

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terkait

Terdapat banyak algoritma dalam teknik klasifikasi dan prediksi dalam data mining. Penelitian terkait yang relevan dengan penelitian ini, telah dibuat berbagai perbandingan dari berbagai algoritma yang menguatkan penggunaan Algoritma *Naive Bayes Classifier*. Penelitian-penelitian itu diantaranya adalah penelitian yang dilakukan oleh Evaristus Didik Madyatmadja dan Mediana Aryuni [8], penelitian ini membahas tentang bagaimana merancang sebuah model dalam data mining yang dapat digunakan oleh pihak bank untuk menentukan penilaian kredit yang dapat mendukung dan meningkatkan kinerja pekerjaan bagi analis kredit. Sebagai analisis kredit sangat penting untuk dapat menentukan penilaian kredit tersebut “disetujui” atau “ditolak”. Untuk mendapatkan metode yang dapat memberikan kinerja terbaik, penelitian ini membandingkan algoritma *Naive Bayes Classifier* dengan Algoritma ID3 dengan menggunakan 1000 data yang mempunyai label kelas yaitu “disetujui” atau “ditolak”. Data dibagi menjadi *data training* dan *data testing* dengan masing-masing jumlah data untuk *data training* yaitu 600 data dan untuk *data testing* 400 data. Untuk mendapatkan hasil akurasi penelitian ini menggunakan pengujian *Confusion Matrix* terhadap *Naive Bayes Classifier* dan ID3. Hasil dari pengujian tersebut didapat nilai akurasi pada *Naive Bayes* yaitu sebesar 82%, dan pada algoritma ID3 didapat nilai akurasi sebesar 76%. Artinya bahwa algoritma *Naive Bayes Classifier* memiliki nilai akurasi yang lebih baik daripada algoritma ID3, sehingga model *Naive Bayes Classifier* dapat diterapkan untuk penilaian kredit yang dapat membantu meningkatkan kinerja pekerjaan analis kredit.

Penelitian oleh Alfa Saleh [9], yaitu penerapan metode *Naive Bayes Classifier* yang digunakan untuk memprediksi besarnya penggunaan listrik rumah tangga

sehingga akan mempermudah dalam pengaturan listrik. Penelitian ini menggunakan 60 data untuk dilakukan pengujian, dan hasil dari pengujian terhadap 60 data penggunaan listrik tersebut didapat hasil bahwa presentase untuk *Correctly Classified Instance* sebesar 78,3333% artinya bahwa dari 60 data penggunaan listrik terdapat 47 data yang berhasil untuk diklasifikasikan dengan benar, sedangkan persentase untuk *Incorrectly Classified Instance* sebesar 21,6667% yang artinya dari 60 data penggunaan listrik terdapat 13 data yang tidak berhasil untuk diklasifikasikan dengan benar dan pada penerapan metode *Naive Bayes Classifier* tersebut didapat nilai akurasi sebesar 78,3333% yang termasuk dalam kategori tingkat akurasi sedang, sehingga metode *Naive Bayes Classifier* dapat digunakan sebagai metode klasifikasi untuk memprediksi penggunaan listrik rumah tangga.

Penelitian yang dilakukan oleh Susanto, Evi Dewi Sri Mulyani, dan Irma Ratnasari Nurhasanah [10], yaitu memprediksi perilaku pola pembelian (untuk tahun 2016) dengan menggunakan data transaksi penjualan dari tahun sebelumnya (tahun 2014-2015). Prediksi perilaku pola pembelian dilakukan untuk mengetahui kemungkinan terhadap waktu transaksi penjualan yang dapat diperoleh, dan merupakan salah satu strategi untuk meningkatkan pendapatan penjualan. Jumlah data yang digunakan untuk sampel adalah 1768 data transaksi penjualan dari "Toko Toha Tasikmalaya" dari tahun 2014 sampai 2015. Penelitian ini menerapkan metode *Naive Bayes Classifier* untuk melakukan prediksi pola pembelian pelanggan yang dimana akan mengklasifikasikan kedalam 2 kelas yaitu kelas pagi dan kelas siang. Hasil dari pengujian algoritma *Naive Bayes Classifier* yaitu didapat nilai akurasi sebesar 97% dan hasil klasifikasi yang didapat berdasarkan data transaksi penjualan yaitu kelas pagi dan kelas siang diperoleh nilai untuk kelas pagi sebesar 0,477 dan kelas siang 0,523 artinya bahwa kemungkinan untuk pembeli melakukan transaksi pembelian adalah diwaktu siang.

Adapun penelitian yang dilakukan oleh Mujib Ridwan, Hadi Suyono, dan M.Sarosa [11], penelitian ini berfokus pada pengklasifikasian kelulusan

mahasiswa dimana dari hasil pengklasifikasian tersebut dijadikan sebagai evaluasi kinerja akademik mahasiswa. Jumlah data yang digunakan sebagai sampel adalah 100 data mahasiswa angkatan 2005-2009 dimana terdapat kelas dengan status kelulusan “tepat” dan “tidak tepat”. Pengujian dengan menggunakan *Confusion Matrix* dilakukan sebanyak 5 kali percobaan dengan jumlah data latih yang berbeda-beda. Pada percobaan 1 jumlah data yang digunakan sebanyak 20 data hasil yang didapat yaitu nilai akurasi 55%, *precision* 66,66%, *recall* 20%. Pada percobaan 2 dengan jumlah data sebanyak 40 data didapatkan hasil nilai akurasi sebesar 55%, *precesion* 55,55%, *recall* 50%, kemudian pada percobaan 3 jumlah data yang digunakan sebanyak 60 data, hasil yang didapat yaitu nilai akurasi 55%, *precision* 55,55%, *recall* 50%. Untuk percobaan yang ke 4 jumlah data yang digunakan sebanyak 80 data, didapat hasil nilai akurasi 65%, *precision* 71,42%, *recall* 50%. Dan untuk percobaan yang terakhir yaitu percobaan ke 5 dengan jumlah data uji sebanyak 100 data, didapat hasil akurasi 70%, *precesion* 83,33%, *recall* 50%. Dari hasil 5 percobaan yang dilakukan tersebut didapat nilai hasil perhitungan tertinggi yaitu nilai akurasi sebesar 70%, nilai *precision* 83%, dan nilai *recall* 50%. Sehingga didapat kesimpulan bahwa hasil dari pengujian tersebut didapat nilai akurasi dalam kategori sedang, maka metode *Naive Bayes Classifier* dapat digunakan untuk mengklasifikasikan kelulusan mahasiswa dan dapat dijadikan sebagai evaluasi untuk kinerja akademik mahasiswa.

Penelitian oleh Alvino [12], yaitu penelitian dengan menerapkan algoritma *Naive Bayes Classifier* berbasis PSO (Particle Swarm Optimization) sebagai *feature selection* atribut untuk menentukan nasabah deposito. Penelitian ini menggunakan algoritma PSO (Particle Swarm Optimization) sebagai cara penentuan atribut yang memiliki nilai optimal. Penerapan PSO sebagai *feature selection* ini berfungsi untuk pembobotan atribut, terbukti pada pemrosesan yang dilakukan oleh Alvino pada pemrosesan di Rapidminer dengan menggunakan 16 atribut yaitu *age*, *job*, *default*, *balance*, *loan*, *contact*, *day*, *month*, *campaign*, *pdays*, *previous*, *marital*, *education*, *haousing*, *duration*, dan *poutcome* didapatkan hasil bahwa 5 atribut yang mendapatkan nilai optimal, yaitu atribut *marital*, *education*,

*housing*, dan *poutcome* yang bernilai 1 dan atribut *duration* yang bernilai 0,856. Artinya bahwa ke- 11 atribut lainnya mendapatkan nilai 0 sehingga atribut tersebut termasuk atribut yang tidak optimal dan tidak dapat digunakan pada proses perhitungan di Rapidminer. Pada penelitian tersebut didapat kesimpulan untuk penerapan algoritma *Naive Bayes Classifier* berbasis PSO didapat peningkatan akurasi yaitu dari 82,19 % menjadi 89,70%.

Dari beberapa penelitian yang telah diuraikan dapat dirangkumkan pada tabel seperti berikut:

**Tabel 2.1 Penelitian Terkait**

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Evaristus Didik Madyatma dja dan Mediana Aryuni, 2014	Bagaimana merancang sebuah model dalam data mining yang dapat digunakan untuk menentukan penilaian kredit	Membandingkan Algoritma <i>Naive Bayes Classifier</i> dengan Algoritma ID3	Pengujian yang dilakukan dengan menggunakan <i>Confusion Matrix</i> didapat nilai akurasi pada <i>Naive Bayes Classifier</i> sebesar 82% dan nilai akurasi pada ID3 sebesar 76%, sehingga algoritma <i>Naive Bayes Classifier</i> lebih baik daripada algoritma ID3, maka dari itu model <i>Naive Bayes Classifier</i> dapat digunakan untuk penentuan penilaian kredit.

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
2.	Alfa Saleh, 2015	Sulitnya memprediksi besarnya penggunaan listrik rumah tangga	Algoritma <i>Naive Bayes Classifier</i>	Dengan metode <i>Naive Bayes Classifier</i> menggunakan data rumah tangga sebagai data training mampu mengklasifikasikan 47 data dari 60 data yang diujikan. Sehingga metode <i>Naive Bayes Classifier</i> berhasil untuk memprediksi besarnya penggunaan listrik rumah tangga, dan didapat hasil nilai akurasi sebesar 78,3333%.
3.	Susanto, Evi Dewi Sri Mulyani, dan Irma Ratnasari Nurhasanah, 2015	Memprediksi perilaku pola pembelian untuk ditahun berikutnya (tahun 2016)	Algoritma <i>Naive Bayes Classifier</i>	Hasil dari penerapan algoritma <i>Naive Bayes Classifier</i> untuk memprediksi perilaku pola pembelian didapat hasil klasifikasi dibagi menjadi 2 kelas yaitu kelas pagi dan kelas siang dengan hasil nilai akurasi sebesar 97%, dan hasil

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
				<p>klasifikasi nilai untuk kelas pagi sebesar 0,477, sedangkan kelas siang sebesar 0,523, maka didapat kesimpulan bahwa kemungkinan perilaku pola transaksi pembelian pada waktu siang.</p>
4.	<p>Mujib Ridwan, Hadi Suyono, dan M. Sarosa, 2013</p>	<p>Bagaimana mengklasifikasi kelulusan mahasiswa yang dapat dijadikan sebagai evaluasi kinerja akademik</p>	<p>Algoritma <i>Naive Bayes Classifier</i></p>	<p>Hasil dari pengujian dengan menggunakan <i>Caonfusion Matrix</i> yang dilakukan pada sampel data 2005-2009 yaitu menghasilkan nilai <i>precision</i> 83%, <i>recall</i> 50%, dan nilai akurasi sebesar 70%. Maka didapat kesimpulan bahwa metode <i>Naive Bayes Classifier</i> dapat digunakan untuk mengklasifikasikan kelulusan mahasiswa dan dapat dijadikan sebagai</p>

No	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
				evaluasi untuk kinerja akademik mahasiswa.
5	Alvino Dwi Rachman Prabowo	Bagaimana <i>Naive Bayes Classifier</i> dapat meningkatkan akurasi dengan PSO sebagai <i>feature selection</i> untuk pembobotan atribut.	<i>Naive Bayes Classifier</i> dengan <i>Naive Bayes Classifier</i> berbasis PSO	Didapat hasil bahwa algoritma PSO dapat digunakan sebagai algoritma <i>feature selection</i> yaitu dengan pembobotan atribut. Dari berawal pengolahan data dengan 16 atribut didapat nilai akurasi 82,19%, setelah dilakukan pembobotan atribut dengan algoritma PSO nilai akurasinya bertambah menjadi 89,70%.

Berdasarkan beberapa penelitian yang sudah diuraikan seperti diatas dapat disimpulkan bahwa metode *Naive Bayes Classifier* mampu dan akurat dalam menyelesaikan masalah prediksi dalam bentuk pengklasifikasian, terlihat dari penelitian Mujib Ridwan, Susanto, dan Alfa Saleh bahwa hasil uji validasi prediksi mendapatkan rata- rata nilai akurasi dalam kategori baik, dan dalam penelitian Alvino menggunakan PSO sebagai *selection feature* atribut dapat digunakan sebagai salah satu alternative untuk pemilihan atribut, sehingga pada

penelitian ini penerapan metode *Naive Bayes Classifier* diharapkan mampu memprediksi nasabah yang berpotensi untuk melakukan simpanan deposito dalam bentuk klasifikasi ke dalam 2 kategori yaitu apakah berpotensi untuk melakukan deposito “no” atau “yes” dan penelitian yang dilakukan Alvino dapat dijadikan sebagai acuan dalam pemrosesan atribut.

## **2.2 Landasan Teori**

Untuk mendukung penelitian yang akan dilakukan perlu adanya landasan teori yang sesuai dengan tujuan penelitian, diantaranya yaitu berasal dari buku, jurnal, paper ataupun prosiding.

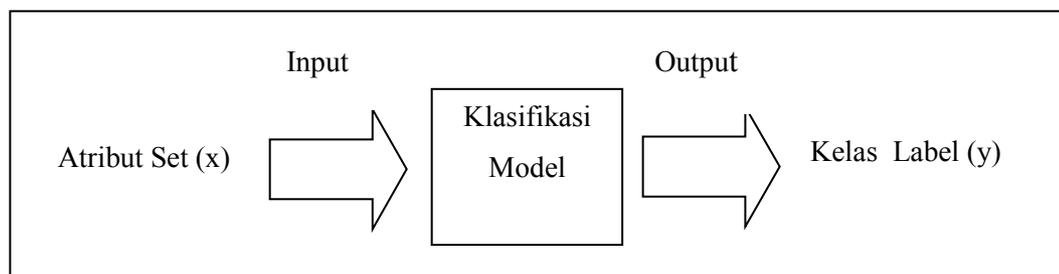
### **2.2.1 Data Mining**

Data Mining merupakan cabang ilmu baru di bidang teknologi yang banyak diterapkan di berbagai bidang ilmu (*artificial intelligence, database, statistik, pemodelan matematika, dan pengolahan citra*) [12]. Informasi dan *knowledge* tersebut dapat digunakan diberbagai bidang. Hal ini menjadikan *data mining* menjadi semakin meluas. *Data mining* banyak disebut *Knowledge Discovery in Database* (KDD) yang artinya proses pengumpulan, menemukan, menggali atau menambang dari dataset dalam jumlah yang besar [13]. Dari definisi data mining yang telah disebutkan, dapat ditarik kesimpulan bahwa data mining adalah disiplin ilmu yang mempelajari metode untuk mengstrak pengetahuan dan menemukan pola, aturan, atau model yang bersifat implisite dan sebelumnya tidak diketahui dari suatu data yang tidak memiliki arti. Berikut adalah karakteristik data mining:

1. Data mining berhubungan dengan sesuatu yang tersembunyi dan menemukan pola data yang sebelumnya tidak diketahui.
2. Data mining biasanya menggunakan data dengan jumlah besar yang digunakan untuk mendapatkan hasil lebih dipercaya.
3. Data mining memiliki kegunaan untuk membuat keputusan yang kritis, lebih utama untuk menentukan strategi.

### 2.2.2 Klasifikasi

Dalam data mining terdapat banyak teknik yang dapat dilakukan untuk mengolah data dalam jumlah yang besar menjadi sebuah informasi yang sangat berguna. Beberapa teknik tersebut antara lain: deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, dan asosiasi. Klasifikasi merupakan metode yang digunakan untuk menemukan suatu fungsi yang bertujuan untuk membedakan suatu kelas data. Tujuan dari klasifikasi adalah untuk memprediksi atau memperkirakan kelas dari suatu data baru yang belum memiliki label. Untuk mencapai tujuan dari klasifikasi tersebut perlu dibuat suatu model untuk membedakan suatu kelas data berdasarkan aturan tertentu [5]. Gambar 2.1 berikut ini mengilustrasikan proses klasifikasi.



**Gambar 2.1 Proses Klasifikasi**

Proses atau model klasifikasi dimulai dari adanya atribut set ( $x$ ) yang kemudian diinputkan dan diproses untuk mendapatkan model klasifikasi yang diharapkan, selanjutnya akan didapatkan output dari proses tersebut berupa kelas label ( $y$ ). Terdapat beberapa pengelompokkan pada metode klasifikasi yaitu, antara lain: Pohon Keputusan, Jaringan Saraf Tiruan, Naive Bayes, dan lainnya. Beberapa komponen yang mendasari proses klasifikasi yaitu:

1. Kelas (*Class*)

Diartikan sebagai variabel dependen dari suatu model, dan biasanya bertipe variabel kategorikal yang merepresentasikan “label” tertentu pada obyek setelah klasifikasi. Contohnya yang dimaksud kelas yaitu terdapat kelas kelulusan mahasiswa, kelas status kredit, dan lainnya.

2. Prediktor (*Predictor*)

Diartikan sebagai variabel independen dari suatu model, direpresentasikan oleh karakteristik (atribut) dari data yang akan diklasifikasikan dan berdasarkan dari klasifikasi yang sudah dibuat. Contohnya: musim, status perkawinan, umur, dan lainnya.

3. Pelatihan dataset (*Training Dataset*)

Kumpulan dari beberapa data yang berisi nilai- nilai dari kedua komponen yang sebelumnya dan biasanya untuk melatih model untuk mengenali kelas yang telah cocok atau sesuai, berdasarkan prediktor yang telah tersedia. Contohnya dataset pada database badai, database penelitian gempa, kelompok pelanggan supermarket, dan lainnya.

4. Dataset pengujian (*Testing Dataset*)

Berisikan data-data baru yang digunakan untuk diklasifikasikan (*classifier*) terhadap suatu model yang sudah dibangun sebelumnya, sehingga nilai akurasi yang didapat dari klasifikasi dapat dievaluasi.

Dibawah ini adalah beberapa model dari klasifikasi yang termasuk populer [13]:

1. *Decision Tree*
2. *Naive Bayes*
3. *Neural Network*
4. *Genetic Algorithm*
5. *Statistical Analysis*
6. *Rough Sets*
7. *K-Nearest Neighbor Classifier*
8. *Rule- based methods*
9. *Support Vector Machines*
10. *Memory Based Reasoning*

### **2.2.3 Naive Bayes Classifier**

*Naive Bayes Classifier* (NBC) adalah cara pengklasifikasian statistik yang dapat dipergunakan untuk melakukan prediksi dengan mengetahui probabilitas tertinggi keanggotaan pada suatu kelas. Kemampuan *Naive Bayes Classifier* didasarkan

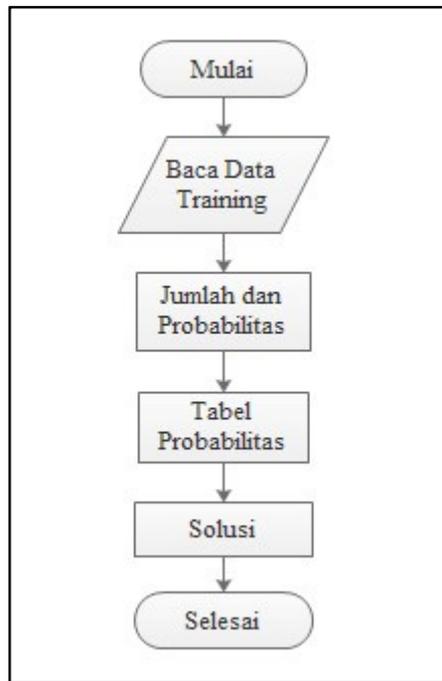
pada suatu teorema Bayes yaitu memiliki definisi pengklasifikasian dengan teknik probabilitas dan statistik untuk memprediksi kejadian dimasa depan berdasarkan kejadian yang sebelumnya [14]. Persamaan Teorema Bayes, berikut ini :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2.1)$$

**Tabel 2.2 Keterangan Persamaan Teorema Bayes**

Rumus	Keterangan
$X$	Data dengan kelas yang belum diketahui
$H$	Hipotesis data $X$ yang merupakan suatu kelas yang spesifik
$P(H X)$	Probabilitas $H$ berdasarkan data $X$
$P(X H)$	Probabilitas $X$ berdasarkan kondisi $H$
$P(H)$	Probabilitas $H$ (prior)
$P(X)$	Probabilitas $X$

Terdapat alur gambaran atau *flowchart* (gambar 2.2) untuk menjelaskan proses *Naive Bayes Classifier*.



**Gambar 2.2** *Flowchart Naive Bayes Classifier*

### 2.2.3.1 Peluang Bersyarat Atribut Kategorikal

Apabila peluang kelas yang bersyarat untuk atribut adalah nol, maka dari seluruh peluang bersyarat untuk kelas hilang. Hal ini akan menjadi masalah karena akan mengakibatkan probabilitas nol, yang menghapus semua probabilitas lain ketika mereka dikalikan dalam persamaan Bayes. Nilai nol ini akan menyebabkan kesulitan dalam analisis.

Masalah ini dapat diselesaikan dengan menggunakan rumus pendekatan estimasi probabilitas untuk frekuensi nol berikut ini [5]:

$$P(X|C_i) = \frac{n_c + n_{equiv}p}{n + n_{equiv}} \quad (2.3)$$

Dimana  $n$  mewakili jumlah keseluruhan dari *record* untuk *class* yang menjadi target  $n_c$  merupakan jumlah dari *nrecord* yang merupakan nilai atribut yang dicari, padahal estimasi prior dari probabilitas yang disetimasi  $p=l/k$  dimana  $k$  adalah jumlah *class* variabel target, dan  $n_{equiv}$  merupakan nilai konstan yang mewakili *equivalent sample size* (dalam penelitian ini,  $n_{equiv} =1$  untuk mengurangi intervensi).

### 2.2.3.2 Peluang Bersyarat Atribut Kontinu

Jika nilai dari  $X_j$  adalah kontinu, maka dapat diasumsikan bahwa nilai- nilai tersebut merupakan distribusi Gaussian dengan *mean*  $\mu$  dan *standar deviasi*  $\sigma$  yang didefinisikan dengan:

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.4)$$

sehingga,

$$P(X_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) \quad (2.5)$$

Maka,  $\mu_{C_i}$  dan  $\sigma_{C_i}$  perlu dikalkulasi karena merupakan *mean* dan standar deviasi dari nilai atribut  $X_j$  sebagai *training samples* dari *class*  $C_i$ . Berikut ini adalah keuntungan utama dari Naive Bayes [5]:

1. Kuat terhadap *noise* yang terisolasi dalam data.
2. Jika ada *missing values*, mengabaikan objek yang berhubungan selama proses probabilitas komputasi.
3. Handal untuk atribut yang tidak relevan.

### 2.2.4 Alat Ukur Evaluasi dan Validasi

Alat ukur evaluasi dan validasi untuk algoritma *Naive Bayes Classifier* yang diterapkan pada penelitian ini yaitu menggunakan alat ukur *Confusion Matrix* dan kurva *Receiver Operating Characteristics* (ROC).

### 2.2.4.1 Confusion Matrix

*Confusion Matrix* [5] berisikan informasi yang aktual dan prediksi pada suatu sistem klasifikasi dan memberikan penilaian *performance* klasifikasi yang berdasarkan pada objek dengan benar atau salah. Serta memberikan keputusan yang diperoleh dalam *training* dan *testing*.

Tabel 2.3 *Confusion Matrix*

		Nilai Sebenarnya	
		Benar	Salah
Nilai Prediksi	Benar	<i>(true positive - tp)</i>	<i>(false positive - fp)</i>
	Salah	<i>(false negative - fn)</i>	<i>(true negative - tn)</i>

Keterangan dari tabel *Confusion Matrix* diatas yaitu:

1. *True Positive* (tp) adalah proporsi positive (benar) dalam dataset yang dikategorikan positive (benar).
2. *False Negative* (fn) adalah proporsi negative (salah) dalam dataset yang dikategorikan negative (salah).
3. *False Positive* (fp) adalah proporsi negative (salah) dalam dataset yang dikategorikan positive (benar).
4. *True Negative* (tn) adalah proporsi positive (benar) dalam dataset yang dikategorikan negative (salah).

Dari keterangan tabel *Confusion Matrix* diatas, berikut ini merupakan persamaan pada *Confusion Matrix*:

1. Nilai Akurasi (*Acc*) merupakan proporsi dari jumlah prediksi yang benar, berikut adalah persamaannya :

$$Acc = \frac{tp+tn}{tp+tn+fp+fn} \quad (2.6)$$

2. *Sensitivity* atau *Recall* biasanya untuk membandingkan proporsi ( $tp$ ) terhadap proporsi yang *positive* (benar), berikut persamaannya:

$$Sensitivity = \frac{tp}{tp+fn} \quad (2.7)$$

3. *Specificity* biasanya untuk membandingkan proporsi ( $tn$ ) terhadap ( $tp$ ) yang *negative* (salah), berikut adalah persamaannya:

$$Specificity = \frac{tn}{tn+fp} \quad (2.8)$$

4. *Positive Predictive Value (PPV)* atau *Precision* merupakan suatu proporsi kasus yang hasil prediksi *positive* (benar), berikut persamaannya:

$$PPV = \frac{tp}{tp+fp} \quad (2.9)$$

5. *Negative Predictive Value (NPV)* adalah suatu proporsi kasus dengan hasil yang diprediksi *negative* (salah), berikut persamaannya:

$$NPV = \frac{tn}{tn+fn} \quad (2.10)$$

#### 2.2.4.2 Kurva Receiver Operating Characteristics (ROC)

*Receiver Operating Characteristic (ROC)* [5] banyak digunakan dalam menilai atau mengevaluasi hasil prediksi. Beberapa model pada klasifikasi (pohon keputusan) yang menghasilkan label kelas diskrit yaitu hanya pada kelas yang diprediksi objeknya. Pengklasifikasi lain (*Neural Network* dan *Naive Bayes Classifier*) menghasilkan output kelas kontinu dimana ambang batas yang berbeda mungkin diterapkan untuk memprediksi keanggotaan kelas. Dalam ROC semakin rendah titik kiri (0,0) maka strategi tidak akan pernah mengeluarkan klasifikasi atau kategori positif (tidak ada ‘positif palsu’ kesalahan tetapi juga tidak ada ‘positif sejati’). Titik kanan atas (1,1) merupakan strategi yang berlawanan dari

klasifikasi positif. Jadi (0,1) merupakan suatu klasifikasi yang sempurna (tidak ada FN dan tidak ada FP). Garis diagonal yang membagi ruang ROC, yaitu:

1. Poin diatas pada garis diagonal merupakan hasil dari klasifikasi yang baik.
2. Poin dibawah pada garis diagonal merupakan hasil dari klasifikasi yang buruk.

Tingkat akurasi dapat didiagnosa atau dikategorikan dalam beberapa tingkat, seperti yang terlihat pada tabel berikut ini [5]:

**Tabel 2.4 Tingkat Kategori Klasifikasi**

Akurasi	Kategori
0.90 – 1.00	Dikategorikan sebagai <i>Excellent Classification</i> .
0.80 – 0.90	Dikategorikan sebagai <i>Good Classification</i> .
0.70 – 0.80	Dikategorikan sebagai <i>Fair Classification</i> .
0.60 – 0.70	Dikategorikan sebagai <i>Poor Classification</i> .
0.50 – 0.60	Dikategorikan sebagai <i>Failure</i> .

### 2.2.5 Prediksi

Prediksi menurut Restiana Putri adalah melakukan suatu perhitungan sistematis yang digunakan untuk mengetahui suatu keadaan dimasa yang akan datang berdasarkan dari pengujian keadaan masa yang sebelumnya [15]. Sedangkan prediksi menurut Haldi Diana dan Cahyo Dwi Raharjo yaitu suatu teknik yang digunakan untuk menganalisa suatu faktor atau beberapa faktor untuk mengetahui peristiwa masa mendatang berdasarkan acuan peristiwa masa lalu [16]. Dari beberapa pendapat yang telah dikemukakan dapat disimpulkan bahwa prediksi adalah suatu proses yang dapat memperkirakan secara sistematis terhadap sesuatu kejadian yang memungkinkan terjadi dimasa yang akan datang berdasarkan pada informasi masa sebelumnya dan masa sekarang yang dimiliki. Hasil dari dilakukannya prediksi tersebut tidak harus memberikan secara pasti

kejadian yang akan terjadi, tetapi sedemikian mungkin untuk memerikan solusi atau jawaban terdekat yang mungkin hal tersebut akan terjadi.

### **2.2.6 Deposito**

Pengertian Deposito menurut Undang- Undang 10 tahun 1998 adalah salah satu tabungan berjangka yang model pengambilannya berdasarkan pada kesepakatan dari pihak bank dengan nasabah deposito [1]. Dari pengertian tersebut dapat disimpulkan bahwa waktu penarikan deposito sudah disepakati terlebih dahulu pada saat pembukaan deposito antara bank dengan nasabah. Menurut Dahlan Siamat yang dikutip oleh [17]. Deposito memiliki ciri- ciri, sebagai berikut :

1. Surat berharga yang diterbitkan oleh bank berdasarkan atas nama, sehingga tidak dapat diperjualbelikan.
2. Jangka waktu penarikannya telah ditentukan terlebih dahulu sesuai dengan yang diperjanjikan.
3. Bunga dibayar setiap bulan pada hari bayarnya atau sekaligus pada saat jatuh tempo.
4. Dapat dijadikan jaminan kredit.

Deposito sendiri juga memiliki fungsi yang sangat penting bagi bank yaitu sebagai salah satu alat untuk mengumpulkan dana dari masyarakat dan sangat berguna sekali untuk pemanfaatan perkreditan bagi bank. Maka fungsi deposito mempunyai peranan penting, hal ini disebabkan karena deposito merupakan salah satu sarana bagi bank untuk mengerahkan dana dari masyarakat. Dimana nantinya oleh bank akan dimanfaatkan kembali dan disalurkan dalam bentuk kredit kepada masyarakat ataupun berupa produk- produk bank yang lain. Dengan demikian berarti deposito merupakan juga suatu cara untuk mengatur kehidupan perekonomian.