

**BAB II**  
**LANDASAN TEORI**

**2.1 Tinjauan Pustaka**

Di dalam penelitian ini, menggunakan sejumlah jurnal acuan dalam penelitian terdahulu, untuk penambahan wawasan penelitian yang sedang dilakukan. Berikut ini merupakan penelitian jurnal terdahulu yang digunakan untuk menentukan Tinjauan Pustaka pada penelitian ini, dapat dilihat pada **Tabel 2.1** sebagai berikut:

**Tabel 2. 1** Tinjauan Pustaka

No.	Peneliti,Tahun	Judul	Hasil
1.	Siska Febriani, Heni Sulistiani ,2021	Analisis data hasil diagnosa untuk klasifikasi gangguan kepribadian menggunakan <i>algoritma c4.5</i>	Hasil dari penelitian ini merupakan prediksi gangguan kepribadian untuk mengaktualkan informasi yang dibutuhkan oleh masyarakat.
2.	Rika Melyanti, Eka Sabna,2022	penerapan klasifikasi <i>data mining</i> untuk prediksi tingkat kemampuan pemahaman skripsi mahasiswa	Hasil prediksi menunjukkan bahwa 55 mahasiswa yang melakukan penelitian web, 13 mahasiswa yang domain dengan kemampuan yang baik.

3.	Joshua Bonardo Junior , Rd. Rohmat Saedudin, Vandha Pradiwiyasma Widharta,2021	perbandingan akurasi algoritma <i>decision tree</i> dan algoritma <i>support vector machine</i> pada penyakit diabetes.	klasifikasi dengan algoritma <i>Decision Tree</i> menghasilkan akurasi sebesar 85.93% dengan rasio data training dan data testing sebesar 75:25. Penerapan <i>K-Fold Cross Validation</i> dengan 10 <i>fold</i> hasil rata-rata yang didapatkan sebesar 80.72%.
4.	Fandi Yulian Pamuji , Viry Puspaning Ramadhan,2021	Komparasi Algoritma <i>Random Forest</i> Dan <i>Decision Tree</i> Untuk Memprediksi Keberhasilan <i>Immunotherapy</i> .	Berdasarkan hasil pengujian dengan menggunakan <i>Cross-Validation</i> maka didapatkan nilai akurasi 844 % dengan menggunakan rancang <i>Decision Tree</i> kemudian nilai akurat 855% dengan memerlukan metode <i>Random Forest</i> .
5.	Novia Hasdyana, Rozzi Kesuma Dinata, 2021.	Analisis <i>Matthew Correlation Coefficient</i> pada <i>K-Nearest Neighbor</i> dalam Klasifikasi Ikan Hias.	Klasifikasi yang ikan hias yang digunakan dalam penelitian ini yaitu terdiri dari 3 kelas antara lain

			<i>premium, medium dan low.</i>
--	--	--	---------------------------------

Penelitian yang telah dijabarkan merupakan sumber dari penelitian sebelumnya yang dapat ditampilkan pada daftar pustaka pada penelitian ini. Berikut merupakan penjelasan tinjauan pustaka antara lain:

Pada penelitian yang dilakukan oleh Siska Febriani, Heni Sulistiani dengan judul Analisis data keputusan diagnosa untuk klasifikasi gangguan kepribadian menggunakan algoritma *C4.5* dengan kiat *data mining* klasifikasi data *C4.5*. pada permasalahan ini mengangkat permasalahan gangguan kepribadian terlebih di masa pandemic *Covid 19* dengan tujuan akurasi yang sesuai. Hasil dari penelitian tersebut yaitu Penerapan algoritma *C4.5* untuk mengklasifikasi gangguan psikologis memiliki nilai akurasi sebesar 57.50%, *mean precission* sebesar 57.50%, *classification error* sebesar 0%, dan *recall* sebesar 57.50% untuk data *training* sedangkan data *testing* memiliki tingkat akurasi sebesar 72.67 %, *precission* Penerapan algoritma *C4.5* untuk mengklasifikasi gangguan psikologis memiliki nilai akurasi sebesar 57.50%, *mean precission* sebesar 57.50%, *classification error* sebesar 0%, dan *recall* sebesar 57.50% untuk data *training* sedangkan data *testing* memiliki tingkat akurasi sebesar 72.67 %, *precission*, nilai klasifikasi akurasi yang didapat pada penerapan algoritma *C4.5* menghasilkan nilai lebih dari 60% yaitu akurasi data 72.67%.

Dalam penelitian ini dilakukan oleh Rika Melyanti, Eka Sabna, pada tahun 2022 dengan judul Penerapan Klasifikasi *Data Mining* Untuk Prediksi Tingkat Kemampuan Pemahaman Skripsi Mahasiswa. Pada penelitian ini mengangkat permasalahan latar belakang pada

pembuatan skripsi mahasiswa, penerapan pada *data mining* pada data skripsi mahasiswa sangatlah bermanfaat guna membantu fakultas maupun universitas untuk memprediksi nilai yang dihasilkan pada pembuatan skripsi tersebut dengan masing-masing judul yang telah dibuat oleh mahasiswa tersebut. Dengan hasil perhitungan Hasil pengujian digunakan *Confusion Matrix*, dengan nilai *Accuracy* Algoritma *C4.5* adalah sebesar 71,00%. Dalam penelitian yang dilakukan oleh Joshua Bonardo Junior , Rd. Rohmat Saefudin, Vandha Pradiwiyasma Widharta, pada tahun 2021 dengan judul Perbandingan Akurasi Algoritma *Decision Tree* Dan Algoritma *Support Vector Machine* Pada Penyakit Diabetes. Pada penelitian ini mengangkat permasalahan pada penyakit diabetes, banyak sekali penderita yang terdiagnosis penyakit diabetes setelah terjadi komplikasi pada dirinya, seharusnya jika terdeteksi lebih awal maka dapat langsung melakukan penanganan sehingga kecil kemungkinan terjadinya hal-hal yang tidak diinginkan. *Data mining* merupakan teknik yang dapat dipakai untuk memprediksi penyakit tersebut, melalui variabel-variabel yang digunakan sehingga menghasilkan prediksi angka diabetes. Pada penelitian ini, hasil akurasi yang didapat algoritma *Decision Tree* sebesar 85.28%, sedangkan hasil akurasi yang didapat dari algoritma *Support Vector Machine* sebesar 83.85%.

Dalam penelitian yang dilakukan oleh Fandi Yulian Pamuji , Viry Puspaning Ramadhan pada tahun 2021 dengan judul Komparasi Algoritma *Random Forest* Dan *Decision Tree* untuk perkiraan keunggulan *Immunotherapy* pada penelitian ini dilaksanakan dengan memprediksi kemakmuran metode pengobatan *immunotherapy* pada penyembuhan penyakit kutil. Arahan penelitian menggunakan dua metode tersebut untuk memahami komparasi metode *Random Forest* dan *Decision Tree* dalam pemulihan penyakit. Hasil dari Komparasi Algoritma *Random Forest* Dan *Decision Tree* Untuk Memprediksi Keberhasilan *Immunotherapy* proses penelitian diperoleh metode *data mining* menggunakan algoritma *Naive Bayes* dengan *10-Fold Cross*

*Validation* menghasilkan nilai akurasi klasifikasi sebesar 81,11%. klasifikasi menggunakan algoritma *Neural Network* dengan nilai akurasi sebesar 80,00%.

Dalam penelitian yang dilakukan oleh Fandi Yulian Pamuji , Viry Puspaning Ramadhan pada tahun 2021 dengan judul penelitian Analisis *Matthew Correlation Coefficient* pada *K-Nearest Neighbor* dalam Klasifikasi Ikan Hias. Pada klasifikasi *K-Nearest Neighbor* menunjukkan bahwa nilai Mathew correlation coefficient pada Euclidean distance diperoleh nilai *MCC* tertinggi pada *class medium* sebesar 0,78642, nilai *MCC* tertinggi kedua pada *class premium* yaitu sebesar 0,56734, dan nilai *MCC* terendah pada *class low* yaitu sebesar 0,435269.

## 2.2 Keaslian Penelitian

Mengenai sejumlah hal yang menjadi pengimbang antara penelitian yang dilangsungkan oleh penulis dengan penelitian yang telah dilakukan sebelumnya sebagaimana tersebut pada tabel tinjauan, antaranya :

1. Melakukan riset literatur secara komprehensif untuk memahami teori-teori yang berkaitan dengan gangguan jiwa pada masyarakat Indonesia, algoritma *Decision Tree*, *K-Nearest Neighbor*. Hal ini dilakukan untuk memastikan bahwa teori yang digunakan dalam penelitian telah dipahami dengan baik dan dapat diaplikasikan dengan benar.
2. Melakukan pengumpulan data dengan cara yang akurat dan valid. Data yang dipakai dalam penelitian berasal dari dataset data.go.id atau satu data Indonesia, yang bersumber langsung dari Kementerian Kesehatan RI yang telah diperoleh dengan cara yang sah dan dapat dipercaya.
3. Menerapkan metodologi penelitian yang tepat dan valid. Dalam penelitian ini, metode *Decision Tree* dengan *K-Nearest Neighbor* dipilih untuk membangun model prediksi

kesehatan jiwa. Metode ini telah terbukti efektif dalam berbagai penelitian sebelumnya dan diperkirakan sebagai salah satu metode yang paling efektif dalam membangun model prediksi.

4. Melakukan analisis data dengan cara yang benar dan valid. Menjabarkan data dalam penelitian ini dilaksanakan dengan menggunakan Algoritma *Decision Tree* dengan *K-Nearest Neighbor* untuk membangun model prediksi kesehatan jiwa. Selanjutnya, model prediksi tersebut diuji dengan memerlukan data uji untuk memeriksa akurasi dan keefektifannya.

### 2.3 Kesehatan Jiwa

Berdasarkan WHO (Organisasi Kesehatan Dunia), kesehatan merupakan perihal yang termasuk aspek fisik, mental dan sosial yang tidak hanya bebas penyakit atau Cacat, jadi kesehatan mental adalah bagian integral darinya kesehatan umum perseorangan, yang tidak berarti hanya bebas dari gangguan mental tapi lebih berkualitas untuk hidup lebih bahagia. Ada empat persoalan kesehatan utama di negara maju, modern dan *industrial* yaitu penyakit *degeneratif*, kanker, gangguan jiwa dan kecelakaan. Salah satu dari empat masalah kesehatan teratas dari negara maju, modern dan berkembang yaitu gangguan kejiwaan, meskipun gangguan jiwa tersebut tidak dianggap sesuatu yang langsung menyebabkan kematian, tetapi gangguan mental dapat mempengaruhi kehidupan seseorang dalam aktivitas sehari-hari. Gangguan jiwa adalah penyakit yang ditimbulkan oleh gangguan kesehatan mental, gangguan pikiran, persepsi dan perilaku yang individu tidak mampu beradaptasi dengan diri sendiri, orang lain, masyarakat dan lingkungan. Saat ini fakta gangguan kesehatan jiwa semakin meningkat yang sangat signifikan, dan jumlah orang yang terkena dampak semakin banyak setiap tahun di berbagai belahan dunia gangguan jiwa semakin meningkat. Ada sekitar 35 juta orang 60 orang dengan depresi mengalami gangguan bipolar, 21

juta menderita *skizofrenia* dan 475 juta. demensia yang didapat , inilah masalahnya sekitar 4785 juta orang di seluruh dunia menderita gangguan jiwa.

(Yuli Permata Sari. 2019).

Kesehatan jiwa merupakan suatu hal yang sangat penting dalam kehidupan seseorang lantaran dengan jiwa yang sehat, manusia mampu mengembangkan fisik dan mental serta memelihara kaitan sosial yang baik ideal, mampu bersosialisasi dengan domain, memenuhi segala kebutuhannya dan kehidupan berkelompok. Kesehatan jiwa merupakan keadaan sejahtera dari badan, jiwa dan kehidupan sosial yang memperbolehkan setiap orang, seseorang hidup secara sosial dan ekonomi. Diasumsikan 2-3% penduduk Indonesia menderita masalah kesehatan jiwa setiap tahunnya. Jika pasien tidak berbahaya, tidak masalah di lingkungan sekitar untuk mencari makanan. Perawatan yang dikhususkan untuk gangguan jiwa yaitu termasuk pengobatan dengan *farmakologis*, tetapi juga pengobatan *psikoterapi* dan bentuk pengobatan yang searah dengan gejala pasien atau kondisi medis yang mendukung perawatan pasien gangguan jiwa. Dengan bentuk terapi ini dukungan keluarga dan sosial diharapkan dapat meningkatkan kesembuhan penderita, dan penderita merasa berguna dalam masyarakat dan tidak merasa terasing dengan penyakitnya. Ketika orang dengan masalah kesehatan mental diperlakukan sebagai pasien rawat jalan atau rawat inap di rumah sakit jiwa, diharapkan keluarga dapat tetap memberikan perhatian dan dukungan sesuai arahan tim medis rumah sakit. Jenis dukungan keluarga seperti dukungan harapan, dukungan nyata, dukungan pengetahuan dan dukungan emosional. Tetapi, banyak keluarga yang tidak memperdulikan hal ini. Banyak keluarga yang menyerahkan kesembuhan pasien sepenuhnya kepada petugas kesehatan. Banyak pasien dengan penyakit mental telah diabaikan oleh keluarga mereka keluarga melupakannya. Banyaknya yang tidak merawatnya sedangkan pasien dimasukkan ke rumah sakit jiwa. Padahal, jika keluarganya

rajin berkunjung dan untuk mendukung pasien dengan masalah kesehatan mental, ini adalah salah satu bentuk pengobatan terbaik untuk pemulihannya. Namun, jika keluarganya tidak merawatnya, penderita akan membutuhkan waktu lebih lama untuk pulih karena pasien merasa diabaikan oleh keluarganya (Suhermi S, 2019).

Sudah banyak orang di dunia ini bahkan di Indonesia yang menderita masalah kesehatan jiwa banyak juga orang yang menderita masalah kesehatan jiwa mulai dari anak muda, dewasa, anak-anak orang tua atau lanjut usia dapat memiliki penyakit ini. Asumsi bahwa kesehatan kejiwaan adalah keadaan emosional, psikologis dan kemasyarakatan terlihat dari hubungan komunikatif antara dua individu yang tidak dipenuhi oleh atau tidak berjalan dengan baik, kemudian sesuatu yang sulit dipahami melalui diri yang baik dan keseimbangan perasaan diri penderita. Selain gangguan jiwa, ada juga masalah psikososial umum terjadi di kalangan masyarakat Indonesia. Kejadian yang terjadi pada psiko sosial merupakan masalah yang sering berjalan di masyarakat. Psikososial adalah suatu kondisi yang ada pada diri seseorang dan mencakup aspek kejiwaan sosial atau sebaliknya. Psikososial mengacu pada hubungan sosial yang melibatkan faktor psikologis. Menurut definisi di atas, masalah psikososial adalah masalah yang timbul secara spiritual dan sosial. Psikososial (psikososial) adalah hubungan antara kesehatan mental atau emosional seseorang dengan kondisi sosial. Istilahnya psikososial kombinasi psikologis dan sosial. Artinya, konsep pembangunan Perkembangan psikososial mengacu pada emosi atau cara berpikir dalam hubungannya dengan orang lain. Begitulah urutan psikososial Perubahan yang dikaitkan dengan hubungan sosial atau serangkaian kepribadian.

Kompleks dan menuntut untuk mendukung orang lain dalam melakukan aktivitas kehidupan fungsional sehari-hari. Ada perubahan model menyebabkan pemikiran, perilaku dan emosi berbagai batasan hidup penting gangguan mental yang serius dapat terjadi merupakan



beban bagi setiap pihak, termasuk pemerintah keluarga dan masyarakat karena terbelang mengurangi produktivitas pasien mengeluarkan biaya yang sangat besar keluarga pengobatan orang dengan masalah kesehatan mental bobot efektif yang lebih rendah dapat menyebabkan hal ini memburuknya gejala gangguan jiwa. Stigma dan konformitas memperburuk gejala kekambuhan pasien setelah operasi perawatan rumah sakit dan kembali untuk tinggal di sana rumah dalam hal ini pemerintah melakukan berbagai tindakan terkait merawat pasien dengan gangguan jiwa berat masyarakat yang meliputi pelayanan kesehatan program hidup dan pembuangan terpadu pasung, yang hidup selama beberapa tahun terakhir. Tapi untuk saat ini masih ada

rantai dan penyalahgunaan orang dengan masalah kesehatan mental yang parah di Indonesia karena stigma dan diskriminasi orang sakit jiwa yang masih sehat. Optimalkan pemulihan dari gangguan jiwa berat membutuhkan pendekatan holistic dan integrasi layanan kesehatan mental dan masyarakat, untuk menyenangkan mereka yang menderita berinteraksi dengan orang lain, dapatkan konsep positif dan meyakinkan diri sendiri kesejahteraan hidup mereka terlepas. Proses pemulihan kesehatan mental bukan hanya tentang itu

tetapi juga dalam keprihatinan menyesuaikan dengan kebutuhan pasien. hiburan

Gangguan mental lebih efektif bila Perhatikan individualitas yang bersangkutan, libatkan keluarga dan masyarakat sebagai sistem pendukung.

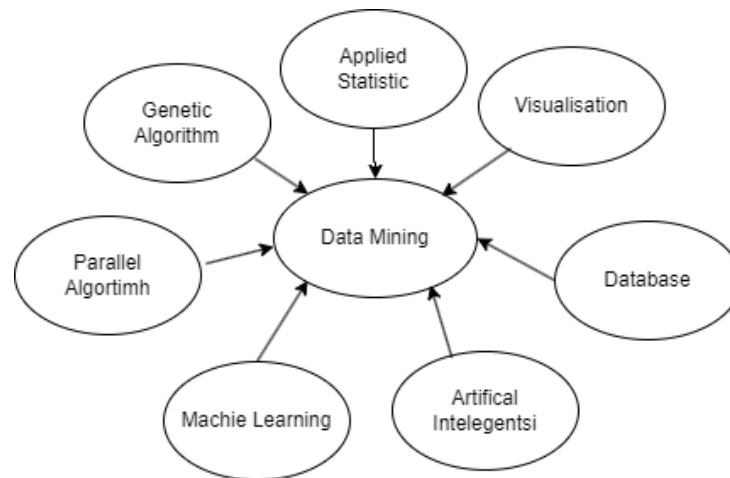
Menurut Edi Syahputra, Kintoro Rochadi, Jek Amidos Pardede, Donal Nababan, Frida Lina Tarigan (2021) Gangguan jiwa adalah masalah jiwa yang serius mengakibatkan hambatan *kognitif, afektif* dan sosial terhadap peran individu tidak dapat melakukan aktivitas sehari-hari Organisasi Kesehatan Dunia pada tahun 2019 mengatakan bahwa gangguan jiwa mencakup depresi dan gangguan kecemasan gangguan bipolar, skizofrenia dan psikosis lainnya, demensia

dan gangguan perkembangan termasuk autisme. Gangguan jiwa dapat menyerang banyak orang, dimulai dari anak-anak. siswa SMA dari orang dewasa hingga lanjut usia. Saat ini gangguan jiwa terus meningkat setiap tahunnya (Maulana et al., 2019). Data yang dikeluarkan oleh WHO tahun 2018 ditunjukkan lebih dari 300 juta orang dunia menderita depresi, lebih dari 60 juta orang di dunia menderita gangguan bipolar dan 23 juta orang. Yang menderita masalah psikologis betul-betul seperti skizofrenia dan psikosis lainnya. Sesuai dengan situasi global, data dasar survei yang dilakukan oleh kementerian kesehatan 2018 (Riskesdas) dipertimbangkan. jumlah rumah naik dari 1,7 pada 2013 menjadi 7 dari seribu pada 2018 artinya dalam 7 dari 1000 rumah tangga terdapat anggota keluarga yang mengalami gangguan kesehatan jiwa. Lebih dari 19 juta orang di atas usia 15 tahun mengalami gangguan jiwa perkiraan 12 juta orang di atas usia 15 tahun menderita depresi.

#### **2.4 Data Mining**

*Data mining* yaitu proses dalam melakukan teknik statistik, kecerdasan buatan, *machine learning*, agar *mengekstraksi* dan mengidentifikasi suatu informasi yang bermanfaat dan *ekstraksi* informasi pengetahuan yang berkaitan dengan data-data besar (Boby Septia Pratama, Dito Putro Utomo, 2020). Namun secara sederhananya *data mining* merupakan suatu pengembangan atau suatu ditemukan informasi dengan mencari sejumlah data tertentu untuk di analisis. *Data mining* dapat disebut juga dengan suatu yang berfungsi untuk menggali informasi yang berupa pengetahuan yang tidak dapat dideteksi secara manual dari serangkaian kumpulan data (Wulan Lestari, Fatoni, Hutrianto, 2020).

*Data mining* adalah langkah analisis terhadap suatu proses wawasan di dalam basis *knowledge discovery in database (KDD)*. *KDD* merupakan hubungan dengan *integrasi* dengan penemuan ilmiah, *interpretasi* dan *visualisasi* dari pola data.



**Gambar 2. 1**Tahapan *Data Mining*

(Sumber: Siska Febriani, Heni Sulistiani, 2021)

Terdapat beberapa faktor yang menjadi tujuan utama untuk menggunakan *data mining*, adapun tujuan tersebut yaitu (Siska Febriani, Heni Sulistiani, 2021):

1. Besarnya data yang dikumpulkan maka dari itu membutuhkan waktu yang lama dan seorang ahli untuk menguraikan.
2. Komputer PC merupakan salah satu pilihan dikarenakan memiliki kemampuan dalam kecepatan, ketepatan, dan tidak akan lelah dalam melakukannya.
3. Kontroversi dari kompetisi bisnis yang akan terus menerus bersaing menjadikan esensial dan harus dimiliki.
4. Mampu mendapatkan pemikiran yang berbeda.

Menurut (H. Rizki and R. M. Agus,2021) *data mining* adalah salah suatu tindakan dalam aspek perangkat program dapat memberikan *ROI (Return of Investment)* yang panjang. Hal yang perlu diamati adalah bahwa data mining berbeda dengan *query tools*. *Query* dan data mining adalah dua hal yang saling membantu. Eksistensi *data mining* bukan untuk menggantikan *query* tetapi ditambahkan sejumlah tambahan yang berarti. Jika menggunakan *query* sederhana maka

informasi yang dapat diakses sekitar 80% dari data yang ada dalam basis data sedangkan 20% serta akan menjadi keterangan yang disembunyikan yang perlu teknik-teknik khusus dalam mengaksesnya. Karena *data mining* adalah suatu kaitan proses maka dibagi menjadi sejumlah tahap antara lain:

- a. Membersihkan data guna menyisihkan data yang tidak selaras *noise*.
- b. *Integritas* data perlu menyambungkan data dari beberapa sumber.
- c. *Transformasi* data untuk merubah data menjadi ukuran yang bertimbang untuk di mining.
- d. Aplikasi teknik *data mining*.
- e. Penilaian rupa yang ditentukan untuk mendapatkan *informan* yang menarik atau pun bernilai.
- f. Penyajian pemahaman dengan sistem visual.

## 2.5 *Decision Tree*

Algoritma *decision tree* yaitu algoritma yang mengklasifikasikan yang digunakan mengeksploitasi data yang akan menemukan jumlah variabel tersembunyi antara jumlah variabel input dengan variabel target (Khusnul Khotimah, 2021). *Decision tree* adalah salah satu algoritma yang digunakan dalam membuat pohon keputusan, yang biasa digunakan untuk memprediksi suatu data dalam jumlah data yang cukup besar sehingga mendapatkan tingkat akurasi yang tinggi dan dapat menghasilkan pohon keputusan yang mudah diinterpretasikan (Christin ndari Dengen, Kursini, Emha Taufiq Luthfi, 2020).

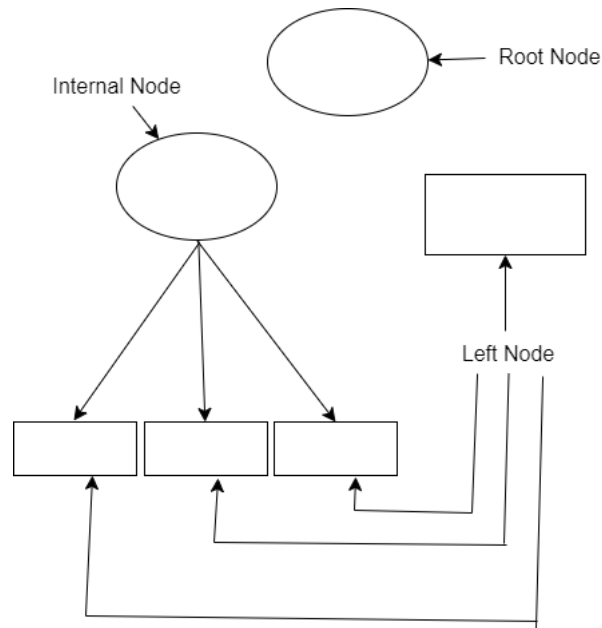
*Decision tree* adalah metode pengklasifikasian dan metode prediksi yang sungguh akurat dan sering dikenal. Metode *decision tree* merubah kenyataan yang banyak menjadi pohon keputusan yang mempresentasikan aturan. Aturan yang gampang dipahami dengan bahasa alami. *Decision tree* dapat juga dalam wujud basis data seperti *Structure Query Language (SQL)* untuk

memperoleh rekaman pada data tertentu. Selain itu, *decision tree* merupakan sebuah struktur yang berfungsi untuk membagi sebuah data yang besar menjadi himpunan *record* yang kecil dengan penerapan dalam rangkaian hasil keputusan (Yusmi Nur Aini, Eva Yulia Puspaningrum, Wahyu Syaifullah Jauharis Saputra, 2021) .

Secara umum algoritma *decision tree* guna membangun pohon keputusan yaitu sebagai berikut ( Meilany Nonsi Tentual, Vicky Fidiantorol, Pradana Feri Ariyanto, 2022) :

1. Memilah atribut sebagai akar.
2. Menjadikan cabang tiap-tiap nilai.
3. Memisah kasus dalam cabang.
4. Menyalin proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

*Decision tree* merupakan pembagian perkumpulan *dataset* yang banyak menjadi himpunan *record* menjadi lebih rendah dengan penerapan kesesuaian susunan keputusan. Setiap simpul daun pada *decision tree* menemui label *class*, dimana simpul yang lain dari simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian *record* yang mendapati karakter yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk persegi empat. Struktur *decision tree* dapat ditampilkan pada gambar sebagai berikut.



**Gambar 2. 2** Struktur *Decision Tree*

Sumber : Popon Handayani, Elah Nurlelah, Mugi Raharjo ,Panji Madya Ramdani (2019)

Menurut Popon Handayani, Elah Nurlelah, Mugi Raharjo ,Panji Madya Ramdani (2019)

Terdapat konsep dalam *decision tree* antara lain yang akan dilakukan pada tahapannya :

1. Data dijelaskan dalam gambaran tabel pada atribut dan *record*.
2. Karakter mencetuskan suatu parameter yang diciptakan dijadikan tolak ukur dalam penyusunan *tree*.
3. Atribut mempunyai skor yang dinamakan *instance*.

Dalam meningkatkan *data mining* teknik yang digunakan dalam *classification* dengan algoritma *decision tree* pada *C4.5* terdapat beberapa tahapan antara lain :

- a. Menyediakan data *training* dan *testing*. Data *training* dan *testing* biasa dikenakan dari dataset yang ada sebelumnya dan sudah dikelompokan dalam kelas tertentu.
- b. Membuat perhitungan pohon. Sumber yang terpilih dari atribut dengan cara perhitungan angka *gain* dari setiap karakter, hasil yang paling tertinggi akan menjadi akar utama. Hasil

dari menghitung *gain* maka terlebih dahulu menghitung *entropy*. Rumus dalam melakukan perhitungan *entropy* antara lain:

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

Pembahasan :

$S$  = Himpunan kasus

$n$  = Jumlah partisi  $S$

$P_i$  = Proporsi  $S_i$  terhadap  $S$

Dilanjutkan dengan dihitung nilai *gain* dengan menggunakan rumus sebagai berikut :

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (2)$$

Pembahasan :

$S$  = Himpunan Kasus

$A$  = Fitur

$N$  = Jumlah Partisi atribut  $A$

$|S_i|$  = Proporsi  $S_i$  terhadap  $S$

$|S|$  = Jumlah kasus dalam  $S$

- c. Mengulangi langkah kedua dan ketiga sampai batas semua *record* terpartisi
- d. Proses partisi *record* akan berhenti pada saat melalui proses :
  - a. *Record* dalam simpul  $N$  mendapatkan klasifikasi persis.
  - b. Atribut pada record tidak ada yang dipartisi.
  - c. Kedapatan cabang kosong pada *record*.

## 2.6 *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah  $K$  data terdekat (tetangganya) sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasikan data berdasarkan *similarity* atau kemiripan atau kedekatannya terhadap data lainnya. Secara umum, cara kerja Algoritma *K-Nearest Neighbor* adalah sebagai berikut:

1. Tentukan jumlah tetangga  $K$  yang akan digunakan untuk pertimbangan menentukan kelas.
2. Hitung jarak dari data baru ke masing-masing data point di dataset.
3. Ambil sejumlah  $K$  data dengan jarak terdekat, kemudian tentukan kelas dari data baru tersebut.

Rumus dari Algoritma *K-Nearest Neighbor* adalah sebagai berikut :

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

keterangan :

$d(x,y)$  : jarak antara data  $x$  dan data  $y$

$x_i$  : data uji ke- $i$

$y_i$  : data latih ke- $i$



## 2.7 Pengujian *Confusion matrix*

Menurut Elsa Paskalis Krisda Orpa, Eva Faja Ripanti , Tursina (2019) *matrix* merupakan suatu metode yang dapat digunakan untuk penghitung akurat pada data dalam *data mining*. *Confusion matrix* ini akhirnya akan dilakukan perhitungan yang melakukan 4 keluaran, adalah *recall* (porsi kasus yang identifikasinya dengan benar), *precision* (proporsi kasus dengan hasil positif yang benar), *accuracy* (perbandingan kasus yang diidentifikasi benar dengan jumlah seluruh kasus) dan *error rate* (kasus yang diidentifikasi salah dengan jumlah seluruh kasus). Pada penelitian ini penulis menggunakan *matrix* berukuran 3x3 yang terdiri dari sembuh, tidak dan perawatan dengan 3 jenis tingkat prediksi.

**Tabel 2. 2** *Confusion Matrix* 3x3

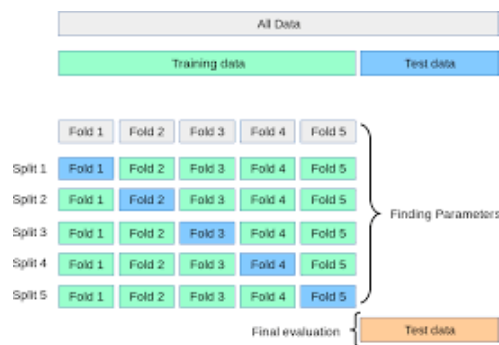
<b>Aktual</b>	<b>Sembuh</b>	<b>Tidak</b>	<b>Perawatan</b>
<b>Sembuh</b>	TPss	FPss	FPsm
<b>Tidak</b>	FNps	TNpp	FNpm
<b>Perawatan</b>	FNms	FNmp	TNmm

Rumus perhitungan dari *Confusion matrix* yaitu :

$$Accuracy = \frac{Tpss+TNpp+TNmm}{Tpss+FPsp+SPsm+FNps+FNpm+FNms+FNms+FNmp+TNmm} \quad (4)$$

## 2.8 *K-fold Cross Validation*

Pengujian *K-Fold* adalah suatu teknik validasi model yang digunakan untuk menghindari *overfitting* pada data pelatihan. Teknik ini membagi data pelatihan menjadi *K-fold* subset yang sama besar dan kemudian melakukan iterasi *K-fold* kali. Pada setiap iterasi, salah satu subset menjadi data uji *K* dan subset lainnya digunakan sebagai data pelatihan. Akurasi model dihitung sebagai rata-rata akurasi pada setiap iterasi. *Cross validation* juga merupakan salah satu pengambilan sampel metode ulang yang diperlukan dalam statistik.



**Gambar 2. 3** Contoh *K-Fold CV*

Farhan Hidayat Zulfallah (2022)