

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Kajian Pustaka

Pada penelitian ini, peneliti mencantumkan beberapa penelitian terdahulu yang relevan sebagai acuan dalam pemahaman bagi peneliti. Kajian pustaka selanjutnya akan disajikan pada Tabel berikut.

Tabel 2.1 Kajian Pustaka

No	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
1	Prediksi Kedatangan Wisatawan Pada Pariwisata Kota Batam Dengan Menggunakan Teknik <i>Knowledge Data Discovery</i>	Penelitian tersebut dilakukan untuk prediksi wisatawan yang datang ke kota menggunakan Data Mining metode C4.5. Penelitian selanjutnya dilakukan untuk prediksi desa wisata yang berpotensi menggunakan Data Mining C4.5.	Penelitian tersebut menggunakan objek wisatawan di kota Batam. Penelitian selanjutnya menggunakan objek seluruh desa wisata jawa tengah tahun 2019.	Data penelitian yang digali menggunakan data mining masih berformat excel	Penelitian ini diharapkan menggunakan aplikasi berbasis knowledge manajemen untuk mempermudah dalam mengelola dataset.	Penelitian tersebut menghasilkan pembentukan model wisatawan yang dibutuhkan agar jumlah kunjungan wisatawan ke kota batam semakin meningkat menggunakan algoritma C4.5.
2	Data Mining Prediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam Dengan Menggunakan	Penelitian tersebut dilakukan untuk memprediksi besarnya penggunaan listrik di Kota Batam dengan menggunakan metode	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut hanya mencakup penggunaan listrik rumah tangga.	Penelitian ini menggunakan <i>desicon tree classifier</i> algoritma C4.5	Penelitian tersebut memberikan kesimpulan bahwa penggunaan Algoritma C4.5 mampu digunakan sebagai metode

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
	Algoritma C4.5	Algoritma C4.5. Penelitian selanjutnya dilakukan untuk mengklasifikasi besarnya penjualan energi listrik di Kota Purwokerto dengan menggunakan metode Algoritma C4.5				klasifikasi untuk memprediksi konsumsi listrik.
3	Klasifikasi Penyakit Jantung Dengan Menggunakan Algoritma C4.5	Penelitian tersebut dan penelitian selanjutnya memiliki persamaan yaitu keduanya menggunakan algoritma C4.5	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut tidak menyantumkan hasil perhitungan <i>entropy</i> dan <i>gain</i>	Penelitian tersebut dapat melakukan penambahan atribut untuk data yang lebih baik lagi.	Penelitian tersebut mendapatkan akurasi 79% sehingga diharapkan bisa menjadi sumber informasi untuk penelitian dengan menggunakan algoritma C4.5.
4	Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah	Penelitian tersebut dan penelitian selanjutnya memiliki persamaan yaitu menggunakan penerapan algoritma C4.5 dengan pemrograman <i>python</i>	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut tidak menyantumkan klasifikasi masing-masing atribut yang digunakan.	Penelitian tersebut perlu membandingkan dengan metode yang berbeda untuk hasil yang lebih baik lagi.	Hasil penelitian mendapatkan tingkat akurasi sebesar 73% dengan menerapkan algoritma C4.5 menggunakan bahasa pemrograman <i>python</i> .

No	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
	Menggunakan <i>Python</i>					
5	Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi <i>Decision Tree</i> dan <i>Support Vector Machine</i>	Penelitian tersebut dan penelitian selanjutnya memiliki kesamaan yaitu keduanya menggunakan metode <i>decision tree</i> algoritma C4.5	Penelitian tersebut melakukan perbandingan metode Algoritma <i>Support Vector Machine</i>	Penelitian tersebut tidak menyantumkan penelitian sebelumnya.	Penelitian tersebut dapat menambahkan atribut untuk menghasilkan hasil pengujian yang lebih baik lagi.	Penelitian tersebut menyatakan bahwa penggunaan algoritma C4.5 menjadi algoritma yang tepat dalam menyelesaikan kasus keterlambatan pembayaran listrik.
6	Implementasi Algoritma C4.5 Untuk Klasifikasi Produk Laris Sepeda Motor Honda Pada Cv Cendana Motor Cepiring	Penelitian tersebut dan penelitian selanjutnya memiliki kesamaan yaitu melakukan pengujian menggunakan pemrograman <i>python</i> untuk mendapatkan hasil prediksi.	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut tidak menyantumkan hasil perhitungan <i>entropy</i> dan <i>gain</i>	Penelitian tersebut dapat dilakukan analisa perbandingan dengan metode lain terhadap prediksi kebutuhan yang ada.	Penelitian tersebut menghasilkan klasifikasi dengan akurasi 99%, menunjukkan bahwa algoritma C4.5 paling cocok untuk mengukur prakiraan penjualan sepeda motor Honda terlaris.
7	Implementasi Metode <i>Decision Tree</i> C4.5 Dalam Pemberian Subsidi Listrik Kepada Masyarakat	Penelitian tersebut memiliki kesamaan dengan peneliti selanjutnya yaitu keduanya menggunakan proses data mining dengan	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut perlu dilakukan pengujian kembali dengan data primer untuk menunjukkan hasil yang nyata.	Penelitian tersebut dapat melakukan perkembangan lebih lanjut untuk data yang lebih besar.	Penggunaan algoritma <i>decision tree</i> pada penelitian tersebut menghasilkan akurasi klasifikasi masyarakat yang berhak dan tidak

No	Judul	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
		metode <i>decision tree</i> algoritma C4.5				berhak menerima subsidi listrik sebesar 1 atau sebesar 100%
8	<i>Application of Classification Method C4.5 on Selection of Exemplary Teachers</i>	Penelitian tersebut dilakukan untuk menganalisis menentukan guru teladan menggunakan algoritma C4.5 dengan menggunakan pohon keputusan. Penelitian selanjutnya dilakukan untuk menentukan wisata yang paling unggul dengan menggunakan pohon keputusan.	Penelitian tersebut dan penelitian selanjutnya memiliki perbedaan mengenai objek penelitian.	Penelitian tersebut hanya menunjukkan empat dari sepuluh yang memiliki korelasi yang signifikan.	Penelitian terkait menentukan guru teladan dapat dilakukan dengan membandingkan metode lainnya.	Penelitian tersebut menghasilkan spesifikasi model guru dalam kategori sangat baik dengan menggunakan model pohon keputusan C4.5 berdasarkan aturan yang dihasilkan.[ 16]
9	<i>C4.5 Classification Rules Model for Determining Students Level of Understanding of the Subject</i>	Penelitian tersebut dan penelitian selanjutnya memiliki persamaan yaitu keduanya menggunakan model aturan klasifikasi C4.5.	Penelitian tersebut untuk mengklasifikasi tingkat pemahaman mahasiswa. Penelitian selanjutnya dilakukan untuk mengklasifikasi keunggulan desa wisata.	Penelitian tersebut tidak menyantumkan penelitian sebelumnya.	Penelitian tersebut dapat dikembangkan lebih lanjut dengan menggunakan metode lain untuk keakuratan hasil penelitian.	Penelitian tersebut menyimpulkan bahwa penggunaan algoritma C4.5 tepat diterapkan untuk mengetahui tingkat pemahaman siswa dengan tingkat akurasi

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summarize
						87,50%. [17]
10	<i>Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method.</i>	Penelitian tersebut dan penelitian selanjutnya memiliki persamaan yaitu menggunakan implemntasi data mining	Penelitian tersebut menggunakan metode k-means clustering. Penelitian selanjutnya menggunakan metode algoritma C4.5	Penelitian tersebut diharapkan dapat memberikan kriteria pembobotan agar data yang dihasilkan lebih akurat.	Penelitian yang sama dapat dilakukan menggunakan metode yang berbeda untuk dijadikan pembanding.	Penelitian tersebut menghasilkan penilaian berdasarkan indeks impor beras dengan dua negara kluster impor tinggi dan menengah sehingga metode tersebut mampu mengetahui jumlah beras yang diimpor oleh negara asal unggulan. [18]

## 2.2. Desa Wisata

Desa Wisata adalah sebuah konsep pengembangan pariwisata yang berfokus pada pengembangan potensi dan kearifan lokal masyarakat desa. Desa Wisata memanfaatkan kekayaan alam, budaya, dan kearifan lokal masyarakat sebagai daya tarik utama untuk menarik wisatawan. Potensi Desa Wisata meliputi kekayaan alam seperti pantai, danau, pegunungan, hutan, dan flora dan fauna yang kaya, serta kearifan lokal masyarakat seperti budaya, tradisi, dan kuliner yang unik. [19] Selain itu, Desa Wisata juga memiliki potensi dalam pengembangan pariwisata yang berkelanjutan karena mengedepankan konservasi lingkungan dan pemberdayaan masyarakat setempat. Oleh karena itu, Desa Wisata mempunyai potensi besar sebagai salah satu destinasi wisata yang unik dan menarik untuk berkelanjutan dalam

pengembangan destinasi wisata yang menjadi sebuah “*trend*” dalam dasawarsa terakhir ini [19].

## 2.3. Data Mining

### 2.3.1. Pengertian Data Mining

Penambangan data, juga disebut sebagai data mining *atau Knowledge Discovery in Database* (KDD) yang mampu mengekstraksi informasi dari basis data. Pencarian informasi ini menemukan pola tren dalam data, yang kemudian dapat ditambang dan diubah menjadi informasi yang mudah dipahami [20]. Penambangan data atau data mining merupakan suatu metode pengolahan data dalam pencarian informasi yang sebelumnya tidak diketahui dari kumpulan data besar. Secara umum data mining juga disebut sebagai sebuah cara atau tehnik dalam menyusun data besar dalam melakukan pencarian sebuah informasi yang bermanfaat bagi pengguna. Adapun beberapa metode dari pengolahan data mining yaitu sebagai berikut[21]:

#### a. *Predictive Modeling*

Pengolahan data mining dengan cara melakukan *forecast* yang bertujuan untuk membuat model prediksi dari suatu nilai yang memiliki karakteristik tertentu.

#### b. *Association*

Teknik data mining yang dilakukan untuk mempelajari hubungan antar data.

#### c. *Clustering*

Teknik pengelompokan data ke dalam suatu kelompok tertentu. Contoh algoritmanya adalah *K-Means*, *K-Medoids*, *Self Organization Map (SOM)*, *Fuzzy C-Means*.

#### d. *Classification*

Teknik yang digunakan untuk mengklasifikasikan data. Perbedaan metode pengelompokkan (*classification*) dengan metode pengelompokan (*clustering*) yaitu terletak pada datanya, dimana pada

metode *clustering* tidak membutuhkan variabel dependennya, tetapi metode klasifikasi membutuhkan variabel dependennya.

Berdasarkan beberapa teknik dalam data mining, penelitian ini akan menggunakan teknik *classification* dengan *decision tree classifier* algoritma C4.5 yang dapat mengklasifikasikan data dengan menemukan pola-pola yang berbeda dari model pohon keputusan yang dihasilkan.

### 2.3.2. Tahapan Data Mining

- a. *Selection* atau pemilihan, mempunyai tujuan utama yaitu membuat data target dari data asli dengan memilih subset variabel atau sampel data.
- b. *Preprocessing* memiliki tujuan utama yaitu membersihkan data dengan melakukan berbagai operasi seperti pemodelan dan penghapusan *noise*, menentukan strategi yang tepat untuk menangani bidang data yang hilang, menghitung informasi data yang dibutuhkan.
- c. *Transformation* memiliki tugas yaitu memetakan seluruh rangkaian nilai atribut yang diberikan ke rangkaian nilai pengganti yang baru. Setiap nilai lama dapat diidentifikasi dengan salah satu nilai baru. Data *transformation* memiliki beberapa metode yaitu *smoothing*, *attribute/feature construction*, *aggregation*, *normalization* dan *discretization*.
- d. Data Mining, tahap ini berhubungan dengan penggalian pola yang unik dengan memilih metode atau data mining tertentu, (misalnya *summarization*, *classification*, *clustering*, *regression*, dan sebagainya), algoritma yang tepat, dan representasi yang sesuai dari hasil keluaran.
- e. *Evaluation* atau *Interpretation* digunakan untuk menafsirkan dan mengekstrak pengetahuan dari pola yang dihasilkan dengan memvisualisasikan pola. Interpretasi ini biasanya dilakukan dengan memvisualisasikan pola, model, atau data yang diberikan model tersebut.

## 2.4. Discretization

*Discretization*, juga dikenal sebagai diskritisasi, adalah proses mengubah variabel, model, atau fungsi kontinu ke dalam bentuk diskrit. Hal ini melibatkan pembuatan satu set interval yang berdekatan yang menjangkau variabel, model, atau rentang fungsi yang diinginkan. Nilai diskrit memiliki peran penting dalam penggalian data dan penemuan pengetahuan. Nilai-nilai tersebut adalah tentang interval angka yang lebih ringkas untuk diwakili dan ditentukan, lebih mudah digunakan dan dipahami karena lebih dekat dengan representasi tingkat pengetahuan daripada nilai kontinu. Meskipun tidak ada aturan khusus, biasanya terdapat empat hingga sepuluh interval yang dibentuk. Karena prosesnya mirip dengan membagi nilai data menjadi interval, teknik ini juga disebut sebagai proses *binning*. Proses *binning* memiliki 2 metode yaitu *equal-width and equal-frequency*. Pada *equal-width* rentang kontinu fitur dibagi rata menjadi interval yang memiliki lebar sama dan setiap interval mewakili satu bin. Dan *equal-frequency* membagi rentang nilai kontinu yang sama ditempatkan di dalam jumlah yang sama di setiap bin. Dalam pembahasan tentang skala pengukuran di statistika, teknik ini mengubah variabel numerik menjadi variabel kategorik, sehingga disebut juga sebagai proses kategorisasi. Dengan demikian, untuk setiap atribut dalam menentukan klasifikasi dapat menggunakan metode *binning* untuk data *smoothing* dengan aritas  $k = 4$ , seperti yang ditunjukkan pada Tabel berikut [22]:

Tabel 2.2 *Partition into Equal-Frequency*

Features	Cut points	Number of points
F1	1, 2, 3, 4	4
F2	5, 6, 7, 8,	4
F3	9, 10, 10, 11	4



## 2.5. Pohon Keputusan (*Decision Tree*)

*Decision tree* merupakan algoritma pendukung dalam mengambil keputusan yang melakukan partisi rekursif dari ruang *instance*, pohon keputusan tipikal terdiri dari *node internal*, *edge*, dan node daun [23]. Setiap *node internal* disebut node keputusan, yang mewakili pengujian sebuah atribut atau subset atribut, dan setiap sisi diberi label dengan nilai tertentu atau rentang nilai atribut *input*. Pengklasifikasi pohon keputusan mencapai akurasi yang serupa dan terkadang lebih baik dibandingkan dengan metode klasifikasi lainnya [24].

Struktur pohon keputusan dibangun menyerupai pohon asli yang terdiri dari beberapa bagian sebagai berikut [25]:

1. Simpul akar yaitu node yang berada di bagian atas *decision tree*.
2. *Node internal* yaitu node ini merupakan cabang yang membutuhkan satu masukan dan menghasilkan paling banyak dua keluaran.
3. Simpul daun yaitu simpul yang berada di ujung model. *Node* ini hanya memiliki satu *input*, tidak ada *output*.

## 2.7. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang sering digunakan untuk membentuk pohon keputusan (*Decision Tree*) untuk memprediksi model struktur yang hierarki [15]. Algoritma C4.5 juga merupakan salah satu metode yang biasa digunakan untuk klasifikasi data [26]. Algoritma C4.5 merupakan pengembangan dari ID3 yaitu *Iterative Dichotomiser 3*. ID3 dikembangkan oleh J. Ross Quinlan. Prosedur dari algoritma ID3 mengambil input berupa contoh pelatihan, label pelatihan dan atribut lainnya. Beberapa proyek pengembangan yang diselesaikan di C4.5 mencakup passing nilai yang hilang, penerusan data berkelanjutan, dan pemangkasan [12].

Algoritma ini dapat menyelesaikan masalah dengan membangun pohon keputusan secara sistematis dengan tahapan-tahapan berikut [27]:

1. Mempersiapkan data latih (*data training*), yaitu bisa dengan mengambil dari data histori yang sudah terjadi sebelumnya dan telah diklasifikasikan ke kelas-kelas tertentu.
2. Menghitung akar dari pohon, yaitu dengan menghitung *root* (akar) yang akan diambil dari atribut yang terpilih. Kemudian masing-masing *gain* dari atribut juga akan dihitung dan akan menghasilkan nilai *gain* yang paling tinggi. Sehingga nilai *gain* yang paling tinggi tersebut dijadikan akar pada pohon keputusan. Sebelum menghitung nilai *gain* dari masing-masing atribut yang dipilih, langkah yang dilakukan dahulu yaitu menghitung nilai *entropy*. Hitung nilai *entropy* menggunakan persamaan rumus berikut:

$$Entropy(S) = \sum_{i=1}^n P(x_i) \log_2 (P(x_i)) \quad (2.1)$$

Persamaan 2.1 digunakan sebagai perhitungan *entropy*, dimana  $S$  merupakan himpunan kasus,  $n$  adalah jumlah partisi  $S$ ,  $p(x_i)$  merupakan proporsi kemunculan kelas ke  $i$  dalam himpunan data  $S$ . Jika nilai *entropy* tinggi, maka himpunan data dianggap tidak teratur.

3. Menghitung nilai *gain*, yaitu saat membangun pohon keputusan, pemilihan atribut *root* didasarkan pada nilai validasi tertinggi dari atribut yang ada. Hitung keuntungan menggunakan rumus dalam persamaan di bawah ini.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.2)$$

Persamaan 2.2 digunakan sebagai perhitungan *gain*, dimana  $S_i$  merupakan subset dari himpunan data  $S$  yang nilainya sama dengan  $i$  pada pada atribut  $A$ , dan  $S_i$  adalah jumlah kasus dalam subset  $S_i$ . Informasi *gain* tertinggi menunjukkan bahwa atribut  $A$  mengandung informasi yang berguna untuk memprediksi kelas dalam himpunan data  $S$ .

4. Ulangi tahapan kedua dan tahapan ketiga hingga *record* terpartisi.

5. Proses partisi pohon keputusan akan berhenti ketika semua *record* pada node N memiliki kelas yang sama, tidak ada lagi atribut pada *record* yang dipartisi dan tidak ada *record* pada cabang yang kosong.

## 2.8. Python

*Python* adalah bahasa pemrograman komputer yang dirancang untuk memfasilitasi pengkodean dan skrip aplikasi dan proyek. Bahasa pemrograman ini termasuk dalam jenis bahasa pemrograman tingkat tinggi, artinya kode yang dibuat dengan *python* mudah dipahami oleh orang. *Python* adalah bahasa pemrograman dikenal sebagai pemrograman yang memiliki banyak keunggulan dalam mendukung pemrograman berorientasi objek dan dapat digunakan di beberapa platform sistem operasi seperti *PC*, *Macintosh*, *UNIX*. Bahasa pemrograman *python* ini memiliki efisiensi yang cukup tinggi namun cukup efektif pada pemrograman berorientasi objek yang lebih sederhana, serta pemrograman *python* ini dapat mendukung bahasa pemrograman yang lain. Beberapa kelebihan bahasa pemrograman *python* diantara yang lain [28]:

1. Pengembangan *software* yang cepat dan membutuhkan lebih sedikit pemrograman.
2. Mendukung banyak *platform*.
3. Memiliki sistem memori otomatis.
4. *Python* adalah pemrograman berorientasi objek (PBO).

## 2.9. Confusion Matrix

*Confusion matrix* adalah metode yang umum digunakan untuk melakukan perhitungan yang akurat untuk konsep-konsep dalam data mining. Evaluasi menurut metode matriks konfusi menghasilkan nilai *accuracy*, *precision* dan *recall* [29]. Akurasi klasifikasi dalam data mining adalah persentase akurasi *record* yang diklasifikasikan dengan benar dan diuji berdasarkan hasil klasifikasi. Ketepatan atau kepercayaan adalah proporsi kasus yang diprediksi

positif padahal data sebenarnya juga positif. *Recall* atau sensitivitas adalah proporsi kasus positif yang diprediksi dengan benar [11].

Untuk menggambarkan analisis kinerja pengklasifikasi C4.5, penelitian ini menerapkan metode *confusion matrix* 3x3. Pada Tabel 2.3 *confusion matrix* terdiri dari tiga kelas yaitu kelas A, B, dan C, sedangkan hasil prediksi adalah hasil identifikasi yang dinilai oleh pengklasifikasi C4.5 yang juga terdiri dari tiga kelas. Kasus-kasus tersebut dibagi menjadi sembilan nilai yaitu TA, FA1, FA2, FB1, TB, FB2, FC1, FC2, dan TC. Berikut *confusion matrix* 3x3 yang ditunjukkan pada Tabel 2.3 berikut [30]:

Tabel 2.3 *Confusion Matrix 3x3*

		ACTUAL		
		A	B	C
Predicted	A	TA	FA1	FA2
	B	FB1	TB	FB2
	C	FC1	FC2	TC

Oleh karena itu, berdasarkan metode *Confusion Matrix* yang ditunjukkan pada Tabel 2.3., pengukuran akurasi pada penelitian ini dilakukan dengan menggunakan formulasi berikut.

$$Accuracy = \frac{T}{T+FA1+FA2+FB1+FB2+FC1+FC2} \times 100\% \quad (2.3)$$

Dimana  $T = TA + TB + TC$ . Dengan TA adalah kelas A yang diklasifikasikan dengan benar, TB adalah kelas B yang diklasifikasikan dengan benar, TC adalah kelas C yang diklasifikasikan dengan benar, FA1 adalah kelas B yang diklasifikasikan ke dalam kelas A, FA2 adalah kelas C yang diklasifikasikan ke dalam kelas A, FB1 adalah kelas A yang diklasifikasikan ke dalam kelas B, FB2 adalah kelas C yang diklasifikasikan

ke dalam kelas B, FC1 adalah kelas A yang diklasifikasikan ke dalam kelas C, FC2 adalah kelas B yang diklasifikasikan ke dalam kelas C.

Berdasarkan Tabel 2.4 menunjukkan perubahan dari *extended confusion matrix* berukuran 3x3 menjadi ukuran 2x2, dengan kelas "A" didefinisikan sebagai kelas positif dan kelas "Not A" sebagai kelas negatif [31].

Tabel 2.4 *Confusion Matrix 2x2*[31]

Actual	PREDICTED		
		A	Not A
A		TP	FN
Not A		FP	TN

Keterangan:

TP : *True Positive* yaitu jumlah data yang memiliki nilai positif dan diprediksi benar sebagai positif.

TN : *True Negative* yaitu jumlah data yang memiliki nilai negatif dan diprediksi sebagai negatif.

FP : *False Positive* yaitu jumlah data yang memiliki nilai negatif tetapi diprediksi sebagai positif.

FN : *False Negative* jumlah data yang memiliki nilai positif tetapi diprediksi sebagai negatif.

Beberapa persyaratan yang ditetapkan untuk matriks penilaian adalah sebagai berikut [28][32]:

a. *Accuracy* adalah rasio jumlah prediksi yang benar dengan menggunakan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.4)$$

- b. *Precision* adalah proporsi kasus positif yang diprediksi dengan benar, yang dihitung menggunakan rumus berikut:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.5)$$

- c. *Recall* adalah proporsi kasus positif yang teridentifikasi dengan benar, yang dapat dihitung dengan menggunakan rumus berikut:

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.6)$$

- d. *F1-Score* adalah proporsi yang didapatkan dari hasil *precision* dan *recall* antara kategori hasil prediksi dengan kategori sebenarnya, yang dapat dihitung dengan menggunakan rumus berikut:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.7)$$