

BAB II

LANDASAN TEORI

2.1 DATA MINING

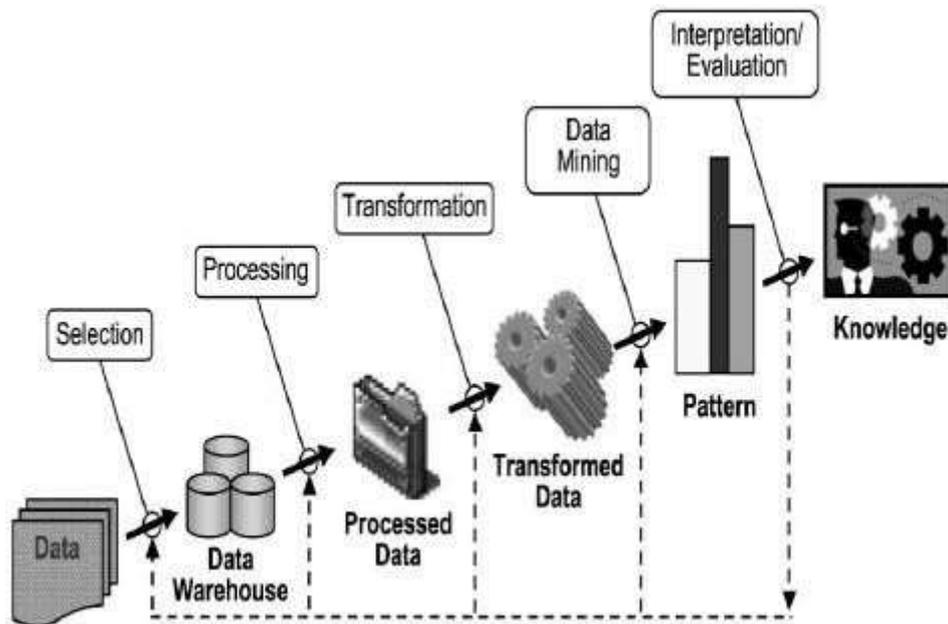
Data mining adalah proses mencari pola dan informasi dalam data dengan menggunakan metode tertentu sesuai dengan kebutuhan. Metode dan algoritma *data mining* sangat bervariasi. Pemilihan metode dan algoritma yang tepat sangat tergantung pada tujuan dan proses penemuan *knowledge discovery in databases* (KDD). KDD adalah istilah lain yang memiliki arti yang sama dengan data mining[12].

Data mining merupakan penggalian atau pengumpulan informasi yang berguna dari sekumpulan data. Informasi yang biasanya dikumpulkan adalah pola tersembunyi dalam data, hubungan antar elemen data, atau model untuk tujuan prediksi data[13].

Berbagai pendapat diatas, maka dapat disimpulkan bahwa *data mining* merupakan sebuah informasi untuk menambang pengetahuan dari sekumpulan data sehingga menjadi sebuah informasi yang bermanfaat.

2.1.1 Tahap-Tahap Data Mining

Data mining biasa disebut juga dengan *Knowledge in database* (KDD) terdapat beberapa proses seperti gambar dibawah ini[14] :



Gambar 2.1 Proses Data Mining

1. *Data selection*

Data selection merupakan proses pengambilan data yang berhubungan dengan analisis. Pada tahapan ini dilakukan teknik perolehan sebuah pengurangan representasi data

2. *Preprocessing/Cleaning,*

Sebelum proses data mining dilakukan, perlu melakukan *cleaning* untuk menghilangkan noise, membuang duplikasi data, memeriksa data yang tidak konsisten dan memperbaiki kesalahan pada data.

3. *Transformation,*

Tranformasi merupakan suatu proses perubahan yang melalui proses bertahap sampai pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

4. *Data mining,*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. *Evaluation*,

Tahap ini merupakan bagian dari proses KDD, mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

6. *Pattern evaluation*

Pattern evaluation merupakan proses mengidentifikasi pola yang hasiln dari tahapan ini dapat berupa pola unik dalam bentuk analisis pengetahuan.

7. *Knowledge presentation*

Menampilkan mined knowledge kepada pengguna.

2.1.2 Pengelompokan Data Mining

Data mining dibagi menjadi enam pengelompokan berdasarkan tugas yang dapat dilakukan yaitu sebagai berikut[15]:

1. Deskripsi

Terkadang peneliti dan analisis mencari cara untuk menggambar suatu pola dan kecenderungan dalam suatu data. Deskripsi dari kecenderungan pola sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Klasifikasi

Pada klasifikasi, terdapat suatu target variabel kategori, contohnya seperti penggolongan yang terdapat sebanyak tiga item, yaitu pendapatan rendah, pendapatan sedang, dan pendapatan tinggi.

3. Estimasi

Estimasi mempunyai kemiripan dengan klasifikasi yang membedakan hanya variabel target estimasi lebih mengarah ke numerik daripada kategori.

4. Prediksi

Hampir sama dengan klasifikasi dan estimasi, bedanya dalam prediksi nilai nilai dari hasil akan ada di masa depan. Metode yang ada di klasifikasi dan estimasi dapat digunakan untuk prediksi.

5. Pengklusteran

Kluster merupakan sekumpulan record atau pengamatan yang memiliki kemiripan satu dengan yang lainnya. Berbeda dengan klasifikasi, pada pengklusteran tidak memiliki target, pengklusteran tidak mencoba melakukan klasifikasi, mengestimasi, ataupun memprediksi, akan tetapi, pengklusteran melakukan pembagian keseluruhan kelompok yang mempunyai kemiripan.

6. Asosiasi

Tugas asosiasi adalah menemukan atribut yang muncul dalam satu waktu. Dalam perDuniaan bisnis sering disebut dengan analisis keranjang belanja atau *market basket analysis*.

2.2 KLASIFIKASI

Klasifikasi adalah proses menemukan pola atau karakteristik yang menggambarkan dan membedakan kelas data atau konsep. Model diturunkan berdasarkan analisis dataset pelatihan, yaitu objek data yang pengidentifikasi kelasnya diketahui. Pola ini digunakan untuk memprediksi pengidentifikasi kelas

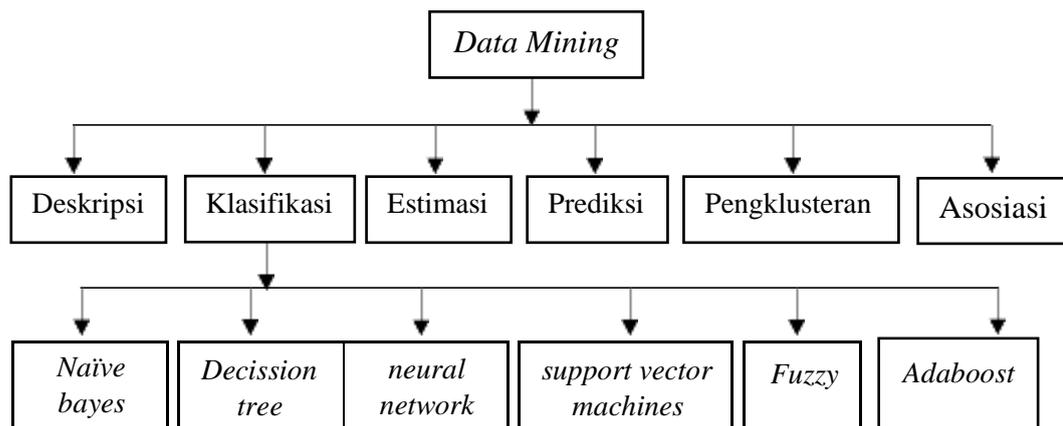
untuk objek yang pengidentifikasi kelasnya tidak diketahui[16]. Klasifikasi mempunyai dua jenis yaitu :

1. Deskriptif

Yakni model klasifikasi yang mempunyai fungsi sebagai alat untuk memperjelas agar dapat membedakan objek dari suatu kelas yang berbeda-beda.

2. Prediktif

Yaitu model klasifikasi yang dapat digunakan sebagai prediksi label kelas untuk record yang belum diketahui.



Gambar 2.2 Metode Data Mining

Dari bagan tersebut, data mining dibagi menjadi beberapa kelompok mulai dari deksripsi, klasifikasi, estimasi, prediksi, pengklusteran, dan asosiasi. Selain pengelompokan, data mining dibagi menjadi beberapa metode, terutama metode yang terdapat pada klasifikasi diantaranya yaitu metode *naïve bayes*, *decision tree*, *neural network*, *support vector machines*, *fuzzy*, dan *adaboost*.

2.3 NAÏVE BAYES

Naive Bayes classifier (NBC) merupakan salah satu metode klasifikasi yang sifatnya independen atau berdiri sendiri dan tidak ada sangkut pautnya dengan variabel yang lain[17].

Naive Bayes Classifier adalah metode klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas untuk menjadi anggota suatu kelas[18]. Probabilitas dilakukan untuk setiap kelas, hasil akhir dianggap nilai tertinggi, Jadi. Algoritma ini tampaknya cukup baik untuk menentukan probabilitas dalam menentukan hasil[19]. *Teorema bayes* memiliki bentuk umum yang dituliskan dengan persamaan 2.1.

$$P(C_i | X) = P(X | C_i)P(C_i) \dots \dots \dots (i)$$

Keterangan:

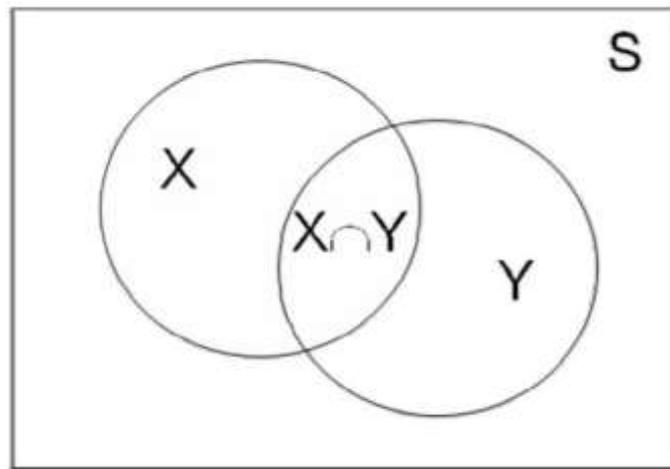
X : Kriteria suatu kasus berdasarkan masukan

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

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

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

C_i P(C_i) : Probabilitas label kelas C_i



Gambar 2. 3 Teorema Bayes

Naïve Bayes merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas.

Untuk menjelaskan metode *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naïve Bayes* di atas disesuaikan dengan persamaan 2.2.

$$P(C|F1 \dots Fn) = \frac{P(C).P(F1...Fn)}{P(F1...Fn)} \dots \dots \dots (2.2)$$

Dimana Variabel C merepresentasikan kelas, sementara *variable F1...Fn* merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga

evidence). Oleh karena itu, rumus 2.2 diatas dapat pula ditulis secara sederhana dengan persamaan 2.3

$$\text{Posterior} = \frac{\text{pior x likelihood}}{\text{evidence}} \dots \dots \dots (2.3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F1, \dots, Fn)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F1, \dots, Fn) &= P(C)P(F1, \dots, Fn|C) \\ &= P(C)P(F1|C)P(F2, \dots, Fn|C, F1) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2) \dots P(Fn|C, F1, F2, F3, \dots, Fn-1) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk $(F1, F2, \dots, Fn)$ saling bebas (independen) satu sama lain. Dari asumsi tersebut, maka berlaku suatu persamaan 2.4.

$$P(C_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i \cap F_j)}{P(F_j)} = P(F_i) \dots \dots \dots (2.4)$$

Untuk $i \neq j$, sehingga :

$$P(F_n|C, F_n) = P(F_n|C) \dots \dots \dots (2.5)$$

Persamaan di atas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss yang ditulis melalui persamaan 2.6.

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

Di mana:

P : Peluang

X_i : Atribut ke i

x_i : Nilai atribut ke i

Y : Kelas yang dicari

y_i : Sub kelas Y yang dicari

μ : Mean, menyatakan rata-rata dari seluruh atribut

σ : Deviasi standar, menyatakan varian dari seluruh atribut

Penjelasan diatas dapat disimpulkan bahwa algoritma *naive bayes* merupakan salah satu metode algoritma yang memiliki independensi yang kuat (naif) yang digunakan untuk memprediksi kemungkinan dimasa yang akan datang melalui pengalaman dari masa lampau.

2.4 ADABOOST

Adaboost digunakan untuk mengklasifikasi data ke dalam kategorinya masing-masing. *Adaboost* mencari kelas class berdasarkan nilai bobot yang dimiliki ole

class tersebut. Proses ini selanjutnya diulang untuk memperbarui catatan kelas. Di Adaboost, dengan setiap iterasi bobot yang salah, bobot terus meningkat di setiap iterasi[20]. *Adaptive boosting (adaboost)* merupakan salah satu dari beberapa varian pada algoritma *boosting*. *Adaboost* merupakan *ensemble learning* yang sering digunakan pada algoritma *boosting*. *Boosting* bisa dikombinasikan dengan *classifier* algoritma yang lain untuk meningkatkan performa klasifikasi. Tentunya secara intuitif, penggabungan beberapa model akan membantu jika model tersebut berbeda satu sama lain. *Adaboost* dan variannya telah sukses diterapkan pada beberapa bidang (domain) karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar[21]. Algoritma *Adaboost* akan menghasilkan suatu keputusan dari *weak learner*. Perhitungan algoritma *Adaboost* ini di tunjukan pada persamaan 2.7.

$$\mathbf{F}(\mathbf{x}) = \text{sign} \sum_{t=1}^T \alpha_t h_t(\mathbf{x}) \dots\dots\dots(2.7)$$

$h_t(\mathbf{x}) = \text{weak atau basic classifier}$

$\alpha_t = \text{tingkat pembelajaran (learning rate)}$

$\mathbf{F}(\mathbf{x}) = \text{final classifier}$

Algoritma ini menerima input berupa pembelajaran lemah (*Weak Learn*). Sebuah *integer* t yang menspesifikasi banyaknya iterasi dan data pelatihan sebanyak $N \{(x_1, y_1) \dots\dots\dots (x_n, y_n)\}$. Setelah itu akan dilakukan inialisasi bobot pada setiap elemen dalam fitur *vector*. Inialisasi bobot ini dirumuskan dengan $D_1(i) = 1/(n)$, dimana n adalah jumlah elemen dalam fitur vector. Perhitungan ini juga menggunakan penentuan jumlah iterasi. *Training* dari *weak classifier* h_t menggunakan bobot dari D_1 sehingga di hasilkan $h_t \in \{-1, +1\}$.

Hasil tersebut akan dilihat error dari perhitungannya. Untuk menghitung kesalahan perhitungan *weak classifier* terkait dengan jumlah bobot yang salah di di klasifikasi dituliskan dengan persamaan 2.8.

$$\text{Error } t = \sum_{i=1}^n D_n(i) |h_t(x_i) - y_i| \dots \dots \dots (2.8)$$

Hasil dari perhitungan ini akan dilakukan perhitungan nilai α . Nilai ini adalah bobot yang diterapkan pada masing-masing *classifier* seperti yang di tentukan oleh algoritma *adaboost*, sehingga akan didapatkan hasil akhir yaitu *strong classifier* dari algoritma *adaboost* seperti pada persamaan(2.7). Perhatikan nilai α ini akan ditunjukkan pada persamaan 2.9.

$$\alpha_t = \log\left(\frac{1 - \epsilon_t}{\epsilon_t}\right) \dots \dots \dots (2.9)$$

Cara ini digunakan untuk menggabungkan *output* dari *weak learner* dan menciptakan *strong learner* yang pada akhirnya meningkatkan kekuatan model prediksi. Algoritma *adaboost* ini memiliki beberapa kelebihan yaitu:

- a. *Adaboost* mampu mengurangi nilai bias maupun perbedaan pada *weak classifier*.
- b. *Adaboost* memiliki kemampuan generalisasi yang baik.
- c. Output *adaboost* mengkonversi ke logaritma rasio terdekat.
- d. *Adaboost* dapat dilihat sebagai pemilih fitur dengan strategi berprinsip (minimalisasi dari batas atas *error*).

2.5 WEKA

WEKA merupakan sebuah perangkat lunak yang menerapkan berbagai algoritma *machine learning* untuk melakukan beberapa proses yang berkaitan dengan sistem temu kembali informasi atau *data mining*.

The Waikato Environment for Knowledge Analysis (WEKA) adalah rangkaian perangkat lunak pembelajaran mesin yang ditulis dalam Java yang dikembangkan di Universitas Waikato di Selandia Baru. Perangkat lunak ini memiliki banyak algoritma pembelajaran mesin untuk data mining[22].



Gambar 2.4 Weka

WEKA memiliki implementasi semua teknik pembelajaran untuk klasifikasi dan regresi, yaitu *decision trees*, *rules set*, pengklasifikasian teorema bayes, *Support Vektor Mechines (SVM)*, *logistik dan linier*, *multi layers perceptrons* dan *metode nearest neighbour*.

Weka memiliki kelebihan dibandingkan dengan aplikasi sejenis yang lainnya. Keunggulan weka adalah yang murah karena weka berlisensi *General Public License*, yang berarti gratis. Weka dapat diinstal hampir di semua sistem operasi, sistem operasi tersebut mendukung *java virtual mechine*[13].

2.6 SERANGAN JANTUNG

Aliran darah menuju jantung dan setelah dari jantung sangat di pengaruhi oleh kondisi jantung dan sistem vaskular setiap orang secara keseluruhan. Kondisi jantung seseorang dapat dilihat dari tekanan darah bagian atas (*diastolik*), serta kondisi pembuluh darah yang merupakan jalur aliran darah keseluruh tubuh penyakit dirasakan ketika kondisi jantung tidak optimal dan pembuluh darah menyempit parah. keluhan diidentifikasi sebagai penyakit jantung. Penyakit jantung dapat dibedakan menjadi penyakit jantung sejak lahir (penyakit bawaan) dan penyakit yang tumbuh saat usia tertentu[23]. Gejala pada pasien yang didiagnosis penyakit jantung antara lain nyeri dada (*chest pain*), tekanan darah tinggi (*tresbps*), kolesterol (*chol*), hasil elektrokardiogram (elektrokardiogram istirahat), detak jantung (*thalamus*), FBS (*fasting blood sugar*)[6].

Gejala serangan jantung dapat ditunjukkan seperti kehilangan kesadaran, menderita sesak nafas hingga henti nafas secara cepat, pusing dan pingsan, rasa sakit di daerah perut dan dada sering dikeluhkan terutama pada penderita dengan *infark miokard*. Kondisi selanjutnya tiba-tiba terjatuh karena kehilangan kesadaran[24].

2.7 PENELITIAN SEJENIS

Untuk mendapatkan informasi dan pengetahuan tentang penelitian sejenis, maka peneliti melakukan tinjauan penelitian sejenis, dimana ringkasan tercantum pada tabel 2.1.

Tabel 2.1 Penelitian Sejenis

No	Penulis	Masalah	Metode	Hasil
1.	Agus Byna, Muhammad Basit[10].	Penyakit stroke kebanyakan sembuh dengan kecacatan. Kota Banjarmasin angka stroke mencapai 4.031 kasus yang sembuh dengan gejala sisa. Gejala sisa ini menjadikan penderita stroke di Kota Banjarmasin memiliki kualitas hidup rendah, terlebih lagi dengan rendahnya dukungan keluarga.	<i>Naïve bayes</i> dan <i>Adaboost</i>	Hasil nilai akurasi Naïve Bayes memiliki nilai 0.976 dengan Split data 80/20, nilai akurasi optimasi adaboost dengan naïve bayes senilai 0.981 split data 70/30 kedua model tersebut memiliki diagnosa Excellent Classification, dalam pengujian prediksi penyakit stroke dengan dengan 11 variabel dengan jumlah data 28,500 untuk data training dan 572 untuk data testing.
2.	Mufti Ari Bianto, Kusrini, Sudarmawan[1].	Kurangnya akses untuk mencari informasi tentang penyakit serangan jantung ini yang menyebabkan keamtian.	<i>Naïve Bayes</i>	Hasil subset 1 pada 60 data nilai akurasi 88,62%, presisi 86,37% dan recall 86,46%, subset 2 pada 120 data dengan nilai akurasi 89,04%, presisi 86,56%, dan recall 87,62%, 91,48%,
3.	Achmad Ridwan, Pulung Nurtantio Andono, Catur Supriyanto[25].	Pemantauan status gizi balita sekarang ini memang sudah ada, tetapi masih ditemukan berbagai kendala yang menghambat seperti terlambatnya laporan yang masuk ke dinas kesehatan Kabupaten. Bahkan kadang-	<i>Naïve Bayes</i> dan <i>Adaboost</i>	NBC+ <i>Adaboost</i> menghasilkan <i>Error</i> Klasifikasi sebesar 11.16%. Dibandingkan dengan algoritma NBC yaitu 11.40%. NBC+ <i>Adaboost</i> mampu menurunkan <i>Error</i> . Klasifikasi sebesar 0,24%. Algoritma NBC+ <i>Adaboost</i> juga menghasilkan akurasi sebesar 88,84%. Untuk itu

		kadang hanya tercatat kasus untuk beberapa Puskesmas saja sehingga jumlah kasus gizi buruk yang ada tidak mencerminkan keadaan sesungguhnya.		pada akhir penelitian disimpulkan bahwa algoritma NBC+Adaboost lebih meningkatkan akurasi klasifikasi Status Gizi Balita berdasarkan Indeks. Antropometri NBC, dengan peningkatan akurasi sebesar 0,24%.
4.	M. Sabransyah, Yuki Novia Nasution, Fidia Deny Tisna Amijaya [23].	penyakit jantung bisa dikenali sejak dini, tapi banyak orang yang belum dibekali pengetahuan yang cukup mengenai penyakit jantung.	<i>Naïve Bayes</i>	Hasil uji yang dilakukan sebanyak dua kali dengan jumlah data testing berturut-turut sebesar 25 dan 50 memperoleh nilai 80% dari data testing pertama, kemudian memperoleh nilai 78% pada hasil tesring yang kedua.
5.	Muhammad Ridho Handoko, Neneng[26].	Penyakit kehamilan adalah sebuah keluhan atau gangguan yang dirasakan dan terjadi pada saat kehamilan. Berdasarkan penjelasan dan pengalaman pakar, terdapat 13 (tiga belas) penyakit pada masa kehamilan, yaitu terjadi pada trimester satu, trimester dua dan trimester tiga.	<i>Naïve Bayes</i>	Dari hasil uji coba yang dilakukan dengan 22 responden secara acak menggunakan sistem pakar diagnosa penyakit selama kehamilan berbasis web mendapat hasil bahwa dari 22 responden, 17 orang mengalami penyakit selama kehamilan. Ketepatan diagnosa yang diperoleh dari perbandingan antara hasil diagnosa sistem yang sama dengan diagnosa dokter adalah dengan presentase nilai 77 %.

6.	Kadek Wibowo, Sfenrianto, Kaman Nainggolan[27]	Keterlambatan gerak motorik kasar pada anak merupakan salah satu gangguan tumbuh kembang anak. Motorik kasar mencakup gerakan otot-otot besar seperti otot tungkai dan lengan pada bayi berupa gerakan menendang,menjejak , meraih, mengangkat leher, dan menoleh. Pertumbuhan kemampuannya harus terus di pantau dan di stimulasi agar anak dapat tumbuh dan berkembang optimal.	<i>Naive Bayes</i> dan <i>Adaboost</i>	Didapat bahwa pengujian Naive Bayes classifier dengan optimasi PSO lebih baik dari pada Naive Bayes classifier berbasis Adaboost. Dengan record data pasien sebanyak 600, pengujian menggunakan Naive Bayes classifier berbasis Adaboost didapat akurasi sebesar 90.00% dengan masih terdapat kesalahan data sebesar 10.00%. Sedangkan akurasi pengujian Naive Bayes classifier dengan optimasi PSO meningkat 8.00% yaitu sebesar 98.00% dengan tingkat kesalahan data rendah, dengan menggunakan data yang sama.
7.	Ginanjari Abdurrahman[28]	Diabetes melitus merupakan penyakit dengan gejala kadar gula darah sewaktu lebih dari 200 mg/dL, dan kadar gula darah puasa lebih dari 126 mg/dL (Misnadiarly(Hestiana, 2017)). International Diabetes Federation (IDF) menjadikan Diabetes sebagai penyakit paling mematikan urutan ke-tujuh di dunia dengan prevalensi 1.9%.	<i>Adaboost</i>	Hasil klasifikasi algoritma <i>Adaboost Classifier</i> pada dataset hasil <i>imputing mean</i> menghasilkan akurasi sebesar 80.09%.Dataset hasil <i>imputing median</i> menghasilkan akurasi sebesar 76.19 %
8.	Rangga Ahsana, Rd. Rohmat	Kementerian Kesehatan telah melakukan beberapa usaha untuk mengendalikan diabetes seperti	<i>Adaboost</i>	AdaBoost adalah metode boosting SAMME dengan nilai akurasi sebesar 91.14% dibandingkan dengan algoritma

	Saedudin, Vandha Pradwiyasma Widarta[29]	membentuk 13.500 Pos Pembinaan Terpadu (Posbindu) agar memudahkan warga untuk melakukan Medical Diagnosis atau Diagnosis Medis. (Kementerian Kesehatan, 2018).		AdaBoost yang menggunakan metode boosting SAMME.R yaitu hanya sebesar 89.58%. Algoritma AdaBoost dengan metode SAMME memiliki nilai precision sebesar 0.93 dan recall sebesar 0.94. Dari nilai precision dan recall tersebut, didapatkan hasil f1-Score sebesar 0.93 dan nilai AUC sebesar 0.9693.
--	---	--	--	--

Berdasarkan tabel 2.1 maka dapat disimpulkan bahwa penelitian yang menggunakan *naïve bayes* saja memiliki nilai akurasi yang cukup baik namun jika menggabungkan dua metode seperti menambahkan metode *naïve bayes* dan *Adaboost* dapat meningkatkan nilai akurasi, dapat mengatasi ketidakseimbangan kelas dan juga untuk mengurangi *error*. Maka dari itu dalam penelitian ini penulis akan menggunakan metode *naïve bayes* dan *adaboost* untuk memprediksi penyakit serangan jantung untuk memperoleh hasil yang maksimal. Penelitian-penelitian yang telah dilakukan sebelumnya akan dijadikan acuan penulis dalam melakukan penelitian.