

**PERBANDINGAN KINERJA *SUPPORT VECTOR
MACHINE* DENGAN OPTIMASI *PARTICLE SWARM*
DAN *BAYESIAN* PADA ANALISIS SENTIMEN
KEBENCANAAN DI INDONESIA**

SKRIPSI



MUHAMMAD SYAMSUL BAHRI

H051191001

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

2023

PERBANDINGAN KINERJA *SUPPORT VECTOR MACHINE* DENGAN OPTIMASI *PARTICLE SWARM* DAN *BAYESIAN* PADA ANALISIS SENTIMEN KEBENCANAAN DI INDONESIA

SKRIPSI

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

MUHAMMAD SYAMSUL BAHRI

H051191001

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

MAKASSAR

JULI 2023

LEMBAR PERNYATAAN KEOTENTIKAN

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

**Perbandingan Kinerja *Support Vector Machine*
Dengan Optimasi *Particle Swarm* Dan *Bayesian* Pada Analisis Sentimen
Kebencanaan Di Indonesia**

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 28 Juli 2023



Muhammad Syamsul Bahri

NIM H051191001

**PERBANDINGAN KINERJA *SUPPORT VECTOR*
MACHINE DENGAN OPTIMASI *PARTICLE SWARM*
DAN *BAYESIAN* PADA ANALISIS SENTIMEN
KEBENCANAAN DI INDONESIA**

Disetujui Oleh:

Pembimbing Utama

Pembimbing Pertama



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.
NIP. 197407131999032001



Siswanto, S.Si., M.Si.
NIP. 199201072019031012

Ketua Program Studi



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 197708082005012002

Pada 28 Juli 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Muhammad Syamsul Bahri
NIM : H051191001
Program Studi : Statistika
Judul Skripsi : Perbandingan Kinerja *Support Vector Machine*
Dengan Optimasi *Particle Swarm* Dan *Bayesian* Pada
Analisis Sentimen Kebencanaan Di Indonesia

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 : Sitti Sahriman, S.Si., M.Si. (.....)
4. Anggota : Anisa, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 28 Juli 2023

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Kinerja *Support Vector Machine* Dengan Optimasi *Particle Swarm* Dan *Bayesian* Pada Analisis Sentimen Kebencanaan Di Indonesia” yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya kepada kepada semua pihak yang terlibat. Oleh karena itu, dengan penuh kesadaran dan kerendahan hati, pada kesempatan ini perkenankanlah penulis menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada yang terhormat:

1. Terima kasih dan penghargaan yang setinggi-tingginya kepada ayahanda Herman dan Ibunda Sumarni atas kasih sayang, pengorbanan, kesabaran hati, dukungan penuh serta doa yang selalu dipanjatkan demi kelancaran setiap langkah dalam hidup penulis.
2. Terima kasih kepada adik tersayang Siti Nurhalizah dan Keluarga besar yang selalu memberikan dukungan dan harapan besar kepada penulis yang membuat penulis selalu termotivasi untuk terus berusaha dan pantang menyerah.
3. Terima kasih kepada Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., selaku Rektor Universitas Hasanuddin beserta seluruh staf jajarannya.
4. Terima kasih Bapak Dr. Eng. Amiruddin, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh staf jajarannya.

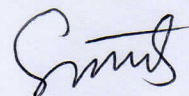
5. Terima kasih kepada Ibu Dr. Anna Islamiyati, S.Si., M.Si. selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf Departemen Statistika yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
6. Terima kasih kepada Ibu Dr. Nurtiti Sunusi S.Si., M.Si., selaku Ketua Departemen Statistika sekaligus Penasihat Akademik penulis pada masa-masa awal perkuliahan penulis yang banyak memberikan nasihat yang sangat berguna dalam kehidupan perkuliahan penulis.
7. Terima kasih kepada Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D sebagai Pembimbing Utama yang telah bersedia untuk meluangkan waktu untuk memberi masukan serta memberikan solusi atas semua permasalahan yang ada pada penelitian ini.
8. Terima kasih kepada Bapak Siswanto, S.Si., M.Si., sebagai Pembimbing Pertama sekaligus Penasihat Akademik penulis yang telah bersedia senantiasa meluangkan waktu memberikan saran, arahan serta solusi terbaik dalam proses perkuliahan hingga proses penyelesaian skripsi penulis.
9. Terima kasih kepada Ibu Sitti Sahrinan, S.Si., M.Si. dan Ibu Anisa, S.Si., M.Si. sebagai Dosen Penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
10. Terima kasih yang setulus-tulusnya kepada Muliana yang selalu ada menemani dan memberikan asupan semangat bagi penulis sehingga dapat menuntaskan masa perkuliahan dan penyusunan skripsi dengan baik.
11. Terima kasih kepada teman-teman dekat di Istana Tercinta yang setia menemani penulis. Terima kasih kanda-kanda: Agus Hermawan, Muhammad Fathurrahman, Arief Rahman Nur, Muhammad Ferdiansyah, dan Sapriadi Rasyid.
12. Terima kasih kepada kanda Fadilah Amirul Adel yang selalu memberikan cerita, pandangan, ide baru tentang kehidupan serta selalu mengingatkan penulis untuk kembali selalu fokus dalam proses penulisan skripsi.
13. Terima kasih kepada teman-teman : Seli Lisnayati, Nur Aisyah, Nurul Hikmah, Andi Muhammad Rajab, Ahmad Firdaus Irianto, Yasmin Pratiwi, Nurazizah,

Nur Syahfika dan Muhammad Rayhan Rifaldi. Terima kasih telah memberikan waktu dan cerita yang menyenangkan di masa perkuliahan penulis.

14. Terima kasih kepada semua teman-teman di angkatan Statistika 2019 yang telah mengisi masa perkuliahan penulis dengan banyak sekali cerita yang akan terus terkenang dalam ingatan penulis.
15. Terima kasih kepada seluruh teman mabar yang telah menemani penulis dalam mengisi waktu luang dan mengambil waktu tidur penulis.
16. Terima kasih yang setinggi-tingginya kepada seluruh pihak yang mungkin tidak sempat penulis sebutkan satu persatu. Terima kasih atas segala dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 28 Juli 2023



Muhammad Syamsul Bahri

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

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

Nama : Muhammad Syamsul Bahri
NIM : H051191001
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 tugas akhir saya yang berjudul:

**“Perbandingan Kinerja *Support Vector Machine*
Dengan Optimasi *Particle Swarm* Dan *Bayesian* Pada Analisis Sentimen
Kebencanaan Di Indonesia”**

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 tanggal 28 Juli 2022.

Yang menyatakan,



(Muhammad Syamsul Bahri)

ABSTRAK

Informasi yang tersebar di masyarakat memiliki peran penting dalam penanganan bencana sebagai gambaran untuk mengetahui perkembangan keadaan setelah terjadinya bencana. Data dari *twitter* yang merupakan salah satu media sosial yang banyak digunakan masyarakat Indonesia dapat menjadi sumber informasi berupa data teks yang berguna pasca bencana. Analisis sentimen memberikan gambaran dari pendapat yang terkandung dalam teks. Analisis sentimen digunakan sebagai proses pengambilan informasi dari keseluruhan teks *twitter*. *Support Vector Machine* (SVM) adalah algoritma klasifikasi yang banyak digunakan dalam analisis sentimen karena kemampuannya dalam menangani data dengan banyak dimensi seperti data teks. Model SVM memiliki parameter C dan γ yang perlu disesuaikan agar model menghasilkan kinerja maksimal. Penyesuaian parameter pada SVM dapat dilakukan menggunakan proses optimasi. Penelitian ini menggunakan dua metode optimasi yaitu Optimasi *Particle Swarm* dan Optimasi *Bayesian*. Optimasi *Particle Swarm* memiliki kemampuan eksplorasi fungsi yang tinggi dan Optimasi *Bayesian* yang mampu menangani berbagai macam bentuk sebaran data. Penelitian ini bertujuan untuk membandingkan hasil kinerja model SVM dengan Optimasi *Particle Swarm* dan *Bayesian*, pada *tweet* dengan kata kunci “bencana alam indonesia”. Hasil dari penelitian ini menunjukkan kinerja berdasarkan klasifikasi sentimen positif dan negatif. Model SVM dengan Optimasi *Particle Swarm* dan Optimasi *Bayesian* memiliki nilai *accuracy* 94,56% dan 94,26%. Model SVM dengan Optimasi *Particle Swarm* dan *Bayesian* tidak berbeda secara signifikan, namun proses optimasi meningkatkan kinerja SVM secara signifikan dibandingkan nilai *accuracy* 59,75% model SVM tanpa optimasi.

Kata Kunci: Bencana, Optimasi *Bayesian*, Optimasi Parameter, *Particle Swarm*, *Support Vector Machine*.

ABSTRACT

Information that spread in the community has important role in disaster management as illustration of the development conditions after disaster. Data from twitter, which is the social media that's widely used by the people of Indonesia, can be a source of information in the form of text data after disaster. Sentiment analysis provides an overview of the opinions contained in the text. Sentiment analysis used as a process of retrieving information from the twitter text. Support Vector Machine (SVM) is a classification algorithm that's widely used in sentiment analysis because the ability to handle multi-dimensional data such as text. The SVM model has parameters C and γ which need to be adjusted so the model produces maximum performance. Parameter adjustment in SVM can be done using the optimization. This study uses two optimization methods, Particle Swarm Optimization and Bayesian Optimization. Particle Swarm Optimization has high function exploration capabilities and Bayesian Optimization is capable of handling various forms of data distribution. This study aims to compare the performance results of the SVM model with Particle Swarm Optimization and Bayesian, on tweets with keywords "Indonesian natural disasters". Results of this study indicate performance based on classification of positive and negative sentiments. SVM model with Particle Swarm Optimization and Bayesian Optimization has accuracy value of 94.56% and 94.26%. The SVM model with Particle Swarm Optimization and Bayesian did not differ significantly, but the optimization process significantly improved SVM performance compared to the 59.75% accuracy value of the SVM model without optimization.

Keywords: *Bayesian Optimization, Disaster, Particle Swarm, Parameter Optimization, Support Vector Machine.*

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	ix
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Text Mining</i>	6
2.2 Praproses Data Teks	7
2.3 <i>Term Frequency Inverse Document Frequency</i>	8
2.4 <i>Support Vector Machine</i>	8
2.5 Optimasi <i>Particle Swarm</i>	13
2.6 Optimasi <i>Bayesian</i>	14
2.7 <i>Performance Evaluation Measure</i>	16
2.8 <i>Twitter</i>	18
2.9 Bencana Alam.....	18
BAB III METODOLOGI PENELITIAN	19
3.1 Data.....	19
3.2 Metode Analisis	20

BAB IV HASIL DAN PEMBAHASAN	22
4.1 Pelabelan Data	22
4.1.1 Proses dan Standar Pelabelan Data Positif.....	22
4.1.2 Proses dan Standar Pelabelan Data Negatif	23
4.2 Deskripsi Data	25
4.3 Praproses Data Teks	26
4.3.1 <i>Case Folding</i>	26
4.3.2 <i>Text Cleaning</i>	26
4.3.3 <i>Stemming</i>	27
4.3.4 <i>Stopword Removal</i>	28
4.3.5 <i>Tokenizing</i>	29
4.4 Visualisasi Teks dengan <i>Word Cloud</i>	30
4.5 Proses <i>Term Frequency Inverse Document Frequency</i>	30
4.6 Pembagian Data Latih dan Data Uji	34
4.7 Klasifikasi <i>Support Vector Machine</i> Tanpa Optimasi	34
4.7.1 Penentuan Nilai <i>b</i> dan <i>w</i>	34
4.7.2 Penentuan Kelas Prediksi	37
4.7.3 Hasil Klasifikasi <i>Support Vector Machine</i>	38
4.8 <i>Support Vector Machine</i> dengan Optimasi <i>Particle Swarm</i>	40
4.8.1 Optimasi Parameter	40
4.8.2 Hasil Klasifikasi <i>Support Vector Machine</i> dengan Optimasi <i>Particle Swarm</i>	41
4.9 <i>Support Vector Machine</i> dengan Optimasi <i>Bayesian</i>	43
4.9.1 Optimasi Parameter	43
4.9.2 Hasil Klasifikasi <i>Support Vector Machine</i> dengan Optimasi <i>Bayesian</i>	48
4.10 Perbandingan Kinerja	50
BAB V KESIMPULAN DAN SARAN	52
5.1 Kesimpulan	52
5.2 Saran	52
DAFTAR PUSTAKA	53
LAMPIRAN.....	59

DAFTAR GAMBAR

Gambar 2.1. Pemetaan Data ke Ruang Berdimensi Tinggi dengan <i>Kernel</i>	11
Gambar 2.2. Perbedaan <i>Polynomial Kernel</i> dan <i>RBF Kernel</i>	12
Gambar 4.1. Perbandingan Jumlah Data dengan Label Positif dan Negatif.....	25
Gambar 4.2. <i>Word Cloud</i> data positif (a) dan <i>Word Cloud</i> data negatif (b).....	30
Gambar 4.3. Grafik Perbandingan Kinerja Model SVM	50

DAFTAR TABEL

Tabel 2.1. Syarat Sentimen Positif dan Sentimen Negatif.....	6
Tabel 2.2. <i>Confusion Matrix</i>	16
Tabel 3.1. Contoh Struktur Data Penelitian	19
Tabel 4.1. <i>Dataset</i>	25
Tabel 4.2. Hasil Tahapan <i>Case Folding</i>	26
Tabel 4.3. Hasil Tahapan <i>Text Cleaning</i>	27
Tabel 4.4. Hasil Tahapan <i>Stemming</i>	28
Tabel 4.5. Hasil Tahapan <i>Stopword Removal</i>	28
Tabel 4.6. Hasil Tahapan <i>Tokenizing</i>	29
Tabel 4.7. Nilai-Nilai n_i, j dan DF_i Data Pertama.....	31
Tabel 4.8. Nilai TF-IDF Data Pertama.....	33
Tabel 4.9. Proporsi Pembagian Data Latih dan Data Uji.....	34
Tabel 4.10. Nilai α_i untuk Tiap Data	35
Tabel 4.11. Data dengan Nilai $0 < \alpha_i < C$	36
Tabel 4.12. Kelas Prediksi untuk Data Uji.....	38
Tabel 4.13. <i>Confusion Matrix</i> Model <i>Support Vector Machine</i> Tanpa Optimasi. 39	
Tabel 4.14. Proses Optimasi Parameter dengan Optimasi PSO.....	40
Tabel 4.15. <i>Confusion Matrix</i> Model SVM dengan PSO	41
Tabel 4.16. Proses Optimasi Parameter dengan Optimasi <i>Bayesian</i>	47
Tabel 4.17. <i>Confusion Matrix</i> Model SVM dengan Optimasi <i>Bayesian</i>	48
Tabel 4.18. Perbandingan Hasil Evaluasi Kinerja Model SVM	50

DAFTAR LAMPIRAN

Lampiran 1. Data <i>Tweet</i> Hasil <i>Crawling</i> dari <i>Twitter</i>	60
Lampiran 2. Vektor Numerik Hasil Pembobotan dengan TF-ID.....	61
Lampiran 3. Nilai $K(\mathbf{x}_i, \mathbf{x}_j)$ menggunakan <i>kernel</i> RBF dan $\gamma = 0.1$	61
Lampiran 4. Nilai $0 < \alpha_i < C$ pada SVM Tanpa Optimasi	62
Lampiran 5. Hasil Prediksi Data Uji dengan SVM Tanpa Optimasi	63
Lampiran 6. Proses Iterasi Optimasi Parameter SVM Menggunakan PSO	64
Lampiran 7. Hasil Prediksi Data Uji Model SVM dengan PSO	66
Lampiran 8. Proses Iterasi Optimasi Parameter SVM Menggunakan Optimasi <i>Bayesian</i>	67
Lampiran 9. Hasil Prediksi Data Uji Model SVM dengan Optimasi <i>Bayesian</i>	68

BAB I PENDAHULUAN

1.1 Latar Belakang

Kondisi geografis di Indonesia menjadikan tiga lempeng tektonik utama dunia bertemu yaitu di daerah Sumatera, Jawa, Bali Nusa Tenggara, Maluku, Sulawesi dan Papua yang merupakan salah satu penyebab terjadinya gempa bumi tektonik dan tsunami. Selain interaksi antar lempeng tektonik utama dunia, terdapat pula interaksi antar lempeng mikro sehingga sering terjadi gempa bumi tektonik dan gelombang tsunami (Tanjung dkk., 2020).

Jaringan informasi yang baik dibutuhkan saat bencana untuk mengatasi kondisi bencana yang terjadi. Salah satu media yang dapat digunakan sebagai jaringan informasi adalah media sosial, dengan adanya media sosial masyarakat dan pemerintah dapat terhubung dalam keadaan bencana. Negara Indonesia merupakan negara dengan jumlah pengguna *twitter* terbanyak kelima di dunia sebanyak 18,45 juta pengguna pada tahun 2022, jumlah tersebut setara dengan 4,23% dari total pengguna *twitter* di dunia yang mencapai 436 juta (Rizaty, 2022).

Media sosial *twitter* dapat berperan sebagai sumber informasi dalam situasi keadaan darurat bencana. Keterbatasan sumber media lain dalam memberikan informasi yang berguna saat bencana menjadi salah satu alasan penggunaan media sosial. *Twitter* dapat berfungsi sebagai media komunikasi yang berharga melalui berbagi teks dan foto selama keadaan darurat, sekaligus media sumber informasi memberikan gambaran daerah yang terdampak bencana (Pourebrahim dkk., 2019). Informasi penting dapat diperoleh menggunakan metode yang tepat dengan mengolah *tweet* pada *twitter*. Metode analisis sentimen atau *opinion mining* merupakan proses untuk mengelompokkan kecenderungan opini yang terdapat pada suatu teks, berbentuk dokumen maupun pendapat yang dikemukakan bersifat positif atau negatif (Nugraha dkk., 2020).

Analisis sentimen dapat dilakukan melalui dua pendekatan, yaitu *machine learning* dan pendekatan *lexicon* atau pendekatan berbasis kamus kata (Hamka & Sari, 2022). Kelebihan dari penggunaan *machine learning* dalam melakukan analisis adalah kemampuan algoritma untuk mempelajari dan mengolah data dengan jumlah yang sangat besar (Wahyu, 2020). Beberapa algoritma *machine*

learning yang dapat digunakan untuk sentimen analisis antara lain adalah *naive bayes*, *logitboost*, *support vector machine*, *augmented latent semantic indexing spaces*, dan *radial basis function network* (Berry & Kogan, 2010). Salah satu algoritma klasifikasi analisis sentimen teks yang sering digunakan untuk proses klasifikasi adalah algoritma *support vector machine* (Tsani dkk., 2020).

Support Vector Machine (SVM) adalah algoritma klasifikasi yang dapat diaplikasikan pada permasalahan *linear* dan *non-linear* dan merupakan salah satu model yang sering digunakan dalam *data mining*. Algoritma SVM menggunakan pemetaan *non-linear* untuk mentransformasikan latih data awal ke dimensi yang lebih tinggi selanjutnya algoritma SVM membuat garis pemisah atau *hyperplane* yang memisahkan data ke dalam masing-masing kelas. Dibandingkan dengan jenis pengklasifikasian lainnya, SVM memiliki kelebihan untuk melakukan klasifikasi dengan akurasi yang tinggi baik pada data yang digunakan dalam pemodelan pada data uji dan aktual (Mechelli & Vieira, 2019).

Performa berupa nilai *accuracy*, *precision*, *recall* dan *F-measure* dari SVM dapat ditingkatkan salah satunya dengan menetapkan nilai parameter yang sesuai pada model, namun jumlah data yang besar dapat menjadi batasan dalam menentukan parameter yang optimal. Penggunaan optimasi dapat menjadi solusi untuk menentukan parameter optimal model SVM. Optimasi *Particle Swarm* atau *Particle Swarm Optimization* (PSO) dan optimasi *bayesian* merupakan dua algoritma dari banyaknya algoritma optimasi yang dapat digunakan. PSO merupakan proses pencarian nilai menggunakan kumpulan partikel acak, setiap partikel tersebut berupa titik yang merupakan solusi dari masalah optimasi. Tujuan dari PSO adalah untuk menemukan posisi partikel yang menghasilkan nilai optimal fungsi objektif (Dista & Abdulloh, 2022). Dibandingkan dengan algoritma optimasi lainnya, pendekatan berbasis PSO bekerja lebih baik untuk mencari struktur optimum global dengan menggunakan jumlah partikel yang sedikit (Mitikiri dkk., 2018). Kemampuan eksplorasi yang tinggi dan kecepatan menemukan hasil konvergen untuk menentukan nilai optimum merupakan kelebihan dari PSO (Zhao dkk., 2021).

Optimasi *bayesian* merupakan proses optimasi dengan membangun distribusi fungsi dari sampel parameter menggunakan proses *gaussian* untuk mengetahui

wilayah distribusi mana dalam ruang parameter yang perlu ditelusuri dan mana yang tidak (Nabila, 2022). Keuntungan utama dari Optimasi *Bayesian* adalah bahwa pendekatan ini dapat menentukan jumlah data parameter yang diperlukan, sehingga memungkinkan untuk lebih efisien dalam mencari solusi optimal dengan menggunakan jumlah percobaan yang lebih sedikit dibandingkan metode optimasi lain yang membutuhkan lebih banyak perulangan (Yamawaki dkk., 2018).

Mulyani (2022) telah melakukan penelitian mengenai klasifikasi sentimen mengenai vaksinasi COVID-19 dengan membandingkan kinerja SVM dan fitur berbasis leksikon pada data *tweet* dengan topik vaksinasi Covid-19 di Indonesia dan memperoleh hasil akurasi 85% pada model SVM. Penelitian yang dilakukan Kurniawan (2022) menggunakan data serupa dengan menggunakan metode klasifikasi *naive bayes* berbasis PSO memperoleh hasil akurasi sebesar 91.28%. Penelitian lainnya yang dilakukan Xie dkk. (2021) mengenai terjadi atau tidaknya longsor pada daerah-daerah di Kota Nangpin Negara Cina menggunakan SVM dengan dan tanpa optimasi *bayesian* hasil penelitian tersebut menunjukkan bahwa SVM teroptimasi memiliki hasil akurasi sebesar 89,53% lebih baik dari SVM tanpa optimasi dengan akurasi 84,91%. Penelitian sebelumnya memperlihatkan model klasifikasi memberikan hasil kinerja yang lebih saat menggunakan optimasi sehingga pada penelitian ini akan dilakukan analisis menggunakan data teks berupa *tweet* dari *user* media sosial *twitter* yang berkaitan dengan kebencanaan yang terjadi di Indonesia lalu membandingkan kinerja antara SVM dengan PSO dan SVM dengan optimasi *bayesian*.

Proses menentukan sentimen data *tweet* yang berbentuk data teks dengan jumlah besar dan beragam serta parameter pada model SVM yang perlu disesuaikan untuk memperoleh kinerja model yang maksimal agar dapat menentukan klasifikasi sentimen yang tepat dan memberikan informasi yang berguna dari data yang digunakan, maka berdasarkan hal tersebut penelitian ini difokuskan dalam menentukan nilai parameter pada model SVM untuk memperoleh kinerja dan klasifikasi model terbaik pada data *tweet* dengan topik kebencanaan di Indonesia. Metode penentuan parameter terbaik yang digunakan adalah metode PSO dan optimasi *bayesian*. Hasil penelitian ini dapat menjadi acuan pemerintah dalam mewujudkan penanganan bencana yang lebih baik.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil klasifikasi analisis sentimen masyarakat terhadap peristiwa bencana di Indonesia menggunakan algoritma SVM dengan PSO dan SVM dengan optimasi *bayesian*?
2. Bagaimana perbandingan kinerja algoritma SVM dengan PSO dan SVM dengan optimasi *bayesian* dalam klasifikasi opini masyarakat terhadap peristiwa bencana di Indonesia?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. *Tweet* yang akan dianalisis bersumber dari aplikasi *twitter* dengan kata kunci “bencana alam indonesia” pada rentang waktu 01 Januari sampai 28 Februari 2023.
2. Sentimen yang akan diklasifikasikan terdiri dari dua kategori yaitu sentimen positif dan sentimen negatif.
3. *Kernel* yang digunakan pada algoritma SVM adalah *Radial Basis Function* (RBF).
4. Metode optimasi akan diterapkan pada parameter C dari persamaan SVM dan parameter γ (*gamma*) dari fungsi *kernel* RBF.
5. Nilai awal parameter yang digunakan sebelum dilakukan proses optimasi yaitu $C = 0.1$ dan $\gamma = 0.1$.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh hasil klasifikasi analisis sentimen masyarakat terhadap peristiwa bencana di Indonesia menggunakan algoritma SVM dengan PSO dan SVM dengan optimasi *bayesian*.
2. Memperoleh hasil perbandingan kinerja algoritma SVM dengan PSO dan SVM dengan optimasi *bayesian* dalam klasifikasi opini masyarakat terhadap peristiwa bencana di Indonesia.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Memberikan informasi tentang pemodelan algoritma SVM menggunakan PSO dan optimasi *bayesian* pada data *tweet* berkaitan bencana di Indonesia.
2. Memberikan inspirasi dan gagasan baru bagi pembaca dalam mengembangkan model klasifikasi SVM dengan optimasi pada kasus lainnya.
3. Sebagai salah satu sumber informasi pemerintah Indonesia terkait kondisi masyarakat pada keadaan bencana sehingga dapat mengambil langkah antisipatif dalam proses penanggulangan dan penanganan bencana.

BAB II

TINJAUAN PUSTAKA

2.1 *Text Mining*

Text mining merupakan proses pencarian informasi dari hasil identifikasi dan eksplorasi pada data yang berupa kumpulan dokumen, dengan data tekstual yang tidak terstruktur. Proses *text mining* bergantung pada praproses data, algoritma untuk pencarian pola, dan elemen visualisasi untuk menyempurnakan penjelajahan kumpulan informasi (Feldman & Sanger, 2006). *Text mining* dalam prosesnya terkhusus ke dalam proses pengkategorian teks dan pengelompokan teks (Ariyanti & Iswardani, 2020). Salah satu penggunaan pengkategorian teks menggunakan *text mining* adalah analisis sentimen yang merupakan metode analisis untuk mengekstrak atribut dari sebuah komentar (opini, sentimen, dan emosi) yang diekspresikan secara tekstual dalam sebuah halaman dokumen. Analisis ini dilakukan untuk melihat opini atau kecenderungan opini (cenderung positif atau negatif) seseorang terhadap sebuah masalah atau objek (Syahadati dkk., 2021).

Menurut Kamus Besar Bahasa Indonesia (KBBI) sentimen positif merupakan reaksi atau sikap yang meningkatkan nilai seseorang atau sesuatu sedangkan sentimen negatif merupakan reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu, jadi kalimat bersentimen negatif akan menyebabkan penyurutan nilai pandang terhadap sesuatu (Ardiani dkk., 2020). Kalimat dengan sentimen positif memiliki perasaan suka terhadap sesuatu sementara sebaliknya untuk kalimat dengan sentimen negatif (Fararo, 1992). Sebuah ungkapan dikategorikan bersentimen negatif apabila ungkapan tersebut merupakan ekspresi ketidaksenangan atau emosi yang tidak mengenakkan terhadap subjek tertentu sementara untuk sentimen positif berupa ungkapan sebaliknya (Fujita dkk., 2020). Berdasarkan rujukan tersebut, maka syarat sebuah teks dikatakan memiliki sentimen positif atau negatif dapat dilihat pada Tabel 2.1.

Tabel 2.1. Syarat Sentimen Positif dan Sentimen Negatif

Sentimen Positif	Sentimen Negatif
Reaksi atau sikap yang meningkatkan nilai seseorang atau sesuatu	Merupakan reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu

Tabel 2.1. Syarat Sentimen Positif dan Sentimen Negatif

Sentimen Positif	Sentimen Negatif
Berisi ungkapan dengan perasaan suka terhadap sesuatu	Berisi ungkapan dengan perasaan tidak suka terhadap sesuatu
Ungkapan yang mengekspresikan emosi yang baik atau mengesankan terhadap suatu subjek	Ekspresi ketidaksenangan atau emosi yang tidak mengesankan terhadap subjek

2.2 Praproses Data Teks

Praproses data teks merupakan proses yang bertujuan untuk mempersiapkan data sebelum dilakukan pelatihan pada model, data akan diolah sehingga dapat digunakan dalam model (Wijaya & Wibawa, 2022). Tahapan dalam praproses data teks adalah sebagai berikut:

1. *Case Folding*

Penyamaan bentuk karakter dalam sebuah dokumen, karena tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. *Case folding* memiliki peran untuk menyeragamkan teks (Alita & Auliya, 2020).

2. *Text Cleaning*

Proses untuk menghapus karakter berupa tanda baca, angka simbol dan karakter spesial lainnya dari dokumen (Aljedani dkk., 2021).

3. *Stemming*

Stemming adalah proses pemotongan imbuhan (*affix*), baik *prefix* maupun *suffix* dari sebuah *term* untuk mendapatkan kata dasar (*root* atau *stem*) dari kata berimbuhan (Wirayasa dkk., 2019).

4. *Stopword Removal*

Proses menghapus kata dari model yang tidak berkontribusi banyak pada model, kata yang termasuk ke dalam *stopword* dihilangkan dari dokumen agar tidak mempengaruhi kinerja model (Alita & Auliya, 2020).

5. *Tokenizing*

Proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token atau bagian-bagian terpisah berupa kata atau kalimat tunggal. Pada saat bersamaan *tokenizing* juga berfungsi untuk membuang beberapa karakter tertentu yang dianggap sebagai tanda baca (Alita & Auliya, 2020).

2.3 *Term Frequency Inverse Document Frequency*

Term Frequency-Inverse Document Frequency (TF-IDF) adalah statistik numerik yang menunjukkan seberapa penting sebuah kata bagi sebuah koleksi dokumen. Metode ini sering digunakan sebagai faktor pembobot dalam *text mining*. TF-IDF meningkat secara proporsional dengan berapa kali sebuah kata muncul dalam dokumen, tetapi diimbangi dengan frekuensi kata dalam kumpulan dokumen, yang membantu mengontrol fakta bahwa beberapa kata lebih umum daripada yang lain (Christian dkk., 2016). Berikut merupakan proses untuk menghitung nilai TF-IDF dari kata dalam dokumen (Chen dkk., 2019):

$$W_{i,j} = TF_i \times \left(\ln \left(\frac{N + 1}{DF_i + 1} \right) + 1 \right) \quad (2.1)$$

Keterangan:

$W_{i,j}$: bobot TF-IDF dari kata i pada dokumen j

N : jumlah total dokumen

DF_i : jumlah dokumen yang mengandung kata i

$TF_{i,j}$: jumlah kemunculan kata i pada dokumen j

Nilai TF-IDF untuk setiap dokumen kemudian dinormalisasi dengan normalisasi $\frac{v}{\|v\|}$. Normalisasi dilakukan agar pengaruh nilai yang besar pada kata-kata yang jarang muncul tidak mendominasi nilai keseluruhan pada dokumen, dan pengaruh nilai yang kecil pada kata-kata yang sering muncul tidak diabaikan (Gambardella & Bernardo, 2019).

2.4 *Support Vector Machine*

Support Vector Machine (SVM) merupakan metode *machine learning* yang bekerja dengan menyeimbangkan kompleksitas model pada data yang diklasifikasikan. SVM bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada kumpulan observasi. Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi *kernel* dan parameter yang digunakan (Parapat, 2017).

Hyperplane terbaik adalah *hyperplane* yang terletak di tengah dua data poin

dari masing-masing kelas, usaha dalam mencari lokasi *hyperplane* adalah inti dari proses pembelajaran SVM. Berdasarkan dari karakteristiknya, metode SVM dibagi menjadi dua, yaitu *linear SVM* untuk data yang dapat dipisahkan secara *linear* kedua oleh *hyperplane*. Sedangkan *non-linear SVM* yang menerapkan fungsi dari *kernel trick* terhadap data pada ruang yang berdimensi tinggi (Parapat, 2017).

Misalkan data untuk membangun model atau data latih berdimensi p dinotasikan sebagai $\mathbf{x}_i \in \mathbb{R}^p$ sedangkan label kelas dinyatakan $y \in \{-1, +1\}$. Model *linear* untuk *hyperplane* yang umum dipakai dalam metode SVM dapat dituliskan sebagai berikut (Casuarina dkk., 2022):

$$y_i = \mathbf{w}^T \mathbf{x}_i + b, i = 1, 2, \dots, n \quad (2.2)$$

Keterangan:

y_i : label kelas dari data ke- i

\mathbf{w} : $[w_1, w_2, \dots, w_p]$ merupakan vektor bobot yang tegak lurus terhadap *Hyperplane* berupa vektor kolom berukuran $p \times 1$ dan q merupakan banyaknya variabel bebas

\mathbf{x}_i : $[x_{i1}, x_{i2}, \dots, x_{ip}]$ merupakan vektor kolom berukuran $p \times 1$

b : *error* atau bias yang berupa skalar

n : banyaknya data

Tujuan utama dari algoritma SVM adalah untuk mencari *hyperplane* yang dapat memisahkan kedua kelas dengan *margin* maksimal (Parapat, 2017). *Margin* merupakan jarak dari kedua kelas, semakin besar *margin* maka semakin rendah kesalahan generalisasi (*error*) pengklasifikasian (Ghosh dkk., 2019). Menurut Suthaharan (2016) jarak *margin* pada SVM memiliki persamaan sebagai berikut:

$$d = \frac{2}{\|\mathbf{w}\|} \quad (2.3)$$

Memaksimalkan $\frac{2}{\|\mathbf{w}\|}$ ekuivalen dengan meminimalkan nilai $\frac{1}{2} \|\mathbf{w}\|^2$, nilai $\frac{1}{2}$ ditambahkan dan fungsi dikuadratkan bertujuan untuk mempermudah perhitungan dan menghilangkan tanda akar namun tetap memberikan hasil yang sama sehingga diperoleh permasalahan optimasi sebagai berikut (Kowalczyk, 2017):

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.4)$$

dengan kendala

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0, i = 1, 2, \dots, n$$

Persamaan (2.4) digunakan pada kasus data yang dapat dipisahkan secara *linear* namun untuk mengklasifikasi dua kelas yang tidak terpisah secara *linear*, dapat dilakukan dengan menambahkan *slack variable* (ξ_i) pada Persamaan (2.4) dengan nilai $\xi_i \geq 0$ sehingga diperoleh permasalahan (Casuarina dkk., 2022):

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (2.5)$$

dengan kendala

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i \geq 0$$

$$\xi_i \geq 0, \text{ untuk setiap } i = 1, 2, \dots, n$$

Parameter C memberikan kontrol antara optimasi jarak *margin* dan klasifikasi ξ , nilai C akan menjadi penentu seberapa penting nilai ξ . Semakin besar nilai C maka semakin besar penalti terhadap kesalahan klasifikasi (Kowalczyk, 2017).

Menurut Scholkopf dan Smola (2002) permasalahan optimasi pada Persamaan (2.5) dapat diselesaikan jika diubah ke dalam bentuk *wolfe dual lagrangian function*, sehingga diperoleh bentuk permasalahan optimasi sebagai berikut:

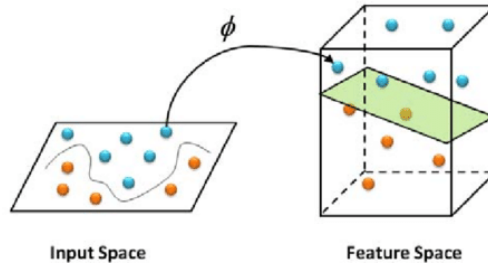
$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (2.6)$$

dengan kendala

$$\alpha_i \geq 0, \text{ untuk setiap } i = 1, 2, \dots, n$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Kasus data dunia nyata pada umumnya bersifat *non-linear* dan prinsip dasar SVM adalah *linear classifier*, sehingga meskipun menambahkan *slack variabel* ke dalam model masih belum cukup untuk memperoleh kinerja terbaik model SVM. Oleh karena itu, dimensi data dimodifikasi sehingga SVM dapat menangani *non-linear classifier* pada ruang berdimensi lebih tinggi. Proses memetakan himpunan data dari *input space* ke dalam *feature space* yang berdimensi lebih tinggi dapat diformulasikan dengan fungsi $\Phi : R^p \rightarrow R^q$ dengan nilai $p < q$. Proses pemetaan diilustrasikan pada Gambar 2.1 (Casuarina dkk., 2022).



Gambar 2.1. Pemetaan Data ke Ruang Berdimensi Tinggi dengan *Kernel*

Pada dasarnya proses pemodelan SVM merupakan proses menemukan *support vector* hasil *dot product* dari data pada *feature space*, yaitu $\Phi_i \Phi_j$. Perhitungan *dot product* dapat digantikan dengan fungsi *kernel* $K(\mathbf{x}_i, \mathbf{x}_j)$ yang mendefinisikan secara implisit hasil *dot product* transformasi Φ tersebut. Inilah yang disebut *kernel trick*, yang dinotasikan sebagai berikut (Fitriyah dkk., 2020):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi_i(\mathbf{x}_i) \cdot \Phi_j(\mathbf{x}_j)$$

Berikutnya dengan memasukkan prinsip *kernel trick* ke permasalahan optimasi pada Persamaan (2.6) maka kita akan memperoleh persamaan sebagai berikut:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (2.7)$$

dengan kendala

$$0 \leq \alpha_i \leq C, \text{ untuk setiap } i = 1, \dots, n$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Penyelesaian permasalahan optimasi menghasilkan nilai α_i yang berbeda untuk tiap observasi. Observasi yang memiliki nilai $0 < \alpha < C$ dikategorikan sebagai *support vector* dan digunakan untuk menghitung nilai \mathbf{w} dan b dengan persamaan berikut (Bishop, 2006):

$$b = \frac{1}{s} \sum_{k=1}^s (y_k - \mathbf{w} \cdot \Phi(\mathbf{x}_k)) \quad (2.8)$$

dengan

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \Phi(\mathbf{x}_i) \quad (2.9)$$

Keterangan:

\mathbf{x}_k : data latih yang termasuk *support vector* dan $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kq}]$

dan merupakan *vector* kolom $q \times 1$

- y_k : label kelas dari himpunan data
- α_k : α yang termasuk *support vector*
- s : jumlah *support vector*
- n : jumlah data latih

Menurut Casuarina dkk. (2022), kelas dari suatu data uji dapat ditentukan dengan persamaan sebagai berikut:

$$f(x) = \sum_{k=1}^s \alpha_k y_k k(x_k, x) + b \tag{2.10}$$

dengan $f(x) \geq 0$ merupakan kelas positif dan $f(x) < 0$ merupakan kelas negatif.

Menurut Kowalczyk (2017) terdapat beberapa jenis *kernel* yang bisa digunakan dalam prose klasifikasi SVM, yaitu sebagai berikut:

1. *Linear Kernel*

Linear Kernel merupakan *kernel* yang paling sederhana, dengan persamaan:

$$K(x_i, x_j) = x_i \cdot x_j \tag{2.11}$$

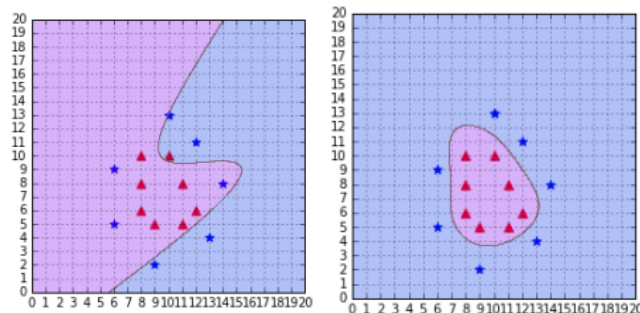
2. *Polynomial kernel*

Polynomial kernel merupakan *kernel* dengan dua parameter c yang merepresentasikan nilai konstan, dan d yang merepresentasikan nilai derajat dari *kernel*. *Polynomial kernel* memiliki persamaan:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \tag{2.12}$$

3. *Radial Basis Function Kernel*

Terdapat jenis data yang sulit dipisahkan meskipun menggunakan *polynomial kernel*. Misalnya pada data yang ditunjukkan Gambar 2.2 terlihat bahwa *kernel* RBF memiliki hasil yang lebih baik.



Gambar 2.2. Perbedaan *Polynomial Kernel* dan RBF *Kernel*

RBF *kernel* memiliki parameter γ (*gamma*) yang jika nilai *gamma* sangat kecil sifat *kernel* RBF hampir sama dengan *linear kernel*, sementara jika nilai *gamma* besar maka model SVM sangat dipengaruhi oleh tiap *support vector*. persamaan RBF *kernel* dituliskan sebagai berikut:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (2.13)$$

RBF *kernel* lebih baik digunakan dalam pembentukan model SVM pada data dengan dimensi tinggi. RBF *kernel* memungkinkan SVM untuk melakukan pemetaan fitur ke ruang dimensi yang lebih tinggi, yang memungkinkan SVM untuk menyelesaikan masalah klasifikasi yang tidak terpisah secara linier di ruang fitur asli. Menggunakan RBF *kernel*, SVM dapat menemukan fungsi hipotesis yang lebih kompleks dan dapat menangani masalah yang lebih kompleks. RBF *kernel* memiliki sifat universal approximation, artinya dengan memilih parameter yang tepat, kernel ini dapat memetakan data ke hampir semua fungsi yang kontinu. Ini berarti SVM dengan kernel RBF dapat dengan presisi tinggi memodelkan fungsi hipotesis yang rumit dan mendekati distribusi data (J. Wang dkk., 2004).

2.5 Optimasi *Particle Swarm*

Optimasi *Particle Swarm* (PSO) adalah algoritma stokastik berbasis populasi, yang diusulkan oleh Kennedy dan Eberhart pada tahun 1995 dan telah digunakan secara luas dalam berbagai masalah optimasi. Algoritma PSO meniru perilaku kawanan burung dan kawanan ikan. Setiap partikel dianggap sebagai titik tanpa volume dalam ruang pencarian (Mo dkk., 2017). Proses PSO diawali dengan populasi yang terdiri dari sejumlah partikel dengan posisi (koordinat) acak. Setiap iterasi selanjutnya akan memperbarui posisi dan kecepatan setiap partikel dalam proses untuk mencari solusi yang lebih baik (Wardhani dkk., 2018).

Setiap partikel mengubah arah pencariannya berdasarkan dua faktor, yaitu pengalaman terbaiknya sendiri sebelumnya dan solusi terbaik dari semua anggota lainnya untuk menemukan solusi optimal. Misalkan n partikel membentuk populasi $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ dalam ruang berdimensi M . Posisi dari partikel j pada iterasi ke d dalam ruang dimensi M dapat dinyatakan sebagai $\mathbf{x}_j^d = \{x_{j1}^d, x_{j2}^d, \dots, x_{jM}^d\}$. Demikian dengan kecepatan atau *velocity* yang juga merupakan vektor berdimensi M untuk partikel j pada iterasi ke d dinyatakan sebagai $\mathbf{v}_j^d = \{v_{j1}^d, v_{j2}^d, \dots, v_{jM}^d\}$.

Solusi optimal yang diperoleh partikel j pada iterasi ke d atau *personal best* dapat dinyatakan sebagai $\mathbf{p}_j^d = \{p_{j1}^d, p_{j2}^d, \dots, p_{jM}^d\}$ dan solusi optimal yang diperoleh dari \mathbf{p}_j^d atau *global best* pada iterasi ke d dinyatakan sebagai $\mathbf{p}_g^d = \{p_{g1}^d, p_{g2}^d, \dots, p_{gM}^d\}$.

Dalam setiap proses iterasi, posisi dan kecepatan setiap partikel diperbarui sesuai dengan nilai *personal best* dan *global best* yang diperoleh dari *fitness function* yang ingin dimaksimalkan. Adapun persamaan untuk pembaruan posisi partikel adalah sebagai berikut (Namrata dkk., 2021):

$$x_{jm}^{d+1} = x_{jm}^d + v_{jm}^{d+1} \quad (2.14)$$

dengan

$$v_{jm}^{d+1} = \left(1, 1 - \frac{p_{gm}^d}{p_{jm}^d}\right) v_{jm}^d + c_1 r_1 (p_{jm}^d - x_{jm}^d) + c_2 r_2 (p_{gm}^d - x_{jm}^d)$$

Nilai $m = 1, 2, \dots, M$, d adalah iterasi saat ini, c_1 dan c_2 adalah konstanta yang disebut koefisien percepatan dengan nilai yang disarankan yaitu $c_1 = 0,7$ dan $c_2 = 1.43$ (Clerc, 2006). r_1 dan r_2 adalah bilangan acak yang mengikuti sebaran acak dengan interval $(0,1)$ dengan x_{jm}^d dan v_{jm}^d adalah posisi dan *velocity* sebelumnya (Gao & Huang, 2017).

2.6 Optimasi Bayesian

Algoritma optimasi *bayesian* merupakan proses optimasi yang terdiri dari tiga komponen utama yaitu model proses *gaussian* untuk fungsi tujuan $f(\mathbf{x})$, proses pembaruan *bayesian* dengan memodifikasi model *gaussian* setelah setiap proses evaluasi dari fungsi tujuan, dan fungsi akuisisi atau *acquisition function*. Fungsi akuisisi ini dimaksimalkan untuk mengidentifikasi titik evaluasi selanjutnya. Peran fungsi ini adalah untuk mengukur peningkatan yang diharapkan dalam fungsi tujuan (Injadat dkk., 2018). Titik pengambilan sampel pada fungsi yang tidak diketahui diperlakukan sebagai variabel acak dari model proses *gaussian*, dengan persamaan sebagai berikut:

$$f(\mathbf{x}) \sim GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}^*)) \quad (2.15)$$

Model proses *gaussian* ditentukan oleh nilai *mean* fungsi $\mu(\mathbf{x})$ dan fungsi kovarians $k(\mathbf{x}, \mathbf{x}^*)$. Biasanya dalam proses *gaussian*, fungsi rata-rata dapat

diasumsikan nol tanpa kehilangan kinerja dari model, sehingga membuat proses gaussian ditentukan oleh fungsi kovarians. Secara umum fungsi kovarians yang digunakan dalam proses gaussian adalah fungsi kovarians eksponensial, dengan persamaan:

$$k(\mathbf{x}, \mathbf{x}^*) = \exp\left(-\frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|^2\right) \quad (2.16)$$

Setiap titik sampel dari fungsi objektif dinyatakan dengan $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$ dengan nilai fungsi yang sesuai $\{y_1, y_2, \dots, y_t\}$ dengan $y_{1:t} = f(\mathbf{x}_{1:t})$. Dalam setiap iterasi pengoptimalan yang dilakukan, menggunakan data sampel untuk proses gaussian diperoleh distribusi *posterior* dari model pengoptimalan. Setiap iterasi menghasilkan titik sampel baru \mathbf{x}_{t+1} menggunakan fungsi akuisisi (*acquisition function*) untuk mengevaluasi nilai $y_{t+1} = f(\mathbf{x}_{t+1})$. Nilai $y_{1:t}$ dan y_{t+1} termasuk dalam distribusi gaussian, dengan persamaan sebagai berikut (Wang dkk., 2019):

$$\begin{bmatrix} y_{1:t} \\ y_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) \end{bmatrix}\right) \quad (2.17)$$

dengan

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_t) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_t, \mathbf{x}_1) & \dots & k(\mathbf{x}_t, \mathbf{x}_t) \end{bmatrix}$$

$$\mathbf{k} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_{t+1}) \\ k(\mathbf{x}_2, \mathbf{x}_{t+1}) \\ \vdots \\ k(\mathbf{x}_t, \mathbf{x}_{t+1}) \end{bmatrix}$$

Distribusi *posterior* dari fungsi objektif dapat dituliskan sebagai berikut:

$$P(y_{t+1} | \mathbf{x}_{1:t}, y_{1:t}) \sim N(\mu(\mathbf{x}_{t+1}), \sigma_t^2(\mathbf{x}_{t+1})) \quad (2.18)$$

dengan

$$\mu(\mathbf{x}_{t+1}) = \mathbf{k}^T \mathbf{K}^{-1} y_{1:t}$$

$$\sigma_t^2(\mathbf{x}_{t+1}) = k(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k}$$

Peran fungsi akuisisi digunakan untuk menghitung nilai potensial maksimal dari fungsi tujuan. Nilai maksimal dari fungsi akuisisi digunakan untuk memilih titik terbaik berikutnya untuk mengevaluasi fungsi objektif. Fungsi akuisisi berbasis ekspektasi peningkatan atau *Expected Improvement* ($EI(x)$) dinyatakan dengan persamaan berikut (Kouziokas, 2020):

$$EI(x) = \begin{cases} (\mu(x) - f(x^+))\phi(Z) + \sigma(x)\varphi(Z), & \text{jika } \sigma(x) > 0 \\ 0, & \text{jika } \sigma(x) = 0 \end{cases} \quad (2.19)$$

dengan

$$Z = \begin{cases} \frac{\mu(x) - f(x^+)}{\sigma(x)}, & \text{jika } \sigma(x) > 0 \\ 0, & \text{jika } \sigma(x) = 0 \end{cases}$$

Keterangan:

$\mu(x)$: mean dari prediksi fungsi *posterior* proses *gaussian* pada x

$\sigma(x)$: standar deviasi dari prediksi fungsi *posterior* proses *gaussian* pada x

$\phi(Z)$: *cumulative distribution function* dari distribusi normal standar nilai Z

$\varphi(Z)$: *probability density function* dari distribusi normal standar nilai Z

x^+ : sampel dengan nilai $f(x)$ paling maksimal

2.7 Performance Evaluation Measure

Performance Evaluation Measure (PEM) atau pengukuran evaluasi performa adalah suatu tahapan yang digunakan untuk mengukur performa suatu model. Ada banyak metode perhitungan untuk mendapatkan nilai PEM, beberapa perhitungan dalam PEM antara lain *precision*, *accuracy*, *recall*, dan *F-measure* (Imron, 2019). Evaluasi performa model dapat dilakukan dengan terlebih dahulu menyusun *confusion matrix*. *Confusion matrix* adalah istilah mendasar dalam pembuatan model pembelajaran mesin untuk menganalisis bagaimana kinerja sebuah model melakukan klasifikasi kelas pada data dengan membandingkan nilai prediksi terhadap nilai sebenarnya (Zeng, 2019). Tabel 2.2 menunjukkan bentuk tabel *confusion matrix*.

Tabel 2.2. *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Negatif	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

True positive adalah total jumlah data yang diprediksi dalam kelas positif dan secara aktual (nilai sebenarnya) masuk dalam kelas positif, *false positive* yaitu total jumlah data yang diprediksi dalam kelas positif dan secara aktual masuk dalam kelas negatif, *true negative* yaitu total jumlah data yang diprediksi dalam kelas

negatif dan secara aktual masuk dalam kelas negatif, lalu *false negative* yaitu total jumlah data yang diprediksi dalam kelas negatif dan secara aktual (nilai sebenarnya) masuk dalam kelas positif. Nilai hasil *confusion matrix* digunakan untuk evaluasi performa sebagai berikut (Imron, 2019):

1. *Accuracy*

Accuracy adalah perbandingan antara nilai data yang dijawab oleh model dengan benar dibandingkan dengan keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.20)$$

2. *Precision*

Precision adalah rasio kelas yang benar diprediksi positif dibandingkan dengan keseluruhan prediksi positif.

$$Precision = \frac{TP}{FP + TP} \times 100\% \quad (2.21)$$

3. *Recall*.

Recall adalah rasio jumlah data yang benar diprediksi positif dibandingkan dengan keseluruhan data kelas positif, *recall* menggambarkan ukuran ketepatan antara informasi yang benar dengan informasi yang diperkirakan. Berikut merupakan persamaan untuk menghitung *recall*:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.22)$$

4. *F-measure*

F-measure merupakan ukuran evaluasi yang mengkombinasikan antara *precision* dan *recall*. Nilai *F-measure* menampilkan ukuran timbal balik antara *recall* dan *precision*. Perhitungan *F-measure* dapat dilihat pada persamaan dibawah:

$$F - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (2.23)$$

Nilai uji performa dikatakan semakin baik jika Nilai *Accuracy*, *Precision*, *Recall* dan *F-measure* mendekati angka 1, semakin mendekati Nilai *Accuracy*, *Precision*, *Recall* dan *F-measure* maka model semakin baik. Pada permasalahan dengan tingkat skala nasional atau regional, nilai evaluasi 80% dapat diterima

namun pada tingkat lokal atau pengambilan keputusan permasalahan yang lebih detail dibutuhkan nilai evaluasi yang lebih tinggi (Olson, 2008). Kinerja dari sebuah model memiliki perbedaan signifikan bergantung dari tingkat evaluasi model, tingkat evaluasi model 50 – 60% memberikan kinerja terendah, 60 – 90% memberikan peningkatan kinerja yang tidak terlalu signifikan dan evaluasi 90 – 98% memberikan perbedaan yang signifikan pada kinerja model (Allan dkk., 2005).

2.8 *Twitter*

Twitter adalah *platform* media sosial untuk komunikasi *online* dengan media komputer, yang memunculkan sebuah struktur sosial. *Platform* komunikasi ini memiliki 1,3 miliar akun dan 336 juta pengguna aktif yang memposting 500 juta *tweet* per hari. Pengguna *twitter* dapat memposting komentar yang dikenal sebagai “*tweet*” masing-masing dibatasi hingga 140 karakter sebelum Oktober 2018 dan saat ini, 280 karakter. *Tweet* tersebut tersedia untuk umum dan pengguna *Twitter* dapat menunjukkan reaksi dan keterlibatan mereka dengan *tweet* dengan membagikannya di profil masing-masing (Karami dkk., 2020).

2.9 **Bencana Alam**

Bencana menurut Undang-Undang Nomor 24 Tahun 2007 tentang penanggulangan bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan, baik oleh faktor alam dan/atau faktor non alam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis (Hardiyanto & Pulungan, 2019). Bencana alam merupakan serangkaian peristiwa yang disebabkan oleh alam, antara lain berupa gempa bumi, tsunami, gunung meletus, banjir, kekeringan, angin topan, dan tanah longsor (Maulida, 2019).

Besarnya dampak bencana alam bagi kehidupan masyarakat, kesiapsiagaan masyarakat harus segera diwujudkan agarantisipasi masyarakat terhadap dampak bencana alam dapat dilakukan. Intensitas bencana yang semakin meningkat tidak terlepas dari pengaruh lingkungan strategis global dan regional berupa pemanasan global (*global warming*) yang menyebabkan peningkatan suhu, perubahan iklim, meningkatnya permukaan air laut, dan perubahan ekologi yang memberikan pengaruh besar kepada peluang terjadinya bencana alam (Pudjiastuti, 2019).