

SKRIPSI

**ANALISIS SENTIMEN VAKSIN COVID-19 DI INDONESIA
MENGUNAKAN ALGORITMA *NAÏVE BAYES* BERBASIS
*PARTICLE SWARM OPTIMIZATION***

Disusun dan diajukan oleh

IWAN KURNIAWAN

H051171514



**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR**

2022

**ANALISIS SENTIMEN VAKSIN COVID-19 DI INDONESIA
MENGUNAKAN ALGORITMA *NAÏVE BAYES* BERBASIS
*PARTICLE SWARM OPTIMIZATION***

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
Pada Program Studi Statistika Departemen Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

IWAN KURNIAWAN

H051171514

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

JANUARI 2022

LEMBAR PENGESAHAN KEONTETIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa Skripsi yang saya buat dengan judul:

**Analisis Sentimen Vaksin COVID-19 Di Indonesia Menggunakan
Algoritma *Naïve Bayes* Berbasis *Particle Swarm Optimization***

Adalah benar hasil kakrya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 03 Februari 2022



IWAN KURNIAWAN

NIM. H051171514

**ANALISIS SENTIMEN VAKSIN COVID-19 DI INDONESIA
MENGUNAKAN ALGORITMA *NAÏVE BAYES* BERBASIS
*PARTICLE SWARM OPTIMIZATION***

Disetujui Oleh:

Pembimbing Utama,



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.

NIP. 197407131999032001

Pembimbing Pertama,



Siswanto, S.Si., M.Si.

NIP. 199201072019031012

Ketua Departemen Statistika


Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 197201171997032002

Pada Tanggal: 03 Februari 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Iwan Kurniawan

NIM : H051171514

Program Studi : Statistika

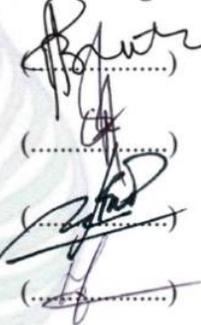
Judul Skripsi : Analisis Sentimen Vaksin COVID-19 Di Indonesia Menggunakan
Algoritma *Naïve Bayes* Berbasis *Particle Swarm Optimization*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D
2. Sekretaris : Siswanto, S.Si., M.Si.
3. Anggota : Drs. Raupong, M.Si
4. Anggota : Sitti Sahrinan, S.Si., M.Si.

Tanda Tangan



(.....)
(.....)
(.....)
(.....)

Ditetapkan di: Makassar

Tanggal: 03 Februari 2022

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus yang senantiasa menjadi penopang dan sahabat terbaik penulis sepanjang hidupnya. Karena kebaikan, kasih dan berkat anugrah-Nya sehingga penulis dapat menyelesaikan skripsi ini, dengan judul “**Analisis Sentimen Vaksin COVID-19 Di Indonesia Menggunakan Algoritma *Naïve Bayes* Berbasis *Particle Swarm Optimization***”. Penelitian ini dimaksudkan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana Sains pada Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Penulis banyak menghadapi kendala dan masalah dalam pelaksanaan penulisan skripsi ini namun dengan kemampuan dan kekuatan Tuhan Yesus berikan serta bantuan dan dukungan dari berbagai pihak, sehingga skripsi ini dapat selesai di waktu terbaik Tuhan.

Penulis menyampikan ucapan terima kasih untuk orangtua penulis atas cinta, doa dan dukungan moril maupun materil yang diberikan kepada penulis kepada Bapak Julius Bunga dan Mama Risma yang telah mendidik penulis dengan penuh kesabaran sampai saat ini. Ucapan terima kasih juga penulis sampaikan kepada saudara penulis yang ada di Makassar, Barru, dan Ciputat atas dukungan doa dan cinta kasih kehangatan keluarga. Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga disampaikan kepada:

1. Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu, MA**, selaku Rektor Universitas Hasanuddin dan Bapak **Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.
2. Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si**, selaku Ketua Departemen Statistika dan Bapak **Andi Kresna Jaya, S.Si., M.Si**, selaku Sekretaris Departemen Statistika dan **segenap dosen pengajar** yang telah memberikan ilmu pengetahuan, pengalaman serta nasihat-nasihat kepada penulis, serta **Staf Akademik** yang telah banyak membantu dalam mengurus berbagai keperluan persuratan.
3. Ibu **Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.** selaku pembimbing utama dan Bapak **Siswanto, S.Si, M.Si.** selaku pembimbing pertama dalam penulisan skripsi ini telah meluangkan waktu, tenaga, dan pikiran untuk membimbing dengan sabar dan turut mendoakan penulis hingga skripsi ini dapat selesai dengan baik.

4. Tim penguji Bapak **Drs. Raupong, M.Si.** dan Ibu **Sitti Sahrinan, S.Si, M.Si.** atas segala saran dan masukan yang diberikan untuk memperbaiki skripsi ini.
5. Teman seperbimbingan penulis **Sri Mulyani** yang telah berjuang bersama dan membantu dalam mengurus segala berkas administrasi skripsi penulis.
6. Untuk **Arya Dwi Putra, Karina Eka Pratiwi,** dan **Shinta Setiyono** terima kasih selalu kebersamaian penulis dalam mengerjakan tugas akhir ini, menjadi tempat mengeluh, dan penyemangat dikala jenuh. Terima kasih selalu ada.
7. Teman-teman **Statistika 2017** atas segala kebersamaan selama 4 tahun, kerja sama, kenangan suka dan duka serta dukungan selama proses perkuliahan.
8. **The Rarities FGL 2021** aka **Fildzah Lamuna, Muh. Jamil Reza,** dan **Brenda Prisyella** selaku *team mates* penulis yang selalu memberi dukungan di tengah padatnya *workload* selama menjadi manager.
9. **AIIESEC in UNHAS, AIIESEC in Indonesia,** dan **Bank BTN Kanwil V Makassar** yang menjadi wadah penulis untuk mengasah kemampuan dan meng-*upgrade* diri penulis menjadi seperti sekarang.
10. Terakhir, penulis ingin berterima kasih kepada **diri sendiri** untuk selalu percaya arti dari kerja keras dan selalu memegang teguh bahwa harapan akan selalu ada.

Penulis menyadari bahwa tugas akhir ini memiliki banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun diperlukan dalam penulisan selanjutnya. Akhirnya, penulis berharap skripsi ini dapat bermanfaat dalam pengembangan wawasan bidang ilmu Statistika dan menjadi berkat untuk penulis dan pembaca, Amin.

Makassar, 3 Februari 2022



Iwan Kurniawan

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Iwan Kurniawan
NIM : H051171514
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (*Non-exclusive Royalty- Free Right*) atas karya ilmiah saya yang berjudul:

**“Analisis Sentimen Vaksin COVID-19 Di Indonesia Menggunakan Algoritma
Naïve Bayes Berbasis *Particle Swarm Optimization*”**

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 3 Februari 2022

Yang menyatakan,



Iwan Kurniawan

ABSTRAK

Pengembangan vaksin COVID-19 akibat dari merebaknya *corona virus* membuat masyarakat Indonesia memberikan pendapat dan opininya melalui berbagai media, salah satunya media sosial *Twitter*, yang apabila dikumpulkan akan dapat dibuat kesimpulan melalui analisis sentimen. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang yang bersifat positif atau negatif. Dalam penelitian ini, Algoritma *Naïve Bayes* digunakan untuk melakukan klasifikasi sentimen ke dalam kelas positif dan negatif yang dioptimasi menggunakan metode *Particle Swarm Optimization* (PSO) mengenai vaksin COVID-19 di Indonesia. Penelitian ini menggunakan 2547 data *tweet* dengan kata kunci 'vaksin covid' yang diperoleh menggunakan *Twitter API*. Percobaan klasifikasi *tweet* menggunakan algoritma *Naïve Bayes* pada 2547 data *tweet* dengan seleksi fitur PSO menghasilkan distribusi sentimen sebanyak 2328 kelas positif dan 219 kelas negatif yang mayoritas berasal dari Pulau Jawa berdasarkan matriks bilangan random yang telah diinisialisasi pada metode PSO. Nilai ketepatan klasifikasi *accuracy* dan *F1 Score* secara berturut-turut sebesar 91.28% dan 95.38%. Kombinasi algoritma *Naïve Bayes* dengan seleksi fitur PSO terbukti memberikan solusi untuk masalah klasifikasi *tweet* dengan hasil yang akurat dan optimal.

Kata Kunci: Sentimen, Klasifikasi, *Naive Bayes*, *Particle Swarm Optimization*, *Twitter*.

ABSTRACT

The development of the COVID-19 vaccine as a result of the outbreak of the corona virus has made the Indonesians give their opinions through various media, one of which is Twitter social media, which when collected, conclusions can be drawn through sentiment analysis. Sentiment analysis is done to see opinions tendencies towards a problem or object by someone who is positive or negative. In this study, the Naïve Bayes Algorithm was used to classify sentiment into positive and negative classes which were optimized using the Particle Swarm Optimization (PSO) method regarding the COVID-19 vaccine in Indonesia. This study uses 2547 tweet data with the keyword 'vaksin covid' obtained using the Twitter API. The tweet classification experiment using the Naïve Bayes algorithm on 2547 tweet data with PSO feature selection resulted in a sentiment distribution of 2328 positive classes and 219 negative classes, the majority of which came from Java Island based on a random number matrix that was initialized in the PSO method. The value of classification accuracy and F1 Score are 91.28% and 95.38%, respectively. The combination of the Naïve Bayes algorithm with PSO feature selection is proven to provide a solution to the tweet classification problem with accurate and optimal results.

Keywords: Sentiment, Classification, Naive Bayes, Particle Swarm Optimization, Twitter.

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
LEMBAR PENGESAHAN KEONTETIKAN	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 <i>Supervised Learning</i>	4
2.2 Analisis Sentimen.....	4
2.3 <i>Algoritma Naïve Bayes</i>	4
2.4 <i>Twitter</i>	11
2.5 <i>Web Crawler</i>	12
2.6 <i>Preprocessing Data</i>	12
2.7 <i>Particle Swarm Optimization</i>	13
2.8 <i>Metode Geo-spasial</i>	17

2.9	<i>Word Cloud</i>	18
2.10	Evaluasi Kinerja Algoritma.....	18
2.11	Vaksin COVID-19.....	19
BAB III METODOLOGI PENELITIAN		21
3.1	Implementasi	21
3.3	Metode Analisis Data	21
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Proses <i>Preprocessing</i> Data.....	24
4.1.1	<i>Case Folding</i>	24
4.1.2	<i>Cleaning</i>	24
4.1.3	<i>Tokenizing</i>	25
4.1.4	<i>Filtering</i>	25
4.2	Proses Seleksi Fitur <i>Particle Swarm Optimization</i> (PSO)	25
4.2.1	Inisialisasi Populasi Partikel	26
4.2.2	Evaluasi Nilai <i>Pbest</i> Partikel	34
4.2.3	<i>Update</i> Komponen Partikel.....	37
4.3	Klasifikasi dengan Algoritma <i>Naïve Bayes</i>	47
4.3.1	Pengklasifikasian Manual Sentimen	48
4.3.2	Pembentukan Fitur Data.....	48
4.3.3	Perhitungan <i>Prior Probability</i> Kelas Sentimen	49
4.3.4	Perhitungan Probabilitas Fitur.....	50
4.3.5	Perhitungan Nilai <i>Hmap</i>	50
4.3.6	Hasil Klasifikasi Sentimen.....	52
4.4	Evaluasi Kinerja Algoritma.....	52
4.5	Visualisasi.....	54
4.5.1	Visualisasi Kelas Sentimen.....	54
4.5.2	Visualisasi Emosi Sentimen.....	55
4.5.3	Visualisasi <i>Word Cloud</i>	56

4.5.4	Visualisasi Geo-Spasial.....	57
BAB V KESIMPULAN DAN SARAN		59
5.1	Kesimpulan.....	59
5.2	Saran.....	59
DAFTAR PUSTAKA		60
LAMPIRAN.....		65

DAFTAR GAMBAR

Gambar 2.1	<i>Conditional Independence</i> pada <i>Naïve Bayes</i>	5
Gambar 2.2	Visualisasi perubahan posisi populasi saat mencari posisi terbaik	14
Gambar 4.1	Visualisasi Kelas Sentimen	54
Gambar 4.2	Visualisasi Emosi Sentimen	55
Gambar 4.3	Visualisasi <i>Word Cloud</i>	56
Gambar 4.4	Visualisasi Geo-Spasial Beberapa Wilayah di Indonesia	57

DAFTAR TABEL

Tabel 2.1	Dataset Golf	8
Tabel 2.2	Probabilitas Keputusan Bermain Golf	9
Tabel 2.3	Probabilitas Faktor Cuaca	9
Tabel 2.4	Probabilitas Faktor Temperatur.....	10
Tabel 2.5	Probabilitas Faktor Kelembaban Udara	10
Tabel 2.6	<i>Confusion Matrix</i>	18
Tabel 4.1	10 Fitur dengan Frekuensi Tertinggi.....	26
Tabel 4.2	Populasi Partikel.....	27
Tabel 4.8	Pemetaan Frekuensi Fitur terhadap Partikel	29
Tabel 4.9	Bilangan Random A.....	30
Tabel 4.10	Bilangan Random B.....	31
Tabel 4.11	Nilai Sigmoid Partikel untuk Iterasi-1	32
Tabel 4.12	Hasil Perbandingan Nilai $X_1^{(1)}$ atau A dengan Nilai Sigmoid.....	33
Tabel 4.13	Nilai P_{best} Partikel pada Iterasi-1	36
Tabel 4.14	Parameter PSO yang digunakan.....	37
Tabel 4.15	Hasil <i>Update</i> Kecepatan Partikel pada Iterasi-2	38
Tabel 4.16	Hasil <i>Update</i> Posisi Partikel pada Iterasi-2	39
Tabel 4.17	Nilai Sigmoid Partikel untuk Iterasi-2	40
Tabel 4.18	Hasil Perbandingan Nilai $X_2^{(2)}$ dengan Nilai Sigmoid	41
Tabel 4.19	Nilai P_{best} Partikel pada Iterasi-2.....	43
Tabel 4.23	Hasil Perbandingan Nilai $X_3^{(3)}$ dengan Nilai Sigmoid	45
Tabel 4.24	Perubahan Nilai P_{best} Partikel untuk setiap Iterasi	47
Tabel 4.25	Hasil Seleksi Fitur PSO.....	49
Tabel 4.26	Klasifikasi Sentimen Manual	49
Tabel 4.27	Pembentukan Fitur Data pada <i>tweet-1</i>	50
Tabel 4.28	Nilai <i>Prior Probability</i>	51

Tabel 4.29 Probabilitas Tiap Fitur	51
Tabel 4.30 Nilai H_{map} untuk 2547 <i>tweet</i>	52
Tabel 4.31 Klasifikasi Kelas Sentimen	53
Tabel 4.32 Hasil <i>Confusion Matrix</i>	53

DAFTAR LAMPIRAN

Lampiran 1 Data <i>tweet</i> hasil <i>crawling</i>	66
Lampiran 2. Hasil Tahap <i>Case Folding</i>	69
Lampiran 3. Hasil Tahap <i>Cleaning</i>	70
Lampiran 4. Hasil Tahap <i>Tokenizing</i>	72
Lampiran 5. Hasil Tahap <i>Filtering</i>	74
Lampiran 6. Data <i>tweet</i> Hasil <i>preprocessing</i>	76
Lampiran 7. Hasil <i>Update</i> Kecepatan Partikel pada Iterasi-3.....	77
Lampiran 8. Hasil <i>Update</i> Posisi Partikel pada Iterasi-3	78
Lampiran 9. Nilai Sigmoid Partikel untuk Iterasi-3.....	79
Lampiran 10. Nilai H_{map} , kelas sentimen, dan emosi sentimen.....	80

BAB I PENDAHULUAN

1.1 Latar Belakang

Awal tahun 2020 ini dunia dikejutkan oleh penyebaran *coronavirus*, virus yang bersifat *strain* baru sehingga tidak dikenal oleh kekebalan tubuh manusia yang sudah beradaptasi dengan penyakit lama dengan *strain* virus yang lama (Wijono, 2020). Penyakit *coronavirus* 2019 (COVID-19) telah menjangkiti lebih dari 30 juta populasi manusia dan telah merenggut lebih dari 1 juta korban jiwa (Bowdle & Munoz-Price, 2020). Masyarakat menunggu hadirnya vaksin yang akan memberikan perlindungan dengan melatih sistem kekebalan orang untuk melawan virus sehingga mereka tidak jatuh sakit (Haseltine, 2020).

Seiring beredarnya informasi perkembangan vaksin COVID-19, muncul berbagai pandangan mengenai pembuatan vaksin ini. Masyarakat memberikan pendapat dan opininya melalui berbagai media, salah satunya media sosial *Twitter*. Opini yang diberikan masyarakat terhadap vaksin COVID-19 beragam, baik itu opini yang pro dan kontra, mengenai efektivitas, kelayakan, target sasaran, bahkan dugaan bisnis dibalik pembuatan vaksin tersebut. Apabila diteliti lebih lanjut terhadap opini tersebut maka akan diperoleh sebuah sentimen yang apabila dikumpulkan akan dapat dibuat kesimpulan (Zadrozny & Kodali, 2013).

Menurut Alsaeedi & Khan (2019), analisis sentimen adalah suatu cara untuk menilai opini tertulis atau lisan untuk menentukan suatu opini bersifat positif, negatif atau netral. Terdapat beberapa metode yang telah digunakan untuk analisis sentimen ini, diantaranya pendekatan *Lexicon Based* (Effendy, 2015), Algoritma *K-Nearest Neighbor* (KNN) (Claudy dkk., 2018), metode *Naïve Bayes* (Fanissa dkk., 2018). Salah satu metode atau algoritma yang populer dalam pengklasifikasian pada analisis sentimen adalah Algoritma *Naïve Bayes*. *Naïve Bayes* merupakan penggolongan menggunakan statistik sederhana berdasarkan Teorema Bayes yang mengasumsikan keberadaan atau ketiadaan dari suatu kelas dengan fitur lainnya (Frank & Bouckaert,

2006). Beberapa keuntungan algoritma *Naïve Bayes* diantaranya adalah sangat cocok untuk digunakan pada klasifikasi data teks, dapat digunakan pada *dataset* yang besar, serta bekerja sangat baik pada *multi-class prediction*. Sifatnya yang sebagai model probabilitas, menyebabkan algoritma *Naïve Bayes* dikategorikan sebagai *supervised learning* (Pramana dkk., 2018).

Suatu algoritma biasanya akan bekerja lebih baik lagi jika dilakukan seleksi fitur pada datanya (Pushpalata & Gupta, 2012). Salah satu metode seleksi fitur yang digunakan pada algoritma *Naïve Bayes* adalah *Particle Swarm Optimization* (PSO). PSO adalah teknik optimasi berdasarkan populasi yang terinspirasi oleh perilaku sosial dari pergerakan burung atau ikan (*bird flocking* atau *fish schooling*) (Rosita & Purwananto, 2012). PSO sebagai alat optimasi menyediakan prosedur pencarian berbasis populasi dengan masing-masing individu yang disebut partikel menyesuaikan kecepatan dan mencari posisi terbaik berdasarkan informasi yang diperoleh dari partikel lain.

Penelitian yang berkaitan tentang analisa sentimen telah dilakukan oleh Effendy (2015) dengan menggunakan pendekatan *Lexicon Based* yang membahas tentang informasi seputar solusi sampah dengan perolehan nilai akurasi metode sebesar 70,68%. Analisa sentimen juga dilakukan oleh Claudy dkk. (2018) menggunakan Algoritma *K-Nearest Neighbor* (KNN) mengenai klasifikasi karakter calon karyawan berdasarkan *tweet* dengan perolehan akurasi sebesar 66%. Fanissa dkk. (2018) juga mengkaji analisa sentimen menggunakan metode *Naïve Bayes* dengan seleksi fitur *Query Expansion Ranking* yang membahas tentang informasi seputar pariwisata di Kota Malang dan memperoleh tingkat akurasi sebesar 86,6%.

Dari berbagai penelitian sebelumnya mengenai analisa sentimen terhadap media sosial *Twitter* yang telah di uraikan, terlihat bahwa algoritma *Naïve Bayes* memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode lainnya. Oleh karena itu pada penelitian ini akan dibahas aplikasi Algoritma *Naïve Bayes* dengan menggunakan seleksi fitur PSO pada analisis sentimen *Twitter* mengenai vaksin COVID-19 di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang akan dibahas pada penelitian ini adalah:

1. Bagaimana hasil klasifikasi opini masyarakat tentang vaksin COVID-19 di Indonesia melalui analisis sentimen menggunakan algoritma *Naïve Bayes* dengan seleksi fitur PSO?
2. Bagaimana kinerja algoritma *Naïve Bayes* dengan seleksi fitur PSO dalam klasifikasi opini masyarakat tentang vaksin COVID-19 tersebut?
3. Bagaimana sebaran lokasi sentimen di wilayah Indonesia?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah menggunakan kumpulan data *tweet* berbahasa Indonesia yang berformat teks dari bulan 1 Januari hingga 30 Juni 2021 serta menggunakan 10 fitur dan 10 partikel dan nilai *learning factor* $c_1 < c_2$ yaitu $c_1 = 0,4$ dan $c_2 = 0,6$ pada metode PSO.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Memperoleh hasil klasifikasi opini masyarakat tentang vaksin COVID-19 di Indonesia melalui analisis sentimen menggunakan algoritma *Naïve Bayes* dengan seleksi fitur PSO.
2. Memperoleh kinerja algoritma *Naïve Bayes* dengan seleksi fitur PSO dalam klasifikasi opini masyarakat tentang vaksin COVID-19.
3. Memperoleh visualisasi peta sebaran lokasi sentimen data *tweet* vaksin COVID-19 di Indonesia.

1.5 Manfaat Penelitian

Manfaat yang diperoleh pada penelitian ini adalah dapat menambah wawasan mengenai klasifikasi kelas sentimen pada data yang berupa teks dengan memanfaatkan algoritma *Naïve Bayes* dan seleksi fitur *Particle Swarm Optimization*.

BAB II

TINJAUAN PUSTAKA

2.1 *Supervised Learning*

Klasifikasi merupakan salah satu metode yang dapat menangani *big data*. Terdapat dua metode dalam klasifikasi yaitu *supervised learning* dan *unsupervised learning*. *Supervised Learning*, disebut juga analisis klasifikasi, merupakan suatu teknik statistik yang bertujuan untuk mengelompokkan data ke dalam kelas-kelas yang telah memiliki label dengan membangun suatu model yang berdasarkan kepada suatu *dataset* serta memprediksi kelas dari suatu data (Pramana dkk., 2018).

2.2 Analisis Sentimen

Menurut Alsaeedi dan Khan (2019), analisis sentimen adalah suatu cara untuk menilai opini tertulis atau lisan untuk menentukan suatu opini bersifat positif, negatif atau netral. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam data, kalimat, atau fitur/tingkat aspek dan menentukan suatu pendapat yang dikemukakan dalam data, kalimat atau fitur entitas/aspek bersifat positif, negatif atau netral. Kata-kata yang tergolong bersifat positif diantaranya seperti baik, senang, membantu, peduli, terima kasih, dsb. Kemudian, kata-kata yang tergolong bersifat negatif diantaranya seperti sakit, panik, khawatir, kehilangan, menakutkan, gila, dsb. Adapun, kata-kata yang tergolong bersifat netral biasanya merupakan sebuah konjungsi, subjek, atau kata depan. Lebih lanjut, analisis sentimen juga dapat menyatakan emosional sedih, gembira, atau marah (Liu, 2012).

2.3 Algoritma *Naïve Bayes*

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas (Annur, 2018). *Naïve Bayes* merupakan salah satu algoritma paling efektif dan efisien pada pembelajaran *machine learning* dan *data mining*. Algoritma ini memiliki performa yang sangat menjanjikan dan sangat kompetitif dengan algoritma lainnya oleh karena asumsi *conditional*

independence yang menjadi dasarnya dan pengaplikasiannya dalam kehidupan yang sudah cukup banyak.

Algoritma *Naïve Bayes* salah satunya digunakan untuk klasifikasi teks serta merupakan metode *machine learning* yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh *Thomas Bayes* (Chai dkk., 2002). Algoritma ini digunakan untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu.

Aturan/Teorema Bayes (Zhang, 2004). adalah sebagai berikut:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)}, \quad P(X) \geq 0 \quad (2.1)$$

dengan:

X : atribut

C : kelas

$P(C|X)$: probabilitas kejadian bersyarat C dengan syarat X terjadi

$P(X|C)$: probabilitas kejadian bersyarat X dengan syarat C terjadi

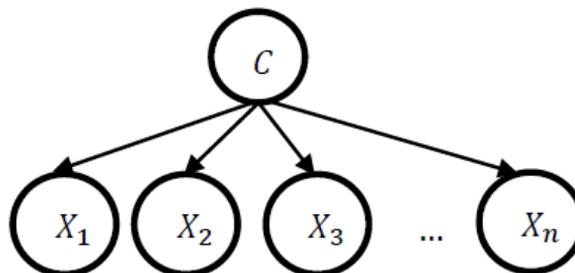
$P(C)$: probabilitas kejadian C

$P(X)$: probabilitas kejadian X

X dapat dituliskan sebagai:

$$X = (x_1, x_2, x_3, \dots, x_n)$$

Hubungan antara C dan X dapat dilihat pada Gambar 2.1.



Gambar 2.1 *Conditional Independence* pada *Naïve Bayes*

Persamaan (2.1) dapat dituliskan sebagai:

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

dengan,

$$\left. \begin{aligned} P(C|x_1) &= \frac{P(x_1 \cap C)}{P(x_1)} \\ P(C|x_1) &= \frac{P(x_1 \cap C)}{P(x_1)} \\ \vdots \\ P(C|x_n) &= \frac{P(x_n \cap C)}{P(x_n)} \end{aligned} \right\} P(C|X) \propto P(X \cap C)$$

Selanjutnya, $P(C|X)$ dapat diuraikan menjadi:

$$\begin{aligned} P(C|X) &= P(X \cap C) \\ &= P(x_1 \cap x_2 \cap \dots \cap x_n \cap C) \\ &= P(x_1|x_2 \cap x_3 \cap \dots \cap x_n \cap C) \cdot P(x_2 \cap x_3 \cap \dots \cap x_n \cap C) \\ &= P(x_1|x_2 \cap \dots \cap x_n \cap C) \cdot P(x_2|x_3 \cap \dots \cap x_n \cap C) \dots \\ &\quad P(x_{n-1}|x_n, \dots \cap C) \cdot P(x_n|C) \cdot P(C) \end{aligned}$$

Sekarang asumsi *conditional independence* diberlakukan yaitu asumsikan bahwa semua fitur di X saling independen, bergantung pada kelas C sehingga $P(C|X)$ dapat dituliskan sebagai:

$$P(C|x_1, x_2, \dots, x_n) = P(C) \prod_{i=1}^n P(x_i|C)$$

Naïve Bayes adalah model penyederhanaan dari metode Bayes yang cocok untuk pengklasifikasian teks (*Naïve Bayes Classifier*). *Naïve Bayes Classifier* dikenal lebih baik daripada beberapa metode klasifikasi lainnya seperti metode *Decision Tree*, *Support Vector Machine*, *Fuzzy*, dll. Ciri utama *Naïve Bayes* adalah asumsi independensi yang sangat kuat dari setiap kondisi atau peristiwa. Kemudian, modelnya sederhana, mudah dibuat dan dapat diimplementasikan untuk kumpulan data yang besar (Salmi & Rustam, 2019). Menurut (Wibawa dkk., 2019) dan (Syarifah & Muslim, 2015) beberapa keuntungan algoritma *Naïve Bayes* diantaranya adalah:

- 1) Sangat cocok untuk digunakan pada klasifikasi data teks.
- 2) Dapat digunakan pada *dataset* yang besar.
- 3) Mudah dan cepat untuk memprediksi klasifikasi dataset dan bekerja sangat baik pada *multi-class prediciton*
- 4) Bekerja dengan baik dengan variabel input yang bersifat kategorik dibandingkan variabel input numerik.
- 5) Bekerja sangat cepat sehingga dapat digunakan untuk memprediksi secara *real time*.

Beberapa persamaan yang dapat digunakan dalam *Naïve Bayes Classifier* adalah yaitu:

$$P(v_j) = \frac{|docs|}{|training|} \quad (2.2)$$

dengan:

$P(v_j)$: Probabilitas setiap data terhadap sekumpulan data
 $|docs|$: Frekuensi data pada setiap kategori

$$P(w_k|v_j) = \frac{n_k+1}{|n+jumlahkata|} \quad (2.3)$$

dengan:

$P(w_k|v_j)$: Probabilitas kemunculan kata w_k pada suatu data
 n_k : Frekuensi kata ke- k setiap kategori

Dalam *Naïve Bayes Classifiers* kita perlu memaksimalkan nilai probabilitas setiap kelas, yang dinyatakan sebagai *Hypothesis Maximum A Posteriori* (H_{map}) (Koller & Friedman, 2009):

$$H_{map} = \underset{\{positif, negatif\}}{argmax} P(w_k|c) \cdot P(c) \quad (2.4)$$

dengan:

H_{map} : Nilai probabilitas tertinggi data dari masing-masing kelas

$P(w_k|c)$: Probabilitas kemunculan kata w_k pada kelas c

$P(c)$: Probabilitas kelas c

Contoh sederhana.

Sekelompok pemuda diketahui sedang berencana untuk bermain golf. Namun, beberapa faktor seperti cuaca, temperatur, dan kelembaban udara dipertimbangkan untuk menentukan keputusan pemuda tersebut akan pergi bermain golf atau tidak. Berikut *dataset* yang diberikan.

Tabel 2.1. *Dataset* Golf

No.	Cuaca	Temperatur	Kelembaban udara	Bermain golf
1.	Hujan	Dingin	Tinggi	Tidak
2.	Hujan	Dingin	Tinggi	Tidak
3.	Mendung	Dingin	Tinggi	Tidak
4.	Terik	Panas	Normal	Ya
5.	Hujan	Sejuk	Normal	Tidak
6.	Terik	Panas	Normal	Ya
7.	Mendung	Sejuk	Normal	Ya
8.	Hujan	Dingin	Tinggi	Tidak

Pada *dataset* diperoleh:

Tabel 2.2. Probabilitas Keputusan Bermain Golf

Bermain golf		Probabilitas
Ya	5	$\frac{5}{8}$
Tidak	3	$\frac{3}{8}$
Total	8	1

Probabilitas pemuda bermain golf berdasarkan dataset adalah $\frac{5}{8}$, sedangkan probabilitas pemuda untuk tidak bermain golf adalah $\frac{3}{8}$. Kemudian dilakukan perhitungan probabilitas keputusan untuk setiap faktor.

Tabel 2.3. Probabilitas Faktor Cuaca

Atribut	Ya	Tidak	P(Ya)	P(Tidak)
Terik	2	0	$\frac{2}{3}$	$\frac{0}{5}$
Mendung	1	1	$\frac{1}{3}$	$\frac{1}{5}$
Hujan	0	4	$\frac{0}{3}$	$\frac{4}{5}$
Total	3	5	1	1

Pada faktor cuaca terlihat bahwa banyaknya data berkategori “Terik” yang berlabel Ya adalah 2 dan berlabel Tidak adalah 0 dengan probabilitas berlabel Ya sebesar $\frac{2}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{0}{5}$. Kemudian, banyaknya data berkategori “Mendung” yang berlabel Ya adalah 1 dan berlabel Tidak adalah 1 dengan probabilitas berlabel Ya sebesar $\frac{1}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{1}{5}$. Untuk data berkategori “Hujan” yang berlabel Ya adalah 0 dan berlabel Tidak adalah 4 dengan probabilitas berlabel Ya sebesar $\frac{0}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{4}{5}$. Total probabilitas untuk faktor cuaca adalah 1.

Tabel 2.4 Probabilitas Faktor Temperatur

Atribut	Ya	Tidak	P(Ya)	P(Tidak)
Panas	2	0	$\frac{2}{3}$	$\frac{0}{5}$
Sejuk	1	1	$\frac{1}{3}$	$\frac{1}{5}$
Dingin	0	4	$\frac{0}{3}$	$\frac{4}{5}$
Total	3	5	1	1

Pada faktor temperatur terlihat bahwa banyaknya data berkategori “Panas” yang berlabel Ya adalah 2 dan berlabel Tidak adalah 0 dengan probabilitas berlabel Ya sebesar $\frac{2}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{0}{5}$. Kemudian, banyaknya data berkategori “Sejuk” yang berlabel Ya adalah 1 dan berlabel Tidak adalah 1 dengan probabilitas berlabel Ya sebesar $\frac{1}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{1}{5}$. Untuk data berkategori “Dingin” yang berlabel Ya adalah 0 dan berlabel Tidak adalah 4 dengan probabilitas berlabel Ya sebesar $\frac{0}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{4}{5}$. Total probabilitas untuk faktor temperatur adalah 1.

Tabel 2.5 Probabilitas Faktor Kelembaban Udara

Atribut	Ya	Tidak	P(Ya)	P(Tidak)
Tinggi	0	4	$\frac{0}{3}$	$\frac{4}{5}$
Normal	3	1	$\frac{3}{3}$	$\frac{1}{5}$
Total	3	5	1	1

Pada faktor kelembaban udara terlihat bahwa banyaknya data berkategori “Tinggi” yang berlabel Ya adalah 0 dan berlabel Tidak adalah 4 dengan probabilitas berlabel Ya sebesar $\frac{0}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{4}{5}$. Kemudian, banyaknya data berkategori “Normal” berlabel Ya adalah 3 dan berlabel Tidak adalah 1 dengan

probabilitas berlabel Ya sebesar $\frac{3}{3}$ dan probabilitas berlabel Tidak sebesar $\frac{1}{5}$. Total probabilitas untuk faktor kelembaban udara adalah 1.

Setelah menghitung probabilitas untuk setiap faktor, dilakukan pengambilan keputusan berdasarkan nilai probabilitas kelas “Ya” dan “Tidak” untuk setiap faktor dengan ketentuan bahwa cuaca sedang terik, temperatur panas, dan kelembaban udara normal.

$$P(\text{"Ya"}) = P(\text{Ya}|\text{terik}) \cdot P(\text{Ya}|\text{panas}) \cdot P(\text{Ya}|\text{normal}) = \left(\frac{2}{3}\right) \left(\frac{2}{3}\right) \left(\frac{3}{3}\right) = 0,444$$

$$P(\text{"Tidak"}) = P(\text{Tidak}|\text{terik}) \cdot P(\text{Tidak}|\text{panas}) \cdot P(\text{Tidak}|\text{normal}) = \left(\frac{0}{5}\right) \left(\frac{0}{5}\right) \left(\frac{1}{5}\right) = 0,000$$

Oleh karena $P(\text{"Ya"}) = 0,444 > P(\text{"Tidak"}) = 0,000$ maka pemuda tersebut memutuskan untuk pergi bermain golf pada hari itu.

2.4 *Twitter*

Twitter adalah alat jejaring sosial gratis yang banyak digunakan yang memungkinkan orang-orang dalam berbagi informasi, dalam umpan berita *real-time* (Mistry, 2011) melalui posting komentar singkat tentang pengalaman dan pemikiran mereka (Bristol, 2010). Pesan publik yang dikirim dan diterima melalui *Twitter* - atau 'tweet' - dibatasi tidak lebih dari 140 karakter dan dapat menyertakan tautan ke blog, halaman web, gambar, video, dan semua materi *online* lainnya. Meskipun alat media ini singkat, penggunaan *Twitter* banyak digunakan dalam berbagai keadaan. Menurut Mollett dkk. (2011) “ribuan akademisi dan peneliti di semua tingkat pengalaman dan disemua disiplin ilmu sudah menggunakan *Twitter* setiap hari”.

Twitter telah menjadi media sosial yang sangat populer dan efektif. *Twitter* telah menyediakan *Application Programming Interface* (API) guna memudahkan para *developer* untuk mengambil data dari *twitter* dan mengolahnnya. Pengumpulan data dari *twitter* dapat digunakan untuk berbagai kebutuhan, salah satunya dalam melakukan analisis sentimen. Pengunduhan data dari *twitter* memerlukan *Keys* dan *Token Twitter API* yang juga terintegrasi dengan banyak *software* pemrograman sehingga

memudahkan para peneliti untuk melakukan suatu analisis opini/sentimen. Format opini yang terdapat pada *tweet* tidak hanya berbentuk tentang teks saja tetapi juga format lain seperti emoji atau *emoticon*. Orang-orang sering menggunakannya ketika lebih sulit untuk menggambarkan ekspresi mereka hanya dengan kata-kata (Ayvaz & Shiha, 2017). *Twitter* bahkan memberikan fitur tambahan lain kepada *user* untuk membuat *tweet* dalam format audio-visual seperti *voice note* maupun gambar (Chu & Roy, 2017).

2.5 Web Crawler

Web Crawler adalah program atau skrip otomatis yang menelusuri *World Wide Web* dengan cara metodis dan otomatis (Gupta & Johari, 2009). Proses ini disebut juga *Web Crawling* atau *Spidering*. *Web Crawler* secara umum digunakan untuk membuat salinan dari semua halaman yang dikunjungi untuk diproses lebih lanjut oleh *search engine*, yang akan mengindeks halaman yang diunduh untuk menyediakan pencarian cepat.

Crawler juga dapat digunakan untuk mengotomatisasi tugas pemeliharaan di situs *Web*, seperti memeriksa tautan atau memvalidasi kode HTML. Selain itu, *Crawler* dapat digunakan untuk mengumpulkan jenis informasi tertentu dari halaman *Web*, seperti memanen alamat *e-mail* (biasanya untuk spam).

2.6 Preprocessing Data

Preprocessing data terdiri dari *manual-labelling* dan *preprocessing* untuk mengubah data menjadi data terstruktur (Rachmat & Lukito, 2016). *Manual-labelling* adalah analisis untuk menentukan sentimen positif dan negatif pada data *tweet*. Proses ini mencakup membaca dengan cermat setiap kalimat di *tweet* dan menyesuaikan sentimen *tweet* berdasarkan kemunculan kata sifat dan kata kerja. Kemudian *preprocessing* adalah proses pembersihan dan penyiapan data untuk proses selanjutnya yaitu proses klasifikasi.

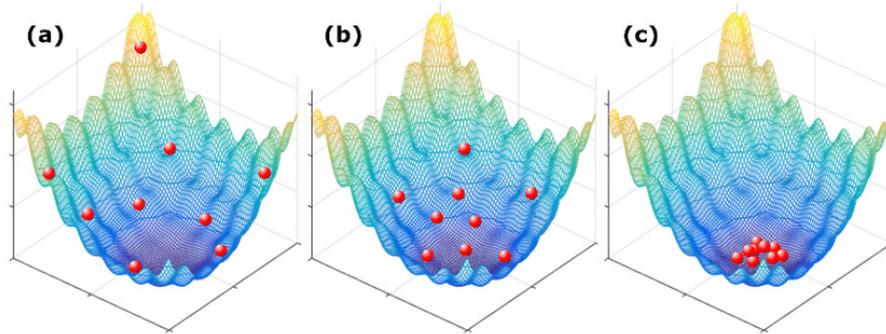
Terdapat empat tahapan pada *preprocessing data* yaitu *case folding*, *cleaning data*, *tokenizing*, dan *filtering* dengan tujuan untuk mengubah data mentah menjadi bentuk yang lebih mudah diproses oleh sistem (Mujilahwati, 2016). Keempat tahapan tersebut yaitu:

1. *Case Folding*, yaitu tahap mengubah semua huruf dalam data menjadi huruf kecil serta menghilangkan tanda baca pada data.
2. *Cleaning Data*, yaitu tahap menghapus setiap karakter selain huruf seperti URL, *mention*, dan *hashtag*.
3. *Tokenizing*, yaitu tahap setiap kata akan dipisahkan berdasarkan spasi yang ditemukan.
4. *Filtering*, yaitu tahap mengambil kata-kata penting dari hasil token.

2.7 Particle Swarm Optimization

Particle Swarm Optimization (PSO) dikembangkan oleh Kennedy dan Eberhart pada tahun 1995. PSO adalah teknik optimasi berdasarkan populasi yang terinspirasi oleh perilaku sosial dari pergerakan kawanan burung atau ikan (*bird flocking* atau *fish schooling*). Kawanan diasumsikan mempunyai ukuran tertentu dengan setiap partikel posisi awalnya terletak di suatu lokasi yang acak dalam ruang multidimensi.

Dalam PSO, setiap partikel diasumsikan memiliki dua karakteristik: posisi dan kecepatan. Setiap partikel bergerak dalam ruang tertentu dan mengingat posisi terbaik yang pernah dilalui atau ditemukan terhadap sumber makanan atau fungsi nilai objektif. Setiap partikel menyampaikan informasi atau posisi terbaiknya kepada partikel lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi tersebut.



Gambar 2.2 Visualisasi perubahan posisi populasi saat mencari posisi terbaik

Visualisasi pergerakan populasi dalam mencari posisi terbaik ditampilkan pada Gambar 2.2. Pada a) terlihat bahwa partikel diinisialisasi dan tersebar secara acak pada ruang dimensi D . Setiap partikel mewakili posisi dan kecepatan tertentu yang kemudian akan bergerak menuju daerah minimum lokal suatu fungsi. Kemudian pada b) setiap partikel akan memperbarui posisi dan kecepatan yang baru berdasarkan informasi yang diperoleh di partikel sekitarnya lalu seluruh partikel akan terus bergerak dan berhenti pada daerah minimum lokal fungsi tersebut seperti yang ditunjukkan pada c). Algoritma PSO menggabungkan metode pencarian lokal (*local/personal search*) dengan metode pencarian global (*global search*). Berikut ini merupakan model matematika yang menggambarkan mekanisme *updating* status partikel (Kennedy & Eberhart, 1995):

Update kecepatan partikel:

$$V_i^{(t)} = V_i^{(t-1)} + c_1 B (P_{best} - X_i^{(t-1)}) + c_2 A (G_{best} - X_i^{(t-1)}) \quad (2.5)$$

dengan:

- $V_i^{(t)}$: Kecepatan partikel ke- i pada iterasi ke- t
- $V_i^{(t-1)}$: Kecepatan partikel ke- i pada iterasi ke- $(t-1)$
- $X_i^{(t-1)}$: Posisi partikel ke- i pada iterasi ke- $(t-1)$
- P_{best} : *Personal Best* partikel ke- i
- G_{best} : *Global Best* partikel ke- i

c_1, c_2 : Konstanta akselerasi (*learning factor*) (0,1)

A, B : Bilangan Random [0,1]

Pemilihan konstanta akselerasi/*learning factor* adalah bebas, sehingga nilai koefisien dapat divariasikan dan untuk menentukan koefisien yang optimal hanyalah melalui prosedur *trial and error* (Wu et al., 2014).

Dalam implementasinya, ditemukan bahwa kecepatan partikel pada Persamaan (2.5) diupdate terlalu cepat dan nilai minimum fungsi tujuan yang dicari sering terlewati. Oleh karena itu dilakukan modifikasi atau perbaikan terhadap Persamaan (2.5). Perbaikan itu berupa penambahan suatu bobot inersia ω untuk mengurangi kecepatan pada formula update kecepatan. Untuk mencapai hal tersebut, Shi & Eberhart (1998) menyatakan bobot inersia yang mengecil nilainya dengan bertambahnya iterasi dapat menggunakan formula:

$$\omega^{(t)} = \omega_{max} - \left(\frac{\omega_{max} - \omega_{min}}{t_{max}} \right) t \quad (2.6)$$

dengan:

t : iterasi

$\omega^{(t)}$: bobot *inertia weight* pada iterasi ke- t

ω_{max} : bobot *inertia weight* maksimum yang bernilai 1,2

ω_{min} : bobot *inertia weight* maksimum yang bernilai 0,8

t_{max} : iterasi maksimum

Shi & Eberhart (1998) secara sistematis menuliskan persamaan tersebut sebagai:

Update kecepatan partikel (modifikasi):

$$V_i^{(t)} = \omega V_i^{(t-1)} + c_1 A (P_{best} - X_i^{(t-1)}) + c_2 B (G_{best} - X_i^{(t-1)}) \quad (2.7)$$

Update posisi partikel:

$$X_i^{(t)} = V_i^{(t)} + X_i^{(t-1)} \quad (2.8)$$

dengan:

- ω : Parameter *inertia weight*
- $X_i^{(t)}$: Posisi partikel ke- i pada iterasi ke- t
- $V_i^{(t)}$: Kecepatan partikel ke- i pada iterasi ke- t
- $X_i^{(t-1)}$: Posisi partikel ke- i pada iterasi ke- $(t-1)$

Pada PSO algoritma dimodifikasi untuk memperbarui solusi dengan nilai diskrit untuk setiap dimensi. Persamaan (2.9) mewakili fungsi Sigmoid yang digunakan untuk menentukan probabilitas iterasi ke- t , dan Persamaan (2.10) digunakan untuk memperbarui posisi setiap partikel (Abualigah & Khader, 2017).

$$S(V_i^{(t)}) = \frac{1}{(1 + \exp^{-V_i^{(t)}})} \quad (2.9)$$

dengan:

- $S(V_i^{(t)})$: Nilai sigmoid kecepatan partikel pada iterasi ke- t

$$M = \begin{cases} 1, & \text{jika } X_i^{(t)} \leq S(V_i^{(t)}) \\ 0, & \text{lainnya} \end{cases} \quad (2.10)$$

dengan:

- M : Kecocokan nilai posisi partikel terhadap nilai sigmoidnya

P_{best} merupakan posisi terbaik yang dimiliki oleh masing-masing partikel. Besarnya nilai P_{best} menggambarkan ukuran ketepatan/kelayakan prediksi posisi partikel pada suatu dimensi tertentu. Semakin tinggi nilai P_{best} maka posisi suatu partikel semakin baik/dekat dengan daerah minimum lokal suatu fungsi (Ibrahim dkk., 2021). Penentuan nilai P_{best} digunakan untuk mengevaluasi setiap solusi yang dihasilkan oleh metode. Dalam setiap iterasi, nilai P_{best} dari setiap solusi dihitung untuk memutuskan ada atau tidaknya perbaikan yang ditemukan dalam solusi. Akhirnya, solusi dengan nilai P_{best} tertinggi adalah solusi terbaik.

$$P_{best} = \frac{1}{n} \sum \left| \frac{v_i^{(t)} - s(v_i^{(t)})}{v_i^{(t)}} \right| \quad (2.11)$$

PSO banyak digunakan untuk menyelesaikan masalah optimasi yang kompleks dari berbagai bidang. Algoritma ini terbilang cukup sederhana karena hanya mengandalkan dua populasi yaitu *personal best* dan *global best*. Hal ini memungkinkan keragaman dan eksplorasi yang lebih besar atas satu populasi. Selain itu, efek momentum pada pergerakan partikel dapat memungkinkan konvergensi yang lebih cepat.

2.8 Metode *Geo-spasial*

Goldstein dkk. (2014), menyebutkan bahwa *Geo-spasial* adalah salah satu bagian umum dari analisis spasial. *Geo-spasial* adalah proses mengubah alamat atau nama tempat menjadi koordinatnya (Alamsyah dkk., 2015). *Geo-spasial* digunakan dalam sistem informasi geografis untuk membantu menemukan koordinat wilayah atau tempat (Goldberg dkk., 2007).

Salah satu *tools* yang dapat digunakan dalam melakukan metode ini adalah layanan *Google Maps API*. *Google Maps API* memungkinkan alamat jalan untuk di-*geocode* secara eksklusif. Jika lokasinya bukan tempat tinggal, tampilan depan *Google Places* ke *Google Maps* menyediakan lapisan tambahan pencocokan tertentu untuk bisnis dan tempat menarik yang tidak diselesaikan melalui *Google Maps* saja, sehingga meningkatkan tingkat keberhasilan pengodean geografis. *Google Maps API* juga dapat digunakan pada berbagai software pemrograman, salah satunya R. Penggunaan metode *Geo-spasial* pada analisis data yang bersifat tekstual memungkinkan untuk melihat sebaran opini masyarakat berdasarkan letak geografisnya.

2.9 Word Cloud

Word cloud adalah sejenis daftar berbobot untuk memvisualisasikan bahasa atau data teks (Kinnaird & Talgam-Cohen, 2012). *Word cloud* merupakan cabang penting dari data mining yang telah mendapatkan perhatian dan lebih banyak peluang aplikasi dalam *Big Data*. Pada abad ke-21, seiring berkembangnya teknologi internet khususnya *website* dan blog, kata cloud digunakan untuk meletakkan tag pada halaman internet sebagai alat bantu navigasi informasi pembaca melalui visualisasi frekuensi setiap kata kunci berdasarkan ukuran font (Mayer-Schönberger & Cukier, 2013).

Saat ini, ada beberapa generator *online* untuk *word cloud* seperti “Wordclouds.com” dan “jasondavies.com”. Menggunakan generator ini, pengguna dapat mengunggah file atau halaman, menyesuaikan daftar kata, font, warna atau bahkan skala dan orientasi. Selain generator *online*, paket dalam bahasa pemrograman menyediakan alat pengembangan yang jauh lebih fleksibel bagi pengguna untuk menghasilkan *word cloud* yang memuaskan sesuai dengan permintaan mereka. Di *platform* Java, R dan Python, tersedia paket untuk *word cloud*.

2.10 Evaluasi Kinerja Algoritma

Evaluasi atau pengukuran kinerja suatu algoritma dapat dilakukan menggunakan bantuan *confusion matrix*. Jurafsky dan Martin (2008) menyatakan bahwa *confusion matrix* merupakan salah satu metrik yang dapat digunakan untuk mengukur kinerja algoritma klasifikasi. Terminologi yang terkait dengan *confusion matrix* bisa jadi agak membingungkan, tetapi matriks ini sendiri mudah dipahami.

Tabel 2.6. *Confusion Matrix*

		<i>Actual</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	<i>Negative</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Perhatikan bahwa 'True' atau 'False' pada Tabel 2.6 menunjukkan bahwa kelas diprediksi dengan benar atau tidak, sedangkan 'Positif' atau 'Negatif' menunjukkan bahwa prediksi masuk ke kelas positif atau negatif. Dari Tabel 2.6, nilai *accuracy* suatu algoritma dapat diperoleh. *Accuracy* menyatakan persentase dari semua pengamatan yang diberi label dengan benar oleh suatu algoritma. Nilai *accuracy* pada *confusion matrix* dapat dihitung sebagai:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.12)$$

Dua metrik yang digunakan untuk mengukur kinerja sistem adalah *precision* dan *recall* (Dalianis, 2018). *Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif sedangkan *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Kombinasi antara *precision* dan *recall* merupakan suatu metrik tunggal yang disebut *F₁ Score* (Van Rijsbergen, 1975). *F₁ Score* menunjukkan kelayakan *precision* dan *recall* dalam melakukan klasifikasi. *F₁ Score* didefinisikan sebagai:

$$F_1 \text{ Score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (2.13)$$

2.11 Vaksin COVID-19

Vaksin COVID-19 adalah salah satu dari beberapa vaksin berbeda yang bertujuan untuk memberikan kekebalan pada tubuh terhadap penyakit *coronavirus* 2019 (COVID-19). Sebelumnya, pengembangan vaksin untuk melawan penyakit virus korona SARS dan MERS dijadikan sebagai landasan pengetahuan tentang struktur dan fungsi virus corona yang mempercepat pengembangan berbagai platform teknologi untuk vaksin COVID-19 selama awal tahun 2020. Hingga Oktober 2020, terdapat 321 kandidat vaksin yang sedang dikembangkan, meningkat 2,5 kali lipat sejak April. Namun, tidak ada kandidat yang menyelesaikan uji klinis untuk membuktikan keamanan dan kemanjurannya. Pada bulan Oktober, sekitar 42 kandidat vaksin berada dalam penelitian klinis: yaitu 33 dalam uji coba Tahap I-II dan 9 dalam uji coba Tahap

II-III.

Generasi pertama vaksin COVID-19 diharapkan mendapat persetujuan paling cepat pada akhir tahun 2020 atau awal 2021. Asumsi yang populer adalah vaksin ini akan memberikan kekebalan penduduk yang dapat mengurangi penularan sindrom pernafasan akut parah coronavirus 2 (SARS-CoV-2) dan mengarah pada dimulainya kembali "keadaan normal" pra-COVID-19.