

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Data Mining**

*Data mining* adalah proses setengah otomatis yang menggunakan teknik matematika, statistika, *machine learning*, dan kecerdasan buatan untuk mengidentifikasi serta mengekstraksi informasi pengetahuan yang terkait dari berbagai basis data yang sangat luas (Rachman & Handayani, 2021). *Data mining* adalah suatu proses penelusuran pengetahuan baru, pola, dan tren yang diidentifikasi dari jumlah data yang besar yang tersimpan dalam repositori atau tempat penyimpanan. Proses ini menggunakan teknik statistika dan matematika untuk memilah informasi yang signifikan (Rifqo & Wijaya, 2017).

*Data mining* juga dapat diartikan sebagai proses ekstraksi informasi baru dari sejumlah besar data, yang membantu dalam pengambilan keputusan. Istilah "data mining" sering juga disebut sebagai "*knowledge discovery*". Salah satu teknik yang dikembangkan dalam data mining adalah cara menelusuri data yang ada untuk membangun model, dan kemudian menggunakan model tersebut untuk mengenali pola data lain yang tidak terdapat dalam basis data yang tersimpan. Teknik ini juga dapat dimanfaatkan untuk kebutuhan prediksi. Dalam konteks data mining, pengelompokan data juga dapat dilakukan dengan tujuan menemukan pola data yang universal. Pentingnya mendeteksi anomali data transaksi juga menjadi faktor kunci untuk menentukan tindak lanjut yang tepat. Semua upaya tersebut diarahkan untuk mendukung kegiatan operasional perusahaan, dengan harapan bahwa tujuan akhir perusahaan dapat tercapai.

##### **2.1.1. Pekerjaan dalam data mining**

Menurut Prasetyo (2012) tugas yang berkaitan dengan data mining dapat disusun menjadi empat kategori utama, yakni pembuatan model prediksi (*prediction modeling*), analisis kelompok (*cluster analysis*), analisis asosiasi (*association analysis*), dan deteksi anomali (*anomaly detection*).

##### **1. Model Prediksi**

Pembuatan model prediksi melibatkan proses menciptakan model yang mampu memetakan setiap himpunan variabel ke targetnya, dan setelahnya

menggunakan model tersebut untuk memberikan nilai target pada himpunan data baru yang diperoleh. Terdapat dua jenis utama model prediksi, yakni klasifikasi dan regresi. *Klasifikasi* digunakan untuk variabel target yang bersifat diskret, sementara regresi digunakan untuk variabel target yang bersifat kontinu.

## 2. Analisis Kelompok

Analisis kelompok melibatkan proses pengelompokan data ke dalam sejumlah kelompok (*cluster*) berdasarkan karakteristik yang serupa di setiap kelompok. Data dengan kesamaan yang signifikan akan dikelompokkan bersama, sementara data yang tidak memiliki kesamaan yang cukup akan dipisahkan dalam kelompok yang berbeda. Proses ini memungkinkan identifikasi pola atau struktur yang mungkin tidak terlihat secara langsung dan membantu dalam pemahaman hubungan antar data.

## 3. Analisis Asosiasi

Analisis asosiasi bertujuan untuk menemukan pola yang mencerminkan kekuatan hubungan antara fitur-fitur dalam data. Pola yang diidentifikasi sering menggambarkan aturan implikasi atau subset fitur. Fokus utamanya adalah mengidentifikasi pola yang menarik dengan efisien, sehingga dapat memberikan wawasan yang berharga tentang interaksi fitur-fitur dalam dataset.

## 4. Deteksi anomaly

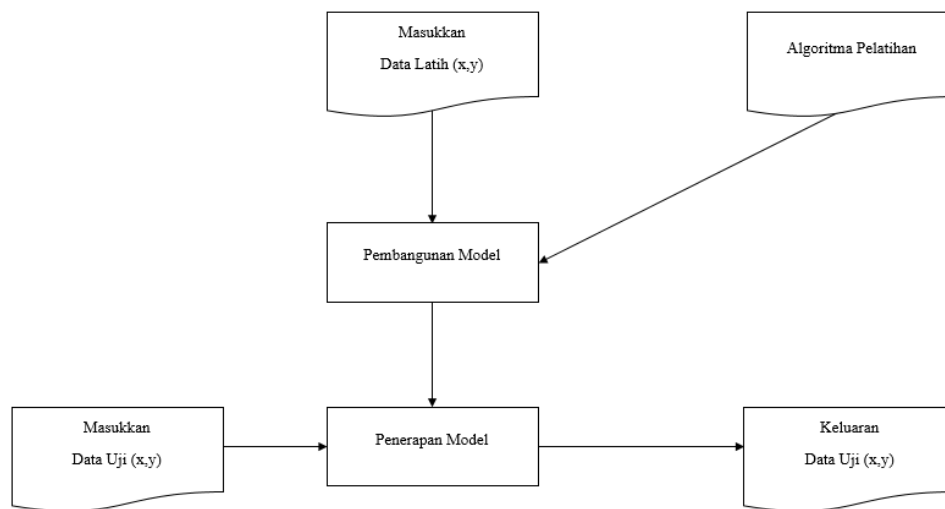
Pekerjaan deteksi anomaly melibatkan pengamatan data untuk mengidentifikasi data yang secara signifikan berbeda dari data lainnya. Data yang menunjukkan perbedaan mencolok dari karakteristik umum data lain disebut sebagai outlier. Algoritma deteksi anomaly yang efektif diharapkan memiliki tingkat deteksi yang tinggi dan tingkat kesalahan yang rendah.

### 2.2. Klasifikasi

*Klasifikasi* atau kategorisasi adalah proses untuk menemukan model atau fungsi yang dapat menjelaskan atau membedakan konsep atau kelas data. Tujuan utamanya adalah untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui, berdasarkan pada pola atau karakteristik yang teridentifikasi dalam data

yang telah dianalisis (Setiawan, et al., 2015). *Klasifikasi* berasal dari bahasa Latin, yaitu "*Classic*," yang artinya proses pengelompokan atau pengumpulan benda yang sejenis, lalu memisahkan benda yang tidak sejenis (Ainum, et al., 2022). Dalam proses *klasifikasi*, pembentukan pengklasifikasi terjadi dengan merujuk pada kumpulan data latih yang sudah memiliki kelas atau label yang telah ditentukan sebelumnya. Evaluasi kinerja pengklasifikasi ini biasanya dilakukan dengan menggunakan metrik akurasi. Akurasi mengukur sejauh mana pengklasifikasi dapat dengan tepat mengidentifikasi kelas dari data yang belum pernah dilihat sebelumnya. Semakin tinggi tingkat akurasi, semakin baik kinerja pengklasifikasi dalam memprediksi kelas objek yang tidak dikenal (Syarli & Muin, 2016).

*Klasifikasi* dapat dijelaskan sebagai tugas yang melibatkan pelatihan atau pembelajaran pada fungsi target  $f$ , yang memetakan setiap himpunan atribut (fitur)  $x$  ke salah satu dari berbagai label kelas  $y$  yang tersedia. Proses pelatihan tersebut menghasilkan suatu model yang kemudian disimpan sebagai memori. Model dalam konteks *klasifikasi* dapat dibandingkan dengan kotak hitam, di mana model tersebut menerima input, melakukan proses pemikiran terhadap input tersebut, dan menghasilkan output sebagai hasil dari pemikirannya. Kerangka kerja *klasifikasi* diilustrasikan pada (Gambar 2.1). Dalam ilustrasi tersebut, tersedia sejumlah data latih  $(x, y)$  yang akan digunakan sebagai basis pembangunan model. Model yang dibuat dari data latih tersebut kemudian digunakan untuk meramalkan kelas dari data uji  $(x, y)$  sehingga dapat diketahui kelas sebenarnya dari kelas  $y$  (Prasetyo, 2012).



Gambar 2. 1 Proses Pekerjaan Klasifikasi

Model yang telah dikembangkan selama tahap pelatihan dapat digunakan untuk memprediksi label kelas pada data baru yang belum dikenal. Dalam proses pembangunan model selama tahap pelatihan, diperlukan suatu algoritma pelatihan yang disebut *learning algorithm*. Algoritma ini dirancang untuk membaca, melatih, dan belajar dari data latihan dengan tujuan memetakan setiap vektor masukan ke label kelas yang sesuai. Pada akhir proses pelatihan, model mampu memetakan semua vektor data uji ke label kelas dengan benar, dan di antara algoritma yang digunakan adalah *Artificial Neural Network (ANN)*, *Support Vector Machine (SVM)*, *Decision Tree*, *Naïve Bayes*, dan sejenisnya (Prasetyo, 2012).

### 2.2.1. Pengukuran Kinerja Klasifikasi

Sistem *klasifikasi* diinginkan mampu mengklasifikasikan semua set data dengan tingkat akurasi, meskipun diketahui bahwa mencapai tingkat kebenaran 100% tidak mungkin. Oleh karena itu, penting untuk melakukan pengukuran kinerja pada sistem *klasifikasi*. Evaluasi kinerja *klasifikasi* secara umum dilakukan dengan menggunakan matriks konfusi (*confusion matrix*) (Prasetyo, 2012).

Matriks konfusi adalah tabel pencatat hasil dari proses *klasifikasi*. Contoh matriks konfusi pada (Tabel 2.1) digunakan dalam klasifikasi masalah biner (dua kelas), yang hanya memiliki dua kelas, yaitu kelas 0 dan 1. Setiap sel  $f_{ij}$  dalam matriks mencatat jumlah rekord atau data dari kelas  $i$  yang hasil prediksinya masuk



ke kelas  $j$ . Sebagai ilustrasi, sel  $f_{11}$  mengindikasikan jumlah data dalam kelas 1 yang secara benar diprediksi masuk ke kelas 1, sementara  $f_{10}$  mencerminkan jumlah data dalam kelas 1 yang diprediksi masuk ke kelas 0 (Prasetyo, 2012).

Tabel 2. 1 Matrik Konfusi untuk Klasifikasi dua kelas

| $f_{ij}$           |           | Kelas hasil prediksi ( $j$ ) |           |
|--------------------|-----------|------------------------------|-----------|
|                    |           | Kelas = 1                    | Kelas = 0 |
| Kelas asli ( $i$ ) | Kelas = 1 | $f_{11}$                     | $f_{10}$  |
|                    | Kelas = 0 | $f_{01}$                     | $f_{00}$  |

Berdasarkan isi matriks konfusi, kita dapat mengetahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu  $(f_{11} + f_{00})$ , dan data yang diklasifikasikan secara salah, yaitu  $(f_{10} + f_{01})$ . Kuantitas dalam matriks konfusi dapat disederhanakan menjadi dua nilai, yakni akurasi dan tingkat kesalahan. Dengan mengetahui jumlah data yang diklasifikasikan dengan benar, akurasi hasil prediksi dapat diidentifikasi, dan dengan mengetahui jumlah data yang diklasifikasikan secara salah, tingkat kesalahan prediksi dapat dihitung. Kedua kuantitas ini berperan sebagai metrik kinerja dalam proses evaluasi klasifikasi.

Untuk menghitung akurasi digunakan formula

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi benar}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2.1)$$

Untuk menghitung laju error (kesalahan prediksi) digunakan formula

$$\text{Laju error} = \frac{\text{Jumlah data yang diprediksi salah}}{\text{jumlah prediksi yang dilakukan}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2.2)$$

Semua algoritma *klasifikasi* bertujuan untuk mengembangkan model dengan tingkat akurasi tinggi (tingkat kesalahan yang rendah). Umumnya, model yang dibuat dapat memprediksi dengan benar pada seluruh data yang digunakan sebagai data latihnya. Namun, ketika model diuji dengan data uji, barulah kinerja sebenarnya dari suatu algoritma klasifikasi dapat dievaluasi (Prasetyo, 2012).

### 2.3. Naïve Bayes Gaussian

Algoritma *Naïve Bayes* merupakan metode *klasifikasi* yang mempresentasikan setiap kelas objek berdasarkan kesimpulan atau ringkasan probabilitas. Algoritma ini bertujuan untuk menemukan kemungkinan besar kelas yang sesuai untuk setiap objek uji berdasarkan atribut-atribut atau variabel yang memiliki nilai yang diketahui (Riany & Testiana, 2023). *Naïve Bayes* merupakan salah satu metode *klasifikasi* yang sederhana namun sering digunakan karena kemudahan penerapannya dan memberikan hasil yang baik dalam berbagai kasus. Meskipun demikian, kelemahan metode ini terletak pada asumsinya, yaitu bahwa atribut atau variabel yang digunakan dalam *klasifikasi* dianggap saling bebas. Asumsi ini dapat mengakibatkan kurangnya akurasi, terutama ketika kemerdekaan antar atribut tidak sepenuhnya terpenuhi dalam data yang digunakan (Siregar, et al., 2020). *Naïve Bayes* menurut definisi lain adalah metode klasifikasi yang menerapkan konsep probabilitas dan statistik yang diajukan oleh ilmuwan Inggris Thomas Bayes. Konsep ini melibatkan prediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. *Naïve Bayes* bergantung pada asumsi sederhana bahwa nilai atribut bersyarat saling bebas jika nilai output sudah diketahui. Dengan kata lain, ketika nilai output sudah diketahui, probabilitas mengamati nilai-nilai atribut tersebut secara bersamaan dianggap sebagai hasil perkalian dari probabilitas masing-masing nilai atribut tersebut secara individual (Rifai, et al., 2019). *Naïve Bayes* merupakan metode yang efektif dan efisien, sehingga sering kali digunakan untuk melakukan klasifikasi atau memberi prediksi secara real-time (Nadira, et al., 2023).

Bayes adalah metode prediksi yang sederhana berbasis probabilitas, didasarkan pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi yang sangat kuat (naif). Dengan kata lain dalam *Naïve Bayes*, model yang digunakan adalah "model fitur independen". Dalam konteks *Naïve Bayes*, independensi yang kuat pada fitur berarti bahwa keberadaan atau ketidakhadiran suatu fitur pada suatu data tidak terkait dengan keberadaan atau ketidakhadiran fitur lain dalam data yang sama. Dasar dari prediksi Bayes adalah Teorema Bayes, yang dirumuskan dengan formula umum berikut ini:

$$P(H | E) = \frac{P(E | H) \times P(H)}{P(E)} \quad (2.3)$$

Penjelasan dari formula tersebut adalah sebagai berikut :

$P(H | E)$  = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

$P(E | H)$  = Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H.

$P(H)$  = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$  = Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Ide dasar dari aturan Bayes adalah bahwa hasil dari suatu hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan beberapa bukti yang diamati (E). Beberapa aspek penting dari aturan Bayes ini mencakup :

1. Sebuah probabilitas awal/priori H atau  $P(H)$  adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau  $P(H | E)$  adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Menurut Mujahidin (2022) Algoritma *Naïve Bayes Gaussian* merupakan salah satu algoritma berbasis nilai kontinu dengan konsep probabilitas yang dapat digunakan dalam menentukan kelas dari dokumen maupun pengolahan data dalam jumlah yang besar dan nilai akurasi yang tinggi. Performa nilai tingkat akurasi yang dihasilkan oleh algoritma *Naïve Bayes Gaussian* bergantung pada data latih yang dipilih, nilai akurasi akan maksimal jika data latih mencakup sebagian besar atau bahkan mencakup keseluruhan dari data yang dimiliki. Dikarenakan data dalam penelitian ini berbentuk kontinu maka akan digunakan metode *Distribusi Gaussian* yang sering kali digunakan untuk tipe data numerik (Amien, et al., 2023).

### 2.3.1. Naïve Bayes untuk klasifikasi

Kaitan antara *Naïve Bayes* dengan klasifikasi, korelasi hipotesis, dan bukti *klasifikasi* dapat diuraikan sebagai berikut: Dalam teorema Bayes, hipotesis mencerminkan label kelas yang menjadi fokus dalam tugas klasifikasi, sedangkan bukti mengacu pada fitur-fitur yang menjadi input dalam model *klasifikasi*. Dalam

notasi matematis, jika  $X$  adalah vektor input yang berisi fitur-fitur, dan  $Y$  adalah label kelas, *Naïve Bayes* dapat diartikan sebagai  $P(Y|X)$ . Representasi ini menunjukkan probabilitas bahwa label kelas  $Y$  ditemukan setelah observasi fitur-fitur  $X$ , yang dikenal sebagai probabilitas akhir atau posterior untuk  $Y$ . Di sisi lain,  $P(Y)$  mengacu pada probabilitas awal atau prior untuk  $Y$ .

Saat proses pelatihan, pembelajaran melibatkan perhitungan probabilitas akhir  $P(Y|X)$  pada model untuk setiap pasangan kombinasi  $X$  dan  $Y$ , menggunakan informasi yang diperoleh dari data latih. Dengan konstruksi model ini, data uji  $X$  dapat diklasifikasikan dengan menentukan nilai  $Y$  yang mengoptimalkan probabilitas  $P(Y'|X')$  yang dihasilkan.

Formulasi *Naïve Bayes* untuk klasifikasi dapat dirumuskan sebagai berikut :

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^q P(X_i | Y)}{P(X)} \quad (2.4)$$

$P(Y|X)$  merupakan probabilitas data dengan vektor  $X$  pada kelas  $Y$ .  $P(Y)$  adalah probabilitas awal kelas  $Y$ .  $\prod_{i=1}^q P(X_i | Y)$  adalah probabilitas independen kelas  $Y$  terhadap semua fitur dalam vektor  $X$ . Nilai  $P(X)$  tetap konstan, sehingga dalam perhitungan prediksi, kita hanya perlu menghitung bagian  $P(Y) \prod_{i=1}^q P(X_i | Y)$  dan memilih yang memiliki nilai terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara itu, probabilitas independen  $\prod_{i=1}^q P(X_i | Y)$  mencerminkan pengaruh semua fitur dari data terhadap setiap kelas  $Y$ , yang dinotasikan dengan

$$P(X|Y=y) = \prod_{i=1}^q P(X_i | Y=y) \quad (2.5)$$

Secara umum, penghitungan metode Bayes dapat dengan mudah dilakukan pada fitur yang bersifat kategoris, contohnya pada klasifikasi hewan dengan fitur "penutup kulit" yang memiliki nilai {bulu, rambut, cangkang}, atau pada fitur "jenis kelamin" dengan nilai {pria, wanita}. Namun, untuk fitur yang bersifat numerik (kontinu), diperlukan penanganan khusus sebelum dimasukkan ke dalam model *Naïve Bayes*. Caranya adalah:



1. Menyusun fitur kontinu dengan melibatkan diskretisasi, menggantikan nilai setiap fitur kontinu dengan interval diskret, dan menggunakan pendekatan ini untuk mentransformasikan fitur kontinu menjadi fitur ordinal.
2. Mengasumsikan bentuk distribusi probabilitas tertentu untuk fitur kontinu dan melakukan estimasi parameter distribusi dengan menggunakan data pelatihan. *Distribusi Gaussian* sering dipilih untuk menggambarkan probabilitas bersyarat dari fitur kontinu dalam suatu kelas  $P(X_i | Y)$ , di mana *distribusi Gaussian* ditandai oleh dua parameter utama: mean ( $\mu$ ) dan varian ( $\sigma^2$ ). Untuk setiap  $y_j$ , probabilitas bersyarat kelas  $y_j$  untuk fitur  $X_i$  adalah:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \left( -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \quad (2.6)$$

Parameter  $\mu_{ij}$  dapat diperoleh dari mean sampel  $X_i$  ( $\bar{x}$ ) dari seluruh data latih yang termasuk dalam kelas  $y_j$ , sedangkan  $\sigma_{ij}^2$  dapat diperkirakan dari varian sampel ( $s^2$ ) dari data pelatihan.

#### 2.4. KNN (K-Nearest Neighbor)

Algoritma *K-Nearest Neighbor* (KNN) adalah merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek data baru berdasarkan (K) tetangga terdekatnya (Mustafa & Simpen, 2019). Algoritma KNN adalah algoritma ketetanggaan terdekat, yang dihitung dari nilai jarak pada pengujian data testing dengan data training, dari nilai terkecil ketetanggaan terdekat (Amien, et al., 2023). Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance, Jarak Euclidean paling sering digunakan dalam menghitung jarak dikarenakan sangat cocok untuk menggunakan jarak terdekat antara dua data (Amilia & Prasetyo, 2020). Perhitungan jarak pada data berdimensi tinggi lebih baik menggunakan Euclidean, Manhattan dan Minkowski sedangkan untuk data berdimensi rendah lebih baik menggunakan Chebychev, selain itu dari perhitungan jarak Euclidean, Manhattan dan Minkowski hanya Manhattan yang tidak menggunakan perhitungan kuadrat sehingga hanya pada  $K=1$  saja mendapatkan hasil terbaik jika dibandingkan dengan Euclidean dan Minkowski yang memiliki akurasi terbaik di beberapa

jumlah  $k$  (Mandita & Pratama, 2024). Pemanfaatan algoritma *K-Nearest Neighbor* diseleksi gara-gara kemampuannya dalam memutuskan group data dengan mengestimasi jarak orang sebelah terdekat (Akmal, et al., 2023).

Adapun tahapan dari algoritme KNN dapat dijelaskan sebagai berikut :

1. Persiapan data training dan data testing
2. Melakukan normalisasi data menggunakan min – max, apabila ada nilai yang memiliki satuan ukuran yang berbeda. Berikut adalah rumusnya

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}} \quad (2.7)$$

3. Menentukan nilai  $K$
4. Menghitung jarak data testing ke setiap data training dengan rumus Euclidean Distance. Berikut adalah rumusnya :

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.8)$$

## 2.5. Python

*Python* merupakan bahasa pemrograman tingkat tinggi yang dirancang oleh *Guido Van Rossum* dan pertama kali diperkenalkan pada tahun 1991. Bahasa ini telah meraih popularitas yang besar dalam beberapa tahun terakhir. *Python* dikenal karena sintaksisnya yang mudah dipahami dan memiliki berbagai pustaka yang lengkap. Salah satu keunggulan utamanya adalah dukungan yang kuat dari komunitas, karena *Python* bersifat open source (Alfarizi, et al., 2023). *Python* dapat dianggap sebagai bahasa pemrograman yang menggabungkan kemampuan dan kapabilitas dengan sintaksis yang jelas. Bahasa ini memiliki pustaka standar yang cukup banyak dan komprehensif. Selain itu, *Python* sering digunakan untuk mengembangkan aplikasi berbasis web (Semendawai, et al., n.d.).

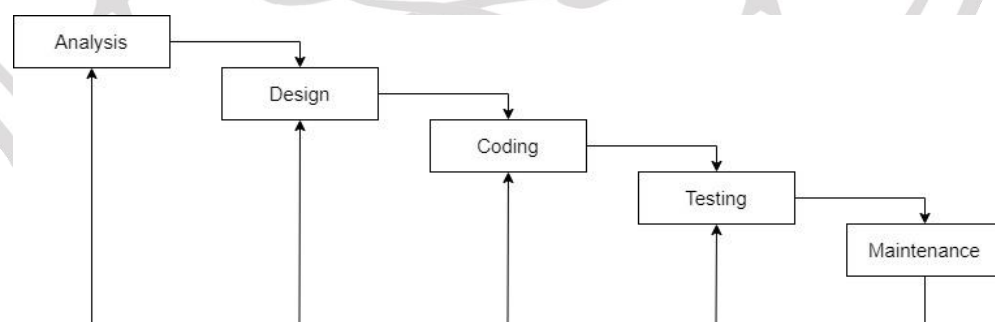
## 2.6. Streamlit

*Streamlit* adalah kerangka kerja (*framework*) sumber terbuka (*open-source*) dari *Python* yang memungkinkan pembuatan aplikasi web dengan menggunakan bahasa pemrograman *Python*, terutama dalam konteks menerapkan model dari bidang *machine learning* atau *data science* (Ferdyandi, et al., 2022). Adrian, pencipta *Streamlit*, menjelaskan bahwa *Streamlit* adalah suatu kerangka aplikasi

(*apps framework*) yang digunakan oleh insinyur *machine learning* dan ilmuwan data, khususnya untuk membuat alat-alat internal atau aplikasi web. Dengan *Streamlit*, pengguna dapat dengan mudah membangun aplikasi web tanpa harus terjebak dalam bahasa pemrograman yang kompleks, memberikan kemudahan penggunaan bagi mereka (Nofyantoro, et al., 2022). Salah satu keunggulan utama dari *Streamlit* adalah bahwa pengembang tidak perlu memikirkan tata letak situs web dengan menggunakan CSS, HTML, dan JavaScript. *Framework Streamlit* telah menyediakan fitur-fitur tersebut melalui fungsi-fungsi terintegrasi, memungkinkan pengembang untuk fokus pada logika aplikasi dan analisis data tanpa harus menyita waktu untuk menangani aspek tampilan dan styling secara terpisah. (Putranto, et al., 2023).

## 2.7. Metode Waterfall

Metode *Waterfall* merupakan metode yang menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau terurut (Kurniawati & Badrul, 2021). Metode air terjun atau yang sering disebut metode *waterfall* sering dinamakan siklus hidup klasik (*classic life cycle*), yang mana pada hal ini menggambarkan pendekatan yang sistematis dan juga berurutan pada pengembangan perangkat lunak (Triayudi & Rodhi, 2018). Menurut Pinatih & Hidayatul (2022) Model *Waterfall* (model air terjun) diaplikasikan pada penelitian ini yang bersifat terstruktur dalam membangun sebuah perangkat lunak. Tahapan metode *Waterfall* ini harus diselesaikan satu persatu yang berarti tidak bisa meloncat ke tahap berikutnya. Berikut adalah gambarnya :



Gambar 2. 2 Alur Metode Waterfall

Adapun keterangan dari gambar diatas adalah :

- 1) Analysis : Pada tahap analisis ini, focus utama adalah memahami dan mengidentifikasi potensi penyakit diabetes mellitus tipe II pada pasien. Yang mana ini mencakup penelitian terhadap gejala, factor dan data medis yang relevan.
- 2) Desain : Langkah selanjutnya adalah merancang kerangka kerja untuk klasifikasi potensi penyakit diabetes mellitus tipe II. Ini melibatkan algoritme Naïve Bayes Gaussian untuk mengolah informasi medis.
- 3) Coding : Langkah selanjutnya adalah mengimplementasikan algoritme Naïve Bayes Gaussian ke dalam kode program, proses pengkodean ini mencakup konsep desain menjadi website yang dapat dieksekusi.
- 4) Testing : Langkah selanjutnya adalah melakukan pengujian yang berguna untuk memastikan bahwa website sudah berfungsi sesuai dengan yang diinginkan, pengujian ini melibatkan uji coba dengan data medis untuk mengevaluasi kinerja algoritme Naïve Bayes Gaussian
- 5) Maintenance : Langkah terakhir adalah pemeliharaan secara rutin.

## 2.8. Penelitian Terkait

Tabel 2. 2 Tabel Penelitian Terkait

| No | Judul Penelitian  | Ringkasan Penelitian   |
|----|---|--|
| 1. | Prediksi Penyakit Diabetes Menggunakan Machine Learning Menggunakan Algoritma Naïve Bayes | <ol style="list-style-type: none"> <li>1. Tujuan Penelitian = Untuk membuat model prediksi menggunakan Algoritma Naïve Bayes yang menghasilkan klasifikasi dan prediksi diagnosa penyakit diabetes mellitus yang dilakukan dengan menggunakan Rapid Miner agar pencegahan terhadap penyakit diabetes dapat dilakukan dengan segera mungkin.</li> <li>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diabetes mellitus dengan atribut Pregnancies, Glucose, Blood Pressure, BMI, Diabetes Pedigree Function dan Age dengan total 1035 data yang didapat dari Apotik Harapan Mulya.</li> <li>3. Perancangan Sistem =               <ol style="list-style-type: none"> <li>a) Data Mining</li> <li>b) Algoritma Naïve Bayes</li> <li>c) Rapid Miner</li> </ol> </li> </ol> |



|    |   |   |
|----|---|---|
|    |   | <p>4. Teknik Pengumpulan Data = Teknik Pengumpulan data yang digunakan adalah kuantitatif yang berasal dari Apotik Harapan Mulya.</p> <p>5. Hasil Penelitian = Hasil penelitian yang didapatkan adalah metode Naïve Bayes dapat digunakan dalam mengklasifikasi dan memprediksi diagnosa penyakit diabetes mellitus berdasarkan atribut Pregnancies, Glucose, Blood Pressure, BMI, Diabetes Pedigree Function dan Age menggunakan RapidMiner. Dengan tingkat akurasi sebesar 95,94%.</p> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini menghasilkan nilai akurasi yang sangat tinggi yaitu 95,94%</li> <li>Terdapat rumus perhitungan akurasi.</li> </ol> <p>7. Kelemahan Penelitian =</p> <ol style="list-style-type: none"> <li>Peneliti tidak membandingkannya dengan metode lain.</li> <li>Peneliti tidak menunjukkan perhitungan secara manual</li> <li>Tidak ada rumus Naïve Bayes</li> </ol> <p>8. Kesimpulan = Tingkat akurasi klasifikasi diagnosa penyakit diabetes mellitus menggunakan algoritma Naïve Bayes sebesar 95,94% dapat digunakan untuk membuat GUI atau aplikasi yang digunakan dokter dan pasien diabetes dalam mendiagnosa penyakit diabetes.</p> |
| 2. | Diabetes Prediction using supervised machine learning | <p>1. Tujuan Penelitian = Penggunaan machine learning untuk mencegah kematian dengan menggunakan model kecerdasan buatan yang dapat memprediksi penyakit diabetes dan metode yang digunakan adalah perbandingan antara KNN dan Naïve Bayes, untuk melihat algoritma mana yang paling sesuai untuk prediksi penyakit diabetes.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diabetes mellitus dengan atribut Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function dan Age dengan total 768 data</p>  |

|    |   |   |
|----|---|---|
|    |   | <p>yang didapat dari Pima Indian diabetes database.</p> <p>3. Perancangan Sistem =</p> <ol style="list-style-type: none"> <li>Data Mining.</li> <li>Machine Learning.</li> <li>Algoritma Naïve Bayes.</li> <li>Algoritma KNN.</li> <li>Google Collaboratory untuk mereplikasi Jupyter Notebook di cloud.</li> </ol> <p>4. Teknik Pengumpulan Data = Teknik pengumpulan data yang digunakan adalah kuantitatif yang berasal dari Pima Indian diabetes database.</p> <p>5. Hasil Penelitian = Hasil penelitian ini adalah akurasi algoritma Naïve Bayes lebih unggul dari algoritma KNN dengan menggunakan split data 80% untuk data training dan 20% untuk data testing, didapatkan nilai akurasi sebesar 78,57% untuk algoritma Naïve Bayes dan akurasi sebesar 77,92% untuk algoritma KNN.</p> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini berhasil menemukan metode terbaik dari dua metode yang digunakan.</li> <li>Terdapat rumus Naïve Bayes.</li> <li>Terdapat contoh persamaan dari k-fold cross validation.</li> <li>Evaluasi model menggunakan Confusion Matrix untuk mencari nilai akurasi, precision dan recall.</li> </ol> <p>7. Kelemahan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini tidak menunjukkan perhitungan secara manual.</li> <li>Tidak ada rumus KNN.</li> </ol> <p>8. Kesimpulan = Algoritma Naïve Bayes lebih baik daripada algoritma KNN untuk memprediksi penyakit diabetes menggunakan dataset suku Indian Pima.</p> |
| 3. | Klasifikasi Penyakit Diabetes Menggunakan Algoritma Naïve Bayes | <p>1. Tujuan Penelitian = Tujuan penelitian ini adalah mencoba menerapkan suatu metode klasifikasi untuk memprediksi apakah seseorang terkena diabetes atau tidak.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diabetes mellitus dengan atribut</p>  |

|       |   |  |
|-------|---|--|
|       |   | <p>Cholesterol, Glucose, HDL Cholesterol, Age, Gender, Weight, Systolic BP, Diastolic BP, dan Diabetes dengan total 390 data yang didapat dari situs web Kaggle Predict diabetes based on diagnostic measure.</p> <p>3. Perancangan Sistem =</p> <ol style="list-style-type: none"> <li>Data Mining.</li> <li>Algoritma Naïve Bayes.</li> </ol> <p>4. Teknik Pengumpulan Data = Teknik pengumpulan data yang digunakan adalah kuantitatif yang berasal dari situs web Kaggle Predict diabetes based on diagnostic measure.</p> <p>5. Hasil Penelitian = Hasil penelitian ini menunjukkan bahwa perempuan yang menderita penyakit diabetes lebih banyak daripada laki-laki. Dengan nilai tingkat akurasi sebesar 91,6%.</p> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Menunjukkan hasil nilai akurasi dari data train dan data test, dengan tingkat akurasi 92,3% untuk data train dan 91,6% untuk data test.</li> <li>Terdapat rumus untuk formula Naïve Bayes.</li> <li>Terdapat rumus untuk menghitung nilai akurasi.</li> </ol> <p>7. Kelemahan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini tidak menunjukkan perhitungan secara manual.</li> <li>Penelitian ini tidak menunjukkan software yang digunakan untuk menganalisis data.</li> </ol> <p>8. Kesimpulan = Klasifikasi pemodelan Algoritma Naïve Bayes terhadap dataset diabetes sudah bagus nilai akurasi nya.</p> |
| 1uil; | <p>Penerapan Algoritma Naïve Byes Untuk Klasifikasi Penyakit Diabetes Mellitus di Rumah Sakit Aisyiah</p> | <p>1. Tujuan Penelitian = Tujuan Penelitian ini adalah untuk melakukan klasifikasi dan evaluasi model algoritma Naïve Bayes untuk mengetahui akurasi algoritma Naïve Bayes dalam mengklasifikasikan penyakit Diabetes Mellitus.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diabetes mellitus dengan atribut Age,</p>   |

|  |  |  |
|--|--|--|
|  |  | <p>Polyuria, Polydipsia, Sudden weight loss, Weakness, Polyphagia, Genital thrush, Visual blurring, Itching, Irritability, Delayed healing, Partial paresis, Muscle stiffness, Alopecia, Obesity dan Class dengan total 520 data yang didapat dari Rumah Sakit Aisyiah Kudus.</p> <p>3. Perancangan Sistem =</p> <ol style="list-style-type: none"> <li>Data Mining.</li> <li>Algoritma Naïve Bayes.</li> <li>Rapid Miner.</li> </ol> <p>4. Teknik Pengumpulan Data = Teknik pengumpulan data yang digunakan adalah kuantitatif yang berasal dari dataset Rumah Sakit Aisyiah Kudus.</p> <p>5. Hasil Penelitian = Hasil penelitian ini menunjukkan beberapa nilai yang didapat dari pemodelan menggunakan algoritma Naïve Bayes adalah sebagai berikut :</p> <ol style="list-style-type: none"> <li>Nilai Precision sebesar 82,35%</li> <li>Nilai Recall sebesar 87,50%</li> <li>Nilai Akurasi sebesar 90,20%</li> <li>Kurva ROC (Receiver Operating Characteristic) menunjukkan algoritma Naïve Bayes memiliki nilai AUC (Area Under the Curve) sebesar 0,9555</li> </ol> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Terdapat rumus untuk menghitung nilai akurasi, recall dan precision.</li> <li>Terdapat rumus teorema Naïve Bayes.</li> <li>Pada evaluasi hasil model menampilkan kurva ROC dari hasil pengklasifikasian dengan Naïve Bayes.</li> </ol> <p>7. Kelemahan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini tidak menunjukkan perhitungan secara manual.</li> </ol> <p>8. Kesimpulan =</p> <ol style="list-style-type: none"> <li>Hasil evaluasi klasifikasi ini menghasilkan nilai akurasi sebesar 90,20% yang berarti klasifikasi pemodelan algoritma Naïve Bayes terhadap dataset diabetes sudah sangat bagus.</li> <li>Hasil kurva ROC menunjukkan bahwa algoritma Naïve Bayes memiliki nilai AUC sebesar 0,955 yang berarti</li> </ol> |
|--|--|--|



|    |  | Axcellent Classification (Sangat Bagus).  |
|----|--|---|
| 5. | Diabetes Mellitus Pada Perempuan Usia Reproduksi Di Indonesia Tahun 2007 | <p>1. Tujuan Penelitian = Untuk mengetahui gambaran Diabetes Mellitus dan faktor resikonya pada perempuan usia reproduksi di Indonesia tahun 2007.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diabetes mellitus dengan atribut Obesitas, Aktivitas fisik, Merokok, Konsumsi lemak, Konsumsi buah dan sayur dan Umur yang didapat dari data Biomedis Riskesdas 2007.</p> <p>3. Perancangan Sistem =</p> <p>4. Teknik Pengumpulan Data = Teknik pengumpulan data yang digunakan adalah kuantitatif dengan desain studi potong lintang (cross sectional). Data yang digunakan adalah data Biomedis Riskesdes 2007, pengambilan data Riskesdes 2007 dilaksanakan oleh Balitbangkes (Badan Penelitian dan Pengembangan Kesehatan) Kementerian Indonesia.</p> <p>5. Hasil Penelitian = Hasil dari penelitian ini adalah terdapat 3,6% penderita Diabetes Mellitus pada perempuan usia reproduksi, presentasi faktor risiko diabetes pada perempuan usia reproduksi yaitu obesitas sebesar 29,6% , kurang aktivitas fisik sebesar 52,5% , merokok sebesar 26,9% , sering mengkonsumsi lemak sebesar 16,7% , kurang konsumsi buah dan sayur sebesar 97,3% dan rata-rata usia perempuan reproduksi adalah 32 tahun.</p> <p>6. Kelebihan Penelitian =</p> <p>a) Menunjukkan persentase Diabetes Mellitus pada wanita sebesar 53,2% dari laki-laki yang sebesar 46,8%.</p> <p>b) Menunjukkan persentase faktor-faktor resiko terkena Diabetes Mellitus pada perempuan yang meliputi Obesitas, Aktivitas fisik, Merokok, Konsumsi lemak, Konsumsi buah dan sayur dan Umur.</p> <p>7. Kelemahan Penelitian =</p> |

|    |   |  |
|----|---|--|
|    |   | <p>a) Jumlah data yang digunakan untuk analisis kurang transparan atau jelas.</p> <p>b) Penelitian ini tidak menunjukkan software yang digunakan untuk menganalisis data.</p> <p>8. Kesimpulan = Persentase Diabetes Mellitus pada perempuan usia reproduksi adalah sebesar 3,6% oleh karena itu perlu diwaspadai untuk penurunan dan pencegahan kejadian Diabetes Mellitus pada perempuan usia reproduksi melalui pencegahan pada faktor resikonya.</p>   |
| 6. | <p>Analisis Algoritma Gaussian Naïve Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung</p> | <p>1. Tujuan Penelitian = Untuk meneliti dengan judul Analisis Algoritma Gaussian Naïve Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data pasien penyakit gagal jantung pada tahun 2020 yang terdiri dari 1000 data dan terbagi atas 508 pasien hidup dan 492 pasien dinyatakan meninggal.</p> <p>3. Perancangan Sistem =</p> <ol style="list-style-type: none"> <li>Data Mining</li> <li>Klasifikasi</li> <li>Naïve Bayes</li> <li>Gaussian Naïve Bayes</li> <li>Confusion Matrix</li> <li>Cross Validation</li> <li>Jupyter Notebook</li> </ol> <p>4. Teknik Pengumpulan Data = Pengumpulan data yang dilakukan dalam penelitian ini diperoleh melalui Kaggle yang terdiri dari data pasien penyakit gagal jantung pada tahun 2020.</p> <p>5. Hasil Penelitian = Hasil dari penelitian ini adalah</p> <ol style="list-style-type: none"> <li>Nilai akurasi sebesar 69%</li> <li>Nilai presisi sebesar 65,73%</li> <li>Nilai recall sebesar 95,91%</li> </ol> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Terdapat rumus untuk menghitung nilai akurasi, recall dan presisi.</li> <li>Terdapat rumus teorema Naïve Bayes, rumus Gaussian Naïve Bayes, rumus</li> </ol> |

|    |  |   |
|----|--|---|
|    |  | <p>mean, rumus probabilistic, dan rumus standar deviasi.</p> <p>c) Pada hasil evaluasi dan validasi terdapat hasil scenario</p> <p>7. Kelemahan Penelitian =</p> <p>a) Variabel yang digunakan untuk analisis kurang transparan atau jelas.</p> <p>b) Penelitian ini tidak menunjukkan perhitungan secara manual.</p> <p>8. Kesimpulan = Nilai akurasi, recall, dan presisi didapatkan sebagai berikut:</p> <p>a) Nilai akurasi sebesar 69%</p> <p>b) Nilai presisi sebesar 65,73%</p> <p>c) Nilai recall sebesar 95,91%</p>  |
| 7. | <p>Klasifikasi Pengidap Diabetes Pada Perempuan Menggunakan Penggabungan Metode Support Vector Machine dan K-Nearest Neighbour</p> | <p>1. Tujuan Penelitian = Untuk mengklasifikasikan objek baru berdasarkan atribut dan training samples.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data diagnosis DM yang berasal dari Dataset Pima Indian, terdiri dari 768 data.</p> <p>3. Perancangan Sistem =</p> <p>a) Algoritma K-Nearest Neighbour (KNN)</p> <p>b) Algoritma Support Vector Machine (SVM)</p> <p>4. Teknik Pengumpulan Data = Pengumpulan data yang dilakukan dalam penelitian ini diperoleh melalui repository database UCI Pima Indian.</p> <p>5. Hasil Penelitian = Hasil dari penelitian ini adalah nilai akurasi sebesar 92%</p> <p>6. Kelebihan Penelitian =</p> <p>d) Terdapat rumus SVM.</p> <p>e) Terdapat rumus KNN.</p> <p>f) Pada hasil dan analisis terdapat tabel hasil uji coba.</p> <p>7. Kelemahan Penelitian =</p> <p>c) Penelitian ini tidak menunjukkan perhitungan secara manual.</p> <p>d) Penelitian ini tidak menunjukkan software yang digunakan untuk menganalisis data.</p> <p>8. Kesimpulan = Dari uji coba yang telah dilakukan, metode usulan terbukti mampu mengoptimalkan pemilihan parameter k</p> |

|    |  |   |
|----|--|---|
|    |  | dari metode KNN, sehingga nilai akurasi yang didapat dapat lebih tinggi dari metode KNN dan SVM tradisional.  |
| 8. | Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Mellitus (Studi Kasus: RS. Siti Khadijah Palembang) | <ol style="list-style-type: none"> <li>1. Tujuan Penelitian = Untuk mengetahui algoritma mana yang tingkat akurasi yang paling akurat untuk klasifikasi penyakit diabetes mellitus antara algoritma Naïve Bayes dan Support Vector Machine.</li> <li>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data yang berasal dari rekam medis Rumah Sakit Khadijah Palembang yang berjumlah 613 record dengan atribut sebanyak 9 atribut dalam kurun waktu hamper 3 tahun.</li> <li>3. Perancangan Sistem = <ol style="list-style-type: none"> <li>a) Data Mining</li> <li>b) Algoritma Naïve Bayes</li> <li>c) Algoritma SVM (Support Vector Machine)</li> <li>d) K-Fold Cross Validation</li> <li>e) Confusion Matrix</li> <li>f) WEKA</li> </ol> </li> <li>4. Teknik Pengumpulan Data = Pengumpulan data yang dilakukan dalam penelitian ini diperoleh dari data Rumah Sakit Khadijah Palembang yang digunakan pada tahun 2017-2019 sebanyak 613 record.</li> <li>5. Hasil Penelitian = Hasil dari penelitian ini adalah <ol style="list-style-type: none"> <li>a) Hasil dari Confusion Matrix Naïve Bayes didapat nilai akurasi sebesar 92,0746% dan eror sebesar 7,9254%</li> <li>b) Hasil dari Confusion Matrix SVM dengan kernel Polynomial didapat nilai akurasi sebesar 96,2704% dan eror sebesar 3,7296%</li> </ol> </li> <li>6. Kelebihan Penelitian = <ol style="list-style-type: none"> <li>a) Terdapat rumus SVM.</li> <li>b) Terdapat rumus Naïve Bayes.</li> <li>c) Terdapat rumus untuk mencari nilai akurasi dan error menggunakan Confusion Matrix.</li> </ol> </li> <li>7. Kelemahan Penelitian = <ol style="list-style-type: none"> <li>a) Penelitian ini tidak menunjukkan perhitungan secara manual.</li> </ol> </li> </ol> |



|     |  |   |
|-----|--|---|
|     |  | 8. Kesimpulan = Algoritma yang akurat dalam klasifikasi penyakit diabetes yaitu algoritma Support Vector Machine (SVM) dengan kernel polynomial   |
| 9.  | Pendekatan Data Science untuk Deteksi Dini Diabetes Menggunakan Naïve Bayes Classifier | <p>1. Tujuan Penelitian = Untuk membantu seseorang mengenali lebih dini mengenai risiko penyakit diabetes dan untuk mengklasifikasikan terhadap gejala yang diberikan apakah gejala tersebut masuk dalam penyakit diabetes atau tidak.</p> <p>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data random hasil survey dari 100 orang..</p> <p>3. Perancangan Sistem =</p> <ol style="list-style-type: none"> <li>Machine Learning.</li> <li>Algoritma Naïve Bayes</li> <li>Algoritma Laplacian Smoothing.</li> <li>Confusion Matrix</li> <li>Jupyter Notebook</li> </ol> <p>4. Teknik Pengumpulan Data = Pengumpulan data yang dilakukan dalam penelitian ini diperoleh dari hasil survey pada 100 orang sehingga diperoleh 100 record data.</p> <p>5. Hasil Penelitian = Hasil dari penelitian ini adalah mendapatkan nilai akurasi sebesar 70% yang berasal dari pembagian dataset sebesar 40% untuk data testing dan 60% untuk data training.</p> <p>6. Kelebihan Penelitian =</p> <ol style="list-style-type: none"> <li>Terdapat rumus Laplacian Smoothing.</li> <li>Terdapat rumus Naïve Bayes.</li> <li>Terdapat rumus untuk mencari nilai akurasi menggunakan Confusion Matrix.</li> <li>Terdapat tabel data real.</li> </ol> <p>7. Kelemahan Penelitian =</p> <ol style="list-style-type: none"> <li>Penelitian ini tidak menunjukkan perhitungan secara manual.</li> </ol> <p>8. Kesimpulan = Penelitian yang telah dilakukan dengan menggunakan algoritma Naïve Bayes dan Laplacian Smoothing mampu untuk digunakan dalam memprediksi deteksi awal dari penyakit diabetes.</p> |
| 10. | Penerapan Metode Naïve Bayes Dalam   | 1. Tujuan Penelitian = Peneliti ingin memprediksi obesitas menggunakan  |

|   |  |
|---|--|
| <p>Mengklasifikasi Tingkat Obesitas Pada Pria</p> | <p>algoritma Naïve Bayes dengan beberapa kriteria yang didapat dari pengukuran antropometri pada pria beserta perhitungan Indeks Massa Tubuh (IMT), dengan mengimplementasikan algoritma Naïve Bayes diharapkan dapat mengklasifikasi dan memprediksi tingkat obesitas pada pria.</p> <ol style="list-style-type: none"> <li>2. Lokasi dan Subject Penelitian = Subject yang digunakan pada penelitian ini adalah data pengukuran antropometri yang didapat dari Kaggle Machine Learning and Data Science Community dengan total 252 data.</li> <li>3. Perancangan Sistem =       <ol style="list-style-type: none"> <li>g) Data Mining</li> <li>h) Algoritma Naïve Bayes</li> </ol> </li> <li>4. Teknik Pengumpulan Data = Pengumpulan data yang dilakukan dalam penelitian ini diperoleh dari Kaggle Machine Learning and Data Science Community dengan total 252 data.</li> <li>5. Hasil Penelitian = Hasil dari penelitian ini adalah akurasi yang didapat sebesar 83,15% dengan menggunakan split data 60% untuk data training dan 40% untuk data testing.</li> <li>6. Kelebihan Penelitian =       <ol style="list-style-type: none"> <li>d) Terdapat rumus Naïve Bayes.</li> <li>e) Terdapat rumus mean, standar deviasi, probabilitas dan distribusi gaussian.</li> <li>f) Terdapat rumus untuk mencari nilai akurasi, precision, recall dan F1-Score.</li> </ol> </li> <li>7. Kelemahan Penelitian =       <ol style="list-style-type: none"> <li>b) Penelitian ini tidak menunjukkan perhitungan secara manual.</li> </ol> </li> <li>8. Kesimpulan = Penerapan algoritma Naïve Bayes dalam memprediksi tingkat obesitas pada pria dapat diterapkan dengan menghasilkan tingkat akurasi yang memuaskan sebesar 84,15% pada rasio split data 60:40.</li> </ol> |
|---|--|