

# Analisis Sentimen Masyarakat Dengan Metode *Naïve Bayes* dan *Particle Swarm Optimization*

*Public Sentiment Analysis using Naïve Bayes Method and Particle Swarm Optimization*

Septiawan Dwi Pramukti<sup>1</sup>, Agung Nugroho<sup>2</sup>, Aswan Supriyadi Sunge<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Universitas Pelita Bangsa

E-mail: <sup>1</sup>septiawandpramukti@gmail.com, <sup>2</sup>agung@pelitabangsa.ac.id,

<sup>3</sup>aswan.sunge@pelitabangsa.ac.id

## Abstrak

Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) merupakan kebijakan yang di terapkan pada 15 Kota diluar pulau Jawa dan pulau Bali yang memiliki status zona merah atau daerah yang memiliki risiko tinggi terhadap paparan kasus *COVID-19*. Pada tanggal 3 Agustus 2021 kebijakan PPKM darurat di beberapa daerah di kepulauan Jawa dan kepulauan Bali di perpanjang sampai 9 Agustus 2021, perpanjangan kebijakan PPKM darurat menuai banyak komentar dari masyarakat sehingga menimbulkan pro dan kontra pada social media *twitter*. Penelitian ini bertujuan untuk mengetahui hasil klasifikasi sentimen masyarakat mengenai perpanjangan kebijakan PPKM darurat pada social media *twitter* dan untuk mengetahui hasil *accuracy*, *precision*, *recall* yang dihasilkan dari metode yang digunakan. Penelitian ini menggunakan metode *Naïve Bayes* dan *Particle Swarm Optimization* sebagai *feature selection*, selain itu terdapat tahap *preprocessing* yang didalamnya meliputi *cleansing*, *remove duplicate*, seleksi data, normalisasi, *case folding*, *tokenizing*, *filtering*, *stopwords*, *stemming*, dan *labeling*. Hasil klasifikasi yang didapat 53,31% pengguna *twitter* setuju dan 46,69% pengguna *twitter* tidak setuju dengan perpanjangan kebijakan PPKM darurat. Nilai *accuracy* yang didapatkan meningkat sebanyak 15,21% dari 77,16% menjadi 92,37%, nilai *precision* yang didapatkan meningkat sebanyak 3,07% dari 87,33% menjadi 90,40%, dan nilai *recall* yang didapatkan meningkat sebanyak 30,96% dari 64,42% menjadi 95,38%.

Kata kunci: *Naïve Bayes*, *Particle Swarm Optimization*, Analisis Sentimen, PPKM Darurat

## Abstract

*The implementation of Community Activity Restrictions (PPKM) is a policy that is applied to 15 cities outside Java and Bali that have red zone status or areas that have a high risk of exposure to COVID-19 cases. On August 3, 2021, the emergency PPKM policy in several areas in the Java and Bali islands was extended until August 9, 2021, the extension of the emergency PPKM policy drew many comments from the public, causing pros and cons on Twitter social media. This study aims to determine the results of the classification of public sentiment regarding the extension of the emergency PPKM policy on Twitter social media and to find out the results of accuracy, precision, recall resulting from the method used. This study uses the Naïve Bayes method and Particle Swarm Optimization as feature selection, besides that there is a preprocessing stage which includes cleansing, remove duplicates, data selection, normalization, case folding, tokenizing, filtering, stopwords, stemming, and labeling. The classification results obtained are 53.31% of twitter users agree and 46.69% of twitter users do not agree with the extension of the emergency PPKM policy. The accuracy value obtained increased by 15.21% from 77.16% to 92.37%, the precision value obtained increased by 3.07% from 87.33% to 90.40%, and the recall value obtained increased by 30.96% from 64.42% to 95.38%.*

Keywords: *Naïve Bayes*, *Particle Swarm Optimization*, Sentiment Analysis, Emergency PPKM

## 1. PENDAHULUAN

*Covid-19* merupakan penyakit yang muncul pertama kali di daerah Wuhan, Tiongkok yang disebabkan oleh *coronavirus* jenis baru dengan tingkat penularan yang tinggi, pada bulan Desember 2019 *World Health Organization* (WHO) memberikan nama untuk *coronavirus* jenis baru dengan sebutan *Severe Acute Respiratory Syndrome Coronavirus 2 (Sars-Cov2)*, *Sars-Cov2* ini lah yang menjadi awal mula penyebab *Covid-19* [1]. *Covid-19* merupakan virus yang sampai saat ini masih menjadi *trending topic* diseluruh negara, termasuk negara Indonesia. Mulai dari awal kasus *Covid-19* sampai saat ini, pasien terkonfirmasi positif *Covid-19* di Indonesia setiap hari terus-menerus bertambah. Hal ini lah yang menjadi faktor penyebab pemerintah melakukan berbagai upaya untuk dapat menangani dan mencegah pertumbuhan kasus *Covid-19* yang baru, dengan mengeluarkan berbagai kebijakan seperti *lockdown*, PSBB, dan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Kebijakan PPKM di terapkan di beberapa daerah di kepulauan Jawa dan kepulauan Bali, yang memiliki status zona merah atau daerah yang memiliki risiko tinggi terhadap paparan kasus *Covid-19*. Pada 25 Juli 2021 pemerintah Indonesia kembali menerapkan kebijakan pelaksanaan PPKM darurat di beberapa daerah di kepulauan Jawa dan kepulauan Bali, yang berlaku sampai 2 Agustus 2021. Pada tanggal 3 Agustus 2021 kebijakan PPKM darurat di beberapa daerah di kepulauan Jawa dan kepulauan Bali di perpanjang sampai 9 Agustus 2021, perpanjangan kebijakan PPKM darurat menuai banyak respon berupa *tweets* dari berbagai kalangan masyarakat, sehingga menghasilkan banyak jejak *tweets* yang di dalamnya berisikan opini masyarakat mengenai perpanjangan kebijakan PPKM darurat pada media sosial *twitter*. *Tweets* adalah sebuah kalimat yang berisikan pesan atau status yang dibuat oleh pengguna akun *twitter* yang di dalamnya mengandung sebuah opini, opini didalam *tweets* dapat mengekspresikan sebuah perasaan atau keadaan dari pengguna akun *twitter* tersebut, *tweets* juga dapat digunakan sebagai penilaian bagi perusahaan atau instansi [2]. Banyaknya respon masyarakat Indonesia dalam menilai kebijakan pemerintah menjadi acuan bagi peneliti untuk melakukan analisis sentimen masyarakat mengenai kebijakan perpanjangan PPKM darurat pada social media *twitter*.

Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) merupakan kebijakan yang di terapkan pada 15 Kota diluar pulau Jawa dan pulau Bali yang memiliki status zona merah atau daerah yang memiliki risiko tinggi terhadap paparan kasus *COVID-19* [3]. Proses pengambilan data yang berisikan informasi penting yang didapatkan dari suatu media penyimpanan (*database*) merupakan definisi *data mining* [4]. Digunakan teknik matematika, statistika, kecerdasan buatan, dan *machine* pada *data mining* [5]. Tujuan *data mining* untuk klasifikasi, klusterisasi, mendapatkan pola asosiasi sampai melakukan sebuah peramalan (*predicting*) [6]. *Association rule mining*, *clustering*, dan klasifikasi merupakan 3 teknik yang sangat populer [7]. *Text mining* biasanya digunakan dalam mengklasifikasi sebuah dokumen yang berisikan teks [8]. Tahapan pada *text mining* terdiri dari teks, pengolahan teks (tokenisasi), perubahan teks (*stemming*), pemilihan teks (*filtering*), *data mining (pattern discovery)*, dan evaluasi [9]. *Preprocessing* adalah tahap penghapusan teks yang tidak dibutuhkan dan tidak digunakan pada saat proses klasifikasi teks [10]. *Preprocessing* dilakukan bertujuan untuk memperbaiki data yang sebelumnya tidak terstruktur menjadi data yang terstruktur, tujuannya untuk dapat mempermudah dan mempercepat pemrosesan data [11]. Terdapat beberapa tahap pada *preprocessing* contohnya seperti *tokenize*, *transform cases*, *filter tokens*, *filter stopword*, dan *stemming* [8]. *Natural Language Processing* (NLP) merupakan bidang penelitian baru dalam analisis sentimen [11]. *Opinion Mining* adalah bidang studi yang dipakai untuk menganalisis suatu pendapat atau pandangan, penilaian atau evaluasi, sikap, dan luapan perasaan terhadap entitas dan aspek-aspek lainnya yang diungkapkan dengan sebuah teks [12]. Tujuan dari analisis sentimen yaitu untuk menentukan polaritas dari suatu teks [13]. Untuk melakukan analisis sentimen terdapat beberapa tahap yang harus di lakukan antara lain menentukan sebuah *dataset*, *preprocessing*, pemilihan *feature selection*, pemberian label sentimen, klasifikasi dan evaluasi [14].

Penelitian yang dilakukan oleh Renaldy Permana Sidiq, Budi Arif Dermawan, & Yunun Umidah [10] yang berjudul “Sentimen Analisis Komentar *Toxic* pada Grup Facebook *Game*

Online Menggunakan Klasifikasi *Naïve Bayes*”, penelitian ini membahas tentang analisis komentar *toxic* pada grup facebook *game online arena of valor*. Tujuan dari penelitian ini untuk melakukan klasifikasi *sentiment toxic*, dan *non-toxic* terhadap data uji, dan juga membandingkan akurasi klasifikasi menggunakan metode *Naïve Bayes* dengan transformasi TF-IDF, tanpa TF-IDF dan seleksi fitur *Information Gain* serta menggunakan rasio pembagian data 80:20. Dataset yang digunakan berjumlah 1.500 yang di ambil dari media social facebook. Dikarenakan jumlah kata *non-toxic* cenderung lebih sering muncul pada *wordcloud* hasil klasifikasi sentimen cenderung *non-toxic* dengan jumlah 1.237 komentar dan *toxic* dengan jumlah 263 komentar. Hasilnya klasifikasi *Naïve Bayes* yang menggunakan transformasi TF-IDF menghasilkan nilai akurasi tertinggi dengan akurasi sebesar 75%, *precision* sebesar 63%, *recall* sebesar 67%, dan *F-measure* sebesar 64%. Selanjutnya penelitian yang dilakukan oleh Hennie Tuhuteru, & Ade Iriani [11] yang berjudul “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode *Support Vector Machine* dan *Naïve Bayes Classifier*”, dataset yang digunakan berjumlah 1492 *tweet* berasal dari *tweet* yang di *mention* ke akun @ambonlima dan *keyword* “PLN Ambon” data diambil menggunakan metode *snipping*. Tujuan dari penelitian ini untuk dapat mengetahui sentimen masyarakat terhadap kondisi kelistrikan, khususnya pemadaman listrik yang terjadi di Pulau Ambon. *Text pre-processing* dilakukan melalui 6 tahapan yaitu *cleansing*, *case folding*, *tokenizing*, *normalization*, *filtering*, dan *stemming*. Pada penelitian ini digunakan 2 metode yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* untuk dibandingkan metode mana yang memberikan hasil terbaik. Hasil dari penelitian menggunakan metode NBC diketahui bahwa *sentiment* masyarakat cenderung positif dengan nilai akurasi 67.20%, namun hasilnya cenderung negatif jika menggunakan metode SVM dengan tingkat akurasi 81.67%. Jika dibandingkan metode SVM lah yang memiliki nilai akurasi yang lebih baik, namun seiring bertambahnya nilai *fold* pada proses validasi, metode SVM membutuhkan waktu yang lebih lama dibandingkan metode NBC. Proses pengujian pada *K-Fold Cross Validation* sangat berpengaruh dalam meningkatkan akurasi metode klasifikasi NBC dan SVM. Hasil analisis *sentiment* dengan menggunakan metode klasifikasi NBC menghasilkan sentimen positif sebesar 67%, sentimen netral 19%, dan *sentiment negative* 14%. Pada metode klasifikasi SVM menghasilkan sentimen positif sebesar 24%, sentimen *negative* 47%, dan sentimen netral 29%.

Penelitian yang di lakukan oleh Yono Cahyono, dan Saprudin [15] yang berjudul “Analisis Sentiment *Tweets* Berbahasa Sunda Menggunakan *Naïve Bayes Classifier* dengan seleksi fitur *Chi Squared Statistic*”, penelitian ini menggunakan metode *Naïve Bayes Classifier* dan seleksi fitur *Chi Squared Statistic*. *Dataset* yang digunakan berjumlah 316 *tweets* yang di dalamnya berisikan kalimat campuran bahasa Sunda dan bahasa Indonesia, *dataset* berasal dari *twitter* yang diambil dengan cara *crawl* data pada sosial media *twitter* dengan menggunakan RapidMiner. Tahap *pre-processing* yang dilakukan pada penelitian ini menggunakan *case folding*, *tokenize*, dan *stopword removal*. Seleksi fitur *chi square statistic* dilakukan *optimize selection* dengan menggunakan *forward selection*, *chi square statistic* dilakukan untuk memilih kata atau *term* apa saja yang dapat dijadikan sebagai wakil penting untuk kumpulan dokumen yang akan dianalisis. Menggunakan standar validasi *10-fold validation* untuk meningkatkan akurasi klasifikasi. Hasil penelitian ini menunjukkan penggunaan seleksi fitur *Chi Square Statistic* dan algoritma *Naïve Bayes Classifier* mendapatkan akurasi sebesar 78.48%. Selanjutnya penelitian yang dilakukan oleh Yono Cahyono [2] dengan judul “Analisis Sentiment Pada Sosial Media *Twitter* Menggunakan *Naïve Bayes Classifier* Dengan *Feature Selection Particle Swarm Optimization* Dan *Term Frequency*”, penelitian ini bertujuan untuk memisahkan sentimen *tweets* berbahasa Indonesia pada media sosial *twitter* dan mengetahui hasil dari akurasi yang didapatkan. Penggunaan metode *Naïve Bayes Classifier* (NBC) dengan *feature selection Particle Swarm Optimization* (PSO) dan *Term Frequency* bertujuan untuk mengurangi atribut yang kurang relevan pada saat proses klasifikasi. Tahap *text processing* yang ada pada penelitian ini antara lain *case folding*, *tokenizing*, *stopword removal*. Untuk mendapatkan hasil validasi yang akurat peneliti menggunakan stratified *10-fold cross-validation*. Hasil penelitian penggunaan seleksi fitur PSO menggunakan parameter *term frequensi* (TF) dengan metode NBC hasil *term* yang terseleksi sebanyak 776 kata, sedangkan seleksi fitur PSO menggunakan parameter TF-IDF

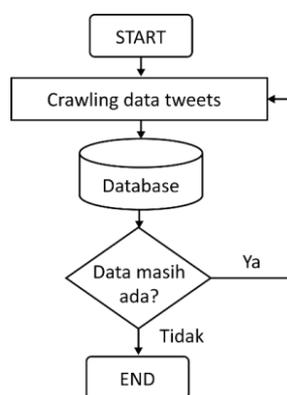
dengan metode NBC *terms* yang didapatkan berjumlah 774. Hasil pengujian seleksi fitur PSO terbukti dapat meningkatkan akurasi algoritma *Naïve bayes classifier*, hasil akurasi yang di dapat setelah pengujian seleksi fitur PSO menggunakan *parameter term frequency* (TF) dengan metode *Naïve Bayes Classifier* sebesar 97,48%.

Penelitian terdahulu yang dilakukan oleh Karno Juni Prayoga, Agung Nugroho, dan Tri Ngudi Wiyatno [12] yang berjudul “Komparasi *Feature Selection Particle Swarm Optimization* (PSO) Dengan *Genetic Algorithm* (GA) Terhadap *Algoritma Naïve Bayes* pada Analisis Sentiment *Twitter*”. Penelitian ini bertujuan untuk membandingkan *feature selection Particle Swarm Optimization* dan *Genetic Algorithm* dalam meningkatkan kinerja akurasi dari algoritma *Naïve Bayes*. *Dataset* yang digunakan adalah komentar yang berasal dari akun twitter @hendralm. Terdapat tahap *preprocessing* sebagai pemrosesan data awal yang terdiri dari seleksi data, *remove duplicate*, *cleansing*, normalisasi kata, dan penentuan kelas atribut. Selain itu terdapat *process document* yang didalamnya terdiri dari *transform case*, *tokenization*, *filter tokens (By Length)*, *filter stopwords*, *stemming*, dan *N-Gram (bigram)*. Hasil dari pengujian yang dilakukan terdapat peningkatan akurasi pada algoritma *Naïve Bayes* dengan *feature selection Particle Swarm Optimization* dari 67.17% menjadi 71.84%, dan terdapat peningkatan pula pada akurasi algoritma *Naïve Bayes* dengan *feature selection Genetic Algorithm* dari 67.71% menjadi 68.96%. Dengan demikian dapat diketahui bahwa *Particle Swarm Optimization* adalah *feature selection* terbaik dengan peningkatan akurasi sebesar 4.00%. Oleh karena itu, peneliti tertarik menggunakan metode *Naïve Bayes* dalam menganalisis sentimen masyarakat pada sosial media *twitter*, dikarenakan berdasarkan hasil literatur menunjukkan bahwa metode *Naïve Bayes* adalah metode sederhana yang mempunyai nilai *accuracy* dan performansi yang tinggi untuk mengklasifikasi sebuah teks. Untuk mendapatkan hasil akurasi yang lebih baik, ditambahkan *Particle Swarm Optimization* sebagai *feature selection*. Penelitian ini dilakukan bertujuan untuk mengetahui hasil klasifikasi sentimen masyarakat mengenai perpanjangan kebijakan PPKM darurat pada social media *twitter* dan untuk mengetahui perbandingan nilai *accuracy*, *precision*, *recall* yang dihasilkan dari metode yang digunakan yaitu metode *Naïve Bayes* dan PSO.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Data set yang digunakan dalam penelitian ini berasal dari social media *twitter* yang diambil melalui situs <https://www.twitter.com>, yang dilakukan dengan cara *crawling* data. Pengumpulan data *twitter* diambil melalui operator *search twitter* yang ada pada *software RapidMiner Studio*. Data *tweet* yang diambil berisikan pendapat masyarakat mengenai perpanjangan kebijakan PPKM darurat di Indonesia yang diambil pada tanggal 2 Agustus 2021, pada jam 18:40 wib sampai 21:11 wib malam dengan *query* “PPKM Diperpanjang”, *result type* “*recent or popular*”, *limit* “3000”, dan *language* “id”.



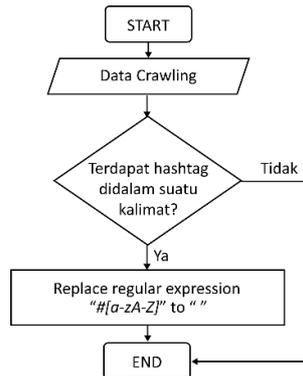
Gambar 1 Alur proses *crawling* data

## 2.2 Preprocessing

Tahap *preprocessing* adalah tahap yang dipakai untuk menghapus kata atau teks yang tidak dibutuhkan dalam proses klasifikasi. Terdapat beberapa proses yang terdapat pada tahap *preprocessing* yang terdiri dari *cleansing*, *remove duplicate*, seleksi data, normalisasi, *transform case*, *tokenizing*, *filtering*, *stopword*, *stemming*, dan pemberian label (*labeling*). Di bawah ini merupakan penjelasan mengenai proses-proses yang ada pada tahap *preprocessing*.

### 1. Cleansing

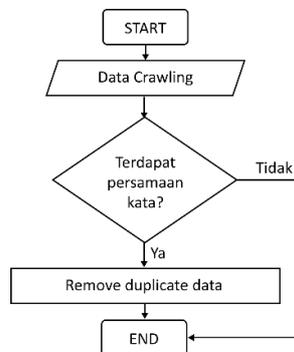
Tahap *cleansing* atau tahap pembersihan data bertujuan untuk menghapus special karakter (simbol), nama *username*, *retweet* (RT), dan link yang terdapat didalam dataset.



Gambar 2 Alur dari proses *cleansing*

### 2. Remove duplicate

Tahap yang berfungsi untuk menghapus atribut yang bernilai sama (*duplicate*) didalam data yang di *crawling*.



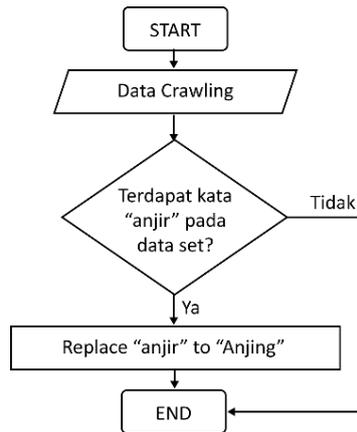
Gambar 3 Alur *remove duplicate*

### 3. Seleksi data

Tahap seleksi data adalah tahap penghapusan beberapa data *tweet* yang tidak berhubungan dengan perpanjangan PPKM darurat, contohnya seperti *tweet* promosi penjualan suatu produk dan *tweet* informasi berita dari berbagai situs yang di *share* kedalam *twitter*. Setelah dilakukan proses seleksi data dari 3000 dataset didapatkan total 302 dataset yang berhubungan dengan perpanjangan PPKM darurat.

### 4. Normalisasi

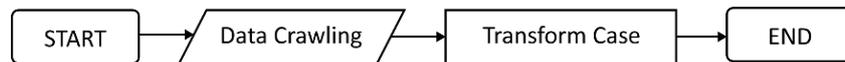
Normalisasi merupakan tahap untuk mengubah kata-kata yang terdapat pada dataset *twitter* yang sebelumnya kata tersebut tidak baku menjadi kata dasar yang baku atau sesuai ejaan kata yang benar. Normalisasi dilakukan dikarenakan kebiasaan pengguna *twitter* yang menggunakan kata singkatan ataupun kata *gaul* dalam menulis sebuah *tweet*.



Gambar 4 Alur proses normalisasi teks

### 5. Transform case

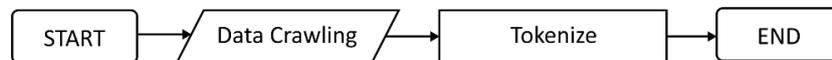
*Transform case* merupakan proses untuk mengubah semua huruf besar atau *uppercase* yang terdapat didalam dataset menjadi huruf kecil atau *lowercase*.



Gambar 5 Alur proses transform case

### 6. Tokenize

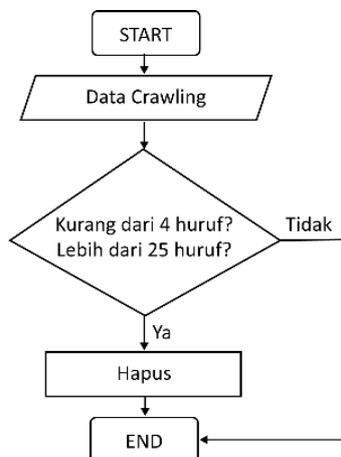
*Tokenize* merupakan tahap untuk memecah kalimat yang ada didalam data *tweet (dataset)*, data yang asalnya berbentuk sebuah kalimat akan di ubah ke bentuk kata perkata.



Gambar 6 Alur proses tokenize

### 7. Filtering

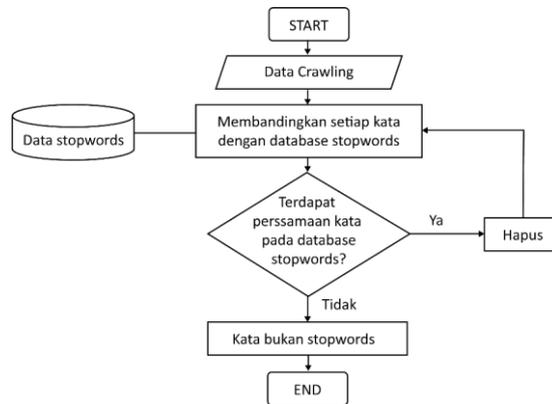
*Filtering* merupakan tahap menghilangkan kata-kata (setelah proses *tokenize*) yang tidak memiliki nilai dalam proses klasifikasi teks. Pada penelitian yang dilakukan, kata yang didalam *dataset* memiliki nilai kurang dari 4 karakter dan panjang melebihi 25 karakter akan hapus atau dihilangkan.



Gambar 7 Alur dari proses filtering

### 8. Stopwords

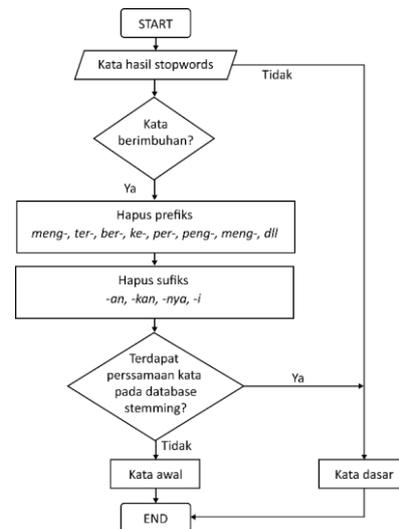
*Stopwords* merupakan proses menghilangkan kata sambung yang ada didalam dataset. Data *stopwords* yang didapatkan bersumber dari website <https://github.com/masdevi/ID-Stopwords> dengan total berjumlah 758 kata.



Gambar 8 Alur dari proses *stopwords*

### 9. Stemming

*Stemming* merupakan tahap akhir dalam tahap *preprocessing*, *stemming* adalah proses menghilangkan semua imbuhan (*prefix* dan *suffixs*) yang ada pada awal dan akhir kata. Kata dasar atau data *stem* didapatkan dari website <https://github.com/sastrawi/sastrawi/tree/master/data> dengan total berjumlah 30342 kata dasar.



Gambar 9 Alur dari proses *stemming*

### 10. Labeling

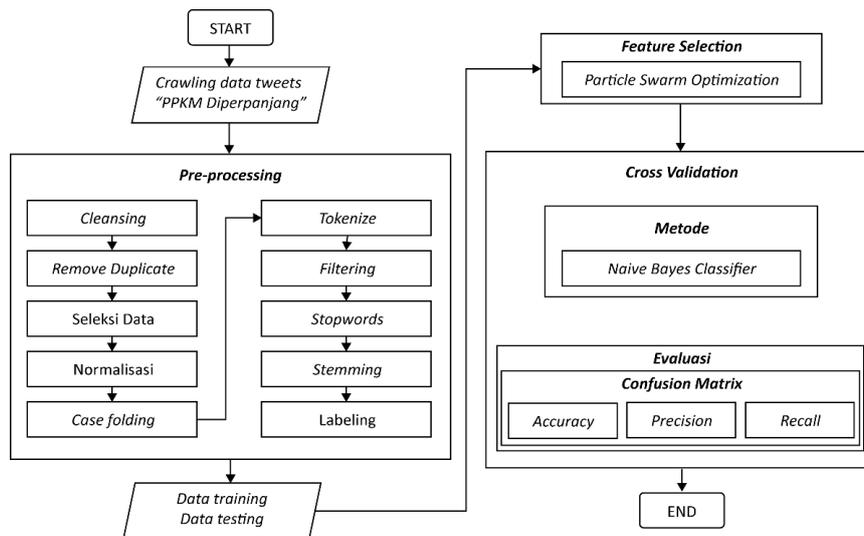
Labeling atau pemberian label merupakan tahap untuk menentukan respon *tweet* yang ada didalam *dataset* yang dilakukan secara manual, pendapat yang ada pada *dataset* akan dibagi dua menjadi pendapat positif dan negatif. Pendapat yang berisikan kalimat pujian atau dukungan terhadap perpanjangan kebijakan PPKM akan dikelompokkan ke dalam sentimen positif. Pendapat yang berisikan kata-kata atau kalimat ketidak setujuan, ataupun kata-kata kasar berdasarkan kamus *bad words* pada website <https://github.com/masdevi/ID-OpinionWords> akan di masukan kedalam sentimen negatif. Didapatkan 302 *tweets* mengandung sentimen yang terdiri dari 152 *tweets* bersentimen positif dan 150 *tweets* bersentimen negatif.

Table 1 Contoh sentimen positif dan negatif

Tweets	Label
selamat atlet berjuang semangat rakyat berjuang pemberlakuan pembatasan kegiatan masyarakat diperpanjang agustus	Positif
anjing pemberlakuan pembatasan kegiatan masyarakat diperpanjang	Negatif
pemberlakuan pembatasan kegiatan masyarakat diperpanjang tolong siaran badminton perbanyak	Positif
bangsat pemberlakuan pembatasan kegiatan masyarakat diperpanjang	Negatif

### 2.3 Metode yang digunakan

Penelitian ini menggunakan metode *Naïve Bayes* dan menambahkan *Particle Swarm Optimization* sebagai *feature selection*. Untuk mengetahui hasil evaluasi *accuracy*, *precision*, *recall* dari penggunaan metode *Naïve Bayes* dan *Particle Swarm Optimization* akan digunakan teknik *Cross Validation* dengan nilai  $K=10$ .



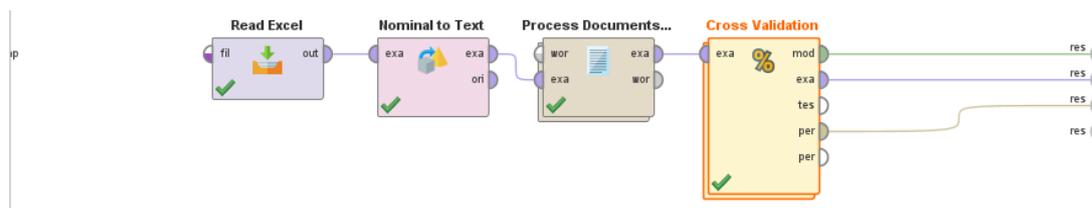
Gambar 10 Metode yang digunakan

## 3. HASIL DAN PEMBAHASAN

Pada tahap proses pengujian, dalam penelitian ini digunakan metode *Naïve Bayes* dan *Particle Swarm Optimization* sebagai *feature selection*. Proses pengujian metode dilakukan sebanyak 2 kali, untuk mengetahui hasil perbandingan *accuracy*, *precision*, *recall* dari penggabungan ke dua metode. Pada proses pengujian pertama dilakukan hanya dengan menggunakan metode *Naïve Bayes* dan pada proses pengujian kedua dilakukan menggunakan metode *Naïve Bayes* dengan *Particle Swarm Optimization*.

### 3.1 Proses pengujian pertama

Pengujian pertama dilakukan dengan metode *Naïve Bayes* tanpa *Particle Swarm Optimization*. Data yang digunakan berjumlah 302 data *tweets*, terdiri dari 152 *tweets* positif, dan 150 *tweets* negatif.



Gambar 11 Model pengujian pertama hanya menggunakan metode *Naïve Bayes*

Untuk mengetahui hasil evaluasi dari *accuracy*, *precision*, *recall* yang dihasilkan, pada pengujian pertama ini digunakan teknik *Cross Validation* menggunakan nilai  $K=10$ . Dibawah ini merupakan hasil *Confusion Matrix* pengujian metode *Naive Bayes* tanpa *Particle Swarm Optimization*.

### 1. Accuracy

Hasil *accuracy* dari penggunaan algoritma *Naive Bayes* tanpa *Particle Swarm Optimization* sebesar 77,16%, yang bisa dilihat pada gambar 12.

accuracy: 77.16% +/- 5.23% (micro average: 77.15%)

	true Negatif	true Positif	class precision
pred. Negatif	135	54	71.43%
pred. Positif	15	98	86.73%
class recall	90.00%	64.47%	

Gambar 12 Hasil *accuracy* dari algoritma *Naive Bayes* tanpa PSO

### 2. Precision

Hasil *precision* dari penggunaan algoritma *Naive Bayes* tanpa *Particle Swarm Optimization* sebesar 87,33%, yang bisa dilihat pada gambar 13.

precision: 87.33% +/- 6.63% (micro average: 86.73%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	135	54	71.43%
pred. Positif	15	98	86.73%
class recall	90.00%	64.47%	

Gambar 13 Hasil *precision* dari algoritma *Naive Bayes* tanpa PSO

### 3. Recall

Hasil *recall* dari penggunaan algoritma *Naive Bayes* tanpa *Particle Swarm Optimization* sebesar 64,42%, yang bisa dilihat pada gambar 14.

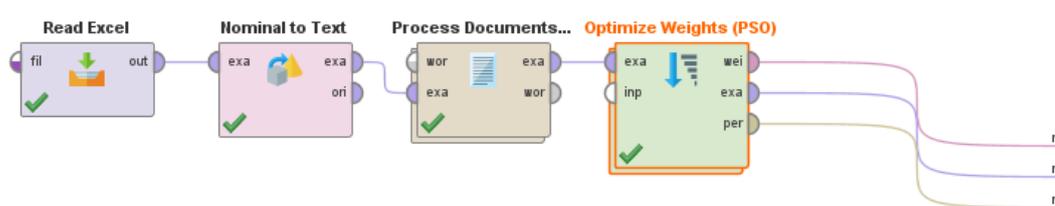
recall: 64.42% +/- 9.18% (micro average: 64.47%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	135	54	71.43%
pred. Positif	15	98	86.73%
class recall	90.00%	64.47%	

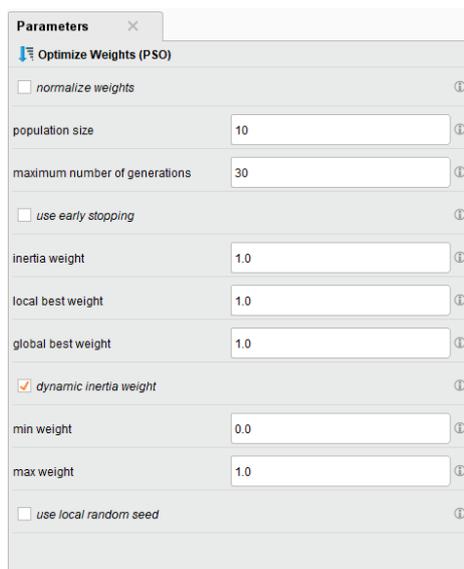
Gambar 14 Hasil *recall* dari algoritma *Naive Bayes* tanpa PSO

### 3.2 Proses pengujian kedua

Pengujian kedua dilakukan dengan metode *Naive Bayes* dan *Particle Swarm Optimization* sebagai *feature selection*. Data yang digunakan sama seperti pengujian pertama yaitu berjumlah 302 data *tweets*, terdiri dari 152 *tweets* positif, dan 150 *tweet* negatif.



Gambar 15 Pengujian kedua dengan metode *Naive Bayes* dan *feature selection* PSO



Gambar 16 Parameters PSO

Untuk mengetahui hasil evaluasi dari *accuracy*, *precision*, *recall* yang dihasilkan, pada pengujian kedua ini digunakan teknik *Cross Validation* menggunakan nilai  $K=10$ . Untuk pengujian ke dua pada parameter *Optimize Weight* (PSO) dilakukan perubahan nilai *population size* menjadi “10” dan nilai *inertia weight* menjadi “1.0” atau dapat dilihat pada gambar 16. Dibawah ini merupakan hasil *Confusion Matrix* pengujian metode *Naïve Bayes* dengan *Particle Swarm Optimization*.

### 1. Accuracy

Hasil *accuracy* dari pengujian algoritma *Naïve Bayes* dengan *Particle Swarm Optimization* sebesar 92,37%, yang bisa dilihat pada gambar 17.

accuracy: 92.37% +/- 4.74% (micro average: 92.38%)

	true Negatif	true Positif	class precision
pred. Negatif	134	7	95.04%
pred. Positif	16	145	90.06%
class recall	89.33%	95.39%	

Gambar 17 Hasil *accuracy* dari algoritma *Naïve Bayes* dengan PSO

### 2. Precision

Hasil *precision* dari pengujian algoritma *Naïve Bayes* dengan *Particle Swarm Optimization* sebesar 90,40%, yang bisa dilihat pada gambar 18.

precision: 90.40% +/- 6.87% (micro average: 90.06%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	134	7	95.04%
pred. Positif	16	145	90.06%
class recall	89.33%	95.39%	

Gambar 18 Hasil *precision* dari algoritma *Naïve Bayes* dengan PSO

### 3. Recall

Hasil *recall* dari pengujian algoritma *Naïve Bayes* dengan *Particle Swarm Optimization* sebesar 95,38%, yang bisa dilihat pada gambar 19.

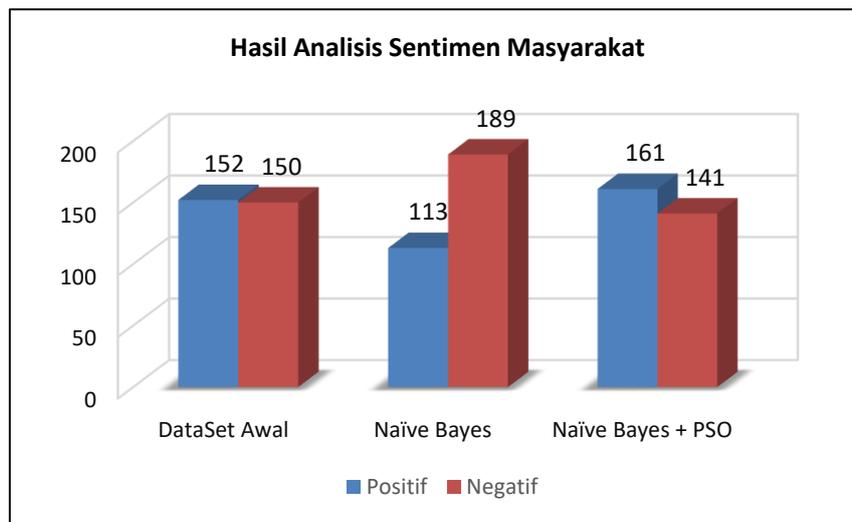
recall: 95.38% +/- 3.19% (micro average: 95.39%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	134	7	95.04%
pred. Positif	16	145	90.06%
class recall	89.33%	95.39%	

Gambar 19 Hasil recall dari algoritma Naive Bayes dengan PSO

### 3.2 Hasil klasifikasi sentimen masyarakat

Hasil klasifikasi setimen masyarakat terhadap perpanjangan kebijakan PPKM darurat yang dihasilkan berasal dari data *training* yang sebelumnya sudah dilakukan pengujian dengan 2 metode yang berbeda. Dibawah ini merupakan gambar grafik perbandingan hasil klasifikasi sentimen masyarakat terhadap perpanjangan kebijakan PPKM darurat.



Gambar 20 Grafik perbandingan hasil analisis sentimen masyarakat

Dapat dilihat pada gambar *grafik* diatas bahwa dari total 302 data *tweets* yang di uji pada proses pengujian pertama hasilnya 113 data *tweets* mengandung sentimen positif dan 189 data *tweets* mengandung sentimen negatif. Pada proses pengujian kedua hasilnya terdapat 161 data *tweets* mengandung sentimen positif dan 141 data *tweets* mengandung sentimen negatif.

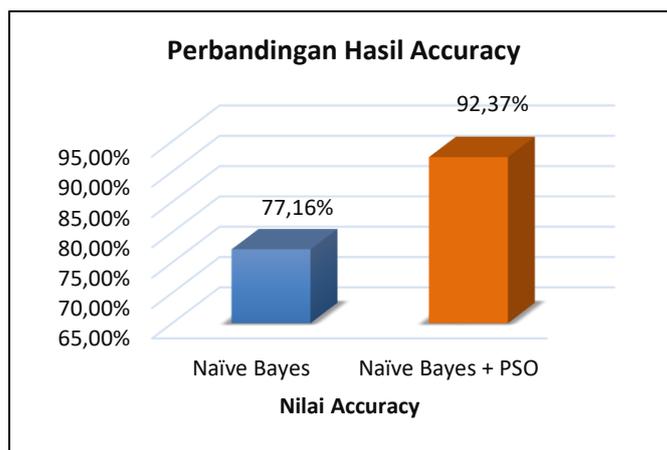
Dari gambar *grafik* diatas dapat diketahui bahwa hasil klasifikasi sentimen masyarakat dari pengujian pertama hasilnya lebih unggul *tweets* negatif dengan persentase 37,42% *tweets* mengandung sentimen positif dan 62,58% *tweets* mengandung sentimen negatif. Dari pengujian kedua hasilnya lebih unggul *tweets* positif dengan presentase 53,31% *tweets* mengandung sentimen positif dan 46,69% *tweets* mengandung sentimen negatif.

Table 2 Hasil analisis sentimen

Pengujian	Accuracy	Precision	Recall	Hasil Klasifikasi Sentimen		Keterangan	
				Positif	Negatif	Positif	Negatif
<i>Dataset</i>	-	-	-	152	150	50,33%	49,67%
<i>NB</i>	77,16%	87,33%	64,42%	113	189	37,42%	62,58%
<i>NB+PSO</i>	92,37%	90,40%	95,38%	161	141	53,31%	46,69%

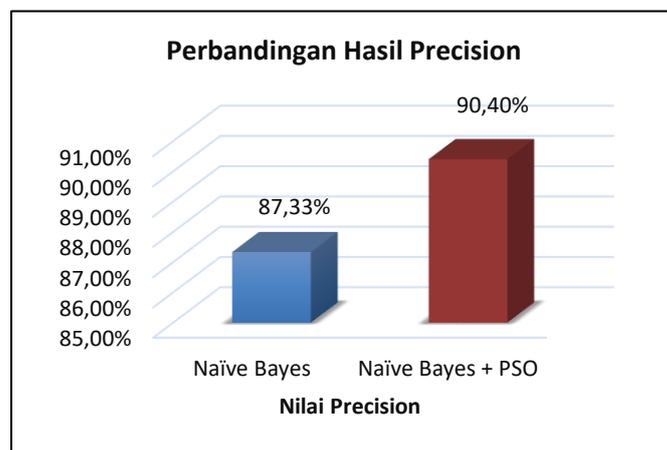
### 3.3 Pembahasan

Dibawah ini merupakan gambar grafik perbandingan hasil *accuracy*, *precision*, *recall* dari metode Naive Bayes tanpa *feature selection* PSO dan metode Naive Bayes dengan *feature selection* PSO.



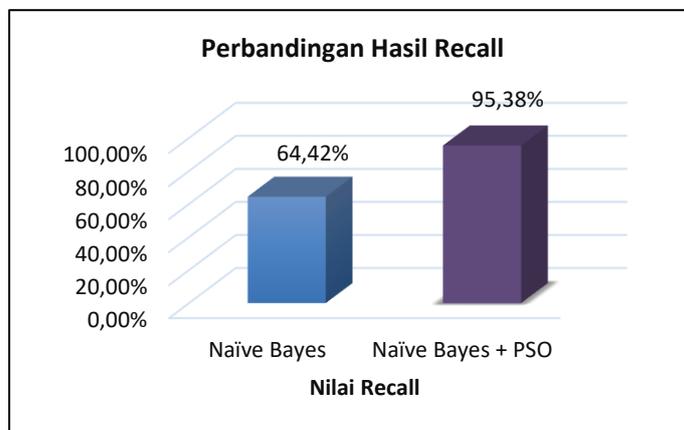
Gambar 21 Grafik perbandingan hasil *accuracy*

Dari hasil pengujian yang telah dilakukan antara pengujian pertama dan kedua, pada gambar 21 bahwa nilai *accuracy* yang didapatkan meningkat sebanyak 15,21% dari 77,16% menjadi 92,37%.



Gambar 22 Grafik perbandingan hasil *recall*

Dari hasil pengujian yang telah dilakukan antara pengujian pertama dan kedua, pada gambar 22 bahwa nilai *precision* yang didapatkan meningkat sebanyak 3,07% dari 87,33% menjadi 90,40%.



Gambar 23 Grafik perbandingan hasil *recall*

Dari hasil pengujian yang telah dilakukan antara pengujian pertama dan kedua, pada gambar 23 bahwa nilai *recall* yang didapatkan meningkat sebanyak 30,96% dari 64,42% menjadi 95,38%.

Table 3 Perbandingan hasil pengujian pertama dan kedua

	Pengujian Pertama	Pengujian Kedua	Keterangan	
	Naïve Bayes	NB+PSO	Peningkatan Nilai	Hasil
<b>Accuracy</b>	77,16%	92,37%	15,21%	Meningkat
<b>Precision</b>	87,33%	90,40%	3,07%	Meningkat
<b>Recall</b>	64,42%	95,38%	30,96%	Meningkat

#### 4. KESIMPULAN DAN SARAN

##### 4.1 Kesimpulan

Berdasarkan hasil pengujian yang sudah dilakukan dapat ditarik kesimpulan yang ada pada penelitian ini bahwa:

1. Pengujian yang dilakukan menggunakan metode *Naïve Bayes* dan *Particle Swarm Optimization* sebagai *feature selection* dalam menganalisis sentimen masyarakat mengenai kebijakan perpanjangan PPKM darurat pada social media *twitter* hasil klasifikasi yang didapat dari total 302 data *tweets*, 161 *tweets* mengandung sentimen positif dan 141 *tweets* mengandung sentimen negatif.
2. Dapat Diketahui bahwa 53,31% pengguna *twitter* setuju dan 46,69% pengguna *twitter* tidak setuju dengan perpanjangan kebijakan PPKM darurat yang di terapkan di beberapa daerah di Kepulauan Jawa dan Kepulauan Bali, yang memiliki status zona merah atau daerah yang memiliki risiko tinggi terhadap paparan kasus COVID-19.
3. Penggunaan *feature selection Particle Swarm Optimization* pada metode *Naïve Bayes* dalam menganalisis sentimen masyarakat mengenai kebijakan perpanjangan PPKM darurat pada social media *twitter* dapat meningkatkan nilai *accuracy*, *precision*, dan *recall*. Nilai *accuracy* yang didapatkan meningkat sebanyak 15,21% dari 77,16% menjadi 92,37%, nilai *precision* yang didapatkan meningkat sebanyak 3,07% dari 87,33% menjadi 90,40%, dan nilai *recall* yang didapatkan meningkat sebanyak 30,96% dari 64,42% menjadi 95,38%.

##### a. Saran

Adapun saran masukan dari penulis untuk penelitian selanjutnya dapat melakukan analisis sentimen dengan menggunakan metode klasifikasi atau teknik *feature selection* yang berbeda supaya dapat diketahui perbandingan nilai *accuracy*, *precision*, dan *recall* yang di hasilkan. Penelitian yang dilakukan tidak diimplementasikan pada suatu sistem, oleh karena itu dibutuhkan sebuah sistem analisis sentimen masyarakat berbasis website atau android.

#### DAFTAR PUSTAKA

- [1] 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," *Sist. J. Sist. Inf.*, vol. 10, no. 1, pp. 228–238, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [2] Y. Cahyono, "Analisis Sentiment pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency," *J. Inform. Univ. Pamulang*, vol. 2, no. 1, pp. 14–19, 2017, doi: 10.32493/informatika.v2i1.1500.
- [3] Humas, "Berlaku Mulai 12 Juli, PPKM Darurat Diberlakukan di 15 Kab/Kota Luar Jawa-Bali," Jul. 09, 2021. <https://setkab.go.id/berlaku-mulai-12-juli-ppkm-darurat-diberlakukan-di-15-kab-kota-luar-jawa-bali/> (accessed Sep. 20, 2021).

- [4] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.
- [5] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [6] E. Ermawati, "Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai," *Sist. J. Sist. Inf.*, vol. 8, no. 3, pp. 513–528, 2019, doi: 10.32520/stmsi.v8i3.576.
- [7] S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS DEHASEN BENGKULU)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [8] R. yana Yanis, "Sentiment Analysis of Bpjs Kesehatan Services To Smk Eklesia and Bina Insani Jailolo Teachers," *JUTEI J. Terap. Teknol. Inf.*, vol. 2, no. 2, pp. 113–122, 2018, doi: 10.21460/jutei.2018.22.105.
- [9] E. Putri and T. Setiadi, "Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes," *J. Sarj. Tek. Inform.*, vol. 2, no. 3, pp. 73–83, 2014, doi: 10.12928/jstie.v2i3.2877.
- [10] R. P. Sidiq, B. A. Dermawan, and Y. Umaidah, "Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," *J. Inform. Univ. Pamulang*, vol. 5, no. 3, pp. 356–363, 2020, doi: 10.32493/informatika.v5i3.6571.
- [11] 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, doi: 10.30591/jpit.v3i3.977.
- [12] K. J. Prayoga, A. Nugroho, and N. Wiyatno, "Komparasi feature selection particle swarm optimization (pso) dengan genetic algorithm (ga) terhadap algoritma naïve bayes pada analisis sentimen twitter," *Pros. Semin. Nas. Teknol. dan Sains*, no. September, pp. 299–314, 2019.
- [13] A. Taufik, "Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel," *J. Tek. Komput.*, vol. IV, no. 2, pp. 112–118, 2018, doi: 10.31294/jtk.v4i2.3461.
- [14] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, pp. 145–148, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [15] Y. Cahyono and S. Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic," *J. Inform. Univ. Pamulang*, vol. 4, no. 3, pp. 87–94, 2019, doi: 10.32493/informatika.v4i3.3186.