

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 Kabupaten Serang**

Kabupaten Serang merupakan salah satu kabupaten yang berada di Provinsi Banten. Serang semula adalah bagian dari wilayah kerajaan atau kesultanan Banten yang berdiri pada abad ke XVI dengan pusat pemerintahannya berada di Serang, dan juga lokasi Serang berada di ujung bagian Barat. Luas wilayah secara administratif tercatat 1.467,35 Km persegi dengan 28 wilayah kecamatan dan juga 320 desa [8].

Kabupaten Serang memiliki perkembangan jumlah penduduk setiap tahunnya. Pada 2019 sendiri jumlah penduduk Kabupaten Serang adalah 1.508.397, sedangkan pada tahun 2020 jumlah penduduk Kabupaten Serang adalah 1.622.630 [9].

#### **2.2 Kriminalitas**

Kriminalitas adalah bentuk tindakan atau perbuatan melanggar hukum yang secara ekonomis dan psikologis dapat merugikan orang lain dan diri sendiri. Kriminalitas sendiri terjadi karena pelaku lemah dalam mengatur dirinya sendiri yang menyebabkan memiliki kecenderungan untuk melakukan tindakan diluar kebiasaan. Tindak kriminal sering terjadi juga dikarenakan adanya kesempatan untuk melakukan tindakan tersebut, lemahnya tingkat kewaspadaan seseorang terhadap lingkungan sekitar dan kurangnya kepedulian betapa pentingnya meningkatkan keamanan terhadap diri sendiri juga menjadi salah satu penyebab terjadinya tindak kriminal.

Banyak terjadi tindak kriminalitas di Negara-negara berkembang, terutama di Indonesia. Dapat dilihat pada Gambar 2.1, bahwa sepanjang 2020 tingkat kriminalitas terus menurun dari tahun-tahun sebelumnya, namun tetap masih ada tindak kriminal yang terjadi sehingga masyarakat tetap harus menaruh penuh perhatian terhadap pentingnya meningkatkan keamanan terhadap diri sendiri dan lingkungannya.

Tindak kriminal yang banyak terjadi di Indonesia ada beberapa jenis atau kategori yang masuk dalam kriminalitas seperti : [10]

- 1.Kejahatan terhadap nyawa: Pembunuhan

- 2.Kejahatan terhadap Fisik/Badan: Penganiayaan berat/ringan dan kekerasan dalam rumah tangga
- 3.Kejahatan terhadap Kesusilaan: Perkosaan dan pencabulan
- 4.Kejahatan terhadap kemerdekaan orang: penculikan dan memperkejakan anak di bawah umur
- 5.Kejahatan terhadap hak milik/barang dengan penggunaan kekerasan: Pencurian dengan kekerasan, pencurian dengan kekerasan menggunakan senjata api, pencurian dengan kekerasan menggunakan senjata tajam
- 6.Kejahatan terhadap hak milik/barang: pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja, penadahan
- 7.Kejahatan terkait narkoba: Narkoba dan Psikotropika
- 8.Kejahatan terkait penipuan, penggelapan dan korupsi: Penipuan/perbuatan curang, penggelapan, korupsi
- 9.Kejahatan terhadap ketertiban umum: Terhadap ketertiban umum



Gambar 2.1. Jumlah kejahatan dan resiko kejahatan

Sumber: [10]

## 2.3 Data Mining

*Data mining* adalah salah satu teknik untuk mengambil atau mencari informasi di dalam data yang besar. Untuk mendapatkan data yang tepat juga akurat diperlukan pengumpulan data dengan melakukan seleksi dan pencarian knowledge. Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dengan sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [11].

Berdasarkan beberapa definisi yang telah dijelaskan tentang *data mining* dapat diketahui bahwa penting sekali melakukan *data mining* untuk dapat mengidentifikasi karakteristik data seperti apa yang nantinya didapatkan, tentunya dengan beberapa tahapan-tahapan pada *data mining*. Berikut beberapa metode *data mining* :

### 1. *Classification*

Proses untuk mengelompokkan suatu kelas dari suatu objek dengan menemukan definisi kesamaan karakteristik dari kelas itu sendiri. Setiap keadaan kelompok diberi tindakan yang nantinya berfungsi untuk menjelaskan kelas tersebut.

### 2. *Association*

*Market basket analysis* (analisa keranjang pasar) atau *association* merupakan metode pemasaran yang memiliki tujuan menganalisa transaksi penjualan yang biasa produk dibeli bersamaan, misalnya pelanggan membeli roti biasanya juga membeli air mineral.

### 3. *Clustering*

*Clustering* atau metode segmentasi digunakan untuk mengelompokkan suatu kelas ke dalam segmen yang memiliki kemiripan atribut. Untuk menentukan atribut harus sesuai dengan kesamaan kelas tadi.

### 4. *Regression*

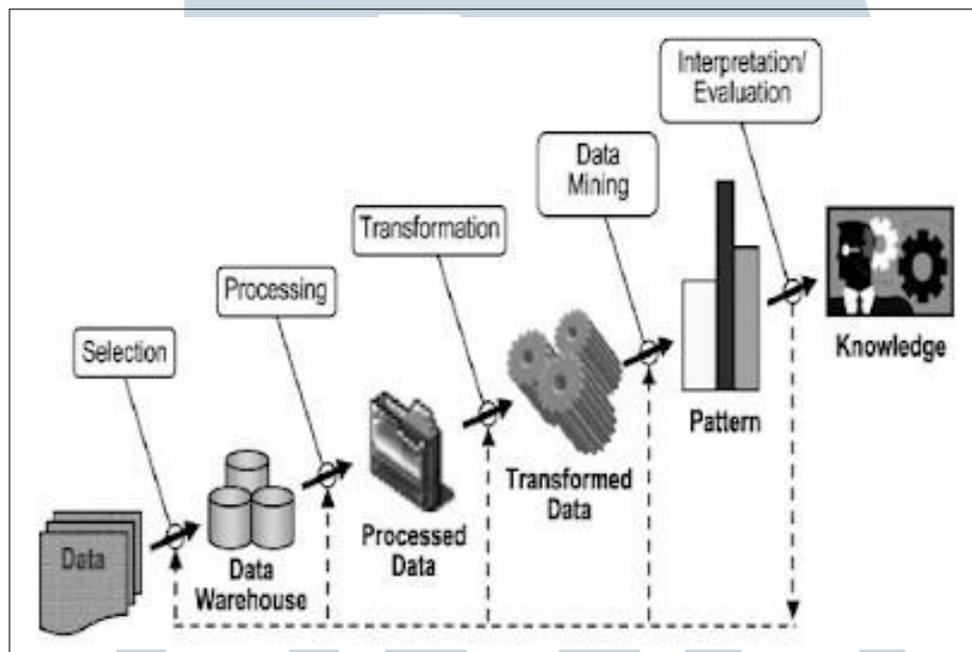
Metode yang penjelasannya hampir mirip dengan *classification*, yang membedakan metode regresi mencari pola yang tidak dijabarkan berdasarkan kelas dan bertujuan untuk mencari pola nilai numeric bukan kelas.

### 5. *Forecasting*

Metode yang digunakan untuk memprediksi nilai dapat dicapai selama satu

periode. Dengan menggunakan metode ini noise dan nilai data di periode sebelumnya menjadi dasar bahan prediksi.

Data mining merupakan bagian dari salah satu tahapan dalam keseluruhan proses KDD (Knowledge discovery in database). KDD adalah metode untuk menemukan dan mendapatkan pengetahuan dari database yang ada. Proses KDD diperlihatkan penjelasannya pada Gambar 2.2 :



Gambar 2.2. Proses knowledge discovery in database

Sumber: [12]

### 1. Data Selection

Melakukan penyeleksian untuk sekumpulan pada data operasional yang diperlukan sebelum masuk tahapan *Knowledge Discovery in Database* (KDD) dimulai untuk penggalian informasi. Hasil dari seleksi pada penggalian informasi kemudian dipakai dalam pemrosesan *data mining*.

### 2. Pre-processing / Cleaning

Tujuan *cleaning* yaitu untuk membuang data ganda atau duplikasi, memeriksa data yang tidak konsisten, memperbaiki kesalahan pada data, dan melakukan proses *enrichment*, yaitu memperluas data yang ada dengan informasi lain yang relevan.

### 3. Transformation

*Coding* yang merupakan proses transformasi pada data yang terpilih, sehingga data yang didapatkan sesuai untuk proses *data mining*. *Coding* dalam KDD merupakan proses kreatif yang sangat tergantung dengan jenis atau pola informasi yang nantinya dicari di basis data.

### 4. Data Mining

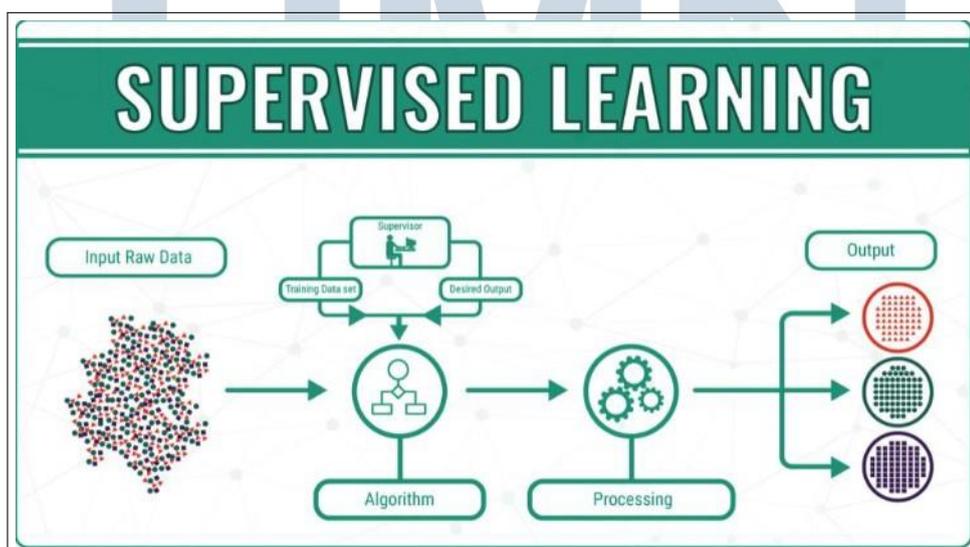
Setelah melakukan transformasi maka dilanjutkan dengan *data mining* yaitu pemrosesan melakukan pencarian pola atau juga informasi menarik yang didapatkan dari data terpilih memakai metode atau teknik tertentu.

### 5. Interpretation / Evaluation

Dari proses *data mining* didapatkan pola informasi kemudian ditampilkan pada bentuk yang bisa mudah dimengerti. Tahapan ini mencakup apakah pemeriksaan pola yang sudah didapatkan atau informasi bertentangan pada hipotesis atau fakta yang sudah ada sebelumnya.

## 2.4 Supervised Learning

*Supervised Learning* adalah teknik *machine learning* menggunakan dataset (data training) yang sudah berlabel dengan melakukan pembelajaran data pada mesin, sehingga dapat melakukan prediksi maupun klasifikasi[13]. Berikut gambaran tentang *supervised learning* yang terdapat pada gambar Gambar 2.3 :



Gambar 2.3. Supervised learning

Sumber: [14]

Tahap pertama yang dilakukan adalah memasukkan data, kemudian tahapan berikutnya disini *supervised learning* mulai memproses dataset (data training) yang kemudian menghasilkan *output*, selanjutnya dilakukan uji coba dengan algoritma yang kemudian melakukan *processing* hingga hasil terakhir mendapatkan *output*.

## 2.5 Klasifikasi

Klasifikasi merupakan proses dalam menemukan model atau fungsi yang menggambarkan juga membedakan konsep dan kelas data. Tujuan klasifikasi untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui atau mengklasifikasikan beberapa data kedalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Ada banyak teknik klasifikasi yang secara luas digunakan, diantaranya adalah *K-nearest neighbor*, *Bayesian classifiers*, *Neural network*, dan lain-lain.

Klasifikasi merupakan bagian dari *data mining*, dimana *data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan di dalam database. Data mining banyak menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat [15].

Proses klasifikasi data, langkah-langkah yang harus dilakukan pertama learning setelah itu klasifikasi. Pada saat proses learning data training dianalisa lalu direpresentasikan berupa rule klasifikasi. Proses klasifikasi, untuk memperkirakan akurasi dari rule klasifikasi perlu menggunakan data tes. Proses klasifikasi didasarkan pada empat komponen [16].

### 1. Kelas

Variabel dependen berupa kategori yang mempresentasikan label yang terdapat pada objek.

### 2. *Predictor*

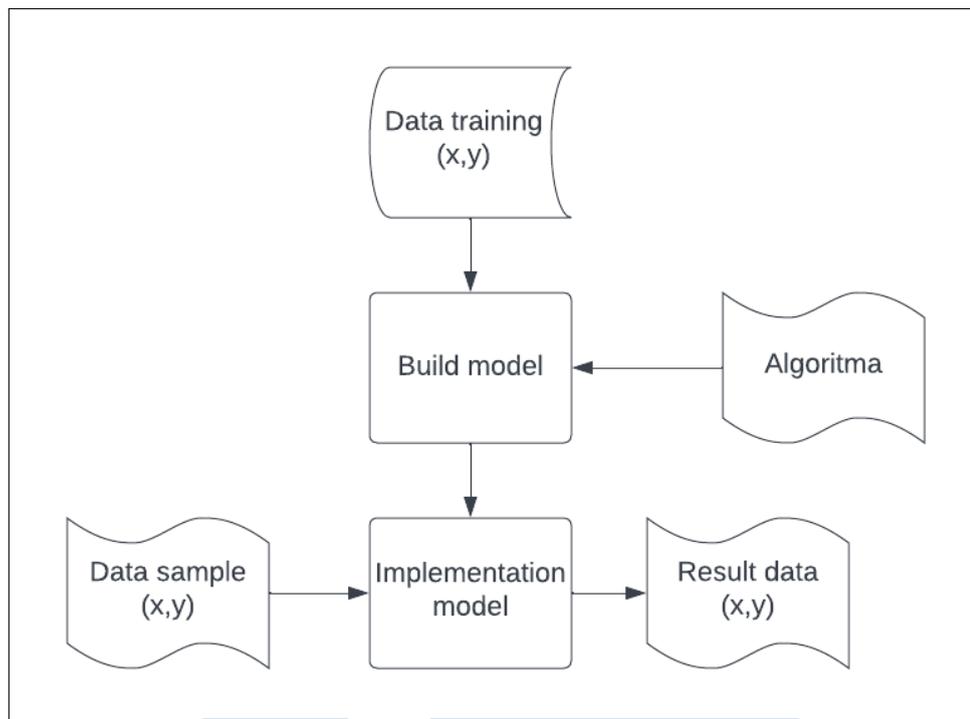
Variabel independen yang direpresentasikan oleh karakteristik data.

### 3. *Training dataset*

Satu set data yang mempunyai nilai dari kedua komponen yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

### 4. *Testing dataset*

Data baru yang diklasifikasikan oleh model data yang telah dibuat dan akurasi klasifikasi dievaluasi.



Gambar 2.4. Tahapan klasifikasi

Sumber: [17]

Pada Gambar 2.4 tahapan data latih yang dimiliki kemudian dimasukkan kedalam pembuatan model dengan algoritma, selanjutnya data uji masuk ke penerapan model dengan hasil algoritma dan juga data latih di pembangunan model yang sebelumnya, maka mendapatkan hasil data pada proses terakhir.

## 2.6 K-Nearest Neighbor

Algoritma *k-nearest neighbor* adalah algoritma yang biasa digunakan untuk implementasi klasifikasi suatu data berdasarkan *train data set* dengan mencari jarak K terdekat untuk data yang telah dievaluasi dan juga dengan K tetangga (neighbor) terdekat.

Cara kerja *k-nearest neighbor* pada saat *training*, algoritma melakukan penyimpanan vektor-vektor fitur juga klasifikasi dari data yang *training*. Ketika melakukan klasifikasi data, untuk data tes selanjutnya dihitung pada fitur-fitur yang sama. Semua vektor data *training* terhadap jarak vektor yang baru kemudian dihitung, k yang jaraknya paling dekat yang kemudian diambil. Klasifikasi titik yang baru dan yang paling banyak diprediksikan pada titik tersebut.

Untuk mendapatkan *k* yang baik didapatkan dari data yang optimal, nilai *k*

yang tinggi secara umum dapat mengurangi efek *noise* saat melakukan klasifikasi tetapi juga dapat membuat setiap batasan klasifikasi lebih kabur. Nilai  $k$  yang baik didapatkan dengan melakukan optimasi parameter, seperti memakai cross-validation. Berikut langkah-langkah penggunaan algoritma K-Nearest Neighbor:

1. Menentukan berapa jumlah dari tetangga yang terdekat

Menentukan jumlah dari  $k$  tetangga yang telah digunakan. Optimal jumlah  $k$  berada diantara angka 5 sampai 10, untuk jumlah  $k$  yang dipakai dapat ditentukan dengan jumlah data yang ada.

2. Menghitung jarak antar nilai data

Untuk menghitung jarak yang paling umum digunakan adalah *Euclidean distance* seperti pada Rumus 2.1

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Dimana:

$d(x, y)$  = jarak antar data

$x_i$  = nilai awal atau koordinat data ke- $i$

$y_i$  = nilai akhir atau koordinat data ke- $i$

$n$  = dimensi atau jumlah data

3. Menentukan kategori objek baru

Setelah melakukan penghitungan jarak antar data, selanjutnya mengambil  $k$  tetangga terdekat yang sesuai perhitungan jarak didapat sebelumnya. Pada  $k$  tetangga ini, melakukan perhitungan jumlah titik data untuk setiap kategori, kemudian pada kategori data baru dimasukkan untuk  $k$  yang tetangganya paling banyak.

## 2.7 Akurasi

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual [11]. Setelah melakukan penghitungan *euclidean*, mulai menghitung uji ketepatan nilai yang dipakai untuk menghitung tingkat akurasi pada dataset yang digunakan. Cara melakukan perhitungan akurasi dengan membagi jumlah data yang diklasifikasikan dengan total data *sample testing* yang sudah diuji, dapat dilihat dari persamaan berikut:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (2.2)$$

Keterangan Variabel:

TP: *true positif*, jumlah data positif yang terprediksi dengan benar

TN: *true negatif*, jumlah data negatif yang terprediksi benar

FP: *false positif*, jumlah data positif terprediksi salah

FN: *false negatif*, jumlah data negatif terprediksi salah

## 2.8 Presisi

Rasio item relevan yang dipilih terhadap semua item yang terpilih [7]. Melakukan perhitungan dengan menghitung nilai presisi membagi jumlah data benar dengan nilai positif atau *True positive*, selanjutnya dibagi dengan jumlah data benar bernilai positif atau *True positive* dan data salah yang bernilai positif atau *False negative*.

$$\text{Presisi} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (2.3)$$

## 2.9 Recall

Rasio dari item relevan dipilih terhadap total jumlah item relevan tersedia [7]. Melakukan perhitungan dengan cara membagi data benar bernilai positif (TP), dengan hasil penjumlahan dari data benar bernilai positif, dan data salah bernilai negatif (FN).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.4)$$

## 2.10 F-Measure

F-Measure atau F1-Score adalah harmonic mean antara nilai presisi dan *recall*, didapatkan dari pengolahan perhitungan dan pembagian hasil perkalian dari *precision* dan *recall* [7]. Hasil dari penjumlahan tadi kemudian dikalikan dua.

$$F \text{ — Measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (2.5)$$