

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Gangguan Mental**

Gangguan mental adalah suatu kondisi yang mempengaruhi pikiran, perasaan, dan suasana hati seseorang. Hal ini juga mempengaruhi bagaimana seseorang menjalani kehidupan sehari-hari dan berinteraksi dengan orang lain [7]. Gangguan mental bisa dialami oleh siapa saja, Data Riskesdas menyatakan bahwa gangguan mental lebih banyak didapati pada usia remaja [8].

Gangguan kesehatan mental juga dapat disebabkan kecanduan bermain game atau durasi bermain game yang lama, karena ketika seseorang hanya fokus pada bermain *game online*, hal itu akan mulai mempengaruhi pertumbuhannya karena seseorang akan mulai menarik diri dari lingkungan sosialnya [9].

#### **2.2 Kesehatan Mental**

Mental adalah kondisi kesehatan dan kesejahteraan pikiran seseorang. Mental yang sehat mencakup keadaan di mana seseorang dapat berfungsi secara efektif dalam kehidupan sehari-hari, termasuk dalam hubungan sosial, pekerjaan, dan aktivitas sehari-hari. Gangguan ketenangan batin, kecemasan, ketakutan, apatis, anti sosial, ketegangan kronis, obsesif kompulsif, depresi, histeria, dan gangguan mental adalah tanda gangguan mental [10].

Kesehatan mental adalah terwujudnya keserasian yang sungguh-sungguh antara fungsi-fungsi kejiwaan dan terciptanya penyesuaian diri antara manusia dengan dirinya sendiri dan lingkungannya [11]. Ciri dari tingkat kesejahteraan mental di mana seseorang dapat berfungsi dengan baik, menikmati hidup secara seimbang, mampu menyesuaikan diri dengan kesulitan hidup, dan berkontribusi pada kehidupan sosial, budaya, dan agama [12].

### 2.3 Algoritma C45

Algoritma C45 adalah salah satu algoritma induksi pohon keputusan, yaitu ID3 (*Iterative Digital Calculator 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam proses algoritma ID3, *input* mengambil bentuk sampel pelatihan, pelatihan label dan atribut [13]. Sebuah pohon keputusan adalah struktur pohon yang seperti *flowchart*, dengan setiap node internal (node nonleaf) yang mewakili tes pada atribut, setiap cabang mewakili hasil tes, dan setiap node daun memegang label kelas [14].

Algoritma C45 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatan dalam proses klasifikasi atau segmentasi dan bersifat prediktif. Algoritma C45 salah satu metode untuk membuat C4.5 berdasarkan training data yang telah disediakan. “Kelebihan algoritma C45 adalah metode pohon keputusan yang dapat menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi keputusan yang dihasilkan” [15]. Adapun langkah-langkah dalam algoritma C4.5 sebagai berikut [16] :

- a. Menghitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu untuk proses pertama syaratnya masih kosong.
- b. Memilih atribut sebagai Node.
- c. Membuat cabang untuk tiap-tiap anggota dari Node.
- d. Memeriksa apakah nilai entropy dari anggota Node ada yang bernilai nol. Jika ada, menentukan daun yang terbentuk. Jika seluruh nilai entropy anggota Node adalah nol, maka proses berhenti.
- e. Jika ada anggota Node yang memiliki nilai entropy lebih besar dari nol, proses diulang lagi dari awal dengan Node sebagai syarat sampai semua anggota dari Node bernilai nol. Node adalah atribut yang mempunyai nilai gain tertinggi dari atribut-atribut yang ada.

Sedangkan rumus yang digunakan dalam penghitungan dengan algoritma C4.5 dapat dijelaskan sebagai berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} * Entropy(S_i)$$

Keterangan :

S : Himpunan Kasus

A : Atribut

N : Jumlah partisi atribut A

|S<sub>i</sub>|: Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sedangkan perhitungan nilai Entropy dapat dilihat pada persamaan berikut:

$$Entropy(A) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan:

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

p<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S

Kriteria yang paling banyak digunakan sebagai pemecah dalam algoritma C4.5 adalah Gain Ratio (Rasio Gain), dengan persamaan berikut:

$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)}$$

Keterangan:

Gain (s,j) : Nilai Information

Gain SplitInfo (s,j) : Nilai Splitinfo

Dari persamaan menyatakan nilai rasio gain pada fitur ke-j. SplitInfo(s,j) didapat dari persamaan 4 berikut :

$$SplitInfo(s,j) = \sum_{i=1}^k p(v \log_2 p(vi|s))$$

Keterangan :

s : Data yang digunakan untuk training / testing

j : Atribut

vi : jumlah data untuk atribut ke-i

## 2.4 Confusion Matrix

*Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [17].

**Tabel 2.1** Confusion matrix

		Kelas Prediksi	
		1	0
Kelas sebenarnya	1	TP	FN
	0	FP	TN

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut:

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

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

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

Keterangan:

TP (*True Positive*) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1

TN (*True Negative*) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (*False Positive*) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (*False Negative*) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

## 2.5 Cross Validation

*Cross validation* atau estimasi rotasi adalah teknik validasi model yang digunakan untuk mengevaluasi kemampuan suatu model statistik dalam menggeneralisasi pada kumpulan data independen. Teknik ini terutama digunakan untuk memprediksi kinerja model dan mengestimasi tingkat akurasi model prediktif saat digunakan dalam situasi praktis [18]. *Cross validation* bekerja dengan membagi total instance data menjadi N bagian. Pada setiap iterasi, satu bagian dari data diambil sebagai data pengujian (testing), sementara bagian lainnya digunakan sebagai data pelatihan (training). Proses ini berulang untuk setiap bagian hingga seluruh data telah digunakan sebagai bagian pengujian sekali [19].

## 2.6 Naïve Bayes

Naïve Bayes ialah tata cara pengklasifikasian probabilistik simpel. Model ini hendak menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi serta campuran nilai dari dataset yang diberikan. Tata cara Naïve Bayes berpendapat seluruh atribut pada tiap jenis tidak mempunyai ketergantungan satu sama lain Naïve Bayes teruji mempunyai akurasi serta kecepatan yang besar dikala diaplikasikan ke dalam database dengan informasi

yang besar. Keuntungan pemakaian Naïve Bayes ialah cuma membutuhkan beberapa kecil informasi latih buat memastikan parameter mean serta varians dari variabel yang dibutuhkan buat klasifikasi [20].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hiotesis H berdasar kondisi X (Posteriori Probabilitas)

P(H) : Probabilitas Hipotesis H (Prior Probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H

P(X) : Probabilitas X

## 2.7 Particle Swarm Optimization (PSO)

*Particle Swarm Optimization (PSO)* merupakan salah satu metode dan teknik untuk mengklasifikasi dan meningkatkan akurasi, dimana *Particle Swarm Optimization (PSO)* terdiri dari sekumpulan partikel yang mencari posisi yang terbaik [21]. *Particle Swarm Optimization (PSO)* didasarkan pada kebiasaan sekelompok burung atau ikan yang berenang bergerombol, ratusan burung atau ikan ini memiliki kemampuan untuk bergerak dengan cepat tanpa gangguan meskipun mereka berada pada jarak yang sangat dekat satu sama lain [22]. Dibandingkan dengan algoritma matematika dan metode optimasi heuristik lainnya, metode *Particle Swarm Optimization (PSO)* lebih unggul, mulai dari konsepnya yang sederhana, mudah digunakan dan efisien dalam perhitungan [23].

## 2.8 Rapidminer

RapidMiner adalah perangkat lunak yang membantu orang belajar ilmu data mining. Perusahaan membuat platform yang didedikasikan untuk semua langkah yang melibatkan banyak data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran [24]. Dengan penerapan RapidMiner, tidak diperlukan keahlian khusus dalam pemrograman, karena segala fasilitasnya telah tersedia dan RapidMiner difokuskan untuk keperluan data mining dengan model yang disediakan cukup banyak [25].

RapidMiner menawarkan antarmuka *drag-and-drop* yang memungkinkan pengguna untuk membangun alur kerja untuk memproses dan menganalisis data. Ini mendukung beragam sumber data, termasuk file datar, basis data, dan platform big data seperti hadoop dan spark. Perangkat lunak ini juga mencakup beragam operator yang sudah dibangun yang merupakan blok bangunan dari alur kerja, yang mencakup semua tahap proses *data mining*, seperti pembersihan data, pemilihan fitur, dan pemodelan [26].

## 2.9 Studi Pustaka

Studi pustaka adalah proses dimana dilakukannya studi pengkajian data tergantung pada metode dan permasalahan yang dibahas. Studi pustaka bisa diperoleh dari membaca buku-buku referensi, artikel ataupun jurnal penelitian yang pernah dilakukan sebelumnya supaya penelitian ini bisa lebih baik dan memiliki dasar yang kuat untuk dilakukan penelitian. Berikut adalah daftar referensi terkait dengan penelitian.

Tabel 2.2 Studi Pustaka

No	Judul	Peneliti	Tahun terbit	Hasil/Kesimpulan
1.	Metode C4.5 Dan Naïve Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja	Meilany Nonsi Tentua, Vicky Fidiantoro dan Pradana Feri Ariyanto	2022	Berdasarkan metode C4.5 atribut berat badan merupakan atribut yang paling berpengaruh dalam penelitian ini. Dibuktikan dengan atribut <i>Leave</i> menduduki <i>root node</i> . Pada perhitungan Naïve Bayes data dibagi menjadi 2 bagian yaitu data <i>training</i> dan data <i>testing</i> . Untuk data kesehatan mental yang berjumlah 1259 kita bagi menjadi dua yaitu 1200 untuk data training dan 59 untuk data testing. Selanjutnya menghitung Probabilitas dari masing masing atribut.

**Table 2.2** Studi Pustaka (Lanjutan)

2.	Tweet Analysis of Mental Illness Using K-Means Clustering and Support Vector Machine	Kartikadyota Kusumaningtyas, Muhammad Habibi, Irmma Dwijayanti, Retno Sumiyarini	2023	<p>Tujuan utama penelitian ini adalah melakukan analisis data tweet terkait gangguan kesehatan mental menggunakan menggunakan K-Means clustering dan Support Vector Machine. Untuk melakukan analisis tersebut, penelitian ini melakukan dua tahapan analisis yaitu pengelompokan data untuk mencari topik pembahasan menggunakan K-Means clustering dan analisis sentimen untuk mencari nilai polaritas positif dan negatif. (1) Hasil clustering, didapatkan 5 cluster pembicaraan terkait gangguan kesehatan mental. Sedangkan hasil analisis sentimen, semua cluster memiliki jumlah kelas sentimen negatif yang lebih banyak dibandingkan dengan positif. (2) Cluster yang memiliki jumlah nilai sentiment negatif paling signifikan adalah cluster 4 dan cluster 5. Cluster tersebut membahas mengenai pentingnya konseling dengan psikiater dan psikolog jika mengalami gangguan kesehatan</p>
----	--------------------------------------------------------------------------------------	----------------------------------------------------------------------------------	------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 2.2 Studi Pustaka (Lanjutan)

				mental serta mengenai pembiayaan pengobatan gangguan kesehatan mental melalui layanan BPJS Kesehatan.
3.	Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5	Siska Febriani, Heni Sulistiani	2021	Penerapan algoritma C4.5 untuk mengklasifikasi gangguan psikologis memiliki nilai akurasi sebesar 57.50%, mean precision sebesar 57.50%, <i>classification error</i> sebesar 0%, dan recall sebesar 57.50% untuk data <i>training</i> sedangkan data <i>testing</i> memiliki tingkat akurasi sebesar 72.67 %, <i>precision</i> sebesar 72.67%, dan <i>recall</i> sebesar 100%.
4.	Lexicon and Naïve Bayes Algorithms to Detect Mental Health Situations from Twitter Data	Sheila Shevira, I Made Agus Dwi Suarjaya dan Putu Wira Buana	2022	Sebagian besar twit dikategorikan sebagai 'normal', oleh karena itu status kesehatan mental tampak aman. Namun, temuan ini perlu diperiksa ulang di masa depan, terutama di Provinsi DKI Jakarta, yang memiliki kasus gangguan mental terbanyak. Studi ini menemukan bahwa algoritma Naïve Bayes lebih akurat (85,5%) daripada algoritma berbasis leksikon. Hal ini dapat

**Table 2.2** Studi Pustaka (Lanjutan)

				ditingkatkan dalam studi mendatang dengan meningkatkan kinerja pada tahap pra-pemrosesan.
5.	Data Mining Algorithms and Techniques in Mental Health: A Systematic Review	Susel Góngora Alonso, Isabel de la Torre-Díez, Sofiane Hamrioui, Miguel López- Coronado, Diego Calvo Barreno, Lola Morón Nozaleda, Manuel Franco	2018	Academic databases that were used to perform the searches are Google Scholar, IEEE Xplore, PubMed, Science Direct, Scopus and Web of Science, taking into account as date of publication the last 10 years, from 2008 to the present. Several search criteria were established such as ‘techniques’ AND ‘Data Mining’ AND ‘Mental Health’, ‘algorithms’ AND ‘Data Mining’ AND ‘dementia’ AND ‘schizophrenia’ AND ‘depression’, etc. selecting the papers of greatest interest. A total of 211 articles were found related to techniques and algorithms of Data Mining applied to the main Mental Health diseases. 72 articles have been identified as relevant works of which 32% are Alzheimer’s, 22% dementia, 24% depression, 14% schizophrenia and 8% bipolar disorders. Many of the papers

**Table 2.2** Studi Pustaka (Lanjutan)

				<p>show the prediction of risk factors in these diseases. From the review of the research articles analyzed, it can be said that use of Data Mining techniques applied to diseases such as dementia, schizophrenia, depression, etc. can be of great help to the clinical decision, diagnosis prediction and improve the patient's quality of life.</p>
--	--	--	--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------