

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **1.1. Penelitian Terdahulu**

##### **2.1.1. Penelitian Terkait**

Metode Naive Bayes merupakan salah satu metode klasifikasi yang telah banyak diimplementasikan pada beberapa kasus, seperti pada penelitian yang dilakukan Putro et al. (2020) mengimplementasikan metode Naive Bayes untuk mengklasifikasikan pelanggan berpotensi atau tidak terhadap suatu produk sehingga dapat digunakan sebagai dasar pertimbangan dalam strategi pemasaran. Penelitian ini menggunakan sebanyak 75 data latih dan 25 data uji, sedangkan atribut atau kriteria yang digunakan sebanyak tiga atribut, yaitu jumlah pembelian, interval waktu, dan lokasi serta target keluaran terdiri dari dua kelas, yaitu pelanggan berpotensi dan tidak berpotensi. Pengujian akurasi sendiri menggunakan *confusion matrix* dan menghasilkan akurasi yang tinggi yaitu sebesar 92% [5]

Penelitian yang dilakukan Sumanto et al. (2021) juga melakukan penelitian terhadap penggunaan metode Naive Bayes. Pada penelitiannya, Naive Bayes digunakan untuk memprediksi kelayakan kredit rumah. Data penelitian yang digunakan merupakan data transaksi penjualan rumah yang diambil antara Maret 2019 hingga Februari 2020 milik PT. Pratama. Kriteria yang digunakan berjumlah enam diantaranya, karakter, jenjang pendidikan, pekerjaan, rumah, pendapatan, serta tanggungan serta target keluaran yaitu layak dan tidak layak. Dengan menggunakan 100 data latih dan 15 data uji, penelitian ini menghasilkan akurasi sebesar 87.50% [6]

Penelitian yang dilakukan Mardhiyah et al. (2020) pada penelitiannya menggunakan algoritma C4.5 untuk mengklasifikasikan pembayaran kartu kredit. Penelitian bertujuan untuk membantu mengatasi pembayaran kartu kredit nasabah yang macet yang merugikan pihak penyedia layanan. Algoritma C4.5 digunakan untuk memprediksi macet atau tidaknya pembayaran kredit oleh nasabah. Atribut

atau kriteria penentu yang digunakan yaitu jumlah kredit, status, usia, dan status pembayaran dari bulan pertama hingga ketiga. Dari hasil penelitian yang menggunakan 4199 data, algoritma C4.5 menghasilkan tingkat akurasi sebesar 70.93% [7].

Penelitian mengenai kredit macet juga dilakukan oleh Nirwana et al. (2022) yang juga menggunakan algoritma C4.5. Pada penelitiannya, Nirwana menggunakan jumlah atribut yang lebih banyak yaitu sebanyak sepuluh atribut. Atribut-atribut yang digunakan diantaranya adalah usia, jumlah tanggungan, pendapatan, total bulan, jumlah tanggungan, status pendapatan, serta status kelancaran pembayaran. Penelitian menggunakan dataset sebanyak 10.127 data yang dibagi menjadi 60% data latih dan 40% data uji. Dengan menggunakan *confusion matrix* untuk perhitungan akurasi, algoritma C4.5 menghasilkan akurasi yang tinggi yaitu sebesar 99,97% [8].

Penelitian terhadap implementasi metode Naive Bayes dan algoritma C4.5 juga dilakukan oleh Desyanita & Wibowo (2020) yang membandingkan kedua metode pada proses klasifikasi kelayakan kredit rumah. Metode klasifikasi dapat memberikan prediksi dalam penentuan calon nasabah yang akan mengajukan kredit sehingga dapat meminimalisir terjadinya kredit macet setelah pengajuan kredit disetujui. Desyanita & Wibowo (2020) menggunakan atribut prediktor sebanyak 15 atribut dan satu atribut sebagai target kelas yaitu diterima dan ditolak. Sedangkan perbandingan data latih dan data uji yang digunakan sebanyak 80% data latih dan 20% data uji dari total data sebanyak 2,861. Dari penelitian ini algoritma C4.5 menghasilkan akurasi yang lebih baik yaitu sebesar 59,54% dan Naive Bayes sebesar 36.36% [9].

### **2.1.2. Perbandingan dengan penelitian terdahulu**

Perbedaan penelitian ini dengan penelitian yang dilakukan Sumanto et al. (2021) memiliki perbedaan dalam penelitian terhadap penggunaan metode Naive Bayes. Pada penelitiannya, Naive Bayes digunakan untuk memprediksi kelayakan kredit rumah. dengan data transaksi penjualan rumah, Kriteria yang digunakan berjumlah enam diantaranya, karakter, jenjang pendidikan, pekerjaan, rumah, pendapatan, serta tanggungan serta target keluaran yaitu layak dan tidak layak, dan tidak menggunakan metode Algoritma C.45, Sedangkan penelitian ini memiliki persamaan yaitu menggunakan data mining dan penelitian berkaitan dengan kredit.

Perbedaan penelitian ini dengan penelitian yang dilakukan Mardhiyah et al. (2020) pada penelitiannya hanya menggunakan algoritma C4.5 untuk mengklasifikasikan pembayaran

kartu kredit. Penelitian bertujuan untuk membantu mengatasi pembayaran kartu kredit nasabah yang macet yang merugikan pihak penyedia layanan. Algoritma C4.5 digunakan untuk memprediksi macet atau tidaknya pembayaran kredit oleh nasabah. Atribut atau kriteria penentu yang digunakan yaitu jumlah kredit, status, usia, dan status pembayaran dari bulan pertama hingga ketiga, Sedangkan penelitian ini menggunakan Naïve Bayes dan Algoritma C4.5.

Perbedaan penelitian ini dengan penelitian yang dilakukan oleh Nirwana et al. (2022) yang juga menggunakan algoritma C4.5. Pada penelitiannya, Nirwana menggunakan jumlah atribut yang lebih banyak yaitu sebanyak sepuluh atribut. Atribut-atribut yang digunakan diantaranya adalah usia, jumlah tanggungan, pendapatan, total bulan, jumlah tanggungan, status pendapatan, serta status kelancaran pembayaran. Penelitian menggunakan dataset sebanyak 10.127 data yang dibagi menjadi 60% data latih dan 40% data uji. Dengan menggunakan *confusion matrix* untuk perhitungan akurasi, algoritma C4.5 menghasilkan akurasi yang tinggi yaitu sebesar 99,97%, sedangkan penelitian ini hanya menggunakan Sembilan atribut, diantaranya adalah jenis kelamin, jenis usaha, status tempat usaha, jangka waktu pengambilan, jumlah pinjaman, jaminan,

Perbedaan penelitian ini dengan penelitian yang dilakukan juga dilakukan oleh Desyanita & Wibowo (2020) yang membandingkan kedua metode pada proses klasifikasi kelayakan kredit rumah. Metode klasifikasi dapat memberikan prediksi dalam penentuan calon nasabah yang akan mengajukan kredit sehingga dapat meminimalisir terjadinya kredit macet setelah pengajuan kredit disetujui. Desyanita & Wibowo (2020) menggunakan atribut prediktor sebanyak 15 atribut dan satu atribut sebagai target kelas yaitu diterima dan ditolak. Sedangkan perbandingan data latih dan data uji yang digunakan sebanyak 80% data latih dan 20% data uji dari total data sebanyak 2,861. Dari penelitian ini algoritma C4.5 menghasilkan akurasi yang lebih baik yaitu sebesar 59,54% dan Naive Bayes sebesar 36.36%. Rincian beberapa penelitian sebelumnya dapat dilihat pada Tabel 2.1 berikut.

**Tabel 1.1.Rangkuman penelitian sebelumnya**

No	Peneliti dan Tahun	Judul	Metode	Keterangan
1.	Desyanita &	Pemodelan Sistem Prediksi Kelayakan	Menggunakan metode Naive Bayes dan	<ul style="list-style-type: none"> <li>Algoritma C4.5 menghasilkan akurasi yang lebih</li> </ul>

	Wibowo (2020)	Pengajuan Kredit	algoritma C4.5 untuk menentukan kelayakan kredit rumah	<p>baik yaitu sebesar 59,54%</p> <ul style="list-style-type: none"> <li>• Naive Bayes memiliki akurasi lebih rendah yaitu sebesar 36.36%.</li> </ul>
2.	Mardhiyah et al., (2020)	Klasifikasi untuk Memprediksi Pembayaran Kartu Kredit Macet	Menggunakan algoritma C4.5 untuk memprediksi macet atau tidaknya pembayaran kartu kredit nasabah	<ul style="list-style-type: none"> <li>• Menggunakan atribut <i>predictor</i> berupa jumlah kredit, status, usia, dan status pembayaran dari bulan pertama hingga ketiga</li> <li>• Menghasilkan akurasi sebesar 70.93%.</li> </ul>

**Tabel 2.1 (lanjutan)**

3.	Putro et al., (2020)	Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan	Menggunakan metode <i>Naive Bayes</i> untuk menentukan potensi pelanggan terhadap penawaran suatu produk	Penerapan metode yang digunakan menghasilkan akurasi yang cukup tinggi yaitu sebesar 92%
4.	Sumanto et al., (2021)	Analisis Kelayakan Kredit Rumah Menggunakan Metode Naive Bayes untuk	Menggunakan metode <i>Naive Bayes</i> untuk memprediksi kelayakan pengajuan kredit rumah	<ul style="list-style-type: none"> <li>• <i>Predictor</i> yang digunakan adalah karakter, jenjang pendidikan, pekerjaan, rumah,</li> </ul>

		Mengurangi Kredit Macet		pendapatan, serta tanggungan <ul style="list-style-type: none"> <li>• Menghasilkan akurasi sebesar 87.50%.</li> </ul>
5.	Nirwana et al. (2022)	Klasifikasi Permasalahan Kredit Macet Pada Bank Menggunakan Algoritma Decision Tree C4.5	Menggunakan algoritma C4.5 untuk menentukan status kelancaran pembayaran kredit pada bank	<ul style="list-style-type: none"> <li>• Dataset dibagi menjadi 60% data latih dan 40% data uji</li> <li>• Menghasilkan akurasi sebesar 99,97%.</li> </ul>

Berdasarkan beberapa penelitian sebelumnya, pada beberapa kasus, metode Naive Bayes maupun algoritma C4.5 memiliki akurasi yang tinggi namun pada beberapa kasus lain juga memiliki akurasi yang rendah. Pada penelitian ini, dilakukan klasifikasi menggunakan kedua metode dengan objek yang sama, yaitu KSPPS. Adil Berkah Sejahtera untuk melihat metode mana yang memiliki akurasi yang lebih tinggi.

## 1.2. Landasan Teori

### 2.2.1. Koperasi

Koperasi menurut Undang-Undang Republik Indonesia Nomor 25 Tahun 1992 Tentang Perkoperasian merupakan badan usaha yang beranggotakan orang-seorang atau badan hukum Koperasi dengan melandaskan kegiatannya berdasarkan prinsi koperasi sekaligus seagai gerakan ekonomi rakyat yang berdasar atas asas kekeluargaan. Koperasi bertujuan memajukan kesejahteraan anggota pada khususnya dan masyarakat pada umumnya serta ikut membangun tatanan perekonomian nasional dalam rangka mewujudkan masyarakat yang maju, adil, dan makmur berlandaskan Pancasila dan Undang-Undang Dasar 1945 [10].

Dalam menjalankan kegiatannya, koperasi dapat menjalankan beberapa lapangan usaha, diantaranya:

- a) Koperasi dapat menghimpun dana dan menyalurkannya melalui kegiatan usaha simpan pinjam dari dan untuk:
  - Anggota Koperasi yang bersangkutan;
  - Koperasi lain dan/atau anggotanya.

- b) Kegiatan usaha simpan pinjam dapat dilaksanakan sebagai salah satu atau satu-satunya kegiatan usaha Koperasi.
- c) Pelaksanaan kegiatan usaha simpan pinjam oleh Koperasi diatur lebih lanjut dengan Peraturan Pemerintah.

### 2.2.2. Koperasi Jasa Keuangan Syariah

Berdasarkan Peraturan Menteri Koperasi dan Usaha Kecil dan Menengah Republik Indonesia Nomor : 11/PER/M.KUKM/XII/2017 Tentang Pelaksanaan Kegiatan Usaha Simpan Pinjam dan Pembiayaan Syariah Oleh Koperasi, Koperasi Simpan Pinjam dan Pembiayaan Syariah yang selanjutnya disingkat KSPPS adalah Koperasi yang kegiatan usaha simpan, pinjam dan pembiayaan sesuai prinsip syariah, termasuk mengelola zakat, infak, sedekah, dan wakaf.

Perbedaan antara koperasi konvensional dengan koperasi berbasis syariah terletak pada akad dimana pada koperasi konvensional tidak ada akad yang disepakati oleh kedua belah pihak. Salah satu jenis akad yang digunakan pada pembiayaan koperasi berbasis syariah adalah akad murabahah dimana mekanisme pembayarannya dapat dilakukan dengan cara kredit atau cicilan. Akad murabahah juga bisa dikatakan sebagai akad jual beli barang dimana harga perolehan dan keuntungan akan disepakati oleh kedua belah pihak yaitu penjual dan pembeli, dan pengembalian dana pinjaman nasabah dapat dilakukan melalui cicilan atau kredit dengan waktu yang disepakati [11].

### 2.2.3. Naive Bayes

Naive Bayes merupakan salah satu teknik prediksi berbasis probabilistik sederhana yang berdasar pada konsep Teorema Bayes. Naive Bayes memprediksi keanggotaan suatu kelas dengan menghitung probabilitas kelas tersebut menggunakan perhitungan “Bayes” sederhana. Pada Naive Bayes diasumsikan fitur merupakan fitur *independent*, tidak terikat atau bergantung pada fitur lain secara kuat (*naive*). Ide dasar dari Teorema Bayes adalah hipotesis atau peristiwa dapat diperkirakan berdasarkan pada beberapa *evidence* atau bukti yang diamati. Naive Bayes mempelajari bukti yang ada dengan menghitung korelasi di antara variabel yang diinginkan dan semua variabel yang lain. Secara umum, persamaan Teorema Bayes dapat dilihat pada persamaan (2.1) berikut.

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (2.1)$$

$$Posterior = \frac{(Likelihood \times Prior)}{Evidence} \quad (2.2)$$

Pada persamaan (3.1)  $H$  merupakan hal yang menyebabkan suatu peristiwa terjadi (*cause*) dan  $E$  adalah akibatnya (*evidence*). Mengacu pada persamaan (3.1),  $P(H/E)$  disebut juga posterior atau menyatakan probabilitas munculnya  $H$  jika diketahui  $E$ . Sedangkan  $P(E/H)$  atau *likelihood* adalah probabilitas sebuah *evidence*  $E$  terjadi akan mempengaruhi hipotesis  $H$ . Pada persamaan (3.2) *prior* atau  $P(H)$  adalah probabilitas awal (priori) hipotesis  $H$  terjadi tanpa memandang *evidence* apapun. Sedangkan *evidence* atau  $P(E)$  adalah hipotesis awal  $E$  terjadi tanpa memandang hipotesis/*evidence* yang lain [12].

#### 2.2.4. Algoritma C4.5

Metode *Decision Tree* atau pohon keputusan merupakan salah satu metode klasifikasi *supervised learning* yang memiliki kinerja dalam bentuk pohon. Pohon dalam *decision tree* terdiri dari node-node yang merepresentasikan atribut yang digunakan serta *leaf* atau node terbawah yang mewakili label atau kelas kategori [13].

Algoritma C4.5 merupakan salah satu algoritma *decision tree* yang dapat memproses *rule* yang mudah diinterpretasi dan tercepat diantara algoritma lainnya [14]. Pada algoritma C4.5 ditambahkan *pruning* untuk memotong atau menghilangkan beberapa cabang yang tidak diperlukan [15]. Node pada *Decision Tree* sendiri terdiri dari tiga bagian, yaitu:

- a) *Root node* atau akar, merupakan node teratas pada sebuah *tree* dimana node ini tidak memiliki *input* tetapi memiliki satu atau lebih cabang. Node ini juga merupakan node yang paling berpengaruh pada *tree*.
- b) *Internal node*, merupakan node yang terbentuk setelah *root node* dimana node ini memiliki satu *input* yang berasal dari *root* dan juga memiliki satu atau lebih cabang.
- c) *Leaf node* atau daun, merupakan node terakhir atau terbawah pada suatu *tree* sehingga node ini tidak memiliki cabang. Node ini merepresentasikan label atau kelas dari *tree* [16].

Algoritma C4.5 merupakan salah satu metode klasifikasi yang menggunakan teknik *Decision Tree* dalam pengambilan keputusan. J. Ross Quinlan (1993) mengembangkan algoritma C4.5 dari pendahulunya yaitu ID3 (*Iterative Dichotomiser*). Pada ID3, induksi *Decision Tree* hanya bisa dilakukan pada data bertipe kategorikal, sedangkan tipe numerik tidak dapat dilakukan [17].

Ide dasar dari algoritma C4.5 adalah membangun pohon keputusan yang didasarkan pada nilai atribut yang memiliki pengaruh tertinggi atau memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Algoritma C4.5 memiliki 2 tahapan kerja, yaitu: membangun pohon keputusan dan membuat aturan-aturan (*rule model*) yang berbentuk *if-else* [18].

Setelah terbentuk pohon keputusan, tahapan kerja algoritma C4.5 dimulai dari akar paling atas, misal diberikan sejumlah data uji  $X$ , dimana label dari data  $X$  belum diketahui, maka pohon keputusan akan menelusuri dimulai dari node teratas hingga node-node di bawahnya, dan setiap nilai dari atribut sesuai data  $X$  diuji apakah sesuai dengan aturan pohon keputusan, kemudian pohon keputusan akan memprediksi kelas dari tupel  $X$  [19]. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu:

- 1) Menyiapkan *data training*. Data training biasanya terdiri dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
- 2) Menentukan akar dari pohon. akar akan diambil dari atribut yang terpilih dengan cara menghitung nilai *Gain* dari masing-masing atribut, nilai *Gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *Gain* dari atribut, hitung dahulu nilai *entropy* yaitu :

$$entropy(S) = \sum_{i=1}^n - \pi * \log_2 \pi \quad (2.3)$$

Keterangan:

$S$  : himpunan kasus.

$n$  : jumlah partisi  $S$ .

$\pi$  : proporsi dari  $S_i$  terhadap  $S$ .

- 3) Kemudian hitung nilai *Gain* dengan metode information gain:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.4)$$

Keterangan :

$S$  : Himpunan Kasus.

$A$  : Atribut.

$n$  : Jumlah Partisi Atribut.

$A | S_i$  : Jumlah Kasus pada partisi ke- $i$ .

$|S|$  : Jumlah Kasus dalam  $S$ .



#### 4) *Gain ratio*

Untuk menghitung *gain ratio* kita perlu ketahui suatu term baru yang disebut *split information*. *Split information* dihitung dengan formula sebagai berikut:

$$SplitInfo(S,A) = -\sum_{i=1}^n \frac{s_i}{s} \log_2 \frac{s_i}{s} \quad (2.5)$$

$S_1$  sampai  $S_c$  adalah  $c$  subset yang dihasilkan dari pemecahan  $S$  dengan menggunakan atribut  $A$  yang mempunyai banyak  $C$  nilai.

Selanjutnya *gain ratio* dihitung dengan cara :

$$Gainratio = \frac{Gain(S,A)}{Splitinformation(S,A)} \quad (2.6)$$

5) Ulangi langkah ke-2 hingga semua semua tupel terpartisi.

6) Proses partisi pohon keputusan akan berhenti saat:

- a. Semua tupel dalam node  $N$  mendapat kelas yang sama.
- b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
- c. Tidak ada tupel di dalam cabang yang kosong.

Secara umum untuk membangun pohon keputusan algoritma C4.5 adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Kelebihan utama Algoritma C4.5 dapat membuat pohon keputusan (*decision tree*) yang efisien menangani atribut tipe diskrit dan tipe diskrit- numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima. Kekurangannya Algoritma C4.5 hanya dapat digunakan untuk menangani sampel-sampel yang dapat disimpan secara keseluruhan dan pada waktu yang bersamaan di memori.

#### **2.2.5. *K-Fold Cross Validation***

*K-Fold Cross Validation* merupakan teknik validasi dengan membagi data menjadi sebanyak k-subset dengan ukuran yang sama, kemudian sebanyak k-1 buah subset digunakan sebagai data uji dan sisanya sebagai data latih. Sebagai ilustrasi terdapat data sebanyak k subset yaitu D1, D2, ..., Dk dengan ukuran yang sama. Data kemudian secara acak dibagi menjadi data latih dan data uji dengan proses pelatihan dan pengujian dilakukan sebanyak k-kali secara berulang-ulang dengan setiap kali perulangan satu buah subset digunakan sebagai data uji dan sisanya digunakan sebagai data latih. Ilustrasi pembagian data latih dan data uji pada tiap subset dapat dilihat pada Gambar 2.1[16]

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10

Training Data
  Testing Data

**Gambar 2.1 Ilustrasi K-Fold Cross Validation**

### 2.2.6. Confusion Matrix

*Confusion Matrix* merupakan suatu metode yang digunakan untuk menghitung performa suatu model. *Confusion Matrix* digambarkan dengan tabel yang menyatakan jumlah hasil klasifikasi yang benar dengan jumlah klasifikasi yang salah.

**Tabel 2.2 Tabel confusion matrix**

Keterangan :

TP (*True Positive*) : data prediksi sesuai dengan data hasil

TN (*True*

Classification	Predicted Class	
	TRUE	FALSE
TRUE	TP	FN
FALSE	FP	TN

*Negative*)

: data yang benar-benar

tidak masuk pada kelompok tertentu

FP (*False Positive*) : data yang seharusnya di kelompok lain, tetapi ada pada kelompok tertentu

FN (*False Negative*) : data yang seharusnya di kelompoknya tetapi ada di kelompok lain.

Dari tabel tersebut nilai performa model dapat menggunakan beberapa perhitungan berikut [20].

a) *Accuracy*, merupakan perhitungan performa model dalam mengklasifikasikan data yang benar, yang dihitung dengan membagi jumlah total data klasifikasi benar dengan total data.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.7)$$

b) *Precision*, merupakan perhitungan performa model dimana ketika hasil klasifikasi positif, seberapa sering prediksi itu benar. Perhitungan *precision* dapat dilihat pada persamaan berikut.

$$Precision = \frac{TP}{TP+FP} \quad (2.8)$$

c) *Recall (Sensitivity/True Positive Rate)*, merupakan perhitungan seberapa sering model memprediksi positif ketika kelas aktualnya positif. Perhitungan *recall* dapat dilakukan menggunakan persamaan berikut.

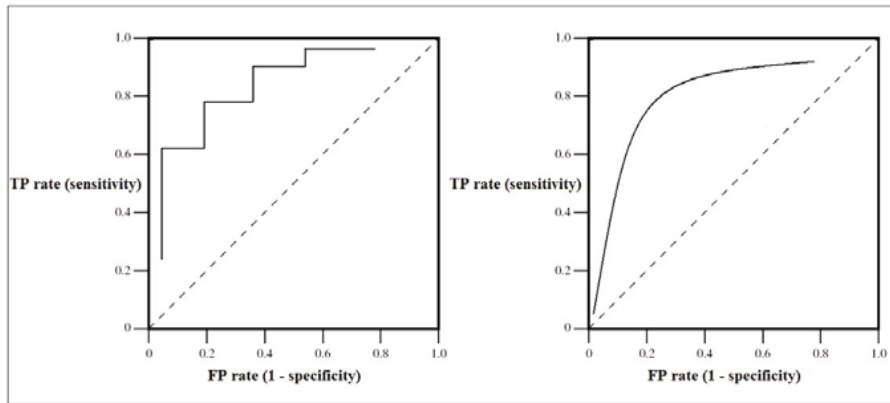
$$Recall = \frac{TP}{TP+FN} \quad (2.9)$$

d) *F1-Score*, merupakan nilai rata-rata harmonik dari *precision* dan *recall*. Formula *F1-Score* dapat dilihat pada persamaan berikut.

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (2.10)$$

### 2.2.7. Kurva ROC dan AUC

Kurva ROC (*Receiver Operating Characteristic*) merupakan metode yang banyak digunakan untuk melihat hasil suatu prediksi. Pada proses klasifikasi, kurva ROC merupakan teknik untuk memvisualisasikan, mengelompokkan, serta memilih metode klasifikasi berdasarkan performa masing-masing metode. Kurva ROC atau dikenal dengan grafik ROC, merupakan grafik dua dimensi dimana TP (*True Positive*) berada di sumbu Y, dan FP (*False Positive*) berada di sumbu X [21]. Gambaran kurva ROC dapat dilihat pada Gambar 2.2.



**Gambar 2.2 Kurva ROC diskrit dan kontinyu[21]**

Untuk menggunakan kurva ROC dalam menentukan performa suatu metode, dapat dilakukan dengan menghitung area di bawah kurva yang biasa disebut AUC atau *area under the ROC curve*. AUC memiliki rentang nilai antara 0 hingga 1. Model dikatakan memiliki performa yang baik bilai memiliki nilai AUC mendekati 1[22].

