

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Analisis Sentimen

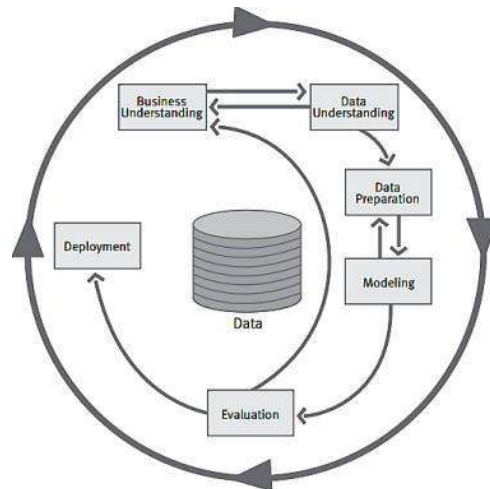
Analisis sentimen termasuk salah satu bidang ilmu dari data mining yang bisa digunakan untuk memproses, menganalisis dan mengekstrak data berupa teks yang biasa diambil dari internet yang berkaitan dengan individu, lingkungan, sosial, budaya, Pendidikan, atau masalah dan topik tertentu. Tujuan dari dilakukannya analisis sentimen yakni untuk mendapatkan informasi mengenai data mengandung sentimen positif, negatif, atau netral.

Teknik klasifikasi sentimen dapat dilakukan menggunakan 3 teknik yakni diantaranya hybrid approach, lexicon based, dan machine learning. Pada saat ini penelitian mengenai analisis sentimen umumnya menggunakan Teknik machine learning karena mampu memprediksi polaritas sentimen menjadi 3 kelas yakni positif, negatif, atau netral.

2.1.2 CRISP-DM

Metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang dirancang untuk memberikan *blueprint* berupa tahapan proses pengumpulan data. CRISP-DM dibagi menjadi 6 tahap yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. Pada gambar 2.1 menunjukkan tahapan kerja proses data mining. Setiap arah panah

menunjukkan siklus dan menggambarkan bahwa data mining dapat menghasilkan pengetahuan, pemahaman dan solusi bisnis.



Gambar 2. 1 Metode CRISP-DM

Tahapan CRISP-DM:

1) *Business Understanding*

Pada tahap awal pemahaman data atau *business understanding* berfokus pada pemahaman tujuan penelitian berdasarkan perspektif bisnis yaitu mengklasifikasikan sentimen masyarakat terhadap Pemekaran Papua.

2) *Data Understanding*

Pada tahap pemahaman data atau data understanding dimulai dengan pengumpulan data awal kemudian mengidentifikasi dan mengeksplorasi data.

3) *Data Preparation*

Pada tahap menyiapkan data atau data *preparation* meliputi Menyusun data mentah sehingga dapat digunakan dalam pemodelan dengan melakukan pemilahan, pembersihan, penyusunan dan pengintegrasian data.

4) *Modelling*

Pada tahap pemodelan merupakan proses penerapan teknik atau algoritma berdasarkan jenis data yang dipilih. Tahapan pemodelan diantaranya pemilihan teknik, pembuatan, dan penilaian model.

5) *Evaluation*

Tahap evaluasi dilakukan untuk melakukan pengujian performa model dari algoritma *Support Vector Machine* (SVM). Evaluasi dilakukan dengan menggunakan confusion matrix. Confusion matrix merupakan tools atau alat yang digunakan untuk mengukur nilai accuracy, precision, recall, dan f-score [7]. Kemudian Grafik ROC yang digunakan untuk mengilustrasikan, menentukan proses klasifikasi, dan mengevaluasi akurasi klasifikasi secara visual. Grafik ROC adalah grafik 2 dimensi dengan skala false positive pada sumbu x atau garis horizontal dan true positive pada sumbu y atau garis vertikal. Kurva ROC merupakan alternatif lain untuk menguji kinerja proses klasifikasi atau mengukur nilai AUC. Menurut Gorunescu performa keakurasian AUC di klasifikasikan menjadi [8]:

0.90 – 1.00 = *Excellent Classification*

0.80 – 0.90 = *Good Classification*

$0.70 - 0.80 = \textit{Fair Classification}$

$0.60 - 0.70 = \textit{Poor Classification}$

$0.50 - 0.60 = \textit{Failure Classification}$

6) *Deployment*

Pada tahap *deployment* melakukan penyusunan proses data mining yang telah dilakukan dalam bentuk laporan untuk menentukan Langkah yang diambil selanjutnya.

2.1.3 *Pre-Processing*

Pre-processing merupakan proses pemeriksaan data yang bersifat tidak terorganisasi menjadi terorganisasi, sehingga hal tersebut membantu dalam pemrosesan data. Pada data yang dikumpulkan dari media sosial tidak semua dapat diproses dengan baik, pada kumpulan data tersebut tidak menutup kemungkinan terdapat data yang sulit terbaca oleh sistem seperti kata tidak baku, mengandung Bahasa daerah atau penulisan kata yang disingkat untuk itu diperlukan tahap *pre-processing* atau normalisasi untuk mengatasi permasalahan tersebut.

Tahapan *pre-processing* terdiri dari 6 tahap:

1) *Cleansing*

Menghilangkan tanda yang tidak diperlukan seperti titik (.), koma (,), tanda petik ("), tanda seru (!), tanda tanya (?).

2) *Case Folding*

Setelah melalui tahap *Cleansing* karakter alfabet akan diubah ke huruf kecil (*lower case*).

3) *Tokenizing*

Memisahkan setiap kata yang tergabung dalam kalimat berdasarkan spasi biasa disebut *term* atau *token*.

4) *Normalization atau Konversi Slangword*

Mengubah kata tidak baku menjadi baku sesuai Kamus Besar Bahasa Indonesia (KBBI) dengan cara memeriksa kata tersebut terdapat kamus *slangword* atau tidak.

5) *Filtering atau Stopword Removing*

Menghapus kata yang tidak penting seperti kata penghubung yang, atau, dengan, di, tetapi.

6) *Stemming*

Mengubah kata berimbuhan menjadi kata dasar.

2.1.4 *Term Weighting*

Term weighting adalah tahapan pembobotan kata untuk meningkatkan kemampuan analisis sentimen pada proses *text mining*. Pada penelitian ini akan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), yang akan diimplementasikan dengan *tools* RapidMiner dengan operator *Process Document from Data*.

TF-IDF berperan untuk mengubah data yang telah melalui tahap *pre-processing* menjadi numerik. Untuk menghitung bobot menggunakan TF-IDF terlebih dahulu hitung nilai TF per kata dengan bobot per kata adalah 1. Sedangkan

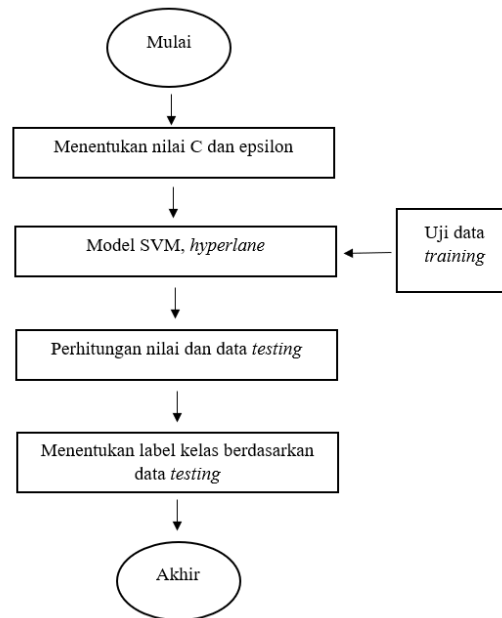
untuk menghitung nilai IDF dari setiap kata yang dicari dapat dilakukan dengan rumus :

$$IDF(word) = \log \frac{td}{df} \dots \dots \dots (1)$$

Td adalah jumlah seluruh dokumen sedangkan df adalah jumlah munculnya suatu kata pada dokumen.

2.1.5 *Support Vector Machine (SVM)*

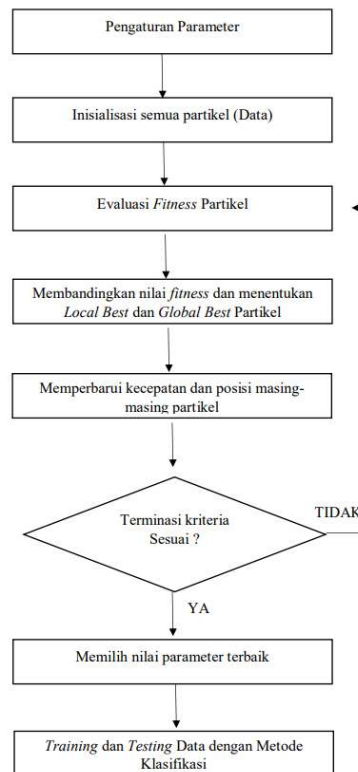
Support Vector Machine (SVM) merupakan suatu metode yang relative baru dan memiliki kemampuan lebih baik dari algoritma klasifikasi lain dalam memisahkan data karena adanya fungsi kernel trick yang mentransformasi data ke ruang dimensi lebih tinggi. Selain itu SVM mampu menjadi solusi seperti pada masalah *over-fitting*, *optimal local*, dan rasio konvergensi rendah. Kinerja SVM sangat bergantung pada pemilihan parameter yang sesuai namun salah satu kelemahan SVM yakni sulitnya pemilihan parameter SVM yang optimal. Tahap awal SVM yakni mengkonversi data teks ke bentuk data vector untuk kemudian digabungkan TF-IDF untuk pembobotan.



Gambar 2. 2 Flowchart SVM

2.1.6 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) dapat digunakan sebagai solusi untuk menyelesaikan permasalahan optimasi pada aplikasi di berbagai bidang [5]. Secara singkatnya PSO merupakan metode optimasi untuk memodifikasi parameter. Cara yang ditempuh untuk menggunakan metode PSO dapat dilakukan dengan melakukan seleksi atribut (*attribute selection*) dan *feature selection*, meningkatkan attribute (*attribute weight*) pada attribute yang akan digunakan [9]. Penerapan PSO pada *Support Vector Machine (SVM)* dapat mendukung pengambilan keputusan serta merupakan metode yang cukup baik dalam mengklasifikasikan data [10]. Kelebihan yang dimiliki metode PSO yakni sederhana, mudah diterapkan, dan kecepatan konvergensinya.



Gambar 2. 3 Penerapan PSO

2.1.7 Genetic Algorithm (GA)

Optimasi merupakan proses pemecahan suatu masalah dengan tujuan mencapai posisi paling menguntungkan berdasarkan suatu sudut pandang. *Genetic Algorithm* (GA) dapat diterapkan pada SVM dengan baik untuk pemilihan parameter dan seleksi fitur. Dengan penerapan GA pada SVM dapat mencapai kinerja yang lebih baik dari segi komputasi dan generalisasi dari grid maupun Teknik lainnya [11].

2.1.8 Pemekaran papua

Provinsi Papua terdaftar sebagai salah satu daerah otonom dengan status khusus dalam pengelolaan pemerintahan daerah dimana rancangan awal sudah

dilakukan sejak masa pemerintahan Presiden Bacharudin Jusuf Habibie. Otonomi khusus papua tercantum dalam UU Nomor 21 Tahun 2001. Diberlakukannya Otonomi Khusus (Otsus) di Papua dan Papua Barat, memiliki tujuan supaya peningkatan kesejahteraan masyarakat papua serta penurunan angka kemiskinan di daerah papua berlangsung lebih cepat [12]. Namun pemberlakuan ketentuan Otsus hanya berlangsung dalam jangka waktu 20 tahun yakni sampai tahun 2022 [13]. Salah satu alternatif sebagai bagian dari perpanjangan UU Otsus adalah pemekaran wilayah Papua [14].

2.1.9 Twitter

Twitter merupakan media sosial yang cukup populer di kalangan masyarakat. Twitter didirikan oleh Evan Williams, Jack Dorsey, Christopher “Biz” Stone, dan Noah Glass pada tahun 2006 dan mulai populer di Indonesia. Publik bisa mengetahui trend terkini melalui twitter karena pada twitter terdapat fitur Trending yang menampilkan # (hashtag) topik yang saat ini banyak diperbincangkan di Twitter. Selain itu media sosial Twitter sering menjadi obyek penelitian dari berbagai keilmuan dan sudut pandang. Salah satu contoh penelitian yang sering digunakan menggunakan twitter yakni analisis sentimen yang diambil dari postingan masyarakat terhadap topik tertentu. Adanya twitter API membantu dalam pengambilan data yang dibutuhkan.

2.2 Penelitian Terkait dan Kebaruan Penelitian

Tabel 2. 1 Penelitian Terkait

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
Siti Rohaidah Ahmad, Muhammad Zakwan Muhammad Rodzi, Nurlaila Syafira S, Nurhafizah Moziyana M Y, Suhaila I.	2019	<i>A Review of Feature Selection and Sentiment Analysis Technique in Issues of Propaganda</i>	Naïve Bayes Classifier, Support Vector Machine (SVM), SVM berbasis Particle Swarm Optimization (PSO)	Algoritma Particle Swarm Optimization (PSO) dapat menghasilkan subset fitur yang relevan serta memiliki kualitas tinggi yang mampu meningkatkan nilai akurasi klasifikasi sentimen. [15]
Nia Kusuma W, Rezkiani, Sigit K, Hendra S, Grace Gata, Siswanto T, Windu Gata, M Wahyudi.	2018	<i>Sentiment Analysis Particle News Coordinator Minister Of Maritime Affairs Using Algorithm Naïve Bayes and Support Vector Machine with Particle Swarm Optimization.</i>	<i>Naïve Bayes and Support Vector Machine with Particle Swarm Optimization.</i>	Akurasi NB berbasis PSO mendapatkan nilai akurasi yang lebih tinggi dari SVM, SVM-PSO, NB. PSO dapat meningkatkan nilai akurasi untuk metode Naïve Bayes (NB) dan SVM. [16]
Abi Rafdi, Herman Mawengkang, Syahril Efendi.	2021	<i>Sentiment Analysis Using Naïve Bayes Algorithm with Feature Selection</i>	<i>Naïve Bayes Algorithm with Feature Selection Particle Swarm Optimization</i>	Hasil penelitian menunjukkann bahwa peningkatan nilai akurasi tertinggi Naïve Bayes diperoleh

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
		<i>Particle Swarm Optimization (PSO) and Genetic Algorithm.</i>	<i>(PSO) and Genetic Algorithm.</i>	menggunakan <i>feature selection Particle Swarm Optimization</i> lebih unggul daripada <i>Genetic Algorithm</i> dengan peningkatan akurasi sebesar 17,24%. [17]
Dwi Andini Putri	2020	<i>Comparison of Naïve Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review.</i>	<i>Naïve Bayes Algorithm and Support Vector Machine using PSO Feature Selection</i>	Hasil penelitian NB tergolong <i>Fair Classification</i> dan SVM tergolong <i>Excellent Classification</i> . Dalam penelitian ini SVM merupakan algoritma terbaik dalam mengklasifikasikan teks. Fitur pemilihan PSO dapat meningkatkan akurasi yang sangat baik pada model klasifikasi algoritma NB dan SVM, karena hasil akurasi kedua model memiliki nilai yang tinggi. [18]

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
Hendry Cipta Husada, Adi Suryaputra P.	2021	Analisis Sentimen pada Maskapai Penerbangan di Platform Twitter menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	<i>Support Vector Machine</i> (SVM)	Hasil klasifikasi terbaik diperoleh oleh SVM kernel RBF dengan nilai akurasi sebesar 84,37% dan 80,41% [7].
Elly Indrayuni	2018	Komparasi Algoritma Naïve Bayes dan <i>Support Vector Machine</i> untuk Analisa Sentimen Review Film	Naïve Bayes Classifier, <i>Support Vector Machine</i> (SVM)	Hasil akurasi terbaik diperoleh menggunakan algoritma SVM dengan nilai akurasi sebesar 90.00% sedangkan Naïve Bayes 84,50% [19].
Lala Nilawati, Yuni Eka Achyani	2019	Optimasi Metode <i>Particle Swarm Optimization</i> (PSO) pada Prediksi Penilaian Apartemen	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO)	Metode SVM berbasis <i>Particle Swarm Optimization</i> (PSO) dapat memecahkan masalah analisa prediksi penilaian apartemen, mendukung pengambilan keputusan, dan pengembangan sistem informasi manajemen [10].

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
Valentino Kevin Sitanayah Que, Ade Iriani, dan Hindriyanto Dwi Purnomo	2020	Analisis Sentimen Transportasi Online menggunakan <i>Support Vector Machine</i> (SVM) Berbasis <i>Particle Swarm Optimization</i> (PSO)	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO)	Penggunaan data <i>training</i> dan <i>testing</i> dapat dilakukan dan SVM-PSO lebih baik daripada SVM biasa, meskipun nilai parameter yang digunakan <i>default</i> [9].
Carol Shofiya dan Samina Abidi	2021	<i>Sentiment Analysis on Covid-19-Related Social distancing in Canada Using Data Twitter</i>	<i>Support Vector Machine</i> (SVM)	Hasil penelitian menunjukkan 40% sentimen netral, 35% sentimen negatif, dan 25 % sentimen positif terkait <i>Socia distancing</i> di Kanada. Nilai akurasi algoritma SVM sebesar 87% [20].
Suhardjono, Ganda Wijaya, dan Abdul Hamid	2019	Prediksi Waktu Kelulusan Mahasiswa menggunakan SVM Berbasis PSO	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO)	SVM berbasis PSO dapat meningkatkan akurasi prediksi dari sebesar 85,81% menjadi 86,43% dengan nilai kenaikan akurasi sebesar 00,62% [21].

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
I Gusti Ngurah Ersania Susena, Muhammad Tanzil Furqon, dan Randy Cahya Wihandika	2018	Optimasi Parameter <i>Support Vector Machine</i> (SVM) dengan <i>Particle Swarm Optimization</i> (PSO) untuk klasifikasi Pendonor Darah dengan Dataset RFMTC.	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO)	Hasil pengujian yang dilakukan menggunakan 748 data adalah nilai akurasi sebesar 90% dengan nilai <i>learning rate</i> SVM yang kecil dan nilai jumlah partikel PSO yang rendah [22].
*Herwinsyah, dan Arita Witanti	2022	Analisis Sentimen Masyarakat terhadap Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	<i>Support Vector Machine</i> (SVM)	Hasil akurasi yang diperoleh 80%(3766) data <i>training</i> dan 20% (942) data <i>testing</i> diperoleh hasil akurasi <i>score</i> sebesar 89%, <i>FI score</i> 93%, <i>precision score</i> 88%, <i>recall score</i> 99% [23].
Satria Wira Yudha, dan Mochamad Wahyudi	2018	Komparasi Algoritma Klasifikasi untuk Analisis Sentimen Review Film	Naïve Bayes, <i>Support Vector Machine</i> (SVM), dan k-Nearest Neighbour (k-NN)	SVM mendapatkan hasil akurasi terbaik sebesar 91,92% dan AUC 0,981 pada dataset berbahasa Inggris, sedangkan pada dataset

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
		Berbahasa Asing		berbahasa jerman nilai akurasi sebesar 90,96% dan AUC 0,978 [24].
Putri Yuniasari, dan Febri Maspiyanti	2021	Analisis Sentimen Data Tweet menggunakan Metode <i>Support Vector Machine</i> (Studi Kasus: Pemandangan Ibukota Baru Republik Indonesia)	<i>Support Vector Machine</i> (SVM)	Hasil analisis sentimen yang telah dilakukan didapatkan hasil Precision 56%, recall 55%, dan tingkat akurasi sebesar 75% [25].
Endang Supriyadi	2017	Metode SVM berbasis PSO untuk meningkatkan Prediksi Ketepatan Waktu Kelulusan Mahasiswa	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO)	PSO dapat meningkatkan akurasi metode SVM dari 80,14% menjadi 82,05%, dengan peningkatan akurasi sebesar 1,91% [26].
Ika Kurniawati dan Hilman F. Pardede	2018	<i>Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of</i>	<i>Support Vector Machine</i> (SVM) berbasis <i>Particle Swarm Optimization</i> (PSO) dan	Hasil penelitian menunjukkan bahwa method yang diajukan dalam penelitian tersebut memperoleh nilai akurasi terbaik yakni

Peneliti	Tahun	Judul Penelitian	Metode	Hasil
		<i>SVM-Based Sentiment Analysis</i>	<i>Information Gain (IG).</i>	94.80% sehingga menunjukkan bahwa method yang diusulkan yakni SVM+PSO+IG mendapatkan hasil lebih baik daripada IG dan PSO diterapkan masing-masing. [3]

Analisis sentimen merupakan proses untuk menentukan isi dari dataset yang berupa teks positif, negatif, atau netral. Salah satu algoritma yang bisa digunakan untuk mengklasifikasikan data teks kedalam 3 kelas tersebut adalah *Support Vector Machine* (SVM). Berdasarkan penelitian sebelumnya dengan menggunakan algoritma *Support Vector Machine* (SVM) menunjukkan hasil akurasi yang baik dan dapat diterapkan untuk analisis sentimen. Setiap algoritma tentu saja memiliki kekurangan dan kelebihan masing-masing. Salah satu kelebihan SVM yakni dalam menentukan jarak menggunakan SVM sehingga proses komputasi menjadi lebih cepat. Kekurangan yang dimiliki SVM yakni kesulitan dalam pemilihan parameter yang tepat sehingga nilai akurasi kurang stabil. Hal tersebut dapat diatasi dengan menerapkan Teknik optimasi. Dua diantara Teknik optimasi yang memiliki performa cukup baik adalah *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA).

Merujuk pada penelitian sebelumnya yakni penelitian yang dilakukan oleh Ika Kurniati dan Hilman P. Pardede pada tahun 2018 dengan judul “*Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis*” menggunakan metode atau percobaan yang dilakukan pada penelitian tersebut dengan menggunakan dua *feature selection* dalam satu proses yakni PSO dan *Information Gain* (IG) pada algoritma SVM dan hasil akurasi dari percobaan tersebut diperoleh hasil akurasi yang lebih tinggi dibandingkan dengan penerapan masing-masing seperti SVM+PSO saja atau SVM+IG saja akan tetapi metode yang diajukan pada penelitian ini yakni SVM+PSO+IG dalam satu proses menggunakan rapidminer.

Pada penelitian ini algoritma yang akan digunakan untuk pemodelan adalah *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen publik menjadi kelas sentimen positif atau negatif pada *tweet* pemekaran Papua, selain itu untuk mendapatkan nilai akurasi pada analisis sentimen tersebut. Untuk mengoptimalkan nilai akurasi yang didapatkan dengan menggunakan algoritma *Support Vector Machine* (SVM) pada penelitian ini akan menerapkan Teknik optimasi *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA). Dengan membandingkan hasil akurasi yang diperoleh dari penerapan SVM, SVM+PSO, SVM+GA, SVM+PSO+GA untuk memperoleh nilai akurasi terbaik.