

Analisis Sentimen Masyarakat Indonesia Terhadap Kebijakan Kominfo Tentang Penyelenggara Sistem Elektronik (PSE) Menggunakan Metode *Naïve Bayes*

Noni Meilawati^{a,1,*}, Sri Winiarti^{a,2}

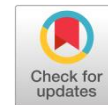
^a Program Studi Informatika Universitas Ahmad Dahlan, Jl. Ringroad Selatan, Kragilan, Tamanan, Bantul, Yogyakarta, dan 55191, Indonesia

¹ noni1600018154@webmail.uad.ac.id; ² sri.winiarti@tif.uad.ac.id;

* Penulis Korespondensi

ABSTRAK

Kementerian Komunikasi dan Informatika (Kominfo) membuat aturan baru, yaitu untuk setiap platform, aplikasi, dan web yang termasuk ke dalam Penyelenggara Sistem Elektronik (PSE) wajib untuk mendaftarkan diri kepada Kominfo menghindari pemblokiran. Pada tanggal 30 juli 2022 memblokir beberapa situs yaitu origin, paypal, steam dan lain-lain menjadi perbicangan di media sosial twitter menggunakan #BlokirKominfo untuk mengungkapkan berbagai opini dan keluh kesahnya. Pengumpulan data *tweet* tanggal 16 juli 2022- 06 Agustus 2022, *data tweet* diambil sebanyak 1500 data dengan *tweet* berbahasa Indonesia. Penelitian ini menggunakan metode klasifikasi *Naïve Bayes*. *Naïve Bayes* Merupakan salah satu algoritma *machine learning* yang penggunaannya mudah serta pemrosesannya memiliki waktu yang cepat dan mudah diimplementasikan. Berdasarkan hasil penelitian yang telah dilakukan menggunakan *data tweet* PSE menunjukkan bahwa sentimen dari *data tweet* tersebut cenderung negatif, *tweet* yang mengandung sentimen negatif ada sebanyak 606 data, *tweet* sentimen positif ada 378 data dan sentimen netral sebanyak 213 data. Proses klasifikasi pada penelitian ini menggunakan 20% data testing dan 80% data training. Pengujian menggunakan *confusion matrix* akurasi sebesar 63%, precision 80%, recall 46% dan F-score 44%.



Kata Kunci
Machine Learning
Analisis Sentimen
Naïve Bayes
Twitter



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Pendahuluan

Kementrian Komunikasi dan Informatika Republik Indonesia (Kominfo) adalah kementrian di Republik Indonesia yang bertanggung jawab urusan ruang lingkup informasi dan komunikasi. Kominfo mempunyai tugas menyelenggarakan urusan pemerintahan dan membantu Presiden dalam menyelenggarakan pemerintahan Negara (Kominfo 2022). Batas waktu terakhir untuk pendaftaran platform, aplikasi, dan web kepada PSE jatuh pada tanggal 20 Juli 2022. Pihak manapun yang tidak mendaftarkan diri hingga batas waktu tersebut menghadapi beberapa tahapan seperti diberi teguran terlebih dahulu lalu diberi sanksi denda hingga akhirnya akan dilakukan pemblokiran terhadap pihak yang bersangkutan. Tentu saja dengan aturan tersebut banyak menuai pro dan kontra, salah satunya adalah banyak masyarakat menganggap PSE dapat mengancam privasi.

Karena pasal-pasal dalam PSE mengandung ancaman hak atas privasi kebebasan berekspresi dalam platform digital [2]. Melalui Siaran Pers No. 308/HM//07/2022, Kominfo memberikan informasi beberapa aplikasi yang akan diblokir yaitu aplikasi Battle Net, Dota, Paypal, *Steam*, *Epic Games*, dan *Origin* [3]. Sampai pada tanggal 30 Juli 2022 pukul 00.00 WIB, aplikasi tersebut sudah tidak bisa di akses. Hal tersebut menimbulkan berbagai macam reaksi dalam media sosial khususnya pada media sosial *twitter*. Pengguna Twitter membuat tagar #BlokirKominfo sebagai bentuk protes, tagar #BlokirKominfo sendiri menduduki posisi teratas di Indonesia dengan jumlah *tweet* mencapai 56,8 ribu *tweet*. Hal ini tentu mengungkapkan perasaan mereka yang kesal dengan tingkah kominfo atau pemerintah [4].

Akibat pemblokiran aplikasi tersebut banyak pengguna aplikasi yang dirugikan, seperti pengguna aplikasi *Steam*, *Epic Games*, dan *Origin* pemblokiran ini mengakibatkan *gamer* tidak bisa mengakses situs serta *platform* aplikasi tersebut. Sanksi tersebut juga menyebabkan *gamer* tidak dapat menemukan informasi, mengunduh, serta melakukan pembelian *game* baru melalui URL steampowered.com (*Steam*) dan epicgames.com (*Epic Games*). Seperti halnya pengguna aplikasi Paypal juga tidak bisa menarik uang yang ada di aplikasi tersebut, Paypal dibutuhkan para pengguna untuk bertransaksi, seperti menerima upah, belanja, karitatif atau amal [5]. *Twitter* merupakan salah satu media dalam berjejaring sosial dimana penggunanya dapat mengirim hingga 110 karakter yang disebut *tweet* atau kicauan [6]. Melalui postingan pada *Twitter*, masyarakat dapat membagikan dan mendapatkan informasi mengenai hal apapun.

Dengan demikian, penting untuk menganalisis *tweet* komentar publik tentang kebijakan saat ini sehingga pemerintah dapat mengetahui reaksi publik terhadap kebijakan tersebut. Umpan balik yang diberikan bisa dalam bentuk apapun bisa positif, netral atau negatif. Sehingga salah satu cara yang dapat digunakan adalah dengan analisis sentimen. Analisis sentimen merupakan salah satu cara untuk mengumpulkan pendapat orang banyak terhadap sesuatu seperti layanan publik, isu, kinerja pemerintahan atau hal lain yang berkaitan. Analisis sentimen dapat digunakan sebagai salah satu cara untuk melakukan evaluasi terhadap layanan yang telah diberikan [6]. Hasil tersebut dapat dijadikan sebagai evaluasi dan acuan dalam penyusunan kebijakan. Selanjutnya, dapat dilakukan peninjauan kembali kebijakan yang ada ataupun mengatasi masalah dalam masyarakat sebagai penerima informasi dan pihak yang melaksanakan kebijakan [7]. Penelitian sebelumnya yang dilakukan oleh Pintoko dkk [8], mendapatkan hasil akurasi yang cukup tinggi yaitu sebesar 86,80%.

Naive Bayes merupakan salah satu metode dalam *machine learning* yang penggunaannya mudah serta pemrosesannya memiliki waktu yang cepat, mudah diimplementasikan dengan struktur yang cukup sederhana dan memiliki tingkat efektifitas yg tinggi [9]. Klasifikasi *Naive Bayes* adalah salah satu pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi ini didasarkan pada teorema *Naive Bayes* yang memiliki kemampuan klasifikasi serupa dengan *Decision Tree* dan *Neural Network* yang terbukti memiliki akurasi dan kecepatan yang tinggi saat diimplementasikan kedalam database dengan dataset yang besar[10].

Dari latar belakang penelitian yang sudah dilakukan, metode *Naive Bayes* mempunyai nilai akurasi dan performansi yang tinggi untuk mengklasifikasi sebuah teks[11]. Sehingga dari permasalahan yang ada akan diteliti tentang opini masyarakat terhadap kebijakan Kominfo di media social twitter dengan mengelompokkan komentar menjadi tiga kelas yakni positif, negatif dan netral.

2. Metode

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu cara untuk mengumpulkan pendapat orang banyak terhadap sesuatu seperti layanan publik, isu, kinerja pemerintahan atau hal lain yang berkaitan. Analisis sentimen dapat digunakan sebagai salah satu cara untuk melakukan evaluasi terhadap layanan yang telah diberikan.

2.2. Crawling Data

Crawling data di *Twitter* adalah suatu proses untuk mengambil atau mengunduh data dari *server Twitter* dengan bantuan *Application Programming Interface (API) Twitter* baik berupa data pengguna maupun *data tweet*. *API Twitter* berfungsi sebagai penghubung antara sistem yang dibangun dengan *Twitter*. *API Twitter* membutuhkan *consumer key*, *consumer access*, *access token* dan *access secret token* yang didapatkan dengan cara mendaftarkan aplikasi *API Twitter* di <http://dev.twitter.com>. Sebelum mendaftar kita harus mempunyai akun *Twitter* terlebih dahulu kemudian melakukan registrasi dan verifikasi data [12]

2.3. Preprocessing

Tahap *preprocessing* adalah tahap yang dipakai untuk menghapus kata atau teks yang tidak dibutuhkan dalam proses klasifikasi, tahapan ini dilakukan agar data menjadi terstruktur, menyeragamkan dan memudahkan dalam pembacaan serta proses klasifikasi [13]

- *Cleaning* adalah menghilangkan karakter pada teks berupa hastag, url, mention dan symbol[14]

- Case Folding, tahap case folding merupakan tahap mengubah semua huruf yang terdapat pada dokumen dari huruf kapital menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima[13]
- Tokenizing, tahap tokenizing atau parsing merupakan tahap pemotongan string input berdasarkan tiap kata yang menyusunnya[13]. Selain itu, spasi digunakan untuk memisahkan antar kata tersebut
- Filtering, tahap filtering merupakan tahap mengambil kata – kata penting dari hasil tokenizing. Proses filtering dapat menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist atau stopword merupakan kata – kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopword adalah “yang”, “dan”, “di”, “dari” dan lain-lain
- Stemming, tahap stemming merupakan proses mengubah kata berimbuhan menjadi kata dasar. Aturan – aturan bahasa diterapkan untuk menanggalkan imbuhan – imbuhan pada kata.

2.4. Naïve Bayes

Naïve Bayes atau *Naïve Bayes Classifier* (NBC) merupakan metode klasifikasi yang berdasar pada teorema *Bayes*. Metode klasifikasi ini cocok digunakan ketika jumlah masukan yang sangat besar. Klasifikasi ini lebih disukai karena kecepatan dan kesederhanaannya. Klasifikasi ini sangat memperhatikan tingginya akurasi serta kecepatan dalam memproses suatu data dalam jumlah yang besar[15]. Teorema *Bayes* yang menjadi dasar dari metode *Naïve Bayes*. Pada teorema *Bayes*, bila terdapat dua kejadian yang terpisah (misalkan A dan B), maka teorema *Bayes* dirumuskan sebagai berikut [14] :

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(A)} \quad (1)$$

Naïve Bayes bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan Metode *Bayes* dalam algoritma *Naïve Bayes classifier* setiap dokumen direpresentasikan dengan pasangan atribut, ”a1, a2, ... , an” dimana a1 adalah kata pertama, a2 adalah kata kedua, dan seterusnya. Sedangkan V adalah himpunan kategori *tweet*. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (*VMAP*) pada persamaan (2) sebagai berikut :

$$V_{MAP} = \arg \max P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (3)$$

Berdasarkan persamaan (3) maka persamaan (4) dapat ditulis

$$V_{MAP} = \arg \max_{v_j \in V} \frac{P(a_1, a_2, a_3, \dots, a_n) | P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (4)$$

Karena nilai $P(a_1, a_2, \dots, a_n)$ untuk semua v_j besarnya sama maka nilainya dapat diabaikan sehingga persamaan diatas menjadi persamaan 5 Dapat diturunkan menjadi:

$$V_{MAP} = \arg \max_{v_j \in V} P(a_1, a_2, a_3, \dots, a_n) | P(v_j) \quad (5)$$

Dengan mengasumsikan bahwa setiap kata dalam (a1,a2,a3,..an) adalah independen, maka $P(a_1, a_2, \dots, a_n | v_j)$ dalam persamaan diatas dapat ditulis sebagai persamaan 6

$$P(a_1, a_2, \dots, a_n | v_j) = \prod P(a_i | v_j) \quad (6)$$

Sehingga persamaan diatas dapat ditulis menjadi persamaan

$$V_{MAP} = \arg \max_{v_j \in V} P(v_j) \times \prod P(a_i | v_j) \quad (7)$$

Nilai $P(V_j)$ ditentukan pada saat pelatihan, yang nilainya didekati dengan persamaan

$$P(v_j) = \frac{|Dok_i|}{|training|} \tag{8}$$

Kemudian untuk nilai $P(a_i|V_j)$ yaitu probabilitas kata a_i dalam kategori j ditentukan dengan persamaan

$$P(a_i|v_j) = \frac{n_i+1}{n+(kosakata)} \tag{9}$$

Keterangan :

$P(v_j)$: Probabilitas setiap dokumen pada sekumpulan dokumen

$P(a_i|v_j)$: Probabilitas kemunculan kata a_i pada suatu dokumen dengan kelas v_j

$|Dok_i|$: Frekuensi dokumen pada setiap kategori

$|traininnng|$: jumlah dokumen training yang ada

n_i : frekuensi kata ke-k pada setiap kategori

$|kosakata|$: jumlah kosakata yang ada pada dokumen

2.5. Pengujian Confusion Matrix

Pengukuran kinerja sebuah sistem sangat penting karena menggambarkan seberapa baik algoritma yang diterapkan dalam suatu sistem. *Confusion Matrix* adalah metode yang digunakan dalam mengukur kinerja dari suatu algoritma klasifikasi yang berbentuk tabel dengan 4 nilai yang merepresentasikan hasil dari klasifikasi. Nilai – nilai tersebut dapat dilihat pada Tabel 1 [16].

Tabel 1. *Confusion Matrix*

		Predicted		
		Class	Positive	Negative
Actual	Positive	True Positive (TP)	False Positif (FP)	
	Negative	False Negatif (FN)	True Negative (TN)	

Keterangan:

- *True Positive* (TP) merupakan jumlah *record* positif dalam dataset yang diklasifikasikan positif.
- *True Negative* (TN) merupakan jumlah *record* negatif dalam dataset yang diklasifikasikan positif.
- *False Positive* (FP) merupakan jumlah *record* negatif dalam dataset yang diklasifikasikan positif.
- *False Negative* (FN) merupakan jumlah *record* positif dalam dataset yang diklasifikasikan negatif.

Untuk mendapatkan nilai dari *accuracy*, *precision*, *recall* dan F1-Score maka digunakan rumus seperti dibawah ini.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{10}$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{11}$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{12}$$

$$\text{F1-Score} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \times 100\% \tag{13}$$

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diambil dari *Twitter* menggunakan *keyword* “PSE” pada tanggal 16 Juli-06 Agustus 2022 dengan proses *crawling* data menggunakan *Twitter API Developer*. Data yang didapatkan berjumlah 1500 data dan disimpan kedalam file .csv, setelah melalui proses *preprocessing* data berubah menjadi 1197 data dan disimpan kedalam file excel. Untuk contoh dataset dapat dilihat pada Tabel 2.

Tabel 2. Dataset

No	tweetcreatedts	text
1	2022-07-16 13:57:34+00:00	Aku baca di CNN itu bakal diblokir kalo mereka gak daftar PSE atau Penyelenggara.
2	2022-07-16 13:56:52+00:00	Banyak tidak paham malah menghujat menkominfo. Ini gertakan biar semua perusahaan tersebut melegalkan perusahaan mereka disini.
3	2022-07-16 13:56:00+00:00	RT @gatorade: Tgl 20 Juli ini, berarti 4 hari lagi, PSE, Penyelenggara Sistem Elektronik, belum mendaftar ke Kominfo, diancam diblokir
4	2022-07-16 13:56:00+00:00	RT @gatorade: Tgl 20 Juli ini, berarti 4 hari lagi, PSE, Penyelenggara Sistem Elektronik, belum mendaftar ke Kominfo, diancam diblokir
5	2022-07-16 13:44:00+00:00	RT @korbankomet: @CNNIndonesia Ini bukan mengancam kalian para penggunya , ini lagi mengancam pemilik aplikasinya , kalau mereka belum daftar pse

3.2. Preprocessing

Setelah melakukan pengumpulan data, selanjutnya dilakukan *preprocessing* untuk menghilangkan karakter angka dan *username* data yang telah diperoleh. Berikut beberapa tahapan dalam melakukan *preprocessing*.

3.2.1. Case folding

Case Folding adalah tahapan mengubah huruf kapital menjadi huruf kecil (*lowercase*). Tahapan ini akan mengubah semua huruf yang terdapat pada data menjadi huruf kecil. Huruf yang berubah yaitu huruf ‘a’ sampai dengan huruf ‘z’. Untuk melakukan proses *case folding* seperti pada tabel 3.

Tabel 3. Hasil Case Folding

No	tweetcreatedts	text
1	2022-07-16 13:57:34+00:00	aku baca di cnn itu bakal diblokir kalo mereka gak daftar pse atau penyelenggara
2	2022-07-16 13:56:52+00:00	banyak tidak paham malah menghujat menkominfo. ini gertakan biar semua perusahaan tersebut melegalkan perusahaan mereka disini.
3	2022-07-16 13:56:00+00:00	rt @gatorade: tgl 20 juli ini, berarti 4 hari lagi, pse, penyelenggara sistem elektronik, belum mendaftar ke kominfo, diancam diblokir
4	2022-07-16 13:56:00+00:00	rt @gatorade: tgl 20 juli ini, berarti 4 hari lagi, pse, penyelenggara sistem elektronik, belum mendaftar ke kominfo, diancam diblokir
5	2022-07-16 13:44:00+00:00	rt @korbankomet: @cnnindonesia ini bukan mengancam kalian para penggunya , ini lagi mengancam pemilik aplikasinya , kalau mereka belum daftar pse

3.2.2. Cleaning

Tahapan selanjutnya adalah melakukan *cleaning*, yaitu menghilangkan atau menghapus karakter, *symbol*, *username* dan data duplikat. Dari data awal 1500 data tweet, setelah dilakukan penghapusan data duplikat menjadi 1197 data, seperti ditunjukkan pada Tabel 4.

Tabel 4. Hasil Cleaning

No	tweetcreatedts	text
1	2022-07-16 13:57:34+00:00	aku baca di cnn itu bakal diblokir kalo mereka gak daftar pse atau penyelenggara
2	2022-07-16 13:56:52+00:00	banyak tidak paham malah menghujat menkominfo. ini gertakan biar semua perusahaan tersebut melegalkan perusahaan mereka disini.
3	2022-07-16 13:56:00+00:00	tgl juli ini, berarti hari lagi, pse, penyelenggara sistem elektronik, belum

No	tweetcreatedts	text
		mendaftar ke kominfo, diancam diblokir
4	2022-07-16 13:44:00+00:00	ini bukan mengancam kalian para penggunanya , ini lagi mengancam pemilik aplikasinya , kalau mereka belum daftar pse

3.2.3. Tokenizing

Tokenizing adalah tahapan memotong kalimat menjadi beberapa kata. Pada proses ini memecah sekumpulan karakter dalam sebuah teks menjadi satuan kata. Untuk contoh *tokenize* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Tokenizing*

No	text
1	[aku, baca, di, cnn, itu, bakal, diblokir, kalo, mereka, gak, daftar, pse, atau, penyelenggara]
2	[banyak, tidak, paham, malah, menghujat, menkominfo, ini, gertakan, biar, semua, perusahaan, tersebut, melegalkan, perusahaan, mereka, disini]
3	[tgl, juli, ini, berarti, hari, lagi, pse, penyelenggara, sistem, elektronik, belum, mendaftar, ke, kominfo, diancam, diblokir]
4	[ini, bukan, mengancam, kalian, para, penggunanya, ini, lagi, mengancam, pemilik, aplikasinya , kalau, mereka, belum, daftar, pse]

3.2.4. Filtering

Filtering adalah tahapan untuk menghapus kata-kata yang tidak memiliki makna. Pada tahap ini menggunakan fungsi yang ada dilibrary NLTK, yaitu *stopword* bahasa Indonesia. *Stopword* adalah tahapan menghilangkan kata yang tidak memiliki makna atau tidak memiliki nilai, kata yang biasa dihilangkan berupa kata ganti, kata penunjuk, kata sapaan, dan sebagainya. Contoh *stopword* yaitu “yang”, “dan”, “atau”, “rt”, “dengan”, “karena”, dan masih banyak lagi. Hasil filtering dapat dilihat pada Tabel 6.

Tabel 6. *Filtering*

No	text
1	[baca, cnn, diblokir, tidak,daftar, pse, penyelenggara]
2	[banyak, tidak, paham, menghujat, menkominfo, gertakan, semua, perusahaan, melegalkan, perusahaan]
3	[pse, penyelenggara, sistem, elektronik, belum, mendaftar, kominfo, diancam, diblokir]
4	[bukan, mengancam, penggunanya, mengancam, pemilik, aplikasinya , belum, daftar, pse]

3.2.5. Stemming

Stemming adalah tahapan merubah kata yang memiliki imbuhan menjadi kata dasar. Perubahan menjadi kata dasar berawal dari kata imbuhan berdasarkan posisinya yaitu awalan (prefiks), akhiran (subfiks), sisipan (infiks), dan gabungan awalan dan akhiran (konfiks). Contohnya seperti kata “ancaman” menjadi “ancam”, “penyelenggara” menjadi “selenggara”. Untuk hasil *stemming* dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Stemming*

No	text
1	[baca, cnn,blokir, tidak,daftar, pse, selenggara]
2	[banyak, tidak, paham, hujat, menkominfo, gertakan, semua, usaha, legal, usaha]
3	[pse, selenggara, sistem, elektronik, belum, mendaftar, kominfo, ancam, blokir]
4	[bukan, ancam, guna, ancam, pemilik, aplikasi , belum, daftar, pse]

3.3. Pelebelan Data

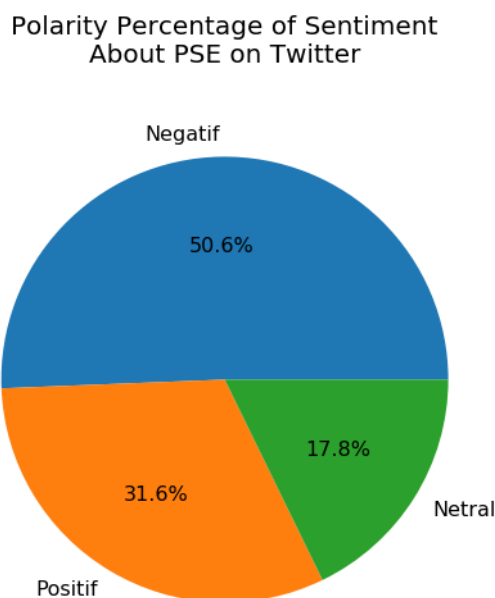
Tahap selanjutnya yaitu labeling, data yang telah dilakukan proses *preprocessing* belum diketahui kelas sentimentnya. Pada penelitian ini terdapat tiga kelas sentiment yaitu positif, netral, dan negatif. Pada proses ini data di klasifikasikan sesuai dengan sentimentnya. *Library* yang digunakan pada proses ini yaitu *VanderSentiment*. *Library VADER (Valence Aware Dictionary and SENTiment Reasoner)* adalah pustaka

penganalisis sentimen sumber terbuka berbasis aturan / leksikon, dilindungi di bawah lisensi MIT. Proses memetakan nilai sentimen untuk data tersebut dibutuhkan kolom baru yaitu ‘Compound Score’, kolom score akan memiliki nilai numerik[17]. Jika nilai lebih dari 0 maka sentiment positif, jika nilai 0 maka sentiment netral, dan jika nilai kurang dari 0 maka akan menghasilkan sentiment negatif. Dalam menggunakan vandersentimen, data yang sudah melalui proses preprocessing diterjemahkan kedalam bahasa inggris untuk mendapatkan hasil yang maksimal. Seperti pada tabel 8.

Tabel 8. Label Data

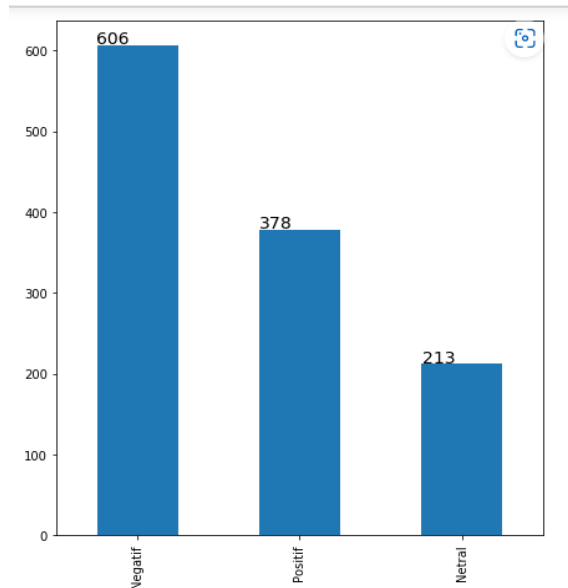
No	Text Translate	Compound Score	Sentimen
1	read cnn block, not list pse organizers	-0.7096	Negatif
2	many do not understand the blasphemy of the minister of communication and information on the bluffing of all legal business ventures	0.1280	Positif
3	PSE Implements Electronic Systems Has Not Registered Kominfo Threatens to Block	-0.6705	Negatif
4	not a threat to threaten the owner of the application has not registered pse	-0.7184	Negatif

Setelah nilai sentimen diketahui selanjutnya dilakukan proses untuk mengelompokkan jumlah sentimen menggunakan circle chart dengan persentase. Seperti pada gambar 1



Gambar 1. Persentase Sentimen

Berdasarkan Gambar 1, terlihat bahwa 50% data tweet bersentimen negatif, 31,6% bersentimen positif dan 17,8% bersentimen netral.. Selanjutnya untuk mengetahui jumlah sentiment setiap *tweet* dengan *chart sentiment* seperti Gambar 2.



Gambar 2. Chart Sentimen

Dari Gambar 2 dapat terlihat bahwa dari 1197 data, yang bersentimen negatif berjumlah data terbanyak yaitu 606 data, sentiment positif berjumlah 378 data dan sentiment netral berjumlah 213 data. Dari data tersebut dapat terlihat bahwa banyak masyarakat Indonesia terutama pengguna *Twitter* tidak suka dengan kebijakan yang dikeluarkan kominfo terkait PSE. Contoh beberapa *tweet* yang termasuk kedalam 3 kelas sentiment yaitu positif, negatif dan netral dapat dilihat pada tabel 9.

Tabel 9. Contoh Tweet dan Kelas Sentimen

Tweet	Sentimen
sangat menyusahkan, membuat ribut saja aturan PSE ini, cabut saja aturannya	negatif
mengatasi data yang bocor saja tidak bisa kominfo ini, apalagi mengurus pse	negatif
Masa Depan Internet indonesia akan semakin memburuk jika pemerintah masih tidak mau cabut Aturan PSE.	negatif
aku sih dukung pse	positif
Untuk melindungi masyarakat dari tindak kejahatan penipuan dan pencurian data yang sering terjadi di Indonesia melalui media sosial,pemerintah resmi menerbitkan aturan PSE lingkup privat yang bisa mencegah terjadinya tindak pidana cyber di Indonesia, terutama penipuan.	netral

3.4. Implementasi Algoritma *Naïve Bayes*

Tahapan ini mengimplementasikan metode *Naïve Bayes* kedalam data yang sudah melalui tahapan preprocessing dan pelebalan data. Sebelum itu kita akan membagi data menjadi 20% data testing dan 80% data training. selanjutnya mengimplementasikan metode *Naïve Bayes* menggunakan library *MultinomialNB*. Pengklasifikasi *MultinomialNB* adalah jenis pengklasifikasi NB dan sering digunakan sebagai dasar untuk klasifikasi teks[18]. Untuk penggunaan *library* nya dapat dilihat pada Kode Program 1.

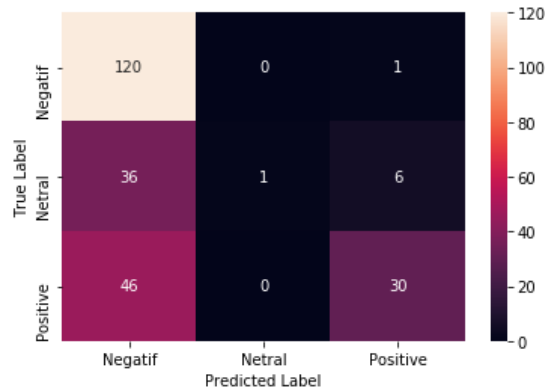
```

1  mnb = MultinomialNB()
2  mnb.fit(X_train, y_train)
3  predict = mnb.predict(X_test)
4  print(predict)
    
```

Kode Program 1. *Naïve Bayes*

3.5. Pengujian Confusion Matrix

Setelah mengimplementasikan ke dalam *Naive Bayes* selanjutnya yaitu melakukan pengujian algoritma untuk mengetahui akurasi dari metode *Naive Bayes*, menggunakan *confusion matrix*. Untuk mengetahui hasil *confusion matrix* dapat dilihat pada Gambar 3.



Gambar 3 Hasil *Confusion Matrix*

Berdasarkan Gambar 3 *Confusion Matrix*, diperoleh jumlah dokumen dari kelas sebenarnya kelas negatif (TN) yang benar diklasifikasikan sebagai kelas negatif berjumlah 120 data, jumlah dokumen dari kelas netral yang salah diklasifikasikan sebagai kelas negatif berjumlah 36 data, jumlah dokumen dari kelas positif yang salah diklasifikasikan sebagai kelas negatif berjumlah 46 data. Selanjutnya jumlah dokumen dari kelas netral yang benar diklasifikasikan sebagai kelas netral berjumlah 1 data, jumlah dokumen kelas positif yang benar diklasifikasikan sebagai kelas positif berjumlah 30 data.

Setelah mendapatkan hasil *confusion matrix*, selanjutnya menghitung akurasi menggunakan persamaan 10, menghitung *precision* menggunakan 11, menghitung *recall* dengan persamaan 12 dan *f1-score* dengan persamaan 13. Hasil perhitungan dapat dilihat pada tabel 10.

Tabel 10. Hasil Akurasi, *Precision*, *Recall* dan *F1-Score*

		Precision	Recall	F1-Score
Kelas Sentimen	<i>Negatif</i>	59%	99%	74%
	<i>Netral</i>	100%	2%	5%
	<i>Positif</i>	81%	39%	53%
Avg		80%	46%	44%
Akurasi				63%

Berdasarkan Tabel 10. akurasi yang diperoleh dari pengujian pengklasifikasian sentimen dengan metode *Naive Bayes* yaitu akurasi sebesar 63%, serta nilai rata-rata dari *precision* adalah 80%, *recall* 46% dan *f1-score* 44%.

4. Kesimpulan

Berdasarkan hasil dari penelitian analisis sentimen pada data yang diperoleh dari *Twitter* dengan *keyword* “PSE” dengan data yang diperoleh dari hasil *crawling* data 1500, maka dapat ditarik kesimpulan. Pada analisis sentiment data *Twitter* yang diperoleh dengan topik PSE menghasilkan nilai sentiment negatif terbanyak yaitu 606 data, sedangkan sentiment positif 378 data, dan sentimen netral sebanyak 213 data. Hasil akurasi yang didapatkan adalah 63%, dengan *precision* 80%, *recall* 46% dan *F-score* 44%.

Daftar Pustaka

- [1] Kominfo, “Sejarah Kominfo,” 2022. [Online]. Available: <https://www.kominfo.go.id/profil>.

- [2] LinovHR, "Mengenal PSE Kominfo dari Pro Kontra Hingga Siapa Saja yang Terdaftar," *LinovHR*, 2022. [Online]. Available: <https://www.linovhr.com/mengenal-pse-kominfo-adalah/>. [Accessed: 18-Aug-2022].
- [3] F. C. Pradipha and O. W. Widayanti, "Alasan Kominfo Blokir Steam, Epic Games, Dota dan Beberapa Aplikasi Lainnya," *Tribunnews.com*, 2022. [Online]. Available: <https://www.tribunnews.com/nasional/2022/07/30/alasan-kominfo-blokir-steam-epic-games-dota-dan-beberapa-aplikasi-lainnya>.
- [4] A. Brilliana, "Trending! Tagar Blokir Kominfo Trending di Twitter Setelah Memblokir Situs," *bekasiurbanjabar.com*, 2022. [Online]. Available: <https://bekasi.urbanjabar.com/news/pr-3114016802/trending-tagar-blokir-kominfo-trending-di-twitter-setelah-memblokir-situs#>. [Accessed: 15-Aug-2022].
- [5] I. N. Jelita, "PayPal Diblokir, Hambat Aktivitas Pekerja," 2022. [Online]. Available: <https://mediaindonesia.com/ekonomi/511149/paypal-diblokir-hambat-aktivitas-pekerja>. [Accessed: 20-Aug-2022].
- [6] S. Suryono, E. Utami, and E. T. Luthfi, "Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naive Bayes Classifier," pp. 9–15, 2018.
- [7] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021.
- [8] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [9] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017.
- [10] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020.
- [11] S. D. Pramukti, "Analisis Sentimen Masyarakat Dengan Metode Naive Bayes dan Particle Swarm Optimization," *Techno.Com*, vol. 21, no. 1, pp. 61–74, 2022.
- [12] V. A. Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *J. Buana Inform.*, vol. 11, no. 2, p. 140, 2020.
- [13] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018.
- [14] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [15] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, p. 32, 2021.
- [16] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naive Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020.
- [17] S. Panchal, "Sentiment Analysis with VADER- Label the Unlabelled Data," 2020. [Online]. Available: Sentiment Analysis with VADER- Label the Unlabelled Data.
- [18] S. Informatika and J. Komputer, "Perbandingan algoritma dalam analisa sentimen krisis evergrande pada kanal berita youtube," vol. 11, no. 2, pp. 72–76, 2021.