

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Penyakit Tiroid**

Penyakit tiroid merupakan gangguan pada kelenjar tiroid yang terletak dibagian leher tepatnya dibawah jakun [6]. Diagnosa penyakit tiroid ini sulit dilakukan, karena gejala penyakit tiroid bisa bermacam-macam tergantung pada naik dan turunnya hormon tiroid. Hormon tiroid meningkatkan penggunaan oksigen oleh sel-sel tubuh. Ketika tiroid memproduksi hormon berlebih, sel tubuh akan bekerja lebih keras dan metabolisme tubuh menjadi lebih cepat, kondisi ini disebut hipertiroid [7].

Krtika tubuh mengalami kekurangan atau kelebihan hormon tiroid, otak akan merangsang kelenjar tiroid untuk menyesuaikan kinerjanya agar kadar hormon tersebut kembali seimbang. Penyebab utama penyakit tiroid adalah kadar hormon tiroid yang terlalu tinggi atau rendah dalam tubuh kita. Hal ini dapat dipicu oleh berbagai faktor yang meliputi:

1. Masalah pada kelenjar pituari di otak.
2. Kelenjar tiroid yang rusak misalnya karena pancaran radiasi.
3. Pengaruh obat litium.
4. Kadar iodin yang berlebihan dalam tubuh.

Gambaran khas ini merupakan suatu tiroid, yang disebabkan oleh pembakaran atau metabolisme tubuh yang melebihi semestinya. Tanda-tanda hipertiroid ini sangat khas, oleh karena itu pasien hipertiroid lebih cepat datang ke dokter untuk memperoleh pengobatan, terutama apabila pasien mengalami pembesaran pada leher [8]. Penyakit tiroid dapat dikategorikan kedalam beberapa jenis, diantaranya adalah :

##### **1. Hipotiroid**

Hipotiroid adalah kondisi terlalu sedikitnya hormon tiroksin yang diproduksi oleh kelenjar tiroid sehingga tubuh mengalami defisiensi. Kondisi ini lebih sering dialami oleh wanita (terutama lansia) dan memiliki gejala-gejala umum

seperti kulit kering contohnya mengelupas, kelelahan, kenaikan berat badan tanpa sebab jelas, serta lebih sensitif terhadap hawa dingin [9].

## 2. Hipertiroid

Hipertiroid ialah hiper fungsi kelenjar tiroid dan sekresi berlebih dari hormone tiroid dalam sirkuit darah. Adapun subklinis hipertiroid, secara definisi diartikan kasus dengan kadar hormon normal tetapi Thyroid Stimulating Hormon (TSH) rendah. Jika hormon berlebihan, berat badan akan merosot tajam, pasien merasa panik, tegang, sulit tidur, jantung berdebar, tangan gemetar, dan mata terbelalak keluar (eksophtalmus) [10].

## 3. Penyakit Gondok

Penyakit gondok atau kekurangan zat beryodium adalah sekumpulan gejala yang timbul karena tubuh seseorang kekurangan unsur yodium secara terus menerus dalam jangka waktu yang lama (Depkes R.I, 2005). Akibat kekurangan yodium dapat berdampak pada penurunan kualitas kesehatan anak dan ibu hamil karena susunan syaraf terdiri dari sel neuron dan sel ganglia yang mulai terbentuk pada stadium embrologis yang terus berlangsung dalam waktu singkat sesudah bayi dilahirkan [11].

## 4. Hashimoto

Penyakit Hashimoto nama lainnya adalah Tiroiditis Hashimoto atau tiroiditis limfatik kronis. Penyakit ini dapat terjadi pada semua usia, tetapi paling sering terjadi pada wanita setengah baya. Terjadi karena proses autoimun, di mana sistem kekebalan tubuh keliru sehingga menyerang dan secara perlahan-lahan menghancurkan kelenjar tiroid. Dengan demikian kemampuan kelenjar tiroid dalam memproduksi hormon menjadi menurun [9].

## 5. Kanker Tiroid

Kanker tiroid merupakan penyakit keganasan yang tidak jarang ditemukan. Sebagian besar pertumbuhan dan perjalanan penyakit lambat, sehingga morbiditas dan mortalitasnya rendah namun ada yang pertumbuhannya sangat cepat dengan prognosa yang fatal. Kanker tiroid merupakan kelenjar tiroid yang berada pada bagian depan leher sedikit di bawah laring berbentuk kupukupu. Terjadi 0,85% dan 2,5% dari seluruh keganasan kanker tiroid pada

laki-laki dan perempuan dengan perbandingan 1:3. Dengan kata lain kanker tiroid lebih sering terjadi pada perempuan [11].

## 2.2. Data Mining

Data mining adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari *database* yang besar. Data mining merupakan metode untuk menemukan informasi tersembunyi dalam *database* dan bagian dari proses *Knowledge Discovery in Database* (KDD) untuk menemukan informasi dan pola yang berguna dalam data [12].

Karakteristik data mining sebagai berikut [13]:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2015).

Tahap-tahap data mining adalah sebagai berikut: (Junianto & Rachman, 2020:119)

1. Pembersihan data (data cleaning)  
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
2. Integrasi data (data integration)  
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
3. Seleksi data (data selection)  
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk di analisis yang akan diambil dari database.
4. Transformasi data (data transformation)  
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.

### 5. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan data.

### 6. Evaluasi pola (pattern evaluation)

Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan.

### 7. Presentasi pengetahuan (knowledge presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

## **2.3. Prediksi**

Prediksi merupakan proses atau hasil perkiraan tentang apa yang mungkin terjadi di masa depan. Ini melibatkan penggunaan informasi yang tersedia saat ini, analisis data historis, tren, pola, dan pemahaman tentang faktor-faktor yang mempengaruhi situasi atau peristiwa tertentu. Tujuan dari prediksi adalah untuk mengestimasi kemungkinan hasil masa depan berdasarkan informasi yang tersedia. Prediksi Kata atau istilah ini bisa menimbulkan berbagai persepsi. Prediksi atau dalam hal ini juga disebut peramalan (forecast) pada dasarnya merupakan dugaan atau peramalan mengenai peristiwa di waktu yang akan datang [14].

## **2.4. Klasifikasi**

Proses pengelompokan objek, entitas, atau data ke dalam kategori atau kelas yang berbeda berdasarkan karakteristik atau atribut yang dimiliki. Tujuan klasifikasi adalah untuk mengidentifikasi pola, memahami hubungan antara data, atau membuat prediksi berdasarkan atribut yang diberikan. Klasifikasi digunakan untuk mengelompokkan data berdasarkan pola atau atribut yang dimiliki olehnya. Ini dapat membantu dalam analisis data dan pengambilan keputusan. Klasifikasi merupakan suatu teknik menemukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah

didefinisikan. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan [4].

## **2.5. Implementasi**

Implementasi dalam Kamus Besar Bahasa Indonesia adalah pelaksanaan, penerapan. Adapun implementasi menurut para ahli yakni, menurut Usman (2002), mengemukakan pendapatnya tentang Implementasi atau pelaksanaan sebagai berikut “implementasi adalah bermuara pada aktivitas, aksi, tindakan, atau adanya mekanisme suatu sistem. Implementasi bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan untuk mencapai tujuan kegiatan”. Pengertian implementasi yang dikemukakan di atas, dapat dikatakan bahwa implementasi adalah bukan sekedar aktivitas, tetapi suatu kegiatan yang terencana dan dilakukan secara sungguh-sungguh berdasarkan acuan norma tertentu untuk mencapai tujuan kegiatan. Oleh karena itu, implementasi tidak berdiri sendiri tetapi dipengaruhi oleh objek berikutnya [15].

## **2.6. Algoritma Support Vector Machine (SVM)**

Algoritma Support Vector Machine (SVM) adalah metode pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM menciptakan sebuah model yang dapat memisahkan dua kelas data yang berbeda dengan menggunakan hyperplane (bidang pemisah) yang optimal di antara kelas-kelas tersebut. Kemampuan SVM dalam mengidentifikasi hyperplane secara terpisah diantara dua kelas berbeda sehingga termaksimalkan (Indrayuni, 2016). SVM menjamin untuk memaksimalkan jarak antara data yang paling dekat dengan hyperplane [16].

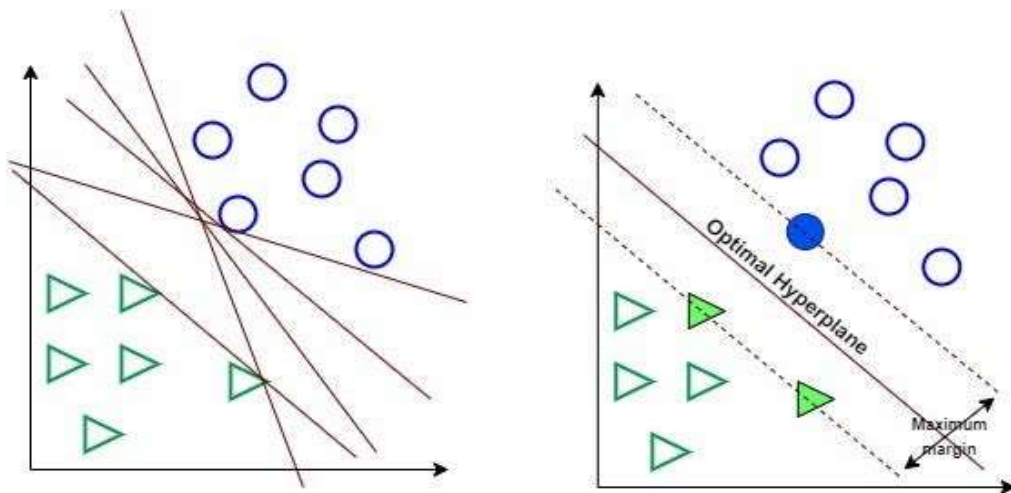
Berikut adalah langkah-langkah umum dalam algoritma SVM:

1. Persiapan data: Mulai dengan mengumpulkan dan mempersiapkan data yang akan digunakan untuk pelatihan model. Data ini harus terstruktur dengan benar dan terbagi menjadi set pelatihan dan set pengujian.
2. Pelatihan model: Dalam tahap ini, SVM akan melatih model menggunakan set pelatihan. Tujuannya adalah menemukan hyperplane yang optimal untuk memisahkan kelas-kelas data. SVM mencari hyperplane yang

memiliki margin terbesar, yaitu jarak terbesar antara hyperplane dan titik-titik pelatihan dari kedua kelas.

3. Transformasi data: jika data tidak dapat dipisahkan secara linear oleh hyperplane, SVM dapat menggunakan fungsi kernel untuk mentransformasi data ke dalam ruang fitur yang memiliki dimensi yang lebih tinggi. Ini memungkinkan SVM untuk menemukan hyperplane yang dapat memisahkan data secara nonlinear.
4. Pengujian dan evaluasi: Setelah pelatihan selesai, model SVM akan diuji dengan menggunakan set pengujian yang tidak terlihat sebelumnya. Performa model dievaluasi berdasarkan metrik evaluasi yang sesuai, seperti akurasi, presisi, recall, atau F1-score.

SVM memiliki beberapa kelebihan, termasuk kemampuan untuk menangani data dengan dimensi yang tinggi, toleransi terhadap overfitting, dan kinerja yang baik bahkan dalam kasus di mana jumlah fitur lebih banyak dari jumlah sampel. Namun, SVM juga memiliki beberapa kelemahan, seperti kurang efisien dalam menangani set data yang sangat besar dan kebutuhan untuk menentukan parameter yang tepat, seperti jenis kernel dan parameter penalti. Algoritma SVM dapat diterapkan dalam berbagai bidang, termasuk pengenalan pola, klasifikasi teks, bioinformatika, dan pengenalan citra. Tujuan Algoritma Support Vector Machine (SVM) dalam menentukan hyperplane terbaik dalam ruang lingkup berdimensi-N (ruang dengan N-jumlah fitur) yang berfungsi sebagai pemisah yang jelas bagi titik-titik data input.



Gambar.1 titik input data hyperplane

Pada gambar sebelah kiri yang menunjukkan beberapa kemungkinan bidang (hyperplane) untuk memisahkan data lingkaran dan data kotak. Algoritma SVM mencari hyperplane terbaik yang dapat memisahkan kedua kelas secara optimal. Terlihat pada gambar. 1 yang berada di sebelah kanan, sebuah hyperplane optimal berhasil dibuat dan mampu memisahkan kedua kelas sehingga memiliki margin yang maksimal.

## 2.7. Algoritma Adaboost

Metode Adaboost, singkatan dari Adaptive Boosting, adalah algoritma ensemble learning yang digunakan untuk meningkatkan kinerja model pembelajaran mesin. Adaboost adalah salah satu metode boosting yang populer dan efektif dalam mengatasi masalah klasifikasi. Adaboost digunakan untuk mengklasifikasi data pada kelasnya masing-masing, mencari kategori kelas berdasarkan dengan nilai bobot yang dimiliki oleh kelas. Metode boosting dapat meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara membangkitkan kombinasi dari suatu model, tetapi hasil klasifikasi atau prediksi yang dipilih adalah model yang memiliki nilai bobot paling besar. Jadi, setiap model yang dibangkitkan memiliki atribut berupa nilai bobot.

Dataset yang telah seimbang akan divalidasi dengan menggunakan 10-fold cross validation. Hasil dari validasi akan menghasilkan data yang diukur yaitu AUC dan Accuracy [17]. Proses ini terus dilakukan berulang sehingga terdapat pembaruan nilai pada kelas. Pada adaboost nilai bobot akan terus bertambah pada setiap iterasinya dari bobot nilai yang salah pada setiap iterasinya [18]. Perhitungan algoritma adaboost ditunjukkan pada persamaan (1) yaitu [19] :

$$F(x) = \text{sign} \sum_t^T \alpha_t h_t(x)$$

$h_t(x)$  = weak atau basic classifier

$\alpha_t$  = tingkat pembelajaran (learning rate)

$F(x)$  = final classifier

Algoritma ini menerima input berupa pembelajaran lemah (Weak Learn). Sebuah integer  $t$  yang menspesifikasi banyaknya iterasi dan data pelatihan sebanyak  $N$   $\{(x_1, y_1), \dots, (x_n, y_n)\}$ . Setelah itu akan dilakukan inisialisasi bobot pada setiap element dalam fitur vector. Inisialisasi bobot ini dirumuskan dengan  $D_1(i) = 1/n$ ,

dimana  $n$  adalah jumlah element 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).

$$error_t = \sum_{i=1}^n D_t(i) h_t(x_i) y_i$$

Hasil dari perhitungan ini akan dilakukan perhitungan nilai alpha ( $\alpha$ ). 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(1). Perhatikan nilai alpha ini akan ditunjukkan pada persamaan(3).

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

Dengan cara ini, akan menggabungkan output dari weak learner dan menciptakan strong learner yang pada akhirnya meningkatkan kekuatan model prediksi.

Adaboost memiliki kelebihan seperti kemampuan untuk mengatasi data yang tidak seimbang, kinerja yang baik dalam klasifikasi, dan tidak adanya parameter kompleks yang harus diatur. Namun, Adaboost juga sensitif terhadap data noise atau pencilan (outliers), dan dapat mengalami overfitting jika kompleksitas model lemah terlalu tinggi.

## **2.8. Feature Selection *Forward Sequential Select (FSS)***

Feature Selection atau seleksi fitur adalah sebuah proses yang biasa digunakan pada Machine Learning dimana sekumpulan dari fitur yang dimiliki oleh data digunakan untuk pembelajaran algoritma [20]. Metode Sequential Forward Selection atau metode seleksi maju adalah algoritma pencarian paling sederhana. Forward Selection didasarkan pada model Regresi Linear. Forward Selection adalah salah satu teknik untuk mereduksi dimensi dataset untuk menghapus atribut-atribut yang tidak relevan. Metode Forward Selection merupakan model yang diawali dengan nol variable, untuk selanjutnya variable dimasukkan satu persatu sampai pada kriterianya terpenuhi. Prosedur Forward



Selection dimulai dengan tidak ada variabel dalam model [21]. Langkah-langkah metode forward selection adalah sebagai berikut [22]:

1. Membuat model dengan meregresikan variabel respon Y dengan setiap variabel prediktor. Kemudian dipilih model yang mempunyai nilai R2 tertinggi. Misal model tersebut adalah yang memuat prediktor Xa, yaitu:

$$Y = b_0 + b_a X_a$$

2. Meregresikan variabel respon Y, dengan prediktor Xa, ditambah dengan setiap prediktor selain Xa dan prediktor lain. Kemudian dipilih model yang nilai R2 nya tertinggi, misal mengandung tambahan prediktor Xb, yaitu model

$$Y = b_0 + b_a X_a + b_b X_b$$

Prediktor terpilih Xb berarti mempunyai Fsequensial tertinggi. Formula Fsequensial untuk Xb adalah

$$F_{seq} = R(\beta_0, \beta_a, \beta_b) / MSE_{db}$$

Nilai Fsequensial untuk Xb juga dapat diperoleh dengan cara mengkuadratkan nilai statistik uji T prediktor Xb.

3. Proses diulang sampai didapatkan  $F_{seq} > F_{in}$ . Nilai  $F_{in} = F(1, v, \alpha)$ , sehingga model terbaik yang dipilih adalah model yang tidak mempunyai prediktor dengan  $F_{seq} < F_{in}$ .

## 2.9. Dataset

Dataset adalah kumpulan data yang diorganisasi secara terstruktur. Biasanya, dataset terdiri dari himpunan objek atau sampel yang disusun dalam bentuk tabel, file teks, atau format data lainnya. Setiap objek atau sampel dalam dataset mewakili entitas atau fenomena tertentu, sedangkan setiap atribut atau kolom mewakili informasi atau fitur yang berkaitan dengan objek tersebut. Dataset digunakan dalam berbagai bidang, seperti ilmu data, statistik, pembelajaran mesin, dan penelitian ilmiah. Tujuan dari penggunaan dataset adalah untuk melakukan analisis, pengolahan, atau pelatihan algoritma dengan menggunakan data yang ada. Dataset dapat beragam dalam ukuran dan kompleksitasnya, mulai dari dataset kecil dengan beberapa ratus sampel hingga dataset besar dengan jutaan atau bahkan miliaran sampel.

### **2.10. Akurasi**

Akurasi adalah ukuran yang menentukan tingkat kemiripan antara hasil pengukuran dengan nilai yang sebenarnya diukur. Akurasi digunakan untuk mengetahui seberapa besar tingkat kesalahan pengukuran yang dapat terjadi pada suatu alat ukur. Biasanya pernyataan terhadap akurasi dalam bentuk persentase dengan skala berdasarkan dengan jenis ukuran yang digunakan [23].

### **2.11. RapidMiner**

Menurut (Fatmawati, 2016) Cepat Miner adalah perangkat lunak itu dibuat oleh Dr. Markus Hofmann dari Institut Teknologi Blanchardstown dan Raif Klinkenberg dari rapid-i.com dengan GUI (Graphical Userm Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini adalah sumber terbuka dan dibuat menggunakan bahasa Java dilisensikan di bawah Lisensi Publik GNU dan Rapid Miner dapat dijalankan pada sistem operasi apapun. Dengan menggunakan Rapid Miner, tidak diperlukan keterampilan coding khusus, karena semua fasilitas sudah disediakan. Rapid Miner khusus untuk penggunaan data mining [24].

### **2.12. Kurva ROC**

Kurva ROC merupakan hasil plot antara true signal (sensitivity) dan false signal (1-specificity) pada keseluruhan rentang titik cut-off yang mungkin. Rentang dari luas area di bawah kurva ROC adalah nol sampai satu. Luas area di bawah kurva mampu mengukur seberapa besar kemampuan model untuk mendiskriminasi observasi yang mengalami kejadian sukses dan tidak mengalami kejadian sukses [25].

### **2.13. Penelitian Terkait**

Beberapa penelitian terkait dengan teknik Data Mining metode untuk memprediksi penyakit tiroid diantaranya:

Tabel 2.13.1. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
1.	Perbandingan kinerja Algoritma Naive Bayes dan K-Nearest Neighbors (KNN) dalam memprediksi penyakit tiroid, Syarifah Jasmine Putri, Adetia Pratama, Kevin Attaqwa, Rahmaddeni, 2023.	Dalam penelitian ini menggunakan data sebanyak 500 data, dengan 21 atributn yaitu Age, sex, onthyroxine , query onthyroxine , on antithyroid medication, sick, pregnant, thyroid surgery, I131 treatment, query hypothyroid , query hyperthyroid, lithium, goiter, tumor, hypopituitar y, psych, TSH, T3, TT4, T4U, FTI.	Metode yang digunakan adalah metode Naive Bayes dan K-Nearest Neighbors (KNN) Dan machine learning menggunakan fremwork python.	Dengan menggunakan freamework python untuk mencari perbandingan sebuah accuracy pada kedua algoritma, algoritma Naive Bayes dengan data testing 63% dan data training 64% sedangkan akurasi yang di hasilkan pada algoritma K-Nearest Neighbors (KNN) adalah 63%.	perbandingan pada akurasi setiap algoritma Naive Bayes dengan data testing 63% dan data training 64% sedangkan akurasi yang di hasilkan pada algoritma K-Nearest Neighbors (KNN) adalah 63% dapat disimpulkan bahwa algoritma yang memiliki klasifikasi sangat baik secara adalah Naive Bayes dan K-Nearest Neighbors (KNN) berdasarkan penilaian akurasi dengan 500 data penyakit tiroid. algoritma Naive Bayes permasalahan mengidentifikasi penyakit tiroid.	Perbandingan Kinerja Algoritma Naive Bayes Dan K-Nearest Neighbors (KNN) Dalam Mempredik si Penyakit Tiroid menghasilkan akurasi algoritma yang paling baik dalam menentukan identifikasi penyakit.

Tabel 2.13.2. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
2.	Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus : Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna), Devi Sartika, Yupianti 2020.	Menggunakan data privad pada RSUD Hasanudin Damrah Manna.	Menggunakan Algoritma Dicison Tree, C4.5 dan metode waterfall.	Hasil penelitian ini adalah menghasilkan rule yang dapat membantu tenaga kesehatan di RSUD Hasanuddin Damrah Manna dalam mengdiagno sa penyakit yang disebabkan oleh kelenjar tiroid.	Perlu adanya pengembangan sistem lebih lanjut.	Sistem Klasifikasi penyakit tiroid menggunakan algoritma C4.5 yang dibuat telah sesuai dengan tahapan atau urutan proses yang semestinya. Dengan adanya sistem ini, maka RSUD Hasanuddin Damrah Manna dapat terbantu dalam melakukan diagnosa penyakit kelenjar tiroid. Mempermudah menampilkan informasi diagnosa penyakit pasien berdasarkan gejala yang dialami atau yang telah dipilih di sistem.

Tabel 2.13.3. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
3.	Perbandingan Metode Algoritma Decision Tree C4.5 Dan Naïve Bayes Untuk Memprediksi Penyakit Tiroid. Leli Safitri, Krista Cahayani Murtiwiyati, Siti Chodidjah, Deasy Indayanti. 2022	21 atribut independen menggunakan dua algoritma data mining yaitu Decision Tree C4.5 dan Naïve Bayes.	Penelitian ini menggunakan metode klasifikasi data mining Algoritma Decision Tree C4.5 dan Naïve Bayes	Hasil pengujian menunjukkan bahwa akurasi lebih baik diperoleh dari Algoritma Decision Tree C4.5 sebesar 97,12% sedangkan nilai akurasi Algoritma Naïve Bayes sebesar 76,02%. Nilai Area Under Curve (AUC) pada kurva Receiver Operating Characteristic (ROC) menunjukkan Algoritma Decision Tree C4.5 memiliki nilai lebih tinggi dari Algoritma Naïve Bayes dengan hasil klasifikasi Good Classificati.	akurasi Algoritma Decision Tree C4.5 adalah sebesar 97,12% dan nilai AUC 0,841, sedangkan nilai akurasi Algoritma Naïve Bayes sebesar 76,02% dan nilai AUC 0,718.	Kedua algoritma ini sudah cukup baik diterapkan dalam memprediksi penyakit tiroid.

Tabel 2.13.4. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
4.	Klasifikasi Support Vector Machine (SVM) untuk menentukan Tingkat Kemanisan Mangga Berdasarkan Fitur Warna, Muhammad Ichwan, Irma Amelia Dewi, Zeni Muharom S. 2018	Data yang digunakan berdasarkan citra warna buah dengan citra buah mangga berjumlah 20 buah dari 24 buah yang diuji	Metode KNN dan Metode SVM	Metode KNN memiliki prosentasi akurasi sebesar 83.3%, sistem mampu mengklasifikasi citra buah mangga dengan benar berjumlah 20 buah dari 24 buah yang diuji dengan pengaturan parameter $k = 7$ , sedangkan akurasi atau berdasarkan menggunakan metode SVM memiliki prosentasi sebesar 87.5%, sistem mampu mengklasifikasi citra buah mangga dengan benar berjumlah 21 buah dari 24 buah		Metode SVM memiliki akurasi lebih tinggi dibandingkan KNN

Tabel 2.13.5. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
5.	Classification of Thyroid Diseases Using Machine Learning and Bayesian Graph Algorithms. Giuseppe Mollica, Daniel Francesconi, Gabriel Costante, Sonia Moretti, Riccardo Giannini, Efisio Puxeddu, Paolo Valigi. 2022.	Such a dataset comprises 730 genes from 60 thyroid samples (53 cancers and 7 normal tissues), which were classified into a healthy patient class (0) and 4 immunophenotype classes (1-4).	Exploratory Data Analysis techniques and oversampling methods for data preprocessing and overfitting reduction.	Results show that the use of Bayesian network frameworks can help in integrating prior expertise knowledge in the classification problem and build new hypotheses about features interaction	the inclusion of more samples and the use of different gene correlation maps could help improve the classification and study of gene-based diagnosis, prognosis, and response to therapy of thyroid tumors.	The results show that the proposed use of PGN yields interesting results over the classical machine learning algorithms by exposing the discovery of new interactions and correlations between features not defined a priori.
6.	Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. Kalpna Guleriaa, Shagun Sharmaa, Sushil Kumarb, Sunita Tiwarib. 2022.	The dataset has 30 attributes and 3772 instances related to hypothyroidism. WEKA 3.8.4 simulator has been used in this work to perform experiments .	In this article, various machine learning-based techniques have been applied to build predictive models, which includes decision tree, random forest, naive Bayes and multiclass classifier	The performance evaluation indicates that the decision tree and random forest provide better results with the highest accuracy of 99.5758% and 99.3107% and very few error rates of 0.0424 and	Further, the dataset can be increased, and academic researchers in the medical area can also use this work to identify more ML or DL based prediction models to detect hypothyroidism.	The performance evaluation indicates that the decision tree and random forest provide better results with the highest accuracy of 99.5758% and 99.3107% and very few error rates of 0.0424 and

			<p>and a deep learning (DL) based model Artificial Neural Network (ANN), which is best known for dealing with text data has been applied to predict the class of hypothyroidism.</p>	<p>0.0689, respectively . Furthermore , a comparison among the presented classifiers has been made, and also the proposed model has been compared with previous works, and it has been found that it shows better accuracy as compared to other related works. The DL-based ANN model also offers a competitive accuracy which is 93.8226%. Furthermore , this study can be useful for researchers to identify a suitable model for hypothyroidism detection and classificatio</p>	<p>0.0689, respectively .</p>
--	--	--	--	--	-------------------------------



Tabel 2.13.6. Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
7.	Exploring the potential of thyroid hormones to predict clinical improvements in depressive patients: A machine learning analysis of the real-world based study. Dan Qiaoa, Huishan Liua, Xuemin Zhanga, Lei Leia, Ning Suna, Chunxia Yanga, Gaizhi Lia, Meng Guoa, Yu Zhanga, Kerang Zhanga, Zhifen Liu. 2022.	Data of 2086 depressive inpatients during period of 2014-2020.	Here, we conducted a real-world based retrospective study using clinical and biochemical	The results showed that lower values of FT3 and FT4 can both predict a poor clinical outcome in depression. Further, a model with the best performance was selected (sensitivity=0.91, specificity=0.79, and ROC-AUC=0.86), including the values of FT3 and FT4, and the scores of Hamilton Depression Scale (HAMD) and Hamilton Anxiety Scale (HAMA) as features.	Our findings present a crucial role of thyroid measurements in predicting clinical outcomes of depression. Assessment of thyroid hormone should be extended to routine practice settings to determine which patients should be most in need of earlier or intensive interventions for preventing continued dysfunction.	In conclusion, this study quantified the effectiveness and contribution of thyroid hormones in predicting the clinical outcomes of depression, and future research can confidently explore deeper insights into endocrinology data, especially thyroid hormones, of depressed patients and more robust prediction models for predicting the response to antidepressant treatment in the real-world setting
8.	Novelectronic health records applied for prediction of pre-	This retrospective study included 5243 pregnant	Machine learning algorithms and performance analysis	The XGboost model had the best prediction performance	Future studies are needed to assess the real-world clinical utility	risk of PE development can be predicted with excellent

	eclampsia: Machine-learning algorithms. Yi-xin Lia,1, Xiao-ping Shena,1, Chao Yangb, Zuo-zeng Caoa, Rui Dua, Min-da Yua, Jun-ping Wanga, Mei Wang. 2021.	women who received antenatal care at Xinhua hospital Chongming branch, affiliated to Shanghai Jiaotong University between July 2016 and December 2019.		e (accuracy = 0.920, precision = 0.447, recall = 0.789, fl_score = 0.571, auROC = 0.955). The most predictive feature of PE development was fasting plasma glucose, followed by mean blood pressure and body mass index. An easy-to-use model that a patient could answer independently still enabled accurate prediction, with auROC of 0.83	of the model.	discriminative ability using ML algorithms based on EHR collected at the early second trimester
--	--	--	--	---	---------------	---

Berdasarkan review jurnal yang telah dilakukan baik jurnal nasional maupun internasional dapat disimpulkan bahwa dataset dapat terakurasi dengan baik maka daya yang dibutuhkan dalam penelitian kurang lebih 500 fitur, dan metode yang digunakan yaitu Algoritma *Support Vector Machine (SVM)*, Adaboost dengan feature selection *Forward Sequential Select (FSS)* .