

## BAB II

### TINJAUAN PUSTAKA

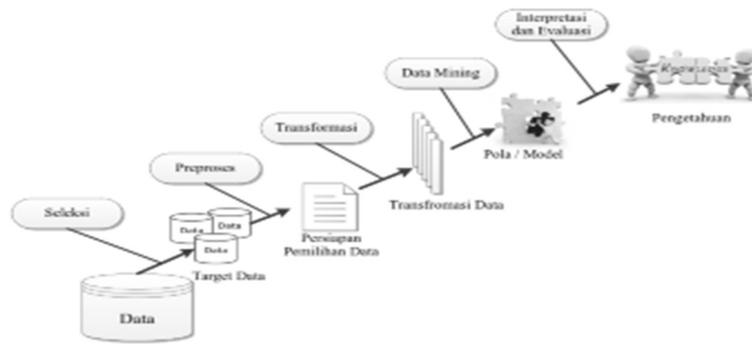
#### 2.1 Data Mining

Data Mining merupakan cabang ilmiah baru yang digunakan untuk mempermudah pengambil keputusan dalam menganalisis dan mengekstraksi data (Astriningtias & Mardhiyah, 2014). Sedangkan menurut Heri Susanto & Sudiyatno, data mining adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar (Susanto & Sudiyanto, 2014). Selain itu, pengertian data mining menurut Berry & Linoff (1997) adalah suatu pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan (Setiyani, Wahidin, Awaludin, & Purwani, 2020).

Data mining merupakan suatu proses otomatis terhadap data yang sudah ada (Kusrini, 2017).

1. Data yang akan diproses berupa data yang sangat besar.
2. Tujuan dari data mining adalah mendapatkan hubungan atau pola yang memungkinkan memberikan indikasi yang bermanfaat.

Istilah *data mining* dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya, kedua istilah tersebut memiliki konsep yang berbeda, akan tetapi memiliki keterkaitan satu sama lain, yang dimana tahapan dalam keseluruhan proses *Knowledge Discovery in Database* (KDD) adalah *data mining*. Secara garis besar proses KDD adalah sebagai berikut:



Gambar 2.1 Tahapan Data Mining  
 Sumber : Kusriani, "Algoritma Data Mining", Andi, Yogyakarta

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan dalam proses data mining tersimpan dalam suatu berkas, terpisah dari basis data operasional.

### 1. *Pre-processing/ Cleaning*

Sebelum proses *data mining* bisa dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Dan dilakukan juga proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

### 2. *Transformation*

*Coding* adalah proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai untuk proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

### 3. Data mining

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 4. *Interpretation/ Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. (Kusrini & E., 2009).

### 2.1.1 Tahap-Tahap Data Mining

Tahapan-tahapan menurut (Han & Kaufmann, 2006) yaitu :

#### 1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

#### 2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

### 3. Seleksi Data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

### 4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima *input* data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

### 5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

### 6. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasilnya berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

#### 7. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Karenanya presentasi dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

### **2.1.2 Pengelompokan Data Mining**

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Kusrini & E., 2009)

#### 1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden.

#### 2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

### 3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

### 4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

#### 5. Pengklusteran

Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku financial dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

#### 6. Asosiasi

Tugas asosiasi dalam Data Mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.
- b. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

## 2.2 Metode Klasifikasi

Metode data mining yang digunakan dalam penelitian ini adalah klasifikasi atau *classification*. Menurut Hamakonda (1982), klasifikasi adalah pengumpulan bahan pustaka baik berupa buku maupun bahan lainnya secara sistematis ke dalam satu kelas atau golongan tertentu berdasarkan ciri-ciri yang sama sehingga memudahkan pemakai menemukannya. Sedangkan menurut Sulistiyo Basuki (1991), klasifikasi merupakan suatu proses pengelompokan, artinya mengumpulkan benda atau entitas yang sama serta memisahkan benda atau entitas yang tidak sama.

Metode-metode klasifikasi merupakan bagian dari *Machine Learning* atau pembelajaran mesin. Pembelajaran mesin merupakan ilmu pengembangan algoritma dan model secara statistik yang digunakan oleh sistem komputer untuk menjalankan tugas tanpa instruksi secara langsung, dan hanya mengandalkan pola serta kesimpulan (what is machine learning, n.d.). Klasifikasi sendiri termasuk pada pendekatan *Supervised Learning* pada penggunaan pembelajaran mesin. *Supervised Learning* merupakan suatu pendekatan yang membuat pembelajaran mesin dilatih untuk mengenali pola antara data yang dimasukkan dan label yang akan menjadi hasilnya. Adapun 5 metode klasifikasi, yaitu : (Wibawa, Purnama, Akbar, & Dwiyanto, 2018)

### 1. *Naïve Bayes*

Merupakan sebuah metode klasifikasi yang berpaku pada teorema Bayes. Ciri utama *classifier* ini adalah asumsi yang sangat kuat atau naif akan independensi dari masing-masing kondisi/kejadian.

### 2.5 *Support Vector Machine*

*Support Vector Machine* (SVM) adalah metode klasifikasi yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space*.

### 2. *Decision Tree Support*

*Decision Tree* adalah algoritma yang paling banyak digunakan untuk masalah klasifikasi dimana algoritma ini terdiri dari beberapa simpul yaitu *tree's root*, *internal node* dan *leafs*.

### 3. *Neural Network*

*Neural Network* atau dalam bahasa indonesia disebut jaringan saraf tiruan merupakan pemrosesan suatu informasi yang terinspirasi oleh sistem saraf biologi, sama seperti otak yang memproses suatu informasi.

### 4. *Fuzzy*

Logika *Fuzzy* adalah logika yang kabur atau mengandung unsur ketidakpastian dan cenderung lebih praktis untuk digunakan karena sederhana, mudah dimengerti, fleksibel, serta lebih baik dan hemat. Logika *Fuzzy* mengenal nilai antara benar dan salah.

Selain itu, terdapat satu metode lain, yaitu K-Nearest Neighbor (KNN).

Dimana KNN merupakan metode yang digunakan untuk mengklasifikasikan K=1 untuk kelas *Single Nearest Neighbour* (Sri, 2019).

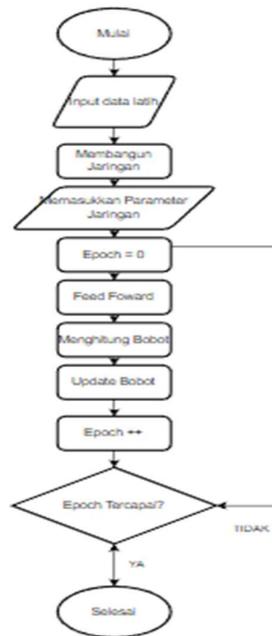
Metode yang akan diimplementasikan dan dibandingkan akurasi dalam mengklasifikasikan Indeks Pembangunan Manusia di kabupaten/kota di Pulau Jawa adalah *Artificial Neural Network* dan *K-Nearest Neighbor*. Sedangkan indikator yang digunakan dalam penelitian ini ada 4, yaitu Umur Harapan Hidup Saat Lahir, Rata-Rata Lama Sekolah, Harapan Lama Sekolah, Pengeluaran Per Kapita Disesuaikan.

### 2.3 *Artificial Neural Network*

*Artificial Neural Network* (ANN) merupakan sistem pengolahan data yang terinspirasi dari konfigurasi otak manusia yang pada dasarnya terbuat dari neuron buatan yang diidentifikasi sebagai konstituen pemrosesan yang saling berhubungan bertindak sama sekali untuk mencapai masalah tertentu (Khademi, Jamal, Deshpande, & Londe, 2016). Sedangkan menurut Haykin (2009), ANN adalah sebuah jaringan yang dirancang untuk menyerupai otak manusia yang bertujuan untuk melaksanakan suatu tugas tertentu.

Metode ANN telah digunakan dalam penelitian-penelitian sebelumnya dengan objek yang berbeda. Salah satu penelitian yang telah dilakukan berjudul Perbandingan Metode *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) pada Klasifikasi Gizi Balita. Penelitian ini dilakukan oleh Harifa Hananti dan Kartika Sari dengan tujuan untuk mendapatkan metode yang terbaik pada pengklasifikasian gizi buruk, khususnya gizi balita, dari kedua metode yang dilakukan. Hasil yang diperoleh menunjukkan bahwa metode ANN memiliki nilai akurasi lebih tinggi yaitu 94,82% dibandingkan metode SVM yang memiliki nilai akurasi sebesar 94,46% (Hananti & Sari, 2021). Penelitian lain yang telah dilakukan adalah penelitian oleh E. T. Lau, L. Sun, dan Q. Yang yang berjudul *Modelling*,

*Prediction and Classification of Student Academic Performance Using Artificial Neural Network.* Penelitian ini bertujuan untuk mengeksplor *modelling* dan analisis dari performa mahasiswa sarjana berdasarkan data yang berisi latar belakang sosial-ekonomi dan hasil ujian masuk yang diambil dari salah satu universitas di China menggunakan metode ANN (Lau, Sun, & Yang, 2019).



Gambar 2.2 *Flowchart* Algoritma *Artificial Neural Network*

Sumber : Implementasi Jaringan Syaraf Tiruan pada Kendali Lampu Sorot Mobil Adaptif Berbasis Python, Farras Timorremboko , Oki Teguh Karya, 2020 (Timorremboko & Karya, 2020)

Salah satu contoh model pelatihan ANN adalah *Feedforward Backpropagation*. Alur dari tahap *training* ANN tergambar pada *flowchart* diatas. Awal dari semua tahap adalah *input dataset training*. Lalu, jaringan dibangun dengan menentukan parameter seperti banyak node *hidden layer* dan *activation function*. Setelah parameter sudah ditentukan, *training* akan dimulai. *Training* dimulai pada perhitungan matematika *feedforward* dan dilanjutkan *backpropagation*, tahap ini dilakukan berulang-ulang sampai maksimum *Epoch*

atau batas maksimum pengulangan telah tercapai. Perhitungan 1 kali *Epoch* adalah ketika seluruh dataset melalui proses *forward* dan *backward*. Setelah maksimum *Epoch* tercapai, program *training* akan menampilkan *error*, nilai bobot dan bias akhir (Timorremboko & Karya, 2020).

Setelah mengetahui *flowchart* ANN, adapun rumus yang digunakan dalam perhitungan ANN sebagai berikut : (Yunirakhman, Prof. Dr. Adiwijaya, & Widi Astuti, 2019)

Pada hidden layer :

$$\begin{aligned} V1 &= P \times W1 + B1 \\ A1 &= \frac{1}{1 + e^{-V2}} \end{aligned} \quad (1)$$

Pada output layer :

$$\begin{aligned} V2 &= A1 \times W2 + B2 \\ A2 &= \frac{1}{1 + e^{-V2}} \\ E &= T - A2 \\ MSE &= \frac{\sum E^2}{N} \end{aligned} \quad (2)$$

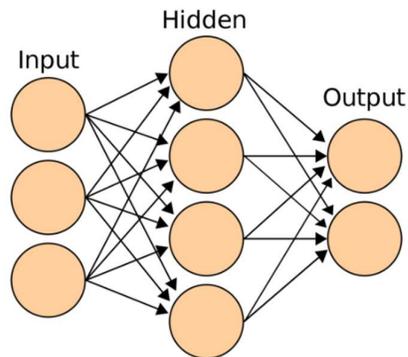
Perhitungan mundur :

$$\begin{aligned} D2 &= A2 \times (1 - A2) \times E \\ D1 &= A1 \times (1 - A1) \times (W2 \times D2) \\ dW1 &= lr \times D1 \times P \\ dW2 &= lr \times D2 \times P \\ dB1 &= lr \times D1 \\ dB2 &= lr \times D2 \\ W1 &= W1 + dW1 \\ W2 &= W2 + dW2 \\ B1 &= B1 + dB1 \\ B2 &= B2 + dB2 \end{aligned} \quad (3)$$

Keterangan :

V1 : nilai pada hidden layer; V2 : nilai pada output layer; P : matriks pada *input* layer; B1 : bias pada bobot; B2 : bias pada output; A1 : nilai pada hidden layer (perhitungan maju); A2 : nilai pada output layer (perhitungan maju); W1 : bobot pada hidden layer; W2 : bobot pada output layer; T : target ; MSE : mean square error; E : error; N : jumlah data; lr : learning rate; D1 : nilai pada hidden layer (perhitungan mundur); D2 : nilai pada output layer (perhitungan mundur); dW1 : delta W1; dW2 : delta W2; dB1 : delta B1; dan dB2 : delta B2.

Menurut (Haykin, 2009), terdapat 3 kelas dasar arsitektur ANN, yaitu *Single Layer Feedforward Network*, *Multi Layer Feedforward Network*, dan *Recurrent Network*. Pada penelitian ini digunakan metode ANN dengan *Multi Layer Feedforward Network*. Jika pada *Single Layer Feedforward Network* hanya terdapat 2 layer yaitu *input layer* dan *output layer*, maka pada *Multi Layer Feedforward Network* terdapat 1 layer tambahan yang disebut *hidden layer*. Dalam satu arsitektur *multi layer*, jumlah *hidden layer* yang digunakan boleh lebih dari satu sesuai dengan kasus ataupun masalah yang akan diselesaikan, seperti yang terlihat pada gambar dibawah ini.



Gambar 2.3 *Multi Layer Feedforward Network*

Sumber : Artificial Neural Network Big Data dan Pemanfaatannya, Aditya Yanuar R., 2018 (R., 2018)

Terdapat 3 (tiga) *layer* utama yaitu *input layer*, *hidden layer*, dan *output layer*.

*Input layer* merupakan lapisan yang membawa data masuk ke dalam sistem untuk kemudian diproses pada *layer* selanjutnya. Sedangkan *hidden layer* merupakan lapisan antara *input layer* dan *output layer*, dimana *artificial neuron* yang memiliki sekumpulan *input* pembobot ‘*weight*’ dan prosedur untuk menghasilkan *output neuron* melalui fungsi aktivasi. Jumlah hidden layer dapat ditentukan dengan rumus  $[n_i + n_o + 1]$ , dimana  $n_i$  sebagai nilai *input* dan  $n_o$  sebagai nilai *output* (Perez, Trier, & Reyes, 2000). Terakhir adalah *output layer* yang merupakan lapisan terakhir dari *neuron* yang menghasilkan *output* sistem (Rahanra, Kusri, & Luthfi, 2022).

#### 2.4 *K-Nearest Neighbor*

*K-Nearest Neighbor* (KNN) merupakan sebuah metode untuk mengklasifikasikan objek berdasarkan contoh data training yang terdekat dengan fitur objek tersebut (Imandoust & Bolandraftar, 2013). Menurut (Khamar, 2013), KNN merupakan sebuah kasus dimana *learning* algoritmanya berdasarkan jarak atau kesamaan fungsi untuk berbagai pengamatan seperti fungsi *Euclidean Distance*.

Beberapa penelitian yang telah menggunakan KNN sebagai metode klasifikasinya adalah penelitian untuk mengidentifikasi jenis kaca, yaitu dengan menghitung nilai akurasi dari nilai  $K=3$  sampai  $K=9$ . Hasilnya nilai  $K$  yang menghasilkan nilai performa paling baik adalah nilai  $K=3$ , dengan nilai akurasi mencapai 64% (Baharuddin, Hasanuddin, & Azis, 2019). Selain itu, terdapat penelitian lain yang menggunakan metode KNN untuk mengklasifikasikan aroma alkohol (Admojo & Ahsanawati, 2020). Dalam penelitian ini, nilai  $K$  yang digunakan adalah  $K=3$  dan  $K=4$ . Dengan hasil nilai akurasi  $K=4$  lebih tinggi yaitu 96,4% dibanding  $K=3$  yang memiliki nilai akurasi 95,8%.

KNN sendiri memiliki algoritma untuk mengolah data yang ada, berikut merupakan *flowchart* klasifikasi KNN.



Gambar 2.4 *Flowchart* Algoritma *K-Nearest Neighbor*

Sumber : Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor, Ikhsan Nuh Atthalla, Adithia Jovandy, dan Hanif Habibie, 2018 (Atthalla, Jovandy, & Habibie, 2018)

Perhitungan KNN dimulai dengan menentukan jumlah tetangga yang akan kita perhitungkan  $K$ . Selanjutnya, bagi data uji dan data latih sesuai presentase

yang diinginkan. Setelah itu, hitung jarak setiap tetangga terhadap objek tersebut kemudian urutkan hasilnya berdasarkan *range*, dari yang terkecil hingga ke yang terbesar. Kemudian, ambil (n) tetangga yang paling dekat, lalu lihat masing-masing dari tetangga tersebut, termasuk pada jenis apa (Athala, Jovandy, & Habibie, 2018).

Fungsi yang dipakai dalam penelitian ini untuk mencari jarak pada metode KNN yaitu fungsi *Euclidean Distance*. Rumus dari *Euclidean Distance* adalah sebagai berikut (Nikmatun & Waspada, 2019).

$$\text{Euc} = \sqrt{\sum_{I=1}^P (X_{2i} - X_{1i})^2} \quad (4)$$

Keterangan :

$X_1$  = data latih  
 $X_2$  = data uji  
 $i$  = variabel data  
 $p$  = dimensi data

## 2.6 *K-Fold Cross Validation*

*Cross validation* merupakan sebuah teknik validasi model untuk menilai hasil statistik analisis akan menyamaratakan kumpulan data independen (Tempola, Muhammad, & Khairan, 2018). Salah satu teknik atau metode dari *cross validation* adalah *k-fold cross validation*, dimana metode ini digunakan untuk membagi data latih dan data uji secara acak, dengan ukuran subset yang sama (Nainggolan, Prasvita, & Bukit, 2021). Proses metode ini yaitu model klasifikasi akan dilatih dengan data latih dan divalidasi menggunakan data validasi sebagai *k-fold* kali (Fuadah, et al., 2022).

Salah satu peneliti sebelumnya menggunakan *K-Fold Cross Validation* untuk memvalidasi data klasifikasi dengan metode KNN dan *Naïve Bayes* dengan nilai  $k\text{-fold} = 3$  (Tempola, Muhammad, & Khairan, 2018). Selain itu, pada penelitian lain, teknik ini juga digunakan untuk memvalidasi data klasifikasi dengan metode KNN dengan jumlah  $k\text{-fold} = 5$  (Azis, Purnawansyah, Fattah, & Putri, 2020).

## 2.7 Metrik Evaluasi untuk Klasifikasi

Evaluasi dilakukan agar mendapatkan metode klasifikasi dengan tingkat akurasi tertinggi. *Confusion Matrix* digunakan sebagai metrik evaluasi pada penelitian ini. *Confusion Matrix* merupakan sebuah metode yang bisa digunakan untuk mengukur akurasi data sehingga data dapat digunakan sebagai pendukung keputusan sebuah sistem (Rahmad, Suryanto, & Ramli, 2020).

Penelitian sebelumnya menggunakan *Confusion Matrix* untuk mengukur tingkat akurasi metode *Logistic Regression*, *Naïve Bayes*, *Decision Tree*, dan KNN (Ortega, et al., 2020). Selain itu, terdapat penelitian lain yang juga menggunakan *Confusion Matrix* sebagai pengukur tingkat akurasi dari teknologi anti-spam pada email (Rahmad, Suryanto, & Ramli, 2020). Berikut merupakan tabel tampilan *Confusion Matrix*.

Tabel 2.1 *Confusion Matrix*  
Sumber : Data Primer tahun 2022

<i>Predicted</i> <i>Actual</i>	<i>Positive (Fakta)</i>	<i>Negative (Palsu)</i>
<i>Positive (Fakta)</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative (Palsu)</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

*Confusion Matrix* terdiri dari 2 (dua) kelas yaitu kelas *actual* dan *predicted*. Dari 2 (dua) kelas tersebut menghasilkan 4 (empat) nilai, yaitu *True Positive* (TP), *False Negative* (FN), *False Positive* (FP), dan *True Negative* (TN). Dimana *True Positive* merupakan data fakta yang diprediksi benar sebagai fakta oleh sistem. *True Negative* merupakan data palsu yang diprediksi benar sebagai data palsu oleh sistem. *False Positive* merupakan data fakta yang diprediksi salah oleh sistem. Terakhir, *False Negative* merupakan data palsu yang diprediksi salah sebagai fakta oleh sistem (Kurniawan & Mustikasari, 2022).

Penelitian ini menganalisis akurasi dari metode KNN dan ANN. Nilai akurasi berupa perbandingan antara data yang berhasil terklasifikasikan benar dengan keseleruhan data yang telah diprediksi. Akurasi sendiri digunakan sebagai ukuran dari seberapa sering algoritma klasifikasi yang dibuat, menghasilkan prediksi yang benar (Kurniawan & Mustikasari, 2022). Adapun rumus perhitungan nilai akurasi yang digunakan pada penelitian ini adalah sebagai berikut.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

Keterangan :

TP : *True Positive*

TN : *True Negative*

FP : *False Postive*

FN : *False Negative*