PENERAPAN ALGORITMA *DECISION TREE* PADA PREDIKSI RISIKO TERSERANG *STROKE*

SKRIPSI

OLEH
BAGAS HARMADI
NIM. 18610113

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

PENERAPAN ALGORITMA *DECISION TREE* PADA PREDIKSI RISIKO TERSERANG *STROKE*

SKRIPSI

Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)

Oleh
Bagas Harmadi
NIM. 18610113

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

PENERAPAN ALGORITMA *DECISION TREE* PADA PREDIKSI RISIKO TERSERANG *STROKE*

SKRIPSI

Oleh
Bagas Harmadi
NIM. 18610113

Telah Diperiksa dan Disetujui Untuk Diuji
Malang, 07 Desember 2022

Dosen Pembimbing I	Dosen Pembimbing II
Hisyam Fahmi, M.Kom.	Erna Herawati, M.Pd.
NIP. 19890727 201903 1 018	NIDT. 19760723201802012222

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PENERAPAN ALGORITMA *DECISION TREE* PADA PREDIKSI RISIKO TERSERANG *STROKE*

SKRIPSI

Oleh
Bagas Harmadi
NIM. 18610113

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)
Tanggal, 21 Desember 2022

Ketua Penguji	: Abdul Aziz, M.Si.	.....
Anggota Penguji I	: Angga Dwi Mulyanto, M.Si	.....
Anggota Penguji II	: Hisyam Fahmi, M.Kom.	.....
Anggota Penguji III	: Erna Herawati, M.Pd	.....

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama	: Bagas Harmadi
NIM	: 18610113
Program Studi	: Matematika
Fakultas	: Sains dan Teknologi
Judul Skripsi	: Penerapan Algoritma *Decision Tree* Pada Prediksi Risiko Terserang *Stroke*.

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar – benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut

Malang, 21 Desember 2022
Yang membuat pernyataan,

Bagas Harmadi
NIM. 18610113

MOTO

"DIMA BUMI DIPIJAK,
DISINAN LANGIK DI JUNJUANG"

PERSEMBAHAN

Bismillahirrohmanirrahim …
Alhamdulillah, Alhamdulillah, Alhamdulillahirobbil alamin …

Berkisah ketika April, 2018 waktu pertama kali saya merantau jauh dari orang tua, Mulut tak bisa berucap, senyum tak lagi terpancar, namun itulah kisah terindah, dan disitu saya menemukan banyak kenangan dan pengalaman yang tak akan terlupakan sampai tua pun.. -Bagas H

Kusembahkan karya ini untuk kedua orang yang sangat kucintai, orang yang kukasihi dan kusayangi.
BAPAK DAN IBU TERCINTA
Terima kasih banyak kepada orang tua saya bapak Desharfius dan Ibu Syafneli yang telah memberikan kasih sayang dan semuanya kepada saya, yang menjadi penyemangat, motivasi sehingga skripsi ini bisa selesai.

KATA PENGANTAR

Assalamu'allaikum Warrahmatullahi Wabarakatuh

Pertama sekali penulis ucapkan segala puji bagi Allah karena telah memberikan rahmat, taufik dan hidayah-Nya, sehingga penulis telah menyempurnakan penulisan proposal skripsi yang dijadikan sebagai syarat untuk menerima gelar sarjana di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada bidang Matematika di Fakultas Sains dan Teknologi.

Kemudian penulis ucapkan banyak terimakasih kepada berbagai pihak yang telah memberikan bimbingan, arahan dan dukungan selama proses penyempurnaan proposal skripsi ini. Terima kasih dan penghargaan sebesar-besarnya penulis ucapkan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc., selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Hisyam Fahmi, M,Kom., selaku dosen pembimbing I yang telah memberikan banyak bimbingan, arahan, dukungan serta perbaikan demi kebaikan penyusunan skripsi.
5. Erna Herawati, M.Pd., selaku dosen pembimbing II yang telah memberikan bimbingan, dukungan serta perbaikan kepada penulis.
6. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.

7. Seluruh keluarga, khususnya untuk kedua orang tua yang selalu mendukung dari segala aspek.
8. Seluruh sahabat, teman-teman kost mariana, yang telah memberikan banyak dukungan motivasi dalam penyusunan skripsi hingga selesai.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca maupun bagi penulis, serta dapat dijadikan sebagai penambah wawasan ilmu matematika terutama dalam bidang matematika komputasi.

Malang, 21 Desember 2022

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

ABSTRAK

Harmadi, Bagas. 2022. **Penerapan Algoritma *Decision Tree* Pada Prediksi Risiko Terserang *Stroke*.** Skripsi. Program Studi Matematika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim, Malang. Pembimbing (1) Hisyam Fahmi, M.Kom. (2) Erna Herawati, M.Pd.

Kata Kunci*:* *Decision Tree*; Prediksi Terserang *Stroke*; RapidMiner

Kesehatan merupakan aspek yang paling penting bagi kehidupan manusia. Saat ini banyak penyakit yang diderita yang disebabkan oleh kuman, virus dan bakteri, tetapi penyebab yang paling utama adalah kebiasaan atau pola hidup yang tidak sehat. *Stroke* menjadi salah satu dari penyakit-penyakit yang diderita tersebut. Oleh karena itu, diperlukan sebuah analisis mengenai prediksi seseorang akan terkena penyakit, seperti penelitian terkait prediksi terserang *Stroke*. Penelitian ini bertujuan untuk mengetahui model dan hasil prediksi risiko terserang *Stroke* pada manusia dengan menggunakan algoritma *Decision Tree*. Metode ini memiliki tingkat akurasi yang baik dan efektif dalam pengambilan keputusan. Dari delapan faktor yang menjadi penyebab seseorang terkena *Stroke*, yaitu Jenis Kelamin, Umur, Hipertensi, Sakit Jantung, Jenis Tempat Tinggal, Kadar Glukosa, Index Mass Tubuh (BMI) dan Status Merokok didapatkan hasil prediksi risiko terserang *Stroke* dengan 360 data menggunakan algoritma *Decision Tree C4.5* dan bantuan *tool RapidMiner* dengan persentase keakuratan 68.89%, *precision* sebesar 68.68%, *recall* sebesar 69.4% dan Kurva ROC diperoleh AUC 0.726. Hasil ini menunjukkan bahwa model pada metode *Decision Tree* termasuk dalam klasifikasi cukup (*fair classification*). Selanjutnya, dengan menggunakan 360 data diperoleh model *tree* dengan hasil 28 aturan keputusan, serta faktor yang paling dominan yang menjadi penyebab *Stroke* adalah Umur pada usia diatas 65 Tahun.

ABSTRACT

Harmadi, Bagas. 2022. **Application of the Decision Tree Algorithm in Predicting the Risk of Having a Stroke.** Thesis. Mathematics Study Program. Faculty Science and Technology. Universitas Islam Negeri Maulana Malik Ibrahim, Malang. Supervisor (1) Hisyam Fahmi, M.Kom. (2) Erna Herawati, M.Pd.

Keywords: Decision Tree; RapidMiner; Stroke Prediction

Health is the most important aspect of human life. Germs, viruses and bacteria cause many illnesses, but the main causes are unhealthy habits or lifestyles. Stroke is one of these diseases. Therefore, an analysis is needed regarding the prediction that someone will get a disease, such as research related to stroke prediction. This study aims to determine the model and predictive results of stroke risk in humans using the Decision Tree algorithm. This method has a good level of accuracy and it is effective in decision-making. From that eight factors that cause a person to have a Stroke are Gender, Age, Hypertension, Heart Disease, Type of Residence, Glucose Levels, Body Mass Index (BMI) and Smoking Status, the risk prediction results for Stroke are obtained with 360 data using the Decision Tree algorith. C4.5 and the result of the RapidMiner tool with an accuracy percentage of 68.89%, precision of 68.68%, recall of 69.4% and the ROC curve obtained AUC 0.726. These results indicate that the model in the Decision Tree method is included in the appropriate classification (fair classification). Furthermore, using 360 data, a tree model was obtained with the results of 28 decision rules, and the most dominant factor that causes Stroke is the age that over 65 years.

مستخلص البحث

حرمادي، باغاس (٢٠٢٢) **تطبيق خوارزمية شجرة القرار في توقع مخاطر الإصابة بسكتة دماغية.** البحث الجامعي. قسم الرياضيات.كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول (١) هشام فهمي، الماجستير. المشرف الثاني (٢) إيرنا هيراواتي، الماجستير.

الكلمات الأساسية: أشجار القرار؛ توقع السكتة الدماغية؛ رابيدماينر

الصحة هي أهم جانب في حياة الإنسان. تسبب الجراثيم والفيروسات والبكتيريا العديد من الأمراض، ولكن الأسباب الرئيسية هي العادات أو أنماط الحياة غير الصحية. السكتة الدماغية هي من إحدى الأمراض. لذلك، هناك حاجة إلى تحليل فيما يتعلق بالتنبؤ بإصابة شخص ما بمرض، مثل البحث المتعلق بالتنبؤ بالسكتة الدماغية. تهدف هذه الدراسة إلى تحديد النموذج والنتائج التنبؤية لمخاطر السكتة الدماغية لدى البشر باستخدام خوارزمية شجرة القرار. تتمتع هذه الطريقة بمستوى جيد من الدقة وهي فعالة في اتخاذ القرار. من بين العوامل الثمانية التي تسبب إصابة الشخص بالسكتة الدماغية، وهي الجنس (*Gender*) والعمر (*Age*) وارتفاع ضغط الدم (*Hypertension*) وأمراض القلب (*Heart Disease*)ونوع المقيم (*Resident Type*) ومتوسط مستوى الجلوكوز (*Avg Glucose Level*) ومؤشر كتلة الجسم (*Body Mass Index*) وحالة التدخين(*Smoking Status*) ، يتم الحصول على نتائج التنبؤ بمخاطر الإصابة بالسكتة الدماغية باستخدام ٣٦٠ البيانات باستخدام خوارزمية شجرة القرار ج٤,٥ (*Decision Tree C4.5*) ومساعدة أداة رابيدماينر (*RapidMiner*) بنسبة دقة تبلغ ٦٨,٨٩٪ ودقة (*Precision*) حصل ٦٨,٦٨٪ واسترجاع (*Recall*)حصل ٦٩,٤٪ وخاصية تشغيل مستقبل كورفا (*Kurva Receiver Operating Characteristic*) على المنطقة تحت المنحنى 0.726 (*Area Under Curve*) تشير هذه النتائج إلى أن النموذج الموجود في طريقة شجرة القرار مدرج في التصنيف المناسب) تصنيف عادل .(*Fair Classification*/علاوة على ذلك، باستخدام بيانات ٣٦٠، تم الحصول على نموذج شجرة بنتائج ٢٨ قاعدة قرار، والعامل الأكثر انتشارًا الذي يسبب السكتة الدماغية هو العمر (*Age*) الذي يزيد عن ٦٥ عامًا.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Decision Tree menjadi salah satu metode yang dapat diterapkan untuk melakukan proses prediksi. Metode ini menghasilkan suatu model yang dapat memprediksi kategori Data dengan cara mempelajari aturan penentuan kategori berdasarkan fitur-fitur yang dimiliki oleh Data (Ceballos, 2019). *Decision Tree* juga memiliki tingkat akurasi yang tinggi saat diterapkan untuk jumlah Data yang besar dibandingkan dengan metode lain. Oleh karena itu, dalam indrustri kesehatan dan medis keakuratan prediksi sebuah penyakit sangatlah penting dan memerlukan keputusan yang efektif dalam mengambil suatu analisa dan keakuratan suatu penyakit yang diderita pasien (Rifai, 2013).

Salah satu penyakit yang bisa di prediksi adalah *Stroke*. *Stroke* adalah penyakit yang terjadi ketika suplai darah ke otak terganggu karena terdapat pembuluh darah yang pecah atau tersumbatnya dalam bentuk darah yang menggumpal. Akibat dari *Stroke* dapat menyebabkan kesulitan dalam melaksanakan pekerjaan ringan, misalnya bergerak, berjalan dan hilangnya keseimbangan, hilangnya kesadaran atau pingsan dan sakit kepala tanpa adanya penyebab. Akibat *Stroke* tergantung seberapa parah kerusakannya pada bagian otak yang terluka. Bahkan pasien yang mengalami *Stroke* yang sangat serius dapat mengalami kematian secara mendadak (American Stroke Association, 2020).

Setelah penyakit jantung dan kanker, *Stroke* menjadi salah satu penyakit yang menyebabkan kecacatan terbanyak dan menjadi penyebab kematian terbesar ketiga didunia. Di Indonesia *Stroke* mempunyai skala *prevalensi* yang cukup

tinggi. Merujuk pada Data yang dikemukakan oleh kementrian kesehatan, sebagaimana diagnosis dokter bahwa skala *prevalensi Stroke* di indonesia pada penduduk yang berusia di atas 15 tahun adalah 10.9%, atau diperkirakan 2.120.362 orang. Berdasarkan kelompok umur, dapat dilihat bahwa kejadian *Stroke* paling tinggi pada kelompok umur 55-64 tahun (33.3%) dan proporsi penderita *Stroke* paling rendah pada kelompok umur 15-24 tahun. Sebagian besar penderita *Stroke* tinggal diperkotaan (29.5%), sedangkan dengan masyarakat pedesaan sebesar 36.1% (InfoDATIN, 2019).

Stroke dapat disebabkan oleh beberapa faktor, antara lain tekanan darah tinggi, riwayat atrial fibrilasi, kolesterol, diabetes mellitus, dan penyakit jantung (Hopkins, n.d.). Pengobatan *Stroke* biasanya dilaksanakan secara manual. Praktiknya, pasien yang terkena *Stroke* mendatangi tempat pemeriksaan dokter spesialis syaraf untuk diperiksa dengan memberikan beberapa pertanyaan seperti gejala yang dialami, dan penyebab-penyebab yang memungkinkan pasien terserang *Stroke*. Kemudian dokter akan mendiagnosis kemungkinan tingkat resiko yang dialami oleh pasien yang tersrang *Stroke*. Pengobatan ini tentu menimbulkan masalah tersendri, seperti masalah dalam pembiayaaan dan waktu yang sangat lama. Maka sangat diperlukan sebuah analisis atau penelitian yang dapat memprediksi kemungkinan seseorang terkena *Stroke*, sehingga dapat melakukan pencegahan sebelum *Stroke* menyerang seseorang.

Menentukan metode yang tepat dalam memprediksi tingkat risiko terserang *Stroke* menjadi sangat penting karena dapat mempengaruhi hasil dan kesimpulan yang diperoleh. Maka *Decision Tree* merupakan suatu metode yang mudah dalam menginterpretasikan suatu hasil prediksi. *Decision Tree* sering kali dipergunakan

dalam proses pengklasifikasian suatu objek yang berdasarkan pada Data dengan nilai selisih kecil dengan jarak tetangga yang sangat dekat dengan objek. *Decision Tree* bertujuan untuk mem-*break down* proses pengambilan keputusan yang komplek menjadi simple dengan pohon keputusan. Prinsip umum dari algoritma *Decision Tree* adalah menentukan atribut label dan membagi Data menjadi *Data training* dan *Data testing*, Data tersebut kemudian diproses menggunakan algoritma *Decision Tree* berdasarkan pada parameter yang telah ditentukan. Sehingga metode ini cocok digunakan dalam memprediksi risiko terserang *Stroke*.

Seperti yang dijelaskan Al-Qur'an Q.S Yusuf ayat 47-49 yang berbunyi (Kemenag, 2002):

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَأَبًا فَمَا حَصَدتُّمْ فَذَرُوهُ فِي سُنۢبُلِهِۦٓ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ (٤٧) ثُمَّ يَأْتِي مِنۢ بَعْدِ ذَٰلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُحْصِنُونَ (٤٨) ثُمَّ يَأْتِيْ مِنۢ بَعْدِ ذَٰلِكَ عَامٌ فِيْهِ يُغَاثُ النَّاسُ وَفِيْهِ يَعْصِرُوْنَ (٤٩)

Artinya:

"Dia (Yusuf) berkata, "Agar kamu bercocok tanam tujuh tahun (berturut-turut) sebagaimana biasa; kemudian apa yang kamu tuai hendaklah kamu biarkan di tangkainya kecuali sedikit untuk kamu makan. 47. Kemudian setelah itu akan datang tujuh (tahun) yang sangat sulit, yang menghabiskan apa yang kamu simpan untuk menghadapinya (tahun sulit), kecuali sedikit dari apa (bibit gandum) yang kamu simpan. 48. Setelah itu akan datang tahun, di mana manusia diberi hujan (dengan cukup) dan pada masa itu mereka memeras (anggur)". 49. (QS. Yusuf ayat 47-49).

Ayat di atas menjelaskan bahwa nabi yusuf AS memprediksi akan datangnya musim paceklik selama tujuh tahun ketika menafsirkan mimpi seorang raja. Nabi Yusuf AS memberitahu raja berdasarkan wahyu dari Allah untuk mempersiapkan perbekalan selama tujuh tahun dalam menghadapi musim tersebut. Prediksi Nabi Yusuf dapat membantu Raja mengetahui akan datangnya suatu musim yang

sangat sulit, sehingga Raja dapat menyiapkan perbekalan yang cukup dalam menghadapi kesulian-kesulitan yang akan terjadi. Ayat ini sesuai dengan tujuan dan manfaat penelitian yang sedang penulis lakukan, yaitu untuk mengetahui atau prediksi kemungkinan sesorang terserang *Stroke*, sehingga seseorang dapat melakukan pencegahan dan menghindari penyebab-penyebab dari kemungkinan terserang *Stroke*.

Kajian terkait prediksi menggunakan algoritma *Decision Tree* pernah dilakukan oleh peneliti sebelumnya. *Pertama*, penelitian yang menggunakan metode *Decision Tree* dengan Penerapan Algoritma C4.5" yang dilakukan oleh Susi Mashlahah pada tahun 2013. Hasil penelitian ini menjelaskan bahwa hasil uji coba menggunakan 60 record Data sampel, menghasilkan sebesar 65.51% tingkat akurasi, 79 record Data sampel yang digunakan memperoleh 70.96% tingkat akurasi dan 90 record Data sampel memperoleh tingkat akurasi sebesar 82.79% (Mashlahah, 2013). *Kedua*, penelitian yang dilakukan oleh Isa Iskandar, dkk pada tahun 2019. Hasil penelitian tersebut dengan model kelulusan mahasiswa menghasilkan nilai akurasi 73.19% dengan nilai AUC 0.806 (Rohman & Rufiyanto, 2019). Berdarkan hal tersebut maka peneliti ingin melakukan prediksi risiko manusia terserang *Stroke* menggunakan algoritma *Decision Tree*.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka diperoleh rumusan masalah yaitu bagaimana model dan hasil prediksi risiko terserang *Stroke* pada manusia dengan menggunakan algoritma *Decision Tree*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas maka diperoleh tujuan penelitian yaitu untuk mengetahui model dan hasil prediksi risiko terserang *Stroke* pada manusia dengan menggunakan algoritma *Decision Tree*.

1.4 Manfaat Penelitian

Manfaat pada penelitian ini yaitu:

1. Memperluas wawasan, pengetahuan serta pemahaman mengenai penyebab kemungkinan manusia dapat terserang *Stroke*.
2. Dapat melakukan upaya preventif terhadap faktor-faktor yang menjadi penyebab terserang *Stroke.*

1.5 Batasan Masalah

Penelitian ini dibatasi pada beberapa hal, yaitu:

1. Data yang di gunakan merupakan data pasien *Stroke* yang berasal dari *World Health Organization* (WHO) yang diambil dari repositori KAGGLE.
2. Aspek lokasi dan waktu penelitian dapat diabaikan pada penelitian *general* karena di anggap tidak berpengaruh dalam melakukan klasifikasi dan pembentukan model.

1.6 Defenisi Istilah

1. *Accuracy*: Tolak ukur yang digunakan untuk mengetahui seberapa tepat suatu pola klasifikasi dalam memprediksi kelas data.
2. Algoritma: Sekumpulan instruksi yang terstruktur dan terbatas yang diimplementasikan kedalam bentuk program komputer untuk menyelesaikan suatu masalah.

3. Atribut: Spesifikasi yang membuktikan properti dari suatu objek, element atau data.
4. C4.5: Algoritma *Decision Tree Learning* (algoritma pembelajaran pohon keputusan) yang dikembangkan oleh Ross Quinlan.
5. CART: Algoritma *Decision Tree learning* (algoritma pembelajaran pohon keputusan) yang dikembangkan oleh Leo Breiman.
6. *Clustering*: Suatu teknik yang dapat digunakan untuk mengenali kelompok-kelompok yang dihasilkan dari pengelompokan unsur-unsur yang lebih kecil berdasarkan suatu kemiripan satu sama lain.
7. CSV: Suatu format data dalam basis data dimana setiap *record* dipisahkan dengan tanda koma (,).
8. *Data Testing*: Data yang digunakan untuk mengetahui performa algortima yang sudah dilatih sebelumnya ketika menemukan data yang belum pernah dilihat sebelumnya.
9. *Data Training*: Data yang digunakan untuk melatih algoritma guna mencari model yang pas.
10. *Database*: Kumpulan data atau informasi yang tersimpan secara sistematis.
11. *Dataset*: Merupakan kumpulan data atau informasi.
12. *Decision Tree*: Metode klasifikasi yang menggunakan struktur pohon untuk mengambil suatu keputusan secara efektif.
13. Diabetes: penyakit kronis yang ditandai dengan tingginya gula darah.
14. *Entropy*: nilai informasi yang menyatakan ukuran ketidakpastian (*impurity*) dari attribut dari suatu kumpulan obyek data dalam satuan bit.

15. Fibrilasi: Denyut jantung tidak teratur yang menyebabkan aliran darah tidak lancar.
16. *Gain*: Ukuran efektifitas suatu variabel dalam mengklasifikasikan data.
17. Hipertensi: Suatu kondisi ketika tekanan darah terhadap dinding arteri terlalu tinggi.
18. ID3: Algoritma *Decision Tree learning* (algoritma pembelajaran pohon keputusan).
19. Interpretasi: Proses pemberian pendapat/gagasan, kesan tentang yang dinilai.
20. Interval: Waktu diantara dua kejadian.
21. KAGGLE: Situs atau platform untuk berkompetisi dalam bidang science.
22. KDD: *Knowledge discovery* in *Database* atau kegiatan yang meliputi pengumpulan, pemakaian data, pola atau hubungan dalam set data besar.
23. Kemenkes: Kementrian kesehatan.
24. Klasifikasi: Penyusunan bersistem dalam berkelompok atau golongan menurut kaidah kaidah atau standart yang ditetapkan.
25. Kolesterol: Senyawa lemak berlilin yang sebagai besar diproduksi di hati.
26. Kuantitatif: Penelitian ilmiah yang sistematis yang menggunakan model-model matematis.
27. mg/dl: miligram per desiliter dan merupakan kisaran kadar gula darah
28. Partisi: Pembagian suatu objek ke dalam beberapa bagian dengan tujuan tertentu.
29. Prediksi: Memperkirakan secara sistematis tentang kemungkinan sesuatu yang paling mungkin akan terjadi dimasa depan.

30. Prevalensi: Proporsi dari populasi yang memiliki karakteristik tertentu dalam jangka waktu tertentu.
31. Query SQL: Syntax atau perintah yang dipakai untuk mengakses dan menampilkan Data.
32. *RapidMiner*: Platform perangkat lunak untuk melakukan analisis data.
33. *STROKE*: Kondisi yang terjadi ketika pasokan darah ke otak terganggu atau bekurang akibat penyumbatan (*Stroke* hemoragik).
34. WHO: Organisasi Kesehatan Dunia.

BAB II
KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 *Data Mining*

1. Pengertian

Data Mining secara sederhananya ialah eksplorasi, penggalian maupun pencarian informasi yang terbaru yang bertujuan untuk menemukan aturan atau pola yang telah ditentukan dari sekian banyak data yang ada (Beynon-Davies, 2004). *Data Mining* dipahami juga sebagai sebuah rangkaian prosedur dalam memperoleh nilai tambah berupa informasi yang belum diketahui secara manual selama ini dari beberapa kumpulan data (Pramudiono, 2003). Istilah *data mining* seringkali dikenal sebagai KKD, yaitu *knowledge discovery in Database*. KDD merupakan suatu kegiatan dengan mengaitkan penggunaan dan pengumpulan data histori untuk menemukan kumpulan *dataset* yang berukuran besar pada desain-desain atau hubungannya.

Data mining merupakan kegiatan untuk memperoleh format yang lebih menarik dari sebagian besar data yang berada dalam *Database*, *Data warehouse* maupun tempat dokumentasi-dokumentasi lainnya yang berbentuk informasi. Dengan kata lain, *Data Mining* diartikan sebagai proses untuk mendapatkan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Namun, pola yang didapatkan harus penuh akan makna dan harus memberikan manfaat atau keuntungan, keuntungan tersebut umumnya dalam hal ekonomi. Sedangkan kebutuhan Datanya dalam jumlah yang sangat besar (Han dkk., 2006). *Data mining* memiliki keterkaitan dengan disiplin ilmu lainnya, misalnya berkaitan dengan *Datawarehousing*, *Database system*, statistik, *information retrieval*,

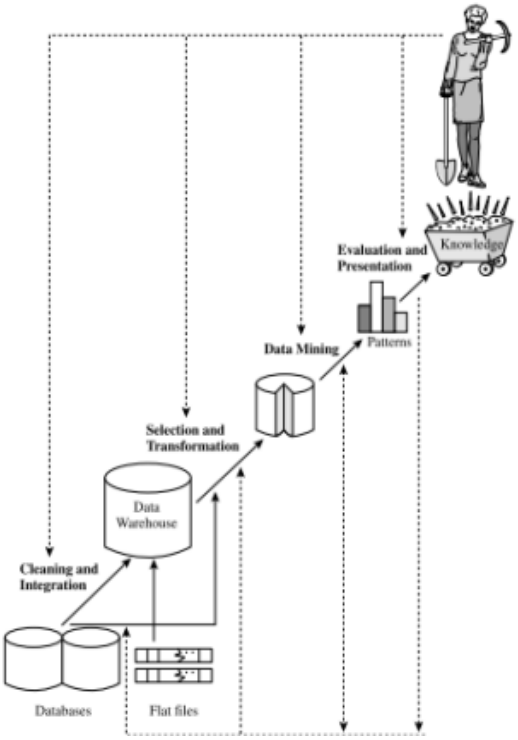
machine learning, dan komputasi tingkat tinggi. *Data Mining* juga didukung oleh disiplin kajian lainnya, seperti pengenalan pola, *neural network*, *spatial Data analysis*, *signal processing*, *image Database* (Han dkk., 2006).

Maka dapat diambil kesimpulan bahwa *data mining* merupakan sebuah cara untuk mengekstrak informasi-informasi yang sangat berharga dan yang tersembunyi atau terpendam dalam *Database* atau sejumlah kumpulan data untuk mengetahui pola yang belum pernah diketahui sebelumnya, dan pola tersebut sangatlah menarik. Karena kata *mining* sendiri memiliki arti suatu usaha dalam memperoleh barang berharga dalam jumlah sedikit dari material dasar yang besar. Oleh karena itu, *Data mining* sejatinya memiliki akar yang lebih luas dan panjang dari disiplin ilmu yang lainnya, seperti *machine learning*, *artificial intelligent* atau kecerdasan buatan, dan statistik serta *Database*. Metode yang selalu dipakai dalam ketika menggunakan *Data mining* antara lain *classification, clustering, association rules mining, genetic algorithm, neural network* dan seba*gain*ya (Pramudiono, 2003).

2. Langkah-Langkah *Data Mining*

Salah satu ketentuan dari *data mining* yang berlaku pada saat diaplikasikan terhadap Data yang berjumlah besar ialah metode sistematis yang digunakan tidak hanya pada saat menganalisis, namun juga saat mempersiapkan data dan ketika melakukan interpretasi dari hasil yang diperoleh sehingga dapat dijadikan sebagai suatu aksi maupun kesimpulan yang dapat memberikan manfaat. *Data mining* sebagai suatu rangkaian proses dapat dikelompokkan menjadi beberapa tahapan proses, seperti pada Gambar 2.1. Langkah ini sifatnya interaktif,

sedangkan pengguna memiliki keterlibatan secara langsung, sebagaiamana basis *knowledge base* berikut (Han dkk., 2006):



Gambar 2.1 Langkah-Langkah *Data Mining*.

Adapun langkah-langkah *data mining* tersebut adalah (Han dkk., 2006):

a. *Data Cleaning*

Data cleaning adalah aktivitas untuk menghilangkan *noise* dan data tidak relevan atau data yang tidak konsisten. Data ini umumnya diperoleh dari hasil penelitian maupun dari perusahaan-perusahaan yang mempunyai isi yang tidak lengkap seperti adanya nilai kosong, nilai tidak valid, ataupun nilai yang salah ketik.

b. *Data Integration*

Integrasi data adalah penggabungan Data dari banyak *database* yang berbeda ke dalam *database* baru. Tidak jarang data yang diperlukan untuk

database berasal dari banyak file. Integrasi data dilakukan pada atribut seperti nama, jenis produk dan lain-lain.

c. *Feature Selection*

Pemilihan fitur merupakan pemilahan variabel yang sesuai dengan kebutuhan penelitian. Biasanya data yang terdapat pada *database* tidak semuanya digunakan dalam penelitian, sehingga data yang tidak dibutuhkan akan dibuang, hanya data yang dapat dianalisis saja yang akan diambil dari Data*base*. Karena dalam penelitian hanya membutuhkan data-data yang diperlukan dan mendukung dalam sebuah penelitian.

d. *Data Transformation*

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Dalam beberapa metode *data mining* dibutuhkan format data yang khusus sebelum bisa dilakukan pengolahan. Menurut (Junaedi et al., 2011), dalam *data transformation*, terdapat beberapa operasi/teknik untuk melakukan transformasi data, yaitu *smoothing, attribute construction, normalization, aggregation*, dan *discretization*.

1) Smoothing

Operasi Smoothing digunakan untuk mengatasi data bersifat *noise*/nilai yang tidak valid untuk proses mining dengan memperhatikan nilai-nilai tentangga. Salah satu metode yang akan digunakan adalah *Clustering*. *Clustering* adalah proses mempartisi atau mengelompokkan serangkaian pola yang diberikan ke dalam cluster yang terpisah (Alsabti et al., 1997). Metode *Clustering* berguna untuk menghilangkan *outliers/noise* (data yang terlalu

menyimpang jauh dari data lainnya). Algoritma yang dipakai adalah k-Means.

2) *Attribute Construction*

Pada *attribute construction*, mengkontruksi atau menambahkan atribut baru untuk meningkatkan ketelitian/ketepatan proses mining.

3) *Normalization*

Normalization adalah proses pengelompokan atribut ke dalam hubungan yang terstruktur dengan baik dan bebas dari anomali (Lee, 1995). *Normalization* digunakan untuk mentransformasi sebuah atribut numerik diskalakan dalam range yang lebih kecil seperti -1.0 sampai 1.0. Teknik yang digunakan untuk operasi ini adalah *Z-Score Normalization.*

4) *Aggregation*

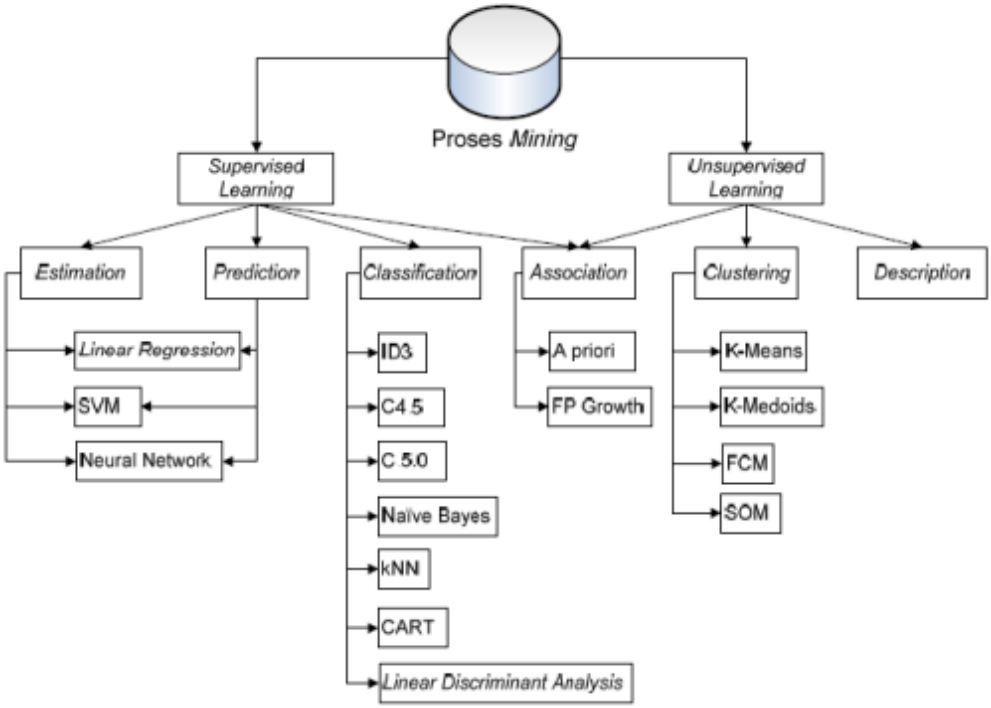
Aggregation merupakan operasi untuk *summary* (peringkasan) yang digunakan untuk data numerik dengan menggunakaan operasi *roll up*.

5) *Discretization*

Discretization adalah digunakan untuk mereduksi sekumpulan nilai yang terdapat pada atribut *continuous*, dengan membagi range dari atribut ke dalam interval. Operasi yang digunakan dalam *discretization* adalah *Binning*.

e. Proses Mining

Merupakan suatu proses utama dan paling penting saat metode diterapkan untuk menemukan pengetahuan baru dan tersembunyi dari data. Beberapa metode tersebut dapat dilihat pada Gambar 2.2.



Gambar 2.2 Metode-Metode *Data Mining*

f. *Pattern evaluation* atau Evaluasi Pola

Evaluasi pola bertujuan untuk merekognisi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

g. *Knowledge presentation* atau presentasi pengetahuan

Presentasi pengetahuan adalah penggambaran dan representasi pengetahuan tentang metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

3. Pengelompokan *Data Mining*

Pengelompokkan *data mining* dibagi menjadi beberapa kelompok, yaitu (Kusrini & Luthfi, 2009):

a. Deskripsi

Deskripsi adalah sarana untuk menggambarkan pola dan tren yang ada dalam data yang dimiliki.

b. Estimasi

Estimasi hampir sama dengan *classifier*, kecuali variabel target pada estimasi lebih bersifat numerik daripada kategoris. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediktor.

c. Prediksi

Prediksi mengasumsikan nilai yang belum diketahui dan juga memperkirakan nilai untuk masa depan.

d. Klasifikasi

Dalam klasifikasi terdapat variabel kategoris sebagai target, misalnya klasifikasi pendapatan dapat dipisahkan menjadi tiga kategori yaitu tinggi, sedang, dan rendah.

e. Pengklasteran

Merupakan pengelompokkan *record*, pengamatan yang membentuk kelas objek-objek yang memiliki kemiripan.

f. Asosiasi

Asosiasi bertugas untuk menemukan atribut yang muncul pada satu waktu tertentu. Dalam dunia bisnis sering disebut dengan *shopping cart analysis*.

4. Karakteristik *Data Mining*

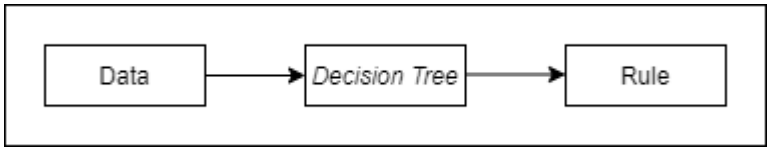
Karakteristik *Data Mining* adalah sebagai berikut (Beynon-Davies, 2004):

a. *Data Mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola tertentu dalam data yang sebelumnya tidak diketahui.
b. *Data Mining* menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat diandalkan.
c. *Data Mining* sangat berguna untuk membuat keputusan yang penting terutama dalam strategi.

2.1.2 *Decision Tree*

Seperangkat aturan untuk membagi populasi yang besar (*heterogen*) dalam model pohon keputusan ini menjadi populasi yang lebih kecil (*homogen*) dengan mempertimbangkan variabel target. Biasanya variabel target digolongkan secara tepat, dan dimaksudkan untuk perhitungan kemungkinan setiap *record* terkait kategorinya dari model pohon keputusan tersebut, atau untuk menggolongkan *record* yang sama masuk ke dalam satu kelas. Pembuatan pohon keputusan dapat menggunakan salah satu di antara beberapa algoritma pohon keputusan, yang bertujuan untuk memodelkan kumpulan data yang belum dikelompokkan berdasarkan kelas (Kusrini & Luthfi, 2009).

Salah satu metode pembelajaran yang paling populer dan banyak digunakan adalah pohon keputusan. Metode ini menjadi upaya untuk mendapatkan fungsi pendekatan melalui nilai diskrit (Suyanto, 2014). Konsep dari pohon keputusan sendiri ialah untuk mengubah Data dalam tabel keputusan menjadi sebuah pohon keputusan dan aturan keputusan yang dikenal dengan *rule-ruke*. Gambaran dari konsep pohon keputusan tersebut adalah:



Gambar 2.3 Konsep Pohon Keputusan

Data dalam pohon keputusan biasanya direpresentasikan dalam bentuk tabel yang berisi atribut dan *record*. Atribut menyatakan parameter yang akan digunakan sebagai dasar dalam pembentukan *Tree*. Misalnya, untuk menentukan timgkat risiko manusia terserang *Stroke*, perlu dipertimbangkan kriteria seperti riwayat sakit jantung, kadar glukosa, pola hidup, jenis pekerjaan, status merokok, jenis kelamin. Salah satu atribut yang digunakan untuk merincikan Data dalam solusi per-Data diistilahkan dengan atribut target. Sedangkan atribut sendiri memiliki nilai yang diistilahkan dengan *instance* (Suyanto, 2014).

Pohon keputusan melalui proses berupa mengalih bentukkan data tabel menjadi model pohon, mengalih bentukkan model pohon menjadi aturan, kemudian menyederhanakan aturan. Langkah pertama dalam membangun pohon keputusan adalah menghitung nilai *entropy* total dari jumlah sampel data, kemudian mengelompokkan variabel untuk nilai *gain* pada setiap atribut. Ketika perhitungan selesai, atribut dengan nilai *gain* tertinggi menjadi akar, atribut lainnya menjadi cabang, cabang tersebut kemudian dihitung ulang untuk melihat atribut lain yang mempunyai nilai *Gain* tertinggi. Langkah perhitungan tersebut dilakukan berulang-ulang secara terus menerus sehingga semua atribut tereksekusi (Suyanto, 2014).

Manfaat utama menggunakan pohon keputusan adalah dapat memecahkan proses pengambilan keputusan yang kompleks menjadi lebih sederhana sehingga. Hal ini memungkinkan proses pengambil keputusan untuk menginterpretasikan

solusi, bukan masalah. Pohon Keputusan juga membantu untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara beberapa variabel input dengan variabel target. Pohon keputusan menggabungkan eksplorasi data dan pemodelan dan merupakan langkah pertama dalam proses pemodelan bahkan ketika digunakan sebagai model akhir untuk teknik lain (Iykra, 2018).

Decission Tree juga dikenal sebagai diagram alur, berbentuk seperti struktur pohon, dengan setiap internal *node* menyatakan pengujian atribut, setiap cabang mewakili output dari hasil pegujian, dan *node* daun (*leaf node*) menyatakan distribusi kelas. *Node* teratas disebut sebagai simpul akar (*root node*). Pohon keputusan digunakan untuk mengklasifikasikan sampel data yang kelasnya belum diketahui ke dalam kelas yang sudah ada. Dalam jalur pengujian data, semua data harus terlebih dahulu melalui *root node* dan terakhir melalui *leaf node*. Ini yang akan menyimpulkan prediksi kelas bagi data tersebut. Atribut data harus berupa data kategorial. Jika kontinu, atribut harus didiskritisasi terlebih dahulu (Han dkk., 2006).

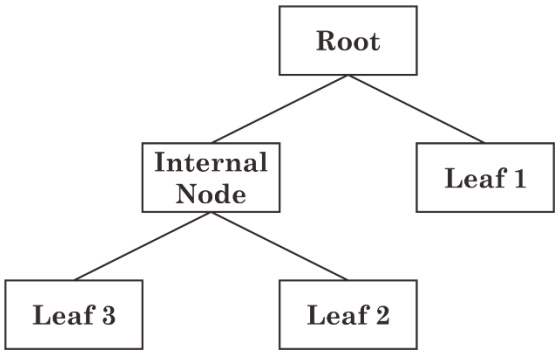
Metode *Decision Tree* memiliki beberapa keunggulan dibandingkan metode lain untuk *Database* besar, yaitu (Han dkk., 2006):

1. Memiliki kecepatan yang relatif lebih cepat.
2. Dapat diubah menjadi *rule* klasifikasi dengan mudah dan sederhana.
3. Dapat menggunakan *query* SQL untuk mengakses *Database.*
4. Akurasi tinggi dibandingkan dengan metode lain.

2.1.3 Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembang dari

algoritma ID3 (Larose, 2005). Tree atau pohon keputusan, dikenal sebagai bagian dari *graph*, yang termasuk dalam irisan bidang ilmu automata dan teori linguistik serta matematika diskrit. *Tree* sendiri merupakan *graph* tak berarah yang terhubung dan tidak mengandung sirkuit (Munir, 2010).



Gambar 2.4 Pohon Keputusan

Secara umum langkah-langkah algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Kusrini & Luthfi, 2009):

1. Persiapkan *dataset training* dan *testing*.
2. Menghitung nilai *gain* yang paling tinggi dari setiap atribut atau berdasarkan pada perhitungan *index entropy* dengan menggunakan rumus sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{n} -pi * \log_2 \ pi \qquad (2.1)$$

Keterangan:

S = himpunan kasus

n = Banyak nilai yang ada pada atribut target (jumlah kelas klasifikasi)

pi = proporsi dari Si terhadapat *S*

3. Menghitung nilai *gain*. Adapun rumus perhitungannya sebagai berikut:

$$Gain(S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \qquad (2.2)$$

Keterangan:

S = himpunan kasus

A = Atribut

n = Banyak nilai yang ada pada atribut target (jumlah kelas klasifikasi)

$|S_i|$ = jumlah kasus pada partisi ke-i

$|S|$ = jumlah kasus dalam S

4. Mengulang langkah no 2 sampai semua *record* terpatisi. Partisi *Decision Tree* akan terhenti jika:

a. Semua *tuple* dalam *record* simpul *m* memperoleh kelas sama.

b. Atribut dalam *record* tidak ada yang dipartisi lagi.

c. pada cabang yang kosong, tidak ada *record*.

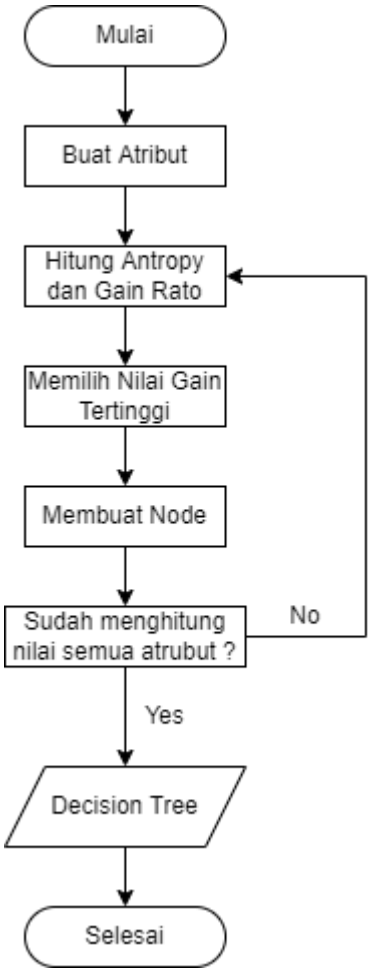
Tree Pruning dilakukan untuk menyederhanakan *tree* sehingga akurasi dapat bertambah. *Pruning* ada dua pendekatan, yaitu:

a. *Pre-pruning*, yaitu menghentikan pembangunan suatu *subtree* lebih awal (yaitu dengan memutuskan untuk tidak lebih jauh mempartisi data training). Saat seketika berhenti, maka *node* berubah menjadi *leaf* (*node* akhir). *Node* akhir ini menjadi kelas yang paling sering muncul di antara subset sampel.

b. *Post-pruning*, yaitu menyederhanakan *tree* dengan cara membuang beberapa cabang subtree setelah *tree* selesai dibangun. *Node* yang jarang

dipotong akan menjadi *leaf* (*node* akhir) dengan kelas yang paling sering muncul.

Algoritma *Decision Tree* C4.5 mengabaikan *missing value*, yaitu nilai elemen yang dapat diprediksi berdasarkan apa yang diketahui tentang nilai atribut di baris lain. Ide dasar dari algoritma ini adalah untuk membagi Data ke dalam rentang berdasarkan nilai atribut dari item yang ditemukan dalam *data training*. Algoritma C4.5 memungkinkan untuk prediksi baik melalui pohon keputusan atau aturan yang dihasilkan dari pembentukan *classifier* (Diwandari & Setiawan, 2015).



Gambar 2.5 *Flowchart* Algoritma C4.5

Berdasarkan *flowchart* pada Gambar 2.5 di atas, dapat diketahui alur kerja algoritma C4.5 yang digunakan dalam penelitian. Persiapan pertama adalah menentukan atribut mana yang akan digunakan dan melakukan perhitungan untuk mencari nilai *gain* tertinggi berdasarkan perhitungan *entropy* untuk semua atribut. Jika nilai *gain* tertinggi ditemukan, maka atribut tersebut akan menjadi *root* awal. Selanjutnya, cabang dibentuk dengan cara yang sama dengan mempertimbangkan nilai *gain* tertinggi dari setiap hasil partisi.

2.1.4 *Data Imbalance*

Data Imbalance / data tidak seimbang merupakan kondisi dimana suatu kelompok kelas memiliki jumlah data yang jauh berbeda dibandingkan dengan kelas lainnya. Kelas yang memiliki jumlah data lebih banyak sering kita sebut dengan majority class dan kelas yang mempunyai jumlah data lebih sedikit disebut dengan minority class[15]. Karakteristik dari data imbalance tentu dapat mempengaruhi terhadap hasil prediksi yang dilakukan oleh algoritma. Untuk mengetahui seberapa besar tingkat ketidak seimbangan data yang ada dapat dihitung menggunakan IR (Imbalanced Ratio) dengan perbandingan sebagai berikut.

$$Imbalanced\ Ratio\ (IR) = \frac{Majority\ Class}{Minority\ Class} \qquad (2,3)$$

Perbandingan di atas menunjukkan besarnya tingkat ketidakseimbangan data berdasarkan perbandingan kelas major dan kelas minor. Metode yang dapat dilakukan untuk mengatasi permasalahan terkait data tidak seimbang (*imbalanced data*) dapat diselesaikan dengan penambahan data sintetik pada kelas minoritas dengan metode *oversampling* dan *undersampling* (Mahmood, 2015).

1. *Oversampling*

   Oversampling merupakan metode *sampling* dengan menambahkan jumlah data pada kelas minoritas sehingga dapat mengimbangi atau mendekati jumlah data pada kelas mayoritas. Konsep penambahan data pada *oversampling* dibagi menjadi dua yaitu: *oversampling* menggunakan data asli, seperti metode *Random Oversampling* dan yang kedua yaitu metode penambahan menggunakan data sintetik seperti *Synthetic Minority Over-sampling Technique* (SMOTE), *SMOTE Borderline*, *Safe Level SMOTE*, *Adaptive Semi-unsupervised Weighted Oversampling* (A-SUWO) (Drummond & Holte, 2003).

2. *Undersampling*

   Undersampling bekerja pada kelas mayoritas dan memiliki kelebihan pada *dataset* berukuran besar. *Undersampling* mengurangi jumlah pengamatan dari kelas mayoritas untuk membuat kumpulan data menjadi seimbang. Namun hal tersebut juga mengakibatkan kelemahan pada kurangnya informasi penting di kelas mayoritas akibat data yang terhapus (Drummond & Holte, 2003). Terdapat dua cara dalam melakukan *undersampling* yaitu *random undersampling* dan *informative undersampling*. Dalam random *undersampling*, data dari kelas mayoritas yang akan dibuang dipilih secara acak sampai *dataset* disebut balance, sedangkan dalam informative *undersampling* data yang akan dibuang dipilih berdasarkan aturan tertentu. Contoh teknik *informative undersampling* diantaranya adalah *Tomek's link, Edited Nearest Neighbors* (ENN), dan *Neighbourhood Cleaning Rule* (NCL).

2.1.5 *Confusion Matrix*

Confusion matrix adalah sebuah tabel yang menunjukkan jumlah *data testing* yang diklasifikasikan dengan benar dan jumlah Data yang diklasifikasikan salah (Indriani, 2014). Evaluasi dengan *confusion matrix* memberikan nilai *accuracy, precison*, dan *recall*. *Accuracy* adalah persentase keakuratan *record* data yang diklasifikasikan dengan benar setelah dilakukan pengujian pada hasil klasifikasi. Sedangkan *precision* adalah persentase kasus yang prediksi positif yang benar-benar positif pada data yang sebenarnya. *Recall* atau *sensitivity* adalah persentase kasus positif yang benar-benar positif yang diprediksikan secara benar (Andriani, 2013).

Tabel 2.1 Contoh Confusion Matrix 2 Kelas

		Actual Value	
		Yes	No
Predicted Value	Yes	TP	FP
	No	FN	TN

Sumber: (Indriani, 2014)

True Positive (TP) adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *False Positive* (FP) adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *False Negative* (FN) adalah jumlah *record* positif yang diklasifikasikan sebagai negatif, *True Negative* (TN) adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif, hasil klasifikasi yang berbentuk confusion matrix digunakan untuk menghitung *accuracy, precision, recall* serta *Receiver Operation Characteristic* (ROC). *Accuracy* merupakan presentase dari total pasien yang diidentifikasi benar, *Precision* merupakan kecocokan antara bagian data yang digunakan dengan informasi yang dibutuhkan, sedangkan *recall*

merupakan tingkat sistem dalam menemukan sebuah informasi. Untuk menghitungnya digunakan persamaan dibawah ini (Indriani, 2014):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2,4)$$

$$Precision = \frac{TP}{TP + FP} \quad (2,5)$$

$$Recall\ (r) = \frac{TP}{TP + FN} \quad (2,6)$$

Keterangan:

TP = Jumlah *true* positif

FP = Jumlah *false* positif

FN = Jumlah *false* negatif

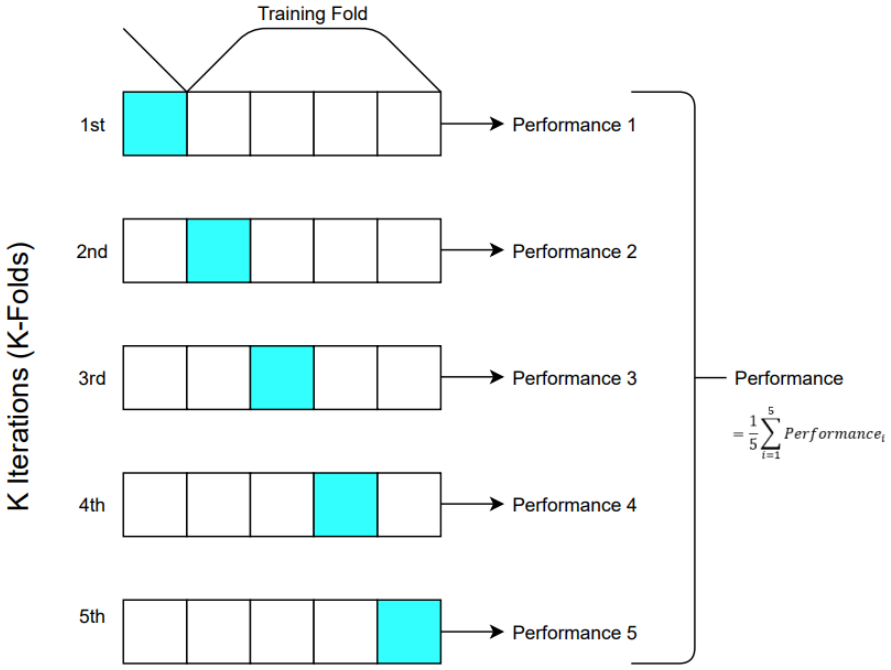
TN = Jumlah *true* negatif

2.1.6 *Cross Validation*

Cross validation adalah metode tambahan dari teknik validasi data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan *k-fold cross validation* dimana percobaan sebanyak *k* kali untuk satu model dengan parameter yang sama (Santosa & Umam, 2018). *Cross-Validation* digunakan untuk mengevaluasi bagaimana hasil analisis statistik digeneralisasikan ke kumpulan data independen. Teknik ini digunakan untuk memprediksi dan memperkirakan seberapa akurat model prediksi ketika diimplementasikan. Dalam masalah prediksi, model diberikan sekumpulan set data (*Dataset*) yang akan digunakan untuk melakukan pelatihan (*training Dataset*), serta satu set data yang tidak diketahui (atau data yang pertama kali ditampilkan) dari model yang diuji (*testing Dataset*) (Santosa & Umam, 2018).

Tujuan dari *Cross-validation* adalah untuk mendefinisikan kumpulan data untuk "menguji" model selama fase pelatihan (yaitu, Data validasi), untuk

membatasi masalah seperti *overfitting* dan memberikan wawasan tentang bagaimana model akan menggeneralisasi independen dari *Dataset* (yaitu, *dataset* tidak diketahui dalam contoh masalah sebelumnya). Berikut contoh ilustrasi proses *cross validation* dengan k=5 pada Gambar 2.6.



Gambar 2.6 Ilustrasi *Cross Validatiton*

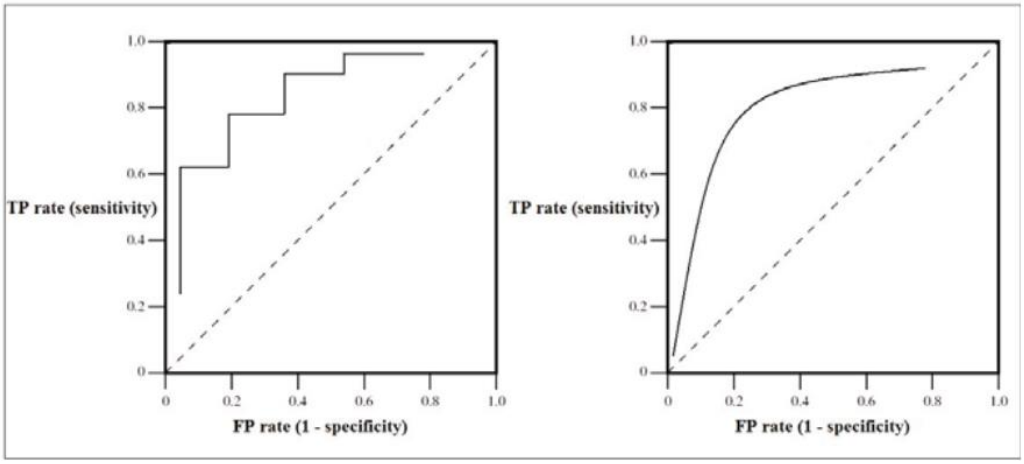
Pada percobaan di atas menggunakan *5-fold cross validation* yang artinya bahwa dilakukan percobaan sebanyak 5 kali, dimana pada percobaan pertama menjadikan partisi pertama menjadi data testing dan partisi lainnya menjadi data training. Begitu dengan percobaan ke-2 sampai ke-5 dengan menjadikan 1 partisi sebagai data testing dan data lainnya sebagai training.

Dalam beberapa penelitian yang sudah dilakukan oleh ahli-ahli data mining, model pengujian atau validasi akan sering dipakai *cross validation* ketimbang *split validation. Cross validation* dengan menerapkan *10-fold cross validation* sudah merupakan standart dari suatu metode validasi yang canggih atau lebih praktis dan mampu meningkatkan nilai akurasi (Santosa & Umam, 2018).

2.1.7 Kurva ROC

Grafik kurva ROC (*Receiver Operating Characteristic*) digunakan untuk mengevaluasi akurasi *classifier* dan untuk membandingkan klasifikasi yang berbeda model. Sebuah grafik ROC adalah grafik dua dimensi dengan proporsi negatif pada sumbu horisontal dan proporsi positif yang benar di sumbu vertikal (Vercellis, 2009). Kegunaan kurva ROC (*Receiver Operating Characteristic*) adalah untuk radar selama Perang Dunia II untuk mendeteksi benda-benda musuh di medan perang, teori deteksi sinyal, dalam psikologi ke rekening untuk deteksi sinyal persepsi, penelitian medis dan dalam mesin pembelajaran dan penelitian data mining, serta masalah klasifikasi (Gorunescu, 2011).

Dalam masalah klasifikasi menggunakan kelas keputusan dua (klasifikasi biner), masing-masing objek dikelompokkan dalam (P.N), yaitu positif atau negatif. Sementara model klasifikasi beberapa (misalnya pohon keputusan) menghasilkan label kelas diskrit (menunjukkan hanya kelas diprediksi objek), pengklasifikasi lainnya (misalnya, naive bayes, jaringan saraf) menghasilkan output yang berkesinambungan, yang ambang batas yang berbeda mungkin diterapkan untuk memprediksi keanggotaan kelas, secara teknis, ROC kurva, juga dikenal sebagai grafik ROC, dua-dimensi grafik dimana tingkat TP diplot pada sumbu Y- dan tingkat FP diplot pada X-sumbu (Gorunescu, 2011).



Gambar 2.7 Grafik ROC (discrete/continuous case)

Pada Gambar 2.7 garis diagonal membagi ruang ROC, yaitu:

1. Pada gambar pertama terdapat point di atas garis diagonal merupakan hasil klasifikasi yan baik.
2. Pada gambar kedua terdapat point dibawah garis diagonal merupakan hasil klasifikasi yang buruk

Digunakan metode yang menghitung luas daerah dibawah kurva ROC yang disebut AUC (*Area Under the ROC Curve*) yang diartikan sebagai probabilitas. AUC merupakan perhitungan untuk mengukur perbedaan performansi metode yang digunakan untuk mengekspresikan data *confusion matrix*. Garis horizontal mewakili nilai *false positives* (FP) dan garis vertikal mewakili nilai *true positives* (TP), AUC terdiri dari AUC *Optimistic*, dan AUC *Pessimistic*. AUC mengukur kinerja diskriminatif dengan memperkirakan probabilitas output dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar AUC, semakin kuat klasifikasi yang digunakan. Karena AUC adalah bagian dari daerah unit persegi, nilainya akan selalu antara 0,0 dan 1,0. Untuk keakurasian nilai AUC

dalam klasifikasi data mining dibagi menjadi lima kelompok (Gorunescu, 2011), yaitu:

1. 0.90 – 1.00 = klasifikasi sangat baik (*excellent classification*)
2. 0.80 – 0.90 = klasifikasi baik (*good classification*)
3. 0.70 – 0.80 = klasifikasi cukup (*fair classification*)
4. 0.60 – 0.70 = klasifikasi buruk (*poor classification*)
5. 0.50 – 0.60 = klasifikasi salah (*failure*)

2.1.8 *Stroke*

1. Defenisi *Stroke*

Definisi yang paling banyak diterima adalah bahwa *Stroke* adalah sindrom tertentu dengan gejala cepat dan adanya tanda-tanda klinis yang berupa gangguan fungsional otak fokal maupun global yang berlangsung lebih dari 24 jam (kecuali ada prosedur bedah atau kematian), yang tidak disebabkan oleh penyebab lain selain pembuluh darah atau vaskuler (Mansjoer, 2000). Menurut Geyer (2009) *Stroke* adalah sindrom klinis yang ditandai dengan perkembangan defisit neurologis persisten fokus sekunder terhadap peristiwa pembuluh darah. *Stroke* adalah penyebab cacat nomor satu di dunia dan nomor dua penyebab kematian di dunia. Duapertiga *Stroke* terjadi di negara-negara berkembang. Di masyarakat Barat, 80% pasien penderita *Stroke* mengalami *Stroke iskemik* dan 20% mengalami *Stroke hemoragik*. Insiden *Stroke* meningkat seiring bertambahnya usia (Dewanto dkk., 2009).

2. Faktor Risiko terjadinya *Stroke*

Faktor-faktor penyebab terjadinya *Stroke* yang tidak dapat diubah, meliputi: usia, jenis kelamin, genetika, ras/etnis. Sedangkan yang dapat diubah, meliputi:

riwayat *Stroke*, hipertensi, penyakit jantung, diabetes, *Transient Ischemic Attack* (TIA), *hiperkolesterol*, obesitas, perokok, alkohol, *hyperureucesemia*, dan *hematokrit*. Banyak faktor yang dapat meningkatkan risiko *Stroke*. Faktor risiko *Stroke* yang berpotensi dapat diobati meliputi (Mayo Clinic Staff, 2022):

a. Faktor risiko gaya hidup
   1) Kelebihan berat badan atau obesitas
   2) Ketidakaktifan fisik
   3) Minum berat atau pesta minuman keras
   4) Penggunaan obat-obatan terlarang seperti kokain dan metamfetamin

b. Faktor risiko medis
   1) Tekanan darah tinggi
   2) Merokok atau paparan asap rokok
   3) Kolesterol Tinggi
   4) Diabetes
   5) Apnea tidur obstruktif
   6) Penyakit kardiovaskular, termasuk gagal jantung, cacat jantung, infeksi jantung atau irama jantung yang tidak teratur, seperti fibrilasi atrium
   7) Riwayat pribadi atau keluarga *Stroke*, serangan jantung atau serangan iskemik transien
   8) Infeksi COVID-19

c. Faktor lain yang terkait dengan risiko *Stroke* yang lebih tinggi meliputi:
   1) Usia - Orang berusia 55 tahun atau lebih memiliki risiko *Stroke* yang lebih tinggi daripada orang yang lebih muda.

2) Ras atau etnis - Afrika Amerika dan Hispanik memiliki risiko lebih tinggi terkena *Stroke* daripada orang-orang dari ras atau etnis lain.
3) Jenis Kelamin Pria memiliki risiko lebih tinggi terkena *Stroke* daripada wanita. Wanita biasanya lebih tua ketika mereka mengalami *Stroke*, dan mereka lebih mungkin meninggal karena *Stroke* daripada pria.
4) Hormon - Penggunaan pil KB atau terapi hormon yang mencakup estrogen meningkatkan risiko.

2.2 Kajian Al-Qur'an

Pada pembahasan sebelumnya telah dipaparkan bahwa Prediksi adalah cara untuk memperkirakan suatu kejadian yang akan terjadi dimasa depan. Dalam bidang ilmu sains prediksi diistilahkan dengan peramalan, yang terbagi menjadi dua macam yaitu peramalan secara ilmiah dan secara non-ilmiah. Islam tidak melarang peramalan dalam ilmu sains secara keseluruhan, seperti permalan secara ilmiah. Allah sendiri memberikan perintah untuk mengikuti pengetahuan bukan hawa nafsu sebagaimana yang difirmankan dalam QS Ar-Rum ayat 29 (Kemenag, 2002):

بَلِ اتَّبَعَ الَّذِيْنَ ظَلَمُوْٓا اَهْوَآءَهُمْ بِغَيْرِ عِلْمٍ فَمَنْ يَّهْدِيْ مَنْ اَضَلَّ اللّٰهُ ۗ وَمَا لَهُمْ مِّنْ نّٰصِرِيْنَ (٢٩)

Artinya:
"Tetapi orang-orang yang zalim, mengikuti keinginannya tanpa ilmu pengetahuan; maka siapakah yang dapat memberi petunjuk kepada orang yang telah disesatkan Allah. Dan tidak ada seorang penolong pun bagi mereka" (QS. Ar-Rum 29).

Ramalan yang berasal dari perkiraan yang berbasiskan ilmu pengetahuan dan keilmuan masih diperbolehkan dan tidak diharamkan selagi memilik manfaat dan kemasalahatan bagi ummat. Berikut ayat al-qur'an yang terkait peramalan atau prediksi yang tertulis pada Q.S Yusuf ayat 47-49 (Kemenag, 2002):

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَأَبًا فَمَا حَصَدتُّمْ فَذَرُوهُ فِي سُنۢبُلِهِۦٓ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ (٤٧) ثُمَّ يَأْتِي مِنۢ بَعْدِ ذَٰلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُحْصِنُونَ (٤٨) ثُمَّ يَأْتِي مِنۢ بَعْدِ ذَٰلِكَ عَامٌ فِيهِ يُغَاثُ النَّاسُ وَفِيهِ يَعْصِرُونَ (٤٩)

Artinya:
"*Dia (Yusuf) berkata, "Agar kamu bercocok tanam tujuh tahun (berturut-turut) sebagaimana biasa; kemudian apa yang kamu tuai hendaklah kamu biarkan di tangkainya kecuali sedikit untuk kamu makan. 47. Kemudian setelah itu akan Datang tujuh (tahun) yang sangat sulit, yang menghabiskan apa yang kamu simpan untuk menghadapinya (tahun sulit), kecuali sedikit dari apa (bibit gandum) yang kamu simpan. 48. Setelah itu akan Datang tahun, di mana manusia diberi hujan (dengan cukup) dan pada masa itu mereka memeras (anggur) ". 49. (QS. Yusuf ayat 47-49).*

Kemudian dijelaskan menurut tafsir ibnu kaitsar bahwa kajian di atas menjelaskan kelak akan Datang musim subur dan banyak hujan kepada kalian selama tujuh tahun berturut-turut. Sapi diibaratkan dengan tahun karena sapilah yang dipakai untuk membajak tanah dan lahan yang akan digarap untuk menghasilkan buah-buahan dan tanam-tanaman, yaiitu bulir-bulir gandum yang hijau dan (subur). Kemudian Yusuf a.s memberikan pengarahan kepada mereka mengenai apa yang harus mereka kerjakan selama tujuh tahun subur itu. Ia berkata: "Maka apa yang kalian panen hendaklah kalian bairkan dibulirnya, kecuali sedikit untuk makan kalian".

Yakni betapapun hasilnya yang kalian peroleh dari panen dimusim subur selama tujuh tahun ini, kalian harus membiarkan hasilnya pada bulir-bulirnya, agar dapat disimpan dalam jangka panjang untuk waktu yang lama untuk menghindari kebusukan. Terkecuali sekedar untuk apa yang kalian makan, maka boleh dipisahkan dari bulirnya. Dan makanlah dalam kadar yang minim, janag berlebihan agar jumlah makanan yang ada cukup menutupi kebutuhan makan kalian selama musim panceklik yang akan Datang.

Penggalan berita lain yang disampaikan Alqur'an tentang gambaran peristiwa masa depan ditemukan dalam ayat pertama Surat Ar-Rum, yang merujuk pada kekaisaran Bizantium, wilayah timur Kekaisaran Romawi. Dalam ayat-ayat ini, disebutkan bahwa Kekaisaran Bizantium telah mengalami kekalahan besar, tetapi akan segera memperoleh kemenangan (Kemenag, 2002).

الٓمّٓ (١) غُلِبَتِ الرُّوْمُۙ (٢) فِيْٓ اَدْنَى الْاَرْضِ وَهُمْ مِّنْۢ بَعْدِ غَلَبِهِمْ سَيَغْلِبُوْنَۙ (٣) فِيْ بِضْعِ سِنِيْنَ ۗ لِلّٰهِ الْاَمْرُ مِنْ قَبْلُ وَمِنْۢ بَعْدُ ۗ وَيَوْمَىِٕذٍ يَّفْرَحُ الْمُؤْمِنُوْنَ (٤)

Artinya:
"Alif, Lam, Mim. Telah dikalahkan bangsa Romawi, di negeri yang terdekat dan mereka sesudah dikalahkan itu akan menang, dalam beberapa tahun (lagi). Bagi Allah-lah urusan sebelum dan sesudah (mereka menang). Dan di hari (kemenangan bangsa Romawi) itu bergembiralah orang-orang yang beriman". (Al Ar-Rum:1-4).

Ayat-ayat ini diturunkan kira-kira pada tahun 620 Masehi, hampir tujuh tahun setelah kekalahan hebat Bizantium Kristen di tangan bangsa Persia, ketika Bizantium kehilangan Yerusalem. Kemudian diriwayatkan dalam ayat ini bahwa Bizantium dalam waktu dekat menang. Padahal, Bizantium waktu itu telah menderita kekalahan sedemikian hebat hingga nampaknya mustahil baginya untuk mempertahankan keberadaannya sekalipun, apalagi merebut kemenangan kembali.

Jadi dapat dipahami didalam quran surah di atas, bahwa allah menggambarkan contoh usaha manusia untuk mempersiapkan menghadapi kemungkinan yang buruk dimasa depan, persiapan tersebut secara langsung menggambarkan proses didalam kehidupan manusia yang tidak tahu bagaimana kehidupan kita selanjutnya (Ar-Rifa'i, 2012).

2.3 Kajian Topik Dengan Teori Pendukung

Prediksi yang akan dilakukan pada penelitian ini menggunakan metode *Decision Tree* yaitu dengan memproses data sehingga membentuk struktur pohon. *Decision Tree* merupakan metode yang paling bagus tingkat akurasinya dibandingkan dengan metode lain pada *data mining* dalam melakukan proses prediksi. Namun hasil yang dihasilkan dengan menggunakan algoritma *Decision Tree* sangat bergantung dalam jumlah dan jenis data, dengan jumlah data yang makin besar dan *type* data tertentu maka tingkat akurasinya akan semakin bagus (Kusrini & Luthfi, 2009).

Pada proses penelitian yang dilakukan menggunakan algoritma C4.5 yaitu algoritma yang digunakan dalam pengolahan data dengan jenis atribut numerik dan kategorial. Data yang digunakan merupakan data pasien *Stroke* dengan *type* data campuran sehingga akan dilakukan tahap *preprocessing* diantaranya proses pembersihan data, seleksi data dan transformasi data. Setelah tahap ini dilakukan langkah selanjutnya akan dilakukan proses *classifier* dengan menggunakan algoritma *Decision Tree* C4.5. Selanjutnya dari hasil yang didapat akan dihasilkan model struktur pohon dan tingkat akurasi. Pada tahap akhir akan dilakukan proses evaluasi dan validasi menggunakkan teknik *k-fold cross validation*. Tahap akhir ini sangat penting karena akan menguji performa dan akurasi pada metode yang digunakan dengan menggunakan *confusion matrix* dengan parameter *accuraccy, precision, recall,* dan AUC (*optimistic*) dan AUC (*pessimistic*) (Kusrini & Luthfi, 2009).

BAB III
METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang dilakukan adalah penelitian kuantitatif. Penelitian ilmiah yang dilakukan secara sistematis dengan menggunakan model-model matematis dan teori-teori yang berkaitan. Penelitian yang dilakukan menggunakan salah satu teknik pada *data mining* yaitu algoritma *Decision Tree* C4.5.

3.2 Data dan Sumber Data

Dataset yang digunakan pada penelitian ini adalah data pasien *Stroke* berupa data sekunder yang bersumber dari *World Health Organization* (WHO) (Fedesoriano, 2021). Data diambil dalam bentuk tabel dengan format file *.CSV. Pengambilan Data dilakukan pada tanggal 16 Maret 2022. Jumlah keseluruhan data 5110 *record* yang terdiri dari 4861 pasien *Stroke* dan 249 pasien tidak *Stroke*. Namun, data yang digunakan setelah melalui proses *preprocessing* sebanyak 360 yang terdiri dari 180 data pasien *Stroke* dan 180 data pasien tidak *Stroke*.

3.3 Teknik Pengumpulan Data

Teknik Pengumpulan Data dalam penelitian ini yaitu:

1. Pengumpulan *Dataset* dilakukan secara online melalui website KAGGLE (Fedesoriano, 2021).
2. Data berasal dari Organisasi Kesehatan Dunia (WHO).

3.4 Instrumen Penelitian

Pada penelitian ini dibutuhkan instrumen pendukung berupa *hardware* dan *software* diantaranya yaitu

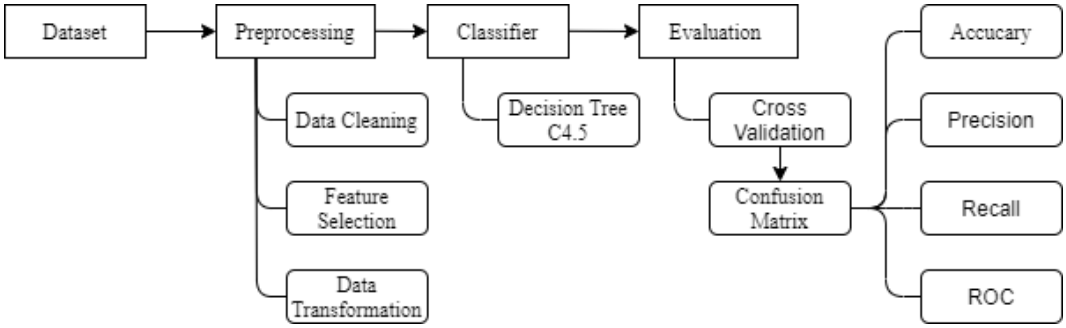
1. Laptop dengan spesifikasi RAM 4, Processor Intel PENTIUM GOLD.
2. Web browser seperti Chrome, Microsoft Edge, dan lai-lain.
3. Software Rapid Miner untuk melakukan proses *Decision Tree*.
4. Microsoft Excel untuk melakukan proses *preprocessing*.

3.5 Variabel Penelitian

Pada Penelitian ini menggunakan dua variabel, diantaranya variabel bebas (*independent variable*) dengan delapan atribut yaitu *Gender, Age, Hypertension, Heart Disease, Resident Type, Avg Glucose Level. BMI* dan *Smoking Status*. Dan variabel terikat (*dependent variable*) yaitu sebagai label/ kelas target output pada penelitian ini yaitu atribut *Stroke*.

3.6 Teknik Analisis Data

Pada penelitian ini memiliki urutan langkah-langkah dalam memecahkan masalah. Mulai dari pemilihan *dataset* yang akan digunakan kemudian melakukan tahap preprocessing *dataset*. Untuk melakukan *classifer*, dilakukan tahap *cleaning* dan *tranformasi data* serta melakukan tahap uji data dengan melakukan pembagian menjadi *data training* dan *data testing*. Setelah melakukan pembagian data, *classifier* dapat diimplementasikan. langkah terakhir adalah melihat hasil prediksinya. Langkah-langkah secara lengkap dapat dijelaskan seperti pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

Berikut penjelasan lebih detail dari Gambar 3.1. Masing-masing langkah akan dijelaskan dalam sub-bab berikut.

3.6.1 *Preprocessing*

Tahap preprocessing adalah proses yang dilakukan sebelum dilakukan proses pengolahan. Tahap ini terdiri dari beberapa langkah yaitu melakukan pembersihan data dan transformasi data. Pembersihan dilakukan untuk meningkatkan kualitas hasil, efisiensi dan kemudahan dalam teknik dan metode *Data mining* (Han dkk., 2006).

1. *Data Cleaning* (Pembersihan Data)

   Pada tahap ini dilakukan proses penghapusan data yang memiliki nilai tidak lengkap, kosong, missing value ataupun menghilangkan inkonsistensi yang muncul dalam *dataset* serta melakukan *balance* pada data yang *unbalance*. Proses *cleaning* dilakukan untuk meningkatkan efisiensi dalam melakukan pengolahan data dan mengurangi tingkat eror yang lebih rendah.

2. *Feature Selection* (Pemilihan fitur)

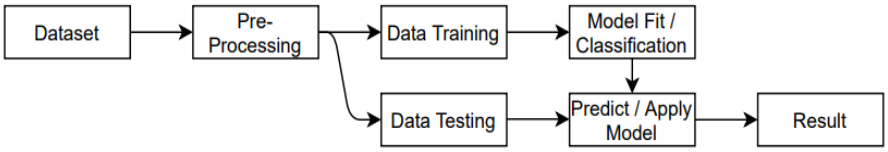
   Merupakan proses pemilihan variabel untuk meminimalkan jumlah data yang digunakan dalam proses mining dengan tetap merepresentasikan data aslinya.

3. *Data Transformation* (Transformasi Data)

   Yaitu proses transformasi data yang sudah dipilih sehingga sesuai dalam bentuk mining procedure. Pada proses ini dilakukan perubahan pada nilai atribut dan *discretization*. Proses *discretization* dilakukan karena data kontinu sulit dalam menafsirkan hasil karena memiliki kelas terlalu banyak.

3.6.2 *Classifier*

Classifier atau proses mining merupakan proses paling utama saat metode diterapkan dalam melakukan proses prediksi. Pada proses ini dilakukan pemilihan metode yang akan digunakan untuk menemukan pengetahuan atau pola-pola baru dan tersembunyi dari data misalnya karakterisasi, klasifikasi, regresi, klustering atapun asosiasi. Pada proses ini juga dilakukan pemilihan teknik, metode ataupun algoritma yang tepat karena sangat bergantung pada tujuan dan proses secara keseluruhan. Pada tahap *classifier* ini peneliti menggunakan salah satu teknik dalam *data mining* klasifikasi yaitu algoritma *Decision Tree* C4.5. Pada proses ini dilakukan pengolahan data dengan bantuan *software RapidMiner*. Berikut pada Gambar 3.2 cara kerja *software RapidMiner* dalam melakukan proses prediksi.



Gambar 3.2 Proses Kerja RapidMiner

Pada tahap ini juga dilakukan pembentukan model pohon keputusan dengan menghitung nilai *entropy total*, *entropy* setiap nilai pada atribut dan *gain* dari setiap atribut. Nilai *gain* tertinggi akan menjadi akar pertama serta selanjutnya sebagai cabang. Proses dilakukan untuk semua atribut sehingga *Decision Tree* terbentuk.

3.6.3 *Evaluation*

Proses evaluasi adalah proses penerjemahan pola-pola yang dihasilkan dari data mining untuk menguji atau mengevaluasi keakuratan dan *performance* dari metode yang digunakan. Pada proses ini menggunakan *cross validation* dengan 10 uji. *k-fold cross validation* merupakan metode uji untuk mengevaluasi kinerja dari algoritma *Decision Tree* C4.5 dimana data akan dibagi sebanyak k menjadi *data training* dan *data testing*. Nilai k diambil *10-fold* sehingga dari 360 data akan menjadi 10 bagian data dengan ukuran yang sama yaitu sekitar 36 data untuk setiap bagiannya. Dari masing-masing 10 bagian data tersebut, 36 data menjadi *data testing* dan 324 data menjadi *data training*. Proses ini menghasilkan hasil berupa *Decision Tree* dan *performance* dari metode yang digunakan yang menghasikan *accucarcy*, *precision*, *recall*, dan *ROC.*

BAB IV
HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk mengetahui karakteristik dari data yang akan diteliti. Pada bagian ini akan dibahas mengenai deskripsi untuk setiap variabel yang digunakan dalam penelitian.

4.1.1. Analisis Data Mentah

Proses analisis dengan mendeskripsikan variabel *dependent Stroke* dan *variabel independent* yaitu *id, ever maried, work type, age, hypertension, heart disease, smoking status, resident type, avg glucose level, BMI* dan *Stroke*. Berikut tabel deksripsi dari variabel data mentah sebagai berikut:

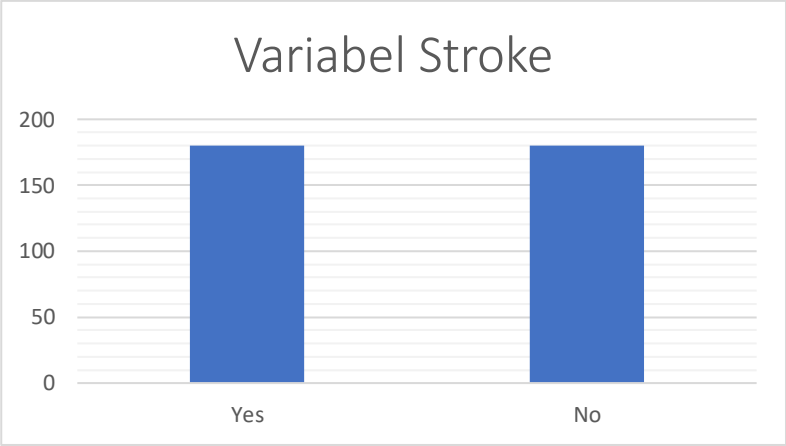
Tabel 4.1 Deskripsi Data Mentah

No	Varaibel	Nilai	Keterangan	Jumlah Data
1	Id	Nominal	Kode user identification pasien *Stroke*	5110
2	Gender	Male	Jenis kelamin pasien	2115
		Female		2994
		Other		1
3	Age	Nominal	Nilai kontinu dengan rentang umur pasien dari balita sampai manula	5110
4	Hypertension	1	Hypertensi	498
		0	Tidak hypertensi	4612
5	Heart Disease	1	Sakit jantung	276
		0	Tidak sakit jantung	4834
6	Ever Maried	Yes	Menikah	3353
		N0	Tidak menikah	1757

No	Varaibel	Nilai	Keterangan	Jumlah Data
7	Work Type	Childrem	Anak-anak	687
		Govt Job	Pegawai pemerintahan	657
		Never Worked	Belum Pernah Bekerja	22
		Private	Pekerjaan pribadi	2925
		Self Employed	Wiraswasta	819
8	Resident Type	Urban	Perkotaan	2596
		Rural	Pedesaan	2514
9	Avg Glucose Level	Nominal	Nilai kontinu kadar glukosa pasien	5110
10	BMI	Nominal, N/A	Nilai kontinu indeks massa tubuh pasien	5110
11	Smoking Status	Formerly Smoked	Pernah merokok	885
		Never Smoked	Belum pernah merokok	1892
		Smokes	Perokok	789
		Unknown	Status merokok tidak ketahui	1544
12	*Stroke*	1	*Stroke*	249
		0	Tidak *Stroke*	4861

4.1.2. Analisis Data Siap

Proses analisis menggunakan tabulasi silang antara variabel *dependent Stroke* dengan *variabel independent gender*, *age*, *hypertension, heart disease, smoking status, resident type, avg glucose level, BMI* dan *Stroke*. Berikut diperoleh diagram dari variabel *Stroke* sebagai berikut:



Gambar 4.1 Diagram Variabel *Stroke* Data Siap

Berdasarkan Gambar 4.1 dapat diketahui bahwa dari total 360 data yang digunakan terdapat 180 data pasien *Stroke* dan 180 pasien yang tidak *Stroke*. Berikut hasil analisis deskriptif menggunakan *cross tabulation*:

Tabel 4.2 Tabulasi Jenis Kelamin terhadap *Stroke*

Stroke	*Gender*		Jumlah
	Male	*Female*	
Yes	75	105	180
	21%	29%	50%
No	70	110	180
	19%	31%	50%
Jumlah	145	215	360
	40%	60%	100%

Berdasarkan Tabel 4.2 pasien *Stroke* terdiri dari jenis kelamin laki-laki sebanyak 21% dan perempuan sebanyak 29% serta pasien yang tidak *Stroke* terdiri dari jenis kelamin laki-laki sebanyak 19% dan perempuan sebanyak 31%.

Tabel 4.3 Tabulasi Umur terhadap *Stroke*

Stroke	***Age***								**Jumlah**
	6-11 th	**12-16 th**	**17-25 th**	**26-35 th**	**36-45 th**	**46-55 th**	**56-65 th**	**Lebih dari 65 th**	
Yes	0	0	0	1	7	24	33	115	180
	0%	0%	0%	0%	2%	7%	9%	32%	50%
No	2	2	20	15	28	30	39	44	180
	1%	1%	6%	4%	8%	8%	11%	12%	50%

Stroke	***Age***								**Jumlah**
	6-11 th	**12-16 th**	**17-25 th**	**26-35 th**	**36-45 th**	**46-55 th**	**56-65 th**	**Lebih dari 65 th**	
Jumlah	2	2	20	16	35	54	72	159	360
	1%	1%	6%	4%	10%	15%	20%	44%	100%

Berdasarkan Tabel 4.3 terdapat pasien *Stroke* yang berumur lebih dari 65 tahun sebanyak 32%, umur 56 - 65 tahun sebanyak 9%, umur 46 - 55 tahun sebanyak 7%, umur 36 - 45 tahun sebanyak 2%, umur 26 - 35 tahun sebanyak 0.0%. sedangkan pasien yang tidak *Stroke* berumur lebih dari 65 tahun sebanyak 12%, umur 46 - 55 tahun sebanyak 8%, umur 56 - 65 tahun sebanyak 11%, umur 36 - 45 tahun sebanyak 8%, umur 26 - 35 tahun sebanyak 4%, umur 17 - 25 tahun sebanyak 6%, umur 12 - 16 tahun sebanyak 1% serta pasien berumur 6 - 11 tahun sebanyak 1%.

Tabel 4.4 Tabulasi Hipertensi terhadap *Stroke*

Stroke	**Hypertensi**		**Jumlah**
	Yes	**No**	
Yes	57	123	180
	16%	34%	50%
No	26	154	180
	7%	43%	50%
Jumlah	83	277	360
	23%	77%	100%

Berdasarkan Tabel 4.4 terdapat pasien *Stroke* dengan tekanan darah tinggi sebanyak 16% dan tekanan darah normal sebanyak 34%, serta pasien yang tidak *Stroke* dengan tekanan darah tinggi sebanyak 7% dan tekanan darah normal sebanyak 43%.

Tabel 4.5 Tabulasi Sakit Jantung terhadap *Stroke*

Stroke	***Heart Disease***		**Jumlah**
	Yes	**No**	
Yes	36	144	180

Stroke	***Heart Disease***		**Jumlah**
	Yes	**No**	
	10%	40%	50%
No	12	168	180
	3%	47%	50%
Jumlah	48	312	360
	13%	87%	100%

Berdasarkan Tabel 4.5 terdapat pasien *Stroke* yang mengalami sakit jantung sebanyak 10% dan tidak mengalami sakit jantung sebanyak 40% serta pasien yang tidak *Stroke* dan mengalami sakit jantung sebanyak 3% dan tidak mengalami sakit jantung sebanyak 47%.

Tabel 4.6 Tabulasi Tipe Tempat Tinggal Terhadap *Stroke*

Stroke	***Resident Type***		**Jumlah**
	Urban	***Rural***	
Yes	94	86	180
	26%	24%	50%
No	84	96	180
	23%	27%	50%
Jumlah	178	182	360
	49%	51%	100%

Berdasarkan Tabel 4.6 terdapat pasien *Stroke* yang tinggal di daerah pedesaan sebanyak 24% dan yang tinggal di perkotaan sebanyak 26% serta pasien yang tidak *Stroke* tinggal dipedesaan sebanyak 27% dan tinggal didaerah perkotaan sebanyak 23%

Tabel 4.7 Tabulasi Kadar Gula Darah terhadap *Stroke*

Stroke	***Avg Glucose Level***			**Jumlah**
	Kurang dari 101	**101 - 125**	**Lebih dari 125**	
Yes	77	26	77	180
	21%	7%	21%	50%
No	110	32	38	180
	31%	9%	11%	50%
Jumlah	187	58	115	360

	52%	16%	32%	100%

Berdasarkan Tabel 4.7 terdapat pasien *Stroke* dengan kadar gula darah normal sebanyak 21%, kadar gula darah pra-diabetes sebanyak 7% serta kadar gula darah diabetes sebanyak 21%, serta terdapat pasien yang tidak *Stroke* dengan kadar gula darah normal sebanyak 31%, kadar gula darah pra-diabetes sebanyak 9% dan kadar gula darah diabetes sebanyak 11%.

Tabel 4.8 Tabulasi Indeks Massa Tubuh terhadap *Stroke*

Stroke	**BMI**				**Jumlah**
	Kurang dari18.6	**18.6-24.9**	**25-29.9**	**Lebih dari29.9**	
Yes	1	29	64	86	180
	0%	8%	18%	24%	50%
No	4	37	56	83	180
	1%	10%	16%	23%	50%
Jumlah	5	66	120	169	360
	1%	18%	33%	47%	100%

Berdasarkan Tabel 4.8 terdapat pasien *Stroke* dengan berat badan kurang sebanyak 0.0%, berat badan normal sebanyak 8%, berat badan lebih sebanyak 18% dan pasien obesitas sebanyak 24%, serta terdapat pasien yang tidak *Stroke* dengan berat badan kurang sebanyak 1%, berat badan normal sebanyak 10%, berat badan lebih sebanyak 16% dan obesitas sebanyak 23%.

Tabel 4.9 Tabulasi Status Merokok terhadap *Stroke*

Stroke	***Smoking Status***			**Jumlah**
	Formerly Smoked	***Never Smoked***	***Smokes***	
Yes	57	84	39	180
	16%	23%	11%	50%
No	47	97	36	180
	13%	27%	10%	50%
Jumlah	104	181	75	360
	29%	50%	21%	100%

Berdasarkan Tabel 4.9 terdapat pasien *Stroke* yang pernah merokok sebanyak 16%, tidak pernah merokok sebanyak 23%, perokok aktif sebanyak 11%. kemudian pasien yang tidak *Stroke* yang pernah merokok sebanyak 13%, tidak pernah merokok sebanyak 27%, perokok aktif sebanyak 10%.

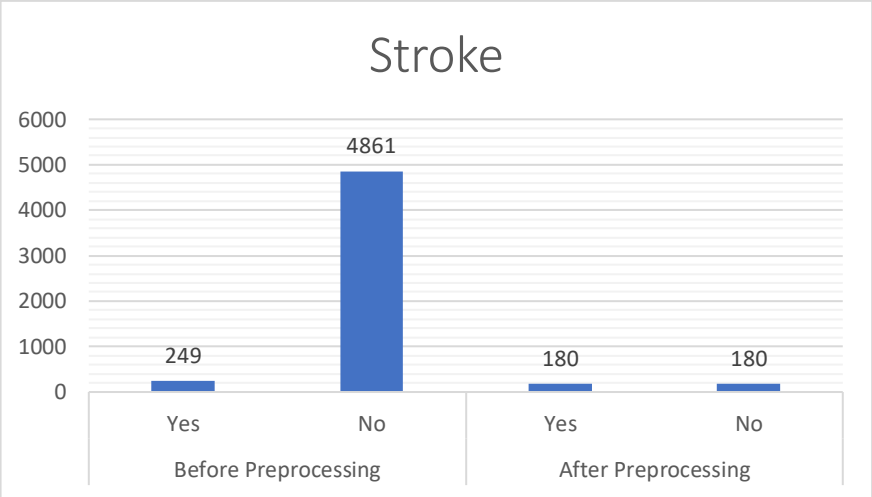
4.2 *Preprocessing*

4.2.1 *Data Cleaning*

Data yang didapat dari *Kaggle* sebelum dilakukan proses *preprocessing* berjumlah 5110. Berikut adalah sampel data yang diambil dari *Kaggle* (Lampiran 1). Pada proses *cleaning* dilakukan pembersihan data terhadap nilai yang kosong sebanyak 201, pembersihan terhadap nilai *unknown* pada variabel *BMI* sebanyak 1482. Data setelah proses penghapusan terhadap nilai yang kosong, tidak diketahui berjumlah 3427. Setelah dilakukan pembersihan data selanjutnya dilakukan penyeimbangan *class* karena terdapat *imbalance dataset.*

$$Imbalanced\ Ratio\ (IR) = \frac{Majority\ Class}{Minority\ Class} = \frac{Stroke\ Yes}{Stroke\ No} = \frac{180}{3247}$$

Jika dilihat dari hasil perbandingan tersebut, nampak terlihat bahwa *dataset* yang digunakan tidak dalam kondisi *balance* (seimbang). Selanjutnya adalah membuat data tersebut menjadi *balance* menggunakan teknik *undersampling* yaitu dengan meratakan sampel data dari kedua *class* dengan mengurangi data dan membuang sampel dari *class* mayoritas. Data yang dihasilkan setelah dilakukan proses *balance* sebanyak 360 data. Tabel data setelah dilakukan proses *cleaning* terdapat pada (Lampiran 2) dan berikut gambaran data sebelum dan setelah proses *prepocessing*.



Gambar 4.2 Data sebelum dan setelah proses *preprocessing*

4.2.2 *Feature Selection*

Proses *feature selection* dilakukan untuk memilih variabel yang akan digunakan dalam penelitian. Pemilihan variabel dilakukan berdasarkan penyebab terjadinya *Stroke* (Hopkins, n.d.). variabel yang digunakan pada penelitian yaitu *gender, age, hypertension, heart disease, resident type, avg glucose level, BMI, smoking status* dan *Stroke*. Atribut yang akan dihapus yaitu *id, ever maried* dan *work type*. Berikut pada Tabel dibawah ini pemilihan atribut yang digunakan dalam penelitian.

Tabel 4.10 Proses *Selection*

No	Atribut	Keterangan	
1	Id	×	No
2	*Gender*	✓	Yes
3	*Age*	✓	Yes
4	Hypertenseion	✓	Yes
5	*Heart Disease*	✓	Yes
6	Ever Maried	×	No
7	Work Type	×	No
8	Residen Type	✓	Yes
9	*Avg Glucose Level*	✓	Yes
10	BMI	✓	Yes
11	*Smoking Status*	✓	Yes

No	Atribut	Keterangan	
12	*Stroke*	✓	Label/Class

Proses pemilihan atribut yang akan digunakan dalam penelitian didasarkan pada faktor-faktor risiko terjadinya *Stroke* (Mayo Clinic Staff, 2022). data yang telah dilakukan proses *selection* terdapat pada (Lampiran 3).

4.2.3 *Data Transformation*

Proses transformasi data dilakukan untuk mengubah data pada *dataset* sesuai dengan format yang dapat diproses oleh software yang digunakan. Pada proses ini terdapat dua proses transformasi yaitu perubahan pada atribut *hypertension*, *heart disease* dan *Stroke* dimana dilakukan pengubahan nilai atribut dari nominal menjadi numerik dan proses *discretization* dengan mengubah nilai atribut dari kontinu ke kategorik. Berikut proses transformasi nominal ke numerik dapat dilihat pada tabel dibawah ini.

Tabel 4.11 Data Setelah Proses *Transformation*

Atribut	Sebelum Transformasi	Setelah Trasnformasi
Hypertension	0	No
	1	Yes
Heart Disease	0	No
	1	Yes
Stroke	0	No
	1	Yes

Tabel 4.11 adalah tampilan data sebelum dan setelah proses mengubah nilai atribut nominal ke numerik, data selengkapnya terdapat pada (lampiran 4). Selanjutnya adalah proses *discretization*, pada proses ini dilakukan pengelompokan dimana nilai-nilai kontinu menjadi data dengan nilai interval.

perubahan nilai dilakukan pada variabel *age*, *avg glucosa level* dan BMI. Proses pengelompokan didasarkan pada ketentuan-ketentuan sebagai berikut:

1. *Age*

   Berdasarkan berbagai kondisi terkait umur, DPR RI kemudian menetapkan rancangan Undang-Undang tentang perubahan atas Undang-Undang Nomor 13 Tahun 1998. Dimana klasifikasi usia menurut kementrian kesehatan sebagai berikut (Hakim, 2020):

   a. 0 – 5 th (masa balita)
   b. 6 – 11 th (masa kanak-kanak)
   c. 12 – 16 th (masa remaja awal)
   d. 17 – 25 th (masa remaja akhir)
   e. 26 – 35 th (masa dewasa awal)
   f. 36 – 45 th (masa dewasa akhir)
   g. 46 – 55 th (masa lansia awal)
   h. 56 – 65 th (masa lansia akhir)
   i. Lebih dari 65 (masa manula)

2. *Avg Glucose Level*

   Level glukosa normal sebelum makan (GDP) dalam satuan miligram per desiliter (mg/dL) sebagai berikut (Lestari, 2021):

   a. Kurang dari 101 mg/dL (Normal)
   b. 101 – 125 mg/dL (Pradiabetes)
   c. Lebih dari 125 mg/dL (Diabetes)

3. BMI

Indeks massa tubuh (*body mass indeks*) adalah metrik standar yang digunakan untuk menentukan siapa saja yang masuk dalam goolongan berat badan sehat dan tidak sehat. Berikut standar kategori berat badan pria dan wanita menurut WHO (Puji, 2022).

a. Kurang dari 18.6 (berat badan kurang)

b. 18.6 – 24.9 (berat badan normal)

c. 25 – 29.9 (berat badan berlebihan)

d. Lebih dari 29.9 (obesitas)

Pada proses *discretization* dari *type* kontinu ke kategorik secara lengkap terdapat pada tabel berikut. Data secara lengkap setelah melalui proses *preprocessing* terdapat pada (Lampiran 5).

Tabel 4.12 Type Atribut Data Pasien *Stroke*

No	Atribut	Row Data Type	Ready Data Type
1	*Gender*	Kategorik Nominal	Kategorik Nominal
2	*Age*	Numerik Rasio	Kategorik Ordinal
3	*Hypertensi*	Kategorik Nominal	Kategorik Nominal
4	*Heart Disease*	Kategorik Nominal	Kategorik Nominal
5	*Resident Type*	Kategorik Nominal	Kategorik Nominal
6	*Avg Glucose Level*	Numerik Rasio	Kategorik Ordinal
7	*BMI*	Numerik Rasio	Kategorik Ordinal
8	*Smoking Status*	Kategorik Nominal	Kategorik Nominal
9	*Stroke*	Kategorik Nominal	Kategorik Nominal

4.3 *Classifier*

Pada proses ini dilakukan pemilihan teknik yang akan digunakan serta pembentukan model. Teknik yang digunakan dalam penelitian ini adalah *Decision Tree C4.5*. Dalam pembentukan model tahap pertama yaitu menentukan *node* akar, kemudian dilanjutkan dengan menentukan cabang dari masing-masing *node*, selanjutnya akan dilakukan pembagian kelas pada cabang yang telah diperoleh dan proses tersebut dilakukan berulang-ulang sampai setiap cabang memiliki kelas.

Langkah pertama dalam proses pembentukan pohon keputusan adalah menghitung nilai *entropy* total dan *entropy* dari nilai setiap atribut pada data di (lampiran 1) dihitung menggunakan persamaan (2.1). adapun sebagai contoh perhitungan *entropy total* dan *entropy* dari variabel jenis kelamin sebagai berikut:

Menghitung *entropy total*:

$$Entropy\ total = \left(\left(-\frac{180}{360}\right) * \log_2\left(\frac{180}{360}\right)\right) + \left(\left(-\frac{180}{360}\right) * \log_2\left(\frac{180}{360}\right)\right)$$

$$= 1$$

Kemudian dihitung nilai *entropy* dan nilai *gain* dari variabel *Gender*

a. *Male*

Jumlah Kasus : 145

Ya : 75

Tidak : 70

$$Entropy\ Male = \left(\left(-\frac{75}{145}\right) * \log_2\left(\frac{75}{145}\right)\right) + \left(\left(-\frac{70}{145}\right) * \log_2\left(\frac{70}{145}\right)\right)$$

$$= 0.9991$$

b. *Female*

Jumlah Kasus : 215

Ya : 105

Tidak : 110

$$Entropy\ Female = \left(\left(-\frac{105}{215}\right) * \log_2\left(\frac{105}{215}\right)\right) + \left(\left(-\frac{110}{215}\right) * \log_2\left(\frac{110}{215}\right)\right)$$

$$= 0.9996$$

Dengan langkah yang sama dilakukan pula perhitungan nilai *entropy* pada setiap variabel bebas lainnya yaitu variabel *Age, Hypertension, Heart Disease, Smoking Status, Resident Type, Avg Glucose Level* dan *BMI*. Kemudian proses selanjutnya dengan menghitung nilai *gain* untuk setiap variabel bebas menggunakan persamaan (2.2). Adapun sebagai contoh perhitungan nilai *gain* pada variabel *Gender* adalah sebagai berikut:

$$Gain(s, A) = 1 - \left(\begin{matrix}\left(\frac{145}{360} entropy\ male\right) + \\ \left(\frac{216}{360} entropy\ female\right)\end{matrix}\right)$$

$$= 0.00057$$

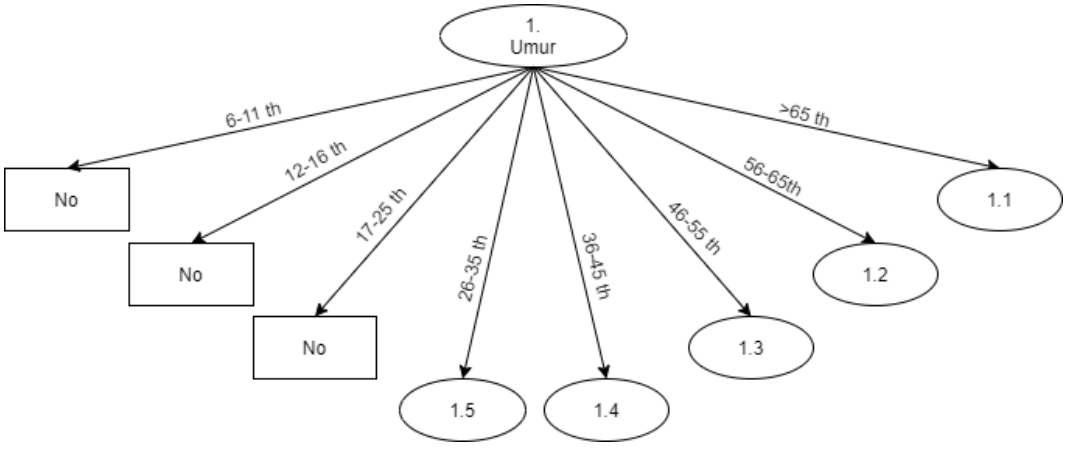
Dengan langkah yang sama dilakukan pula perhitungan nilai *gain* pada setiap variabel bebas lainnya yaitu variabel *Age, Hypertension, Heart Disease, Smoking Status, Resident Type, Avg Glucose Level* dan *BMI*. Adapun hasil perhitungan nilai *entropy* dan *gain* untuk *node* akar disajikan pada Tabel 4.13 berikut:

Tabel 4.13 Perhitungan *Node* Akar

No	Variabel	Nilai	Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Total		360	180	180	1	
1	*Gender*						0.00057
		Male	145	75	70	0.9991	
		Female	215	105	110	0.9996	
2	*Age*						0.19132
		6 - 11 th	2	0	2	0	
		12 - 16 th	2	0	2	0	
		17 - 25 th	20	0	20	0	
		26 - 35 th	16	1	15	0.3372	
		36 - 45 th	35	7	28	0.7219	
		46 - 55 th	54	24	30	0.9910	
		56 - 65 th	72	33	39	0.9949	
		Lebih dari 65 th	159	115	44	0.8509	
3	*Hypertension*						0.03073
		Yes	83	57	26	0.8968	
		No	277	123	154	0.9909	
4	*Heart Disease*						0.02886
		Yes	48	36	12	0.81128	
		No	312	144	168	0.99573	
5	*Smoking Status*						0.00772
		Formerly Smoked	104	57	47	0.9933	
		Never Smoked	181	84	97	0.9962	
		Smokes	75	39	39	0.9811	
6	*Resident Type*						0.00222
		Urban	178	94	84	0.9977	
		Rural	182	86	96	0.9978	
7	*Avg Glucose Level*						0.05424
		Kurang dari 101	187	77	110	0.9774	
		101 - 125	58	26	39	0.9039	
		Lebih dari 125	115	77	38	0.9153	
8	BMI						0.00698
		Kurang dari 18.6	5	1	4	0.7219	
		18.6 - 24.9	66	29	37	0.9893	

No	Variabel	Nilai	Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		25 - 29.9	120	64	56	0.9967	
		Lebih dari 29.9	169	86	83	0.9997	

Berdasarkan Tabel 4.13 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Age* yaitu sebesar 0.27169. Dengan demikian *Age* akan menjadi *node* akar dalam pohon keputusan. Ada 8 nilai variabel pada *Age* yaitu 6 - 11 tahun, 12 - 16 tahun, 17 - 25 tahun, 26 - 35 tahun, 36 - 45 tahun, 46 - 55 tahun, 56 - 65 tahun, Lebih dari 65 tahun. Dari kedelapan nilai tersebut, nilai variabel 6 - 11 tahun diklasifikasikan kasus menjadi keputusan No, nilai variabel 12 - 16 tahun diklasifikasikan kasus menjadi keputusan No, dan nilai variabel 17 - 25 tahun diklasifikasikan kasus menjadi keputusan No, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel 26 - 35 tahun, 36 - 45 tahun, 46 - 55 tahun, 56 - 65 tahun dan Lebih dari 65 tahun masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.3 Pohon Keputusan *Node* Akar

Setelah *node* akar terbentuk, langkah selanjutnya melakukan perhitungan ulang untuk *node* selanjutnya namun data yang di gunakan adalah sisa data terhadap komposisi kelas yang masuk dalam *node* selanjutnya.

1. Perhitungan nilai variabel umur lebih dari 65 tahun

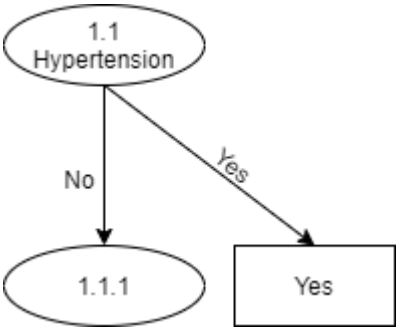
   a. *Node* 1.1

Tabel 4.14 Perhitungan nilai variabel lebih dari65 tahun *Node* 1.1

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Age lebih dari 65 Tahun		159	115	44	0.8510	
1	*Gender*						0.0012
		Male	63	47	16	0.8175	
		Female	96	68	28	0.8709	
2	*Hypertension*						0.0180
		Yes	51	42	9	0.6723	
		No	108	73	35	0.9088	
3	*Heart Disease*						0.0024
		Yes	35	27	8	0.7755	
		No	124	88	36	0.8691	
4	*Smoking Status*						0.0037
		Formerly Smoked	48	37	11	0.7766	
		Never Smoked	90	63	27	0.8813	
		Smokes	21	15	6	0.8631	
5	*Resident Type*						0.0018
		Urban	84	59	25	0.8784	
		Rural	75	56	19	0.8165	
6	*Avg Glucose Level*						0.0160
		Kurang dari 101	71	48	23	0.9086	
		101 - 125	19	12	7	0.9495	
		Lebih dari 125	69	55	14	0.7277	

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
7	BMI						0.0133
		Kurang dari 18.6	2	1	1	1.0000	
		18.6 - 24.9	31	25	6	0.7088	
		25 - 29.9	54	42	12	0.7642	
		Lebih dari 29.9	72	46	26	0.9436	

Berdasarkan Tabel 4.14 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Hypertension* yaitu sebesar 0.0180. Dengan demikian *Hypertension* akan menjadi cabang 1.1 dalam pohon keputusan. Ada 2 nilai variabel pada *Heart Disease* yaitu *Yes* dan *No*. Dari kedua nilai tersebut, nilai variabel *Yes* diklasifikasikan kasus menjadi keputusan *Yes* sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel *No* masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.4 *Node* 1.1

b. *Node* 1.1.1

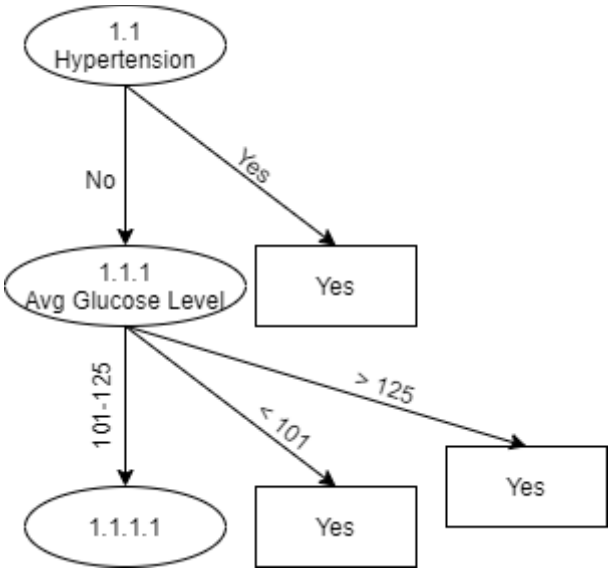
Tabel 4.15 Perhitungan nilai variabel lebih dari 65 tahun *Node* 1.1.1

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Hypertension (No)		108	73	35	0.9088	
1	*Gender*						0.0001

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		Male	47	32	15	0.9035	
		Female	61	41	20	0.9127	
2	*Heart Disease*						0.0053
		Yes	24	18	6	0.8113	
		No	84	55	29	0.9297	
3	*Smoking Status*						0.0031
		Formerly Smoked	32	23	9	0.8571	
		Never Smoked	60	39	21	0.9341	
		Smokes	16	11	5	0.8960	
4	*Resident Type*						0.0028
		Urban	60	39	21	0.9341	
		Rural	48	34	14	0.8709	
5	*Avg Glucose Level*						0.0229
		Kurang dari 101	51	32	19	0.9526	
		101 - 125	16	9	6	0.9976	
		Lebih dari 125	41	32	9	0.7593	
6	BMI						0.0121
		Kurang dari 18.6	2	1	1	1	
		18.6 - 24.9	22	16	6	0.8454	
		25 - 29.9	36	27	9	0.8113	
		Lebih dari 29.9	48	28	20	0.9799	

Berdasarkan Tabel 4.15 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Avg Glucose Level* yaitu sebesar 0.0229. Dengan demikian *Avg Glucose Level* akan menjadi cabang 1.1.1 dalam pohon keputusan. Ada 3 nilai variabel pada *Avg Glucose Level* yaitu Kurang dari 101, 101 – 125 dan lebih dari 125. Dari ketiga nilai tersebut, nilai variabel Kurang dari 101

diklasifikasikan kasus menjadi keputusan *Yes*, nilai variabel lebih dari 125 diklasifikasikan kasus menjadi keputusan *Yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel 101 – 125 masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.5 *Node* 1.1.1

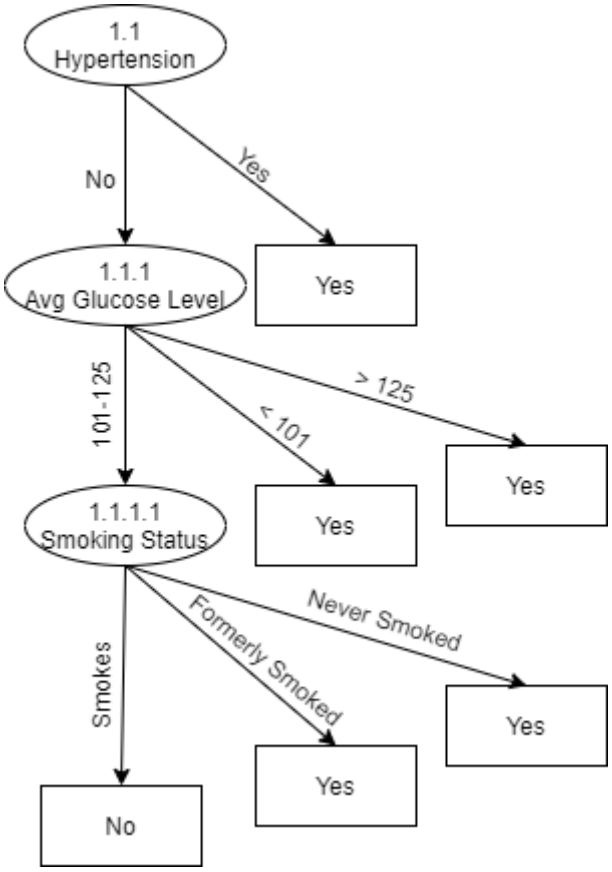
c. *Node* 1.1.1.1

Tabel 4.16 Perhitungan nilai variabel lebih dari 65 tahun *Node* 1.1.1.1

No	**Atribut**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	Avg Glucose Level (101 - 125)		16	9	7	0.9887	
1	*Gender*						0.1381
		Male	6	5	1	0.6500	
		Female	10	4	6	0.9710	
2	*Heart Disease*						0.0075
		Yes	3	2	1	0.9183	
		No	13	7	6	0.9957	
3	*Smoking*						0.2081

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Status						
		Formerly Smoked	6	3	3	1	
		Never Smoked	8	6	2	0.8113	
		Smokes	2	0	2	0	
4	*Resident Type*						0.0038
		Urban	12	7	5	0.9799	
		Rural	4	2	2	1	
5	BMI						0.1853
		Kurang dari 18.6	1	0	1	0	
		18.6 - 24.9	5	3	2	0.9710	
		25 - 29.9	2	2	0	0	
		Lebih dari 29.9	8	4	4	1	

Berdasarkan Tabel 4.16 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Smoking Status* yaitu sebesar 0.1875. Dengan demikian *Smoking Status* akan menjadi cabang 1.1.1.1 dalam pohon keputusan. Ada 3 nilai variabel pada *Smoking Status* yaitu *Formerly Smoked*, *Never Smoked*, dan *Smokes*. Dari ketiga nilai tersebut, nilai variabel *Formerly Smoked* diklasifikasikan kasus menjadi keputusan Yes, nilai variabel *Never Smoked* diklasifikasikan kasus menjadi keputusan Yes, nilai variabel *Smokes* diklasifikasikan kasus menjadi keputusan No sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.6 *Node* 1.1.1.1

2. Perhitungan nilai variabel *age* 56 - 65 tahun

a. *Node* 1.2

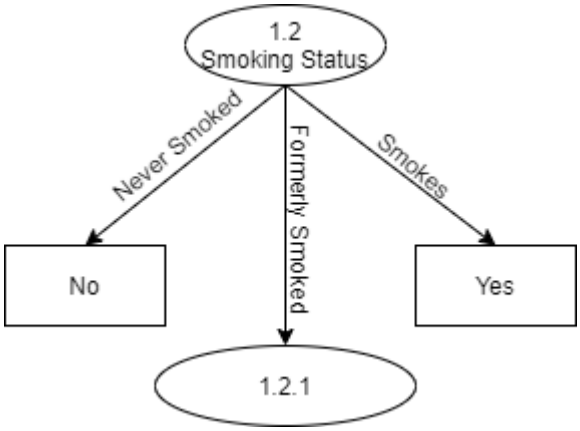
Tabel 4.17 Perhitungan nilai variabel 56 - 65 tahun *Node* 1.2

No	**Atribut**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	Age 56-65		72	33	39	0.9949	
1	*Gender*						0.02764
		Male	36	20	16	0.9910	
		Female	36	13	23	0.9436	
2	*Hypertension*						0.01329
		Yes	20	7	13	0.9340	
		No	52	26	26	1	
3	*Heart Disease*						0.03859
		Yes	11	8	3	0.8453	
		No	61	25	36	0.9764	

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
4	*Smoking Status*						0.17989
		Formerly Smoked	26	14	12	0.9957	
		Never Smoked	30	6	24	0.7219	
		Smokes	16	13	3	0.6962	
5	*Resident Type*						0.01316
		Urban	34	18	16	0.9975	
		Rural	38	15	23	0.9677	
6	*Avg Glucose Level*						0.07461
		Kurang dari 101	32	9	23	0.8571	
		101 - 125	15	9	6	0.9709	
		Lebih dari 125	25	15	10	0.9709	
7	BMI						0.02772
		Kurang dari 18.6	1	0	1	0	
		18.6 - 24.9	7	2	5	0.8631	
		25 - 29.9	17	7	10	0.9774	
		Lebih dari 29.9	47	24	23	0.9996	

Berdasarkan Tabel 4.17 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Smoking Status* yaitu sebesar 0.179899. Dengan demikian *Smoking Status* akan menjadi cabang 1.2 dalam pohon keputusan. Ada 3 nilai variabel pada *Smoking Status* yaitu *Formerly Smoked*, *Never Smoked*, dan *Smokes*. Dari ketiga nilai tersebut, nilai variabel *Never Smoked* diklasifikasikan kasus menjadi keputusan No, nilai variabel *Smokes* diklasifikasikan kasus menjadi keputusan Yes, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel *Formerly Smoked* masih perlu dilakukan

perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.7 *Node* 1.2

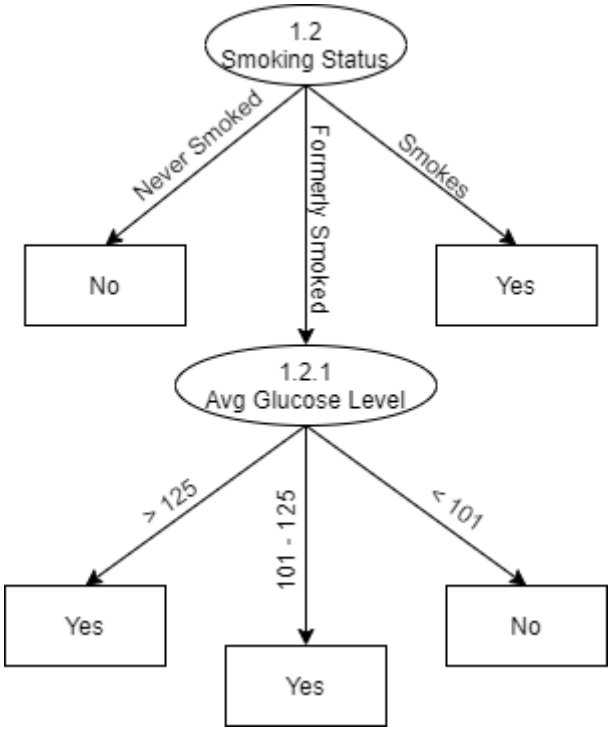
b. *Node* 1.2.1

Tabel 4.18 Perhitungan nilai variabel 56 - 65 tahun *Node* 1.2.1

No	**Atribut**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	Formerly Smoked		26	14	12	0.9957	
1	*Gender*						0.0172
		Male	13	8	5	0.9612	
		Female	13	6	7	0.9957	
2	*Hypertension*						0.1789
		Yes	7	1	6	0.5916	
		No	19	13	6	0.8997	
3	*Heart Disease*						0.0063
		Yes	3	2	1	0.9183	
		No	23	12	11	0.9986	
5	*Resident Type*						0.0172
		Urban	13	8	5	0.9612	
		Rural	13	6	7	0.9957	
6	*Avg Glucose Level*						0.5792
		Kurang dari 101	15	3	12	0.7219	
		101 - 125	4	4	0	0	

No	Atribut		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		Lebih dari 125	7	7	0	0	
7	BMI						0.0245
		18.6 - 24.9	1	0	5	1	
		25 - 29.9	7	3	4	0.9852	
		Lebih dari 29.9	18	11	7	0.9640	

Berdasarkan Tabel 4.18 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Avg Glucose Level* yaitu sebesar 0.5792. Dengan demikian *Avg Glucose Level* akan menjadi cabang 1.2.1 dalam pohon keputusan. Ada 3 nilai variabel pada *Smoking Status* yaitu Kurang dari 101, 101 – 125 dan Lebih dari 125. Dari ketiga nilai tersebut, nilai variabel *never* Kurang dari 101 diklasifikasikan kasus menjadi keputusan *No*, nilai variabel 101 - 125 diklasifikasikan kasus menjadi keputusan *Yes*, nilai variabel Lebih dari 125 diklasifikasikan kasus menjadi keputusan *Yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.8 *Node* 1.2.1

3. Perhitungan nilai variabel *age* 46 - 55 tahun

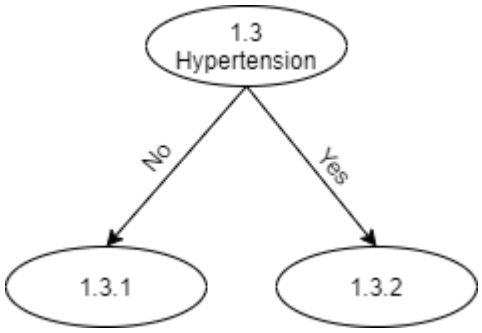
a. *Node* 1.3

Tabel 4.19 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.3

No	**Variabel**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	46-55		54	24	30	0.9910	
1	*Gender*						0.0154
		Male	20	7	13	0.9340	
		Female	34	17	17	1	
2	*Hypertension*						0.0666
		Yes	9	7	2	0.7642	
		No	45	17	28	0.9564	
3	*Heart Disease*						0.0003
		Yes	2	1	1	1	
		No	52	23	29	0.9903	
4	*Smoking Status*						0.0122
		Formerly Smoked	12	4	8	0.9183	

		Never Smoked	24	12	12	1	
		Smokes	18	8	10	0.9910	
5	*Resident Type*						0.0002
		Urban	31	14	17	0.9932	
		Rural	23	10	13	0.9876	
6	*Avg Glucose Level*						0.0026
		Kurang dari 101	30	13	17	0.9871	
		101 - 125	12	5	7	0.9798	
		Lebih dari 125	12	6	6	1	
7	BMI						0.0589
		18.6 - 24.9	11	2	9	0.6840	
		25 - 29.9	18	10	8	0.9910	
		Lebih dari 29.9	25	13	12	0.9988	

Berdasarkan Tabel 4.19 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Hypertension* yaitu sebesar 0.0666. Dengan demikian *Hypertension* akan menjadi cabang 1.3 dalam pohon keputusan. Ada 2 nilai variabel pada *Smoking Status* yaitu *Yes* dan *No*. Dari kedua nilai tersebut masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.9 *Node* 1.3

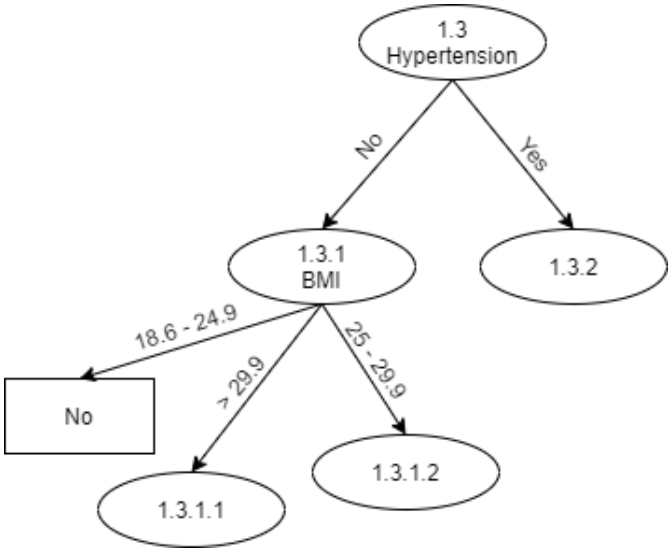
b. *Node* 1.3.1

Tabel 4.20 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.3.1

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	HP - No		45	17	28	0.9565	
1	*Gender*						0.0012
		Male	17	6	11	0.9367	
		Female	28	11	17	0.9666	
3	*Heart Disease*						0.0154
		Yes	1	0	1	0.0000	
		No	44	17	27	0.9624	
4	*Smoking Status*						0.0198
		Formerly Smoked	9	2	7	0.7642	
		Never Smoked	19	8	11	0.9819	
		Smokes	17	7	10	0.9774	
5	*Resident Type*						0.0042
		Urban	26	9	17	0.9306	
		Rural	19	8	11	0.9819	
6	*Avg Glucose Level*						0.0050
		Kurang dari 101	28	11	17	0.9666	
		101 - 125	10	4	6	0.9710	
		Lebih dari 125	7	2	5	0.8631	
7	BMI						0.0970
		18.6 - 24.9	11	1	10	0.4395	
		25 - 29.9	16	8	8	1.0000	
		Lebih dari 29.9	18	8	10	0.9911	

Berdasarkan Tabel 4.20 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah BMI yaitu sebesar 0.0970. Dengan demikian BMI akan menjadi cabang 1.3.1 dalam pohon keputusan. Ada 3 nilai variabel pada BMI

yaitu 18.6 – 24.9, 25 – 29.9 dan Lebih dari 29.9. Dari ketiga nilai tersebut, nilai variabel 18.6 – 24.9 diklasifikasikan kasus menjadi keputusan *No*, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel 25 – 29.9 dan Lebih dari 29.9 masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.10 *Node* 1.3.1

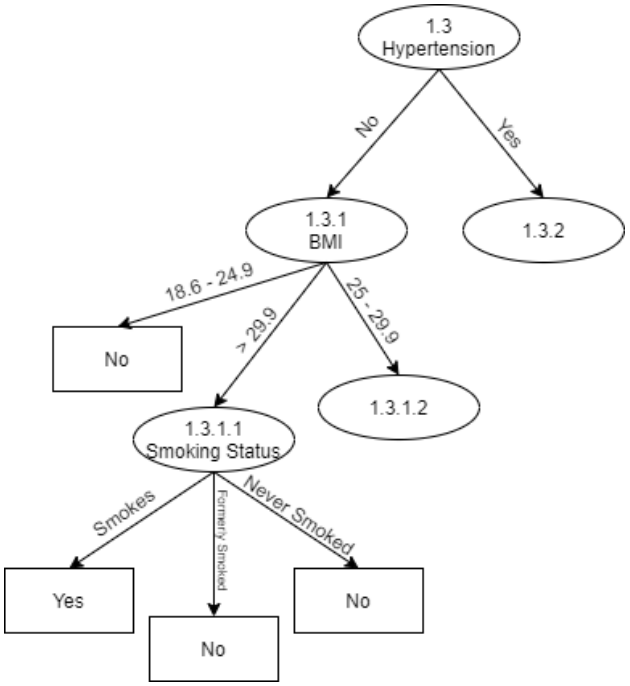
c. *Node* 1.3.1.1

Tabel 4.21 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.3.1.1

No	**Variabel**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	BMI Lebih dari 29.9		18	8	10	0.9911	
1	*Gender*						0.0045
		Male	6	3	3	1.0000	
		Female	12	5	7	0.9799	
3	*Heart Disease*						0.0490
		Yes	1	0	1	0.0000	
		No	17	8	9	0.9975	
4	*Smoking Status*						0.1488

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		Formerly Smoked	5	1	4	0.7219	
		Never Smoked	6	2	4	0.9183	
		Smokes	7	5	2	0.8631	
5	*Resident Type*						0.0072
		Urban	10	4	6	0.9710	
		Rural	8	4	4	1.0000	
6	*Avg Glucose Level*						0.0364
		Kurang dari 101	9	4	5	0.9911	
		101 - 125	3	2	1	0.9183	
		Lebih dari 125	6	2	4	0.9183	

Berdasarkan Tabel 4.21 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Smoking Status* yaitu sebesar 0.1488. Dengan demikian *Smoking Status* akan menjadi cabang 1.3.1.1 dalam pohon keputusan. Ada 3 nilai variabel pada *Smoking Status* yaitu *Formerly Smoked*, *Never Smoked* dan *Smokes*. Dari ketiga nilai tersebut, nilai variabel *Formerly Smoked* diklasifikasikan kasus menjadi keputusan *No*, nilai variabel *Never Smoked* diklasifikasikan kasus menjadi keputusan *No*, nilai variabel *Smokes* diklasifikasikan kasus menjadi keputusan *Yes* sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.11 *Node* 1.3.1.1

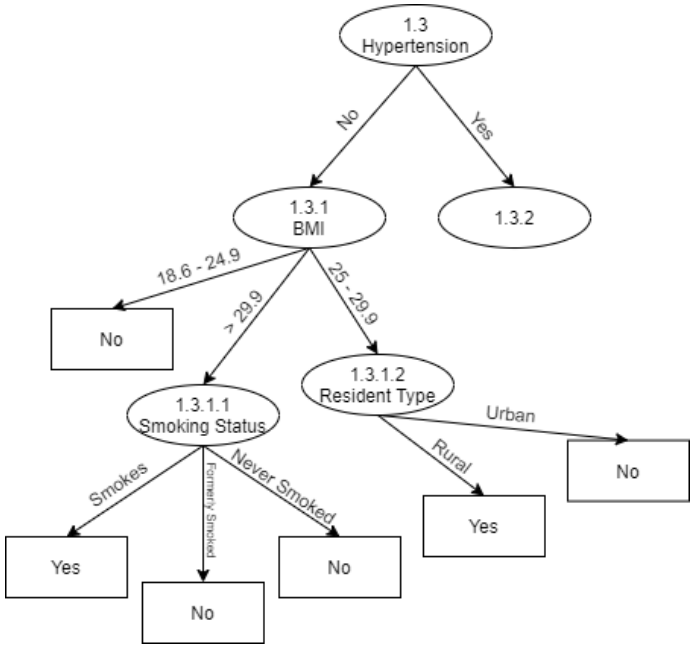
d. *Node* 1.3.1.2

Tabel 4.22 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.3.1.2

No	**Variabel**		**Jumlah (S)**	**Ya (Si)**	**Tidak (Si)**	***Entropy***	***Gain***
	BMI 25-29.9		16	8	8	1	
1	*Gender*						0
		Male	6	3	3	1	
		Female	10	5	5	1	
3	*Heart Disease*						1
		No	16	16	0	0	
4	*Smoking Status*						0.0716
		Formerly Smoked	1	1	0	0	
		Never Smoked	10	5	5	1	
		Smokes	5	2	3	0.9710	
5	*Resident Type*						1
		Urban	8	4	3	1	
		Rural	8	4	5	1	
6	*Avg Glucose*						0.0666

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Level						
		Kurang dari 101	11	6	5	0.9940	
		101 - 125	4	2	2	1	
		Lebih dari 125	1	0	1	0	

Berdasarkan Tabel 4.22 dapat diketahui bahwa variabel dengan *gain* tertinggi ada 2 yaitu *Heart Disease* dan *Resident Type*, karena terdapat 2 nilai variabel dengan *gain* yang sama maka dilakukan perhitungan lebih lanjut sehingga didapatkan nilai *gain* tertinggi yaitu pada variabel *Resident Type*. Dengan demikian *Resident Type* akan menjadi cabang 1.3.1.2 dalam pohon keputusan. Ada 2 nilai variabel pada *Resident Type* yaitu *Urban* dan *Rural.* Dari kedua nilai tersebut, nilai variabel *Urban* diklasifikasikan kasus menjadi keputusan No, nilai variabel *Rural* diklasifikasikan kasus menjadi keputusan Yes, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.12 *Node* 1.3.1.2

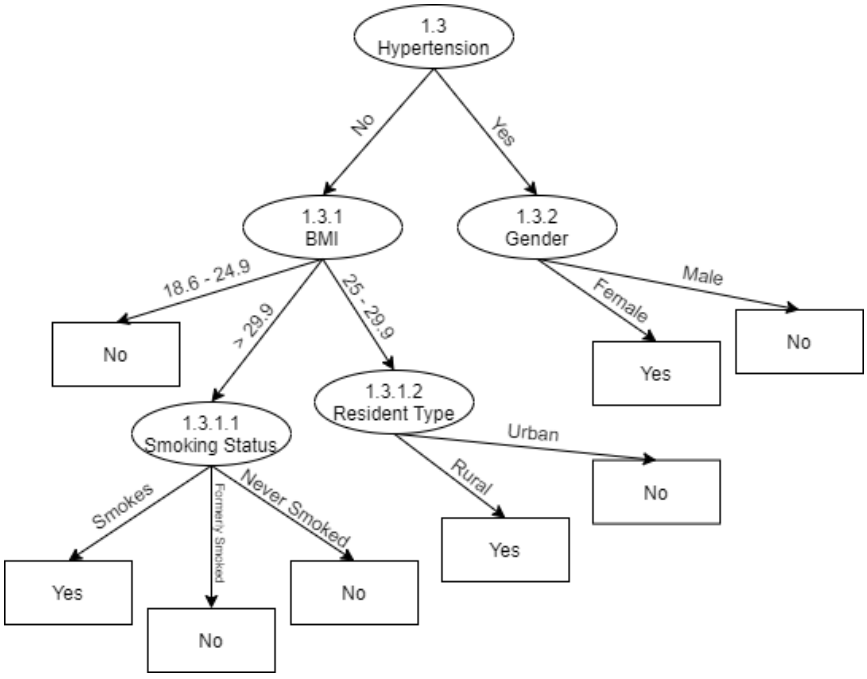
e. *Node* 1.3.2

Tabel 4.23 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.3.2

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	HP - Yes		9	7	2	0.7642	
1	*Gender*						0.4581
		Male	3	1	2	0.9183	
		Female	6	6	0	0.0000	
2	*Heart Disease*						0.0431
		Yes	1	1	0	0.0000	
		No	8	6	2	0.8113	
3	*Smoking Status*						0.0570
		Formerly Smoked	3	2	1	0.9183	
		Never Smoked	5	4	1	0.7219	
		Smokes	1	1	0	0.0000	
4	*Resident Type*						0.3198
		Urban	5	5	0	0.0000	
		Rural	4	2	2	1.0000	
5	*Avg Glucose Level*						0.1409
		Kurang dari 101	2	2	0	0.0000	

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		101 - 125	2	1	1	1.0000	
		Lebih dari 125	5	4	1	0.7219	
6	BMI						0.0929
		25 - 29.9	2	2	0	0.0000	
		Lebih dari 29.9	7	5	2	0.8631	

Berdasarkan Tabel 4.23 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Gender* yaitu sebesar 0.4581. Dengan demikian *Gender* akan menjadi cabang 1.3.2 dalam pohon keputusan. Ada 2 nilai variabel pada *Gender* yaitu *Male* dan *Female*. Dari kedua nilai tersebut, nilai variabel *Male* diklasifikasikan kasus menjadi keputusan *No*, nilai variabel *Female* diklasifikasikan kasus menjadi keputusan *Yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.13 *Node* 1.3.2

4. Perhitungan nilai variabel 36 – 45 tahun

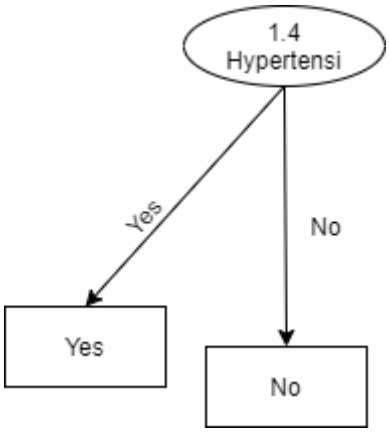
a. *Node* 1.4

Tabel 4.24 Perhitungan nilai variabel 46 - 55 tahun *Node* 1.4

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Age 36 - 45 Tahun		35	7	28	0.7219	
1	*Gender*						0.0359
		Male	12	1	11	0.4138	
		Female	23	6	17	0.8281	
2	*Hypertension*						0.0198
		Yes	2	1	1	1	
		No	33	6	27	0.6840	
3	*Heart Disease*						0
		Yes	30	6	24	0.7219	
		No	5	1	4	0.7219	
4	*Smoking Status*						0.0008
		Formerly Smoked	11	2	9	0.6840	
		Never Smoked	14	3	11	0.7496	
		Smokes	10	2	8	0.7219	
5	*Resident Type*						0.0149
		Urban	15	4	11	0.7219	
		Rural	20	3	17	0.7219	
6	*Avg Glucose Level*						0.0089
		Kurang dari 101	27	6	21	0.7642	
		101 - 125	4	0	4	0.0000	
		Lebih dari 125	4	1	3	0.8113	
7	BMI						0.0101
		18.6 - 24.9	8	1	7	0.5436	
		25 - 29.9	15	3	12	0.7219	
		Lebih dari	12	3	9	0.8113	

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		29.9					

Berdasarkan Tabel 4.24 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Hypertension* yaitu sebesar 0.0198 Dengan demikian *Hypertension* akan menjadi cabang 1.5 dalam pohon keputusan. Ada 2 nilai variabel pada *Hypertension* yaitu *Yes* dan *No*. Dari kedua nilai tersebut, nilai variabel *yes* diklasifikasikan kasus menjadi keputusan *Yes*, nilai variabel *No* diklasifikasikan kasus menjadi keputusan *No*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.14 *Node* 1.4

5. Perhitungan nilai variabel *age* 26 – 35 tahun

a. *Node* 1.5

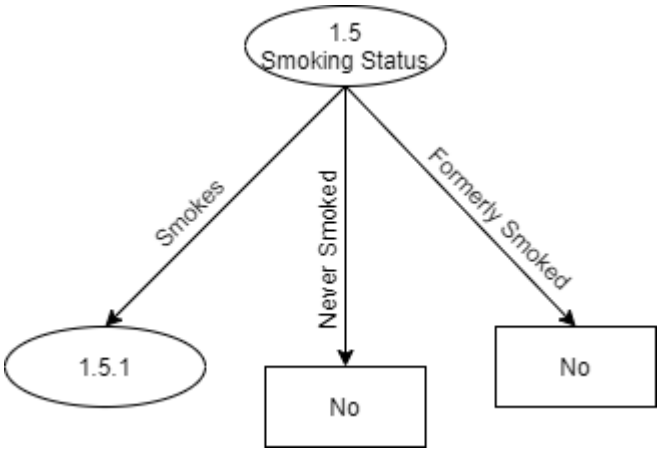
Tabel 4.25 Perhitungan nilai variabel 26 - 35 tahun *Node* 1.5

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	Age 26 - 35 Tahun		16	1	15	0.3373	
1	*Gender*						0.0269
		Male	4	0	4	0.0000	

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
		Female	12	1	11	0.4138	
2	*Hypertension*						0.0060
		Yes	1	0	1	0.0000	
		No	15	1	14	0.3534	
3	*Heart Disease*						0.0000
		Yes	0	0	0	0.0000	
		No	16	1	15	0.3373	
4	*Smoking Status*						0.1345
		Formerly Smoked	2	0	2	0.0000	
		Never Smoked	10	0	10	0.0000	
		Smokes	4	1	3	0.8113	
5	*Resident Type*						0.0194
		Urban	3	0	3	0.0000	
		Rural	13	1	12	0.3912	
6	*Avg Glucose Level*						0.0351
		Kurang dari 101	11	1	10	0.4395	
		101 - 125	2	0	2	0.0000	
		Lebih dari 125	3	0	3	0.0000	
7	BMI						0.1117
		Kurang dari 18.6	1	0	1	0.0000	
		18.6 - 24.9	1	0	1	0.0000	
		25 - 29.9	5	1	4	0.7219	
		Lebih dari 29.9	9	0	9	0.0000	

Berdasarkan Tabel 4.25 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Smoking Status* yaitu sebesar 0.1345. Dengan demikian *Smoking Status* akan menjadi cabang 1.5 dalam pohon keputusan. Ada 3 nilai variabel pada *Smoking Status* yaitu *Formerly Smoked*, *Never Smoked* dan *Smokes.* Dari ketiga nilai tersebut, nilai variabel *Formerly Smoked* diklasifikasikan kasus menjadi keputusan *No*, nilai variabel *Never Smoked*

diklasifikasikan kasus menjadi keputusan *No*, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai pada variabel *Smokes* masih perlu dilakukan perhitungan lagi. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.15 *Node* 1.5

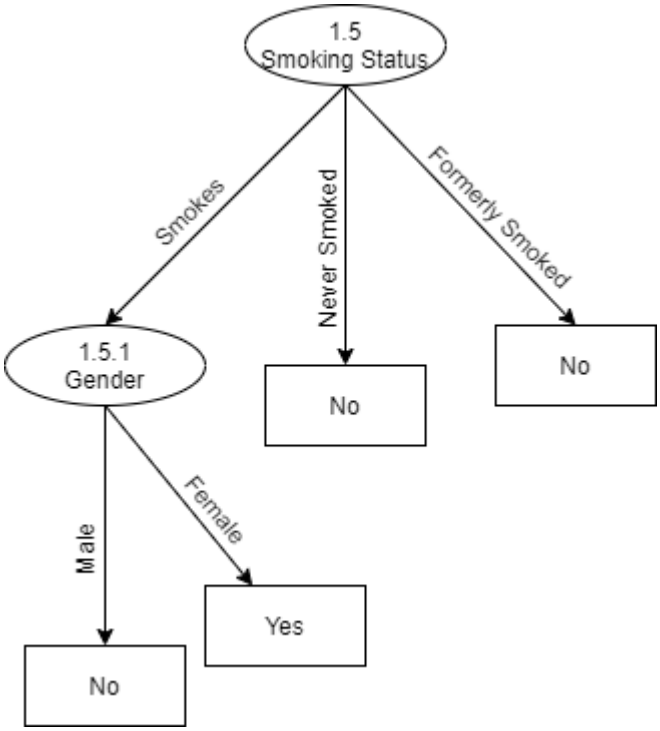
b. *Node* 1.5.1

Tabel 4.26 Perhitungan nilai variabel 26 - 35 tahun *Node* 1.5.1

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
	ST - *Smokes*		4	1	3	0.8113	
1	*Gender*						0.3113
		Male	2	0	2	0.0000	
		Female	2	1	1	1.0000	
2	*Hypertension*						0.0000
		No	4	1	3	0.8113	
3	*Heart Disease*						0.0000
		No	4	1	3	0.8113	
5	*Resident Type*						0.0000
		Urban	0	0	0	0.0000	
		Rural	4	1	3	0.8113	

No	Variabel		Jumlah (S)	Ya (Si)	Tidak (Si)	*Entropy*	*Gain*
6	*Avg Glucose Level*						0.3113
		Kurang dari 101	2	1	1	1.0000	
		101 - 125	1	0	1	0.0000	
		Lebih dari 125	1	0	1	0.0000	
7	BMI						0.3113
		25 - 29.9	2	1	1	1.0000	
		Lebih dari 29.9	2	0	2	0.0000	

Berdasarkan Tabel 4.26 dapat diketahui bahwa variabel dengan *gain* tertinggi adalah *Gender* yaitu sebesar 0.3113. Dengan demikian *Gender* akan menjadi cabang 1.5.1 dalam pohon keputusan. Ada 2 nilai variabel pada *Gender* yaitu *Male* dan *Female*. Dari kedua nilai tersebut, nilai variabel *Male* diklasifikasikan kasus menjadi keputusan *No*, nilai variabel *Female* diklasifikasikan kasus menjadi keputusan *Yes*, sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dari hasil perhitungan di atas dapat digambarkan pohon keputusan sementara sebagai berikut:



Gambar 4.16 *Node* 1.5.1

Dengan memperhatikan pohon keputusan pada Gambar 4.2 sampai pada Gambar 4.16, diketahui bahwa semua kasus telah diklasifikasikan. Dengan demikian, pohon keputusan terbentuk secara lengkap terdapat pada (Lampiran 5). Dari pohon keputusan yang terbentuk hingga terakhir menghasilkan 28 aturan atau *rule*. Keputusan yang dicapai yaitu *Stroke* dengan kelas *yes* dan *no*. *Rule* tersebut diantaranya sebagai berikut:

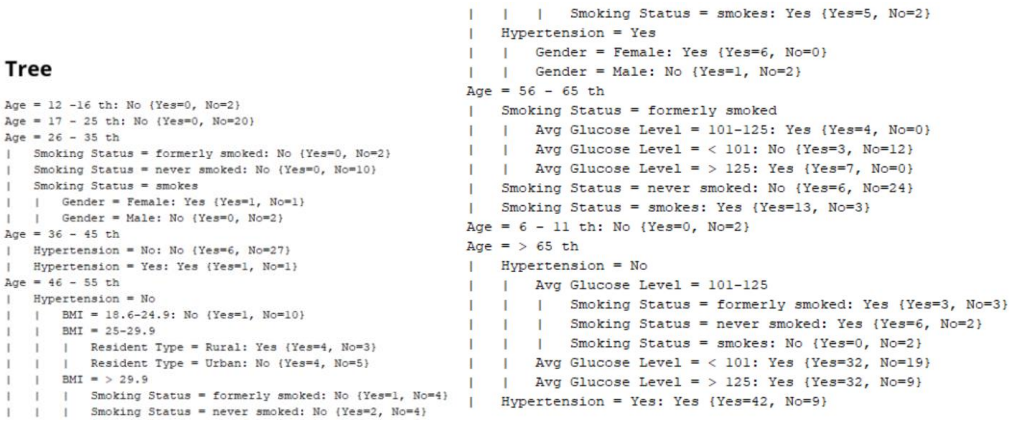
Tabel 4.27 *Rule Tree*

Rule	**Keterangan *Rule***	**Prediksi**
1.	Jika pasien berumur 6 -11 Tahun	Tidak *Stroke*
2.	Jika pasien berumur 12 – 16 Tahun	Tidak *Stroke*
3.	Jika pasien berumur 17 – 25 Tahun	Tidak *Stroke*
4.	Jika pasien berumur 26 – 35 Tahun, pernah merokok	Tidak *Stroke*
5.	Jika pasien berumur 26 – 35 Tahun, tidak pernah merokok	Tidak *Stroke*

Rule	**Keterangan *Rule***	**Prediksi**
6.	Jika pasien berumur 26 – 35 Tahun, perokok aktif, jenis kelamin perempuan	*Stroke*
7.	Jika pasien berumur 26 – 35 Tahun, perokok, jenis kelamin laki-laki	Tidak *Stroke*
8.	Jika pasien berumur 36 – 45 Tahun, tekanan darah normal	Tidak *Stroke*
9.	Jika pasien berumur 36 – 45 Tahun, tekanan darah tinggi	*Stroke*
10.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan normal	Tidak *Stroke*
11.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan berlebihan, dan tinggal di pedesaan	*Stroke*
12.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan berlebihan, dan tinggal di perkotaan	Tidak *Stroke*
13.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, pernah merokok	Tidak *Stroke*
14.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, tidak pernah merokok	Tidak *Stroke*
15.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, perokok aktif	*Stroke*
16.	Jika pasien berumur 46 – 55 Tahun, tekanan darah tinggi, jenis kelamin perempuan	*Stroke*
17.	Jika pasien berumur 46 – 55 Tahun, tekanan darah tinggi, jenis kelamin laki-laki	Tidak *Stroke*
18.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah pradiabetes	*Stroke*
19.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah normal	Tidak *Stroke*
20.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah diabetes	*Stroke*
21.	Jika pasien berumur 56 – 65 Tahun, tidak pernah merokok	Tidak *Stroke*
22.	Jika pasien berumur 56 – 65 Tahun, perokok aktif	*Stroke*
23.	Jika pasien berumur di atas 65 Tahun, tekanan	*Stroke*

Rule	Keterangan *Rule*	Prediksi
	darah normal, gula darah pradiabetes dan pernah merokok	
24.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah pradiabetes dan tidak pernah merokok	*Stroke*
25.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah pradiabetes dan perokok aktif	Tidak *Stroke*
26.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah normal	*Stroke*
27.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah diabetes	*Stroke*
28.	Jika pasien berumur di atas 65 Tahun, tekanan darah tinggi	*Stroke*

Untuk lebih jelasnya dapat dilihat pada model aturan terbentuk teks seperti Gambar 4.16.



Tree

```
Age = 12 -16 th: No (Yes=0, No=2)
Age = 17 - 25 th: No (Yes=0, No=20)
Age = 26 - 35 th
|   Smoking Status = formerly smoked: No (Yes=0, No=2)
|   Smoking Status = never smoked: No (Yes=0, No=10)
|   Smoking Status = smokes
|   |   Gender = Female: Yes (Yes=1, No=1)
|   |   Gender = Male: No (Yes=0, No=2)
Age = 36 - 45 th
|   Hypertension = No: No (Yes=6, No=27)
|   Hypertension = Yes: Yes (Yes=1, No=1)
Age = 46 - 55 th
|   Hypertension = No
|   |   BMI = 18.6-24.9: No (Yes=1, No=10)
|   |   BMI = 25-29.9
|   |   |   Resident Type = Rural: Yes (Yes=4, No=3)
|   |   |   Resident Type = Urban: No (Yes=4, No=5)
|   |   BMI = > 29.9
|   |   |   Smoking Status = formerly smoked: No (Yes=1, No=4)
|   |   |   Smoking Status = never smoked: No (Yes=2, No=4)
|   |   |   Smoking Status = smokes: Yes (Yes=5, No=2)
|   Hypertension = Yes
|   |   Gender = Female: Yes (Yes=6, No=0)
|   |   Gender = Male: No (Yes=1, No=2)
Age = 56 - 65 th
|   Smoking Status = formerly smoked
|   |   Avg Glucose Level = 101-125: Yes (Yes=4, No=0)
|   |   Avg Glucose Level = < 101: No (Yes=3, No=12)
|   |   Avg Glucose Level = > 125: Yes (Yes=7, No=0)
|   Smoking Status = never smoked: No (Yes=6, No=24)
|   Smoking Status = smokes: Yes (Yes=13, No=3)
Age = 6 - 11 th: No (Yes=0, No=2)
Age = > 65 th
|   Hypertension = No
|   |   Avg Glucose Level = 101-125
|   |   |   Smoking Status = formerly smoked: Yes (Yes=3, No=3)
|   |   |   Smoking Status = never smoked: Yes (Yes=6, No=2)
|   |   |   Smoking Status = smokes: No (Yes=0, No=2)
|   |   Avg Glucose Level = < 101: Yes (Yes=32, No=19)
|   |   Avg Glucose Level = > 125: Yes (Yes=32, No=9)
|   Hypertension = Yes: Yes (Yes=42, No=9)
```

Gambar 4.17 Model Aturan *Text Decision Tree*

Dari Gambar 4.17 di atas dapat dijelaskan bahwa faktor yang paling mempengaruhi pada *node* pertama adalah *age*, *node* kedua adalah *hypertension*, *node* ketiga adalah *avg glucose level, node* keempat adalah *smoking status.* Untuk *Stroke* model aturan yang terbaik adalah apabila pasien berumur di atas 65 Tahun dengan tekanan darah tinggi dengan jumlah atribut 42. Sedangkan untuk tidak

Stroke model aturan terbaik adalah jika pasien berumur 36 – 45 Tahun dengan tekanan darah normal dengan jumlah atribut 27. Untuk melihat *rule* dengan *gain rasio Stroke* dapat dilihat pada Tabel 4.28 dengan ketentuan *Stroke = Yes*, Tidak *Stroke = No.*

Tabel 4.28 Keterangan *Rule Text* dengan *Gain Ratio*

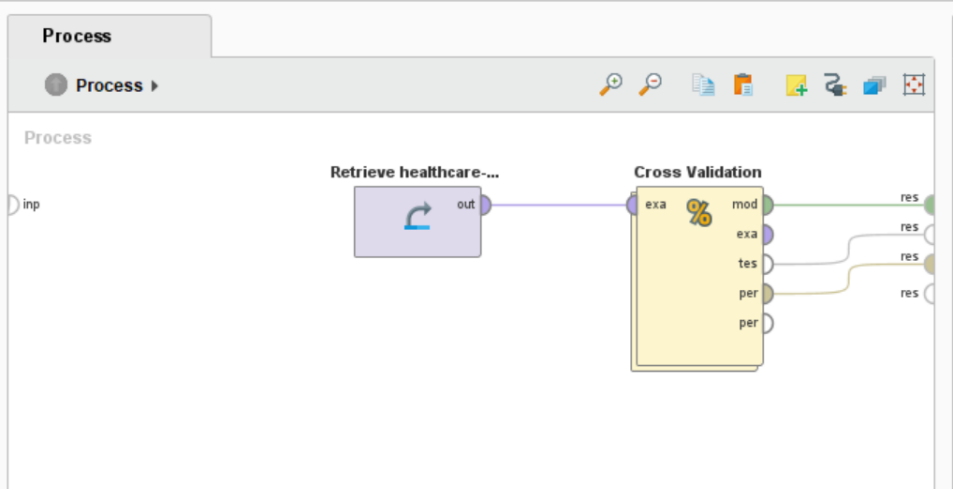
Rule	**Keterangan *Rule***	**Predikat *Gain Rasio***	
		Yes	**No**
1.	Jika pasien berumur 6 -11 Tahun maka Tidak *Stroke*	0	2
2.	Jika pasien berumur 12 – 16 Tahun maka Tidak *Stroke*	0	2
3.	Jika pasien berumur 17 – 25 Tahun maka Tidak *Stroke*	0	20
4.	Jika pasien berumur 26 – 35 Tahun, pernah merokok maka Tidak *Stroke*	0	2
5.	Jika pasien berumur 26 – 35 Tahun, tidak pernah merokok maka Tidak *Stroke*	0	10
6.	Jika pasien berumur 26 – 35 Tahun, perokok aktif, jenis kelamin perempuan maka *Stroke*	1	1
7.	Jika pasien berumur 26 – 35 Tahun, perokok, jenis kelamin laki-laki maka Tidak *Stroke*	0	2
8.	Jika pasien berumur 36 – 45 Tahun, tekanan darah normal maka Tidak *Stroke*	6	27
9.	Jika pasien berumur 36 – 45 Tahun, tekanan darah tinggi maka *Stroke*	1	1
10.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan normal maka Tidak *Stroke*	1	10
11.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan berlebihan, dan tinggal di pedesaan maka *Stroke*	4	3

Rule	**Keterangan *Rule***	**Predikat *Gain Rasio***	
		Yes	**No**
12.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan berlebihan, dan tinggal di perkotaan maka Tidak *Stroke*	4	5
13.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, pernah merokok maka Tidak *Stroke*	1	4
14.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, tidak pernah merokok maka Tidak *Stroke*	2	4
15.	Jika pasien berumur 46 – 55 Tahun, tekanan darah normal, berat badan obesitas, perokok aktif maka *Stroke*	5	2
16.	Jika pasien berumur 46 – 55 Tahun, tekanan darah tinggi, jenis kelamin perempuan maka *Stroke*	6	0
17.	Jika pasien berumur 46 – 55 Tahun, tekanan darah tinggi, jenis kelamin laki-laki maka Tidak *Stroke*	1	2
18.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah pradiabetes maka *Stroke*	4	0
19.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah normal maka Tidak *Stroke*	3	12
20.	Jika pasien berumur 56 – 65 Tahun, pernan merokok, guna darah diabetes maka *Stroke*	7	0
21.	Jika pasien berumur 56 – 65 Tahun, tidak pernah merokok maka Tidak *Stroke*	6	24
22.	Jika pasien berumur 56 – 65 Tahun, perokok aktif maka *Stroke*	13	3
23.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah pradiabetes dan pernah merokok maka *Stroke*	3	3

Rule	**Keterangan *Rule***	**Predikat *Gain Rasio***	
		Yes	**No**
24.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah pradiabetes dan tidak pernah merokok maka *Stroke*	6	2
25.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah pradiabetes dan perokok aktif maka Tidak *Stroke*	0	2
26.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah normal maka *Stroke*	32	19
27.	Jika pasien berumur di atas 65 Tahun, tekanan darah normal, gula darah diabetes maka *Stroke*	32	9
28.	Jika pasien berumur di atas 65 Tahun, tekanan darah tinggi maka *Stroke*	42	9

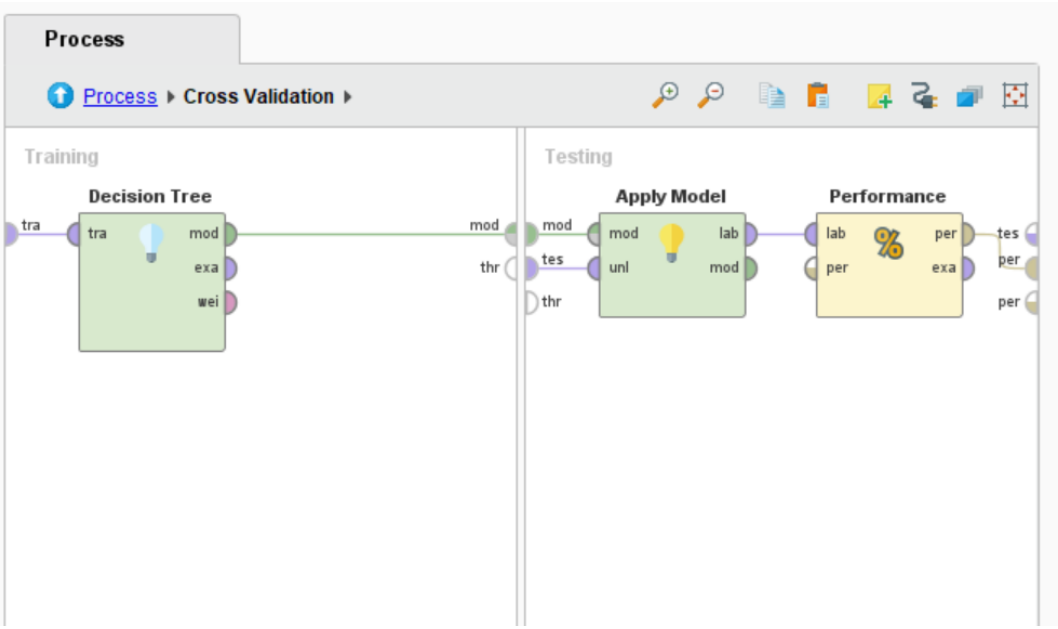
4.4 *Evaluation*

Proses evaluasi dilakukan untuk menganalisis hasil klasifikasi. Pengukuran data dilakukan dengan *confusion matrix* dengan mengevaluasi hasil dari algoritma *Decision Tree C4.5*. Proses evaluasi dilakukan dengan menggunakan *software rapidminer* untuk mengetahui kebenaran dari hasil perhitungan yang telah dilakukan. Atribut yang digunakan sebagai label adalah *Stroke*, penulis menganalisis 360 record data yang akan dipakai sebagai *data training* dan *data testing* dengan teknik *cross validation 10-fold*. Berikut adalah desain pengujian dengan *software RapidMiner* v.9.10.



Gambar 4.18 Proses Awal Uji Validasi

Pada proses awal ini akan dilakukan *import* atau memasukkan data dengan format CSV kedalam *tool* RapidMiner, setelah proses *Import* data selesai, data akan di drop ke panel proses, kemudian dihubungkan ke drop *cross validation*. Setelah itu klik dua kali pada *operator cross validation*, maka akan muncul tampilan sebagai berikut:



Gambar 4.19 Operator *Performance*

Setelah muncul desain proses seperti di atas, masukkan operator *Decision Tree* pada bagian training, kemudian masukkan operator *apply model* dan *performance* pada bagian *testing* yang nantinya akan menghasilkan hasil berupa *Decision Tree* dan *performance* dari metode yang digunakan. Proses validasi menggunakan 2 pengukuran yaitu *confusion matrix* dan Kurva ROC/AUC (*Area Under Curve*).

1. *Confusion Matrix*

   Berikut tabel *confusion matrix* hasil pengujian menggunakan *tool* RapidMiner dengan jumlah data 360. Hasil dari pengujian menghasilkan 248 data teridentifikasi benar dan 112 data teridentifikasi salah.

Tabel 4.29 Confusion Matrix

	True Yes	True No
Pred. Yes	125	57
Pred. No	55	123

Tabel 4.29 adalah tabel *confusion matrix* dari data testing menggunakan *cross validation 10-fold* pada *software RapidMiner*. jumlah *true yes* (TY) sebanyak 125 *record*, *false yes* (FY) sebanyak 57 *record*, *false no* (FN) sebanyak 55 *record*, dan jumlah *true no* (TN) sebanyak 123 *record*. Setelah dilakukan pengujian dengan *confusion matrix* selanjutnya dilakukan perhitungan tingkat *accuracy*, *precision* dan *recall* dengan menggunakan persamaan (2.4), (2.5) dan persamaan (2.6) seperti berikut:

a. Accuracy

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total data sampel yang diuji.

$$Accuracy = \frac{TY+TN}{TY+TN+FY+FN}$$

$$= \frac{125 + 123}{125 + 123 + 55 + 57}$$

$$= 0.6889$$

$$= 68.89\ \%$$

Akurasi merupakan tingkat kedekatan antara nilai prediksi dengan nilai yang sebenarnya. Akurasi yang didapatkan dari model 68.89% menyatakan bahwa dari 360 data pasien, terdapat 248 berhasil diprediksi.

b. *Precision*

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai yes (*true yes*) dibagi dengan jumlah data benar yang bernilai yes (*true yes*) dan data salah yang bernilai yes (*false yes*).

$$Precision = \frac{TY}{TY+FY}$$

$$= \frac{125}{125 + 57}$$

$$= 0.6868$$

$$= 68.68\ \%$$

Precision merupakan peluang kasus yang diprediksi positif yang pada kenyataanya termasuk kedalam kasus kategori positif. Dalam hal ini tingkat *precision* 68.8% yang berarti dari 182 *data predic* pasien *Stroke* model berhasil memprediksi 125 *data actual Stroke* dengan benar.

c. *Recall*

Recall dihitung dengan cara membagi data benar yang bernilai yes (*true yes*) dengan hasil penjumlahan dari data benar yang bernilai yes (*true yes*) dan data salah yang bernilai no (*false no*)

$$Recall = \frac{TY}{TY + FN}$$

$$= \frac{125}{125 + 55}$$

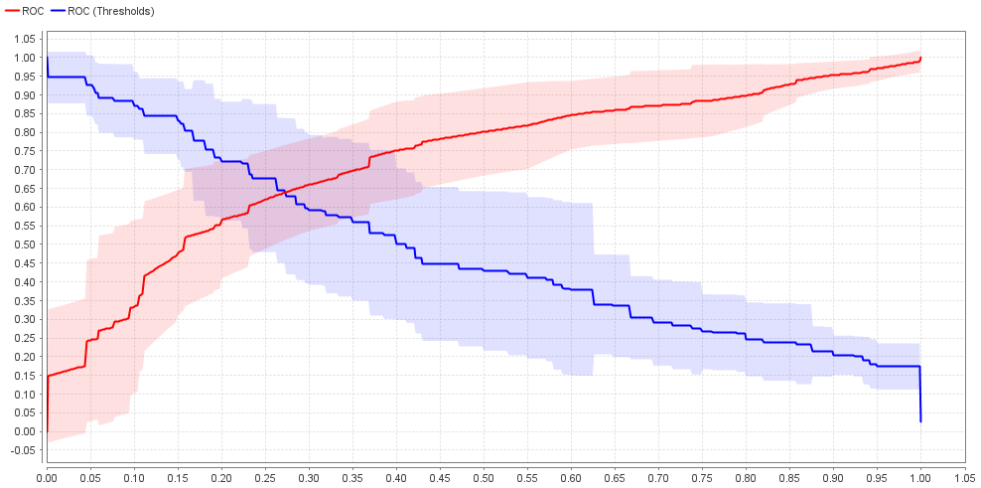
$$= 0.694$$

$$= 69.4\ \%$$

Recall merupakan peluang kasus dengan kategori positif yang dengan tepat diprediksi positif. Dalam hal ini tingkat *recall* 69.4% yang berarti dari *data actual* 180 pasien *Stroke*, model berhasil memprediksi 125 *data predic Stroke.*

2. Kurva ROC/AUC

Hasil analisis *ROC Curve* menggunakan tool RapidMiner dapat dilihat pada Gambar 4.19.



Gambar 4.20 *ROC Curve*

AUC (*Area Under Curve*) dihitung untuk mengukur akurasi dan perbandingan klasifikasi secara virtual dengan *false positif* sebagai garis horizontal (garis biru) dan *true positif* sebagai garis vertikal (garis merah). Semakin dekat kurva ROC pada garis Y(1,0) maka semakin baik model yang dihasilkan. Selama prediksi benar untuk contoh, kurva mengambil satu langkah ke atas (peningkatan TP). Jika prediksi salah, kurva

mengambil satu langkah ke kanan (FP meningkat). Dari data di atas analisis menggunakan tool RapidMiner dengan pengukuran *Decision Tree* didapatkan hasil hubungan antara false yes dengan true yes sebesar 0.726 yang termasuk ke dalam kategori cukup (*Fair Classification*).

4.5 Kajian Keislaman dengan Hasil Penelitian

Kajian terkait prediksi terserang *Stroke* dapat ditinjau dari perspektif Islam. Dalam Islam prediksi mengenai sesuatu dapat dilihat pada surat Yusuf ayat 47-49. Ayat tersebut berisi ramalan Nabi Yusuf mengenai suatu bencana yang akan menimpa umatnya berdasarkan wahyu dari Allah, sebagaimana yang terdapat pada kutipan ayat berikut (Kemenag, 2002):

قَالَ تَزْرَعُونَ سَبْعَ سِنِينَ دَأَبًا فَمَا حَصَدتُّمْ فَذَرُوهُ فِي سُنۢبُلِهِۦٓ إِلَّا قَلِيلًا مِّمَّا تَأْكُلُونَ (٤٧) ثُمَّ يَأْتِي مِنۢ بَعْدِ ذَٰلِكَ سَبْعٌ شِدَادٌ يَأْكُلْنَ مَا قَدَّمْتُمْ لَهُنَّ إِلَّا قَلِيلًا مِّمَّا تُحْصِنُونَ (٤٨) ثُمَّ يَأْتِيْ مِنۢ بَعْدِ ذٰلِكَ عَامٌ فِيْهِ يُغَاثُ النَّاسُ وَفِيْهِ يَعْصِرُوْنَ (٤٩)

Artinya:

"*Dia (Yusuf) berkata, "Agar kamu bercocok tanam tujuh tahun (berturut-turut) sebagaimana biasa; kemudian apa yang kamu tuai hendaklah kamu biarkan di tangkainya kecuali sedikit untuk kamu makan. 47. Kemudian setelah itu akan Datang tujuh (tahun) yang sangat sulit, yang menghabiskan apa yang kamu simpan untuk menghadapinya (tahun sulit), kecuali sedikit dari apa (bibit gandum) yang kamu simpan. 48. Setelah itu akan Datang tahun, di mana manusia diberi hujan (dengan cukup) dan pada masa itu mereka memeras (anggur) ". 49. (QS. Yusuf ayat 47-49).*

Menurut tafsir Al-Muyassar, Yusuf berkata kepada penanyanya soal mimpi raja tersebut, "tafsir mimpi ini adalah hendaknya kalian menanam selama tujuh tahun berturut-turut dengan tekun agar hasil panen menjadi berlimpah. Lalu hasil panen yang kalian hasilkan darinya setiap kali, maka simpanlah dan biarkan dalam bulir-bulirnya, supaya sempurna proses penyimpanannya dari gangguan

ulat dan lebih bertahan lama, kecuali sebagian kecil saja yang kalian makan dari hasil-hasil biji-bijian itu (Ashim & Karimi, 2016).

Ayat di atas menjelaskan tentang bagaimana Nabi Yusuf menafsirkan mimpi Raja Firaun dan meramalkan apa yang akan terjadi di masa depan. Hal tersebut berkorelasi dengan penelitian yang peneliti lakukan yaitu melakukan peramalan risiko seseorang terserang *Stroke*. Dalam penelitian ini diharapkan menghasilkan faktor-faktor dan penyebab seseorang bisa terserang *Stroke*. Hasil yang diperoleh dengan melakukan perkiraan menggunakan algoritma C.45 adalah model dan *rule* yang digunakan untuk menentukan hasil prediksi serta akurasi sebagai tingkat ketepatan hasil.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan dari rumusan masalah di atas beserta hasil dan pembahasannya, maka kesimpulan dari penelitian ini yaitu model prediksi *Stroke* menggunakan 360 data dengan algoritma *Decision Tree C4.5* berhasil memprediksi *Stroke* dengan persentase keakuratan (*Accuracy*) sebesar 68.89% dengan rincian 248 yang berhasil terprediksi dari 360 data yang digunakan. Adapun peluang kasus yang terprediksi positif (*Precision*) sebesar 68.68% dengan rician 125 data *actual Stroke* dari 182 data pasien terprediksi *Stroke*, sedangkan peluang kasus dengan kategori positif (*Recall*) sebesar 69.4% dengan rincian 125 yang berhasil terprediksi dari 180 data *actual Stroke*. Kemudian model *Tree* menghasilkan 28 aturan yang mana 13 memprediksi *Stroke* dan 15 memprediksi tidak *Stroke* serta faktor yang paling berpengaruh berdasarkan *Tree* yang terbentuk adalah variabel Umur, dan setelahnya adalah Hipertensi, Kadar Glukosa dan Status Merokok. Sementara menggunakan perhitungan *ROC curve* diperoleh luas garis di bawah kurva (*Area Under Curve*) sebesar 0.726. Hal ini menunjukkan bahwa model prediksi menggunakan algoritma *Decision Tree C4.5* diklasifikasikan cukup (*fair classification*).

5.2 Saran

Pada penelitian ini peneliti menggunakan metode *Decision Tree* C4.5 sehingga untuk penelitian selanjutnya dapat menggunakan metode yang berbeda sebagai komparasi metode untuk mendapatkan hasil dan model yang lebih baik, serta menggunakan data dan variabel yang lebih banyak.

DAFTAR PUSTAKA

Alsabti, K., Ranka, S., & Singh, V. (1997). *Electrical Engineering and Computer Science An efficient k-means clustering algorithm*.

American Stroke Association. (2020). *About Stroke*. Stroke.Org. https://www.stroke.org/en/about-stroke

Andriani, A. (2013). Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa. *Seminar Nasional Teknologi Informasi Dan Komunikasi 2013*, 163–168.

Ar-Rifa'i, M. N. (2012). *Ringkasan Tafsir IBNU KATSIR* (3rd ed.). GEMA INSANI.

Ashim, A. S. bin M. S. M., & Karimi, I. (2016). Tafsir Al-Muyassar: Memahami Al-Qur'an dengan Terjemahan dan Penafsiran Paling Mudah. In A. Al-Qarni (Ed.), *Tafsir Al-Muyassar* (pp. 1–960). Daarul Haq.

Beynon-Davies, P. (2004). *Database Systems* (3rd ed., Vol. 3). PALGRAVE MACMILLAN. https://doi.org/10.1007/978-0-230-00107-7

Ceballos, F. (2019). *Scikit-Learn Decision Trees Explained*. Medium.Com. https://towardsdatascience.com/scikit-learn-decision-trees-explained-803f3812290d

Dewanto, G., Suwanto, J. W., Riyanto, B., & Turana, Y. (2009). *Panduan Praktis Diagnosis & Tata Laksana Penyakit Saraf*. EGC.

Diwandari, S., & Setiawan, N. A. (2015). PERBANDINGAN ALGORITME J48 DAN NBTREE UNTUK KLASIFIKASI DIAGNOSA PENYAKIT PADA SOYBEAN. In *Seminar Nasional Teknologi Informasi dan Komunikasi*.

Drummond, C., & Holte, R. C. (2003). *C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats OverSampling*. https://www.researchgate.net/publication/245593532

Fedesoriano. (2021). *Stroke Prediction Dataset*. Kaggle.Com. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer. https://doi.org/10.1007/978-3-642-19721-5

Hakim, L. N. (2020). Urgensi Revisi Undang-Undang tentang Kesejahteraan Lanjut Usia. *Jurnal Masalah-Masalah Sosial*, *11*. https://doi.org/10.22212/aspirasi.v11i1.1589

Han, J., Kamber, M., Melton, J., Buxton, S., Teorey, T. J., Lightstone, S. S., Nadeau, T. P., Celko, J., Witten, I., Frank, E., Simsion, G. C., Witt, G. C., Schiller, J., Voisard, A., Halpin, T., Evans, K., Hallock, P., Maclean, B.,

Ceri, S., … Voisard, A. (2006). *Data Mining: Concepts and Techniques* (A. Stephan, Ed.; 2nd ed., Vol. 770). Diane Cerra.

Hopkins, J. (n.d.). *Risk Factors for Stroke*. Hopkinsmedicine.Org. Retrieved March 20, 2022, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/stroke/risk-factors-for-stroke

Indriani, A. (2014). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 5–10.

InfoDATIN. (2019). *Pusat Data dan Informasi - Kementerian Kesehatan Republik Indonesia*. Pusdatin.Kemkes.Go.Id. https://pusdatin.kemkes.go.id/article/view/20031000003/infodatin-stroke.html

Iykra. (2018). *Mengenal Decision Tree dan Manfaatnya*. Medium.Com. https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d

Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). DATA TRANSFORMATION PADA DATA MINING. *Prosiding Konferensi Nasional "Inovasi Dalam Desain Dan Teknologi."*

Kemenag. (2002). *Al-Quran Kemenag* (1st ed.). LPMQ. https://quran.kemenag.go.id/

Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining* (1st ed., Vol. 216). C.V ANDI OFFSET.

Lee, H. (1995). JUSTIFYING DATABASE NORMALIZATION: A COST/BENEFIT MODEL. In *Information Processing & Management* (Vol. 31, Issue I).

Lestari, D. A. (2021). *Berapa Kadar Gula Darah Normal dalam Tubuh?* Hellosehat.Com. https://hellosehat.com/diabetes/kadar-gula-darah-normal/

Mahmood, A. M. (2015). Class Imbalance Learning in Data Mining – A Survey. *International Journal of Communication Technology for Social Networking Services*, *3*(2), 17–36. https://doi.org/10.21742/ijctsns.2015.3.2.02

Mashlahah, S. (2013). *Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree dengan Penerapan Algoritma C4.5*. http://etheses.uin-malang.ac.id/7505/1/07650150.pdf

Mayo Clinic Staff. (2022). *Stroke - Symptoms and causes*. Mayoclinic.Org. https://www.mayoclinic.org/diseases-conditions/stroke/symptoms-causes/syc-20350113

Munir, R. (2010). *Matematika Diskrit* (3rd ed., Vol. 561). Informatika Bandung.

Pramudiono, I. (2003). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*.

Puji, A. (2022, May 20). *Cara Mengukur IMT Orang Dewasa dengan Rumus*. Hellosehat.Com. https://hellosehat.com/nutrisi/cara-menghitung-indeks-massa-tubuh/

Rifai, B. (2013). ALGORITMA NEURAL NETWORK UNTUK PREDIKSI PENYAKIT JANTUNG. In *Techno Nusa Mandiri* (Vol. 1). http://www.bsi.ac.id

Rohman, A., & Rufiyanto, A. (2019). PENERAPAN ALGORITMA DECISION TREE ID3 UNTUK PREDIKSI KELULUSAN MAHASISWA JENJANG PENDIDIKAN D3 DI FAKULTAS TEKNIK UNIVERSITAS PANDANARAN. In *Jurnal NeoTeknika* (Vol. 5, Issue 2).

Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics* (2nd ed., Vol. 387). Penebar Media Pustaka.

Suyanto. (2014). *Artificial Intelligence: Searching, Reasoning, Planning and Learning* (2nd ed.). Informatika, Bandung, Indonesia.

Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*.

LAMPIRAN

Lampiran 1. Sampel Data *ROW Stroke Prediction* dari WHO

No	id	*Gender*	*Age*	*Hyper tension*	*Heart disease*	*Ever married*	*Work type*	*Residence type*	*Avg Glucose Level*	*BMI*	*Smoking Status*	*Stroke*
1	9046	*Male*	67	0	1	Yes	Private	*Urban*	228.69	36.6	*Formerly Smoked*	1
2	51676	*Female*	61	0	0	Yes	Self-employed	*Rural*	202.21	N/A	*Never Smoked*	1
3	31112	*Male*	80	0	1	Yes	Private	*Rural*	105.92	32.5	*Never Smoked*	1
4	60182	*Female*	49	0	0	Yes	Private	*Urban*	171.23	34.4	*Smokes*	1
5	1665	*Female*	79	1	0	Yes	Self-employed	*Rural*	174.12	24	*Never Smoked*	1
6	56669	*Male*	81	0	0	Yes	Private	*Urban*	186.21	29	*Formerly Smoked*	1
7	53882	*Male*	74	1	1	Yes	Private	*Rural*	70.09	27.4	*Never Smoked*	1
8	10434	*Female*	69	0	0	No	Private	*Urban*	94.39	22.8	*Never Smoked*	1
9	27419	*Female*	59	0	0	Yes	Private	*Rural*	76.15	N/A	Unknown	1
10	60491	*Female*	78	0	0	Yes	Private	*Urban*	58.57	24.2	Unknown	1
11	12109	*Female*	81	1	0	Yes	Private	*Rural*	80.43	29.7	*Never Smoked*	1
12	12095	*Female*	61	0	1	Yes	Govt_job	*Rural*	120.46	36.8	*Smokes*	1
13	12175	*Female*	54	0	0	Yes	Private	*Urban*	104.51	27.3	*Smokes*	1
14	8213	*Male*	78	0	1	Yes	Private	*Urban*	219.84	N/A	Unknown	1
15	5317	*Female*	79	0	1	Yes	Private	*Urban*	214.09	28.2	*Never Smoked*	1
16	58202	*Female*	50	1	0	Yes	Self-employed	*Rural*	167.41	30.9	*Never Smoked*	1
17	56112	*Male*	64	0	1	Yes	Private	*Urban*	191.61	37.5	*Smokes*	1
18	34120	*Male*	75	1	0	Yes	Private	*Urban*	221.29	25.8	*Smokes*	1
19	27458	*Female*	60	0	0	No	Private	*Urban*	89.22	37.8	*Never Smoked*	1
20	25226	*Male*	57	0	1	No	Govt_job	*Urban*	217.08	N/A	Unknown	1
21	70630	*Female*	71	0	0	Yes	Govt_job	*Rural*	193.94	22.4	*Smokes*	1
22	13861	*Female*	52	1	0	Yes	Self-employed	*Urban*	233.29	48.9	*Never Smoked*	1
23	68794	*Female*	79	0	0	Yes	Self-employed	*Urban*	228.7	26.6	*Never Smoked*	1
24	64778	*Male*	82	0	1	Yes	Private	*Rural*	208.3	32.5	Unknown	1
25	4219	*Male*	71	0	0	Yes	Private	*Urban*	102.87	27.2	*Formerly Smoked*	1
26	70822	*Male*	80	0	0	Yes	Self-employed	*Rural*	104.12	23.5	*Never Smoked*	1
27	38047	*Female*	65	0	0	Yes	Private	*Rural*	100.98	28.2	*Formerly Smoked*	1
28	61843	*Male*	58	0	0	Yes	Private	*Rural*	189.84	N/A	Unknown	1
29	54827	*Male*	69	0	1	Yes	Self-employed	*Urban*	195.23	28.3	*Smokes*	1
30	69160	*Male*	59	0	0	Yes	Private	*Rural*	211.78	N/A	*Formerly Smoked*	1
31	43717	*Male*	57	1	0	Yes	Private	*Urban*	212.08	44.2	*Smokes*	1
32	33879	*Male*	42	0	0	Yes	Private	*Rural*	83.41	25.4	Unknown	1
…	…	…	…	…	…	…	…	…	…	…	…	…
5110	44679	*Female*	44	0	0	Yes	Govt_job	*Urban*	85.28	26.2	Unknown	0

Lampiran 2. Data Setelah Proses *Cleaning*

No	id	Gender	age	Hyper tension	Heart disease	Ever married	Work type	Residence type	Avg Glucose level	BMI	Smoking status	Stroke
1	9046	*Male*	67	0	1	Yes	Private	*Urban*	228.69	36.6	*Formerly Smoked*	1
2	31112	*Male*	80	0	1	Yes	Private	*Rural*	105.92	32.5	*Never Smoked*	1
3	60182	*Female*	49	0	0	Yes	Private	*Urban*	171.23	34.4	*Smokes*	1
4	1665	*Female*	79	1	0	Yes	Self-employed	*Rural*	174.12	24	*Never Smoked*	1
5	56669	*Male*	81	0	0	Yes	Private	*Urban*	186.21	29	*Formerly Smoked*	1
6	53882	*Male*	74	1	1	Yes	Private	*Rural*	70.09	27.4	*Never Smoked*	1
7	10434	*Female*	69	0	0	No	Private	*Urban*	94.39	22.8	*Never Smoked*	1
8	60491	*Female*	78	0	0	Yes	Private	*Urban*	58.57	24.2	*Never Smoked*	1
9	12109	*Female*	81	1	0	Yes	Private	*Rural*	80.43	29.7	*Never Smoked*	1
10	12095	*Female*	61	0	1	Yes	Govt_job	*Rural*	120.46	36.8	*Smokes*	1
11	12175	*Female*	54	0	0	Yes	Private	*Urban*	104.51	27.3	*Smokes*	1
12	5317	*Female*	79	0	1	Yes	Private	*Urban*	214.09	28.2	*Never Smoked*	1
13	58202	*Female*	50	1	0	Yes	Self-employed	*Rural*	167.41	30.9	*Never Smoked*	1
14	56112	*Male*	64	0	1	Yes	Private	*Urban*	191.61	37.5	*Smokes*	1
15	34120	*Male*	75	1	0	Yes	Private	*Urban*	221.29	25.8	*Smokes*	1
16	27458	*Female*	60	0	0	No	Private	*Urban*	89.22	37.8	*Never Smoked*	1
17	70630	*Female*	71	0	0	Yes	Govt_job	*Rural*	193.94	22.4	*Smokes*	1
18	13861	*Female*	52	1	0	Yes	Self-employed	*Urban*	233.29	48.9	*Never Smoked*	1
19	68794	*Female*	79	0	0	Yes	Self-employed	*Urban*	228.7	26.6	*Never Smoked*	1
20	64778	*Male*	82	0	1	Yes	Private	*Rural*	208.3	32.5	*Never Smoked*	1
21	4219	*Male*	71	0	0	Yes	Private	*Urban*	102.87	27.2	*Formerly Smoked*	1
22	70822	*Male*	80	0	0	Yes	Self-employed	*Rural*	104.12	23.5	*Never Smoked*	1
23	38047	*Female*	65	0	0	Yes	Private	*Rural*	100.98	28.2	*Formerly Smoked*	1
24	54827	*Male*	69	0	1	Yes	Self-employed	*Urban*	195.23	28.3	*Smokes*	1
25	43717	*Male*	57	1	0	Yes	Private	*Urban*	212.08	44.2	*Smokes*	1
26	33879	*Male*	42	0	0	Yes	Private	*Rural*	83.41	25.4	*Never Smoked*	1
27	39373	*Female*	82	1	0	Yes	Self-employed	*Urban*	196.92	22.2	*Never Smoked*	1
28	54401	*Male*	80	0	1	Yes	Self-employed	*Urban*	252.72	30.5	*Formerly Smoked*	1
29	14248	*Male*	48	0	0	No	Govt_job	*Urban*	84.2	29.7	*Never Smoked*	1
30	712	*Female*	82	1	1	No	Private	*Rural*	84.03	26.5	*Formerly Smoked*	1
31	47269	*Male*	74	0	0	Yes	Private	*Rural*	219.72	33.7	*Formerly Smoked*	1
32	24977	*Female*	72	1	0	Yes	Private	*Rural*	74.63	23.1	*Formerly Smoked*	1
…	…	…	…	…	…	…	…	…	…	…	…	…
360	44679	*Male*	3	0	0	Yes	Govt_job	*Urban*	97.86	17.5	*Smokes*	0

Lampiran 3. Data Setelah Proses *Feature Selection*

No	Gender	age	Hyper tension	Heart disease	Residence type	Avg Glucose level	BMI	Smoking status	Stroke
1	*Male*	67	0	1	*Urban*	228.69	36.6	*Formerly Smoked*	1
2	*Male*	80	0	1	*Rural*	105.92	32.5	*Never Smoked*	1
3	*Female*	49	0	0	*Urban*	171.23	34.4	*Smokes*	1
4	*Female*	79	1	0	*Rural*	174.12	24	*Never Smoked*	1
5	*Male*	81	0	0	*Urban*	186.21	29	*Formerly Smoked*	1
6	*Male*	74	1	1	*Rural*	70.09	27.4	*Never Smoked*	1
7	*Female*	69	0	0	*Urban*	94.39	22.8	*Never Smoked*	1
8	*Female*	78	0	0	*Urban*	58.57	24.2	*Never Smoked*	1
9	*Female*	81	1	0	*Rural*	80.43	29.7	*Never Smoked*	1
10	*Female*	61	0	1	*Rural*	120.46	36.8	*Smokes*	1
11	*Female*	54	0	0	*Urban*	104.51	27.3	*Smokes*	1
12	*Female*	79	0	1	*Urban*	214.09	28.2	*Never Smoked*	1
13	*Female*	50	1	0	*Rural*	167.41	30.9	*Never Smoked*	1
14	*Male*	64	0	1	*Urban*	191.61	37.5	*Smokes*	1
15	*Male*	75	1	0	*Urban*	221.29	25.8	*Smokes*	1
16	*Female*	60	0	0	*Urban*	89.22	37.8	*Never Smoked*	1
17	*Female*	71	0	0	*Rural*	193.94	22.4	*Smokes*	1
18	*Female*	52	1	0	*Urban*	233.29	48.9	*Never Smoked*	1
19	*Female*	79	0	0	*Urban*	228.7	26.6	*Never Smoked*	1
20	*Male*	82	0	1	*Rural*	208.3	32.5	*Never Smoked*	1
21	*Male*	71	0	0	*Urban*	102.87	27.2	*Formerly Smoked*	1
22	*Male*	80	0	0	*Rural*	104.12	23.5	*Never Smoked*	1
23	*Female*	65	0	0	*Rural*	100.98	28.2	*Formerly Smoked*	1
24	*Male*	69	0	1	*Urban*	195.23	28.3	*Smokes*	1
25	*Male*	57	1	0	*Urban*	212.08	44.2	*Smokes*	1
26	*Male*	42	0	0	*Rural*	83.41	25.4	*Never Smoked*	1
…	…	…	…	…	…	…	…	…	…
360	*Male*	3	0	0	*Urban*	97.86	17.5	Smoked	0

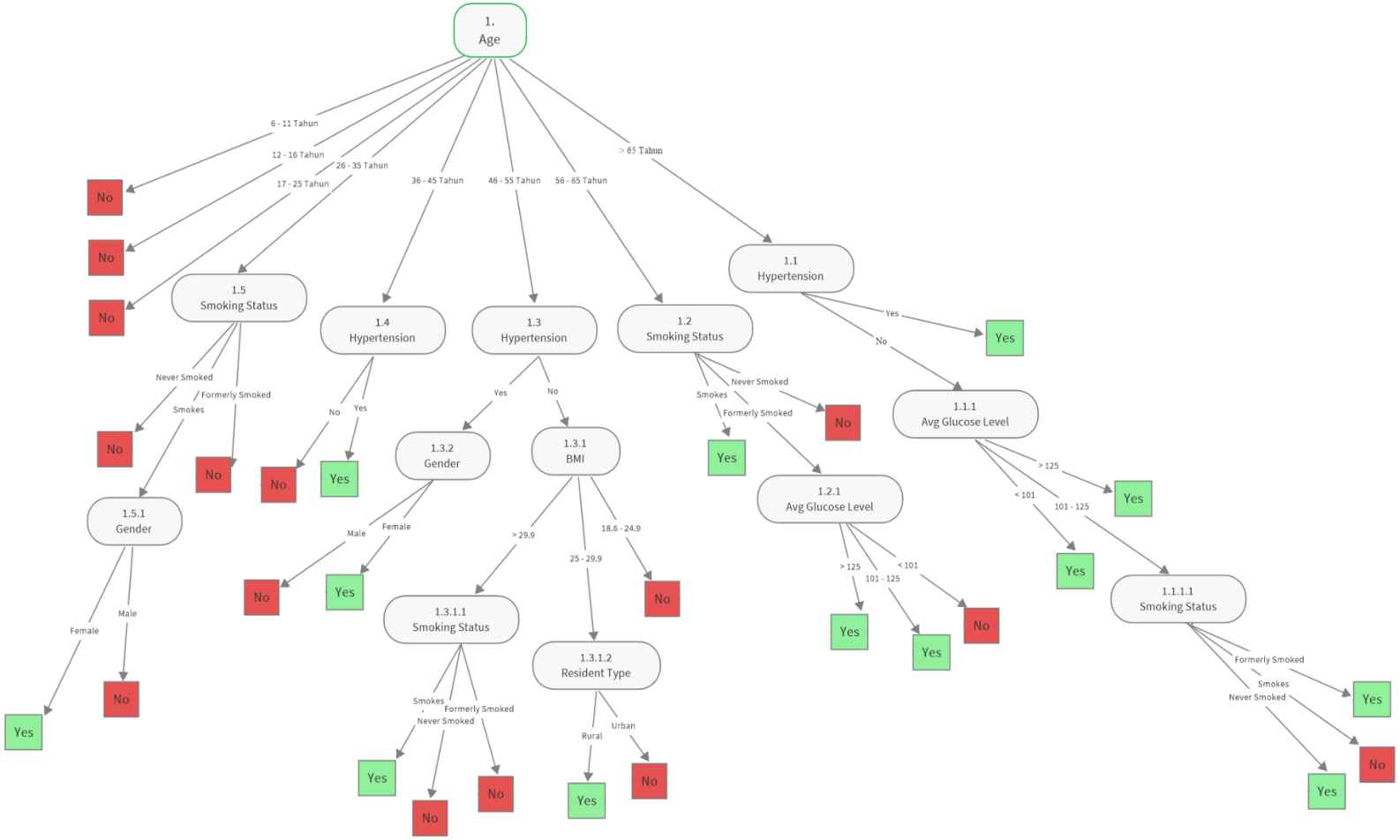
Lampiran 4.a Data Setelah Proses *Transformation* 1

No	*Gender*	*age*	*Hyper tension*	*Heart disease*	*Residence type*	*Avg Glucose level*	BMI	*Smoking status*	*Stroke*
1	*Male*	67	No	Yes	*Urban*	228.69	36.6	*Formerly Smoked*	Yes
2	*Male*	80	No	Yes	*Rural*	105.92	32.5	*Never Smoked*	Yes
3	*Female*	49	No	No	*Urban*	171.23	34.4	*Smokes*	Yes
4	*Female*	79	Yes	No	*Rural*	174.12	24	*Never Smoked*	Yes
5	*Male*	81	No	No	*Urban*	186.21	29	*Formerly Smoked*	Yes
6	*Male*	74	Yes	Yes	*Rural*	70.09	27.4	*Never Smoked*	Yes
7	*Female*	69	No	No	*Urban*	94.39	22.8	*Never Smoked*	Yes
8	*Female*	78	No	No	*Urban*	58.57	24.2	*Never Smoked*	Yes
9	*Female*	81	Yes	No	*Rural*	80.43	29.7	*Never Smoked*	Yes
10	*Female*	61	No	Yes	*Rural*	120.46	36.8	*Smokes*	Yes
11	*Female*	54	No	No	*Urban*	104.51	27.3	*Smokes*	Yes
12	*Female*	79	No	Yes	*Urban*	214.09	28.2	*Never Smoked*	Yes
13	*Female*	50	Yes	No	*Rural*	167.41	30.9	*Never Smoked*	Yes
14	*Male*	64	No	Yes	*Urban*	191.61	37.5	*Smokes*	Yes
15	*Male*	75	Yes	No	*Urban*	221.29	25.8	*Smokes*	Yes
16	*Female*	60	No	No	*Urban*	89.22	37.8	*Never Smoked*	Yes
17	*Female*	71	No	No	*Rural*	193.94	22.4	*Smokes*	Yes
18	*Female*	52	Yes	No	*Urban*	233.29	48.9	*Never Smoked*	Yes
19	*Female*	79	No	No	*Urban*	228.7	26.6	*Never Smoked*	Yes
20	*Male*	82	No	Yes	*Rural*	208.3	32.5	*Never Smoked*	Yes
21	*Male*	71	No	No	*Urban*	102.87	27.2	*Formerly Smoked*	Yes
22	*Male*	80	No	No	*Rural*	104.12	23.5	*Never Smoked*	Yes
23	*Female*	65	No	No	*Rural*	100.98	28.2	*Formerly Smoked*	Yes
24	*Male*	69	No	Yes	*Urban*	195.23	28.3	*Smokes*	Yes
25	*Male*	57	Yes	No	*Urban*	212.08	44.2	*Smokes*	Yes
26	*Male*	42	No	No	*Rural*	83.41	25.4	*Never Smoked*	Yes
…	…	…	…	…	…	…	…	…	…
360	*Male*	3	No	No	*Urban*	97.86	17.5	*Smokes*	No

Lampiran 4.b Data Setelah Proses *Transformation* 2

No	*Gender*	*Age*	*Hyper tension*	*Heart disease*	*Residence type*	*Avg Glucose level*	*BMI*	*Smoking status*	*Stroke*
1	*Male*	Lebih dari65th	No	Yes	*Urban*	Lebih dari125	Lebih dari 29.9	*Formerly Smoked*	Yes
2	*Male*	Lebih dari65th	No	Yes	*Rural*	101-125	Lebih dari 29.9	*Never Smoked*	Yes
3	*Female*	46-55th	No	No	*Urban*	Lebih dari 125	Lebih dari 29.9	*Smokes*	Yes
4	*Female*	Lebih dari65th	Yes	No	*Rural*	Lebih dari 125	18.6-24.9	*Never Smoked*	Yes
5	*Male*	Lebih dari65th	No	No	*Urban*	Lebih dari 125	25-29.9	*Formerly Smoked*	Yes
6	*Male*	Lebih dari65th	Yes	Yes	*Rural*	Kurang dari 101	25-29.9	*Never Smoked*	Yes
7	*Female*	Lebih dari65th	No	No	*Urban*	Kurang dari 101	18.6-24.9	*Never Smoked*	Yes
8	*Female*	Lebih dari65th	No	No	*Urban*	Kurang dari 101	18.6-24.9	*Never Smoked*	Yes
9	*Female*	Lebih dari65th	Yes	No	*Rural*	Kurang dari 101	25-29.9	*Never Smoked*	Yes
10	*Female*	56-65th	No	Yes	*Rural*	101-125	Lebih dari 29.9	*Smokes*	Yes
11	*Female*	46-55th	No	No	*Urban*	101-125	25-29.9	*Smokes*	Yes
12	*Female*	Lebih dari65th	No	Yes	*Urban*	Lebih dari 125	25-29.9	*Never Smoked*	Yes
13	*Female*	46-55th	Yes	No	*Rural*	Lebih dari 125	Lebih dari 29.9	*Never Smoked*	Yes
14	*Male*	56-65th	No	Yes	*Urban*	Lebih dari 125	Lebih dari 29.9	*Smokes*	Yes
15	*Male*	Lebih dari65th	Yes	No	*Urban*	Lebih dari 125	25-29.9	*Smokes*	Yes
16	*Female*	56-65th	No	No	*Urban*	Kurang dari 101	Lebih dari 29.9	*Never Smoked*	Yes
17	*Female*	Lebih dari65th	No	No	*Rural*	Lebih dari 125	18.6-24.9	*Smokes*	Yes
…	…	…	…	…	…	…	…	…	…
360	*Male*	0-5th	No	No	*Urban*	Kurang dari 101	Lebih dari 18.6	*smokes*	No

Lampiran 5: Model *Tree*



RIWAYAT HIDUP

Bagas Harmadi adalah nama penulis skripsi ini. Penulis lahir dari orang tua **Desharfius** dan **Syafneli** sebagai anak kedua dari tiga bersaudara. Penulis lahir di Painan, Kecamatan IV Jurai, Kabupaten Pesisir Selatan, Sumatera Barat pada tanggal 25 April 1999. Penulis pertama kali menempuh Pendidikan di Sekolah Dasar (SD) pada SDN 21 Bunga Pasang (*lulus tahun 2011*), dan pada Tahun yang sama Penulis melanjutkan Pendidikan ke Sekolah Menengah Pertama (SMP) di MTsN Salido (*lulus tahun 2014*), dan kemudian Pada Tahun yang sama Penulis melanjutkan Pendidikan di Sekolah Menengah Atas (SMA) pada MAN Salido dengan mengambil Jurusan IPA dan selesai Pada Tahun 2017. Pada tahun 2018 Penulis melanjutkan Pendidikan di salah satu Perguruan Islam Negeri Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Berkat petunjuk dan pertolongan Allah SWT, usaha dan disertai doa dari kedua orang tua dalam menjalani aktivitas akademik di Perguruan Tinggi Universitas Islam Negeri Maulana Malik Ibrahim Malang Penulis dapat menyelesaikan tugas akhir dengan Skripsi yang berjudul "PENERAPAN ALGORITMA *DECISION TREE* PADA PREDIKSI RISIKO TERSERANG *STROKE*".

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Bagas Harmadi
NIM : 18610113
Fakultas / Program Studi : Sains dan Teknologi / Matematika
Judul Skripsi : Penerapan Algoritma *Decision Tree* Pada Prediksi Risiko Terserang *Stroke*
Pembimbing I : Hisyam Fahmi, M.Kom
Pembimbing II : Erna Herawati, M.Pd

No	Tanggal	Hal	Tanda Tangan
1.	25 Februari 2022	Konsultasi BAB I	1.
2.	04 Maret 2022	Konsultasi Revisi BAB I	2.
3.	17 Maret 2022	Konsultasi BAB II dan III	3.
4.	25 Maret 2022	Konsultasi Revisi BAB II dan III	4.
5.	11 April 2022	Konsultasi Kajian Agama	5.
6.	14 April 2022	Konsultasi Revis Kajian Agama	6.
7.	20 April 2022	ACC Seminar Proposal	7.
8.	23 September 2022	Konsultasi BAB IV dan V	8.
9.	14 Oktober 2022	Konsultasi Revisi BAB IV dan V	9.
10.	27 Oktober 2022	Konsultasi Kajian Agama	10.
11.	01 November 2022	Konsultasi Revisi Kajian Agama dan TTD Untuk Seminar Hasil	11.
12.	02 November 2022	ACC Seminar Hasil	12.
13.	16 November 2022	Konsultasi Revisi Seminar Hasil	13.
14.	21 November 2022	Konsultasi Kajian Agama dan Abstrak Arab	14.
15.	02 Desember 2022	ACC Sidang Skripsi	15.
16.	21 Desember 2022	ACC Skripsi untuk Syarat Yudisium	16.

Malang, 21 Desember 2022
Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005