

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 DATA MINING**

*Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* guna mengekstraksi serta mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

*Data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. *Data mining* didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar [9].

*Data mining* adalah suatu cara yang bertujuan dalam penemuan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam database atau sumber data lain yang dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai aturan proses [10].

*Data mining* membahas mengenai pencarian data serta informasi penting yang bermanfaat dan bersumber pada sekumpulan data. Informasi yang biasanya terkumpul ialah pola-pola tersembunyi yang ada pada data, hubungan antar elemen data, ataupun pembuatan model guna keperluan peramalan data [11].

Dapat disimpulkan bahwa *data mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau proses penentuan pola

dalam data yang berjumlah besar untuk menggali pengetahuan dan informasi yang belum diketahui sebelumnya serta digunakan untuk membuat suatu keputusan.

### **2.1.1 Tahapan *Data mining***

Secara sederhana *data mining* sering disebut sebagai proses pengolahan, penyaring serta menambang pengetahuan dari sejumlah data menjadi pengetahuan. Istilah *data mining* sering disebut *Knowledge Discovery in Database (KDD)* digunakan untuk menjelaskan proses mencari dan mengidentifikasi informasi tersembunyi dalam suatu basis data. Tahapan dalam proses KDD secara garis besar dapat dijelaskan sebagai berikut [12] :

#### 1. *Data Selection*

Seleksi data dari sejumlah data yang telah tersedia perlu dilakukan sebelum tahap penggalian informasi *Knowledge Discovery in Database* dimulai. Dengan adanya *data selection*, proses pengolahan akan menjadi lebih baik sesuai dengan tujuan yang akan dicapai. Data hasil seleksi yang digunakan untuk proses *data mining* disimpan pada suatu berkas, terpisah dari basis data operasional.

#### 2. *Preprocessing / Cleaning*

Pada tahap *preprocessing* dilakukan proses cleaning data antara lain membuang duplikasi data, memeriksa data yang tidak konsisten dan memperbaiki *tipografi*. Dilakukan juga proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan KDD.

### 3. *Transformation*

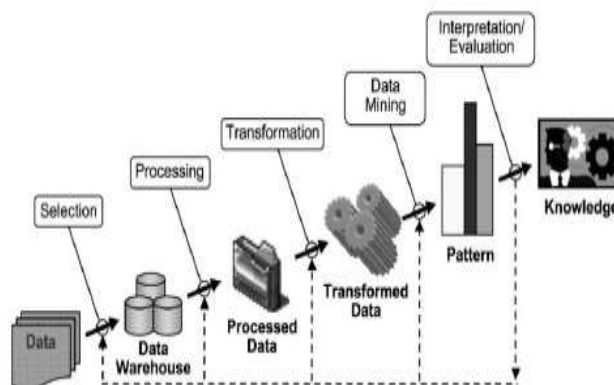
Proses tranformasi dilakukan pada data yang telah dipilih, dengan tujuan untuk menyesuaikan data yang diolah berdasarkan algoritma dan software yang digunakan di dalam pengolahan data.

### 4. *Data mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan 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.

### 5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan sama atupun bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



**Gambar 2. 1 Tahapan *Data Mining* [12]**

Berdasarkan dari gambar 2.1 dapat disimpulkan bahwa *data mining* merupakan salah satu tahapan dari proses *Knowledge Discovery in Database* (KDD). Sebelum data dapat diolah menjadi sebuah informasi yang bermanfaat, diperlukan adanya proses demi mencapai tujuan yang akan dicapai.

### **2.1.2 Pendekatan *Data mining***

Secara garis besar metode yang digunakan dalam teknik-teknik *data mining* dibedakan ke dalam dua pendekatan [13], yaitu :

#### **1. *Unsupervised Learning***

*Unsupervised learning* yaitu pendekatan tanpa adanya latihan atau data mentraining. Data hasil pengamatan yang diperoleh akan dikelompokkan atau dimasukkan dalam beberapa kelas sesuai dengan yang dikehendaki. Hal ini terjadi karena data yang dimiliki tidak memiliki label yang menandai. Teknik *data mining* yang termasuk ke dalam bagian *unsupervised learning* adalah *Clustering*, dan *Self Organization Map* (SOM).

#### **2. *Supervised Learning***

*Supervised learning* yaitu metode belajar dengan adanya latihan dan pelatih. Dalam *supervised learning* disyaratkan agar data analisis telah mengidentifikasi atribut tujuan. Dalam pendekatan ini, data yang memiliki output atau label selama proses training, digunakan untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi. Teknik *data mining* yang termasuk ke dalam bagian *Supervised learning* adalah *Clasification* (*Categorical*) dan *Regression* (*Numerical*), *Logistic Regression*, *Artificial Neural Networks* (ANN), *Support Vector Machine* (SVM) .

### 2.1.3 Metode *Data mining*

Secara umum, metode *data mining* dapat dibagi menjadi dua, yaitu deskriptif dan prediktif. Deskriptif berarti *data mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi [14]. Metode yang ada dalam *data mining* adalah sebagai berikut :

1. *Classification*

Klasifikasi merupakan proses untuk menemukan sekumpulan model yang dijelaskan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih. Sedangkan data uji digunakan untuk mengetahui tingkat akurasi dan model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai dari suatu objek data.

2. *Clustering*

Pengelompokan data yang tidak diketahui label kelasnya kedalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.

3. *Association*

Tujuan dari metode ini yaitu untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang terhubung kuat dengan yang lainnya.

#### 4. *Regression*

*Regression* mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diproduksi nilai yang kontinyu.

#### 5. *Forecasting*

Prediksi (*forecasting*) berfungsi untuk melakukan prediksi kejadian yang akan diproses berdasarkan data sejarah yang ada.

#### 6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit sebagai contoh adalah menemukan kelompok gen dengan tingkat ekspresi yang mirip.

#### 7. *Deviation Analysis*

Tujuan dari metode ini adalah untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut sebagai outlier detection. Sebagai contoh adalah apakah sudah terjadi penipuan terhadap pengguna kartu kredit dengan melihat catatan transaksi yang tersimpan dalam basis data perusahaan tersebut.

## **2.2 KLASIFIKASI**

Klasifikasi berguna sebagai pengumpulan data atau metode mengorganisasi data yang telah dikumpulkan, agregasi dan disagregasi serta untuk memberi makna pada suatu set data untuk maksud analisis termasuk untuk penyusunan indeks. Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*.

Klasifikasi adalah teknik yang dilakukan untuk memprediksi class atau properti dari setiap instance data. Model prediksi memungkinkan untuk memprediksi nilai-nilai variabel yang tidak diketahui berdasarkan nilai variabel lainnya. Klasifikasi memetakan data ke dalam kelompok-kelompok kelas yang telah ditetapkan sebelumnya. Klasifikasi disebut juga dengan supervised learning karena kelas data telah ditentukan sebelumnya.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya. Teknik klasifikasi bekerja dengan mengelompokkan data berdasarkan *data training* dan nilai atribut klasifikasi [15].

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan klasifikasi atau prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya [16].

Beberapa algoritma yang dapat digunakan dalam klasifikasi *data mining* adalah sebagai berikut :

1. *Neural Network*

*Neural Network* (Jaringan Saraf Tiruan) adalah prosesor tersebar paralel yang sangat besar dan memiliki kecenderungan untuk menyimpan

pengetahuan yang bersifat pengalaman dan membuatnya siap untuk digunakan. Keunggulan dari *Neural Network* yaitu memiliki nilai numerik yang memungkinkan untuk melakukan lebih dari satu tugas secara bersamaan, data yang digunakan disimpan di seluruh jaringan sehingga proses kerja jaringan tidak akan terhambat bila terjadi data hilang, mampu menoleransi kesalahan, menghasilkan *output* bahkan dengan data yang tidak memadai.

## 2. *Decision Tree*

*Decision tree* sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak di minati. Dalam *decision tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit . Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut *numeric*. Keunggulan *Decision Tree* yaitu *output* yang mudah dibaca dan ditafsirkan, berguna untuk membantu memecahkan masalah terkait keputusan, lebih sedikit pembersihan data yang diperlukan.

## 3. *Naive Bayes*

*Naive Bayes* merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Naive Bayes* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. Teknik *Naive Bayes* (NB) adalah salah satu bentuk sederhana dari Bayesian untuk klasifikasi.



Kelebihan dari teorema *bayesian* adalah klasifikasi *Naïve Bayes* mudah dan cepat diimplementasikan, algoritma *Naive Bayes* bekerja sangat baik dengan variabel input kategori, dibandingkan variabel numerik, *Naïve Bayes* membutuhkan lebih sedikit data pelatihan.

#### 4. *K-Nearest Neighbor*

Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. KNN memiliki beberapa keunggulan yaitu, pelatihan yang sangat cepat, sederhana dan mudah dipelajari, efektif jika data pelatihan besar.

#### 5. *Logistic Regression*

Regresi Logistik (*Logistic Regression*) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi. Salah satu keunggulan dari *Logistic Regression* yaitu algoritma yang fleksibel, mengambil segala jenis input serta mendukung beberapa tugas analitik yang berbeda.

### 2.3 **NAIVE BAYES**

*Naïve Bayes* merupakan salah satu penerapan teorema *Bayes*. *Naïve Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output,

probabilitas mengamati secara bersama adalah produk dari hipotesis pada sebuah sampel, biasa disebut dengan priori.

Algoritma *Naïve Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve Bayes* juga merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan asal inggris Thomas Bayes, yaitu dengan memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu sehingga dikenal sebagai Teorema Bayes [17] .

Metode *Naïve Bayes* hanya membutuhkan jumlah yang kecil dari data yang diperlukan untuk proses pengklasifikasian merupakan suatu keuntungan penggunaan algoritma *Naïve Bayes* [9].

*Naïve Bayes* merupakan algoritma yang memanfaatkan teori probabilitas, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya dengan memanfaatkan pengetahuan pakar. Metode ini dipilih karena mudah diterapkan bekerja secara independen yakni sebuah fitur di dalam sebuah data tidak berkaitan dengan ada atau tidaknya fitur yang lain dalam data yang sama [16].

*Naïve Bayes* mempunyai keunggulan yaitu cepat, sederhana, dan memiliki akurasi tinggi. Dibandingkan dengan *classifier* lain *Naïve Bayes* bekerja lebih baik dan memiliki tingkat akurasi yang baik dari Teorema Bayes. Nilai *Bayes* yang dipilih merupakan persentase tertinggi. Persamaan dari *Naïve Bayes* sebagai berikut [18] :

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

Dimana :

$X$  : Kriteria suatu kasus berdasarkan masukan.

$C_i$  : Kelas solusi pola ke- $i$ , dimana  $i$  adalah jumlah label kelas.

$P(C_i|X)$  : Probabilitas kemunculan label kelas  $C_i$  dengan kriteria masukan  $X$

$P(X|C_i)$  : Probabilitas kriteria masukan  $X$  dengan label kelas  $C_i$ .

$P(X)$  : Probabilitas label kelas  $C_i$ .

Berdasarkan pemaparan diatas, maka disimpulkan bahwa algoritma *Naïve Bayes* merupakan salah satu metode algoritma yang memiliki independensi yang kuat (naif) yang digunakan untuk memprediksi kemungkinan di masa yang akan datang melalui pengalaman dimasa lampau.

*Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengatur kinerja suatu metode pada klasifikasi. Pada dasarnya *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya [19].

Pada pengukuran kinerja menggunakan *Confusion matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif.

Pada jenis klasifikasi *binary* yang hanya memiliki 2 keluaran kelas, *Confusion matrix* dapat dilihat pada Tabel 2.1 :

**Tabel 2.1 Confusion Matrix**

| <b>Kelas</b>   | <b>Terklasifikasi Positif</b> | <b>Terklasifikasi Negatif</b> |
|----------------|-------------------------------|-------------------------------|
| <b>Positif</b> | <i>TP (True Positive)</i>     | <i>FN (False Negative)</i>    |
| <b>Negatif</b> | <i>FP (False Positive)</i>    | <i>TN (True Negative)</i>     |

Berdasarkan nilai *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*, dan *True Positive (TP)* diperoleh nilai *akurasi*, *presisi* dan *recall*. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan 2.2. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan persamaan 2.3. Sementara itu, *recall* menunjukkan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Nilai *recall* diperoleh dengan persamaan 2.4

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

$$Presisi = \frac{TP}{FP+TP} * 100\% \quad (2.3)$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad (2.4)$$

Keterangan :

- TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi benar oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

## **2.4 TOOLS DATA MINING**

Dalam melakukan pengolahan data, diperlukan sebuah alat bantu (*tools*) untuk mempermudah dalam melakukan pengolahan data. *Tools* dalam *data mining* diantaranya sebagai berikut :

### **2.4.1 Weka**

*WEKA* merupakan perangkat lunak pembelajaran mesin yang populer dan dikembangkan oleh Universitas Waikato, Selandia Baru menggunakan bahasa pemrograman Java. *WEKA* merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. Dengan mengadopsi konsep *open source software*, menjadikan *WEKA* dapat digunakan dan dimodifikasi siapapun secara gratis. *WEKA* berisikan kumpulan algoritma beserta visualisasinya untuk analisis data dan pemodelan prediktif. Algoritma-algoritma pembelajaran mesin pada *WEKA* dimanfaatkan untuk pemecahan masalah pada bidang *data mining*.

### **2.4.2 RapidMiner**

*RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data*

*mining*, text mining dan analisis prediksi. *RapidMiner* merupakan solusi analisis prediksi, *data mining*, dan text mining yang menggunakan macam-macam teknik prediksi dan deskriptif untuk membuat keputusan yang paling baik bagi penggunanya.

*RapidMiner* mengolah data dengan cara mengekstrak suatu pola dari dataset lalu menggabungkannya dengan database, kecerdasan buatan, dan metode statistika yang bertujuan untuk mendapatkan informasi yang bermutu tinggi dari sebuah teks yang sedang diolah [20] .

*RapidMiner* memiliki beberapa sifat sebagai berikut [21] :

1. Ditulis dengan bahasa pemrograman java sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.
3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.

Beberapa fitur dari *RapidMiner* , antara lain [21] :

1. Banyaknya algoritma *data mining*, seperti decision tree dan *selforganization map*.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D scatter plots.

3. Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
4. Menyediakan prosedur *data mining* dan machine learning termasuk: ETL (extraction, transformation, loading) data preprocessing, visualisasi, modeling dan evaluasi.
5. Proses *data mining* tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
6. Mengintegrasikan proyek *data mining* Weka dan statistika R.

## **2.5 BAHASA PEMROGRAMAN *PHYTON***

*Phyton* adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Phyton* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Phyton* dapat secara khusus membuat *source code* yang mudah dibaca. *Phyton* memiliki *library* yang lengkap sehingga memungkinkan programmer untuk membuat aplikasi yang mutakhir dengan menggunakan source code tampak sederhana [22].

## **2.6 KANKER**

Kanker atau tumor ganas terjadi akibat adanya pertumbuhan sel-sel jaringan tubuh yang tidak normal, disebabkan *neoplasia*, *displasia*, dan *hiperplasia*. *Neoplasia* adalah kondisi sel yang terdapat pada jaringan berproliferasi secara tidak normal dan invasif, *dysplasia* yaitu kondisi sel yang tidak berkembang normal dengan indikasi adanya perubahan pada *nucleus* (inti

sel), *hyperplasia* merupakan kondisi sel normal pada jaringan mengalami pertumbuhan berlebihan [23].

Kanker adalah penyakit yang disebabkan oleh pertumbuhan sel-sel jaringan tubuh yang tidak normal, berkembang dengan cepat, tidak terkendali dan terus membelah diri [24].

Kanker dapat muncul di semua sel atau jaringan tubuh, seperti jaringan ikat, sel paru, sel darah, sel otak, sel kulit, sel hati, dan lain sebagainya, oleh karena itu kanker menurut tempat pertumbuhannya dibagi beberapa jenis yaitu [23] :

1. Karsinoma yaitu sel yang muncul pada lapisan pembatas organ (misalnya kanker kulit, kanker kolon, dan kanker mamae).
2. Sarcoma yaitu kanker yang timbul dari jaringan ikat (misalnya kanker tulang).
3. Leukemia yaitu jumlah sel dalam darah putih meningkat (misalnya kanker darah putih).
4. Lymphoma yaitu kanker yang timbul pada jaringan limfa (misalnya limfosarkoma).

Pada usia tertentu, kanker menjadi lebih mungkin terjadi. Usia yang dikatakan rawan terjadinya kanker yaitu usia >40 tahun, karena pada usia tersebut fungsi – fungsi organ tubuh manusia mengalami penurunan dan memungkinkan adanya pemicu kanker yang berasal dari faktor – faktor internal ataupun eksternal itu sendiri. Selain itu, jenis kelamin juga merupakan faktor yang dapat menjadi pemicu kanker. Perempuan memiliki resiko lebih tinggi terhadap terjadinya



beberapa penyakit dari pada laki – laki diantaranya kanker payudara, kanker serviks dan tiroid.

## **2.7 KANKER PAYUDARA**

Kanker payudara adalah tumor ganas yang dimulai pada sel-sel payudara. Sebuah tumor ganas adalah sekelompok sel-sel kanker yang dapat tumbuh menjadi (menyerang) jaringan sekitarnya atau menyebar (metastasis) ke daerah yang jauh dari tubuh. Penyakit ini terjadi hampir seluruhnya pada wanita, tetapi pria bisa mendapatkannya juga [25].

Kanker payudara merupakan salah satu jenis kanker yang kejadiannya bermula dari sel-sel di payudara yang tidak normal dan terus tumbuh berlipat ganda dan pada akhirnya membentuk benjolan pada payudara. Pertumbuhan sel yang terus- menerus akan menyebabkan tingkat keparahan yang terus berlanjut pada payudara karena sel-sel akan menyebar (*metastasis*) pada bagian tubuh lainnya sehingga berpeluang menyebabkan kematian.

### **2.7.1 Pengobatan Kanker Payudara**

Ada beberapa pengobatan kanker payudara yang harus didahului dengan diagnosa yang lengkap dan akurat. Berikut beberapa cara pengobatan kanker payudara [23] :

#### **1. Pembedahan**

Tergantung pada stadium dan jenis tumornya, lumpektomi mungkin yang diperlukan atau pengangkatan jaringan payudara juga biasa dilakukan. Pembedahan pengangkatan seluruh jaringan payudara di sebut mastektomi.

Selama operasi kelenjar getah bening di ketiak juga dipertimbangkan untuk dilakukan pengangkatan.

## 2. Terapi Radiasi

Terapi radiasi adalah pengobatan tambahan untuk sebagian besar wanita setelah lumpektomi atau mastektomi. Tujuan dari radiasi adalah untuk 16 mengurangi kemungkinan terjadinya kekambuhan. terapi melibatkan penggunaan sinar-X berenergi tinggi atau sinar gamma yang menargetkan tumor. Radiasi ini sangat efektif membunuh sel kanker yang mungkin tersisa setelahnya operasi atau kambuh di mana tumor telah diangkat.

## 3. Terapi Sistemik

Terapi sistemik menggunakan obat-obatan untuk mengobati sel kanker ke seluruh tubuh. Perawatan sistemik termasuk kemoterapi, terapi imun dan terapi hormonal. Kemoterapi dapat digunakan sebelum operasi dan sesudah operasi.

## 4. Kemoterapi

Pengobatan kemoterapi bertujuan menjangkau sel-sel kanker yang menyebar ke bagian tubuh lain dengan cara menghambat dan mengontrol pertumbuhan sel kanker.

### **2.7.2 Perawatan Pasien Setelah Operasi**

Perawatan setelah operasi adalah perawatan yang dilakukan untuk meningkatkan proses penyembuhan luka dan mengurangi rasa nyeri dengan cara merawat luka serta memperbaiki asupan makanan tinggi protein dan vitamin.

Perawatan pasien setelah operasi dimulai saat pasien dipindahkan ke ruang pemulihan dan berakhir sampai evaluasi selanjutnya [26].

Adapun tujuan dari perawatan pasien setelah operasi yaitu :

1. Mengurangi atau menghilangkan risiko dan komplikasi.
2. Mengajarkan pasien melakukan perawatan setelah operasi.
3. Mengantisipasi hal-hal yang mungkin akan dibutuhkan pasien.
4. Menjalankan terapi untuk membantu pasien menjadi lebih mandiri.
5. Memantau kemajuan proses penyembuhan.

Hal-hal yang harus diperhatikan dalam melakukan perawatan pasien setelah operasi sebagai berikut :

1. Pemeriksaan kondisi pasien (suhu, nadi, tekanan darah, pernafasan, saturasi dan tingkat kesadaran).
2. Kebersihan diri dan lingkungan.
3. Pemenuhan nutrisi dan cairan yang baik.
4. Perawatan luka.
5. Istirahat yang cukup.
6. Kontrol secara teratur.
7. Minum obat sesuai anjuran dokter.

## **2.8 KAJIAN PENELITIAN SEJENIS**

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik *data mining* untuk menggali informasi dari sebuah data. Penelitian berikut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang akan digunakan oleh penulis.

Tabel 2. 2 Penelitian Terkait

| Peneliti   | Masalah   | Metode             | Hasil Penelitian  |
|--|---|--------------------|---|
| Hardian Oktavianto, Rahman Puji Handri [27]      | Kanker payudara merupakan salah satu faktor penyebab kematian tertinggi di kalangan wanita, penyakit ini menempati urutan kedua penyebab kematian setelah kanker paru – paru. Pada penelitian ini akan dilakukan klasifikasi kanker payudara dengan menggunakan algoritma <i>Naive Bayes</i> untuk melakukan pengelompokan jenis kanker, apakah jinak atau ganas. | <i>Naive Bayes</i> | Kesimpulan dari penelitian ini adalah performa algoritma naive bayes untuk klasifikasi kanker payudara menghasilkan nilai yang baik, dimana rata – rata persentase data yang terklasifikasi dengan benar mencapai 96.9% dan rata – rata persentase data terklasifikasi salah hanya 3.1%. Sedangkan tingkat efektivitas klasifikasi dengan naive bayes ini termasuk tinggi, dimana rata – rata nilai precision dan recall berada di sekitar 0.96. Nilai precision dan recall paling tinggi yaitu ketika data uji menggunakan percentage split 40% dengan nilai masing – masing secara berurutan mencapai 0.974 dan 0.973 |
| Ilham Mubarog, Arief Setyanto, Heri Sismoro [28] | Angka kematian akibat kanker payudara meningkat karena kurangnya informasi tentang gejala awal dan bahaya dari kanker payudara itu sendiri maka dibutuhkan  | <i>Naive Bayes</i> | Hasil akurasi metode <i>Naive Bayes</i> dilakukan pengujian validasi menggunakan 5-Fold Cross Validation dan mendapatkan nilai  |

|  |   |   |   |
|--|---|---|---|
|  | <p>sebuah sistem yang dapat memberikan informasi tentang penyakit kanker payudara dan cara penanggulangan seperti diagnose secara dini dan penanganannya. Sistem berbasis komputer yang dapat menyelesaikan masalah tersebut adalah sistem klasifikasi <i>Naïve Bayes</i>.</p>  |   | <p>akurasi sebesar 80%, nilai <i>presisi</i> sebesar 83%, dan nilai <i>recall</i> sebesar 83% pada 116 dataset.</p>   |
| <p>Johan Taruna Wijaya, Hardian Oktavianto, Habibatul Azizah Al Faruq [29]</p> | <p>Kanker payudara didefinisikan sebagai suatu penyakit neoplasma ganas yang berasal dari parenchyma dan menghasilkan frekuensi kematian yang menjadi penyebab utama kekhawatiran di dunia. Kanker payudara merupakan kanker kedua yang paling banyak diderita dan penyebab kelima kematian kanker di seluruh dunia dengan presentase 6,4% dari semua penyebab kematian. Pada penelitian ini dilakukan klasifikasi terhadap data kanker payudara, dimana data tersebut terdapat adalah data darah pengidap kanker payudara.</p> | <p><i>K-Nearest Neighbor (Knn) dan Gaussian Naïve Bayes</i></p> | <p>Dari penelitian ini didapatkan hasil pada 116 data darah kanker payudara, metode KNN menghasilkan akurasi 86,9% lebih baik dari pada GNB, dan untuk presisi dan recall, metode KNN menghasilkan presisi sebesar 87,3%, dan recall sebesar 86,7%, pengujian pada metode KNN menggunakan nilai K=4.<br/>Kata</p> |
| <p>Lastri Widya Astuti, Imelda Saluza, Faradilla, M. Fadhiel Alie [30]</p>     | <p>Jumlah penderita kanker payudara meningkat setiap tahunnya sehingga apabila dideteksi atau diketahui lebih awal</p>  | <p><i>Forward Selection pada Naïve Bayes</i></p>                | <p>Simpulan hasil penelitian ada peningkatan nilai akurasi setelah menggunakan metode forward</p>   |

|                       |   |   |  |
|-----------------------|---|---|--|
|                       | bisa dilakukan pencegahan sejak dini sehingga dapat menekan angka penderita kanker payudara. Untuk mengurangi resiko peningkatan jumlah penderita kanker perlu dilakukan deteksi dini, beberapa metoda dapat digunakan untuk membantu proses pendeteksian diawal seperti cancer screening, atau dengan metode komputasi.                                  |   | selection yang digunakan untuk memilih fitur/mereduksi dimensi. Selisih akurasi meningkat sebesar 2,92%, precision meningkat sebesar 3,66% dan recall sebesar 0,93 % setelah menggunakan metode forward selection pada metode naive bayes untuk mengklasifikasi kanker payudara. |
| Laily Hermawanti [31] | Penyakit kanker payudara merupakan salah satu penyakit berbahaya dan penyebab kematian di seluruh dunia. Maka dari itu, penyakit kanker payudara perlu didiagnosis. Algoritma yang digunakan dalam penelitian ini adalah penggabungan algoritma Backward Elimination dan Naive Bayes untuk meningkatkan akurasi dalam diagnosis penyakit kanker payudara. | <i>Backward Elimination dan Naive Bayes</i> | Algoritma Naive Bayes menggunakan dataset kanker payudara menghasilkan akurasi sebesar 96.14% +/- 2.13%, sedangkan algoritma Backward Elimination- Naive Bayes menghasilkan akurasi sebesar 97.00% +/- 2.56% sehingga mengalami peningkatan akurasi.                             |

### 2.8.1 Perbedaan dengan Penelitian Terdahulu

Penelitian yang dilakukan oleh Hardian Oktavianto dkk. mengenai klasifikasi kanker payudara menggunakan algoritma *Naive Bayes* untuk mengelompokkan jenis kanker, apakah jinak atau ganas. Pada penelitian ini

menggunakan metode *Naïve Bayes* yang memiliki keunggulan yaitu cepat, sederhana, dan memiliki akurasi tinggi. Dibandingkan dengan *classifier* lain *Naïve Bayes* bekerja lebih baik dan memiliki tingkat akurasi yang baik dari Teorema Bayes. Perbedaan penelitian Hardian Oktavianto dkk. dengan Penulis ialah terdapat pada data yang digunakan.

Penelitian yang dilakukan oleh Ilham Mubarog dkk. bertujuan untuk memberikan informasi tentang penyakit kanker payudara dan cara penanggulangan seperti diagnosa secara dini dan penanganannya menggunakan metode *Naïve Bayes*. Metode *Naïve Bayes* dinilai sebagai metode klasifikasi yang cukup mudah serta akurat untuk diterapkan pada suatu permasalahan klasifikasi atau data mining. Perbedaan penelitian Ilham Mubarog dkk. dengan Penulis ialah terdapat pada atribut *dataset* yang digunakan dan objek penelitian.

Penelitian yang dilakukan oleh Johan Taruna Wijaya dkk. tentang klasifikasi terhadap data kanker payudara, dimana data tersebut merupakan data darah pengidap kanker payudara dengan menggunakan metode *K-Nearest Neighbor* dan *Gaussian Naïve Bayes*. Algoritma *K-Nearest Neighbor* (K-NN) memiliki beberapa keunggulan yaitu, pelatihan yang sangat cepat, sederhana dan mudah dipelajari, efektif jika data pelatihan besar. *Gaussian Naïve Bayes* digunakan untuk mendistribusikan nilai kontinu yang terkait dengan setiap kelas. Secara keseluruhan kinerja dari dua algoritma ini dalam mengklasifikasikan Breast Cancer sudah cukup baik, terbukti dengan menghasilkan nilai *akurasi*, *presisi*, dan *recall* yang cukup tinggi, Tetapi KNN mampu menghasilkan nilai presisi dan recall lebih tinggi dari pada GNB. Perbedaan penelitian Johan Taruna

Wijaya dkk. dengan Penulis ialah terdapat pada lokasi objek *dataset* yang digunakan dan juga penelitian menggunakan metode *K-Nearest Neighbor* dan *Gaussian Naïve Bayes*.

Penelitian yang dilakukan oleh Lastri Widya Astuti dkk., yang bertujuan mengurangi resiko peningkatan jumlah penderita kanker perlu dilakukan deteksi dini, seperti *cancer screening* atau dengan metode komputasi menggunakan algoritma *Forward Selection* pada *Naïve Bayes*. *Naïve Bayes* dinilai sebagai model yang mudah untuk dibangun dan tidak *complicated*, sehingga tepat untuk *database* yang berukuran kecil sampai ukuran besar. *Forward Selection* digunakan untuk mereduksi dimensi *dataset* dengan menghilangkan atribut-atribut yang tidak relevan atau redundan. Perbedaan penelitian Lastri Widya Astuti dkk. dengan Penulis ialah terdapat pada atribut *dataset* yang digunakan dan juga penelitian menggunakan metode *Forward Selection* pada *Naïve Bayes*.

Penelitian yang dilakukan oleh Laily Hermawanti, untuk meningkatkan akurasi dalam diagnosis penyakit kanker payudara menggunakan algoritma *Backward Elimination* dan *Naïve Bayes*. Algoritma *Naïve Bayes* dinilai mampu dan efisien dalam melakukan klasifikasi dan algoritma *Backward Elimination* berfungsi untuk menghilangkan atribut-atribut yang tidak relevan sehingga dengan penggabungan kedua algoritma ini menghasilkan tingkat akurasi yang lebih tinggi. Perbedaan penelitian Laily Hermawanti dengan Penulis ialah terdapat pada objek penelitian yang digunakan dan juga metode yang digunakan dalam melakukan penelitian.