

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

##### 2.1.1 Penelitian Terkait

Beberapa penelitian sebelumnya yang berkaitan dengan penelitian ini, dan dijadikan acuan dalam penelitian ini, diantaranya adalah :

- 1) Penelitian yang dilakukan oleh Atang Saepudin tentang komparasi algoritma *Support Vector Machine* dan *K-Nearest Neighbor* Berbasis *Particle Swarm Optimization* (PSO) pada Analisis Sentimen tentang pergantian presiden hasil komparasi algoritma, SVM menghasilkan akurasi sebesar 88.00% dan AUC 0.964, kemudian dibandingkan dengan SVM berbasis PSO dengan akurasi 92.75% dan AUC 0.973. Hasil pengujian data untuk algoritma K-NN akurasi adalah 88.50% dan AUC 0.948, kemudian dibandingkan akurasi dengan K-NN berbasis PSO sebesar 75.25% dan AUC 0.768. Hasil dari pengujian algoritma PSO dapat meningkatkan akurasi SVM, namun tidak mampu meningkatkan akurasi pada algoritma k-NN. Algoritma SVM berbasis PSO terbukti dapat memberikan solusi terhadap permasalahan klasifikasi tweet/komentar yang menggunakan tagar #2019GantiPresiden pada media sosial Twitter agar lebih akurat dan optimal.
- 2) Penelitian yang dilakukan oleh Valentino Kevin Sitanayah, Ade Iriani, dan Hindriyanto Dwi Purnomo tentang Analisis Sentimen Transportasi Online Menggunakan *Support Vector Machine* Berbasis *Particle Swarm Optimization*. Penelitian ini bertujuan mengetahui sentimen masyarakat terhadap transportasi online dan membandingkan akurasi SVM dan SVM-PSO dengan nilai parameter default. Data penelitian adalah data tweet dengan metode scraping menggunakan *Octoparse*. Total 1.852 data tweet dari 1/1/2019 hingga 15/10/2019 yang dibagi

menjadi data testing 1.130 tweet dan training 722 tweet serta RapidMiner digunakan untuk proses analisis. Analisis sentimen positif menggunakan SVM adalah sebesar 62% dan sentimen negatif sebesar 38%, sedangkan pada SVM-PSO, opini positif sebesar 53% dan negatif 47%. Hasil penelitian menggunakan 10 k-fold CV menghasilkan akurasi pada SVM sebesar 95,46% dan AUC 0,979 (*excellent classification*), sedangkan pada SVM-PSO sebesar 96,04% dan AUC 0,993 (*excellent classification*). Hasil menunjukkan bahwa penggunaan data training dan testing dapat dilakukan dan terbukti bahwa SVM-PSO lebih baik daripada SVM biasa, meskipun menggunakan nilai parameter default[8].

- 3) Penelitian yang dilakukan oleh Faisal et al. yang meneliti tentang Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis *Particle Swarm Optimization* dengan menggunakan metode *Support Vector Machine* dan *Naïve Bayes*. Penelitian ini dilakukan untuk mengukur pendapat atau memisahkan antara sentimen positif dan sentimen negatif terhadap DPR RI. Data yang digunakan didapatkan melalui crawling pada media sosial twitter. Dengan Hasil pengujian *k-fold cross validation* SVM dan NB mendapatkan nilai accuracy 71,04% dan 70,69% dengan nilai Area Under the Curve (AUC) 0,817 dan 0,661. Sedangkan hasil pengujian *k-fold cross validation* dengan menggunakan PSO, untuk SVM dan NB masing-masing mendapatkan nilai accuracy 75,03% dan 73,49% dengan nilai AUC 0,808 dan 0,719. Penggunaan PSO mampu meningkatkan nilai accuracy algoritma SVM sebesar 3,99% dan 2,8% pada algoritma NB. Hasil dari pengujian kedua algoritma tersebut nilai accuracy tertinggi adalah SVM dengan PSO sebesar 75,03%[9].
- 4) Penelitian oleh Ardiansyah et al. yang meneliti tentang Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. Pada penelitian ini, analisis sentimen diambil dari komentar pengguna media sosial Twitter terhadap aplikasi Ruang Guru sebanyak

513 tweet, setelah dilakukan data cleaning, dengan sentimen positif sebanyak 338 tweet dan sentimen negatif sebanyak 175 tweet. Data tersebut diekstraksi menggunakan algoritma Naive Bayes (NB), *Support Vector Machine* (SVM), K-Nearest Neighbour (K-NN), dan feature selection dengan algoritma *Particle Swarm Optimization* (PSO). Penelitian ini membandingkan metode NB, SVM, K-NN tanpa menggunakan *feature selection* dengan metode NB, SVM, K-NN yang menggunakan *feature selection* serta membandingkan nilai Area Under Curve (AUC) dari metode-metode tersebut untuk mengetahui algoritma yang paling optimal. Hasil pengujian mendapatkan hasil bahwa aplikasi optimasi terbaik dalam model ini adalah algoritma PSO berbasis SVM dengan nilai akurasi sebesar 78,55% dan AUC sebesar 0,853. Penelitian ini berhasil mendapatkan algoritma yang efektif dan terbaik dalam mengklasifikasikan komentar positif dan komentar negatif terkait dengan aplikasi Ruang Guru[10].

- 5) Kajian Fuad Nur Hasan tentang Analisis Sentimen Artikel Berita Sepak Bola Dunia Menggunakan Support Vector Machines dan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. Hasil perhitungan metode SVM memberikan nilai akurasi sebesar 78,50 ± nilai AUC sebesar 0,893 dan hasil perhitungan SVM (PSO) memberikan nilai akurasi sebesar 80,00 ± nilai AUC sebesar 0,91. Setelah menggunakan fungsi pemilihan PSO, peningkatan nilai AUC sebesar 0,021 meningkatkan nilai akurasi sebesar 5,50. Hasil perhitungan metode NB memberikan nilai AUC akurasi 76,50 ± 0,633 dan hasil perhitungan NB (PSO) memberikan nilai AUC akurasi 83,00 ± 0,653. Artinya, ada peningkatan tingkat akurasi. Setelah menggunakan fungsi pemilihan PSO, nilai AUC meningkat dari 6,50 n menjadi 0,020. Akurasi SVM dan NB memiliki selisih nilai sekitar 2,00%, sedangkan SVM (PSO) dan NB (PSO) memiliki selisih nilai sekitar 1,00%. Bandingkan hasil yang diperoleh

dengan menguji data NB, NB (PSO), SVM, dan SVM (PSO). Akurasi SVM (PSO) lebih akurat dibandingkan dengan SVM, NB, dan NB (PSO). Dari sini, kita dapat menyimpulkan bahwa aplikasi optimalisasi terbaik dalam model ini adalah mesin vektor pendukung berdasarkan particle swarm optimization (PSO). Hal ini dapat memberikan solusi terhadap masalah klasifikasi pada kasus analisis sentimen sepakbola dunia. Sosok Lionel Messi[11].

Rincian beberapa penelitian sebelumnya dapat dilihat pada table 2.1 berikut ini:

**Tabel 2.1 Rangkuman penelitian sebelumnya**

No	Penelitian Sebelumnya	Judul	Metode	Keterangan
1	Atang Saepudin (2020)	komparasi algoritma <i>Support Vector Machine</i> dan <i>K-Nearest Neighbor</i> Berbasis <i>Particle Swarm Optimization</i> (PSO) pada Analisis Sentimen tentang pergantian presiden	Metode yang digunakan penulis tersebut adalah <i>Support Vector Machine</i> dan <i>K-Nearest Neighbor</i> berbasis <i>Particle Swarm Optimization</i> (PSO)	<ul style="list-style-type: none"> <li>• <i>SVM</i> mendapat <i>Accuracy</i> sebesar 88.00% dan AUC 0.964,</li> <li>• <i>SVM</i> berbasis <i>PSO</i> mendapat akurasi 92.75% dan AUC 0.973.</li> <li>• metode <i>K-NN</i> mendapat nilai <i>Accuracy</i> 88.50% dan AUC 0.948,</li> <li>• <i>K-NN</i> berbasis <i>PSO</i> dengan nilai 75.25% dan AUC 0.768.</li> </ul>
2	Valentino Kevin Sitanayah et al., (2020)	Analisis Sentimen Transportasi Online Menggunakan <i>Support Vector Machine</i> Berbasis	Metode yang digunakan adalah <i>Support Vector Machine</i> Berbasis	<ul style="list-style-type: none"> <li>• <i>SVM</i> mendapat <i>Accuracy</i> sebesar 95,46% dan AUC 0,979</li> </ul>

No	Penelitian Sebelumnya	Judul	Metode	Keterangan
		<i>Particle Swarm Optimization</i>	<i>Particle Swarm Optimization</i>	<p>(<i>excellent classification</i>),</p> <ul style="list-style-type: none"> <li>• <i>SVM-PSO</i> dengan nilai 96,04% dan AUC 0,993 (<i>excellent classification</i>)</li> </ul>
3	Faisal et al, (2020)	<p>Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis <i>Particle Swarm Optimization</i> dengan menggunakan metode <i>Support Vector Machine</i> dan <i>Naive Bayes</i>.</p>	<p>Metode yang digunakan adalah <i>Support Vector Machine</i> dan <i>Naive Bayes</i> berbasis <i>Particle Swarm Optimizatio</i></p>	<ul style="list-style-type: none"> <li>• <i>Support Vector Machine</i> mendapatkan nilai <i>accuracy</i> 71,04% dan <i>Naive Bayes</i> 70,69%</li> <li>• Sedangkan hasil pengujian menggunakan <i>PSO</i>, untuk <i>Support Vector Machine</i> dan <i>NB</i> masing-masing mendapatkan nilai <i>accuracy</i> 75,03% dan 73,49%.</li> </ul>
4	Ardiansyah et al., (2020)	<p>Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi</p>	<p>Metode yang digunakan adalah <i>Naive Bayes</i>, <i>Support Vector Machine</i>, dan <i>K-Nears Neighbor</i></p>	<ul style="list-style-type: none"> <li>• <i>Support Vector Machine</i> mendapat nilai <i>Acuracy</i> sebesar 76,93%</li> <li>• <i>Naive Bayes</i> sebesar 65,40%</li> <li>• <i>K-NN</i> sebesar 70,42%</li> </ul>

No	Penelitian Sebelumnya	Judul	Metode	Keterangan
5	Fuad Nur Hasan (2018)	Analisis Sentimen Artikel Berita Tokoh Sepak Bola Dunia Menggunakan Algoritma <i>Support Vector Machine</i> Dan <i>Naive Bayes</i> Berbasis <i>Particle Swarm Optimization</i>	Metode yang digunakan adalah Algoritma <i>Support Vector Machine</i> Dan <i>Naive Bayes</i> Berbasis <i>Particle Swarm Optimization</i>	<ul style="list-style-type: none"> <li>• <i>Support Vector Machine</i> mendapatkan nilai akurasi 78.50% dan nilai AUC 0.893</li> <li>• <i>Support Vector Machine</i> berbasis (PSO) mendapatkan nilai akurasi 84.00% dan nilai AUC 0.914</li> <li>• <i>Naive Bayes</i> nilai akurasi 76.50% dan nilai AUC 0.633,</li> <li>• <i>Naive Bayes</i> (PSO) nilai akurasi 83.00% dan nilai AUC 0.653</li> </ul>

### 2.1.2 Perbandingan

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh Atang Saepudin adalah penggunaan algoritma *Support Vector Machine (SVM)* dan *K-Nearest Neighbor* Berbasis *Particle Swarm Optimization (PSO)* akan tetapi pada penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* dan *Naive Bayes* Berbasis *Particle Swarm Optimization (PSO)*. Sedangkan penelitian ini memiliki kesamaan yaitu digunakan untuk melakukan analisis sentiment.

Selanjutnya perbedaan penelitian ini dengan penelitian yang dilakukan oleh Valentino Kevin Sitanayah yang membandingkan algoritma *Support Vector Machine* dengan algoritma *Support*

*Vector Machine* berbasis *Particle Swarm Optimization* (PSO) sedangkan penelitian ini membandingkan algoritma *Support Vector Machine* dengan algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO), dan memiliki kesamaan digunakan untuk melakukan analisis sentiment.

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh Faisal et al. adalah penggunaan algoritma yang digunakan untuk analisis sentiment tentang DPR sedangkan penelitian ini digunakan untuk analisis tentang kurikulum merdeka dan memiliki kesamaan pada penggunaan algoritma untuk menganalisis opini masyarakat.

Kemudian perbedaan penelitian ini dengan penelitian yang dilakukan oleh Ardiansyah et al yang membandingkan algoritma *Support Vector Machine*, *K-NN* dan *NB* berbasis *Particle Swarm Optimization* (PSO), sedangkan *experiment* ini membandingkan algoritma *Support Vector Machine* dengan algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO), dan memiliki kesamaan digunakan untuk melakukan analisis sentiment.

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh Fuat Hasan adalah penggunaan algoritma *Support Vector Machine* dan *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO) yang digunakan untuk analisis sentiment tentang Sepakbola sedangkan penelitian ini digunakan untuk analisis tentang kurikulum merdeka dan memiliki kesamaan pada penggunaan algoritma untuk menganalisis opini masyarakat.

## **2.2 Landasan Teori**

### **2.2.1 Teks Mining**

Menurut Jiawei Han dalam Atang Saepudin, 2018 menjelaskan *Text Mining* atau analitik teks adalah istilah yang menjelaskan teknologi yang dapat menganalisis data tekstual semi-terstruktur dan tidak terstruktur, membedakannya dari penambangan data, di mana penambangan data secara

inheren memproses data terstruktur[5]. Pada dasarnya, penambangan teks adalah bidang interdisipliner yang terkait dengan pencarian informasi, penambangan data, pembelajaran mesin, statistik, dan linguistik komputasi.

Menurut Feldman dan Sanger, fase-fase *text mining* biasanya berupa *text preprocessing* dan *feature selection*. Fase *preprocessing* teks adalah fase pertama dari penambangan teks. Fase ini mencakup semua rutinitas dan proses untuk menyiapkan data untuk digunakan dalam operasi penemuan pengetahuan dari sistem penambangan teks. Penambangan teks Ini berfokus pada menemukan model, tren, pola, atau aturan yang berguna dalam data teks terstruktur seperti file teks, file HTML, pesan obrolan, dan email. dikemukakan oleh Abdous.et,al [12].

Tujuan utama *Text Mining* adalah mendapatkan informasi yang berguna dari data yang di olah. Permasalahan yang sering di temui dalam teks mining adalah data dalam jumlah besar, berdimensi tinggi, data yang berubah serta noise dan hal ini merupakan tantangan dalam pengerjaan *text mining*. Didalam *text mining* terdapat fitur pendukung yang sering digunakan antara lain:

- a. **Character**, Komponen individu dari domain penambangan teks, seperti huruf, angka, karakter khusus, dan spasi, membentuk level tertinggi dalam membentuk fungsi semantik.
- b. **Word**, Berarti kata yang dipilih langsung dari dokumen asli yang membentuk dasar atau tingkatan dasar, meskipun kata fitur mungkin terdapat dalam dokumen asli itu sendiri.
- c. **Term**, Ditafsirkan sebagai kata tunggal dan multi-kata yang dipilih langsung dari korpus. Representasi termbase dari suatu dokumen terdiri dari subset dari istilah-istilah dalam dokumen.
- d. **Concept**, Fitur yang diturunkan dari dokumentasi secara manual, berbasis aturan, atau lainnya.

### 2.2.2 Analisis Sentimen



Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negative [11], analisis sentiment bertujuan untuk menganalisa pendapat, penilaian dan emosi seseorang berkenaan dengan suatu topic atau produk maupun kegiatan tertentu. Analisis sentiment adalah proses pengelompokan teks yang ada pada sebuah kalimat atau dokumen kedalam sifat yang positif atau negative.

*Opinion mining* atau analisis sentimen merupakan suatu bidang ilmu dari data mining yang berguna untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu [13]. Analisis ini bertujuan untuk mendapatkan sebuah informasi dari suatu himpunan data. Analisis sentimen adalah penelitian yang baru pada *Natural Language Processing* (NLP) dan bertujuan menemukan subjektivitas dalam teks maupun mengekstraksi dan menjalankan klasifikasi sentimen pada opini[8].

Terdapat beberapa teknik dalam metode analisis sentiment diantaranya adalah dengan cara *hybrid approach*, *lexicon based*, dan *Machine learning*, saat ini banyak penelitian analisis sentiment yang menggunakan *Machine Learning* karena di anggap dapat memprediksi polaritas sentiment yakni positif, negative, dan netral berdasarkan data *training* dan *testing* [14]. Proses analisis sentimen pada teks mencakup proses *tokenisasi*, *stopword removal*, *stemming*, identifikasi sentiment, dan klasifikasi sentiment.

Sentimen analisis bermanfaat juga dalam dunia usaha seperti untuk melakukan analisa tentang sebuah produk yang dilakukan secara cepat dan menggunakan alat bantu untuk melihat respon konsumen, sehingga dapat membuat langkah strategis untuk mengantisipasi hal tersebut. Pada tugas akhir penelitian ini menganalisis sentiment masyarakat terhadap penerapan kurikulum

merdeka menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* yang dikhususkan pada dokumen teks berbahasa Indonesia.

### **2.2.3 Algoritma Naïve Bayes**

*Naive Bayes* adalah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dalam kumpulan data tertentu. Algoritma tersebut menggunakan teorema *Bayes* dan mengasumsikan bahwa semua atribut adalah independen atau bebas dengan nilai variabel kelas. [15]. Definisi lain menyatakan bahwa *Naive Bayes* adalah klasifikasi yang menggunakan metode probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes, metode yang memprediksi peluang masa depan berdasarkan pengalaman masa lalu. [16].

*Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Mujib dkk, 2013). Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Alur kerja dari *Naïve Bayes* adalah sebagai berikut:

1. Membaca data training
2. Menghitung jumlah probabilitas.
3. Mendapatkan nilai dalam tabel mean, standar deviasi, dan probabilitas.

*Naive Bayes Classifier* termasuk ke dalam pembelajaran supervised, *Naive Bayes* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat

yang diberikan dengan label  $y$ , Asumsi independen bersyarat dapat dinyatakan dalam bentuk berikut (Suyanto,2017). Adapun perhitungan

*Naive Bayes* menggunakan persamaan 2.1.

$$P(H|X) = \frac{p(H|X)P(H)}{P(x)} \quad (2.1)$$

Keterangan:

X = Data dengan class yang belum diketahui (bukti)

H = Hipotesis data X merupakan suatu class spesifikasi

$P(H|X)$  = Probabilitas hipotesis H benar untuk kondisi X

$P(H)$  = Probabilitas hipotesis H

PX = Probabilitas prior bukti X

#### **2.2.4 Algoritma *Support Vector Machine***

*Support Vector Machine* (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*[17].

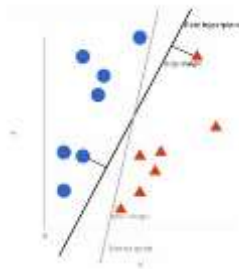
Konsep dasar metode SVM sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang telah ada pada tahun sebelumnya, seperti margin *hyperplane* (Dyda dan Hart, 1973; Cover, 1965; Vapnik, 1964), kernel diperkenalkan oleh Aronszajn tahun 1950, Lagrange Multiplier yang ditemukan oleh Joseph Louis Lagrange pada tahun 1766, dan demikian juga dengan konsep-konsep pendukung lain.

*Support vector machine* adalah sistem pembelajaran yang dilatih dengan algoritma pembelajaran berdasarkan teori optimisasi, menggunakan hipotesis berupa fungsi linier pada fitur berdimensi tinggi.[18]. Keakuratan model yang dihasilkan oleh proses switching SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan. Karena karakteristik metode SVM, dapat dibagi menjadi linier dan nonlinier.SVM linier adalah data yang dipisahkan secara linier, yaitu data

dengan tepi lunak yang memisahkan dua kelas pada hyperplane. Meskipun non-linier, ini adalah fungsi trik kernel di ruang berdimensi tinggi.

Dukungan mesin vektor sangat cepat dan efektif untuk masalah klasifikasi teks. Secara geometris, klasifikasi biner dapat dilihat sebagai hyperplane dalam ruang fitur yang memisahkan titik yang mewakili contoh positif dari kategori yang mewakili keadaan negatif. Pengklasifikasi ini dipilih selama pelatihan sebagai *hyperplane* unik yang memisahkan instance positif yang diketahui dari instance negatif. Ini karena mesin vektor pendukung klasifikasi memiliki keunggulan signifikan dibandingkan pendekatan teoretis yang membenarkan masalah *overfitting* agar dapat bekerja dengan baik.[19].

*Support Vector Machine* Ini bekerja dengan menemukan dan menentukan nilai terbaik dari hyperplane (fungsi pemisahan kelas). Nilai *hyperplane* yang optimal adalah nilai tengah antara dua set objek dari dua kelas. Menemukan *hyperplane* optimal berbanding lurus dengan memaksimalkan margin. Margin adalah jarak tegak lurus antara *hyperplane* dengan obyek terdekat sebagaimana ditunjukkan oleh Gambar 2.1 berikut ini:



Gambar 2.1 *Support Vector Machine*

Sebagai salah satu fungsi linear, bentuk SVM secara umum dinyatakan dalam persamaan 2.2 :

$$(x) = w T x + b \quad (2.2)$$

di mana  $x$  adalah vektor input,  $w$  adalah parameter bobot, dan  $b$  adalah deviasi. Fungsi linear dengan nilai 0 disebut limit keputusan. Fungsi linear dengan nilai 1 disebut hyperplane positif, dan

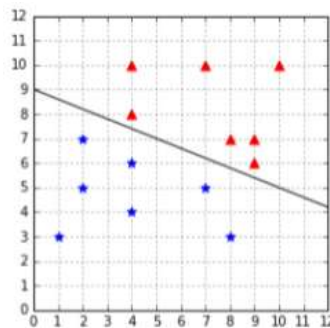
fungsi linear dengan nilai -1 disebut hyperplane negatif. Berdasarkan fungsi *hyperplane* tersebut, margin dapat dihitung menggunakan persamaan 2.3:

$$margin = \frac{\frac{1}{2} * 2}{\|w\|} = \frac{1}{\|w\|} \quad (2.3)$$

Guna memaksimalkan nilai margin, dapat dilakukan dengan meminimalkan pembaginya yaitu  $\|w\|$ , sehingga bentuk optimasi margin sebagaimana dinyatakan dalam persamaan 2.4 :

$$margin = \frac{1}{2} \|w\|^2 \quad (2.4)$$

*Support Vector Machine* menggunakan 2 titik (vector) yang selanjutnya dua titik tersebut akan membentuk garis pembatas (sisi pembatas jika 3 dimensi atau lebih) garis pembatas yang dibentuk dari dua buah vector ini disebut *hyperplane*. *Hyperplane* dapat dilihat pada gambar 2.2 berikut :



Gambar 2.2 Garis *Hyperplane* memisahkan dua kelas.

Dua titik yang menjadi titik acuan *hyperplane* disebut *support vector*. Dapat dilihat keberadaan dua set data disebut klasifikasi. Tugas dari mesin vektor pendukung kemudian untuk menentukan subdivisi terbaik dari kedua grup ini, atau *hyperplane optimal* (subdivisi yang batasnya dapat memisahkan kedua grup). Dengan garis saja [13]. Masalah nonlinier dapat diatasi dengan merubah trik kernel SVM untuk memisahkan kelas atau *hyperplane* menjadi dua kelas dalam ruang vektor. Pada penelitian ini, kernel yang digunakan adalah kernel linier.

### **2.2.5 Particle Swarm Optimization (PSO)**

Algoritma *Particle Swarm Optimization* (PSO) diperkenalkan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995 (He, Jie & Hui, Guo, 2013). *Particle Swarm Optimization* (PSO) merupakan salah satu dari teknik komputasi evolusioner, dimana populasi pada PSO didasarkan pada penelusuran algoritma dan diawali dengan suatu populasi yang random yang disebut dengan *particle* (Tuegeh, Soeprijanto, & Purnomo, 2009).

*Particle Swarm Optimization* (PSO) merupakan algoritma berbasis populasi yang mengeksplorasi individu dalam pencarian, pada PSO populasi disebut *swarm* dan individu disebut *particle* (Kusmarna et al., 2015). PSO yaitu pencarian solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel dengan cara melakukan seleksi terhadap atribut yang ada (Achyani, 2018). Setiap partikel bergerak dalam ruang tertentu kemudian mencari posisi terbaik yang dilaluinya dengan memperbaharui posisinya, setiap partikel menyampaikan informasi posisi terbaik ke partikel lain dan menyesuaikan posisi dan kecepatan berdasarkan informasi posisi terbaik yang diterima (Istighfarin et al., 2020).

### **2.2.6 Kurikulum Merdeka**

Dalam Kamus Besar Bahasa Indonesia Indonesia (KBBI), kurikulum diartikan sebagai sistem pembelajaran yang diterapkan di lembaga pendidikan Indonesia. Lebih lanjut Rani (2020) menyatakan bahwa kurikulum adalah indikator sistem manajemen yang terstruktur secara sistematis yang mencakup beberapa proses perencanaan terperinci, pengembangan program, dan penyampaian pembelajaran. Kurikulum Indonesia memiliki sejarah yang panjang, jelas Loeziana Uce (dalam Ritonga, 2018). Perubahan kurikulum dimulai dari tahun 1997 hingga tahun 2013 dan mengalami banyak perkembangan.

Menteri Pendidikan dan Kebudayaan Nadiem Anwar Makarim meluncurkan dan mengumumkan kurikulum baru bernama Kurikulum Belajar Bebas (Kemendikbud, 2019). Banyak aspek yang mendorongnya untuk memperbarui kurikulum. Salah satunya adalah hasil penelitian PISA (*International Student Assessment Program*) dan fokus pemerintah, pelajar Indonesia menduduki peringkat ke-6 dari 79 negara di dunia dalam literasi dan numerasi (Mustagfiroh, 2020). Ini pukulan telak bagi dunia pendidikan nasional. Kementerian Pendidikan dan Kebudayaan telah menerapkan konsep literasi dan numerasi dalam kebijakannya menghapuskan Ujian Nasional (UN) 2020 dan menggantinya dengan asesmen kemampuan minimal dan asesmen karakter. Kriteria penilaian didasarkan pada PISA. Kurikulum pembelajaran merdeka terdiri dari empat komponen utama yaitu:

1. Ujian Sekolah Berstandar Nasional (USBN) diganti dengan asesmen berupa ujian tertulis dan/atau bentuk ujian lain yaitu penugasan dan portofoli (seperti tugas kelompok, karya tulis, tugas project, dan lain-lain).
2. Tahun 2020 Ujian Nasional (UN) dihapus dan diganti dengan Survei Karakter serta Asesmen kompetensi Minimum.
3. Implementasi perihal Rencana Pelaksanaan Pembelajaran (RPP) satu lembar.
4. Menerapkan sistem zonasi pada Penerimaan Peserta Didik Baru (PPDB).

Kemendikbud menginisiasi empat komponen di atas dengan tujuan memberikan keleluasaan yang luas kepada setiap siswa, guru, dan sekolah dalam menentukan langkah kebijakan. Sekolah, seperti yang dirumuskan oleh Ki Hadjar Dewantara, merupakan tempat bermain yang paling nyaman bagi siswa. Kenyamanan mempengaruhi proses pembelajaran sehingga siswa dapat memahami dengan jelas informasi yang mereka tangkap, terutama ketika pengalaman disampaikan melalui materi yang diberikan guru.

Kurikulum Merdeka Belajar diharapkan dapat mengikuti pesatnya globalisasi abad 21, dan tuntutan zaman mendorong lembaga pendidikan untuk mengadopsi pendekatan kurikulum yang adaptif dan berorientasi solusi secara konsisten. Secara kontekstual, banyak hal yang mempengaruhi perubahan, seperti perubahan kurikulum suatu negara. Selama ini kebutuhan masyarakat tidak menentu atau bahkan sulit diprediksi. Tidak terkecuali kebutuhan akan pendidikan, termasuk matematika. Banyak aspek penting yang dibutuhkan masyarakat, khususnya siswa Sekolah Menengah Atas (SMA).

### 2.2.7 Confusion Matrix

Pengujian dengan *Confusion Matrix* Pada tahap ini pengujian model penelitian dilakukan dengan metode *Confusion Matrix* yang mempresentasikan hasil evaluasi model dengan menggunakan tabel matrik, Jika dataset terdiri dari 2 kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Menguji menggunakan confusion matrix menghasilkan accuracy, presisi, recall, dan nilai-F. Akurasi klasifikasi adalah akurasi kumpulan data yang diklasifikasikan dengan benar setelah memeriksa hasil klasifikasi. Akurasi juga berlaku untuk data nyata. Recall merupakan jumlah nilai positif yang sebenarnya diprediksi positif secara benar. *True Positive* (TP) merupakan jumlah record positif dalam dataset yang diklasifikasikan positif. *True Negative* (TN) merupakan jumlah record negative dalam dataset yang diklasifikasikan positif. *False Positive* (FP) merupakan jumlah record negatif dalam dataset yang diklasifikasikan positif. *False Negative* (FN) merupakan jumlah record positif dalam dataset yang diklasifikasikan negatif. Berikut adalah persamaan model *Confusion Matrix*. [13]

*Accuracy* adalah jumlah perbandingan data yang benar dengan jumlah keseluruhan data. Untuk mencari nilai *Accuracy* dapat menggunakan persamaan 2.5, dengan persamaan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$



*Precision* digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif. Untuk mencari nilai *Precision* dapat menggunakan persamaan 2.6, dengan persamaan sebagai berikut:

$$Precision = \frac{TP}{FP+TP} \quad (2.6)$$

*Recall* difungsikan untuk menunjukkan berapa kelas data positif yang telah diprediksi benar dari seluruh data kelas positif. Untuk mencari nilai *Recall* dapat menggunakan persamaan 2.7, dengan persamaan sebagai berikut:

$$Recall = \frac{TP}{FN+TP} \quad (2.7)$$