

Bab II KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining

Vulandari (2017, hal. 1) menyatakan bahwa data mining artinya proses menganalisis untuk menemukan pengetahuan yang disimpan pada *database*; informasi yang dihasilkan didapatkan dengan mengekstraksi dan mengenali pola yang penting dan menarik dari data yang terdapat pada basis data; data mining terutama digunakan dalam menggali pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut dengan *Knowledge Discovery In Database* yang disingkat KDD; proses KDD memiliki beberapa tahapan di antaranya sebagai berikut :

- (1) *Data Cleansing*, yaitu proses pembersihan data untuk memilih data yang dianggap bisa dipakai atau melepaskan data yang tidak selaras (data yang tidak konsisten dan *noise*);
- (2) *Data Integration*, yaitu proses menggabungkan data yang dianggap berulang akan digabungkan menjadi satu atau kombinasi data dari beberapa sumber;
- (3) *Data Selection*, Pemilihan data yaitu proses seleksi atau pemilihan data yang dianggap relevan terhadap analisis.
- (4) *Data Transformation*, yaitu proses transformasi data terpilih ke dalam bentuk mining procedure atau merubah data menjadi bentuk yang sesuai agar dapat di mining;
- (5) *Data Mining*, yaitu proses dimana dilakukan beragam teknik untuk mengekstrak pola-pola potensial menghasilkan data yang berguna atau proses ekstrasi pola dari data yang ada;
- (6) *Pattern Evolution*, yaitu proses dimana pola-pola yang telah diidentifikasi berdasarkan measure yang diberikan atau menafsirkan pola menjadi pengetahuan yang dapat digunakan untuk mengambil keputusan;
- (7) *Knowledge Presentation*, yaitu data yang sudah diproses divisualisasikan ke dalam bentuk gambar, tulisan, grafik, dan lain-lain agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis.

2. Unified Modelling Language (UML)

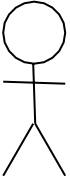


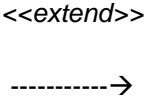
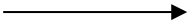
Unified Modelling Language (UML) merupakan suatu metode dalam pemodelan secara visual yang digunakan sebagai fasilitas perancangan sistem berorientasi objek. Awal mulanya, UML diciptakan oleh Object Management Group dengan tipe awal 1.0 pada bulan Januari 1997.

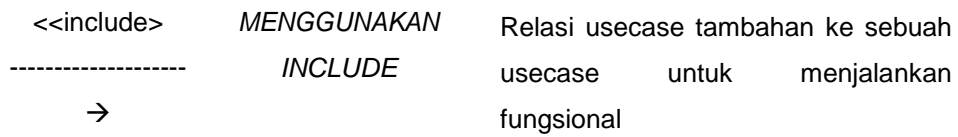
Menurut Rosa (2015, p 113) memberikan batasan bahwa, "UML adalah salah satu standar bahasa yang banyak digunakan didunia industri untuk mendefinisikan requirement, membuat analisis dan desain, serta menggambarkan arsitektur dalam pemograman berorientasi objek".

Uml pun bisa digunakan sebagai alat transfer ilmu tentang sistem aplikasi yang hendak dikembangkan dari developer satu ke developer lainnya. UML sangat penting bagi sebagian orang sebab UML berfungsi sebagai bridge atau jembatan penerjemah antara pengembang sistem dengan pengguna. Di sinilah pengguna dapat memahami sistem yang nantinya akan dikembangkan.

a. Use Case Diagram

Menurut Rosa and Shalahuddin (2016, p. 155),Use Case Diagram adalah pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Use Case mendeskripsikan sebuah interaksi antara satu atau lebih actor dengan sistem informasi yang akan dibuat.

SIMBOL	NAMA	KETERANGAN
	ACTOR	Orang proses atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat diluar sistem informasi dibuat itu sendiri, jadi walaupun simbol dari actor adalah gambar orang, biasanya dinyatakan menggunakan kata benda di awal frase nama actor
	USECASE	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau actor biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama usecase.
	ASOSIASI/ ASSOCIATION	Komunikasi antara actor dan usecase yang berpartisipasi pada usecase atau usecase memiliki interaksi dengan actor.
	EKSTENSI/ EXTEND	Relasi usecase tambahan ke sebuah usecase dimana usecase yang ditambahkan dapat berdiri sendiri walau tanpa usecase tambahan memiliki nama depan yang sama dengan usecase yang <i>ditambahkan</i> .
	GENERALISASI / GENERALIZATION	Hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah usecase dimana fungsi yang satu adalah fungsi yang lebih <i>umum dari lainnya</i> .



Gambar 2. 1 Use Case


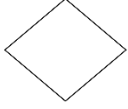
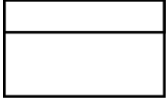

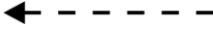
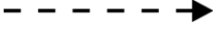

b. Class Diagram

Apabila use case diagram berbicara mengenai alur dari sebuah sistem atau program, maka class diagram merupakan alur database dalam sebuah program. Dimana pada setiap laporan sistem yang hendak dibuat, class diagram harus diikuti sertakan, karena dari sinilah kita dapat melihat aliran data sebuah program.

Selain class diagram, dalam sebuah UML harus ada juga yang namanya sequence diagram. Apabila class diagram menjelaskan aliran data, sequence diagram ini menjelaskan mengenai proses program.

Menurut Sukanto dan Shalahuddin (2013, p 141), mereka berdua mengartikan class diagram adalah penggambaran dari struktur sistem serta mendefinisikan kelas yang hendak dibangun sistem.

Menurut Satzinger (2011, p 28) memaparkan, pengertian class diagram atau diagram kelas ialah alat untuk membantu menggambarkan serta memvisualisasikan struktur kelas yang terdapat dalam suatu sistem atau program class diagram ini mendeskripsikan kelompok objek bersama properti, operasi dan relasi yang sama.

Simbol	Nama	Keterangan
	Generalization	Hubungan dimana objek anak (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor)
	Navy Association	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek
	Class	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama
	Collaboration	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
	Realization	Operasi yang benar-benar dilakukan oleh suatu objek
Simbol	Nama	Keterangan
	Dependency	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung pada elemen yang tidak mandiri.
	Association	Yang menghubungkan antara objek satu dengan objek lainnya.



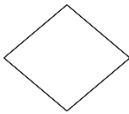
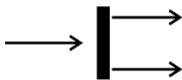

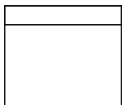
Gambar 2. 2 Class Diagram

c. Activity Diagram

Menurut Sugiarti (2018, Hal. 133) diagram activity adalah suatu dasar-dasar dalam pemrograman atau diagram yang menggambarkan workflow aktivitas dari sebuah sistem atau proses bisnis. Hal yang perlu diperhatikan bahwa diagram aktivitas sistem bukan apa yang dilakukan aktor, jadi aktivitas yang dapat dilakukan oleh sistem.

Alur atau aktivitas berupa bisa berupa runtutan menu-menu atau proses bisnis yang terdapat di dalam sistem tersebut. Dalam buku Rekayasa Perangkat Lunak karangan Rosa A.S mengatakan, "Diagram aktivitas tidak menjelaskan kelakuan aktor. Dapat diartikan bahwa dalam pembuatan activity diagram hanya dapat dipakai untuk menggambarkan alur kerja atau aktivitas sistem saja."

Activity diagram mesti digunakan sejajar (horizontal) dengan teknik pemodelan lainnya, seperti diagram Use Case dan diagram State. Kamu bisa menggunakan activity diagram agar dapat memodelkan alur kerja sistem dengan baik. Activity diagram berfungsi juga untuk menganalisis diagram use case dengan cara mendeskripsikan aktor.

Simbol	Nama	Keterangan
	Status Awal/Initial	Status awal aktivitas sistem
	Activity	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	Decision	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	Join	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	Status Akhir/Final	Status akhir yang dilakukan sistem,.
	Swimlane	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi.

Gambar 2. 3 Activity Diagram

3. Klasifikasi

Klasifikasi ialah suatu proses menemukan definisi kesamaan karakteristik dalam suatu kelompok maupun kelas (*class*). Klasifikasi data mining jadi salah satu metode yang paling umum untuk digunakan. Metode ini dilakukan bertujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui.

4. Business Process Model And Notation (BPMN)

Menurut Jeston dan Nelis (2006, p. 196) “Business Process Modeling Notation (BPMN) adalah suatu notasi standar yang dapat berupa ikon atau gambar untuk digunakan di dalam pemodelan proses bisnis.”(Business Process Model and Notation) adalah flowchart berbasis notasi untuk mendefinisikan dan menggambarkan logika dari langkah-langkah dalam proses bisnis. Sebuah standar untuk menggambarkan proses bisnis yang dikeluarkan oleh Open Management Group. BPMN menyediakan bahasa umum untuk visualisasi, dokumentasi, komunikasi, analisis, perbaikan, simulasi, dan eksekusi proses bisnis. Hingga saat ini, sudah banyak organisasi besar yang menerapkan standar BPMN untuk mewujudkan proses bisnisnya.

Berikut adalah aturan dalam memodelkan suatu proses bisnis :

- (a) Memodelkan kejadian-kejadian yang memulai proses, proses yang dilakukan dan hasil akhir dari aliran proses;
- (b) Keputusan bisnis atau percabangan aliran dimodelkan dengan gateways. Sebuah gateway mirip dengan simbol keputusan dalam flowchart;
- (c) Sebuah proses dalam aliran dapat mengandung sub-proses, yang secara grafis dapat ditunjukkan dengan BPD (Business Process Diagram) lain yang tersambung melalui sebuah hyperlink ke simbol proses;
- (d) Jika sebuah proses tidak diletakkan ke dalam sub proses, maka dianggap sebagai sebuah task – yaitu level proses paling rendah;
- (e) Sebuah tanda ‘+’ pada simbol proses menunjukkan bahwa proses ini didekomposisi, jika tidak ada tanda ‘+’, maka proses ini disebut sebuah task.

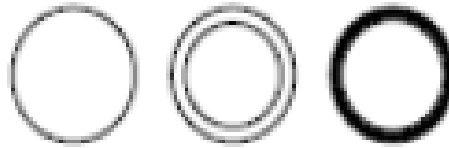
BPD terdiri dari beberapa elemen grafis dan kategori dasar elemen BPD adalah:

a. Flow Objects

Flow Object dibagi menjadi 3, yaitu event, activity dan gateway.

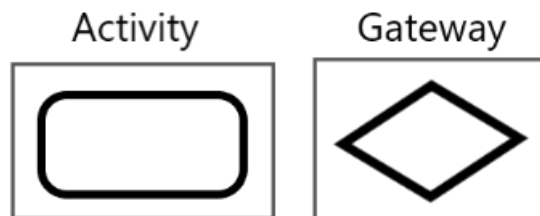
- (1) *Event* digambarkan dengan sebuah lingkaran dan merupakan sesuatu yang “terjadi” selama berlangsungnya proses bisnis. *Event-event* ini mempengaruhi aliran proses dan biasanya memiliki penyebab (*trigger*) atau hasil (*result*). Event adalah lingkaran dengan pusat terbuka untuk memungkinkan pembedaan *trigger* dan *result* yang berbeda. Terdapat 3 tipe event

berdasarkan kapan mereka mempengaruhi aliran yaitu *Start*, *Intermediate*, dan *End*;



Gambar 2. 4 simbol event

- (2) *Activity* ditunjukkan dengan persegi panjang dengan ujung-ujung bulat dan merupakan bentuk umum untuk pekerjaan yang dilakukan oleh perusahaan. Sebuah aktivitas dapat berdiri sendiri atau gabungan. Tipe dari aktivitas adalah task dan sub process yang dibedakan dengan tanda + pada bagian tengah bawah dari bentuk tersebut;
- (3) *Gateway* digambarkan dengan bentuk seperti belah ketupat dan digunakan untuk mengontrol percabangan dan penggabungan Sequence Flow. Jadi, *gateway* menentukan keputusan tradisional, penggabungan, dan penggabungan aliran. Internal Markers akan menentukan perilaku dari kontrol.

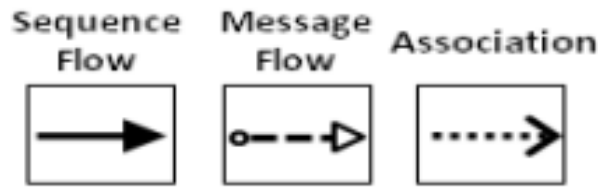


Gambar 2. 5 simbol activity dan gateway

b. Connecting Objects

Connecting Object adalah elemen yang menghubungkan flow object. *Connecting Object* juga memiliki 3 jenis elemen yaitu :

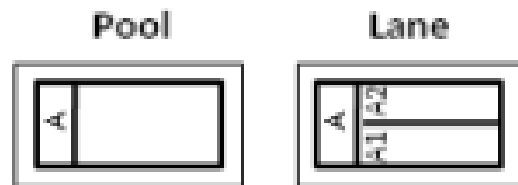
- 1) Alur Sequence (Sequence flow) digunakan untuk menunjukkan urutan yang kegiatan akan yang dilakukan dalam sebuah proses;
- 2) Alur Pesan (Message Flow) digunakan untuk menunjukkan aliran pesan antara dua entitas yang siap untuk mengirim dan menerima;
- 3) Asosiasi (Association) digunakan untuk asosiasi data, informasi dan artefak dengan aliran benda.



Gambar 2. 6 Simbol Sequence Flow, Message Flow dan Association

c. Swimlanes

Swimlanes digambarkan dengan bentuk garis yang memisahkan dan mengelompokkan aktor (pelaku yang berinteraksi dengan system). BPMN mendukung swimlanes dengan dua bentuk *swimlane objects* yaitu pool yang mewakili partisipan dalam sebuah proses dan lane yaitu sub-bagian dalam sebuah pool dan akan menambah panjang dari pool baik vertikal ataupun horizontal. Lanes digunakan untuk mengatur dan mengkategorikan aktivitas.



Gambar 2. 7 Simbol Pool dan Lane dalam Swimlane

d. Artifacts

Artifacts adalah elemen yang digunakan untuk memberikan informasi tambahan dari sebuah proses. BPMN dirancang untuk memungkinkan pemodel dan alat pemodelan fleksibilitas untuk memperluas notasi dasar dan menyediakan kemampuan untuk konteks tambahan yang tepat untuk situasi pemodal tertentu.

5. Python (Bahasa Pemograman)

Python merupakan bahasa pemrograman yang berorientasi objek dengan semantik dinamis. Struktur data bawaan tingkatan besar, dikombinasikan dengan pengetikan dinamis serta pengikatan dinamis, membuatnya sangat menarik untuk Pengembangan Aplikasi Kilat, dan untuk digunakan selaku bahasa skrip ataupun lem untuk menghubungkan komponen yang bersama-sama.

6. Naïve Bayes

Naïve Bayes Menurut Olson dan Delen (2008, p 102) *menjelaskan Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

Kelebihan menggunakan metode *Naïve Bayes* ini yaitu Menangani kuantitatif dan data diskrit, Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata-ratakan ketika mengestimasi peluang bersyarat data. Kekurangan yang didapat pada *Naïve Bayes* ini yaitu Tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga dan Mengasumsikan variabel bebas.

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

(Sumber Khoiriyah, Muharni, Perdana, 2019:Vol 3)

Dimana :

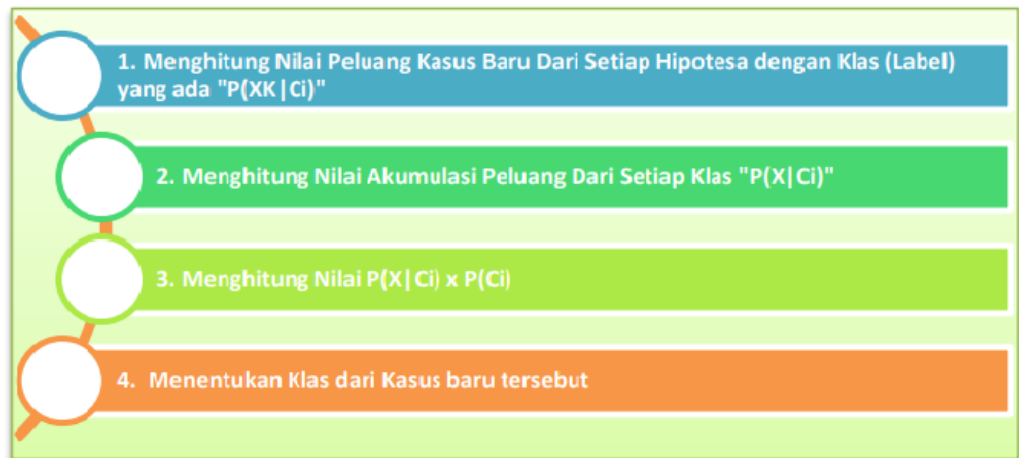
- a. $P(C_i | X)$ yaitu, probabilitas C_i terjadi jika X sudah terjadi.
- b. $P(C_i)$ adalah kemungkinan C_i didata, bersifat independent terhadap X .
- c. X adalah kumpulan atribut.
- d. $P(X | C_i)$ adalah probabilitas X terjadi jika C_i benar atau sudah terjadi berdasarkan data pelatihan (data training).

Langkah-langkah perhitungan naive bayes

- (a) Menghitung jumlah kelas.
- (b) Menghitung jumlah kasus perkelas.
- (c) Mengalikan semua variable kelas.
- (d) Membandingkan hasil perkelas

Penentuan *class* dilakukan dengan cara membandingkan nilai probabilitas suatu sampel berada di *class* yang lain. Untuk menentukan *class* yang cocok dari suatu sampel dilakukan dengan cara membandingkan nilai posterior untuk masing-masing *class*, dan mengambil *class* dengan nilai posterior yang tertinggi.

Adapun algoritma penyelesaian dari Metode Naïve Bayes dapat dilihat pada gambar berikut :



Gambar 2. 8 Naive Bayes

(Sumber : (Nofriansyah, 2015, p. 46))

Menurut Suyanto (2017, p 112) contoh kasus ini yaitu keputusan Ya dan Tidak dalam penerimaan bonus, data training yang digunakan yaitu 11 data dan data yang telah di klasifikasikan berdasarkan kartu, panggilan dan blok, adapun data training yang digunakan dibawah ini berbentuk tabel yaitu :

Tabel 2. 1 Data Training

No.	Pelanggan	Kartu	Panggilan	Blok	Bonus
1.	Andi	Prabayar	Sedikit	Sedang	Tidak
2.	Budi	Pascabayar	Banyak	Sedang	Ya
3.	Citra	Prabayar	Banyak	Sedang	Ya
4.	Dedi	Prabayar	Banyak	Rendah	Tidak
5.	Evan	Pascabayar	Cukup	Tinggi	Ya
6.	Feni	Pascabayar	Cukup	Sedang	Ya
7.	Gito	Prabayar	Cukup	Sedang	Ya
8.	Hani	Prabayar	Cukup	Rendah	Tidak
9.	Jodi	Pascabayar	Sedikit	Tinggi	Ya
10.	Kafi	Pascabayar	Banyak	Tinggi	Ya
11.	Linda	Pascabayar	Sedikit	Rendah	Ya

Pelanggan	Kartu	Panggilan	Blok	Bonus
Mira	Pascabayar	Cukup	Rendah	?

Penyelesaian :

(1) menghitung probabilitas prior untuk kelas pertama;

$$P(\text{Bonus} = \text{Ya}) = 8/11 = 0,7273$$

$$P(\text{Bonus} = \text{Tidak}) = 3/11 = 0,2727$$

(2) menghitung probabilitas bersyarat untuk setiap kelas;

$$P(\text{Kartu} = \text{Pascabayar} | \text{Bonus} = \text{Ya}) = 6/8 = 0,75$$

$$P(\text{Kartu} = \text{Pascabayar} | \text{Bonus} = \text{Tidak}) = 0/3 = 0$$

$$P(\text{Panggilan} = \text{Cukup} | \text{Bonus} = \text{Ya}) = 3/8 = 0,375$$

$$P(\text{Panggilan} = \text{Cukup} | \text{Bonus} = \text{Tidak}) = 1/3 = 0,3333$$

$$P(\text{Blok} = \text{Rendah} | \text{Bonus} = \text{Ya}) = 1/8 = 0,125$$

$$P(\text{Blok} = \text{Rendah} | \text{Bonus} = \text{Tidak}) = 2/3 = 0,6667$$

(3) menghitung probabilitas untuk setiap kelas;

$$P(X | \text{Bonus} = \text{Ya}) = 0,75 \times 0,375 \times 0,125 = 0,0352$$

$$P(X | \text{Bonus} = \text{Tidak}) = 0 \times 0,3333 \times 0,6667 = 0$$

(4) menghitung probabilitas posterior;

$$P(X | \text{Bonus} = \text{Ya}) \times (P | \text{Bonus} = \text{Ya}) = 0,0352 \times 0,7273 = 0,256$$

$$P(X | \text{Bonus} = \text{Tidak}) \times (P | \text{Bonus} = \text{Tidak}) = 0 \times 0,2727 = 0$$

Pelanggan	Kartu	Panggilan	Blok	Bonus
Mira	Pascabayar	Cukup	Rendah	Ya

Mengingat $P(X | \text{Bonus} = \text{Ya}) \times (P | \text{Bonus} = \text{Ya}) = 0,256$ lebih besar dari $P(X | \text{Bonus} = \text{Tidak}) \times (P | \text{Bonus} = \text{Tidak}) = 0$, maka Naïve Bayes Classifier mengklasifikasikan pelanggan bernama Mira ke dalam kelas bonus = "Ya".

B. Tinjauan Studi (Penelitian rujukan)

(1) Menurut (Coastera, Yusa, Lediwara, & Sari, 2020), Dengan judul "Analisis Performa Algoritma Naïve Bayes Untuk Penentuan Kelayakan Pendorong Darah", Berdasarkan hasil dari ujicoba implementasi model algoritma Naive Bayes terhadap dataset klasifikasi kelayakan calon pendonor darah PMI, performanya model tergolong sangat baik yaitu berada di jumlah 80% dari 100 records. Tahap data preprocessing melibatkan 7 atribut yang menjadi indikator penilaian dalam hal mempertimbangkan atribut umur, berat badan, kadar hemoglobin, tekanan darah sistolik dan distolik, serta label atau klasifikasi calon pendonor. Ditinjau dari performanya Recall dan Precisionnya, Algoritma

Naive Bayes juga memiliki nilai kinerja yang sangat baik yaitu 80,61% untuk nilai precision dan 80,405% untuk nilai Recall.

- (2) **Menurut (Sahar, 2020), dengan judul “Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung”** , Penyakit Jantung merupakan salah satu bentuk penyakit kardiovaskular yang menjadi penyebab kematian nomor satu di dunia. Dalam proses pengklasifikasian ini untuk mengetahui apakah termaksud penyakit jantung atau non penyakit jantung dengan menggunakan rumus dari metode *K-Nearest Neighbor* dan *NBC Classifier* yang menggunakan *library scikit learn*. Penerapan metode NBC pada dataset Penyakit jantung diperoleh nilai performa terbaik pada data testing 10% yaitu akurasi sebesar 58%,presisi 55%, recall 90%, f-masure 68%.
- (3) **Menurut (Ridwan A, 2020), dengan judul “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus”**, Diabetes Mellitus atau kencing manis adalah penyakit metabolisme disebabkan oleh kadar gula tinggi didalam darah. ketika terserang Diabetes, pada tubuh manusia insulin tidak biasa dihasilkan secara cukup bahkan tubuh tidak dapat menggunakan insulin tersebut secara benar sesuai kebutuhan. Diabetes Mellitus terdaftar sebagai penyakit penyumbang kematian terbesar terbesar didunia. Diabetes Mellitus dapat diklasifikasikan berdasarkan kemungkinan terkenanya dari atribut gejala diawal fasenya. penyakit ini bisa dideteksi karena banyak gejala yang terdeteksi. Terdapat 16 atribut yang mempengaruhi Klasifikasi dataset Early stage diabetes risk prediction yaitu Age, Gender, Polyuria, Polydipsia, sudden weight loss, weakness, Polyphagia, Genital thrush, visual blurring, Itching, Irritability, delayed healing, partial paresis, muscle stiffness, Alopecia, Obesity, class.
- (4) **Menurut (Doni et al., 2021), dengan judul “PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT HEPATOCELLULAR CARCINOMA MENGGUNAKAN ALGORITMA NAÏVE BAYES”**,Hasil penelitian menunjukkan bahwa hasil performa algoritma Naïve Bayes yang diukur menggunakan Confussion Matrix dengan nilai yang dihasilkan berupa nilai Akurasi sebesar 70,30%, Presisi sebesar 73,53% dan Recall sebesar 77,32% serta hasil performa yang diukur menggunakan Kurva ROC (Receiver Operating Characteristic) dengan nilai yang dihasilkan berupa nilai AUC sebesar 0.783 yang termasuk dalam kategori Fair Classification atau kategori Klasifikasi Cukup.

- (5) Menurut (Jackins et al., 2021), dengan judul “**AI based smart prediction of clinical disease using random forest classifier and Naive Bayes**”, Praktik perawatan kesehatan termasuk mengumpulkan semua jenis data pasien yang akan membantu Dokter mendiagnosis kondisi kesehatan pasien dengan benar. Data ini bisa menjadi gejala sederhana yang diamati oleh subjek, diagnosis awal oleh dokter atau hasil tes terperinci dari laboratorium. Dengan demikian, data tersebut hanya digunakan untuk analisis oleh dokter yang kemudian memastikan penyakit menggunakan keahlian medis pribadinya. Kecerdasan artificial telah digunakan dengan klasifikasi Naive Bayes dan acak algoritma pengklasifikasian hutan untuk mengklasifikasikan banyak kumpulan data penyakit seperti diabetes, jantung penyakit, dan kanker untuk memeriksa apakah pasien terkena penyakit itu atau tidak. Analisis kinerja data penyakit untuk kedua algoritma dihitung dan dibandingkan dengan Hasil simulasi menunjukkan efektifitas klasifikasi.
- (6) Menurut (Valdiviezo-Diaz et al., 2019), dengan judul “**A Collaborative Filtering Approach Based on Naïve Bayes Classifier**”, Sistem pemberi rekomendasi adalah alat penyaringan informasi yang digunakan untuk mengurangi kelebihan informasi untuk pengguna di web. Pemfilteran kolaboratif merekomendasikan item kepada pengguna berdasarkan peringkat historis mereka. Informasi Eksperimen yang dilakukan menggunakan empat set data menyajikan hasil yang baik dibandingkan dengan beberapa baseline canggih, mencapai kinerja terbaik menggunakan Normalized Discounted Cumulative Gain (nDCG) ukuran kualitas dan juga meningkatkan akurasi prediksi dalam beberapa set data.
- (7) Menurut (Utomo & Mesran, 2020), dengan judul “**Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung**”, Berdasarkan hasil penelitian dan pengujian yang telah dilakukan maka dapat ditarik kesimpulan adalah Kinerja Algoritma Naïve Bayes Classifier (NBC) lebih baik.
- (8) Menurut (Setiyani & Prasetyaningrum, 2021), dengan judul “**PENERAPAN METODE NAIVE BAYES CLASSIFIER PADA SISTEM PAKAR DIAGNOSA PENYAKIT LAMBUNG**”, Sistem yang dirancang dengan mengimplementasikan metode naïve bayes classifier dapat digunakan untuk membantu dalam proses diagnosa awal jenis penyakit lambung pada Puskesmas Depok III dan masyarakat secara umum.
- (9) Menurut (Byna & Basit, 2020), dengan judul “**PENERAPAN METODE ADABOOST UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE**”

DENGAN ALGORITMA NAÏVE BAYES", Stroke atau kecelakaan serebrovaskular, melibatkan cedera pada sistem saraf pusat sebagai akibat dari penyebab vaskular, dan merupakan penyebab utama kecacatan di seluruh dunia. Stroke merupakan kematian tertinggi kedua di seluruh dunia dengan penyebab utama adalah kecacatan.

(10) Menurut (Maulana et al., 2022) dengan judul "Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit Stunting Pada Balita", Berdasarkan hasil pengujian menggunakan data sampel pada sistem pakar untuk mendiagnosa penyakit stunting pada balita menggunakan metode naive bayes, maka di dapatkan hasil yang sesuai dengan perncangan dan sistem dapat berjalan dengan baik.

(11) Menurut (Zaw et al., 2019) dengan judul "Brain tumor detection based on Naïve Bayes Classification",

Kanker otak disebabkan oleh populasi sel abnormal yang disebut sel glial yang terjadi di otak. Kanker otak disebabkan oleh populasi sel abnormal yang disebut sel glial yang terjadi di otak. Selama bertahun-tahun, jumlah pasien yang menderita kanker otak meningkat sehubungan dengan populasi yang menua, ini adalah masalah kesehatan di seluruh dunia. Dalam makalah ini, klasifikasi Naïve Bayes digunakan untuk mengenali daerah tumor secara akurat yang berisi semua jaringan kanker yang menyebar. Database MRI otak, preprocessing, operasi morfologi, pengurangan piksel, ambang entropi maksimum, ekstraksi fitur statistik dan algoritma prediksi berbasis pengklasifikasi Naïve Bayes digunakan dalam penelitian ini. Ketika diuji dengan 50 Gambar MRI (Magnetic resonance imaging), metode ini mengembangkan tingkat deteksi 81,25% pada Gambar tumor dan tingkat deteksi 100% pada Gambar yang bukan tumor dengan akurasi keseluruhan 94%.

(12) Menurut (Kamel et al., 2019) dengan judul "Cancer Classification Using Gaussian Naive Bayes Algorithm",

Kanker adalah salah satu penyakit paling mematikan di dunia dan itu adalah subjek yang memprihatinkan karena sampai saat ini mereka tidak dapat menemukan pengobatan yang cocok untuk penyakit ini. Jika dan hanya jika penyakit ini terdeteksi pada tahap awal, pasien yang memiliki penyakit ini dapat diselamatkan. Dalam pengerjaan makalah, algoritma Gaussian Naive Bayes digunakan untuk klasifikasi kanker. Algoritma diuji dengan menerapkannya pada dua dataset di mana yang pertama adalah dataset Kanker Payudara Wisconsin dan yang kedua adalah dataset kanker paru-

paru. Hasil evaluasi dari algoritma yang diusulkan telah mencapai akurasi 98% memprediksi kanker payudara dan 90% memprediksi kanker paru-paru.

(13) Menurut (Maheswari & Pitchai, 2019) dengan judul “Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm”,

Informasi besar data perawatan kesehatan dikumpulkan dari industri perawatan kesehatan yang sayangnya tidak "ditambang" untuk membuat pengambilan keputusan yang efektif untuk identifikasi informasi tersembunyi. Pengguna mendahului proses dengan memeriksa detail spesifik dan gejala penyakit jantung. Pohon keputusan dan teknik naive Bayes dalam penambangan data digunakan untuk mengambil detail yang terkait dengan setiap pasien. Berdasarkan prediksi hasil yang akurat, kinerja sistem dianalisis.

(14) Menurut (Alotaibi, 2019) dengan judul “Implementation of machine learning model to predict heart failure disease”,

Di era saat ini, Gagal Jantung merupakan salah satu penyakit umum yang dapat menyebabkan situasi berbahaya. Setiap tahun hampir 26 juta pasien terkena penyakit semacam ini. Dari sudut pandang konsultan jantung dan ahli bedah, ini adalah memprediksi kompleks untuk gagal jantung pada waktu yang tepat. Selanjutnya, hasil dan studi banding menunjukkan bahwa, Hasil saat ini meningkatkan skor akurasi sebelumnya dalam memprediksi penyakit jantung. Integrasi model pembelajaran mesin yang disajikan dalam penelitian ini dengan sistem informasi medis akan berguna untuk memprediksi HF atau penyakit lain menggunakan data langsung yang dikumpulkan dari pasien.

(15) Menurut (Almustafa, 2021) dengan judul “Prediction of chronic kidney disease using different classification algorithms”,

Penyakit ginjal kronis adalah masalah fungsi ginjal umum yang menyebabkan kerusakan bentuk ginjal dan menyebabkan gagal ginjal. Metodologi diagnostik dini untuk menentukan fungsi ginjal sangat penting dan sangat penting dalam banyak kasus. Algoritma yang diterapkan menggunakan *random tree*, *decision table* (DT), *K-nearest neighbor* (K-NN), J48, *stochastic gradient descent* (SGD) dan pengklasifikasi Naïve Bayes, dan model prediksi diusulkan berdasarkan pemilihan fitur untuk memprediksi kasus CKD secara efisien. Hasil penelitian menunjukkan bahwa J48 dan pengklasifikasi tabel keputusan mengungguli pengklasifikasi lain dengan akurasi 99%, ROC sama dengan 0,999 dan 0,992, MAE 0,0225 dan 0,1815, dan RMSE masing-masing 0,0807 dan 0,2507. Naïve Bayes dan *decision table* klasifikasi ditingkatkan menjadi 99,75%, 98,25% dan 99,25%.

(16) Menurut (Almustafa, 2021) dengan judul “Prediction of heart disease and classifiers' sensitivity analysis”,

Penyakit jantung adalah salah satu penyakit yang paling umum saat ini, dan diagnosis dini penyakit semacam itu merupakan tugas penting bagi banyak penyedia layanan kesehatan untuk mencegah pasien mereka untuk penyakit semacam itu dan untuk menyelamatkan nyawa. Set berisi 76 atribut termasuk atribut kelas, untuk 1025 pasien yang dikumpulkan dari Cleveland, Hongaria, Swiss, dan Long Beach, tetapi dalam makalah ini, hanya subset dari 14 atribut yang digunakan, dan setiap atribut memiliki nilai set yang diberikan. Hal ini menunjukkan bahwa menggunakan algoritma klasifikasi yang berbeda untuk klasifikasi dataset HD memberikan hasil yang sangat menjanjikan dalam hal akurasi klasifikasi untuk K-NN ($K = 1$), pohon keputusan J48 dan pengklasifikasi JRip dengan akurasi klasifikasi 99,7073, 98,0488 dan 97,2683% masing-masing.

(17) Menurut (Priya et al., 2020) dengan judul “Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithm”,

Di dunia modern yang serba cepat, masalah kesehatan psikologis seperti di antara massa. Dalam makalah ini, prediksi kecemasan, depresi dan stres dibuat menggunakan algoritma pembelajaran mesin. Di Kecemasan, depresi dan stres telah menjadi sangat umum di antara massa. Dalam makalah ini, prediksi kecemasan, depresi dan stres dibuat menggunakan algoritma pembelajaran mesin. Kecemasan, depresi dan stress itu Diprediksi seperti yang terjadi pada lima tingkat keparahan oleh lima algoritma pembelajaran mesin yang berbeda - karena ini sangat akurat, mereka sangat cocok untuk memprediksi masalah psikologis. Setelah menerapkan metode berbeda Ditemukan bahwa kelas tidak seimbang dalam confusion matrix. Dengan demikian, ukuran skor f1 ditambahkan, yang membantu mengidentifikasi yang terbaik. mengungkapkan bahwa algoritma juga sangat sensitif terhadap hasil negatif. model akurasi di antara lima algoritma yang diterapkan sebagai pengklasifikasi *Random Forest*. Selanjutnya, parameter spesifisitas mengungkapkan bahwa algoritma juga sangat sensitif terhadap hasil negatif.

(18) Menurut (Alkushayni et al., 2022) dengan judul “The Application of Differing Machine Learning Algorithms and Their Related Performance in Detecting Skin Cancers and Melanomas”,

Kanker kulit, dan bentuk melanoma yang kurang umum, adalah penyakit yang mempengaruhi berbagai macam orang. Karena biasanya terdeteksi pada awalnya dengan inspeksi visual, itu membuat kandidat yang baik untuk

penerapan pembelajaran mesin. Kami menguji kombinasi model, termasuk convolutional neural networks (CNNs) dan berbagai lapisan manipulasi data, seperti penerapan fungsi Gaussian dan pemangkasan gambar untuk meningkatkan akurasi. Kami juga membuat model data yang lebih tradisional, termasuk klasifikasi support vector, K-nearest neighbor, Naïve Bayes, random forest dan gradient boosting algorithms, dan membandingkannya dengan model berbasis CNN yang telah kami buat.

(19) Menurut (TR et al., 2022) dengan judul “Predictive Analysis of Heart Diseases With Machine Learning Approaches”,

Machine Learning (ML) digunakan di sektor kesehatan di seluruh dunia. Metode ML membantu dalam perlindungan penyakit jantung, gangguan alat gerak dalam kumpulan data medis. Penemuan data penting tersebut membantu para peneliti mendapatkan wawasan berharga tentang bagaimana memanfaatkan diagnosis dan pengobatan mereka untuk pasien tertentu. Sekitar 14 dari 76 properti ini dipilih untuk pengujian, yang diperlukan untuk memvalidasi kinerja metode yang berbeda. Analisis ini didasarkan pada metode *supervised learning*, yaitu, Naive Bayes, SVM, regresi Logistik, Pengklasifikasi Pohon Keputusan, Hutan Acak, dan K- Tetangga Terdekat. Hasil eksperimen menunjukkan kekuatan KNN dengan delapan tetangga untuk menguji efektivitas, sensitivitas, presisi, dan akurasi, F1-score; dibandingkan dengan metode lain, yaitu, Naive Bayes, SVM (Linear Kernel), *Decision Tree Classifier* dengan 4 atau 18 fitur, dan pengklasifikasi *Random Forest*.

(20) Menurut (Perveen et al., 2019) Dengan judul ” Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques”,

Analisis kami bertentangan dengan pandangan bahwa diabetes umumnya dikaitkan dengan rendahnya tingkat high-density lipoprotein (HDL). Sebaliknya, hasil kami menunjukkan bahwa peningkatan kadar HDL berkorelasi positif dengan onset diabetes, terutama pada wanita. Dataset yang digunakan dalam penelitian ini berisi 667.907 catatan untuk periode mulai dari 2003 hingga 2013. Ini dicapai rata-rata 79% kinerja karakteristik operasi penerima dengan peningkatan tingkat positif sejati. Hasil makalah ini menyarankan penelitian lebih lanjut untuk mengklarifikasi signifikansi patofisiologis HDL dan jalur dalam pengembangan diabetes.

(21) Menurut (Kohsasih & Situmorang, 2022) Dengan judul “Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular”,

Penyakit Cerebrovascular atau stroke merupakan salah satu penyebab utama kematian di dunia. stroke adalah penyakit yang disebabkan oleh gangguan pada pembuluh darah yang mensuplai darah ke otak. Tujuan dalam penelitian ini yaitu untuk membandingkan akurasi dan kinerja dua algoritma untuk memprediksi Penyakit Cerebrovascular atau stroke. Berdasarkan hasil penelitian didapatkan bahwa algoritma C4.5 memperoleh tingkat akurasi yang lebih tinggi yaitu 95% sedangkan algoritma Naive Bayes memperoleh tingkat akurasi 91%.

(22) Menurut (Utomo & Mesran, 2020) Dengan judul “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung”,

Penyakit jantung merupakan salah satu penyakit dengan tingkat kematian yang tinggi, tercatat 12 juta jiwa kematian setiap tahunnya di seluruh dunia. Hal ini yang menyebabkan perlu diagnosa dini untuk mengetahui penyakit jantung tersebut. Penelitian ini melakukan perbandingan tingkat akurasi dengan menggunakan algoritma C5.0 dan algoritma Naïve Bayes Classifier (NBC), hasil yang didapatkan baik sesudah dan sebelum reduksi adalah algoritma Naïve Bayes Classifier (NBC) memiliki kinerja yang lebih baik dari pada algoritma C5.0.

(23) Menurut (Tiana & Wahyuni, 2020) Dengan judul “Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara”,

Kanker payudara atau Carcinoma Mammarum adalah pertumbuhan sel yang tidak terkendali pada kelenjar penghasil susu (lobular), saluran kelenjar dari lobular ke puting payudara (duktus), dan jaringan penunjang payudara yang mengelilingi lobular, duktus, pembuluh darah dan pembuluh limfe, tetapi tidak termasuk kulit payudara. dapat ditarik kesimpulan bahwa klasifikasi kekambuhan pasien kanker payudara menggunakan algoritma Naïve Bayes 72,7%. Proses imputasi berdampak baik pada algoritma Naïve Bayes yaitu penambahan nilai akurasi sebesar 1,02%.

(24) Menurut (Sulastri et al., 2020) Dengan judul “Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network”,

Hepatitis merupakan penyakit yang diderita oleh banyak orang, bahkan bisa menyebabkan kematian. Prediksi awal dapat mencegah kematian tersebut yaitu dengan mengumpulkan data pasien hepatitis yang dilihat dari faktor - faktornya. Faktor-faktor tersebut antara lain Protime, Alk Phosphat, Albumin, Bilirubin dan Usia.

(25) Menurut (Muthohhar & Prihanto, 2023) Dengan judul “Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung”,

Penyakit jantung menjadi penyebab tertinggi kematian didunia setelah stroke. Dengan kemajuan teknologi yang sangat pesat mendorong berbagai penelitian untuk mempermudah dalam mendiagnosa penyakit jantung menggunakan klasifikasi machine learning. Model klasifikasi yang digunakan yaitu Decision Tree, Naive Bayes dan Random Forest Classifier. Berdasarkan rata-rata hasil pengujian model decision tree menggunakan evaluasi random search mendapatkan nilai 0.844 dan grid search 0.84. Kemudian hasil pengujian model naïve bayes tidak ada perbedaan antara evaluasi menggunakan random search maupun grid search yaitu sebesar 0.85. Selanjutnya hasil pengujian model random forest classifier menggunakan evaluasi random search mendapatkan nilai 0.852 dan grid search 0.868.

(26) Menurut (Dwi Yulianto et al., 2023) Dengan judul “Penerapan Machine Learning Dalam Analisis Stadium Penyakit Hati Untuk Proses Diagnosis dan Perawatan”, Penyakit hati merupakan penyakit yang sudah lama ada dan cukup umum di masyarakat. Penyakit ini terjadi dikarenakan organ hati tidak dapat bekerja dengan optimal dikarenakan adanya peradangan atau virus. Oleh karena itu salah satu cara yang digunakan untuk mengetahui penyakit hati adalah dengan melakukan tes darah di laboratorium sehingga mendapatkan informasi berupa kadar enzim, akan tetapi tes darah di laboratorium memerlukan biaya yang cukup mahal sehingga prediksi menggunakan machine learning sangat diperlukan untuk kasus ini, karena gejala pada penyakit hati perlu penanganan secara cepat. Perhitungan akurasi dilakukan menggunakan algoritma *Random Forest*, *Decision Tree*, *Naive Bayes*, *Logistic Regression*, *Support Vector Machine*, *KNN*, *Gradient Boosting* dan *Artificial Neural Network*, Hasil perhitungan akurasi menggunakan algoritma *Random Forest* dengan PCA antara sebelas dan dua fitur mengalami penurunan sebesar 0,6%, sedangkan menggunakan *features selection* meningkat sebesar 0,7%, didapati akurasi tertinggi menggunakan algoritma *Random Forest* dengan 2 fitur sebesar 72,2%.

(27) Menurut (Taufik et al., 2023) Dengan judul “Teknologi Sistem Informasi dan Sistem Komputer TGD”,

ISPA (Infeksi Saluran Pernafasan Akut) adalah suatu penyakit gangguan saluran pernapasan yang dapat menimbulkan berbagai spektrum penyakit mulai dari penyakit tanpa gejala, infeksi ringan sampai penyakit yang parah dan mematikan akibat faktor lingkungan. Banyak anak yang menderita penyakit ISPA sementara jumlah dokter anak (pakar) terbatas. Keterbatasan jam kerja atau praktik dokter, jarak rumah sakit atau klinik yang jauh dan biaya yang mahal menjadi faktor permasalahan. Dari permasalahan tersebut dibutuhkanlah suatu sistem yang dapat mengadopsi keilmuan seorang pakar dalam hal ini yaitu dokter spesialis anak, dimana sistem tersebut adalah sistem pakar menggunakan metode *naïve bayes*. Hasil yang diperoleh adalah terciptanya sistem pakar yang dapat mendiagnosa gangguan tidur pada remaja berbasis website yang dapat digunakan masyarakat dimana pun dan kapan pun selama terhubung dengan jaringan internet.

(28) Menurut (Riany & Testiana, 2023a) Dengan judul “Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes”,

Penyakit jantung koroner merupakan salah satu penyakit penyebab utama kematian pada semua kelompok umur setelah stroke. Penyakit jantung koroner adalah penyakit yang disebabkan oleh plak yang menumpuk di arteri koroner yang memasok oksigen ke otot jantung. Pentingnya deteksi dini gejala penyakit jantung koroner merupakan pencegahan awal. Oleh karena itu, penelitian ini bertujuan untuk menganalisis data terkait penyakit jantung koroner untuk menentukan apakah pasien tersebut sepuluh tahun mendatang menderita penyakit jantung koroner dan untuk mengetahui performa algoritma *Naïve Bayes*. Hasil penelitian menunjukkan bahwa klasifikasi dengan menggunakan algoritma *Naïve Bayes* dapat diterapkan dalam mengklasifikasi data penyakit jantung koroner menghasilkan akurasi sebesar 79,10% dalam kategori cukup akurat.

(29) Menurut (Riany & Testiana, 2023b) Dengan judul “Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes”,

Stroke merupakan suatu gangguan fungsi otak baik lokal maupun menyeluruh yang terjadi secara mendadak progresif, dan cepat akibat gangguan peredaran darah otak non traumatik yang berlangsung lebih dari 24 jam atau berakhir dengan kematian. Stroke juga merupakan salah satu jenis penyakit yang paling mematikan di Indonesia. Hasil penelitian menunjukkan

bahwa klasifikasi dengan algoritma *Naïve Bayes* dapat diterapkan dalam mengklasifikasi data penyakit stroke menghasilkan nilai akurasi sebesar 92,48% dalam kategori *Good Classification*.

30. Menurut (Ida Maryani, 2022), Dengan judul “**Penerapan Algoritma Naïve Bayes Untuk Penentuan Diagnosa Obesitas Pada Peserta Sosialisasi Deteksi Dini Penyakit Tidak Menular (PTM)**”, Obesitas merupakan penumpukan lemak yang berlebih akibat ketidakseimbangan asupan energi (energy intake) dengan energi yang digunakan (energy expenditure) dalam waktu lama. Obesitas dapat menyebabkan berbagai penyakit tidak menular, diantaranya penyakit jantung, stroke, diabetes, tekanan darah tinggi, batu empedu, masalah pernapasan, kanker, *osteoarthritis*, termasuk *infertilitas*. Aplikasi yang dibangun berupa *prototype* yang memanfaatkan Bahasa pemrograman PHP. Penelitian ini menggunakan data pemeriksaan kesehatan peserta sosialisasi deteksi dini resiko PTM dengan jumlah peserta 60 orang. Variabel yang digunakan ada 10 yaitu Jenis Kelamin, Usia, Tinggi Badan, Berat Badan, Lingkar Perut, Kurang Aktifitas Fisik, Pola Makan Gula Berlebih, Garam Berlebih, Lemak Berlebih dan Kurang Makan dan Buah sedangkan kelasnya ada 2 yaitu diagnosa obesitas dan normal. Berdasarkan hasil perhitungan akurasi dengan *Confusion Matrix* didapatkan nilai akurasi sebesar 86,6%.

Tabel 2. 2 Tinjauan Studi

No	Peneliti	Judul	Jurnal	Kontribusi
1.	(Sahar, 2020)	Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung	Indonesian Journal of Data and Science (IJODAS) ISSN: 2715-9930 Vol 1, No 3, Decembr 2020, pp. 79-86	Kontribusi penelitian ini adalah Menggunakan metode Naive Bayes untuk klasifikasi

No	Peneliti	Judul	Jurnal	Kontribusi
			https://jurnal.yoctobrain.org/index.php/ijodas/article/view/20/19	
2.	(Ridwan A, 2020)	Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus	Jurnal Sistem Komputrdan Kecerdasan Buatan Volume IV, Nomor 1, pp.15-21, September 2020, ISSN: 2621-2927, https://jurnal.tau.ac.id/index.php/skom-kb/article/view/169/146	Kontribusi penelitian ini adalah Menggunakan metode Naive Bayes untuk klasifikasi.
3.	(Coastera, Yusa, Lediwara, & Sari, 2020)	Analisis Performa Algoritma Naïve Bayes Untuk Penentuan Kelayakan	Jurnal Voice Of Information, Vol 9, No 1 (2020),	Sebagai acuan peneliti untuk penentuan kelayakan pendonor darah

No	Peneliti	Judul	Jurnal	Kontribusi
		Pendonor Darah	ISSN: 2579-3489, https://voistatikmalaya.ac.id/index.php/voistatikmalaya/article/view/185	
4.	(Doni et al., 2021)	penerapan data mining untuk klasifikasi penyakit <i>hepatocellular carcinoma</i> menggunakan algoritma naïve bayes	JURNAL RESPONSI F, Vol.3 No.1 Februari 2021, pp. 12~19, E-ISSN:2685-6964, http://ejournal.ars.ac.id/index.php/jti/article/view/403/275	Sebagai acuan peniliti untuk penerapan data mining untuk klasifikasi.
5.	(Valdiviezo-Diaz et al., 2019)	A Collaborative Filtering Approach Based on Naïve Bayes Classifier	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=8787761	Kontribusi penelitian ini adalah Menggunakan metode Naive Bayes untuk klasifikasi.

No	Peneliti	Judul	Jurnal	Kontribusi
6.	(Jackins et al., 2021)	AI-based smart prediction of clinical disease using random forest classifer and Naive Bayes	The Journal of Supercomp uting, https://link.springer.com/content/pdf/10.1007/s11227-020-03481-x.pdf?pdf=button	Sebagai acuan perhitungan confusion matrix.
7.	(Utomo & Mesran, 2020)	Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung	jurnal media informatika budidarma, Volume 4, Nomor2, April 2020, Page 437-444, ISSN:2548-8368 http://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2080	Untuk klasifikasi menggunakan metode naïve bayes

No	Peneliti	Judul	Jurnal	Kontribusi
8.	(Setiyani & Prasetyaningrum, 2021)	penerapan metode naive bayes classifier pada sistem pakar diagnosa penyakit lambung	Jurnal Sistem Informasi Dan Bisnis Cerdas (SIBC) Vol. 14, No. 2. Agustus 2021	Sebagai acuan untuk perhitungan Naive Bayes.
9.	(Byna & Basit, 2020)	PENERAPAN METODE ADABOOST UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE DENGAN ALGORITMA NAIVE BAYES	Jurnal SISFOKOM (Sistem Informasi dan Komputer), Volume 09, Nomor 03, PP 407-411, ISSN : 2087-3891, http://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/1023	Sebagai acuan perhitungan confusion matrix.
10.	(Maulana et al., 2022)	Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit	jurnal sistem informasi tgd, Volume 1, Nomor 4,	Kontribusi penelitian ini adalah Menggunakan metode Naive Bayes.

No	Peneliti	Judul	Jurnal	Kontribusi
		Stunting Pada Balita	Juli 2022, Hal 425-438 P-ISSN : 2828-1004 ; E-ISSN : 2828-2566 https://ojs.trigunadharma.ac.id/index.php/jsi	
11.	(Almustafa, 2021)	Prediction of chronic kidney disease using different classification algorithms	Informatics in Medicine Unlocked, Volume 24, tahun 2021,p 100631, ISSN : 23529148, https://www.sciencedirect.com/science/article/pii/S2352914821001210?via%3Dihub	Kontribusi penelitian ini adalah Menggunakan metode Naive Bayes.
12.	(TR et al., 2022)	PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE	Malaysian Journal of Computer Science, Volume 2022, tahun	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode Naive bayes

No	Peneliti	Judul	Jurnal	Kontribusi
		LEARNING APPROACHES	2022, p 132-148, ISSN : 01279084, https://www.mendeley.com/catalogue/45ff4db3-336a-338c-9710-5394146c52ea/?utm_source=desktop	
13	(Alotaibi, 2019)	Implementation of machine learning model to predict heart failure disease	International Journal of Advanced Computer Science and Applications, Volume 10, Tahun 2019, p 261-268, ISSN : 21565570, https://www.mendeley.com/catalogue/a42eb5bb-134d-3e4a-aeb8-1b25f5b47b	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode Naïve bayes

No	Peneliti	Judul	Jurnal	Kontribusi
			88/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7B604c51f7-db13-4943-a281-1f34bcc07505%7D	
14.	(Priya et al., 2020)	Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms	Procedia Computer Science, Volume 167, Tahun 2020, p 1258-1267, ISSN 18770509, https://www.sciencedirect.com/science/article/pii/S1877050920309091?pes=vor	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode Naïve bayes
15.	(Alkhushayni et al., 2022)	The Application of Differing Machine Learning	Journal of Skin Cancer, Volume	Kontribusi Penelitian ini adalah penerapan metode Naïve bayes

No	Peneliti	Judul	Jurnal	Kontribusi
		Algorithms and Their Related Performance in Detecting Skin Cancers and Melanomas	2022, Tahun 2022, ISSN : 20902913, https://www.mendeley.com/catalogue/d1693e48-0c9d-3542-9f6d-de5458c5eccca/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7B20319c19-e4aa-4fe7-b92a-68ab0e5c3f8a%7D	
16.	(Maheswari & Pitchai, 2019)	Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm	Current Medical Imaging Formerly Current Medical Imaging Reviews, Volume 15,	Kontribusi Penelitian ini adalah penerapan metode Naïve bayes

No	Peneliti	Judul	Jurnal	Kontribusi
			Tahun 2019, p 712-717, ISSN 15734056, http://www.eurekaselect.com/160671/article	
17.	(Kamel et al., 2019)	Cancer Classification Using Gaussian Naive Bayes Algorithm	Proceedings of the 5th International Engineering Conference, IEC 2019, Tahun 2019, p 165-170, ISSN 9781728143774, https://ieeexplore.ieee.org/document/8950650	Kontribusi Penelitian ini adalah penerapan metode Naïve bayes
18.	(Zaw et al., 2019)	Brain tumor detection based on Naïve Bayes classification	Proceeding - 5th International Conference on Engineering, Applied	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode Naïve bayes

No	Peneliti	Judul	Jurnal	Kontribusi
			Sciences and Technology , ICEAST 2019, Tahun 2019, p 1-4, ISSN 9781728100678, https://ieeexplore.ieee.org/document/8802562	
19.	(TR et al., 2022)	PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES	Malaysian Journal of Computer Science, Volume 2022, Tahun 2022, p 132-148, ISSN 01279084 https://ejournal.um.edu.my/index.php/MJCS/article/view/35980	Kontribusi Peneltian ini adalah penerapan pembelajaran mesin metode Naïve bayes
20.	(Kohsasih &	Analisis Perbandingan Algoritma C4.5	Jurnal Informatika, Volume 9,	

No	Peneliti	Judul	Jurnal	Kontribusi
	Situmoraning, 2022)	Dan Naïve Bayes Dalam Memprediksi Penyakit <i>Cerebrovascular</i>	Tahun 2022, p 13-17, ISSN: 2355-6579, https://www.researchgate.net/publication/359649144	Kontribusi Peneltian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
21.	(Utomo & Mesran, 2020)	Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung	JURNAL MEDIA INFORMATIKA BUDIDARMA, Volume 4, Tahun 2020, p 437, ISSN : 2548-8368, http://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2080/1541	Kontribusi Peneltian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
22.	(Tiana & Wahyuni, 2020)	Hasil Analisis Teknik Data Mining dengan Metode <i>Naive Bayes</i> untuk	Jurnal Sistem Komputer dan Informatika	

No	Peneliti	Judul	Jurnal	Kontribusi
		Mendiagnosa Penyakit Kanker Payudara	(JSON), Volume 1, Tahun 2020, p 130, e-ISSN 2685-998X, http://www.ejurnal.stmik-budidarma.ac.id/index.php/JSON/article/view/1766/1448	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
23.	(Sulastriet al., 2020)	Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network	Dinamik, Volume 24, Tahun 2020, p 82-91, ISSN : 0854-9524, https://www.unisbank.ac.id/ojs/index.php/fti1/article/view/7867/2820	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
24.	(Muthohhar & Prihanto, 2023)	Analisis Perbandingan Algoritma Klasifikasi	Journal of Informatics and Computer Science	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>

No	Peneliti	Judul	Jurnal	Kontribusi
		untuk Penyakit Jantung	(JINACS), Volume 04, Tahun 2023, p 298-304, https://ejournal.unesa.ac.id/index.php/jinacs/article/view/50952/41725	
25.	(Dwi Yulianto et al., 2023)	Penerapan Machine Learning Dalam Analisis Stadium Penyakit Hati Untuk Proses Diagnosis dan Perawatan	RESOLUSI : Rekayasa Teknik Informatika dan Informasi, Volume 3, Tahun 2023, p 303-313, http://www.djournals.com/resolusi/article/view/709/470	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
26.	(Ida Maryani, 2022)	Penerapan Algoritma Naïve Bayes Untuk Penentuan Diagnosa Obesitas Pada Peserta	Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains, Volume 13, Nomor 2,	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>

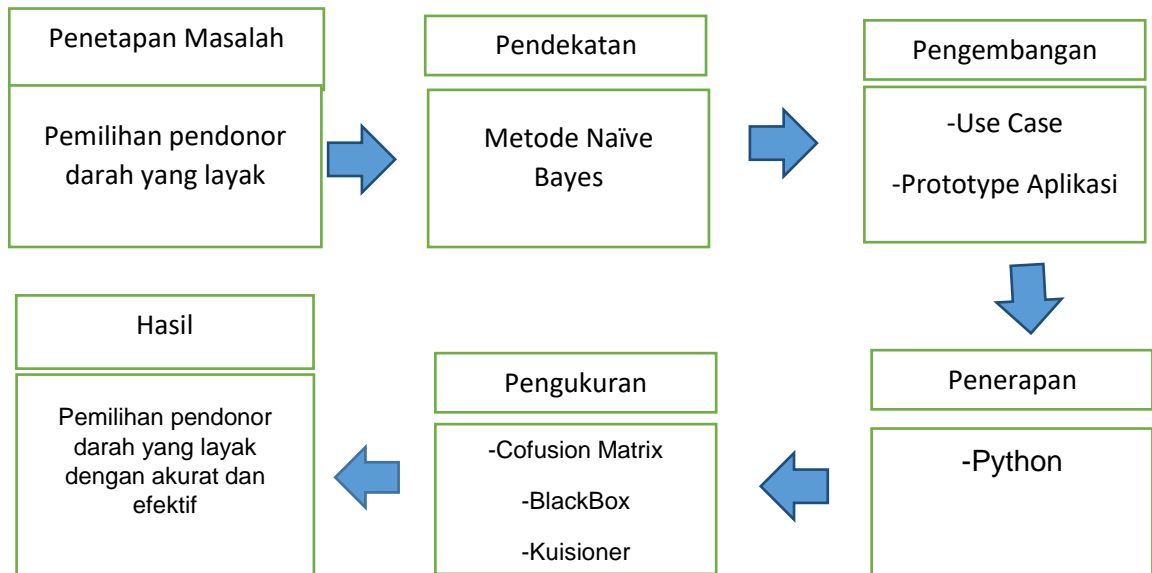
No	Peneliti	Judul	Jurnal	Kontribusi
		Sosialisasi Deteksi Dini Penyakit Tidak Menular (PTM)	July 2023, Page. 234-248, ISSN 2087-3891, http://teknois.stikombinaniaga.ac.id/index.php/JBS/article/view/200/145	
27.	(Taufik et al., 2023)	Penerapan Metode Naive Bayes Untuk Mendiagnosa Penyakit ISPA Pada Anak	Teknologi Sistem Informasi dan Sistem Komputer TGD, Volume 6, Tahun 2023, p 351-361, DOI : 10.53513/jurisi.v1i4.5446	Kontribusi Peneltian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
28.	(Riany & Testiana, 2023a)	Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan	MDP Student Conference , Volume 2, Tahun 2023, p 297-305, DOI : 10.35957/m	Kontribusi Peneltian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>

No	Peneliti	Judul	Jurnal	Kontribusi
		Algoritma Naïve Bayes	dp-sc.v2i1.4388	
29.	(Riany & Testiana, 2023b)	Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes	Jurnal SAINTEKO M, Volume 13, Tahun 2023, p 42-54, ISSN: 2088-1770, DOI : 10.33020/saintekom.v13i1.352	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>
30.	(Perveen et al., 2019)	Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques	IEEE Access, Volume 7, Tahun, 2019, p 1365-1375, DOI 10.1109/ACCESS.2018.2884249, https://ieeexplore.ieee.org/ielx7/6287639/8600701/08585030.pdf	Kontribusi Penelitian ini adalah penerapan pembelajaran mesin metode <i>Naïve bayes</i>

Berdasarkan penelitian rujukan diatas, maka penelitian ini dilakukan dengan permasalahan yang serupa serta dengan metode penyelesaian yang sama yaitu Algoritma Naïve Bayes. Tetapi perbedaan penelitian kali ini dilakukan dengan menggunakan variabel yang berbeda dengan jurnal sebelumnya yaitu,

jk, hb, Berat badan, Sistolik dan distolik, Jenis pendonor, Baru/ulang dan variabel prediktor yaitu status. Prototype sistem aplikasi yang berbasis website sebagai pengembangannya, lalu diuji kembali dengan blackbox.

C. Kerangka Berfikir



Gambar 2. 9 Kerangka Berfikir

Kerangka pemikiran pada gambar 2.9 dapat dijelaskan sebagai berikut:

- Identifikasi masalah terkait yaitu belum akurat dalam pemilihan pendonor darah yang layak.
- Pendekatan penelitian menggunakan Algoritma Naïve Bayes
- Pengembangan yaitu tahap melakukan analisis dan perancangan gambar untuk use case dan Prototype Aplikasi.
- Kemudian pengembangan menentukan bahasa pemrograman dengan konstruksi Python dan melakukan evaluasi serta validasi.
- Pengukuran menggunakan cofusion matrix, BlackBox dan Kuisisioner.
- Hasil yaitu Prototype yang menampilkan Hasil pemilihan pendonor darah yang layak dengan akurat dan efektif.

D. Hipotesis Penelitian

Dalam penerapannya Naïve Bayes merupakan algoritma klasifikasi yang banyak digunakan karena memiliki kelebihan yang lebih baik dibandingkan algoritma lainnya yang memungkinkan algoritma naïve bayes dapat menemukan dan menentukan suatu kemungkinan ataupun peluang di masa yang akan datang berdasarkan dari riwayat atau histori data di masa lampau, dimana nantinya algoritma naïve bayes ini akan mempelajari kriteria kelayakan pendonor darah pada periode sebelumnya sehingga diharapkan dapat memperkirakan kelayakan pendonor darah dan mempengaruhi keputusan di dalam penentuan kelayakan pendonor darah berdasarkan kriteria pendonor darah yang dibutuhkan. Serta didukung dengan keberhasilan penelitian sebelumnya yang berjudul “Analisis Performa Algoritma Naïve Bayes Untuk Penentuan Kelayakan Pendonor Darah” (coastera et al., 2020) dan berdasarkan analisa data dari permasalahan diatas. Maka dapat diduga dan ditetapkan hipotesis berupa metode dan algoritma Naïve Bayes tepat dan efektif untuk dimanfaatkan dan digunakan di dalam penentuan Penerapan Metode Naïve Bayes untuk Penentuan kelayakan pendonor darah di PMI.