

KLASIFIKASI EMOSI PUBLIK TERHADAP LARANGAN PENGGUNAAN OBAT SIRUP MENGGUNAKAN ALGORITMA NAIVE BAYES

Adella Putri Riani[✉], Nina Sulistyowati, Taufik Ridwan, Apriade Voutama

Program Studi Sistem Informasi, Universitas Singaperbangsa, Karawang, Indonesia

Email: adella.putri19001@student.unsika.ac.id

DOI: <https://doi.org/10.46880/jmika.Vol7No2.pp325-339>

ABSTRACT

In October 2022 BPOM withdrew the circulation of syrup drugs and banned the public from using syrup drugs due to the increasing cases of kidney failure in children with a mortality rate of up to 59%. The phenomenon of mass death has caused psychological trauma that threatens. The purpose of this study was to find out how public comments on Twitter and Instagram social media regarding the prohibition of using syrup drugs by classifying emotions using the Naive Bayes algorithm with Particle Swarm Optimization (PSO) feature selection. The methodology used is the AI Project Cycle which consists of problem scoping, data acquisition, data exploration, modeling and evaluation. The amount of data in this study was 1213 comments with 381 surprised comments, 318 angry comments, 277 sad comments, 137 scared comments, and 100 happy comments. Happy and afraid comments have fewer comments than other emotions, so the imbalance dataset will be handled using SMOTE. The results of this study are to compare the classification results of the application of PSO and SMOTE to the Naive Bayes algorithm to determine the best model. Based on the classification results, the Naive Bayes, PSO, and SMOTE models produced the highest accuracy values of 76.48%, with a recall value of 76.47%, a precision of 76.20%, and a fmeasure of 75.62%.

Keyword: AI Project Cycle, Imbalance Dataset, Naive Bayes, PSO, SMOTE.

ABSTRAK

Pada bulan Oktober 2022 BPOM menarik peredaran obat sirup dan melarang masyarakat untuk menggunakan obat sirup buntut kasus meningkatnya penderita gagal ginjal pada anak dengan angka kematian yang mencapai 59%. Fenomena kematian massal telah menimbulkan trauma yang mengancam psikologis. Tujuan penelitian ini adalah untuk mengetahui bagaimana komentar publik pada media sosial twitter dan instagram mengenai larangan penggunaan obat sirup dengan melakukan klasifikasi emosi menggunakan algoritma Naive Bayes dengan seleksi fitur Particle Swarm Optimization (PSO). Metodologi yang digunakan adalah AI Project Cycle yang terdiri dari problem scoping, data acquisition, data exploration, modelling dan evaluation. Jumlah data dalam penelitian ini yaitu 1213 komentar dengan 381 komentar terkejut, 318 komentar marah, 277 komentar sedih, 137 komentar takut, dan 100 komentar senang. Komentar senang dan takut memiliki komentar lebih sedikit dibandingkan emosi lain, sehingga akan dilakukan penanganan imbalance dataset menggunakan SMOTE. Hasil penelitian ini adalah membandingkan hasil klasifikasi penerapan PSO dan SMOTE pada algoritma Naive Bayes untuk mengetahui model terbaik. Berdasarkan hasil klasifikasi, model Naive Bayes, PSO, dan SMOTE menghasilkan nilai accuracy tertinggi yaitu 76.48%, dengan nilai recall 76.47%, precision 76.20%, dan fmeasure 75.62%.

Kata Kunci: AI Project Cycle, Imbalance Dataset, Naive Bayes, PSO, SMOTE.

PENDAHULUAN

Pada bulan Oktober 2022, kasus Gagal Ginjal Akut pada Anak (GGAPA) di Indonesia menjadi bulanan masyarakat. Kasus ini memiliki angka kematian yang cukup tinggi dengan nilai *Case Fatality Rate* (CFR) mencapai 59% (Kementerian Kesehatan RI, 2022). Berdasarkan kelompok umur, kasus GGAPA paling banyak diderita oleh anak berusia 1-5 tahun yaitu sebanyak 190 kasus dan 130 diantaranya

meninggal dunia (Kementerian Kesehatan RI, 2022). Menteri Kesehatan Budi Gunadi Sadikin menyebutkan terdapat tiga zat pelarut pada obat sirup sebagai penyebab meningkatnya kasus gagal ginjal akut pada anak (Kompas, 2022). Pemerintah mengambil langkah cepat dengan melarang masyarakat untuk menggunakan obat sirup untuk sementara (Kementerian Kesehatan RI, 2022).

Fenomena kematian massal telah menimbulkan trauma, ketakutan, bahkan kecemasan yang mengancam fisik dan kesejahteraan psikologis publik (Abdullah, 2020). Larangan penggunaan obat sirup pada anak pun menimbulkan banyak opini dari publik di media sosial. Salah satunya yaitu *Twitter* dan *Instagram*. Berdasarkan survei yang dilakukan oleh Reuters Insititute dengan Universitas Oxford 68% masyarakat Indonesia mengakses berita melalui media sosial (Digital News Report, 2022). Berdasarkan hal tersebut, akan dilakukan penelitian mengenai klasifikasi emosi publik terhadap larangan penggunaan obat sirup pada anak di media sosial *twitter* dan *instagram*.

Emosi merupakan bagian dalam kecerdasan manusia yang berkaitan langsung dengan hati untuk membantu pengambilan keputusan dalam kehidupan sehari-hari (Akbar et al., 2021). Dengan melakukan klasifikasi emosi, maka dapat diketahui kumpulan pendapat dan emosi yang dirasakan publik saat menyampaikan perasaan mereka di media sosial. Menurut Paul Ekman (dalam Ainurrochman et al., 2020), emosi dasar manusia dapat dikategorikan menjadi enam kategori yaitu senang, sedih, marah, jijik, terkejut, dan takut. Namun dalam penelitian ini emosi akan dibagi menjadi lima kategori saja yaitu senang, sedih, marah, terkejut, dan takut. Hal tersebut dilakukan karena proses klasifikasi emosi jijik cukup sulit dan cenderung diklasifikasikan ke dalam emosi marah (Meilani et al., 2018).

Metodologi yang digunakan dalam penelitian ini yaitu *AI Project Cycle*. *AI Project Cycle* memiliki lima tahap yaitu *problem scoping*, *data acquisition*, *data exploration*, *modelling*, dan *evaluation*. Pada tahap *modelling*, diperlukan sebuah algoritma yang digunakan untuk membantu penemuan pola. Salah satu algoritma yang dapat digunakan yaitu algoritma *Naive Bayes*. Beberapa kelebihan yang dimiliki algoritma *Naive Bayes* yakni dapat menghasilkan performa baik pada data berkategori teks, proses eksekusi cukup cepat, serta atribut pada algoritma ini berdiri sendiri dan tidak bergantung dengan nilai-nilai pada kelas lain (Wibawa et al., 2018). Namun algoritma ini juga memiliki kelemahan yaitu dapat berakurasi rendah apabila memiliki fitur yang terlalu banyak (Nurlaela, 2020). Oleh karena itu, untuk mengatasi kelemahan pada algoritma *Naive Bayes* penelitian ini akan menggunakan metode seleksi fitur *Particle Swarm Optimization* (PSO).

Pada penelitian Oky Irnawati dan Kusmayanti Solecha dalam judul “Analisis Sentimen Ulasan Aplikasi Flip Menggunakan *Naive Bayes* dengan

Seleksi Fitur PSO” penggunaan PSO, dapat meningkatkan nilai akurasi dari 82% menjadi 88,24% (Irnawati & Solecha, 2022). Namun dalam proses klasifikasi terdapat kemungkinan data yang digunakan tidak seimbang (*imbalance dataset*) yang dapat menyebabkan model bias dan berkinerja buruk di kelas minoritas (Sabilla & Vista, 2021). Oleh karena itu, pada penelitian ini akan dilakukan penanganan terhadap data tidak seimbang menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE).

Pada penelitian yang dilakukan Berniawan, dll dalam klasifikasi sentimen terhadap Kemkominfo, penggunaan algoritma *Naive Bayes* dan SMOTE dapat meningkatkan nilai accuracy dari 65% menjadi 91% daripada algoritma *Naive Bayes* tanpa SMOTE (Berniawan et al., 2023). Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka dilakukan penelitian mengenai klasifikasi emosi publik terhadap larangan penggunaan obat sirup pada anak di media sosial *twitter* dan *instagram* menggunakan algoritma *Naive Bayes* dengan seleksi fitur PSO dan penanganan terhadap data tidak seimbang menggunakan SMOTE.

TINJAUAN PUSTAKA

Emosi

Secara bahasa emosi berasal dari bahasa latin yaitu “*Emovere*” yang berarti bergerak keluar (Asri & Chusniah, 2016). Menurut Stevenson (dalam Sarasati & Nurvia, 2021) emosi secara harfiah adalah setiap tindakan atau gejala pikiran, perasaan, gairah atau setiap keadaan pikiran yang intens/menggebu-gebu. Emosi adalah suatu kondisi yang disebabkan oleh keadaan khusus dan melibatkan ekspresi fisik (Fitriah, 2018). Menurut Paul Ekman (dalam Meilani et al., 2018) emosi dasar pada manusia dapat dibagi menjadi enam kategori yaitu:

- Senang, yaitu perasaan berhasil atau kemajuan dalam mencapai suatu tujuan
- Sedih, yaitu perasaan kehilangan atau kegagalan suatu tujuan
- Marah, yaitu perasaan kecewa atau kesal terhadap suatu peran atau tujuan
- Jijik, yaitu perasaan tidak nyaman dan ingin menghindari pada suatu objek
- Takut, yaitu perasaan mengerikan menghadapi sesuatu yang dianggap mengancam fisik atau batin
- Terkejut, yaitu emosi sesaat ketika menghadapi kejadian tidak terduga

Preprocessing Text

Preprocessing merupakan proses menghapus inkonsistensi data seperti data kosong (null), duplikasi

- data, dan data yang tidak relevan (Syarifah et al., 2022). Proses ini sangat penting dilakukan karena kehadiran data yang tidak konsisten dapat mempengaruhi kualitas datau kinerja model yang dihasilkan (Febriyani et al., 2021). *Preprocessing* memiliki beberapa tahapan yaitu:
- Convert emoji/emoticon*, yaitu proses untuk mengubah emoji atau emoticon dalam teks menjadi sebuah ekspresi yang digambarkan secara tekstual
 - Cleansing data*, yaitu proses untuk menghilangkan simbol-simbol atau tanda baca tidak penting pada teks
 - Case folding*, yaitu proses untuk menyamakan bentuk huruf dengan huruf kapital (*uppercase*) semua atau huruf kecil (*lowercase*) semua
 - Normalize*, yaitu proses untuk mengubah kata gaul dan singkatan menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI)
 - Convert Negation*, yaitu proses untuk menggabungkan kalimat yang mengandung negasi menjadi satu kata, seperti kata “tidak senang” menjadi “tidaksenang”
 - Tokenizing*, yaitu proses untuk memotong atau memecah kalimat menjadi penggalan kata tunggal
 - Filtering*, yaitu proses untuk menyeleksi atau menghilangkan kata yang tidak penting seperti kata depan, kata hubung atau biasa dikenal dengan istilah *stopwords*.
 - Stemming*, yaitu proses untuk mengubah kata yang memiliki imbuhan menjadi kata dasar atau menghapus imbuhan pada awal atau akhir kata.

Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik pemberian bobot pada setiap kata yang muncul dalam suatu dokumen untuk menentukan sebesar apa pentingnya sebuah kata mewakili isi dalam dokumen (Maulana et al., 2023). Perhitungan bobot TF-IDF dapat dilakukan dengan menghitung jumlah kemunculan kata dalam dokumen (TF) kemudian menghitung nilai invers dari jumlah dokumen yang mengandung sebuah kata tertentu (IDF) menggunakan persamaan (1) berikut

$$IDF_t = \log \frac{n}{df(t)} \tag{1}$$

Langkah terakhir pembobotan TF-IDF merupakan hasil perkalian TF dan IDF yang dapat dijelaskan pada persamaan (2)

$$W_t = TF_t \times IDF_t \tag{2}$$

Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE mensintetis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* (Sulistiyowati & Jajuli, 2020). Adapun langkah-langkah teknik SMOTE sebagai berikut:

- Pilih satu sampel secara acak dari kelas minoritas
- Tentukan jumlah tetangga (k) untuk menghitung jarak tetangga terdekat dengan sampel menggunakan rumus *Euclidean Distance* seperti pada persamaan (3)

$$d_{xy} = \sqrt{\sum_{k=1}^n (x_i - y_i)^2} \tag{3}$$

- Pilih salah satu dari k tetangga terdekat secara acak.
- Hasilkan data sampel sintetis baru menggunakan rumus pada persamaan (4)

$$P1' = P1 + rand(0,1)x(P2 - P1) \tag{4}$$

- Ulangi langkah ini hingga jumlah data pada kelas minoritas sama dengan jumlah data pada kelas mayoritas

Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan teknik klasifikasi dengan pendekatan *supervised* berdasarkan peluang atau kemungkinan suatu peristiwa terjadi yang disebut dengan probabilitas (Andriansyah et al., 2021). Adapun tahap perhitungan klasifikasi algoritma *Naive Bayes* sebagai berikut (Meilani et al., 2018):

- Menghitung probabilitas tiap kelas terhadap seluruh dokumen (*prior*) menggunakan persamaan (5) yaitu

$$P(C) = \frac{N_c}{N} \tag{5}$$

Keterangan:

- N_c = Jumlah kelas c pada seluruh dokumen
- N = Total seluruh dokumen

- Menghitung probabilitas tiap kata untuk setiap kelas menggunakan persamaan (6) yaitu

$$P(X_n|C) = \frac{\sum tf(X_n, d \in C) + \alpha}{\sum N_{d \in C} + V} \tag{6}$$

Keterangan:

$$\sum tf(X_n, d \in C) = \text{Jumlah bobot TF-IDF kata } X_n \text{ dari seluruh dokumen pada data latih} \quad f = \frac{1}{h + 0.001} \quad (8)$$

$\sum N_{d \in C}$ = Jumlah bobot seluruh term pada data latih dari kelas C
 V = Jumlah total kata pada data latih seluruh kelas

Keterangan:

h = Total nilai posisi pada partikel n

- c. Menghitung probabilitas tiap kata pada data uji untuk setiap kelas emosi menggunakan persamaan (7) yaitu

$$P(C|doc D) = P(X_1|C) \times P(X_n|C) \times P(C) \quad (7)$$

Keterangan:

$P(C)$ = Probabilitas prior kelas C
 X_n = Kata pada dokumen d ke- n
 $P(C|doc D)$ = Probabilitas suatu dokumen termasuk kelas C
 $P(X_n|C)$ = Probabilitas kata ke- n pada kelas C

- d. Menentukan kelas pada data uji yang memiliki nilai probabilitas tertinggi. Apabila data uji memiliki probabilitas pada kelas positif lebih tinggi daripada probabilitas pada kelas negatif, maka data uji termasuk ke dalam kelas positif dan sebaliknya.

Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan teknik optimasi dengan mengubah beberapa parameter yaitu menambah bobot atribut dan memilih atribut atau fitur yang digunakan (Hasan & Wahyudi, 2018). Optimasi fitur perlu dilakukan terhadap *data training* dengan memperkecil dimensi *data training* namun tetap mempertahankan akurasi model yang tinggi (Antonio et al., 2021). Setiap partikel memiliki fitur dan setiap fitur memiliki kecepatan (*velocity*) dan posisi (*position*) yang perlu di update pada setiap iterasi untuk mendapatkan nilai *gbest* (*global best*) dan *pbest* (*local best*). Secara garis besar, algoritma Particle Swarm Optimization (PSO) dapat dijabarkan sebagai berikut:

- Inisialisasi beberapa parameter seperti jumlah partikel (i), jumlah fitur (s), bobot inersia maksimal serta minimal (w), dan nilai konstanta (c)
- Inisialisasi kecepatan (v) awal partikel
- Inisialisasi posisi (x) awal partikel
- Menghitung nilai *fitness* pada setiap partikel untuk menentukan nilai *pbest* untuk setiap iterasi menggunakan persamaan (8) yaitu

- Mencari nilai *pbest* dengan membandingkan nilai *fitness* tertinggi pada iterasi saat ini dengan iterasi sebelumnya
- Mencari nilai *gbest* dengan membandingkan nilai *fitness* tertinggi pada *pbest*. Nilai *pbest* baru dengan *fitness* tertinggi akan menjadi nilai *gbest* baru
- Memperbarui nilai kecepatan (v) setiap partikel menggunakan persamaan (9) yaitu

$$V_{i,j}^{t+1} = wV_{i,j}^t + c_1r_1(x_{best_i}^t - x_i^t) + c_2r_2(g_{best_i}^t - x_i^t) \quad (9)$$

Keterangan:

$v_{i,j}^{t+1}$ = Kecepatan partikel i dimensi j pada iterasi selanjutnya
 $v_{i,j}^t$ = Kecepatan partikel i dimensi j pada iterasi saat ini
 w = Bobot inersia
 c_1, c_2 = Konstanta
 r_1, r_2 = Nilai acak antara 0 sampai 1
 $x_{best_{i,j}}^t$ = Nilai *pbest* dari partikel i dimensi j pada iterasi saat ini
 $g_{best_{i,j}}^t$ = Nilai *gbest* dari partikel i dimensi j pada iterasi saat ini
 $x_{i,j}^t$ = Posisi partikel i dimensi j pada iterasi saat ini

- Memperbarui nilai posisi (x) setiap partikel menggunakan persamaan (10) yaitu

$$X_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^t \quad (10)$$

Keterangan:

$x_{i,j}^{t+1}$ = Posisi partikel i dimensi j pada iterasi selanjutnya
 $x_{i,j}^t$ = Posisi partikel i dimensi j pada iterasi saat ini
 $v_{i,j}^t$ = Kecepatan partikel i dimensi j pada iterasi saat ini

- Ulangi langkah kelima hingga jumlah maksimum iterasi tercapai

Confusion Matrix

Confusion Matrix merupakan teknik evaluasi yang digunakan untuk mengukur performa pada suatu model klasifikasi (Saraswati et al., 2021). Tabel 1 merupakan *confusion matrix* berukuran n x n (Meilani et al., 2018)

Tabel 1. *Confusion Matrix*

	<i>Class 1</i>	...	<i>Class n</i>	<i>Total</i>
<i>Class 1</i>	<i>True Positive</i>	...	<i>Error</i>	<i>Actual Class 1</i>
...
<i>Class n</i>	<i>Error</i>	<i>Error</i>	<i>True Positive</i>	<i>Actual Class n</i>
<i>Total</i>	<i>Predicted Class 1</i>	...	<i>Predicted Class n</i>	

Rumus yang dapat digunakan untuk menghitung *accuracy*, *recall*, *precision*, dan *fmeasure* dapat dilakukan seperti pada persamaan (11), (12), (13) dan (14).

$$Accuracy = \frac{TP(Class 1) + \dots + TP(Class n)}{Actual(Class 1) + \dots + Actual(Class n)} \quad (11)$$

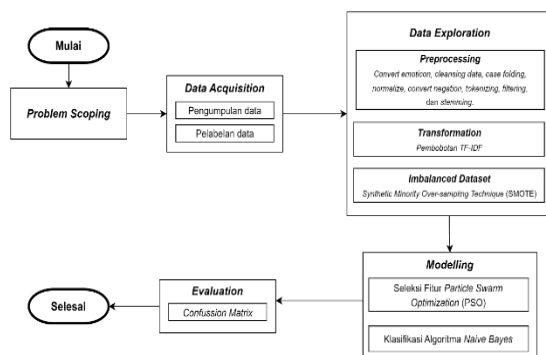
$$Recall(Class i) = \frac{TP(Class i)}{Actual(Class i)} \quad (12)$$

$$Precision(Class i) = \frac{TP(Class i)}{Predicted(Class i)} \quad (13)$$

$$Fmeasure = \frac{2 \times recall \times precision}{recall + precision} \quad (14)$$

METODE PENELITIAN

Penelitian ini menggunakan tahapan AI Project Cycle yang terdiri dari lima tahap yaitu *problem scoping*, *data acquisition*, *data exploration*, *modelling*, dan *evaluation*. Gambar 1 merupakan alur penelitian yang dilakukan



Gambar 1. Alur Penelitian

Problem Scoping

Permasalahan dalam penelitian ini, akan diidentifikasi menggunakan konsep *4W problem scoping* yaitu *who* (siapa), *what* (apa), *where* (dimana), dan *why* (kenapa) (CBSEskillededucation, 2022).

Data Acquisition

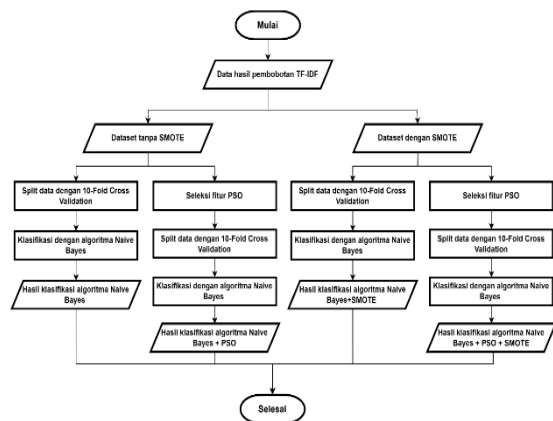
Data acquisition merupakan tahap menetapkan dan mengumpulkan data apa saja yang dibutuhkan dari database sesuai dengan kebutuhan analisis. Tahap *data acquisition* ini meliputi tahap pengumpulan data dan pelabelan data.

Data Exploration

Data exploration merupakan tahap untuk memahami karakteristik data yang telah didapatkan serta membersihkan data agar lebih mudah diolah (Pratama et al., 2021). *Data exploration* yang terdiri dari tiga tahap yaitu *preprocessing*, *transformation* dengan pembobotan TF-IDF, dan penanganan *imbalance dataset* dengan SMOTE.

Modelling

Modelling merupakan tahap membangun dan mengembangkan model dengan menerapkan teknik atau algoritma pembelajaran mesin (*machine learning*) untuk pencarian pola tersembunyi (Lagman et al., 2020). Pembangunan model dilakukan menggunakan algoritma *Naive Bayes*. Selain itu, akan dilakukan optimasi pada klasifikasi algoritma *Naive Bayes* menggunakan seleksi fitur PSO. Berikut merupakan tahap *modelling* yang dapat dilihat pada gambar 2.



Gambar 2. Tahap Modelling

Evaluation

Evaluation merupakan proses mengukur kinerja model yang dihasilkan dari algoritma pembelajaran mesin yang telah dipilih apakah sudah sesuai dengan tujuan yang ditetapkan (Asih et al., 2022). Teknik evaluasi yang digunakan untuk mengukur performa

algoritma *Naive Bayes* dan seleksi fitur PSO adalah *confusion matrix*.

HASIL DAN PEMBAHASAN

Hasil penelitian ini adalah melakukan analisa perbandingan performa klasifikasi algoritma *Naive Bayes* dengan seleksi fitur PSO dan SMOTE. Setelah itu, akan divisualisasikan ke dalam *word cloud* untuk menampilkan istilah yang paling sering digunakan pada *dataset*.

Problem Scoping

Hasil dari tahap ini adalah mendefinisikan lingkup permasalahan dalam penelitian ini menggunakan konsep *4W problem scoping* seperti di bawah ini.

- Who:** Siapa saja yang terlibat dalam permasalahan ini?
Masyarakat yang mengungkapkan pendapat di *twitter* dan *instagram* mengenai larangan penggunaan obat sirup serta BPOM selaku pemberi izin dan pengawas peredaran obat sirup yang dijual bebas di masyarakat.
- What:** Apa permasalahan yang terjadi?
Kebijakan yang dilakukan BPOM dengan menarik penjualan obat sirup sementara dan mencabut izin edar beberapa produk obat sirup menjadi ramai diperbincangkan oleh publik di *twitter* dan *instagram*. BPOM menjadi *trending topic twitter* pada tanggal 20 Oktober 2022 saat mengumumkan penarikan 5 obat sirup (Viva, 2022).
- Where:** Dimana permasalahan ini terjadi?
Melalui *twitter* dan *instagram*, masyarakat dapat menyampaikan keluhan atau aspirasi kepada lembaga pemerintah, pejabat pemerintah, atau sesama masyarakat sipil. *Twitter* dan *instagram* dipilih menjadi sumber data dalam penelitian ini karena menjadi salah satu media sosial populer di Indonesia dengan 18,45 juta pengguna *twitter* dan 97,38 juta pengguna *instagram* (Digital News Report, 2022).
- Why:** Mengapa permasalahan ini perlu diselesaikan dan apa manfaatnya?
Angka kematian pada penderita gagal ginjal akut ini cukup tinggi yang diduga disebabkan oleh penggunaan obat sirup pada anak menjadi ramai diperbincangkan oleh publik. Fenomena kematian massal telah menimbulkan trauma yang mengancam psikologis publik (Abdullah, 2020). Sehingga akan dilakukan klasifikasi emosi publik terhadap larangan penggunaan obat sirup pada anak pada media sosial *twitter* dan *instagram*.

Data Acquisition

Hasil dari tahapan *data acquisition* ini meliputi tahap pengumpulan data awal dan pelabelan data menjadi lima kategori emosi yaitu senang, sedih, marah, takut, dan terkejut.

Pengumpulan Data

Hasil dari tahap pengumpulan data yaitu memperoleh komentar publik di media sosial *twitter* dan *instagram* yang berkaitan dengan larangan penggunaan obat sirup pada anak dari tanggal 1 November sampai dengan 30 November 2022. Komentar pada media sosial *twitter* diambil menggunakan operator *search twitter* pada *software RapidMiner* dengan kata kunci “obat sirup”.

Komentar pada media sosial *instagram* diambil menggunakan *tools web scraper* pada 13 unggahan berita mengenai larangan penggunaan obat sirup yang terdiri dari 8 unggahan akun @detikcom, 2 unggahan akun @kompascom, dan 3 unggahan akun @cnnindonesia pada unggahan mulai tanggal 1 November 2022 sampai dengan 30 November 2022. Akun berita tersebut merupakan 3 media daring dengan konsumsi terbesar di Indonesia (Digital News Report, 2022).

Dari data komentar yang telah terkumpul akan dilakukan proses *filtering* dengan menghapus komentar duplikat. Berikut merupakan jumlah hasil pengumpulan data awal dan setelah proses *filtering* yang dapat dilihat pada tabel 2.

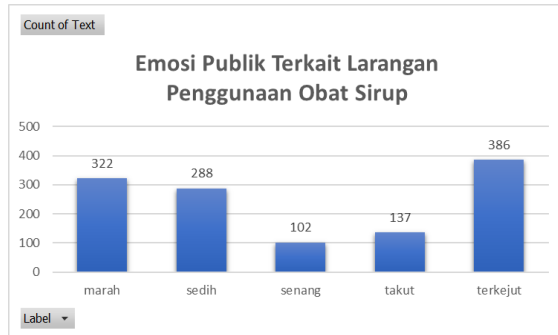
Tabel 2. Jumlah Hasil Pengumpulan Data Awal dan Hasil *Filtering*

Kategori	Data Awal	Data Hasil <i>Filtering</i>
<i>Twitter</i>	8.855	2.155
<i>Instagram</i>	2.585	2.558
Campuran	11.440	4.713

Pelabelan Data

Hasil dari tahap pelabelan data adalah memberikan label pada komentar publik di media sosial *twitter* dan *instagram* yang dilakukan secara manual oleh penulis dan divalidasi oleh ahli Bahasa Indonesia. Komentar yang memiliki berbagai emosi akan diberi label sesuai dengan emosi yang paling dominan. Kemudian komentar yang tidak memiliki emosi dan tidak berberhubungan dengan larangan penggunaan obat sirup akan diberi label bukan emosi dan dihapus. Sehingga jumlah data yang akan digunakan dalam penelitian ini adalah 1235 komentar yang terdiri dari 386 komentar terkejut, 322 komentar

marah, 288 komentar sedih, 137 komentar takut, dan 102 komentar senang.



Gambar 3. Perbandingan Jumlah Label Emosi

Tabel 3 merupakan data sampel komentar yang telah diberi label secara manual oleh penulis dan sudah divalidasi oleh ahli bahasa Indonesia untuk menentukan kelas emosi pada data komentar penelitian ini.

Tabel 3. Data Komentar dan Label

Text	Label
@PartaiSocmed @BPOM_RI ketuanya saja mak lampir. pantas gak berani ngomong yang benar. vaksin yang matiin orang indon. bukan sirup obat	Marah
Gue marah banget abis baca berita masalah obat sirup anak yg tercemar itu, lari dari tanggung jawab??	Marah
Huhu obat sirup ditarik semua dari apotik. aku lg batuk sedih bgtt enakan pake obat batuk sirup	Sedih
obat sirup kesukaan ku :(https://t.co/qZlnhuRXRG	Sedih
alhamdulillah akhirnya ozen sirup dinyatakan aman sm bpom jd ga khawatir lagi aku kalo main sama ayang alergi kambuh tuh, obat sirupnya adaa???? https://t.co/30HqKcxhe6	Senang
Waa, Terima kasih infonyaaa	Senang
Aduh yg udh terlanjur bagaimana dong?	Takut
Percuma org ud parno bro. Apalagi kasiin ke anak	Takut
Aneh aja dah ribuan taun baru ketauan 😞	Terkejut
Bapak suami waktu sakit lambung sulcralfate berboto12 diminum 🤔🤔🤔 @amiefahmie	Terkejut

Data Exploration

Hasil dari tahapan *data exploration* meliputi *preprocessing*, *data transformation* dengan pembobotan TF-IDF, dan penanganan *imbalance dataset* dengan SMOTE.

Preprocessing

Tahap *preprocessing* terdiri dari delapan tahap yaitu *convert emoji/emoticon*, *cleansing data*, *case folding*, *normalize*, *convert negation*, *tokenizing*, *filtering*, dan *stemming*. Tabel 4 merupakan contoh tahapan sebelum dan setelah *preprocessing*.

Tabel 4. Hasil Sebelum dan Setelah Preprocessing

Tahapan Preprocessing	Hasil
Sebelum Preprocessing	Bayar pajak cuma bayarin orang' ga guna 🤔
Convert Emoticon	Bayar pajak cuma bayarin orang' ga guna emuntah
Cleansing Data	Bayar pajak cuma bayarin orang ga guna emuntah
Case Folding	bayar pajak cuma bayarin orang ga guna emuntah
Normalize	bayar pajak hanya bayarin orang tidak guna emuntah
Convert Negation	bayar pajak hanya bayarin orang tidakguna emuntah
Tokenizing	'bayar', 'pajak', 'hanya', 'bayarin', 'orang', 'tidakguna', 'emuntah'
Filtering	'bayar', 'pajak', 'bayarin', 'orang', 'tidakguna', 'emuntah'
Stemming	'bayar', 'pajak', 'bayar', 'orang', 'tidakguna', 'emuntah'

Transformation

Hasil tahapan *transformation* yaitu merubah data teks ke dalam vektor numerik (Rianti et al., 2021). Proses pembobotan dilakukan menggunakan TF-IDF setelah data melalui tahap *preprocessing*. Tabel 5 merupakan data sampel yang digunakan untuk perhitungan TF-IDF.

Tabel 5. Data Sampel Perhitungan TF-IDF

No	Data Sampel
1	mundur percuma gaji bpom
2	sedih anak minum melulu esedih esedih
3	terima kasih informasi
4	percuma orang sudah parno apalagi beri anak
5	waduh kok bisa

Setiap baris pada tabel 5 mengandung beberapa kata dan setiap kata memiliki bobot. Berikut merupakan tahap perhitungan manual bobot kata menggunakan TF-IDF.

- Menghitung jumlah kemunculan kata dalam dokumen (TF). Hasil perhitungan TF disajikan pada tabel 6.
- Menghitung nilai invers dari jumlah dokumen yang mengandung sebuah kata tertentu (IDF) menggunakan persamaan (1) yaitu

$$idf_{mundur} = \log\left(\frac{5}{1}\right) = 0.698$$

$$idf_{percuma} = \log\left(\frac{5}{2}\right) = 0.397$$

Nilai D merupakan total seluruh dokumen dimana berdasarkan data pada tabel 5 total dokumen yang digunakan adalah 5 dokumen. Proses perhitungan dilakukan untuk seluruh term atau kata. Hasil perhitungan IDF disajikan pada tabel 6.

Tabel 6. Hasil Perhitungan TF-IDF Pada Data Sampel

Term	TF					DF	IDF
	D1	D2	D3	D4	D5		
mundur	1	0	0	0	0	1	0.698
percuma	1	0	0	1	0	2	0.397
gaji	1	0	0	0	0	1	0.698
bpom	1	0	0	0	0	1	0.698
sedih	0	1	0	0	0	1	0.698
anak	0	1	0	1	0	2	0.397
minum	0	1	0	0	0	1	0.698
melulu	0	1	0	0	0	1	0.698
esedih	0	2	0	0	0	2	0.397
terima	0	0	1	0	0	1	0.698
kasih	0	0	1	0	0	1	0.698
informasi	0	0	1	0	0	1	0.698
orang	0	0	0	1	0	1	0.698
sudah	0	0	0	1	0	1	0.698
parno	0	0	0	1	0	1	0.698
apalagi	0	0	0	1	0	1	0.698
beri	0	0	0	1	0	1	0.698
waduh	0	0	0	0	1	1	0.698
kok	0	0	0	0	1	1	0.698
bisa	0	0	0	0	1	1	0.698

- Menghitung bobot TF-IDF menggunakan persamaan (2) yaitu

$$W_{mundur(D_1)} = 1 \times 0.698 = 0.698$$

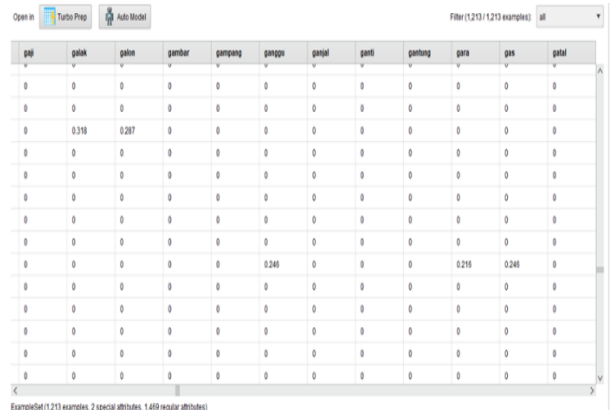
$$W_{mundur(D_2)} = 0 \times 0.698 = 0$$

Hasil perhitungan TF-IDF setiap kata pada data sampel disajikan pada tabel 7.

Tabel 7. Hasil Perhitungan TF-IDF Pada Data Sampel

Term	TF-IDF				
	D1	D2	D3	D4	D5
mundur	0.698	0	0	0	0
percuma	0.397	0	0	0.397	0
gaji	0.698	0	0	0	0
bpom	0.698	0	0	0	0
sedih	0	0.698	0	0	0
anak	0	0.397	0	0.397	0
minum	0	0.698	0	0	0
melulu	0	0.698	0	0	0
esedih	0	0.794	0	0	0
terima	0	0	0.698	0	0
kasih	0	0	0.698	0	0
informasi	0	0	0.698	0	0
orang	0	0	0	0.698	0
sudah	0	0	0	0.698	0
parno	0	0	0	0.698	0
apalagi	0	0	0	0.698	0
beri	0	0	0	0.698	0
waduh	0	0	0	0	0.698
kok	0	0	0	0	0.698
bisa	0	0	0	0	0.698

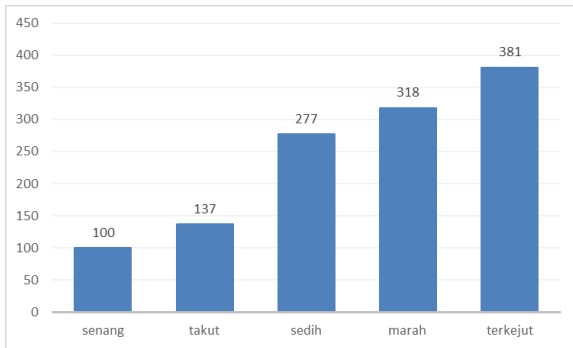
Gambar 4 merupakan hasil perhitungan TF-IDF menggunakan aplikasi *RapidMiner*. Berdasarkan gambar 4 terbentuk matriks berukuran 1213 x 1469, dimana terdapat 1213 dokumen dan 1469 term atau kata.



Gambar 4. Hasil perhitungan TF-IDF menggunakan aplikasi *RapidMiner*

Imbalance Dataset

Jumlah data komentar publik pada penelitian ini memiliki masalah ketidakseimbangan pada *dataset*. Dimana jumlah komentar pada kelas senang dan takut memiliki jumlah yang jauh lebih sedikit dibandingkan dengan kelas lainnya yaitu kelas sedih, marah dan terkejut seperti yang dapat dilihat pada gambar 5.



Gambar 5. Imbalance Dataset Komentar Publik

Oleh karena itu, pada tahap ini akan dilakukan penanganan terhadap data yang tidak seimbang menggunakan metode SMOTE dan diharapkan dapat meningkatkan performa algoritma klasifikasi. Berikut contoh perhitungan SMOTE dengan dua data sampel yang dapat dilihat pada tabel 8.

Tabel 8. Data Sampel Perhitungan SMOTE

	takut	beli	obat	sirup
P1	0.891	0.322	0.057	0.154
P2	0.9	0	0	0

P1 merupakan sampel yang akan di resampling dan P2 merupakan salah satu tetangga terdekat dari lima tetangga lainnya. Adapun rumus menghitung jarak antara P1 dan P2 yaitu menggunakan rumus *Euclidean Distance* seperti pada persamaan (3) yaitu

$$\begin{aligned}
 d &= \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \\
 &= \sqrt{(0.891 - 0.9)^2 + (0.322 - 0)^2} \\
 &\quad + (0.057 - 0)^2 + (0.154 - 0)^2 \\
 &= 0.361
 \end{aligned}$$

Setelah mendapat jarak tetangga paling dekat yaitu P2, tahap selanjutnya adalah melakukan resampling data P1 menjadi data baru (P1') dengan persamaan (4), dimana $\text{rand}(0,1)$ adalah nilai *random* antara angka 0 sampai dengan 1 (dalam contoh ini nilai *random* yang akan digunakan adalah 0.1).

$$\begin{aligned}
 P2(x_1) - P1(x_1) &= 0.9 - 0.891 = 0.009 \\
 P2(x_2) - P1(x_2) &= 0 - 0.322 = -0.322 \\
 P2(x_3) - P1(x_3) &= 0 - 0.057 = -0.057 \\
 P2(x_4) - P1(x_4) &= 0 - 0.154 = -0.154
 \end{aligned}$$

$$\begin{aligned}
 P1' &= (0.891, 0.322, 0.057, 0.154) + 0.1 \\
 &\quad \times (0.009, -0.322, -0.057, -0.154) \\
 &= (0.891, 0.289, 0.513, 0.138)
 \end{aligned}$$

Langkah tersebut akan terus diulang hingga jumlah data pada kelas minor memiliki jumlah yang sama dengan kelas mayor. Nilai P1' merupakan hasil resampling data baru dari perhitungan menggunakan teknik SMOTE seperti terlihat pada tabel 9.

Tabel 9. Hasil Perhitungan SMOTE pada Data Sampel

	takut	beli	obat	sirup
P1	0.891	0.322	0.057	0.154
P2	0.9	0	0	0
P1'	0.891	0.289	0.513	0.138

Hasil penyeimbangan data dengan SMOTE disajikan pada tabel 10.

Tabel 10. Hasil Penyeimbangan Data

Kelas	Jumlah Data Asli	Jumlah Data Hasil SMOTE
Terkejut	381	381
Marah	318	381
Sedih	277	381
Takut	137	381
Senang	100	381

Pada tabel 10 jumlah data emosi senang sebelum diresampling berjumlah 100 data, setelah di resampling maka emosi senang berjumlah 381 data yakni setara dengan kelas mayoritas yaitu emosi terkejut. Begitupun dengan kelas emosi lain, jumlah data akan disamakan sesuai dengan jumlah data pada kelas terkejut yaitu 381 data.

Modelling

Hasil dari tahapan *modelling* yaitu membangun model dengan menerapkan teknik klasifikasi menggunakan algoritma *Naive Bayes* dan klasifikasi algoritma *Naive Bayes* yang dioptimasi menggunakan seleksi fitur PSO. Pengujian setiap metode dilakukan menggunakan *dataset* berbeda, yaitu pada *dataset* yang sudah diolah dengan SMOTE dan *dataset* yang tidak diolah dengan SMOTE.

Pembagian data uji dan data latih dilakukan dengan menggunakan teknik *10-Fold Cross Validation*. Tabel 11 merupakan data sampel yang digunakan untuk perhitungan manual klasifikasi dengan algoritma *Naive Bayes*.

Tabel 11. Data Sampel Perhitungan Algoritma *Naive Bayes*

Data Sampel	Kelas
mundur percuma gaji bpom	Marah
sedih anak minum mulu esedih esedih	Sedih
terima kasih info	Senang
percuma orang sudah parno apalagi beri anak	Takut
waduh kok bisa	Terkejut

Tahap perhitungan manual klasifikasi menggunakan algoritma *Naive Bayes* adalah sebagai berikut.

- a. Menghitung probabilitas tiap emosi terhadap seluruh emosi (*prior*) menggunakan persamaan (5) sebagai berikut:

$$P(\text{marah}) = \frac{1}{5} = 0.2$$

$$P(\text{sedih}) = \frac{1}{5} = 0.2$$

$$P(\text{senang}) = \frac{1}{5} = 0.2$$

$$P(\text{takut}) = \frac{1}{5} = 0.2$$

$$P(\text{terkejut}) = \frac{1}{5} = 0.2$$

- b. Menghitung probabilitas tiap kata untuk setiap kelas emosi menggunakan persamaan (6). Nilai $\sum N_{dec}$ didapatkan sesuai dengan bobot setiap kata hasil perhitungan TF-IDF pada tabel 5 sebagai berikut:

$$\begin{aligned} \text{marah} &= 0.698 + 0.698 + 0.698 + 0.698 \\ &= 2.792 \end{aligned}$$

$$\begin{aligned} \text{sedih} &= 0.698 + 0.397 + 0.698 + 0.698 \\ &\quad + 0.794 = 3.285 \end{aligned}$$

$$\text{senang} = 0.698 + 0.698 + 0.698 = 2.094$$

$$\begin{aligned} \text{takut} &= 0.397 + 0.698 + 0.698 + 0.698 \\ &\quad + 0.698 + 0.698 + 0.397 = 4.284 \\ \text{terkejut} &= 0.698 + 0.698 + 0.698 = 2.094 \end{aligned}$$

Sehingga probabilitas tiap kata untuk setiap kelas emosi adalah sebagai berikut:

$$P(\text{mundur}|\text{marah}) = \frac{0.698 + 1}{2.792 + 20} = 0.074$$

$$P(\text{percuma}|\text{marah}) = \frac{0.397 + 1}{2.792 + 20} = 0.061$$

Tabel 12 merupakan hasil perhitungan probabilitas setiap kata untuk setiap kelas emosi pada data sampel.

Tabel 12. Probabilitas Setiap Kata dan Kelas Emosi

$P(X_n C)$	Nilai P	$P(X_n C)$	Nilai P
P(mundur marah)	0.074	P(waduh marah)	0.043
P(mundur sedih)	0.042	P(waduh sedih)	0.042
P(mundur senang)	0.045	P(waduh senang)	0.045
P(mundur takut)	0.041	P(waduh takut)	0.041
P(mundur terkejut)	0.045	P(waduh terkejut)	0.076
P(percuma marah)	0.061	P(kok marah)	0.043
P(percuma sedih)	0.042	P(kok sedih)	0.042
P(percuma senang)	0.045	P(kok senang)	0.045
P(percuma takut)	0.057	P(kok takut)	0.041
P(percuma terkejut)	0.045	P(kok terkejut)	0.076
P(bpom marah)	0.074	P(bisa marah)	0.043
P(bpom sedih)	0.042	P(bisa sedih)	0.042
P(bpom senang)	0.045	P(bisa senang)	0.045
P(bpom takut)	0.041	P(bisa takut)	0.041
P(bpom terkejut)	0.045	P(bisa terkejut)	0.076
...
P(lolos marah)	0.043	P(lolos takut)	0.041
P(lolos sedih)	0.042	P(lolos terkejut)	0.045
P(lolos senang)	0.045		

- c. Menghitung probabilitas tiap kata pada data uji untuk setiap kelas emosi menggunakan persamaan (7).

Data latih yang digunakan dalam contoh perhitungan algoritma *Naive Bayes* ini adalah “kok bisa lolos bpom” dengan perhitungan probabilitas setiap kelas emosi sebagai berikut:

$$P(\text{marah}|\text{kok bisa lolos bpom}) = 0.043 \times 0.043 \times 0.043 \times 0.074 \times 2.792 = 1.64 \times 10^{-5}$$

$$P(\text{sedih}|\text{kok bisa lolos bpom}) = 0.042 \times 0.042 \times 0.042 \times 0.042 \times 3.285 = 1.02 \times 10^{-5}$$

$$P(\text{senang}|\text{kok bisa lolos bpom}) = 0.045 \times 0.045 \times 0.045 \times 0.045 \times 2.094 = 0.85 \times 10^{-5}$$

$$P(\text{takut}|\text{kok bisa lolos bpom}) = 0.041 \times 0.041 \times 0.041 \times 0.041 \times 4.284 = 1.21 \times 10^{-5}$$

$$P(\text{terkejut}|\text{kok bisa lolos bpom}) = 0.076 \times 0.076 \times 0.045 \times 0.045 \times 2.094 = 2.44 \times 10^{-5}$$

- d. Menentukan kelas emosi pada data uji yang memiliki nilai probabilitas tertinggi. Berdasarkan perhitungan probabilitas pada langkah ketiga, kelas terkejut memiliki probabilitas tertinggi yaitu 2.44×10^{-5} . Sehingga, komentar “kok bisa lolos bpom” termasuk dalam kelas terkejut.

Setelah proses klasifikasi algoritma *Naive Bayes*, dilakukan optimasi menggunakan seleksi fitur PSO. Pada penelitian ini jumlah iterasi yang dilakukan adalah 100 kali iterasi. *Tools* yang digunakan untuk membangun model klasifikasi dalam penelitian ini adalah *RapidMiner*. Adapun proses perhitungan manual seleksi fitur PSO sebagai berikut:

- a. Inisialisasi parameter
 Jumlah partikel: $i=5$
 Jumlah fitur: $s=7$
 Bobot Inersia: $w_{\max}=0.9$ dan $w_{\min}=0.3$
 Nilai konstanta: $c_1=0.5$ dan $c_2=0.5$

Tabel 13 merupakan data sampel yang digunakan dalam perhitungan manual PSO dengan 5 partikel.

Tabel 13. Inisialisasi Partikel

P	X1 beli	X2 bisa	X3 kok	X4 lolos	X5 obat
P1	0.322	0	0	0	0.057
P2	0	0.698	0.698	0.698	0
P3	0.289	0	0	0	0.513
P4	0	0.153	0.568	0.632	0.397
P5	0	0	0.544	0.452	0

- b. Inisialisasi kecepatan (v) awal partikel secara acak. Tabel 14 merupakan nilai kecepatan awal setiap partikel.

Tabel 14. Inisialisasi Kecepatan (v)

P	X1	X2	X3	X4	X5
P1	0.032	0	0	0	0.005
P2	0	0.069	0.069	0.069	0
P3	0.028	0	0	0	0.051
P4	0	0.015	0.056	0.063	0.039
P5	0	0	0.054	0.045	0

- c. Inisialisasi posisi (x) awal partikel secara acak. Tabel 15 merupakan nilai posisi awal setiap partikel.

Tabel 15. Inisialisasi Posisi (x)

P	X1	X2	X3	X4	X5
P1	0.35	0	0	0	0.06
P2	0	0.76	0.76	0.76	0
P3	0.31	0	0	0	0.56
P4	0	0.16	0.62	0.69	0.43
P5	0	0	0.59	0.49	0

- d. Menghitung nilai *fitness* pada setiap partikel untuk menentukan nilai *pbest* untuk setiap iterasi menggunakan persamaan (8). Perhitungan untuk mencari nilai *fitness* setiap partikel adalah sebagai berikut:

$$f[1] = \frac{1}{0.41 + 0.001} = 2.433$$

$$f[2] = \frac{1}{2.28 + 0.001} = 0.438$$

$$f[3] = \frac{1}{0.87 + 0.001} = 1.148$$

$$f[4] = \frac{1}{1.9 + 0.001} = 0.526$$

$$f[5] = \frac{1}{1.08 + 0.001} = 0.925$$

- e. Mencari nilai *pbest* dengan membandingkan nilai *fitness* tertinggi pada iterasi saat ini dengan iterasi sebelumnya. Karena saat ini merupakan iterasi ke-0, maka nilai *pbest* akan disamakan dengan nilai posisi awal partikel seperti yang terlihat pada tabel 16.

Tabel 16. Nilai *pbest*

P	X1	X2	X3	X4	X5	fitness
P1	0.35	0	0	0	0.06	2.433
P2	0	0.76	0.76	0.76	0	0.438
P3	0.31	0	0	0	0.56	1.148
P4	0	0.16	0.62	0.69	0.43	0.526
P5	0	0	0.59	0.49	0	0.925

- f. Mencari nilai *gbest* dengan membandingkan nilai *fitness* tertinggi pada *pbest*.

Pada iterasi ke-0, nilai *pbest* tertinggi terletak pada partikel 1. Sehingga nilai *gbest* adalah 2.433 seperti terlihat pada tabel 17.

Tabel 17. Nilai *gbest*

P	X1	X2	X3	X4	X5	fitness
P1	0.35	0	0	0	0.06	2.433

- g. Memperbarui nilai kecepatan (*v*) setiap partikel menggunakan persamaan (9)

Perhitungan untuk memperbarui nilai kecepatan (*v*) adalah sebagai berikut:

$$V_1^1 = 0.5 \times 0.032 + 0.5 \times 0.3(0,35 - 0,35) + 0.5 \times 0.2(0.35 - 0,35) = 0.016$$

$$V_2^1 = 0.5 \times 0 + 0.5 \times 0.3(0 - 0) + 0.5 \times 0.2(0 - 0) = 0$$

$$V_3^1 = 0.5 \times 0 + 0.5 \times 0.3(0 - 0) + 0.5 \times 0.2(0 - 0) = 0$$

Perhitungan untuk memperbarui nilai kecepatan dilakukan pada setiap partikel. Tabel 18 merupakan hasil nilai kecepatan yang telah diperbarui.

Tabel 18. Hasil Update Kecepatan (*v*)

P	X1	X2	X3	X4	X5
P1	0.016	0	0	0	0.002
P2	0.035	-0.041	-0.041	-0.041	0.006
P3	0.018	0	0	0	-0.024
P4	0.035	-0.008	-0.034	-0.037	-0.017
P5	0.035	0	-0.032	-0.026	0.006

- h. Memperbarui nilai posisi (*x*) setiap partikel menggunakan persamaan (10)

Perhitungan untuk memperbarui nilai posisi (*x*) adalah sebagai berikut:

$$X_1^1 = 0.35 + 0.016 = 0.366$$

$$X_2^1 = 0 + 0 = 0$$

$$X_3^1 = 0 + 0 = 0$$

Tabel 19 merupakan hasil nilai posisi yang telah diperbarui pada setiap partikel

Tabel 19. Hasil Update Posisi (*x*)

P	X1	X2	X3	X4	X5
P1	0.366	0	0	0	0.062
P2	0.035	0.718	0.718	0.718	0.006
P3	0.328	0	0	0	0.535
P4	0.035	0.151	0.586	0.652	0.412
P5	0.035	0	0.558	0.463	0.006

- i. Menghitung nilai *fitness* setiap partikel dengan nilai kecepatan dan posisi yang telah diperbarui menggunakan persamaan (8) sebagai berikut:

$$f[1] = \frac{1}{0.428 + 0.001} = 2.331$$

$$f[2] = \frac{1}{2.196 + 0.001} = 0.455$$

$$f[3] = \frac{1}{0.863 + 0.001} = 1.157$$

$$f[4] = \frac{1}{1.837 + 0.001} = 0.544$$

$$f[5] = \frac{1}{1.062 + 0.001} = 0.940$$

Tabel 20 merupakan hasil nilai *fitness* dengan nilai kecepatan dan posisi yang telah diperbarui pada setiap partikel.

Tabel 20. Perbandingan Nilai *Fitness* Lama dan Baru

P	<i>Fitness</i> Lama	<i>Fitness</i> Baru
P1	2.433	2.331
P2	0.438	0.455
P3	1.148	1.157
P4	0.526	0.544
P5	0.925	0.940

- j. Memperbarui nilai *pbest* dengan membandingkan nilai *fitness* tertinggi pada iterasi saat ini dengan iterasi sebelumnya.

Tabel 21 merupakan nilai *pbest* yang telah diperbarui pada setiap partikel.

Tabel 21. Hasil Update *pbest*

P	X1	X2	X3	X4	X5	fitness
P1	0.35	0	0	0	0.06	2.433
P2	0.035	0.718	0.718	0.718	0.006	0.455
P3	0.328	0	0	0	0.535	1.157
P4	0.035	0.151	0.586	0.652	0.412	0.544
P5	0.035	0	0.558	0.463	0.006	0.940

k. Memperbarui nilai *gbest* dengan membandingkan nilai fitness tertinggi pada *pbest* baru.
 Berdasarkan tabel 22 nilai *fitness* tertinggi terletak pada partikel 1 Sehingga nilai *gbest* pada iterasi pertama adalah 2.433

Tabel 22. Hasil Update *gbest*

P	X1	X2	X3	X4	X5	fitness
P1	0.35	0	0	0	0.06	2.433

Evaluation

Hasil tahapan *evaluation* yaitu melakukan evaluasi pada model klasifikasi yang telah terbentuk menggunakan *confusion matrix*. Perbandingan hasil *accuracy*, *recall*, *precision*, dan *fmeasure* klasifikasi algoritma *Naive Bayes* dan algoritma *Naive Bayes* dengan seleksi fitur PSO pada *dataset* yang tidak diolah dengan SMOTE serta *dataset* yang telah diolah dengan SMOTE disajikan pada tabel 23.

Tabel 23. Perbandingan Nilai *Accuracy*, *Recall*, *Precision*, dan *Fmeasure*

Model	Accuracy	Recall	Precision	Fmeasure
NB	51.77%	51.81%	49.28%	49.66%
NB +PSO	53.67%	53.63%	51.01%	51.45%
NB +SMOTE	76.22%	76.22%	75.84%	75.27%
NB +PSO +SMOTE	76.48%	76.47%	76.20%	75.62%

Berdasarkan tabel 23, model *Naive Bayes*, PSO, dan SMOTE menghasilkan nilai *accuracy* tertinggi yaitu sebesar 76.48%, dengan nilai *recall* sebesar 76.47%, *precision* sebesar 76.20%, dan *fmeasure* sebesar 75.62%. Sedangkan model *Naive Bayes* tanpa seleksi fitur dan tanpa SMOTE menghasilkan nilai *accuracy* terendah yaitu sebesar 51.77%, dengan nilai *recall* sebesar 51.81%, *precision* sebesar 49.20%, dan *fmeasure* sebesar 49.66%. Apabila dibandingkan hasil *accuracy* pada *dataset* yang telah diolah dengan SMOTE lebih tinggi karena telah dilakukan penanganan terhadap data yang tidak seimbang sehingga sistem lebih mudah melakukan klasifikasi pada data emosi minoritas. Hal tersebut menunjukkan bahwa penggunaan data seimbang dapat meningkatkan nilai *accuracy* yang cukup signifikan.

Kemudian penerapan seleksi fitur PSO pada data yang telah diolah dengan SMOTE tidak menunjukkan peningkatan *accuracy* yang begitu signifikan. Hasil *accuracy* pada *dataset* yang telah

diolah dengan SMOTE hanya meningkat sebesar 0.26% setelah menerapkan seleksi fitur PSO pada algoritma *Naive Bayes*. Dimana sebelumnya menghasilkan *accuracy* sebesar 76.22% pada algoritma *Naive Bayes* dan SMOTE kemudian meningkat menjadi 76.48% pada algoritma *Naive Bayes*, PSO, dan SMOTE. Namun penerapan seleksi fitur PSO berkinerja baik pada *dataset* yang belum diolah dengan SMOTE dengan peningkatan nilai *accuracy* sebesar 1.9% yakni menghasilkan *accuracy* sebesar 51.77% pada algoritma *Naive Bayes* kemudian meningkat menjadi 53.67% pada algoritma *Naive Bayes* dan PSO. Selanjutnya istilah yang sering muncul dalam *dataset* akan divisualisasikan dalam bentuk *word cloud*. Gambar 6 merupakan *word cloud* komentar dengan label emosi terkejut.



Gambar 6. Word Cloud Label Emosi Terkejut

Berdasarkan gambar 6 istilah ‘bpom’ merupakan istilah yang paling sering dibicarakan oleh publik dengan ukuran yang cukup besar pada *word cloud*. Semakin besar ukuran istilah dalam *word cloud* maka semakin tinggi juga frekuensi istilah tersebut digunakan oleh publik dalam memberikan komentar. Selain istilah ‘bpom’, *word cloud* dengan emosi terkejut didominasi oleh istilah ‘obat’, ‘kenapa’, ‘edar’, ‘apa’, ‘kok bisa’, ‘baru sekarang’, ‘bisa lolos’. Istilah yang muncul pada emosi terkejut memberi informasi bahwa publik merasa heran serta kaget mengapa obat yang ditarik di pasaran dapat beredar dan lolos dari BPOM.



Gambar 7. Word Cloud Label Emosi Marah

Pada gambar 7 istilah ‘bpom’ juga menjadi istilah yang sering digunakan publik pada komentar

- (2021). Menggunakan Metode Naïve Bayes Berbasis Website Web-Based Expert System of Covid-19 Early Detection Using Naïve Bayes Method. *Journal of Information Technology and Computer Science (INTECOMS)*, 4(2), 446–455.
- Antonio, V. D., Effendi, S., & Mawengkang, H. (2021). Analisis Kinerja Ekstraksi Fitur TF-IDF (Term Frequency – Inverse Document Frequency) Untuk Algoritma Klasifikasi Stochastic Gradient Descent pada Analisis Sentimen Teks Indonesia.
- Asih, N. M., Jaman, J. H., & Umidah, Y. (2022). Analisis Sentimen Terhadap Bantuan Kuota Internet Dari Kemendikbud Dimasa Covid-19 Menggunakan Algoritma C5.0. *Journal of Information Technology and Computer Science (INTECOMS)*, 5(2).
- Asri, D. N., & Chusniah, T. (2016). Emosi ditinjau dari perspektif multibudaya. *Jurnal Prosiding Seminas Nasional Psikologi Indigenous Indonesia*, 167–174.
- Berniawan, D., Amri, & Tinaliah. (2023). Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Sentimen Pengguna Twitter Terhadap Kemkominfo Di Indonesia. *Multi Data Palembang Student Conference*, 2(1), 24–31.
- CBSEskillededucation. (2022). *AI Project Cycle Class 10 Notes*.
- Digital News Report. (2022). *Digital News Report*.
- Fitriana, R. D. (2018). Keseimbangan Emosi Dan Kesehatan Mental Manusia Dalam Perspektif Psikologi Agama. *Jurnal Ilmiah Syi'ar*, 18(1), 91. <https://doi.org/10.29300/syr.v18i1.1285>
- Hasan, F. N., & Wahyudi, M. (2018). Analisis Sentimen Artikel Berita Tokoh Sepak Bola Dunia Menggunakan Algoritma Support Vector Machine Dan Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal Akrab Juara*, 3(November), 42–55.
- Irnawati, O., & Solecha, K. (2022). Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naïve Bayes dengan Seleksi Fitur PSO. 4(02), 189–199.
- Kementerian Kesehatan RI. (2022). *Press Conference Update Perkembangan Gangguan Ginjal Akut Pada Anak (AKI) di Indonesia*.
- Kompas. (2022). *Kemenkes Rilis 3 Zat Berbahaya Diduga Penyebab Gagal Ginjal Akut pada Anak*.
- Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. A. P., Magcuyao, J. P. H., & Vicente, H. N. (2020). Classification algorithm accuracy improvement for student graduation prediction using ensemble model. *International Journal of Information and Education Technology*, 10(10), 723–727. <https://doi.org/10.18178/ijiet.2020.10.10.1449>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi Mypertamina pada Google Play Store Menggunakan Algoritma NBC. 9(1), 42–48.
- Meilani, F., Cahyani, D. E., & Sihwi, S. W. (2018). Perbandingan Emosi Pengguna Commuterline dan Transjakarta Berbasis Data Twitter Menggunakan Metode Multinomial Naïve Bayes. Universitas Sebelas Maret.
- Nurlaela, D. (2020). Penerapan Adaboost Untuk Meningkatkan Akurasi Naive Bayes Pada Prediksi Pendapatan Penjualan Film. *INTI Nusa Mandiri*, 14(2), 181–188. <https://doi.org/10.33480/inti.v14i2.1220>
- Pratama, A. Y., Umidah, Y., & Voutama, A. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 897–910.
- Rianti, D. L., Umidah, Y., & Voutama, A. (2021). Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 6(1), 98. <https://doi.org/10.30998/string.v6i1.9993>
- Sabilla, W. I., & Vista, C. B. (2021). Implementation of SMOTE and Under Sampling on Imbalanced Datasets for Predicting Company Bankruptcy. *Jurnal Komputer Terapan*, 7(2), 329–339.
- Sarasati, B., & Nurvia, O. (2021). Emosi Dalam Tulisan. *Jurnal Psibermetika*, 14(1), 40–48. <https://doi.org/10.30813/psibermetika>.
- Saraswati, E., Umidah, Y., & Voutama, A. (2021). Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19. *Generation Journal*, 5(2), 109–118. <https://doi.org/10.29407/gj.v5i2.16125>
- Sulistiyowati, N., & Jajuli, M. (2020). Integrasi Naive Bayes Dengan Teknik Sampling Smote Untuk Menangani Data Tidak Seimbang. *Nuansa Informatika*, 14(1), 34. <https://doi.org/10.25134/nuansa.v14i1.2411>
- Viva. (2022). *BPOM jadi Trending Topic Twitter, Warganet Bahas Obat Batuk*.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1), 134.