

## Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid

Sarifah Agustiani<sup>1</sup>, Ali Mustopa<sup>2</sup>, Andi Saryoko<sup>3</sup>, Windu Gata<sup>4</sup>, Siti Khotimatul Wildah<sup>5</sup>

<sup>1</sup>STMIK Nusa Mandiri  
e-mail: 14002398@nusamandiri.ac.id

<sup>2</sup>STMIK Nusa Mandiri  
e-mail: 14002370@nusamandiri.ac.id

<sup>3</sup>STMIK Nusa Mandiri  
e-mail: andi.asy@nusamandiri.ac.id

<sup>4</sup>STMIK Nusa Mandiri  
e-mail: windu@nusamandiri.ac.id

<sup>5</sup>STMIK Nusa Mandiri  
e-mail: 14002399@nusamandiri.ac.id

**Abstrak** - Gangguan fungsi tiroid seringkali sulit diidentifikasi karena gejalanya tidak spesifik. Gejala gangguan tiroid sangat mirip dengan berbagai keluhan akibat gaya hidup modern sehingga sangat sering diabaikan. Akibatnya pasien seringkali tidak menyadari ada masalah dan tidak memeriksakan diri ke dokter. Untuk itu, diperlukan sebuah penelitian yang menerapkan metode memprediksi penyakit tersebut yang nantinya akan mempermudah pasien dalam mendiagnosa dan deteksi dini terhadap kadar tiroid. Penelitian ini bertujuan untuk melakukan prediksi terhadap penyakit tiroid dengan data yang digunakan adalah data sekunder yang diperoleh dari UCI repository, data ini berisi tentang data pasien yang terkena penyakit tiroid, sedangkan metodenya menggunakan algoritma J48 karena dalam beberapa penelitian, algoritma J48 terbukti memiliki performa yang baik dalam mendeteksi suatu penyakit, serta menghasilkan nilai accuracy dan AUC yang tinggi. Tahapan analisa data dilakukan berdasarkan metode CRISP-DM sedangkan pengujian algoritma dilakukan dengan tools Weka. Hasil dari pengujian tersebut diperoleh nilai akurasi sebesar 99.645%, dan nilai AUC sebesar 0,992 dengan demikian akurasi memiliki tingkat Excellent Classification.

**Kata Kunci:** Tiroid, Algoritma J48, Tahapan CRISP-DM, Weka.

**Abstract** - Impaired thyroid function is often difficult to identify because the symptoms are not specific. The symptoms of thyroid disorder are very similar to various complaints due to modern lifestyles so it is often overlooked. As a result, patients often do not notice a problem and do not have to consult a doctor. Therefore, there is a study that implements methods to predict the disease which will facilitate the patient in diagnosing and early detection of thyroid levels. This research aims to predict against thyroid disease with the data used is the secondary data obtained from the UCI repository, this data is about the patient data affected by thyroid disease, while the method uses the J48 algorithm because in some studies, the J48 algorithm is proven to have good performance in detecting an illness, as well as producing high value of Accuracy and AUC. The stage of data analysis is based on the CRISP-DM method while algorithm testing is done with Weka tools. Results of the test obtained an accuracy value of 99.645%, and a AUC value of 0.992 thus the accuracy has Excellent Classification level.

**Keywords:** Thyroid, The algorithm J48, Stages of CRISP-DM, Weka.

### PENDAHULUAN

Tiroid merupakan kelenjar endokrin murni terbesar dalam tubuh manusia yang terletak di leher bagian depan, terdiri atas dua bagian (lobus kanan dan lobus kiri). Panjang kedua lobus masing-masing 5 cm dan menyatu di garis tengah, berbentuk seperti kupu-kupu (Kementerian Kesehatan, 2015). Tiroid berfungsi untuk menghasilkan hormon yang berguna untuk menjaga metabolisme tubuh. serta meningkatkan jumlah oksigen pada sel dan

rangsangan jaringan tubuh dalam menghasilkan protein (Wijonarko, 2018).

Menurut kelainan fungsinya, gangguan tiroid dibedakan dalam 3 jenis, yaitu, Hipotiroid yaitu kumpulan manifestasi klinis akibat berkurang atau berhentinya produksi hormon tiroid. Hipertiroid atau disebut juga tirotoksikosis yang merupakan kumpulan manifestasi klinis akibat kelebihan hormontiroid. Eutiroid yaitu keadaan tiroid yang berbentuk tidak normal tapi fungsinya normal (Kementerian Kesehatan, 2015).

Faktor risiko penyakit atau gangguan tiroid dapat dipengaruhi oleh berbagai faktor seperti usia di atas 60 tahun maka semakin berisiko terjadinya hipotiroid atau hipertiroid. Jenis kelamin perempuan biasanya lebih berisiko terjadi gangguan tiroid. Di antara banyak faktor penyebab auto imunitas terhadap kelenjar tiroid, genetik dianggap merupakan factor pencetus utama. Merokok juga dapat menyebabkan kekurangan oksigen di otak dan nikotin dalam rokok dapat memacu peningkatan reaksi inflamasi. Selain itu, stres juga berkorelasi dengan antibodi terhadap antibodi TSH-reseptor (Kementerian Kesehatan, 2015).

Gangguan fungsi tiroid biasanya sulit diidentifikasi karena gejalanya sangat mirip dengan keluhan akibat gaya hidup modern sehingga gejala tersebut sering diabaikan. Akibatnya pasien tidak menyadari ada masalah dalam dirinya dan tidak memeriksakannya ke dokter (Putri, 2019). Untuk itu, diperlukan sebuah penelitian yang menerapkan metode untuk memprediksi penyakit tersebut yang nantinya akan mempermudah pasien dalam mendiagnosa dan deteksi dini.

Salah satu metode yang dapat digunakan untuk memprediksi penyakit adalah dengan menggunakan data mining. Data mining adalah sekumpulan proses untuk mencari pola atau alur yang selama ini tidak diketahui secara manual dari sebuah kumpulan data yang nantinya akan menghasilkan berupa pengetahuan baru (Handayani, Nurlelah, Raharjo, & Ramdani, 2019).

Berdasarkan fungsionalitasnya tugas data mining dapat dikelompokkan menjadi enam kelompok yaitu Klasifikasi (*classification*), Klusterisasi (*clustering*), Regresi (*regression*), Deteksi anomai (*anomaly detection*), Pembelajaran aturan asosiasi (*association rule learning*), dan Perangkuman (*summarization*) (Suyanto, 2017). Dalam penelitian ini data mining yang digunakan adalah klasifikasi yang merupakan teknik *data mining* yang mempelajari sekumpulan data sehingga menghasilkan aturan untuk dapat mengklafikasi atau mengenali data-data baru yang belum pernah dipelajari. Proses pembelajaran memerlukan data *input* berupa data *training* yang memiliki atribut kelas dan mengeluarkan *output* berupa model klasifikasi (Suyanto, 2017).

Sedangkan metode atau algoritma yang digunakan dalam penelitian ini adalah J48. J48 adalah salah satu jenis classifier pada metode klasifikasi dalam data mining dan bagian dari C4.5 *decision tree* yang sederhana. C4.5 membangun sebuah pohon keputusan berdasarkan pada seperangkat input data yang berlabel. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (Diwandari & Setiawan, 2015).

Penerapan algoritma J48 juga telah banyak dilakukan dalam beberapa penelitian, salah satunya penelitian yang membahas tentang Penerapan Algoritme J48 Untuk Prediksi Penyakit Demam Berdarah dan berdasarkan hasil penelitian, disimpulkan bahwa, algoritme J48 mudah dalam mendeteksi penyakit demam berdarah pada pasien opname (Astuti, Mujiati, Ayu, Ristianah, & Lestari, 2016). Dalam penelitian lain mengenai Diagnosis Penyakit Parkinson Berdasarkan Kombinasi Algoritme Data Mining Dan Seleksi Fitur dijelaskan bahwa algoritme J48 memiliki nilai akurasi terbaik yaitu sebesar 96.923% dengan, ROC 0.982 dan running time 0.10 detik. Sedangkan untuk algoritme klasifikasi dengan performa terendah dimiliki oleh algoritme Naive Bayes dengan nilai 76.923% (Lombok et al., 2016). Penelitian selanjutnya mengenai Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda Comparison yang menyimpulkan bahwa berdasarkan hasil analisis dan pembahasan, dapat disimpulkan bahwa hasil ketepatan klasifikasi penyakit stroke pada data pasien di RSUD Abdul Wahab Sjahranie bulan November dan Desember 2017 dengan metode naive Bayes adalah 81,25% dan metode decision tree algoritma (J48) diperoleh tingkat akurasi sebesar 87,5%. Hal ini menunjukkan bahwa pada penelitian ini, metode *decision tree* algoritma (J48) memberikan ketepatan prediksi klasifikasi yang lebih baik (Lishania, Goejantoro, & Nasution, 2019). Sehingga dapat disimpulkan bahwa penggunaan metode J48 merupakan metode yang cukup baik dalam memprediksi penyakit .

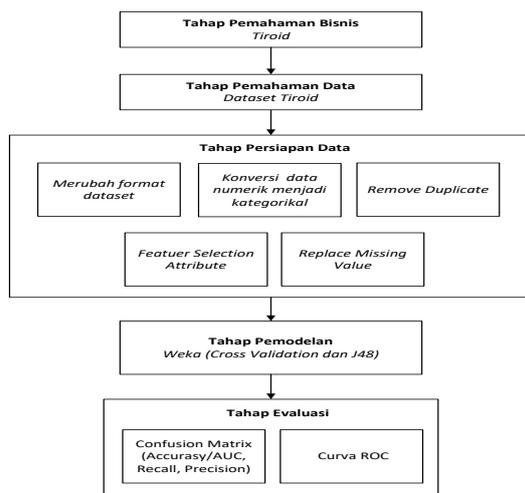
Penelitian penyakit tiroid juga telah banyak dilakukan mengingat beragamnya jenis tiroid. Salah satunya penelitian yang dilakukan oleh Puji Widayati, Agus Ariyanto, Sri Setiyowati, Sutari, V Yulianti mengenai Preparasi Pereaksi Kit Immunoradiometricassa (IRMA) Thyroglobulin (TGB) Untuk Deteksi Kanker Tiroid (Widayati et al., 2019). Penelitian yang dilakukan oleh Garri Prima Decroli, Eva Decroli mengenai Krisis Tiroid pada Wanita Multipara Usia 42 Tahun (Decroli & Decroli, n.d.), penelitian yang dilakukan oleh Litta Septina Mahmelia Zaid Endah Zuraidah Agnes Stephanie mengenai Peran Rearrangement RET/PTC pada Karsinoma Papiler Tiroid (Septina et al., 2019), Pengembangan Modul Catu Daya Tegangan Tinggi Perangkat Scintigrafi Untuk Tiroid Sc-12 dilakukan oleh Wiranto Budi Santoso dkk (Santoso, Santoso, Romadhon, & Sukandar, 2016), selanjutnya Diagnosis Tiroid yang dilakukan oleh Mansyur Arif (Kurniawan & Arif, 2015) yang tentunya memiliki berbagai macam metode. Dalam hal ini juga terdapat penelitian yang menggunakan data mining salah satunya penelitian yang dilakukan oleh Bambang Wijonarko (2018) mengenai Perbandingan Algoritma Data Mining *Naive Bayes* dan *Bayes*

*Network* Untuk Mengidentifikasi Penyakit Tiroid. Dalam penelitian ini melakukan komparasi algoritma diantaranya *Naïve Bayes* dan *Bayes Network* yang menyatakan bahwa *Bayes Network* memiliki Akurasi yang lebih tinggi dibandingkan *Naïve Bayes* (Wijonarko, 2018). Penelitian lainnya dilakukan oleh Umar Sidiq, Dr. Syed Mutahar Aaqib, dan Dr. Rafi Ahmad Khan (2019) mengenai Diagnosis Berbagai Penyakit Tiroid menggunakan Data Mining Teknik klasifikasi. Dalam penelitian ini dinyatakan bahwa *Decision Tree* memiliki akurasi yang lebih tinggi daripada algoritma lainnya (Sidiq, Aaqib, & Khan, 2019).

Berdasarkan latar belakang yang telah diuraikan dari penelitian sebelumnya yaitu penelitian yang menggunakan berbagai metode klasifikasi meliputi KNN, SVM, *Decision Tree*, dan *Naïve Bayes* masih dapat dilakukan peningkatan dalam hal akurasi yang dihasilkan. Hal tersebut dapat dilihat karena terjadinya beberapa masalah seperti ketidakseimbangan data akibat terlalu banyaknya atribut dalam sebuah dataset. Oleh sebab itu, untuk mengatasi permasalahan tersebut, penelitian ini dimaksudkan untuk meningkatkan akurasi dengan menerapkan algoritma J48 untuk prediksi penyakit tiroid.

## METODOLOGI PENELITIAN

Metodologi penelitian ini dilakukan dengan tahapan CRISP-DM yang merupakan upaya standarisasi proses data mining yang terdiri dari 6 tahapan (Haryati, Sudarsono, & Suryana, 2015). Sedangkan algoritma yang digunakan adalah algoritma J48. Berikut merupakan gambar tahapan penelitian:



Sumber: (Agustiani et al., 2020)

Gambar 1. Metodologi Penelitian

### 1. *Business Understanding*

Tahap awal ini berfokus pada pemahaman tujuan proyek dan persyaratan dari perspektif bisnis, kemudian mengubah pengetahuan ini menjadi

definisi masalah penambangan data dan rencana awal yang dirancang untuk mencapai tujuan (T & Larose, 2015). Sasaran utama dari penelitian yang ingin dicapai adalah bagaimana memprediksi penyakit tiroid.

### 2. *Data Understanding*

Tahap ini dimulai dengan pengumpulan data awal dan berlanjut dengan kegiatan yang memungkinkan untuk menjadi memahami data, mengidentifikasi masalah kualitas data, menemukan wawasan pertama ke dalam data, dan/atau mendeteksi himpunan bagian yang menarik untuk membentuk hipotesis mengenai informasi tersembunyi (Haryati et al., 2015).

### 3. *Data Preparation*

Tahapan ini terdiri dari beragam operasi untuk membersihkan dan mentransformasi data yang telah disiapkan untuk langkah pemodelan berikutnya. Beberapa jenis masalah pada data yang paling umum diantaranya: ketidakseimbangan kelas, data tidak konsisten, outlier dan nilai-nilai yang hilang. Adapun jumlah data yang diteliti sebanyak 7200 data dengan 22 atribut.

### 4. *Modeling*

Pada tahap modeling dalam *data mining* dilakukan proses pemilihan algoritma dengan parameter nilai yang optimal. Hal ini bertujuan untuk representasi komputasi dari hasil pengamatan yang meruakan hasil dari pencarian pola-pola yang terkandung pada data. Pada penelitian ini model algoritma yang dipilih yaitu algoritma J48 dengan metode pengujian *split validation*, *cross validation*. *Cross Validation* salah satu metode yang mampu menemukan parameter terbaik dengan cara menguji besarnya error pada data testing (Gunawan, Palit, & Handojo, 2018).

### 5. *Evaluation*

Pada tahapan ini, pengujian dilakukan terhadap model untuk mendapatkan informasi model yang paling akurat. Evaluasi dan validasi menggunakan metode *cross validation* sebagai metode pengujian, *confusion matrix*. *Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining* (Rahman, Alamsah, Darmawidjadja, & Nurma, 2017) selain itu digunakan sebagai evaluasi dari penelitian ini yang nantinya akan menghasilkan *Accuracy*, dimana *Accuracy* ini akan menjadi tolak ukur keberhasilan suatu metode, karena semakin tinggi nilai *accuracy* semakin baik pula metode yang digunakan. Metode yang digunakan kali ini adalah metode J48.

## HASIL DAN PEMBAHASAN

### 1. Tahap Pemahaman Bisnis (*Business Understanding*)

Berdasarkan hasil pemeriksaan pasien penyakit tiroid yang diperoleh dari situs UCI *Machine Learning Repository* yang dibuat oleh the

Garavan Institute and J. Ross Quinlan, New South Wales Institute, Sydney, Australia. Dari data tersebut terdapat data pasien yang diperiksa sebanyak 7200 pasien dimana sebanyak 166 pasien terdeteksi normal, 368 pasien terdeteksi hipertiroid, dan 6666 pasien terdeteksi hipotiroid, terdiri dari 21 atribut sebagai atribut *predictor* dan 1 atribut hasil. Penelitian ini dilakukan untuk mengurangi jumlah resiko penyakit tiroid, karena gangguan fungsi tiroid seringkali sulit diidentifikasi sehingga sangat sering diabaikan. Untuk itu, diperlukan sebuah penelitian yang menerapkan metode untuk memprediksi penyakit tersebut yang nantinya akan mempermudah pasien dalam mendiagnosa dan deteksi dini. Pemanfaatan teknik klasifikasi dengan tingkat prediksi dan nilai akurasi yang tinggi dapat membantu mengatasi permasalahan tersebut, sehingga hasil diagnosis yang didapat menjadi lebih mudah, cepat dan akurat. Pada penelitian ini, untuk dapat menghasilkan tingkat prediksi dan nilai akurasi yang tinggi dilakukan penerapan algoritma J48 untuk memprediksi penyakit tiroid.

2. Tahap Pemahaman Data (*Data Understanding*)

Tahap *Data Understanding* dilakukan setelah menentukan tujuan dari penelitian, pada tahap ini dilakukan proses pengumpulan, analisa dan evaluasi kualitas data pada data tiroid *seperti memahami setiap atribut predictor dari penyakit tiroid, menentukan atribut manakah yang memang sangat berpengaruh pada hasil*, setelah itu, dilakukan penulisan data kembali dalam bentuk yang sama dan disimpan dengan format *Excel 97-2003 Worksheet (.xls)* agar dapat mempermudah pembacaan data untuk tahapan selanjutnya. Berikut atribut yan terdapat pada data penyakit tiroid:

Tabel 1. Atribut dan Nilai Kategori dalam Memprediksi Penyakit Tiroid

Nama Atribut	Keterangan
Age	Age of the patient (0.01–0.97). Continuous variable.
Sex	Sex of the patient, 0 (Male) 1 (Female). Binary variable.
On_thyroxine	0 (False) 1 (True). Binary Variable.
Query_on_thyroxine	0 (False) 1 (True). Binary Variable.
On_antithyroid_medication	0 (False) 1 (True). Binary Variable.
Sick	0 (False) 1 (True). Binary Variable.
Pregnant	0 (False) 1 (True). Binary Variable.
Thyroid_surgery	0 (False) 1 (True). Binary Variable.
I131_treatment	0 (False) 1 (True). Binary Variable.
Query_hypothyroid	0 (False) 1 (True). Binary Variable.
Query_hyperthyroid	0 (False) 1 (True). Binary Variable.

Lithium	0 (False) 1 (True). Binary Variable.
Goitre	0 (False) 1 (True). Binary Variable.
Tumor	0 (False) 1 (True). Binary Variable.
Hypopituitary	0 (False) 1 (True). Binary Variable.
Psych	0 (False) 1 (True). Binary Variable.
TSH	Real (0.0, 0.53)
T3	Real (0.0005, 0.18)
TT4	Real (0.0020, 0.6)
T4U	Real (0.017, 0.233)
FTI	Real (0.0020, 0.642)
Class	1. (normal) 2. (hyperthyroidism) 3. (hypothyroidism). Class variable.

Sumber: (UCI Repository, 2020)

3. Tahap Persiapan Data (*Data Preparation*)

Pada tahap *Data Preparation* dilakukan pengolahan data tiroid, mulai dari merubah format penulisan, mengecek data duplikat dan menghapusnya, begitupun dengan data kosong. Setelah itu dilakukan normalisasi data dengan mengubah data numerik menjadi kategori hal ini bertujuan agar data tersebut lebih mudah dipahami.

Tabel 2. Konversi Data Kedalam Kategorikal

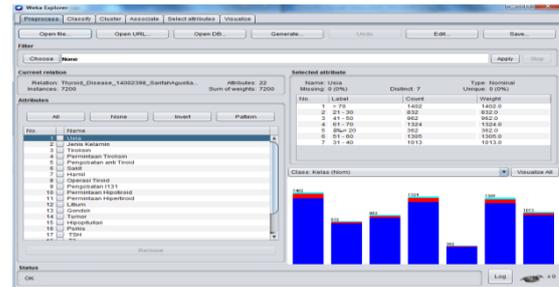
Nama Atribut	Keterangan	Sumber
Usia	≤ 20 Tahun	(Henry Crosby & Kandidat, 2016:432)
	21 – 30 Tahun	
	31 – 40 Tahun	
	41 – 50 Tahun	
	51 – 60 Tahun	
	61 – 70 Tahun	
Jenis Kelamin	0 (Laki-Laki)	UCI Repository
	1 (Perempuan)	
Tiroksin	0 (Tidak)	UCI Repository
	1 (Ya)	
Permintaan Tiroksin	0 (Tidak)	UCI Repository
	1 (Ya)	
Pengobatan anti Tiroid	0 (Tidak)	UCI Repository
	1 (Ya)	
Sakit	0 (Tidak)	UCI Repository
	1 (Ya)	

Hamil	0 (Tidak)	UCI Repository
	1 (Ya)	
Operasi Tiroid	0 (Tidak)	UCI Repository
	1 (Ya)	
Pengobatan I131	0 (Tidak)	UCI Repository
	1 (Ya)	
Permintaan Hipotiroid	0 (Tidak)	UCI Repository
	1 (Ya)	
Permintaan Hipertiroid	0 (Tidak)	UCI Repository
	1 (Ya)	
Litium	0 (Tidak)	UCI Repository
	1 (Ya)	
Gondok	0 (Tidak)	UCI Repository
	1 (Ya)	
Tumor	0 (Tidak)	UCI Repository
	1 (Ya)	
Hipopituitari	0 (Tidak)	UCI Repository
	1 (Ya)	
Psikis	0 (Tidak)	UCI Repository
	1 (Ya)	
	< 10,3 pmol/L > 35 pmol/L	
TSH	0.0 s/d 0.53	UCI Repository
T3	0.0005 s/d 0.18	UCI Repository
TT4	0.0020 s/d 0.6	UCI Repository
T4U	0.017 s/d 0.233	UCI Repository
FTI	0.0020 s/d 0.642	UCI Repository
Class	1 (Normal)	UCI Repository
	2 (Hipertiroid)	
	3 (Hipotiroid)	

Sumber: (Agustiani et al., 2020)

#### 4. Tahap Pemodelan (*Modeling*)

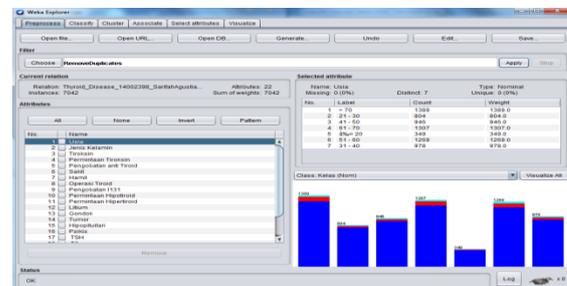
Tahap pemodelan ini diawali dengan menentukan dataset yang akan digunakan dalam penelitian ini yaitu data pasien penyakit tiroid dengan menggunakan format.csv seperti pada gambar berikut:



Sumber: (Agustiani et al., 2020)

Gambar 2. Menentukan Dataset

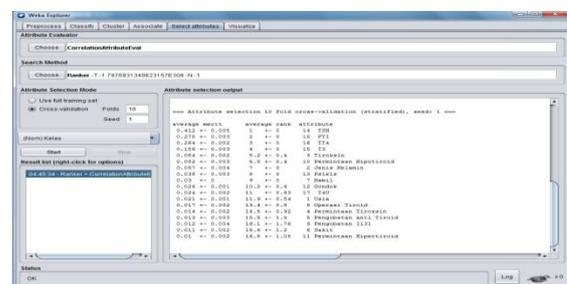
Selanjutnya dilakukan proses *remove duplicate* terhadap dataset untuk mengetahui apakah terdapat data duplikat dalam data tersebut. dalam hal ini dari 7200 data terdapat sekitar 158 data duplikat sehingga data yang digunakan menjadi 7042 data. berikut merupakan proses *remove duplicate*:



Sumber: (Agustiani et al., 2020)

Gambar 3. Proses *Remove Duplicate*

Proses berikutnya adalah melakukan *feature selection* dengan dua tahap yaitu *correlation attribute* dan *ranking selection* hal ini dilakukan untuk melihat atribut mana saja yang tidak memiliki pengaruh dalam pencarian model. Dalam hal ini semua atribut memiliki pengaruh terhadap pemodelan, sehingga semua atribut tetap digunakan. Berikut merupakan gambaran proses *feature selection attribute*:

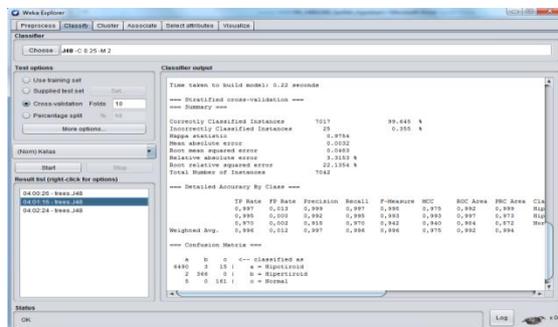


Sumber: (Agustiani et al., 2020)

Gambar 4. *Featur Selection Attribute*

Selanjutnya dilakukan *replace missing values*. *Replace missing values* digunakan untuk mengganti nilai yang hilang pada record. Nilai yang hilang akan digantikan dengan rata-rata seluruh data pada atribut tertentu. Data yang telah melalui tahap *replace missing values* kemudian diproses untuk selanjutnya dilakukan proses pemodelan dengan

menerapkan algoritma *J48* dan pengujiannya menggunakan metode *10 fold cross validation* sehingga menghasilkan model evaluasi yang diukur dengan nilai *Accuracy* dan *Curva ROC* sebagai pengukuran performa algoritma. Berikut merupakan pemodelan dalam penelitian ini:



Sumber: (Agustiani et al., 2020)  
 Gambar 5. Menentukan Algoritma dan Metode Pengujian

5. Evaluasi (*Evaluation*)

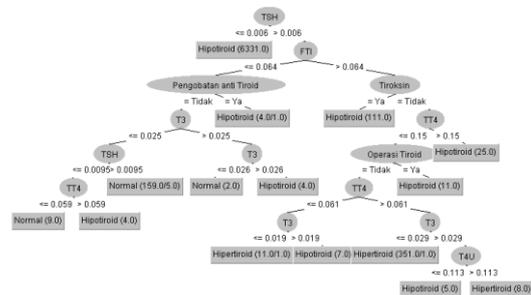
Tahap evaluasi merupakan hasil uji dari pemodelan yang dilakukan terhadap data pasien penyakit tiroid. Dalam hal ini proses evaluasi akan dilakukan dengan *confusion matrix* dan *curve ROC*. Model *confusion matrix* akan membentuk *matrix* yang terdiri dari *true positive*, *true negative*, *false positive*, dan *false negative*. Proses pemodelan dilakukan sebanyak dua kali percobaan dengan algoritma dan metode yang sama namun dengan pemilihan atribut yang berbeda. Percobaan pertama semua atribut digunakan dalam proses pemodelan, percobaan kedua hanya menggunakan 20 atribut dengan menghilangkan atribut hamil karena hasil correlation atribut menunjukkan nilai 0, namun kedua percobaan ini menghasilkan nilai akurasi yang sama dengan demikian semua atribut dalam data penyakit tiroid ini memiliki pengaruh terhadap pemodelan. Berikut merupakan gambaran dari hasil *confusion matrix* yang memperoleh nilai accuracy sebesar 99.645%:

Tabel 3. *Confusion Matrix*

99.645 %	Hipotiroid	Hipertiroid	Normal
Hipotiroid	6490	3	15
Hipertiroid	2	366	0
Normal	5	0	161

Sumber: (Agustiani et al., 2020)

Pemodelan ini juga menghasilkan role atau pola dari penerapan algoritma *J48* terhadap data penyakit tiroid dalam bentuk pohon keputusan seperti terlihat pada Gambar 6.



Sumber: (Agustiani et al., 2020)  
 Gambar 6. Visualisasi Tree Algoritma *J48*

Selain menghasilkan role atau pola dalam bentuk tree atau pohon keputusan, pemodelan ini juga menghasilkan role berupa teks. Dimana role ini dapat digunakan untuk pengimplementasian pada program, berikut merupakan role teks dari pemodelan *J48* terhadap data penyakit tiroid:

TSH <= 0.006: Hipotiroid (6331.0)  
 TSH > 0.006  
 | FTI <= 0.064  
 | | Pengobatan anti Tiroid = Tidak  
 | | | T3 <= 0.025  
 | | | | TSH <= 0.0095  
 | | | | | TT4 <= 0.059: Normal (9.0)  
 | | | | | TT4 > 0.059: Hipotiroid (4.0)  
 | | | | | TSH > 0.0095: Normal (159.0/5.0)  
 | | | | | T3 > 0.025  
 | | | | | | T3 <= 0.026: Normal (2.0)  
 | | | | | | T3 > 0.026: Hipotiroid (4.0)  
 | | | | | | Pengobatan anti Tiroid = Ya: Hipotiroid (4.0/1.0)  
 | | | | | FTI > 0.064  
 | | | | | | Tiroksin = Ya: Hipotiroid (111.0)  
 | | | | | | Tiroksin = Tidak  
 | | | | | | | TT4 <= 0.15  
 | | | | | | | | Operasi Tiroid = Tidak  
 | | | | | | | | | TT4 <= 0.061  
 | | | | | | | | | | T3 <= 0.019: Hipertiroid (11.0/1.0)  
 | | | | | | | | | | T3 > 0.019: Hipotiroid (7.0)  
 | | | | | | | | | | | TT4 > 0.061  
 | | | | | | | | | | | | T3 <= 0.029: Hipertiroid (351.0/1.0)  
 | | | | | | | | | | | | T3 > 0.029  
 | | | | | | | | | | | | | T4U <= 0.113: Hipotiroid (5.0)  
 | | | | | | | | | | | | | T4U > 0.113: Hipertiroid (8.0)  
 | | | | | | | | | | | | | | Operasi Tiroid = Ya: Hipotiroid (11.0)  
 | | | | | | | | | | | | | | TT4 > 0.15: Hipotiroid (25.0)

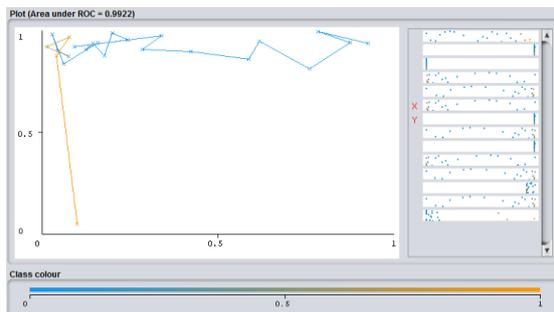
Dari pemodelan ini juga menghasilkan beberapa nilai *precision*, *recall*, dan ROC untuk masing-masing *class* dan juga nilai rata-rata dari semua kelas. Dimana ROC yang di hasilkan adalah 0.992. Berikut merupakan tabel nilai *precision*, *recall* dan ROC:

Tabel 3. Nilai *Precision*, *Recall* Dan ROC

99.645 %	Precision	Recall	ROC
Hipotiroid	0,999	0,997	0,992
Hipertiroid	0,992	0,995	0,997
Normal	0,915	0,970	0,984
Rata-Rata	0,997	0,996	0,992

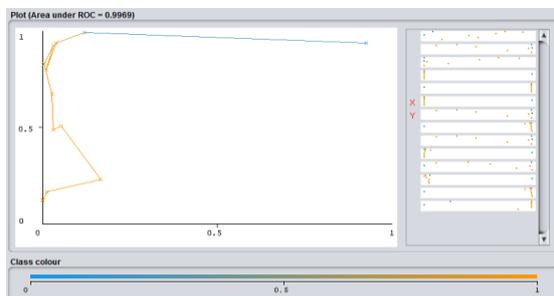
Sumber: (Agustiani et al., 2020)

Sedangkan untuk kurva ROC setiap kelasnya dapat dilihat pada Gambar 7, Gambar 8, dan Gambar 9. Dimana X merupakan *False Positive* dan Y merupakan nilai *True Positive* untuk setiap *class*



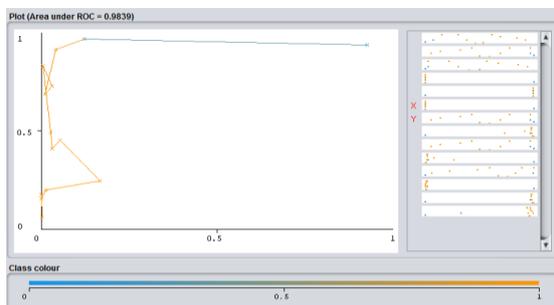
Sumber: (Agustiani et al., 2020)

Gambar 7. Curve ROC untuk *Class* Hipotiroid



Sumber: (Agustiani et al., 2020)

Gambar 8. Curve ROC untuk *Class* Hipertiroid



Sumber: (Agustiani et al., 2020)

Gambar 9. Curve ROC untuk *Class* Normal

## KESIMPULAN

Berdasarkan hasil penelitian terhadap penerapan algoritma *J48* dalam memprediksi tiroid dengan data sebanyak 7200 *record* diperoleh nilai akurasi sebesar 99.645%. Sedangkan nilai AUC

yang dihasilkan adalah 0,992 yang berarti akurasi ini memiliki tingkat klasifikasi yang sangat baik. Dengan demikian dapat disimpulkan bahwa algoritma *J48* memiliki tingkat akurasi yang baik dalam mendeteksi penyakit tiroid. Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Dalam penelitian selanjutnya disarankan untuk menggunakan dataset yang berbeda dengan jumlah data dan atribut yang lebih banyak, sehingga hasil pengukuran dapat lebih baik lagi.
2. Mengimplementasikan metode ke dalam *Graphic User Interface* (GUI) yang mengikuti kemajuan IPTEK, seperti GUI yang berbasis *mobile computing* atau aplikasi yang didukung oleh android dan yang lainnya

## REFERENSI

- Agustiani, S., Mustopa, A., Saryoko, A., Gata, W., & Wildah, S. K. (2020). Penerapan Algoritma *J48* Untuk Deteksi Penyakit Tiroid. *Paradigma – Jurnal Informatika Dan Komputer*, 22(2).
- Astuti, T., Mujiati, I., Ayu, D., Ristianah, V., & Lestari, W. A. (2016). Penerapan Algoritme *J48* Untuk Prediksi. *Jurnal Telematika*, 9(2), 1–10.
- Decroli, G. P., & Decroli, E. (n.d.). Laporan Kasus Krisis Tiroid pada Wanita Multipara Usia 42 Tahun, 8(1), 178–181.
- Diwandari, S., & Setiawan, N. A. (2015). Perbandingan Algoritme *J48* Dan *Nbtree* Untuk Klasifikasi Diagnosa Penyakit Pada Soybean. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2015*(Sentika), 205–212.
- Dua, D. and Graff, C. (2019). UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml]. (Dua, D. and Graff, 2019): University of California, School of Information and Computer Science..
- Gunawan, A., Palit, H. N., & Handojo, A. (2018). Aplikasi Rekomendasi Metode Analisis Sesuai dengan Karakter Data. *Jurnal Infra*, 6(1).
- Handayani, P., Nurlelah, E., Raharjo, M., & Ramdani, P. M. (2019). PREDIKSI PENYAKIT LIVER DENGAN MENGGUNAKAN METODE, 4(1), 55–59.
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA *C4*. 5. *Jurnal Media Infotama*, 11(2), 130–138.
- Henry Crosby, V. P., & Kandidat, M. A. M. (2016). Pola kelainan tiroid di RSUP Prof. Dr. R. D. Kandou Manado periode Januari 2013 -

- Desember 2015. *Jurnal E-Clinic*, 4(1), 430–437.
- Kementerian Kesehatan. (2015). Infodatin Tiroid.
- Kurniawan, L. B., & Arif, M. (2015). DIAGNOSIS TIROID (Diagnosis of Thyroid) Liong. *INDONESIAN JOURNAL OF CLINICAL PATHOLOGY AND MEDICAL LABORATORY*, 7(1), 2–7.
- Lishania, I., Goejantoro, R., & Nasution, N. (2019). Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma ( J48 ) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda Comparison of the Classification for Naive Bayes Method and the Decision Tree Algorithm ( , 10, 135–142.
- Lombok, H., Mataram, R., Astuti, T., Ferinanto, T., Informatika, J. T., & Informasi, J. S. (2016). Diagnosis Penyakit Parkinson Berdasarkan Kombinasi, 28–29.
- Putri, A. W. (2019). Gangguan Tiroid sebagai Beban Baru Negara.
- Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- Santoso, W. B., Santoso, B., Romadhon, & Sukandar. (2016). PERANGKAT SCINTIGRAFI UNTUK TIROID SC-12 pada organ tersebut . Proses pencitraan menggunakan perangkat scintigradi dilakukan untuk mendiagnosis fungsi kelenjar tiroid ( gondok ) [ 1 ] . Proses pencitraan tiroid radionuklida ke pasien . Sistem deteksi pera. *JURNAL PERANGKAT NUKLIR*, 10(1978), 1–10.
- Septina, L., Zaid, M., Zuraidah, E., Stephanie, A., Septina, L., Zaid, M., ... Stephanie, A. (2019). Pratista Patologi Peran Rearrangement RET / PTC pada Karsinoma Papiler Tiroid Pratista Patologi, 6(1).
- Sidiq, U., Aaqib, S. M., & Khan, R. A. (2019). Diagnosis of Various Thyroid Ailments using Data Mining Classification Techniques, 5(1), 131–136.
- Suyanto. (2017). *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung.
- T, L. D., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (Second Edi). Canada: Simultaneously.
- Widayati, P., Ariyanto, A., Setiyowati, S., Yulianti, V., Teknologi, P., Setu, K. P., & Selatan, T. (2019). Preparasi Pereaksi Kit Immunoradiometricassa ( IRMA ) Thyroglobulin ( TGB ) Untuk Deteksi Kanker Tiroid, (April), 1–8.
- Wijonarko, B. (2018). Perbandingan Algoritma Data Mining Naive Bayes Dan Bayes Network Untuk Mengidentifikasi Penyakit Tiroid. *PILAR Nusa Mandiri*, 14(1), 21–26. Retrieved from <https://media.neliti.com/media/publications/227524-perbandingan-algoritma-data-mining-naive-2eb3d17d.pdf>

## PROFIL PENULIS

### Sarifah Agustiani

Tahun 2019 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi Universitas Bina Sarana Informatika dan Sekarang sedang melanjutkan pendidikan Program Magister (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

### Ali Mustopa

Tahun 2019 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Nusa Mandiri dan Sekarang sedang melanjutkan pendidikan Program Magister (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

### Andi Saryoko

Tahun 2008 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Nusa Mandiri dan tahun 2011 lulus dari Magister (S2) Program Studi Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.

### Windu Gata

Tahun 1999 lulus dari Program Strata Satu (S1) Program Studi Teknik Informatika Universitas Budi Luhur, Tahun 2008 lulus dari Program Magister (S2) Program Studi Ilmu komputer Universitas Budi Luhur dan Tahun 2015 lulus dari program Doktor (S3) Program Studi Manajemen Pendidikan Universitas Negeri Jakarta.

### Siti Khotimatul Wildah

Tahun 2019 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi Universitas Bina Sarana Informatika dan Sekarang sedang melanjutkan pendidikan Program Magister (S2) Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.