

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Pengertian Obat**

Obat merupakan bahan yang ditujukan untuk menyembuhkan atau mencegah penyakit, menghilangkan penyakit dan gejalanya serta kegunaan lain pada manusia maupun hewan. Menurut Athijah menyebutkan bahwa sesuai dengan definisi tersebut vitamin dan mineral esensial dapat dikategorikan sebagai obat karena berfungsi untuk memperbaiki keadaan akibat defisiensi suatu zat penting bagi tubuh. Dalam Undang-Undang Kesehatan No. 36 tahun 2009 disebutkan bahwa "obat adalah bahan atau paduan bahan, termasuk produk biologi yang digunakan untuk memengaruhi atau menyelidiki sistem fisiologi atau keadaan patologi dalam rangka penetapan diagnosis, pencegahan, penyembuhan, pemulihan, peningkatan kesehatan dan kontrasepsi untuk manusia". Obat tradisional ialah bahan atau ramuan bahan yang berupa bahan tumbuhan, bahan hewan, bahan mineral, sediaan sarian (galenik), atau campuran dari bahan tersebut yang secara turun-temurun telah digunakan untuk pengobatan, dan dapat diterapkan sesuai dengan norma yang berlaku di masyarakat (Athijah Umami, 2011).

Seiring dengan adanya kehidupan, maka muncul pula penyakit serta telah ada upaya untuk mengobatinya. Upaya pengobatan tersebut tumbuh secara naluriah untuk mempertahankan diri. Bahan yang dipakai untuk pengobatan adalah bahan-bahan yang tersedia di alam baik dari tanaman, hewan maupun mineral. Saat ini telah berkembang obat-obat *modern* yang dihasilkan secara sintesis maupun rekayasa genetik. Ilmuwan yang banyak memberikan kontribusinya dalam bidang obat dan pengobatan, yaitu: Galen seorang dokter Yunani (130 M-200 M) dan Ibnu Sinna (Avicenna) seorang ilmuwan dan ahli filsafat dari Arab (980 M-1037 M) (Zaman-Joenoes, 2001), (Umami Athijah, 2011).

Ada beberapa golongan obat yaitu :

1.) Obat Bebas (OB)

Dalam beberapa peraturan Perundang-undangan yang dikeluarkan oleh Depkes pengertian obat bebas jarang didefinisikan, namun pernah ada salah satu Peraturan Daerah Tingkat II Tangerang yakni Perda Nomor 12 tahun 1994 tentang Izin Pedagang Eceran Obat memuat pengertian Obat Bebas. Obat Bebas adalah obat yang dapat dijual bebas kepada umum tanpa resep dokter dan tidak membahayakan bagi sipemakai. Obat bebas dapat dibeli bebas tanpa resep dokter dan dapat dibeli di apotek dan toko obat berizin untuk mengatasi problem ringan (*minor illness*) yang bersifat nonspesifik. Obat bebas relatif paling aman, boleh digunakan untuk menangani penyakit simptomatis ringan yang banyak diderita masyarakat luas yang penanganannya dapat dilakukan sendiri oleh penderita atau *self medication* (penanganan sendiri).

2.) Obat Bebas Terbatas

Menurut keputusan Menteri Kesehatan RI yang menetapkan obat-obatan kedalam daftar obat "W" memberikan pengertian obat bebas terbatas adalah obat keras yang dapat diserahkan kepada pemakainya tanpa reep dokter, bila penyerahannya memenuhi persyaratan sebagai berikut:

- a. Obat tersebut hanya boleh dijual dalam bungkus asli dari pabriknya atau pembuatnya.
- b. Pada penyerahannya oleh pembuat atau penjual harus mencantumkan tanda peringatan. Tanda peringatan tersebut memuat pemberitahuan berwarna putih. (Evita Aurilia N. 2021)

## 2.2 Data Mining

Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (*massive database*). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan diwaktu yang akan datang (Sikumbang, 2018).

Data mining adalah rangkaian proses pengambilan keputusan dalam mengenali nilai tambah dalam suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Tambunan & Zetli, 2020)

Data mining dapat disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *set* data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa akan datang (Bulolo, 2020). Data mining dapat didefinisikan sebagai proses penambangan data yang menghasilkan sebuah *output* (keluaran) yaitu berupa pengetahuan.

Menurut para ahli Pramudiono mengatakan data mining merupakan salah satu teknik analisis otomatis dari data yang berjumlah besar atau kompleks yang bertujuan untuk menemukan pola yang biasanya tidak disadari keberadaannya, sedangkan menurut Larose menyatakan bahwa data mining adalah teknik analisis dalam meninjau suatu kumpulan data yang berguna bagi untuk menemukan hubungan yang tidak diduga dan dengan cara yang berbeda dengan sebelumnya agar dapat dipahami dan bermanfaat bagi pemilik data (Nofriansya, dan Nurcahyo, 2019).

### 2.2.1 Karakteristik Data Mining

Data mining memiliki karakteristik sebagai berikut : (Ginantara, 2020)

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining menggunakan data yang berukuran besar yang bertujuan agar hasil atau keakuratan data dapat lebih dipercaya.
3. Data mining digunakan sebagai teknik dalam mengambil keputusan di masa depan berdasarkan informasi atau data yang diperoleh melalui data masa lalu.

### 2.2.2 Pengelompokan Data Mining

ada 6 kelompok data mining sebagai berikut, (Buulolo, 2020).

#### 1. Deskripsi

Pada proses deskripsi ini bertujuan untuk menemukan pola yang sering muncul dan mengubah pola tersebut menjadi aturan yang dapat digunakan untuk memudahkan suatu aktivitas.

#### 2. Klasifikasi

Klasifikasi merupakan pengelompokan yang berdasarkan pada hubungan antar variabel kriteria dengan variabel target. Sebagai contoh, penggolongan pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

#### 3. Prediksi

Prediksi merupakan salah satu fungsi data mining yang sering digunakan dan secara umum prediksi hampir sama dengan klasifikasi. Nilai yang didapat dari hasil prediksi diperoleh dari data-data sebelumnya yang digunakan dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

#### 4. Estimasi

Estimasi adalah prediksi atau perkiraan yang hampir sama dengan klasifikasi, yang membedakannya adalah bentuk pengelompokan, di mana estimasi pengelompokan ke arah numerik dan bukan ke arah kategori..

#### 5. Pengklasteran

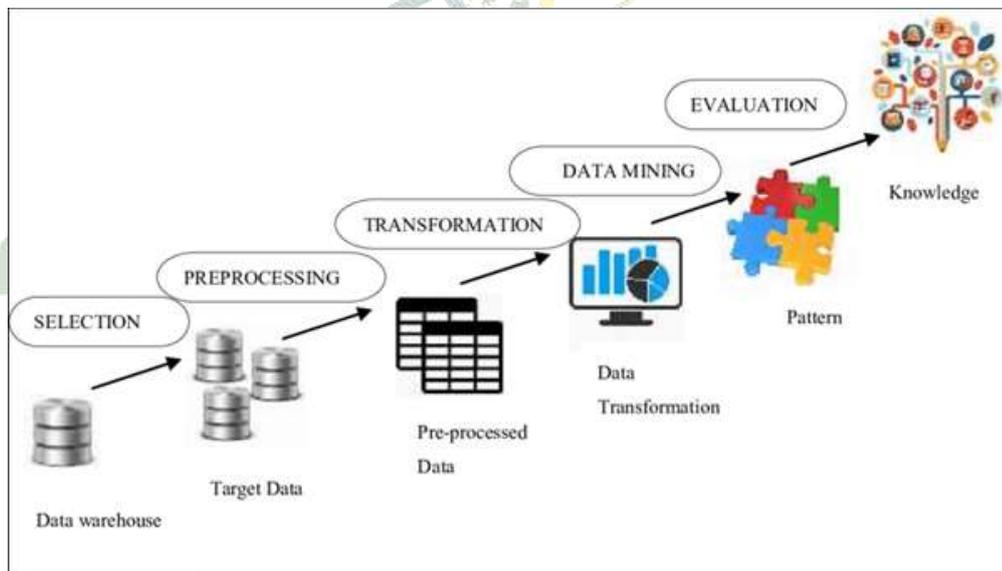
Pengklasteran merupakan pengelompokan data yang memiliki kemiripan nilai atau homogen. Bentuk data yang dapat dikelompokkan dalam pengklasteran adalah hasil pengamatan.

## 6. Asosiasi

Asosiasi adalah kumpulan, himpunan, atau persekutuan. Dalam data mining proses asosiasi merupakan pencarian *attribute* yang sering muncul dalam waktu yang bersamaan.

### 2.2.3 Knowledge Discovery Database (KDD)

Pada proses Data Mining yang biasa disebut *Knowledge Discovery Database* (KDD), terdapat beberapa proses seperti gambar dibawah :



**Gambar 2. 1** Proses *Knowledge Discovery in Database*

(Sumber: Nofriansyah & Nurcahyo, 2019)

## SUMATERA UTARA MEDAN

Pada proses KDD memiliki beberapa fase, antara lain :

### 1. Seleksi Data

Fase seleksi atau pemilihan data dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi akan digunakan untuk proses data mining lalu disimpan dalam suatu berkas terpisah dari basis data operasional

### 2. Pemilihan Data (*Preprocessing Cleaning*)

Proses pemilihan data yaitu, membuang data ganda, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data

3. Transformasi (*Transformation*)

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining.

4. Data Mining

Fase data mining ini dilakukan untuk menerapkan algoritma atau metode pencarian pengetahuan.

5. Interpretasi/ Evaluasi (*Interpretation Evaluation*)

Fase evaluasi merupakan fase terakhir yang berguna dalam proses pembentukan output agar sumber pada proses data mining pola informasi mudah di pahami.

### 2.3 Machine Learning

*Machine learning* merupakan salah satu sub disiplin ilmu dari AI. Metode- metode AI dapat diistilahkan *soft computing*, *machine learning* merupakan salah satu bagian dari itu. *Machine learning* sebagai salah satu bagian dari disiplin ilmu *soft computing* berinsan dengan berbagai disiplin ilmu lainnya, seperti *mathematics*, *statistic programming*, *data science*, *big data*, *data mining database*, *information retrieval computer vision*, *robotic game programming*, *lol expert system decision support witem information system*, dan sebagainya. Salah satu disiplin ilmu yang populer di era boy data saat ini, yaitu *data science*. Secara umum *data science* adalah penggalian atau mengekstrak atau analisis data untuk menemukan data yang benar sehingga menghasilkan suatu informasi atau pengetahuan yang akurat tepat. (Azminuddin.2020)

*Machine Learning* adalah salah satu bidang cabang ilmu komputer yang memberikan kemampuan kepada komputer untuk dapat belajar tanpa diprogram secara eksplisit (Arthur Samuel, 1959). *Machine learning* juga dapat diartikan sebuah komputer yang memiliki kemampuan belajar tanpa diprogram secara eksplisit. Program tersebut memanfaatkan data untuk membangun model dan

mengambil keputusan berdasarkan model yang telah dibangun. Jika kita tinjau kembali, bagaimana cara manusia untuk mengambil keputusan bukan hanya berdasarkan pengalaman tetapi juga dapat berdasarkan logika. Seperti yang telah dijelaskan, *Machine learning* adalah bagian dari *artificial intelligence* yang berfungsi untuk membuat komputer memiliki kemampuan untuk belajar tentang data baru tanpa harus diprogram secara eksplisit. Fokus utamanya adalah membangun sebuah aplikasi komputer yang dapat mempelajari data, lalu membuat sebuah model yang siap digunakan untuk memecahkan kasus tertentu. Berbeda dengan Data Mining, dimana prosesnya dimulai dari pengumpulan database yang berukuran besar yang sebelumnya tidak dapat dipahami dan tidak diketahui dan kemudian menggunakan informasi itu untuk membuat keputusan bisnis yang relevan. Sederhananya, Data Mining dapat diartikan sejumlah metode yang digunakan untuk menemukan pengetahuan berupa relasi dan pola yang berasal dari database yang besar yang sebelumnya tidak diketahui. (Ibnu Daqiqil.2021)

#### **2.4 Algoritma *Naiïve Bayes***

Metode *Naiïve bayes* ini juga sudah banyak digunakan oleh peneliti lainnya, *Naiïve bayes* merupakan sebuah pengklasifikasian probabilistik dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independent atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas.

*Naiïve bayes* juga merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya (Manalu et al., 2017). *Naive Bayes* adalah bentuk paling sederhana dari pengklasifikasi jaringan Bayesian. Dalam *Naive Bayes*, setiap node fitur memiliki simpul kelas sebagai induknya, tetapi tidak memiliki orangtua dari node fitur lainnya. Namun, sejumlah besar pekerjaan dalam pembelajaran terbimbing telah

menunjukkan bahwa pengelompokan *naive bayes* yang begitu sederhana (Fadila, 2020).

Adapun Keuntungan pengguna algoritma *Naïve Bayes* ini yaitu pengguna hanya membutuhkan beberapa jumlah data pelatih (Training Data) yang kecil karena untuk menentukan estimasi parameter yang di perlukan untuk proses pengklasifikasikan karena yang diamsusikan sebagai variable Independent, maka dari itu varians dari suatu variable yang dibutuhkan dalam sebuah kelas untuk menentukan klasifikasi *naive bayes* tersebut Adapun tahap-tahap pemrosesan algoritma *Naive Bayes* diantara lain yaitu :

- a.) Dapat menghitung sejumlah kelas dan label
- b.) Dapat menghitung sejumlah kasus perkelas
- c.) Kemudian di kalikan semua variable kelas tersebut
- d.) Dan terakhir membandingkan hasil perkelas tersebut

Adapun kelebihan dan kekurangan dari algoritma *naive bayes* im yaitu diantara lain:

- 1.) Kelebihan
  - a. Algoritma ini sangat mudah dibuat
  - b. Dan algoritma ini hasilnya juga sangat bagus
- 2.) Kekurangan
  - a. Asumsi *independence* ini sangat atribut untuk membuat akurasi yang berkurang

Persamaan *naïve bayes* :

$$P(C | X) = \frac{P(x | c) * P(c)}{P(x)} \quad (1)$$

Keterangan :

X : Vektor Input

C : Sebuah class spesifik

P(C|X) : Probabilitas class berdasarkan vektor input yang diketahui (*posterioro probability*)

P(c) : Probabilitas class yang dicari (*prior probability*) dari keseluruhan data

P(x|c) : Probabilitas tiap input berdasarkan kondisi pada class

$P(x)$  : Probabilitas suatu input dari keseluruhan data

Berikut adalah contoh kasus sederhana untuk perhitungan manual naïve bayes adalah penentuan jadi main atau tidak main dari faktor-faktor penentu. Seperti dengan melihatnya faktor cuaca

**Tabel 2. 1** Dataset Bermain Golf

(Sumber: Sarwo & Aisyah, 2017)

| No | Cuaca   | Angin     | Temperatur | Kondisi Lapangan | Klasifikasi |
|----|---------|-----------|------------|------------------|-------------|
| 1. | Cerah   | Tidak Ada | 34         | Kering           | Ok          |
| 2. | Hujan   | Tidak Ada | 20         | Licin            | Tidak       |
| 3. | Cerah   | Ada       | 30         | Licin            | Ok          |
| 4. | Mendung | Ada       | 20         | Kering           | Ok          |
| 5. | Cerah   | Tidak Ada | 34         | Berair           | Tidak       |

| No  | Cuaca   | Angin     | Temperatur | Kondisi Lapangan | Klasifikasi |
|-----|---------|-----------|------------|------------------|-------------|
| 6.  | Hujan   | Ada       | 20         | Berair           | Tidak       |
| 7.  | Mendung | Ada       | 31         | Kering           | Ok          |
| 8.  | Cerah   | Tidak Ada | 34         | Licin            | Ok          |
| 9.  | Cerah   | Ada       | 30         | Kering           | Ok          |
| 10. | Hujan   | Ada       | 22         | Berair           | Tidak       |
| 11. | Cerah   | Ada       | 34         | Kering           | Ok          |
| 12. | Mendung | Tidak Ada | 20         | Kering           | Ok          |
| 13. | Mendung | Ada       | 24         | Licin            | Ok          |
| 14. | Hujan   | Ada       | 21         | Berair           | Tidak       |
| 15. | Hujan   | Ada       | 22         | Licin            | Ok          |
| 16. | Cerah   | Ada       | 32         | Kering           | Ok          |
| 17. | Cerah   | Ada       | 30         | Licin            | Ok          |
| 18. | Cerah   | Ada       | 33         | Kering           | Ok          |
| 19. | Hujan   | Tidak Ada | 20         | Berair           | Tidak       |
| 20. | Hujan   | Ada       | 23         | Licin            | Tidak       |

Data rekomendasi bermain golf terdiri dari cuaca, angin, temperature, kondisi lapangan, dan klasifikasi. Berikut pemahaman dari data bermain golf adalah :

a) Cuaca

Merupakan variabel jenis cuaca yang dikelompokkan kedalam tiga kategori yaitu hujan, cerah, dan mendung.

b) Angin

Merupakan variabel jenis angin yang dikelompokkan kedalam dua kategori yaitu ada dan tidak ada.

c) Temperatur

Merupakan variabel jenis temperatur yang dikelompokkan kedalam dua kategori yaitu panas dan dingin. Dimana panas dengan diatas suhu  $25^{\circ}$  dan dingin dibawah suhu  $25^{\circ}$ .

d) Kondisi lapangan

Merupakan variabel jenis kondisi lapangan yang dikelompokkan kedalam tiga kategori yaitu kering, licin, dan berair. e. Klasifikasi Merupakan variabel jenis kondisi lapangan yang dikelompokkan kedalam dua kategori yaitu ya main dan tidak main.

Berikut perhitungan manual Naïve Bayes dengan menggunakan dataset pada tabel data training :

$P / C$

$P$  : Jumlah Data

$C$  : Class

Menghitung jumlah kelas

$P(Y= Ok) = 16/20$  "Jumlah data ok main pada data training dibagi dengan jumlah keseluruhan data"  $P(Y= Tidak) = 4/20$  "Jumlah data tidak main pada data training dibagi dengan jumlah keseluruhan data"

Menghitung jumlah kasus yang sama dengan kelas yang sama

$P(\text{Cuaca} = \text{Hujan} | Y= Ok) = 5/16 = 0,312$   $P(\text{Cuaca} = \text{Hujan} | Y= Tidak) = 2/4 = 0,5$

$P(\text{Cuaca} = \text{Cerah} | Y= Ok) = 7/16 = 0,437$   $P(\text{Cuaca} = \text{Cerah} | Y= Tidak) = 2/4 = 0,5$

$P(\text{Cuaca} = \text{Mendung} | Y= Ok) = 4/16 = 0,25$   $P(\text{Cuaca} = \text{Mendung} | Y= Tidak) = 0/4 = 0$

$$P(\text{Angin} = \text{Ada} | Y = \text{Ok}) = 14/16 = 0,875$$

$$P(\text{Angin} = \text{Ada} | Y = \text{Tidak}) = 1/4 = 0,25$$

$$P(\text{Angin} = \text{Tidak} | Y = \text{Ok}) = 2/16 = 0,125$$

$$P(\text{Angin} = \text{Tidak} | Y = \text{Tidak}) = 3/4 = 0,75$$

$$P(\text{Temperatur} = \text{Panas} | Y = \text{Ok}) = 8/16 = 0,5$$

$$P(\text{Temperatur} = \text{Panas} | Y = \text{Tidak}) = 2/4 = 0,5$$

$$P(\text{Temperatur} = \text{Dingin} | Y = \text{Ok}) = 8/16 = 0,5$$

$$P(\text{Temperatur} = \text{Dingin} | Y = \text{Tidak}) = 2/4 = 0,5$$

$$P(\text{Kondisi Lapangan} = \text{Kering} | Y = \text{Ok}) = 8/16 = 0,5$$

$$P(\text{Kondisi Lapangan} = \text{Kering} | Y = \text{Tidak}) = 0/4 = 0$$

$$P(\text{Kondisi Lapangan} = \text{Licin} | Y = \text{Ok}) = 4/16 = 0,25$$

$$P(\text{Kondisi Lapangan} = \text{Licin} | Y = \text{Tidak}) = 3/4 = 0,75$$

$$P(\text{Kondisi Lapangan} = \text{Berair} | Y = \text{Ok}) = 4/16 = 0,25$$

$$P(\text{Kondisi Lapangan} = \text{Berair} | Y = \text{Tidak}) = 1/4 = 0,25$$

Kalikan semua hasil variabel dengan bermain ok dan bermain tidak

$$P = (X|C1) * (X|C2) \dots\dots\dots (2)$$

P = Probabilitas kemunculan kelas Ci dengan kondisi X

X = Kejadian

Ci = Kelas yang tersedia

$$P(\text{Cuaca} = \text{Hujan} | Y = \text{Ok}) * P(\text{Cuaca} = \text{Cerah} | Y = \text{Ok}) * P(\text{Cuaca} = \text{Mendung} | Y = \text{Ok}) * P(\text{Angin} = \text{Ada} | Y = \text{Ok}) * P(\text{Angin} = \text{tidak} | Y = \text{Ok}) * P(\text{Temperatur} = \text{Panas} | Y = \text{Ok}) * P(\text{Temperatur} = \text{Dingin} | Y = \text{Ok}) * P(\text{Kondisi Lapangan} = \text{Kering} | Y = \text{Ok}) * P(\text{Kondisi Lapangan} = \text{Licin} | Y = \text{Ok}) * P(\text{Kondisi Lapangan} = \text{Berair} | Y = \text{Ok})$$

$$C1 = 0,312 * 0,437 * 0,25 * 0,875 * 0,125 * 0,5 * 0,5 * 0,5 * 0,25 * 0,25$$

$$C1 = 0,00003$$

$$P(\text{Cuaca} = \text{Hujan} | Y = \text{Tidak}) * P(\text{Cuaca} = \text{Cerah} | Y = \text{Tidak}) * P(\text{Cuaca} = \text{Mendung} | Y = \text{Tidak}) * P(\text{Angin} = \text{Ada} | Y = \text{Tidak}) * P(\text{Angin} = \text{tidak} | Y = \text{Tidak}) * P(\text{Temperatur} = \text{Panas} | Y = \text{Tidak}) * P(\text{Temperatur} = \text{Dingin} | Y = \text{Tidak}) *$$

$P(\text{Kondisi Lapangan} = \text{Kering} | Y = \text{Tidak}) * P(\text{Kondisi Lapangan} = \text{Licin} | Y = \text{Tidak}) * P(\text{Kondisi Lapangan} = \text{Berair} | Y = \text{Tidak})$

$C2 = 0,5 * 0,5 * 0 * 0,25 * 0,75 * 0,5 * 0,5 * 0 * 0,75 * 0,25$

$C2 = 0$

Kemudian bandingkan hasil kriteria pengerjaan kelas Main “Ok” dengan Main “Tidak”, dari hasil diatas terlihat bahwa  $P(X|C1) > P(X|C2)$ . Nilai probabilitas tertinggi adalah pada kelas (P|OK) Maka kesimpulannya X1 adalah kriteria kelas Main “Ok”.

## 2.5 PHP (*Hypertext Preprocessor*)

PHP adalah PHP (*Hypertext Preprocessor*) adalah bahasa pemrograman web berbasis server (server side) yang mampu memarsing kode PHP dari kode dengan ekstensi PHP sehingga menghasilkan tampilan website yang dinamis disisi clien (Sitohang et al., 2018). PHP (*Hypertext Preprocessor*) merupakan bahasa pemrograman yang digunakan menerjemahkan baris kode program menjadi kode mesin yang dapat dimengerti oleh komputer yang bersifat *server-side* yang dapat ditambahkan ke dalam HTML (Supono dan Putratama, V., 2018).

PHP adalah bahasa pelengkap HTML yang memungkinkan dibuatnya aplikasi dinamis yang memungkinkan adanya pengolahan data dan pemrosesan data. Semua syntax yang diberikan akan sepenuhnya dijalankan pada server sedangkan yang dikirimkan ke browser hanya hasilnya saja. Kemudian merupakan bahasa berbentuk script yang ditempatkan dalam server dan diproses di server. Hasilnya akan dikirimkan ke client, tempat pemakai menggunakan browser. PHP dikenal sebagai sebuah bahasa scripting, yang menyatu dengan tag HTML, dieksekusi di server, dan digunakan untuk membuat halaman web yang dinamis seperti halnya *Active Server Pages (ASP) atau Java Server Pages (JSP)*. PHP merupakan sebuah *software Open Source* (Hermiati et al., 2021)



**Gambar 2. 2** Logo PHP

(Sumber: [www.php.net](http://www.php.net))

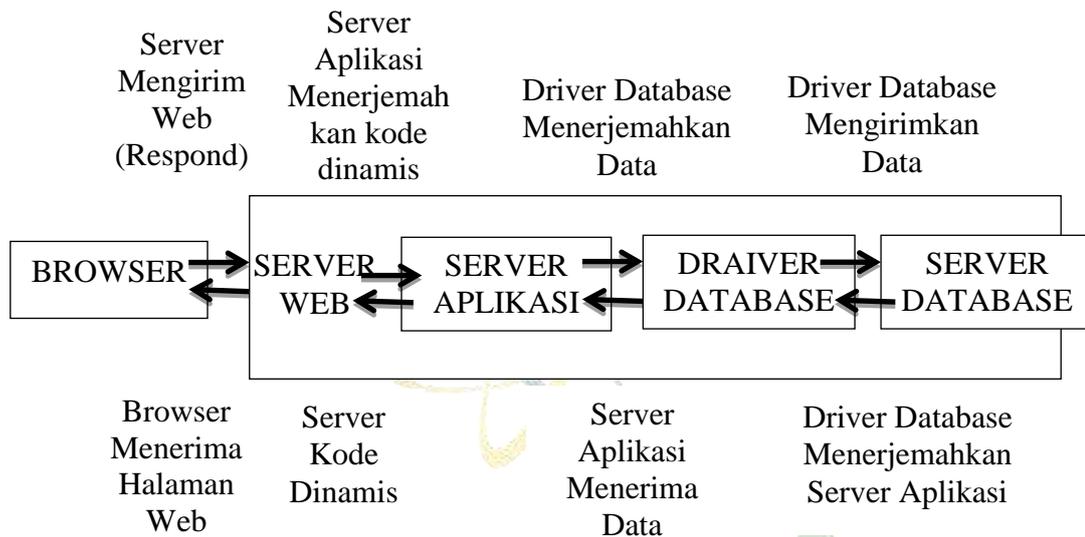
### **2.5.1 Sejarah PHP (*Hypertext Preprocessor*)**

PHP awalnya singkatan dari *Personal Home Page* (Situs Personal). PHP pertama kali diciptakan oleh Rasmus Lerdorf pada tahun 1995. Saat itu PHP masih disebut FI (*From Interpreted*), yang bentuknya berupa sekumpulan script yang digunakan untuk mengolah data form dari web.

Pada tahun 1997, pada perusahaan yang bernama Zend menulis ulang interpreter PHP menjadi lebih baik, kemudian pada 1998 tersebut merilis interpreter baru untuk PHP dan meresmikan rilis tersebut sebagai PHP 3.0 dan singkatan PHP diubah menjadi akronim berulang PHP: *Hypertext Preprocessor*. PHP difokuskan pada *scripting server-side*, maka dapat melakukan apa yang bisa dilakukan CGI dengan menggunakan PHP seperti mengambil data inputan form, meng-*generate* konten halaman dinamis, dan sebagainya. Kemampuan dan *support*-nya untuk *database* juga sangat dapat diandalkan (Supono dan Putratama, V., 2018).

### 2.5.2 Prinsip Kerja PHP

Bahasa pemrograman PHP merupakan Bahasa pemrograman yang dikategorikan kepada *Web Server* untuk menjalankannya. Berikut adalah gambar dari Prinsip kerja PHP :



**Gambar 2. 3** Prinsip Kerja PHP  
(Sumber: Supono dan Putratama, V. 2018)

Penjelasan dari gambar :

1. *Client/user* mengirimkan file PHP (menggunakan *browser*) melalui *Web Server* (seperti *Internet Explorer*, *Mozilla Firofox*, *Google Chrome*, dan lain-lain).
2. *Web Server* mendapatkan *request* atau permintaan dari *user* lalu meneruskan ke *Server* melalui jaringan internet.
3. *Web Server* lalu meneruskan permintaan file PHP tersebut ke *PHP Processor* dapat berupa modul (bagian dari *web-server*) atau terpisah sebagai *CGI*.
4. Permintaan diproses oleh PHP dan diteruskan ke *database* (jika terdapat permintaan ke *database*), kemudian hasilnya dikirim kembali ke *web-server*.
5. *Web server* memaket kemabli hasil tersebut dengan menambahkan *HTTP header* dan dikirim kembali ke browser melalui jaringan Internet.

6. *Browser* memproses HTTP paket dan menampilkannya kembali kepada *user* sebagai file HTML.

## 2.6 MySQL

MySQL merupakan salah satu *database server* yang terkenal. MySQL menggunakan SQL sebagai bahasa dasar agar dapat mengakses *database* nya menggunakan SQL. MySQL bersifat *open source* sehingga dapat digunakan secara gratis oleh penggunanya (Guntara et al., 2020). PHP dan MySQL merupakan dua tool yang akan digunakan untuk membuat aplikasi berbasis *website*. PHP dipakai untuk mengatur interaksi antara *user* dan *database*, sementara MySQL berfungsi sebagai server penyimpanan data.

MySQL (*My Structure Query Language*) adalah sebuah perangkat lunak sistem manajemen basis data SQL (*Database Management System*) atau DBMS dari sekian banyak DBMS, seperti *Oracle*, *MS SQL*, *Postagre SQL*, dan lain-lain. MySQL merupakan DBMS yang multithread, multi-user yang bersifat gratis di bawah lisensi GNU *General Public Licence* (GPL). Tidak seperti *Apache* yang merupakan *software* yang dikembangkan oleh komunitas umum, dan hak cipta untuk kode sumber dimiliki oleh penulisnya masing-masing.

Pada perkembangannya, MySQL disebut juga SQL yang merupakan singkatan dari *Structured Query Languange*. SQL merupakan bahasa terstruktur yang khusus digunakan untuk mengolah *database*. SQL pertama kali didefinisikan oleh *American National Standards Institute* (ANSI) pada tahun 1986. MYSQL adalah sebuah sistem manajemen *database* yang setbersifat *open source*. MYSQL merupakan sistem manajemen *database* yang bersifat relational. Artinya, data yang dikelola dalam *database* yang akan diletakkan pada beberapa tabel yang terpisah sehingga manipulasi data akan jauh lebih cepat. MYSQL dapat digunakan untuk mengelola *database* mulai dari yang kecil sampai dengan yang sangat besar (Novendri, 2019).



**Gambar 2. 4** Logo MySQL

(Sumber: [www.mysql.com](http://www.mysql.com))

### 2.6.1 Kelebihan MySql

Ada beberapa kelebihan dari MySql sebagai berikut : (Anhar, 2010)

1. MySQL dapat berjalan dengan stabil pada berbagai sistem operasi, seperti *Windows, Linux, FreeBSD, Mac Os X Server, Solaris*, dan masih banyak lagi.
2. Bersifat *Open Source*, MySQL didistribusikan secara *open source* (gratis), di bawah lisensi *GNU General Public Licence (GPL)*.
3. Bersifat Multiuser, MySQL dapat digunakan oleh beberapa *user* dalam waktu yang bersamaan tanpa mengalami masalah.
4. MySQL memiliki kecepatan yang baik dalam menangani *query* (perintah SQL). Dengan kata lain, dapat memproses lebih banyak SQL per satuan waktu.
5. Dari segi *security* atau keamanan data, MySQL memiliki beberapa lapisan sekuriti, seperti level *subnet mask*, nama *host*, dan izin akses *user* dengan sistem perizinan yang mendetail serta *password* yang terenkripsi.
6. Selain MySQL bersifat *fleksibel* dengan berbagai pemrograman, MySQL juga memiliki *interface* (antarmuka) terhadap berbagai

aplikasi dan bahasa pemrograman dengan menggunakan fungsi API (*Application Programming Interface*).

## 2.7 Notepad++



**Gambar 2. 5** Logo Notepad++

(Sumber: [www. Notepad++.com](http://www.Notepad++.com))

*Notepad++* adalah sebuah text editor yang sangat berguna bagi setiap orang dan khususnya bagi para developer dalam membuat program. Notepad++ menggunakan komponen Scintilla untuk dapat menampilkan dan menyuntingan 30 teks dan berkas kode sumber berbagai bahasa pemrograman yang berjalan diatas sistem operasi *Microsoft Windows*. Selain manfaat dan kemampuannya menangani banyak bahasa pemrograman, Notepad ++ juga dilisensikan sebagai perangkat free. Jadi, setiap orang yang menggunakannya tidak perlu mengeluarkan biaya untuk membeli aplikasi ini karena *sourceforge.net* sebagai layanan yang memfasilitasi Notepad ++ membebaskannya untuk digunakan (Saifudin & Setiaji, 2019).

## 2.8 Website

*Website* merupakan kumpulan dari halaman-halaman web yang berhubungan dengan file-file lain yang saling terkait. Dalam sebuah *website* terdapat suatu halaman yang dikenal dengan sebutan *home page*. *Home page* adalah sebuah halaman yang pertama kali ketika seseorang mengunjungi sebuah *website*. Dari

*home page*, pengunjung dapat mengklik *hyperlink* untuk pindah ke halaman lain yang terdapat dalam *website* tersebut. Sebuah *home page* biasanya merupakan sebuah file dengan nama *index.htm* atau *index.html* (Novendri, 2019).

*Website* dapat diartikan sebagai kumpulan halaman yang berisi informasi data digital baik berupa teks, gambar, animasi, suara dan video atau gabungan dari semuanya yang disediakan melalui jalur koneksi internet sehingga dapat diakses dan dilihat oleh semua orang di seluruh dunia. Halaman *website* dibuat menggunakan bahasa standar yaitu HTML. Skrip HTML ini akan diterjemahkan oleh *web browser* sehingga dapat ditampilkan dalam bentuk informasi yang dapat dibaca oleh semua orang (Muhammad Romzi & Kurniawan, 2020).

## 2.9 XAMPP

*Xampp* merupakan salah satu *Software Package* dalam pengembangan web. *Xampp* mengumpulkan beberapa perangkat lunak pengembangan web seperti PHP, *Apache*, MySQL dan PHPMyAdmin merupakan salah satu buah *installer* saja. Proses konfigurasi juga sudah dilakukan secara otomatis, sehingga sangat mudah dan praktis. Untuk menjalankan *xampp* pertama kali pastikan bahwa *xampp* beserta PHP, *Apache* dan MySQL berjalan dengan baik, bukalah *browser* dan ketikkan alamat <http://localhost> atau <http://localhost/xampp>. (Solichin A. 2016). *Xampp* terdapat beberapa fitur yang digunakan untuk menjalankan kode PHP, termasuk *web server Apache*. *Web server apache* berguna untuk memilih cara menjalankan kode *script* yang telah ditulis, sehingga *apache* akan memberitahu *web server* bahwa kode yang sedang dijalankan kode PHP. (Jannah M., 2019). *XAMPP* adalah salah satu paket *installer* yang berisi *Apache* yang merupakan *web server* tempat menyimpan file-file yang diperlukan *website*, dan *Phpmyadmin* sebagai aplikasi yang digunakan untuk perancangan database MySQL. (Hermiati et al., 2021)

*XAMPP* merupakan perangkat lunak bebas, yang mendukung banyak sistem operasi, dan merupakan kompilasi dari beberapa program. Seperti *Apache*, MySQL, PHPP, dan *Perl*. *Xampp* adalah tool yang menyediakan paket perangkat lunak dalam satu buah paket. Dalam paket *Xampp* sudah terdapat *Apche*

(*Web Server*), *MySQL (Database)*, *PHP (server side scripting)*, *Perl*, *FTP server*, *PhpMyAdmin*, dan berbagai pustaka bantu lainnya. (Novendri, 2019).

Kepanjangan dari *Xampp* yaitu :

**X** : Program ini dapat dijalankan dibanyak sistem operasi, seperti *Windows*, *Linux*, *Mac OS* dan juga *Solaris*.

**A** : *Apache* merupakan aplikasi *web server*. Tugas utama dari *Apache* adalah menghasilkan halaman *web* yang benar kepada *user* berdasarkan kode *PHP* yang dituliskan oleh pembuat *web* atau *user*.

**M** : *MySQL*, merupakan aplikasi data *server*. Perkembangannya disebut juga *Sql* yang merupakan kepanjangan dari *Structured Query Language*. *Sql* merupakan bahasa terstruktur yang digunakan untuk mengolah *database*.

**P** : *PHP*, merupakan bahasa *pemrograman web*, dimana *user* dapat menggunakan bahasa pemrograman ini untuk membuat web yang bersifat *server-side scripting*.

**P** : *Perl*, yaitu merupakan bahasa pemrograman untuk segala keperluan, dan dikembangkan pertama kali oleh *Larry Wall* di mesin *Unix*

### 2.10 Confusion matrix

Sebuah sistem yang melakukan klasifikasi tentunya diharapkan bisa memprediksi dengan benar. Namun tidak dapat dipungkiri jika pada penerapannya sistem tidak akan bisa 100% benar. Hal ini yang mengharuskan pengukuran kinerja dari sistem tersebut untuk mengetahui berbagai aspek pengujiannya. Menentukan suatu model dari sistem baik atau buruk memerlukan elemen-elemen kunci diantaranya

- a) Akurasi Prediksi, yang menentukan seberapa akurat suatu model dalam memprediksi keluaran.
- b) Kecepatan, yang menunjukkan seberapa cepat suatu model dalam memproses data masukan.

- c) Robustness, menggambarkan kemampuan suatu model melakukan prediksi yang akurat walau dalam kondisi ekstrim dan banyak gangguan terjadi
- d) d. Skalabilitas, kemampuan suatu model memproses data baik dalam ukuran yang lebih besar maupun data dari bidang lain yang berbeda.

Umumnya, pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix*. *Confusion matrix* merupakan alat pengukuran yang dapat digunakan untuk menghitung kinerja atau tingkat kebenaran proses klasifikasi (Lubis et al., 2022). Dengan *confusion matrix* dapat dianalisa seberapa baik *classifier* dapat mengenali *record* dari kelas-kelas yang berbeda. Tabel *confusion matrix* ditunjukkan pada gambar berikut ini:

|                  |              | Actual Values   |  |
|------------------|--------------|---|--|
|                  |              | 1 (Positive)  | 0 (Negative)   |
| Predicted Values | 1 (Positive) | <b>TP</b><br>(True Positive)                          | <b>FP</b><br>(False Positive)<br><i>Type I Error</i> |
|                  | 0 (Negative) | <b>FN</b><br>(False Negative)<br><i>Type II Error</i> | <b>TN</b><br>(True Negative)                         |

**Gambar 2. 6 Confusion matrix (Lubis et al., 2022)**

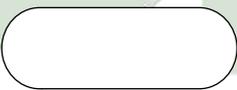
Keterangan gambar 2.6 berikut ini

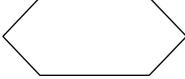
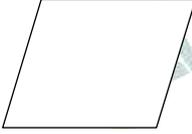
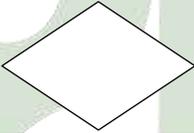
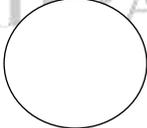
- a) TP (*True Positive*) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas positif.
- b) FN (*False Negative*) merupakan banyaknya data yang kelas aktualnya adalah kelas positif dengan kelas prediksinya merupakan kelas negatif.
- c) FP (*False Positive*) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas positif.
- d) TN (*True Negative*) merupakan banyaknya data yang kelas aktualnya adalah kelas negatif dengan kelas prediksinya merupakan kelas negatif.

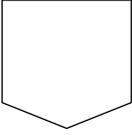
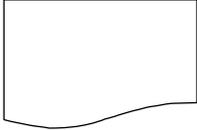
## 2.11 Flowchart

*Flowchart* atau bagan alur adalah metode untuk menggambarkan tahap-tahap penyelesaian masalah beserta aliran data dengan simbol-simbol standar yang mudah dipahami. Tujuan dari *flowchart* adalah untuk menyederhanakan rangkaian proses untuk memudahkan pemahaman pengguna pada informasi tersebut, maka desain sebuah *flowchart* harus ringkas, jelas, dan logis. (Soeherman dan Pinontoan, 2008). *Flowchart* atau sering disebut dengan diagram alir merupakan suatu jenis diagram yang merepresentasikan algoritma atau langkah-langkah instruksi yang berurutan dalam sistem. seorang analis sistem menggunakan *flowchart* sebagai bukti dokumentasi untuk menjelaskan gambaran logis sebuah sistem yang akan dibangun kepada programmer. Dengan begitu, *flowchart* dapat membantu untuk memberikan solusi terhadap masalah yang bisa saja terjadi dalam membangun sistem. Pada dasarnya, *flowchart* digambarkan dengan menggunakan simbol-simbol. Setiap simbol mewakili suatu proses tertentu. Sedangkan untuk menghubungkan satu proses ke proses selanjutnya digambarkan dengan menggunakan garis penghubung.

**Tabel 2. 2** Simbol *Flowchart*  
(Sumber: Ajar et al., n.d.)

| No. | Simbol  | Keterangan   |
|-----|---|--|
| 1.  | <br>Terminator | Terminator untuk permulaan program dan akhir program |
| 2.  | <br>Garis Alur | Garis alur untuk arah aliran program                 |

|    |   |  |
|----|---|--|
| 3. | <br>Simbol Persiapan Program | Yaitu untuk proses instalasi dan pemberian nilai awal  |
| 4. | <br>Simbol Proses            | Proses penghitungan atau proses pengolahan data  |
| 5. | <br><i>Input dan Output</i>  | Untuk proses <i>Input</i> dan <i>Output</i> data, parameter dan informasi                                      |
| 6. | <br>Simbol Sub Proses      | Simbol ini untuk permulaan sub-program dan proses menjalankan sub-program                                      |
| 7. | <br>Simbol Keputusan       | Untuk membandingkan pernyataan yang ada dan penyeleksian data yang memberikan pilihan untuk langkah sebelumnya |
| 8. | <br><i>Connector</i>       | Untuk menghubungkan bagian Flowchart yang berada pada satu halaman   |

|     |  |  |
|-----|--|--|
| 9.  | <br>Simbol Penghubung | Untuk menghubungkan bagian flowchart yang berbeda                                    |
| 10. | <br>Document          | Input/Output dalam format yang dicetak   |
| 11. | <br>Disk Storage      | Input/Output ini untuk menggunakan penyimpanan akses                                 |
| 12  | <br>Punche Card     | Untuk menyatakan input yang berasal dari kartu dan output yang ditulis didalam kartu |

## 2.12 Penelitian Terkait

**Tabel 2. 3** Penelitian Terdahulu

| No | Peneliti                     | Judul   | Kesimpulan   |
|----|------------------------------|---|--|
| 1. | (Ratnawati, 2018)<br>Sinta 3 | Implementasi Algoritma <i>Naive Bayes</i> Terhadap Analisis Sentimen Opini Film Pada <i>Twitter</i> .             | Data opini film pada penelitian ini diambil dari komentar film yang ditulis di twitter. Banyaknya opini yang dituliskan di twitter membutuhkan pengklasifikasian sesuai sentimen yang dimiliki agar mudah untuk mendapatkan kecenderungan opini tersebut terhadap film apakah cenderung beropini positif atau negatif. Algoritma yang akan digunakan pada penelitian ini adalah Algoritma Naive Bayes. Berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90 % dengan rincian nilai precision 92%, recall 90% dan f-measure 90%..                     |
| 2. | (Fitriani, 2019)<br>Sinta 2  | Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur | Penelitian ini akan melakukan pengolahan Data yang diperoleh dari Komisi Pemilihan Umum (KPU) Sidoarjo. Data tersebut di uji menggunakan metode klasifikasi Naive Bayes dengan Tools Weka dan Websitedengan 6 variabel yang sudah ditentukan. Dataset di ambil sebanyak 300 data dibagi 2 yaitu sebanyak 65% dari 195 data Training dan sebanyak 35% dari 105 data Testing. Hasil prediksi partisipasi pemilu dari dataset yang diambil sebanyak 300 data dibagi 2 yaitu sebanyak 65% dari 195 data Training dan sebanyak 35% dari 105 data Testing. Hasil prediksi berdasarkan set atribut kehadiran, untuk kehadiran |

| No | Peneliti                       | Judul   | Kesimpulan   |
|----|--------------------------------|---|--|
|    |                                |   | dengan 105 data diperoleh nilai 97% prediksi kebenarannya dan diperoleh nilai 3% prediksi kesalahannya   |
| 3  | (Widaningsih, 2019)<br>Sinta 3 | Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm | Penelitian ini melakukan prediksi nilai dengan. Teknik yang akan digunakan untuk model data mining klasifikasi ini terdiri dari empat algoritma yaitu C4.5, Support vector machine (SVM), k-nearest neighbor (kNN) dan Naïve Bayes. Metoda klasifikasi terdiri dari variabel-variabel prediktor dan satu variabel target. Variabel-variabel prediktor terdiri dari jenis kelamin dan indeks prestasi dari semester 3 hingga 6. Perangkat lunak yang digunakan untuk mengolah data yaitu software Rapid Miner. Hasil akhir dari keempat algoritma tersebut diperoleh bahwa algoritma Naïve Bayes merupakan algoritma terbaik untuk memprediksi kelulusan mahasiswa yang tepat waktu dan $IPK \geq 3$ dengan nilai accuracy (76,79%), error (23,17%), dan AUC (0,850). |

| No | Peneliti                    | Judul   | Kesimpulan  |
|----|-----------------------------|---|---|
| 4  | (Annur,<br>2018)<br>Sinta 2 | Klasifikasi<br>Masyarakat<br>Miskin<br>Menggunakan<br>Metode Naive<br>Bayes | <p>Penelitian akan melakukan klasifikasi berdasarkan data penduduk miskin yang diperoleh dari Kecamatan Tibawa dengan menggunakan teknik data mining. Atribut yang akan digunakan dalam melakukan klasifikasi penduduk adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin/Belum Kawin). Metode yang akan digunakan adalah metode Naive Bayes Classifier, Berdasarkan hasil pengujian confusion matrix dengan teknik split validasi, penggunaan metode klasifikasi naive bayes terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai Precision sebesar 92% dan Recall sebesar 86%</p> |

| No | Peneliti                         | Judul  | Kesimpulan  |
|----|----------------------------------|--|---|
| 5  | (Damuri et al., 2021)<br>Sinta 4 | Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako | Salah satu aspek penting untuk mendukung Strategi Penanggulangan Kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran. Naïve Bayes merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. Hasil klasifikasi yang dilakukan nantinya akan membantu pengelola bantuan untuk mengambil keputusan terkait klasifikasi penentuan penerima bantuan sembako. Prediksi penerima bantuan sembako yang digunakan terdapat dua kelas, yaitu layak dan tidak layak. Data yang digunakan untuk prediksi yaitu data sampel dari desa XYZ. Pada penelitian ini algoritma Naïve Bayes diimplementasikan dan dianalisa menggunakan aplikasi yang dikembangkan berbasis web. Dari hasil evaluasi menggunakan confusion matrix didapatkan akurasi yang dihasilkan untuk 135 data training dengan 40 data testing dan tujuh atribut yang digunakan menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%. |

| No | Peneliti                           | Judul  | Kesimpulan   |
|----|------------------------------------|--|--|
| 6  | (Abdullah et al., 2022)<br>Sinta 4 | Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti) | Penulis mengimplementasikan penelitian ini menggunakan bahasa pemrograman web yaitu PHP dengan menggunakan framework codeIgniter dengan basis data MySQL. Sistem yang dibangun dengan algoritma naïve bayes ini meliputi data penjualan produk, perhitungan naïve dari masing-masing atribut serta pelaporan.. Hasil prediksi yang diperoleh dengan menggunakan metode algoritma naïve bayes menghasilkan informasi yang dapat digunakan oleh toko untuk mengidentifikasi produk terlaris yang dibeli konsumen sehingga dapat membantu toko prapti untuk menemukan dan menentukan target pasar dengan lebih akurat. Sumber data yang diambil dari 1 tahun sebelumnya dengan keakuratan sistem menggunakan confusion matrix menghasilkan accuracy 83,3%, precision 84,2% dan recall 88,9% |

| No | Peneliti                           | Judul  | Kesimpulan  |
|----|------------------------------------|--|---|
| 7  | (Prasetya, 2022)<br>Sinta 4        | Analisis Data Mining<br>Klasifikasi Berita Hoax COVID 19<br>Menggunakan Algoritma Naive Bayes            | Dari sini peneliti mencoba melakukan eksperimen pada algoritma klasifikasi naïve bayes untuk mengklasifikasikan berita hoax covid 19. Berdasarkan hasil penelitian yang telah dilakukan yang model naïve bayes dan cross validation dapat melakukan klasifikasi berita hoax dengan baik, akurasi yang dihasilkan sebesar 86.3% dimana 80-90% masuk pada kriteria good classification. Data yang diprediksi salah juga tidak terlalu banyak dari total 300 dataset hanya 41 yang dinyatakan salah dalam pelabelan tidak sampai 2% dari keseluruhan total dataset   |
| 8  | (Gusderia et al., 2022)<br>Sinta 3 | Data Mining Untuk Klasifikasi Data Penjualan Alat Teknik<br>Menggunakan Metode Naive Bayesian Classifier | PT. Faedah Teknik dituntut untuk mencari solusi dalam pengelompokan produk dan menemukan segmentasi jenis produk guna untuk meningkatkan usaha dibidang penjualan. Maka dari itu teknik yang akan digunakan pada karya ilmiah ini adalah klasifikasi dengan menggunakan algoritma Naïve Bayesian Classifier. Metode Naïve Bayesian Classifier ini pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar Hasil penelitian merupakan terciptanya |

| No | Peneliti                             | Judul  | Kesimpulan   |
|----|--------------------------------------|--|--|
|    |                                      |  | sebuah aplikasi yang dapat klasifikasi data penjualan alat teknik di Faedah Teknik dengan menggunakan Metode Naive Bayes.  |
| 9  | (Persediaan et al., 2022)<br>Sinta 4 | Penerapan <i>Naive Bayes</i> Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang | Untuk dapat meramalkan penjualan dan persediaan dengan tepat diperlukan metode penelitian yang baik. Penelitian ini dilakukan dengan menggunakan aplikasi Rapidminer versi 9.9 berdasarkan data penjualan dan persediaan pada tahun 2018 sampai 2020, dengan metode <i>Naive Bayes</i> . Hasil Akurasi pada penelitian ini menunjukkan 89.47 % dengan rincian Hasil Prediksi Meningkat dan Menurun.  |
| 10 | (Tyas et al., 2020)<br>Sinta 2       | Implementasi Algoritma <i>Naive Bayes</i> Dalam Penentuan Rating Buku  | Penelitian ini akan melakukan pengolahan rating buku dengan hasil penelitian ini dibantu oleh tools rapidminer dan Python sebagai alat bantu mengelolah data. Hasil yang diperoleh adalah hasil penentuan rating buku menggunakan metode <i>naive bayes</i> memiliki accuracy 66,98%, precision 74,47% dan recall 62,47% dan hasil analisis ini di dapatkan dari dataset yang ada pada situs <a href="http://www.kaggle.com">www.kaggle.com</a> menunjukkan bahwa mayoritas prediksi rating buku cenderung rendah. |

| No | Peneliti                           | Judul   | Kesimpulan   |
|----|------------------------------------|---|--|
| 11 | (Lianda & Atmaja, 2021)<br>Sinta 4 | Prediksi Data Buku Favorit Menggunakan Metode Naïve Bayes (Studi Kasus: Universitas Dehasen Bengkulu) | Penelitian ini menggunakan menerapkan rumus serta langkah dari metode naïve Bayes, maka metode naïve Bayes dapat diterapkan untuk prediksi buku favorit. Dengan menggunakan data peminjaman buku setiap tahun, maka dengan metode naïve Bayes dapat memprediksi data buku pada tahun selanjutnya.  |
| 12 | (Munawir & Iqbal, 2019)<br>Sinta 4 | Prediksi Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes (Studi Kasus 5 PTS di Banda Aceh)      | Penelitian ini bertujuan untuk data mining dengan menggunakan tools rapidminer untuk mengumpulkan data mahasiswa dari laman aplikasi Feeder dari angkatan 2010-2014 yang diasumsikan bahwa mahasiswa angkatan tersebut sudah dinyatakan lulus pada tahun 2018. Data tersebut di dikumpulkan dari 5 (lima) Perguruan Tinggi Swasta di Kota Banda Aceh.. Hasil penelitian didapatkan bahwa algoritma data mining untuk prediksi kelulusan berdasarkan atribut ketepatan lulus yang dipilih mengungkapkan bahwa tingkat prediksi seragam dengan algoritma yang digunakan yaitu Naïve Bayes, akurasi prediksi sebesar 84%. |

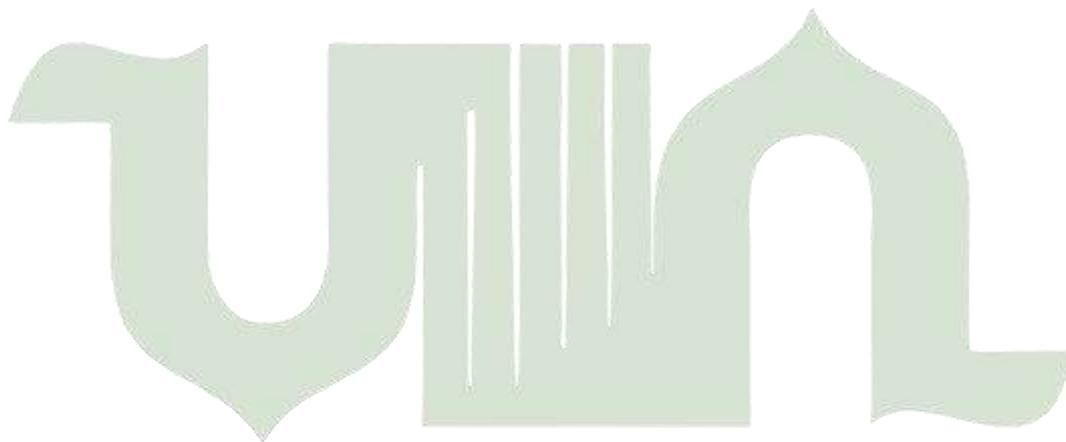
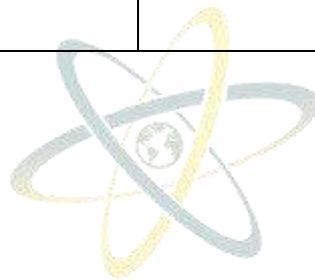
| No | Peneliti                                | Judul   | Kesimpulan  |
|----|---|---|---|
| 13 | (Abdurrahman & Wijaya, 2019)<br>Sinta 4 | Analisis Klasifikasi Kelahiran Caesar Menggunakan Algoritma Naive Bayes   | Pada penelitian ini dilakukan klasifikasi kelahiran Caesar menggunakan algoritma Naive Bayes untuk menentukan apakah kelahiran akan diberikan tindakan operasi Caesar atau melahirkan secara normal. Dataset yang digunakan terdiri dari 80 data ibu hamil dengan 5 atribut, yakni: umur, jumlah tenaga medis, waktu melahirkan, tekanan darah, masalah hati. Hasil penelitian ini diharapkan dapat membantu dunia kesehatan khususnya untuk menentukan proses kelahiran perlu diberi tindakan operasi Caesar ataukah tidak dengan menyediakan model klasifikasi untuk melakukan prediksi secara medis.   |
| 14 | (Ginting & Rakhman, 2019)<br>Sinta 3    | Analisis Data Penjualan Onderdil Sepeda Motor Untuk Estimasi Jumlah Stok Barang Di Perusahaan XYZ Menggunakan Data Mining Dengan Metode Naive Bayes | Tujuan penelitian ini adalah untuk membuat aplikasi yang dapat melakukan analisis data mining untuk memprediksi keputusan restocked onderdil sepeda motor di perusahaan XYZ. Prediksi restocked ini sangat perlu dilakukan agar dapat menghindari masalah-masalah yang sering terjadi ketika proses restocked di perusahaan. Misalnya keputusan yang tidak tepat dalam menentukan onderdil yang harus restocked sehingga terjadinya penumpukan barang di dalam gudang. Masalah lain yang dapat timbul dari penumpukan barang di dalam gudang adalah penurunan kualitas produk dikarenakan terlalu lama disimpan, ruang penyimpanan menjadi penuh, dan terkadang produk yang |

| No | Peneliti                           | Judul   | Kesimpulan  |
|----|------------------------------------|---|---|
|    |                                    |   | laku tidak cukup tersedia di gudang. Dikarenakan hal tersebut dibutuhkan sebuah aplikasi yang dapat menerapkan metode Naïve Bayes Classifier (NBC) untuk memprediksi restocked.   |
| 15 | (Iskandar et al., 2022)<br>Sinta 3 | Model Klasifikasi Naïve Bayes untuk Prediksi Indeks Harga Produsen Naïve Bayes Classification Model for the Producer Price Index Prediction | Penelitian ini untuk menentukan Harga Indeks Prediksi selanjutnya dibutuhkan suatu algoritma Prediksi, yakni Naïve Bayes berdasarkan data dari Triwulan I dan II tahun 2021. Hasil prediksinya yakni Algoritma Naïve Bayes bisa digunakan untuk melakukan prediksi Indeks Harga Produsen (IHP). Prediksi ini dilakukan untuk memberikan gambaran terhadap IHP Triwulan III, mengingat dalam 1 tahun BPS (Badan Pusat Statistika) melakukan update data IHP sampai Triwulan IV dalam 1 tahun. Prediksi yang didapatkan adalah adanya kenaikan untuk Triwulan III dengan nilai maksimal antara 0.961 – 0.980 berdasarkan data dari Triwulan I dan II. |

| No | Peneliti                                | Judul   | Kesimpulan  |
|----|---|---|---|
| 16 | (Riyannah & Fatmawati, 2021)<br>Sinta 4 | Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu | peneliti melakukan teknik pengolahan data dengan observasi, studi pustaka dan kuesioner sedangkan proses pengolahan data dengan menggunakan data mining untuk menentukan status layak atau tidak layak dalam penerima bantuan surat keterangan tidak mampu yaitu dengan menggunakan algoritma Naive Bayes sedangkan penggunaan aplikasi Rapidminer yaitu bertujuan untuk pengujian akurasi dataset. Dataset penerima bantuan surat keterangan tidak mampu yang digunakan dalam penelitian ini yaitu 35 record dengan 8 atribut berupa nama, pekerjaan, usia, status, penghasilan, kendaraan, kepemilikan, dan atap bangunan. Penelitian ini bertujuan untuk memprediksi dan menghasilkan nilai tingkat keakuratan dalam pemberian bantuan surat keterangan tidak mampu kepada masyarakat rukun warga 002 Kelurahan Meruya Selatan dengan menggunakan metode naive bayes. Hasil uji coba menunjukkan bahwa tingkat akurasi sistem sebesar 62.86 %, recall 78.57% dan percision 52.38%. |

| No | Peneliti                            | Judul  | Kesimpulan   |
|----|-------------------------------------|--|--|
| 17 | (Syukri Mustafa et al., 2017)       | Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier  | Penelitian ini difokuskan untuk mengevaluasi kinerja akademik mahasiswa STMIK Dipanegara Makassar pada dua tahun pertama dengan menggunakan teknik data mining algoritma Naive Bayes Classifier (NBC). Hasil yang diperoleh dari penelitian ini menunjukkan bahwa faktor yang paling mempengaruhi penentuan klasifikasi kinerja akademik seorang mahasiswa adalah Indeks Prestasi (IP) pada semester 1,2,3,4 dan jenis kelamin, sehingga faktor tersebut dapat menjadi bahan evaluasi terhadap pihak pengelola STMIK Dipanegara. Pengujian pada beberapa data mahasiswa angkatan 2008-2011 yang diambil secara acak, algoritma NBC menghasilkan nilai akurasi 92,3%. |
| 18 | (Kurniawan & Susanto, 2019) Sinta 3 | Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019 | Penelitian ini melakukan klasifikasi sentimen dengan Tujuan untuk memperoleh analisis dokumen text untuk mendapatkan sentimen positif atau negatif. Metode yang digunakan K-Means untuk melakukan klustering pada data latih dan Naive Bayes classifier untuk mengklasifikasi pada data testing. Hasil dari pembobotan ini berupa sentimen positif dan negatif. Data diambil dari Twitter mengenai pemilu presiden 2019  |

| No | Peneliti | Judul | Kesimpulan  |
|----|----------|-------|---|
|    |          |       | sebanyak 500 data tweet. Dari hasil pengujian 100 dan 150 data uji diperoleh akurasi rata-rata 93.35% dan error rate sebesar 6.66%. |



UNIVERSITAS ISLAM NEGERI  
SUMATERA UTARA MEDAN