

MBKM Riset

# ANALISIS SENTIMEN PADA SOSIAL MEDIA TWITTER TERHADAP PEMILU 2024 DENGAN METODE NAIVE BAYES CLASSIFIERS

I Wayan suardi  
NIm:20210009





## Latar belakang

1. Banyaknya opini dari masyarakat melalui media sosial twitter tentang pemilu 2024
2. Setiap opini dari masyarakat memiliki sentimen positif, negatif dan netral.

## Rumusan masalah



Bagaimana pengklasifikasian opini masyarakat sehingga didapatkan sentimen kelas positif dan negatif dengan menggunakan metode Naive bayes classifier

## METODE PENELITIAN

### Crawling Data

Proses pengambilan data menggunakan library Tweepy. Library ini memanfaatkan API Twitter berupa search yang bisa mencari tweet yang memiliki kecocokan dengan kata kunci yang diberikan. Peneliti memberikan kata kunci berupa pemilu 2024. Data tweet yang terkumpul disimpan dalam format CSV.

### Preprocessing Data

Data yang terkumpul masih berbentuk data yang belum terstruktur dengan isi dari setiap tweet masih dalam bahasa yang tidak baku. Tahap ini bertujuan untuk menghilangkan karakter yang tidak relevan dan mengurangi kualitas model selain itu juga dapat meningkatkan kualitas data. Preprocessing data pada penelitian ini mengalami 5 tahapan yaitu:

- Pembersihan Data
- Tokenizing
- Normalisasi
- Penghapusan Stopword
- Stemming Data



## METODE PENELITIAN

### Naïve Bayes Classifier

Metode Naïve Bayes Classifier (NBC), merupakan suatu metode klasifikasi sederhana berdasarkan aplikasi teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen). Teorema bayes berfungsi untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode pengklasifikasian tersebut dikemukakan oleh ilmuwan Inggris Thomas Bayes.

persamaan klasifikasi data dengan Naïve Bayes:

$$P(C|X) = \frac{p(X|C)p(C)}{P(X)}$$

Dimana,

X = peubah acak

C = kelas

$P(C | X)$  = probabilitas C didasarkan pada kondisi X

$P(X | C)$  = probabilitas X didasarkan pada kondisi C



## hasil tahapan preprocessing

### Mengumpulan Data Tweet

Proses Crawling Data menggunakan API Twitter dilakukan tanggal 6 maret 2023 dengan kata kunci „PEMILU 2024“ Dari proses crawling data twitter didapatkan data sebanyak 11157 data tweet yang belum diketahui label sentimennya.

### Preprocessing Data

Setelah proses pengumpulan data, kemudian data diolah dengan tahapan text preprocessing. Berikut merupakan tahapan text preprocessing yang dilakukan dengan mengambil satu sampel tweet.

Tahapan	Input	Output
cleaning data	@dara_darmayu @HelmiFelis_ Dirumah juga sdh di datangi Pantarlih, catat dan tempel stiker pemilu masa mau di tunda??	dirumah juga sdh di datangi pantarlih catat dan tempel stiker pemilu masa mau di tunda
Tokenizing	dirumah juga sdh di datangi pantarlih catat dan tempel stiker pemilu masa mau di tunda	dirumah, juga, sdh, di, datangi, pantarlih, catat, dan, tempel, stiker, pemilu, masa, mau, di, tunda
Normalisasi	dirumah, juga, sdh, di, datangi, pantarlih, catat, dan, tempel, stiker, pemilu, masa, mau, di, tunda	dirumah, juga, sudah, di, datangi, pantarlih, catat, dan, tempel, stiker, pemilu, masa, mau, di, tunda
remove stopwords	dirumah, juga, sudah, di, datangi, pantarlih, catat, dan, tempel, stiker, pemilu, masa, mau, di, tunda	dirumah, datangi, pantarlih, catat, tempel, stiker, pemilu, tunda
stemming	dirumah, datangi, pantarlih, catat, tempel, stiker, pemilu, tunda	rumah, datang, pantarlih, catat, tempel, stiker, milu, tunda

## Tahap Pelebelan data

Data yang sudah melalui tahapan text preprocessing dan dikelompokkan per kata kemudian dicocokkan dengan kamus data untuk mengetahui polaritas text apakah positif, negatif, atau netral. Kamus data yang digunakan, sebagai berikut:

Positif		Negatif	
acungan jempol	Empati	Absurd	Dicuri
adaptif	ekonomis	Acuh	Dosa
adil	Enak	Alergi	Egois
ahli	Gagah	Ambigu	Gagap
akurat	Gembira	Aneh	Gatal
asyik	Gesit	Angkuh	Gila

Dari hasil training set tersebut menghasilkan polaritas text. Sebuah text/tweet dianggap positif atau negatif jika kata tersebut mempunyai keterkaitan dengan kamus kata, dan dianggap netral jika text/tweet tidak terdapat kata yang terkait dalam kamus kata.

	tweet	positive_counts	negative_counts	label
0	meta tps khusus milu lapas narkotika karang in...	1	0	positive
1	rumah datang pantarlih catat tempel stiker mil...	0	0	netral
2	rangka wujud prabowo presiden menang gerindra ...	1	0	positive
3	kader simpatisan gerindra gerak songsong milu ...	1	0	positive
4	dasco rakorda gerindra sultra instruksi kader ...	1	0	positive
5	ketua pn amk tolak tunda milu alas	0	0	netral
6	progres giat coklit milu pantarlih desa sanan ...	1	0	positive
7	lupa kader grndr insyaallah satu tenggelam grn...	0	2	negative
8	milu jalan sesuai jadwal	1	0	positive
9	putus adil negeri pn jakarta pusat jakpus peri...	1	1	netral

# Tingkat Akurasi Pada Klasifikasi Naive Bayes

Selanjutnya dilakukan pengklasifikasian dengan algoritma naive bayes untuk mengetahui probabilitas sentimen. Hasil pengklasifikasian dapat dilihat pada gambar dibawah:

accuracy_score : 0.6418289585097375				
	precision	recall	f1-score	support
<b>negative</b>	<b>0.54</b>	<b>0.85</b>	<b>0.66</b>	<b>305</b>
<b>neutral</b>	<b>0.77</b>	<b>0.41</b>	<b>0.54</b>	<b>498</b>
<b>positive</b>	<b>0.67</b>	<b>0.78</b>	<b>0.72</b>	<b>378</b>
<b>accuracy</b>			<b>0.64</b>	<b>1181</b>
<b>macro avg</b>	<b>0.66</b>	<b>0.68</b>	<b>0.64</b>	<b>1181</b>
<b>weighted avg</b>	<b>0.68</b>	<b>0.64</b>	<b>0.63</b>	<b>1181</b>

hasil klasifikasi didapatkan 305 data dengan sentimen negatif dan 378 data sentimen positif. mendapatkan precision sebesar 0.54 pada data sentimen negatif dan 0.67 pada sentimen positif, recall sebesar 0.85 pada sentimen negatif dan 0.78 pada sentimen positif, dan f1-score sebesar 0.66 pada data sentimen negatif dan 0.72 pada sentimen positif. Akurasi yang didapatkan pada klasifikasi ini yaitu 0.64 atau 64% .

“sekian presentasi dari saya maaf  
jika ada salah kata atau salah  
menaruh rasa seperti hubungan  
saya dengan dia yang selesai tanpa  
pernah dimulai”

I Wayan Suardi

