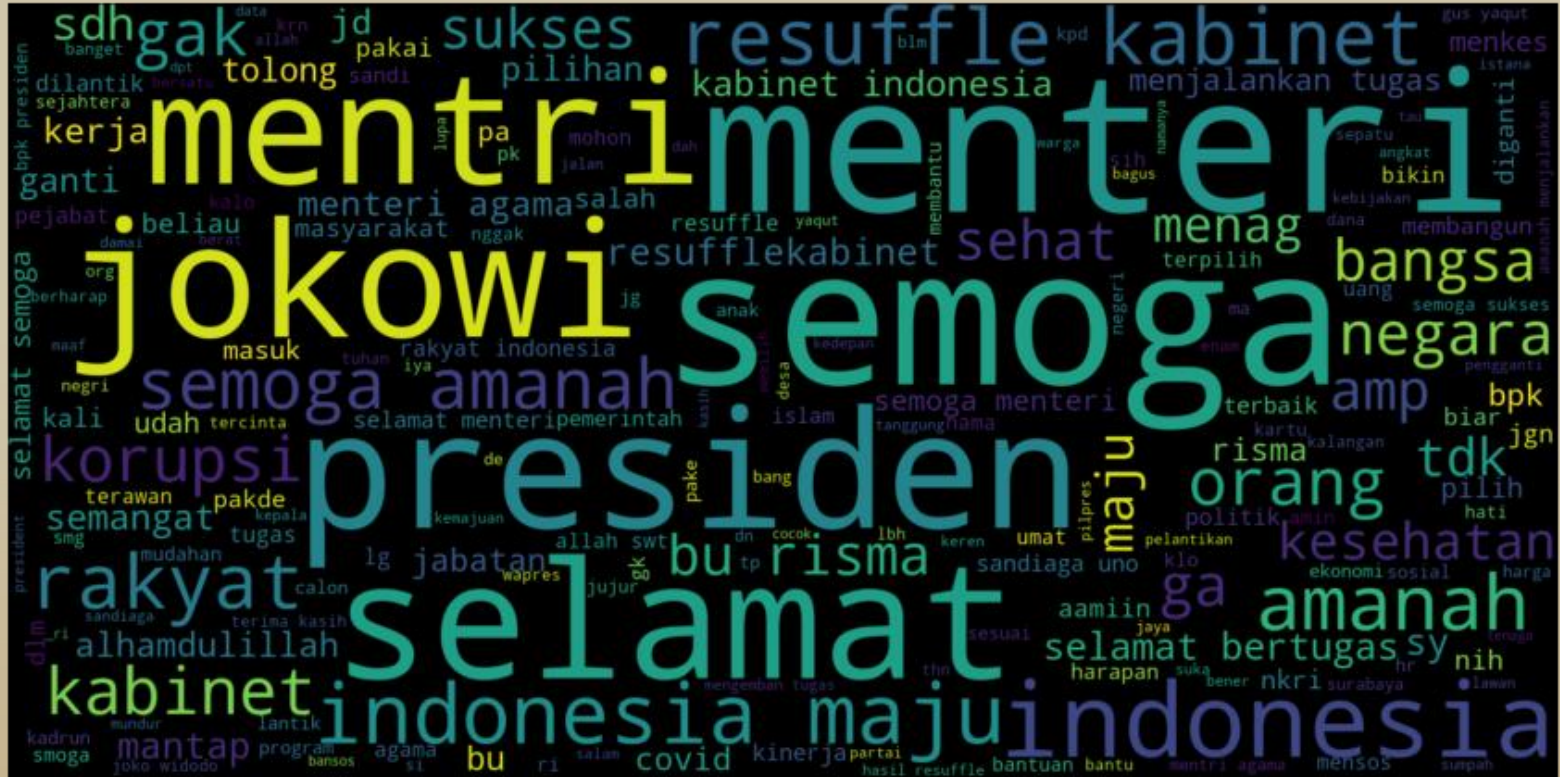


#### Jumlah Kata Sentimen

Positif : 1304, Negatif : 3523



# WORKCLOUD





# PROSEDUR PENGOLAHAN DATA

## ANALISIS SENTIMEN

### Training

Komposisi dataset yang digunakan sebesar 80% data train, dan 20% data test.

#### Pembobotan

Term Frequency-Inverse Document  
Frequency (TF-IDF)



#### Model

Gaussian, Bernoulli, Complement dan Multinomial  
Naive Bayes

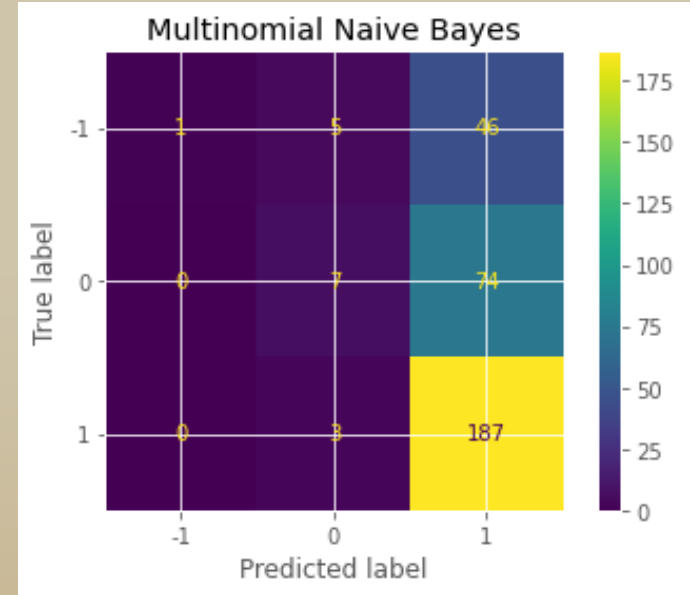




# HASIL DAN PEMBAHASAN EVALUASI

Multinomial Naive Bayes				
	Precision	Recall	F1-Score	Accuracy
Negatif	0.02	1.00	0.04	0.60
Netral	0.09	0.47	0.15	
Positif	0.98	0.61	0.75	

Classification Report



Confusion Matrix

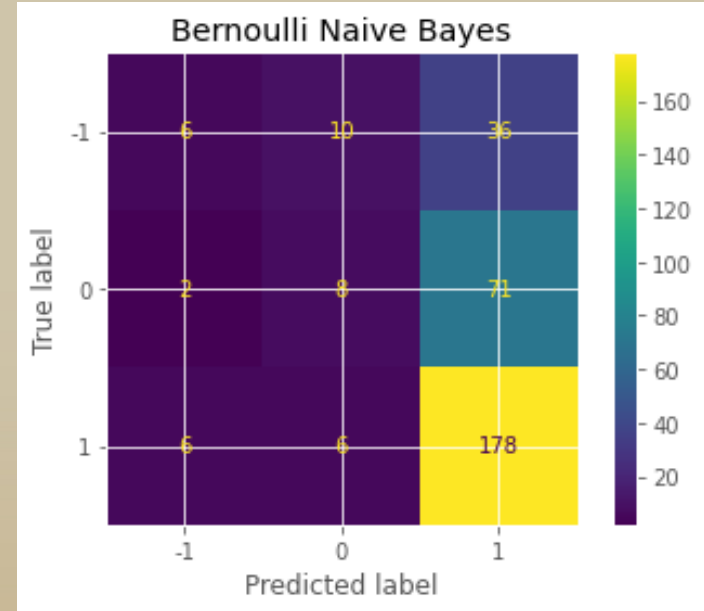


# HASIL DAN PEMBAHASAN

## EVALUASI

Bernoulli Naive Bayes				
	Precision	Recall	F1-Score	Accuracy
Negatif	0.12	0.43	0.18	0.59
Netral	0.10	0.33	0.15	
Positif	0.94	0.62	0.75	

### Classification Report



### Confusion Matrix

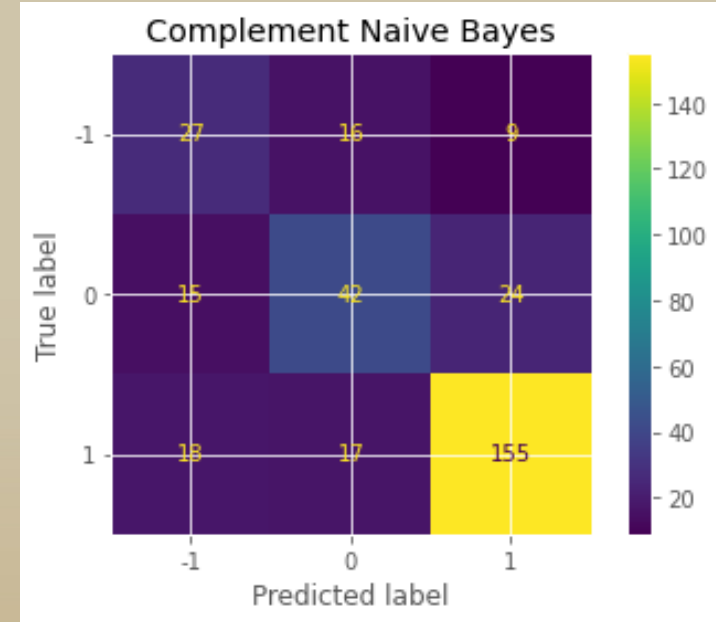


# HASIL DAN PEMBAHASAN

## EVALUASI

Complement Naive Bayes				
	Precision	Recall	F1-Score	Accuracy
Negatif	0.52	0.45	0.48	0.69
Netral	0.52	0.56	0.54	
Positif	0.82	0.82	0.82	

### Classification Report



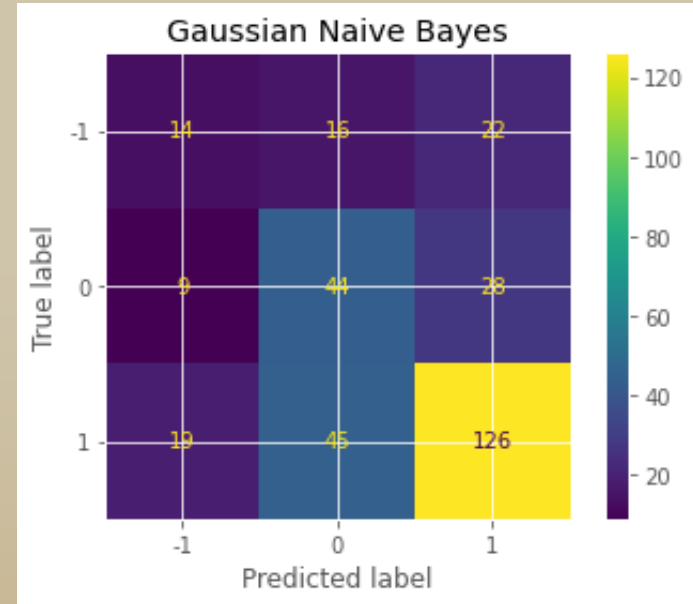
### Confusion Matrix



# HASIL DAN PEMBAHASAN EVALUASI

Gaussian Naive Bayes				
	Precision	Recall	F1-Score	Accuracy
Negatif	0.27	0.33	0.30	0.57
Netral	0.54	0.43	0.47	
Positif	0.66	0.72	0.69	

**Classification Report**



**Confusion Matrix**



# Pembahasan



## Classification Report

Dari keempat model Naive Bayes, akurasi tertinggi dimiliki oleh Complement Naive Bayes (69%), dan terendah dimiliki oleh Bernoulli Naive Bayes (59%). Nilai akurasi didapatkan dari rasio prediksi benar dengan keseluruhan data. Hal tersebut menyebabkan nilai akurasi tidak bisa menjadi satu-satunya penilaian untuk mengukur kinerja sistem untuk data yang tidak seimbang (imbalance).

Metric lainnya yang bisa digunakan apakah suatu model sudah bekerja dengan baik adalah recall, dan precision. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Kenaikan nilai recall akan menyebabkan penurunan nilai precision, begitupun sebaliknya. Untuk mendapatkan nilai precision dan recall yang seimbang dapat dilakukan dengan mencari nilai F1 Score. F1 Score adalah rata-rata precision dan recall dengan menggunakan harmonic mean. F1 Score yang memiliki nilai paling seimbang dari keempat model adalah Complement Naive Bayes (Negatif= 48, Netral=54, Positif=82).

Nilai recall dan precision kategori negatif lebih rendah daripada kategori positif diduga akibat adanya imbalance dataset, dimana dataset kategori positif lebih banyak daripada kategori lainnya. Nilai standar deviasi dari hasil sentiment scoring adalah 2.86, dan mean adalah 1.62. Hal tersebut berarti nilai mean merupakan representasi yang buruk dari keseluruhan data.

# KENAPA COMPLEMENT NAIVE BAYES MEMILIKI CLASSIFICATION REPORT YANG PALING BAIK ?



## KELEBIHAN

- Performa baik untuk imbalanced dataset dibandingkan naive bayes lainnya

Regular Naive Bayes

$$\operatorname{argmax} p(y) \cdot \prod p(w|y)^{f_i}$$

Complement Naive Bayes

$$\operatorname{argmin} p(y) \cdot \prod \frac{1}{p(w|\hat{y})^{f_i}}$$

Pada regular naive bayes kelas dengan nilai tertinggi adalah kelas prediksi, sedangkan CNB kebalikannya.

# KESIMPULAN DAN SARAN

**Model Terbaik**  
Nilai accuracy, precision, dan recall terbaik dimiliki oleh Complement Naive Bayes.

**Data TXT**  
Data txt positif dan negatif disesuaikan dengan topik sehingga hasilnya lebih relevan.

**HASIL SENTIMEN 1**

**MODEL TERBAIK 2**

**WORKCLOUD 3**

**DATASET 4**

## Hasil Sentimen

Berdasarkan data yang diambil dari twiiter, topik reshuffle kabinet yang dilakukan di tengah pandemi covid-19 mendapatkan **respon positif lebih tinggi**.

## Workcloud

Kata positif didominasi oleh kata 'selamat', dan 'semoga'.







Thank you