

BAB 2
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Berikut merupakan beberapa penelitian sebelumnya yang akan dijadikan acuan tugas akhir ini.

Gambar 2.1 Penelitian Terdahulu

No.	Penulis (Tahun)	Judul	Kontribusi
1.	(Rizkiah et al., 2020).	Sistem Pendeteksi Dini Kesehatan Mental Emosional Anak Usia 4-17 Tahun Menggunakan Metode Forward Chaining.	Penilaian mental emosional pada penelitian ini memanfaatkan <i>Strength and Difficulties Questionnaire</i> (SDQ) dan diolah menggunakan Metode <i>Forward Chaining</i> , klasifikasi kesehatan mental emosional sebagai hasil pemeriksaan pada aplikasi ini adalah, abnormal borderline, atau normal, yang dibuktikan dengan pengujian 30 data anak Yayasan Sahabat Anak Jakarta dengan nilai akurasi 100%.
<p>Perbedaan: Penelitian ini menggunakan metode <i>Strength and Difficulties Questionnaire</i> (SDQ) untuk mengukur kesehatan mental emosional dan pengolahan data kuesioner pada penelitian ini menggunakan algoritma <i>Forward Chaining</i>.</p>			

2.	(Kurniawaty et al., 2018).	Klasifikasi Gangguan Jiwa Skizofrenia Menggunakan <i>Algoritme Support Vector Machine</i> (SVM).	Pada penelitian ini digunakan 75 data yang dibagi dalam dua jenis gangguan jiwa skizofrenia yaitu simplex dan paranoid. Metode yang digunakan adalah algoritme SVM. Hasil pengujian dengan menggunakan rasio perbandingan 9:1, menghasilkan akurasi sebesar 100%, dengan menggunakan <i>kernel polynomial of degree</i> nilai $\lambda = 3$, $\gamma = 0,00001$, nilai $C = 0,01$.
Perbedaan: Pada penelitian ini studi kasus yang diterapkan adalah klasifikasi pada Gangguan Jiwa Skizofrenia.			
3.	(Hasibuan & Prahutama, 2017).	Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (Dbd) Menggunakan <i>Support Vector Machine</i> (SVM) Berbasis GUI Matlab.	Pada penelitian ini klasifikasi didasarkan pada jumlah hemoglobin, trombosit, leukosit dan hematokrit dalam tubuh pasien. Klasifikasi dilakukan dengan menggunakan algoritma <i>Support Vector Machine</i> (SVM). Pada penelitian ini menggunakan data sekunder di RSI Sultan Agung pada data rekam medis pasien Demam Berdarah Dengue (DBD), data terbagi dalam data training dan data testing dengan perbandingan 80:20. Hasil nilai ketepatan terbaik yaitu sebesar 96,4286 % pada fungsi kernel Polynomial.
Perbedaan: Pada penelitian ini proses klasifikasi digunakan untuk diagnosa penyakit			

	demam berdarah.		
4.	(Sugara & Subekti, 2019).	Penerapan <i>Support Vector Machine</i> (SVM) Pada <i>Small Dataset</i> Untuk Deteksi Dini Gangguan Autisme.	Pada penelitian ini digunakan algoritma SVM dan <i>k-fold cross validation</i> untuk mengukur nilai keakuratan <i>small dataset</i> serta teknik <i>ensemble</i> untuk mengetahui seberapa pengaruhnya terhadap algoritma <i>Support Vector Machine</i> . Dataset yang diaplikasikan dalam pengujian ini sebesar 67 yang didapat dari hasil wawancara dan kuesioner online. Hasil nilai akurasi tertinggi yaitu 85% pada <i>normalized poly kernel</i> .
Perbedaan: Pada penelitian ini algoritma <i>Support Vector Machine</i> digunakan untuk deteksi dini gangguan autisme.			

2.2 Gangguan Kesehatan Mental Emosional

Gangguan mental emosional adalah kondisi yang mengindikasikan seseorang mengalami suatu perubahan emosi yang apabila terus berkepanjangan dapat berkembang menjadi keadaan *abnormal*, sehingga perlu diantisipasi guna menjaga kesehatan jiwa masyarakat (Mubasyiroh et al., 2017). 16% dari beban penyakit global dan cedera pada orang yang berusia 10-19 tahun adalah gangguan mental (Trisni Handayani, Dian Ayubi, 2020).

Masalah mental emosional dapat digambarkan dengan eksternalisasi, temperamental, gangguan perhatian, ketidakmampuan memecahkan masalah, perilaku bertentangan (tidak suka ditegur/diberi masukan positif, tidak mau ikut aturan), hiperaktifitas, dan biasanya timbul perilaku agresi. Pada remaja perkembangan mental emosional yang kurang baik seperti merasa cemas atau khawatir terhadap apapun, lebih suka menyendiri, tertekan atau menangis, sering

merasa tidak bahagia, mudah takut terhadap sesuatu, dan memiliki fokus dan perhatian yang kurang baik(Susanti et al., 2018).

Berdasarkan hasil Riskedes(2013) gejala gangguan mental emosional lebih fokus kepada gangguan neurosis, dua diantaranya adalah :

a. Depresi

Depresi adalah gangguan yang berhubungan dengan kesedihan manusia. Gejala depresi meliputi perubahan nafsu makan, pola tidur, konsentrasi, psikomotor, kelelahan, anhedonia, putus asa, dan ketidakberdayaan terhadap ide bunuh diri. (Dianovinina, 2018).

b. Ansietas

Ansietas atau kecemasan ialah suatu kekhawatiran yang tidak spesifik dan umum terkait dengan ketidakberdayaan dan ketidakpastian. Indikator kecemasan meliputi ketegangan, cepat lelah, mengantuk, sulit berkonsentrasi, sulit tidur, dan kehilangan memori (Dinuriah, 2015).

2.3 Data Mining

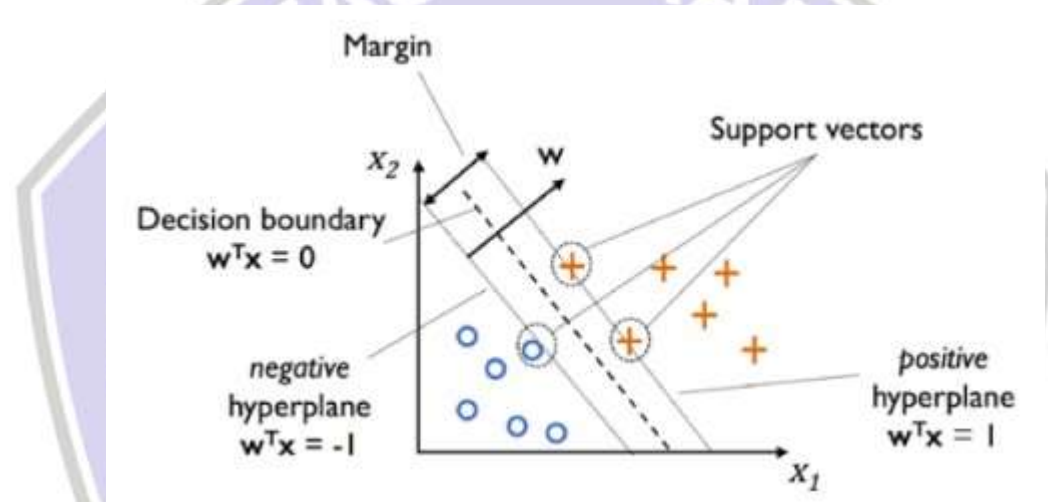
Data mining adalah proses mengekstraksi dan menganalisis informasi secara otomatis menggunakan satu atau lebih teknik pembelajaran komputer (*machine learning*) (MURRAY & SCIME, 2015). Data mining memuat pencarian pola pada *database* untuk membantu pengambilan keputusan (Yanto, 2018).

2.4 Klasifikasi

Klasifikasi adalah teknik untuk mengamati kelakuan berdasarkan atribut dari data yang telah digolongkan. Tujuan teknik klasifikasi adalah untuk menggolongkan sebuah data melalui data yang diklasifikasikan sebelumnya sehingga membentuk sebuah aturan. Tahap dari klasifikasi dibagi menjadi dua, yaitu *learning* dan *testing* (Samodra et al., 2009). Beberapa metode pengklasifikasian antara lain, *Support Vector Machine (SVM)*, *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor* dan *Neural Network*.

2.5 Support Vector Machine

Dasar teori SVM sudah ada sejak tahun 1960-an, namun sejak diperkenalkan oleh Vapnik, Boser, dan Guyon pada tahun 1992, SVM berkembang pesat. SVM memiliki kinerja yang baik di berbagai bidang, contoh bidang aplikasinya adalah pengenalan tulisan tangan, bidang bioinformatika, dan klasifikasi teks (Sembiring, 2007). Prinsip SVM adalah bagaimana menggunakan fungsi-fungsi yang diperoleh dari data selama proses pelatihan untuk memisahkan dua kategori. Konsep klasifikasi menggunakan algoritma SVM dapat diartikan sebagai upaya mencari hyperplane terbaik yang berperan sebagai pemisah antara dua kelas data (Hasibuan & Prahutama, 2017).



Gambar 2.1 Support Vector

Pada Gambar 2.1 ditunjukkan ilustrasi bagaimana hyperplane ditemukan. *Hyperplane* adalah garis pemisah terbaik antara kedua kelas dapat ditemukan yaitu dengan mengukur *margin hyperplane* kemudian dicari titik maksimalnya. Margin ialah jarak yang memisahkan antara *hyperplane* dengan *pattern* yang terdekat dari masing-masing kelas. *Pattern* terdekat ini disebut sebagai *support vector* (Nugroho et al., 2003). Untuk mencari *hyperplane* dari dua kelompok atau kelas positif dan kelas negatif dapat dirumuskan dengan fungsi berikut $f(w, b) = x_i \cdot w + b$.

$$x_i \cdot w + b \geq 1 \text{ untuk } y_i = 1 \quad (2.1)$$

$$x_i \cdot w + b \leq 1 \text{ untuk } y_i = -1 \quad (2.2)$$

Keterangan:

x_i = data ke i

w = nilai bobot vektor

b = nilai bias

y_i = kelas data ke-i yang mempunyai nilai sama dengan persamaan atau $f(x)$.

Nilai bobot Vektor dapat dihitung dengan persamaan 2.3 dan nilai bias dapat dihitung dengan persamaan 2.4.

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (2.3)$$

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (2.4)$$

Keterangan:

$w \cdot x^+$ = nilai bobot untuk kelas data positif

$w \cdot x^-$ = nilai bobot untuk kelas data negatif

α_i = nilai bobot data ke-i

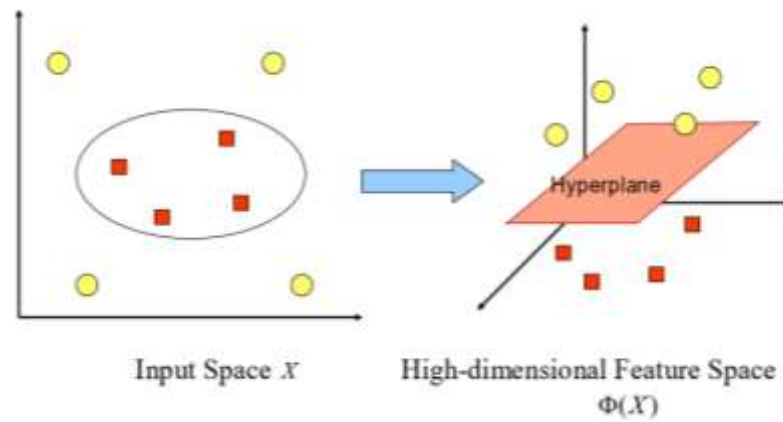
y_i = kelas data ke i

x_i = data ke i

a. Fungsi Kernel

Pada beberapa kasus terdapat data yang tidak dapat dikelompokkan secara *linear*. Dalam SVM dengan data non *linear* dapat dipetakan ke ruang vektor yang berdimensi lebih tinggi dengan menggunakan fungsi kernel. Pada ruang vektor yang baru ini, didapatkan sebuah konstruksi dari *hyperplane* yang memisahkan kedua kelas tersebut. Hal tersebut sesuai dengan teori *Cover* yang menyatakan “Jika suatu transformasi bersifat non linear dan dimensi dari *feature space* cukup tinggi, maka data pada *input space* dapat dipetakan ke *feature space* yang baru, dimana *pattern-pattern* tersebut pada probabilitas tinggi dapat dipisahkan secara

linear” (Nugroho et al., 2003). Ilustrasi dari konsep ini dapat dilihat pada gambar 2.2.



Gambar 2.2 Memetakan Data ke Ruang Vektor yang Berdimensi Lebih Tinggi
Fungsi kernel yang umum digunakan adalah sebagai berikut :

1. *Gaussian Radial Basic Function (RBF)*

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (2.5)$$

2. *Polynomial*

$$K(x, y) = ((x \cdot y) + c)^d \quad (2.6)$$

3. *Linear*

$$K(x, y) = (x \cdot y)^d \quad (2.7)$$

Keterangan:

$K(x, y)$ = fungsi kernel

x = data

y = kelas data

c = konstanta C

σ = sigma

d = konstanta d

b. Training SVM

Terdapat beberapa metode perhitungan *training* yang bisa digunakan pada Support Vector Machine, yaitu metode *Sequential Training*, *Quadratic Programming* dan *Sequential Minimal Optimization* (SMO). Metode *Sequential Training* memiliki algoritma yang sederhana dibandingkan dengan metode *training* yang lain dan tidak memerlukan waktu lama. Berikut adalah metode *Sequential Training* (Kurniawaty et al., 2018).

1. Tahap inisiasi nilai $\alpha_1, \gamma, C, \varepsilon$.

Penjelasan:

α_1 = alpha, digunakan untuk mencari nilai support vector

γ = gamma untuk mengontrol penyebaran

C = cost, digunakan untuk mengontrol trade off, antara margin dengan kesalahan klasifikasi

ε = untuk mencari nilai eror

2. Menghitung *Hessian matrix* berikut persamaan yang digunakan:

$$D_{ij} = y_i y_j (K(x_i x_j) + \lambda^2) \quad (2.8)$$

Dengan $i, j = 1, \dots, n$

Penjelasan:

y_i = kelas data ke-i

y_j = kelas data ke-j

x_i = data ke-i

x_j = data ke-j

$K(x_i x_j)$ = fungsi kernel

λ = lambda

n = jumlah data

3. Selanjutnya untuk mencari nilai alfa dapat dihitung menggunakan persamaan berikut:

$$E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (2.9)$$

$$\delta a_i = \min((\max[\gamma(1 - E_i), a]), C - a_i) \quad (2.10)$$

$$a_i = a_i + \delta a_i \quad (2.11)$$

Penjelasan:

E_i = Error rate

α_j = alfa ke-j

D_{ij} = Matriks Hessian

δa_i = delta alfa ke-i

4. Dilakukan pengulangan pada langkah ke-3 sampai mendapatkan iterasi maksimum atau mencapai nilai $\delta a_i \leq \varepsilon$.
5. Setelah mencapai iterasi maksimum maka akan didapatkan nilai hyperplane atau a_i .

c. Testing Support Vector Machine

Untuk melakukan *testing* atau pengujian pada *Support Vector Machine* dapat dilakukan dengan langkah-langkah berikut.

1. Pertama menghitung nilai $f(x)$, dengan rumus berikut.

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (2.12)$$

Penjelasan:

α_i = alfa ke-i

y_i = kelas data ke-i

m = jumlah data yang merupakan SV

$K(x_i, x)$ = fungsi kernel yang digunakan

b = nilai bias

Nilai bias dapat dicari dengan rumus berikut.

$$b = -\frac{1}{2} \left(\sum_{i=1}^m \alpha_i y_i K(x_i, x^+) + \sum_{i=1}^m \alpha_i y_i K(x_i, x^-) \right) \quad (2.13)$$

Penjelasan:

$K(x_i, x^+)$ = fungsi kernel terhadap data kelas positif

$K(x_i, x^-)$ = fungsi kernel terhadap data kelas negatif

2.6 Confusion Matrix

Dalam mengevaluasi *performance* algoritma dari *Machine Learning* khususnya *supervised learning*, *Confusion Matrix* digunakan untuk merepresentasikan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma (Putra, 2019). *Confusion Matrix* dapat dijadikan acuan untuk

menentukan nilai *Accuracy*, *Precision*, *Recall* dan *Specificity*. Bentuk dari *confusion matrix* ditunjukkan pada Gambar 2.3.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<p>TP (True Positive)</p>	<p>FP (False Positive) <i>Type I Error</i></p>
	0 (Negative)	<p>FN (False Negative) <i>Type II Error</i></p>	<p>TN (True Negative)</p>

Gambar 2.3 *Confusion matrix*

Saat menggunakan *Confusion matrix* untuk mengukur kinerja, ada empat istilah untuk mewakili hasil proses klasifikasi, yaitu benar positif (TP), benar negatif (TN), positif palsu (FP) dan negatif palsu (FN).

2.7 Akurasi

Akurasi merupakan pengukuran kinerja *machine learning* paling sederhana yang digunakan untuk mengevaluasi kinerja model. Akurasi didefinisikan sebagai proporsi prediksi yang benar dibagi dengan banyaknya sampel (Putra, 2019).

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.14)$$

Keterangan :

TP = Nilai *True Positive*, merupakan jumlah data dari kelas positif yang terdeteksi benar.

TN = Nilai *True Negative*, merupakan jumlah data dari kelas negatif yang terdeteksi dengan benar.

FP = Nilai *False Positive*, merupakan jumlah data yang merupakan data negatif namun terdeteksi sebagai data positif.

FN = Nilai *False Negative*, merupakan jumlah data yang merupakan data positif namun terdeteksi sebagai data negatif

2.8 *User Experience Questionnaire (UEQ)*

Menurut Alan Dix, evaluasi dalam desain sistem berperan dalam memastikan bahwa sistem akan bekerja sesuai dengan yang diharapkan dan memenuhi permintaan (Kurniawan, 2019). Salah satu metode untuk mengevaluasi desain sistem adalah dengan metode UEQ. Metode UEQ sangat mudah untuk diaplikasikan, efisien, dan memiliki keunggulan karena mempunyai *Data Tools Analysis* yang dapat membandingkan tingkat pengalaman dari setiap responden dengan mudah (Juniantari et al., 2021). UEQ dapat digunakan untuk menentukan kualitas pragmatik dan kualitas hedonic. Kualitas pragmatik terdiri dari kejelasan, efisiensi, dan ketepatan sedangkan kualitas hedonic terdiri dari stimulasi dan kebaruan.

