**PENERAPAN ALGORITMA DECISION TREE UNTUK PREDIKSI PENYAKIT GINJAL BERDASARKAN DATA MINING**

Diajukan untuk memenuhi Tugas Akhir Semester   
Mata Kuliah Data Mining



Disusun Oleh :

Bryant Fawwaz Bernhard (A11.2023.15114)

Krisna Nindya Kambara (A11.2023.15109)

Ilham Gozali (A11.2023.15105)

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DIAN NUSWANTORO**

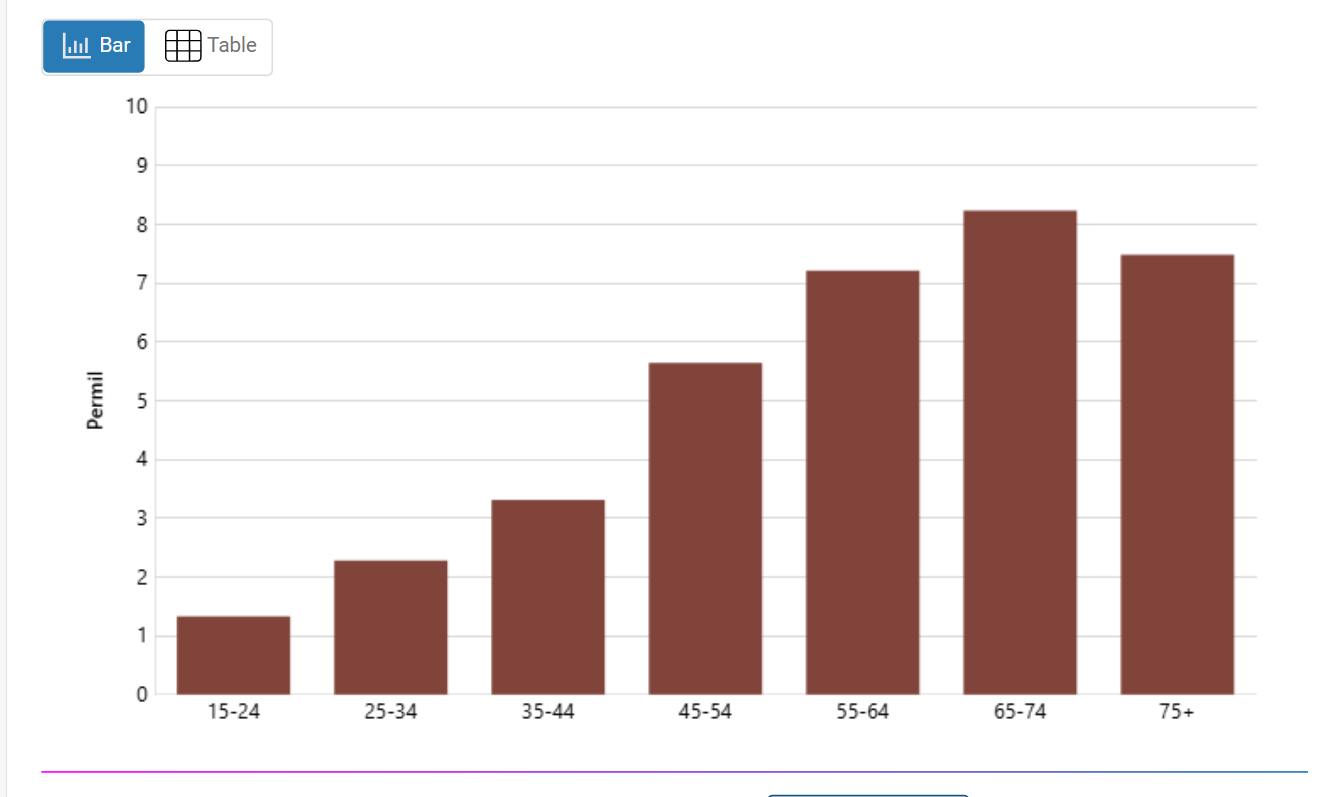
**SEMARANG**

**2024**

**BAB 1  
PENDAHULUAN**

**1.1. Latar Belakang**

Tingkat Penyakit Ginjal Kronis (PGK) dapat menimbulkan keresahan di masyarakat Indonesia, hal itu dibuktikan oleh Data Riset Kesehatan Dasar (Riskesdas) pada 2018 menunjukkan bahwa lebih dari 700 ribu orang, atau sekitar 3,8% penduduk Indonesia yang di indikasi bahwa penyakit yang akan terus meningkat di Indonesia. Prevalensi tertinggi terdapat pada kelompok usia 65-74 tahun (8,23%), diikuti oleh usia 75 tahun ke atas (7,48%) dan 55-64 tahun (7,21%). Peningkatan jumlah penderita PGK berdampak signifikan terhadap kebutuhan pelayanan kesehatan terutama prosedur cuci darah (hemodialisis) yang diselenggarakan oleh Badan Penyelenggara Jaminan Sosial(BPJS) mencatat bahwa pada tahun 2023, jumlah pasien gagal ginjal secara nasional mencapai 1 juta lebih kasus. Selain itu, pada tahun 2021, penyakit gagal ginjal menjadi diagnosis sekunder tertinggi kedua.

  
Gambar 1. Prevalensi penyakit Ginjal Kronis berdasarkan Diagnosis Dokter (Usia>15)

Deteksi dini PGK sangat penting untuk mencegah komplikasi lebih lanjut dan meningkatkan kualitas hidup pasien. Dalam konteks ini, penerapan teknik data mining, khususnya algoritma Decision Tree, menjadi relevan. Algoritma Decision Tree mampu mengolah data medis yang kompleks dan menghasilkan model prediksi yang mudah diinterpretasikan, sehingga dapat membantu tenaga medis dalam proses diagnosis. Penelitian sebelumnya() menunjukkan bahwa model prediksi PGK menggunakan algoritma Decision Tree mencapai tingkat akurasi hingga 98,12%, dengan nilai presisi 65,34% dan recall 65,88%. Hasil ini menunjukkan potensi algoritma Decision Tree dalam meningkatkan akurasi deteksi dini PGK. Dengan demikian, pengembangan aplikasi prediksi PGK berbasis algoritma Decision Tree diharapkan dapat menjadi alat bantu yang efektif bagi tenaga medis dalam melakukan screening awal dan diagnosis PGK secara lebih efisien, serta berkontribusi dalam upaya penanggulangan peningkatan prevalensi PGK di Indonesia.

Data mining merupakan salah satu cabang ilmu pengetahuan yang berkembang pesat di era big data. Menurut (Syahara, et al., 2021), data mining adalah proses pengumpulan informasi penting dari suatu dataset besar atau big data. Data mining berperan sebagai solusi untuk membantu berbagai sektor, termasuk dunia medis, dalam menganalisis data untuk mendukung pengambilan keputusan strategis. Metode-metode dalam data mining, seperti estimasi, prediksi, klasifikasi, clustering, dan asosiasi, memungkinkan pengolahan data mentah menjadi informasi yang berguna. Penyakit ginjal merupakan salah satu masalah kesehatan global yang memerlukan perhatian khusus. Diagnosa dini dan prediksi penyakit ini menjadi penting untuk memberikan intervensi medis yang cepat dan tepat. Dalam konteks ini, algoritma Decision Tree menawarkan pendekatan sistematis untuk menganalisis data medis dan membuat model prediktif. Algoritma ini mampu mengidentifikasi pola dari dataset yang besar dan mudah dipahami.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijabarkan, rumusan masalah dalam penelitian ini yaitu, menentukan pola klasifikasi dalam suatu *dataset* dengan menggunakan *Algoritma Decision Tree* guna memprediksi penyakit ginjal yang berguna kuat untuk mengidentifikasi pola dari *dataset* yang besar dan penerapan output model prediktif dapat memberikan tujuan yang memadai.

**1.3 Batasan Masalah** Dari pembuatan penelitian yang berjalan, adapun batasan-batasan penelitian yang akan dilakukan. Masalahnya sebagai berikut:  
 1. Penelitian ini menggunakan *datasets* dari luar objek penelitian dan juga *datasets* ini dalam kurun waktu 2018.

2.

**1.4 Tujuan Penelitian**

Dari latar belakang yang telah dijabarkan, Adapun tujuan dari penelitian ini yaitu mendapatkan penerapan Algoritma Decision Tree pada dataset medis untuk memprediksi penyakit ginjal dan mengevaluasi performa model prediktif yang dihasilkan berdasarkan metrik akurasi, presisi, dan *recall.*

**1.5 Manfaat Penelitian**  
  
 Dari penelitian ini, manfaat yang bisa didapatkan sebagai berikut:

**1.5.1 Bagi Penulis** Penelitian ini diharapkan dapat berkontribusi dalam memperkaya pemahaman mengenai pengolahan data melalui metode Algoritma Decision Tree. Selain itu, penelitian ini juga berperan sebagai wadah untuk menerapkan ilmu yang telah dipelajari selama masa perkuliahan, khususnya dalam bidang Data Mining.

**1.5.2 Bagi Tenaga Medis**  Penelitian ini diharapkan dapat memberikan wawasan baru dan mendalam mengenai penerapan algoritma Decision Tree dalam dunia medis. Algoritma ini dapat digunakan untuk mempermudah proses diagnosa awal, meningkatkan efisiensi dalam pengolahan data medis, dan membantu tenaga medis untuk mengambil keputusan berdasarkan analisis data yang valid dan terstruktur. Dengan demikian, penelitian ini dapat mendukung implementasi teknologi yang lebih canggih dalam upaya deteksi dini penyakit ginjal kronis.

**1.5.3 Bagi Pembaca** Penelitian ini dapat menjadi salah satu referensi yang komprehensif dalam bidang kajian data mining, khususnya terkait penerapan algoritma Decision Tree untuk prediksi penyakit ginjal. Selain itu, penelitian ini juga diharapkan dapat memberikan inspirasi bagi pengembangan sistem pendukung keputusan berbasis data mining yang lebih inovatif dan efektif, baik di dunia medis maupun sektor lainnya.

# **BAB II**

# **LANDASAN TEORI**

**2.1 Tinjauan Pustaka**  
 Dalam melakukan penelitian ini, penulis menggunakan berbagai studi sebelumnya yang memiliki keterkaitan dengan objek atau metode yang digunakan. Studi-studi tersebut berfungsi sebagai referensi dan acuan perbandingan untuk meningkatkan efisiensi dan kualitas hasil penelitian. Berikut adalah beberapa penelitian relevan yang mendasari langkah-langkah penelitian ini:

1. **Implementasi Decision Tree pada Data Medis:** Penelitian oleh (Aji Thofiq Madani, et al., 2020) membahas penerapan Decision Tree pada dataset penyakit kronis. Studi ini menjelaskan pentingnya preprocessing data untuk menghasilkan model prediktif yang lebih akurat.
2. **Evaluasi Algoritma Decision Tree dalam Prediksi Penyakit:** Studi (Dikan Ismafillah, et al., 2021) mengevaluasi performa algoritma Decision Tree berdasarkan metrik akurasi, presisi, dan recall. Penelitian ini menjadi acuan dalam mengevaluasi hasil model prediktif yang dikembangkan.

**2.1.1 Kajian Jurnal Pertama** “Diagnosis Dan Prediksi Penyakit Ginjal Kronis Dengan Menggunakan Pendekatan Stacked-Generalization” Merupakan Penelitian dalam konteks prediksi penyakit ginjal, algoritma Decision Tree juga telah diterapkan. Sebuah penelitian menunjukkan bahwa metode ini efektif untuk mengetahui jumlah penderita dan memprediksi risiko terkena penyakit ginjal kronis.  
  
**2.1.2 Kajian Jurnal Kedua** “Analisis algoritma pohon keputusan untuk memprediksi penyakit diabetes menggunakan *oversampling smote*” merupakan penelitian yang dilakukan pada sebuah *datasets* yang didapat dari situs data dunia kaggle yang berasal dari National Institute of Diabetes and Digestive and Kidney Diseases sebanyak 768 data yang terdiri dari 8 variabel dan 1 class target. analisis algoritma pohon keputusan digunakan untuk memprediksi penyakit ginjal, dengan menggunakan model Random Forest (RF) dan Decision Tree (DC) yang dikombinasikan dengan teknik SMOTE untuk meningkatkan akurasi prediksi.

# 

# **BAB III**

# **METODE PENELITIAN**

**3.1 Instrumen Penelitian**

Dalam penelitian ini ada beberapa instrumentasi peralatan yang mendukung dalam proses penelitian, diantaranya :

1. Bahan

Bahan yang dibutuhkan dalam melakukan penelitian ini adalah dataset Chronic Kidney Disease pada tahun 2015 pada dari datasets global di situs archive.ics.uci.edu.

1. Kebutuhan Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Sistem Operasi : Microsoft Windows 11 Versi 64 bit Penelitian ini menggunakan sistem Operasi Microsoft windows 11 untuk menjalankan program data mining.

2. Microsoft Excel Microsoft Excel merupakan tools yang digunakan untuk mengelola data-data transaksi yang akan diolah.

3. Google Documents Google Documents digunakan untuk membuat dan mengelola laporan Tugas Akhir.

4. Visual Studio Code Visual Studio Code digunakan untuk melakukan proses implementasi algoritma.

## **3.2** **Alur Penelitian**

Untuk mendapatkan hasil penelitian yang diinginkan, maka terdapat alur penelitian yang digunakan sebagai acuan setiap langkah-langkah yang diambil dalam penelitian. Alur penelitian yang digunakan yaitu sebagai berikut :

## 

## **3.3** **Objek Penelitian**

Pada penelitian ini, objek yang digunakan yaitu sekumpulan data *check up* di Rumah sakit di India. Data yang diambil untuk diproses dalam penelitian ini merupakan data *check up* periode 2016 . Data *check up* tersebut berjumlah 400 baris excel.

## **3.4. Metode Pengambilan dan Pengumpulan Data**

## **3.5. Jenis Data**

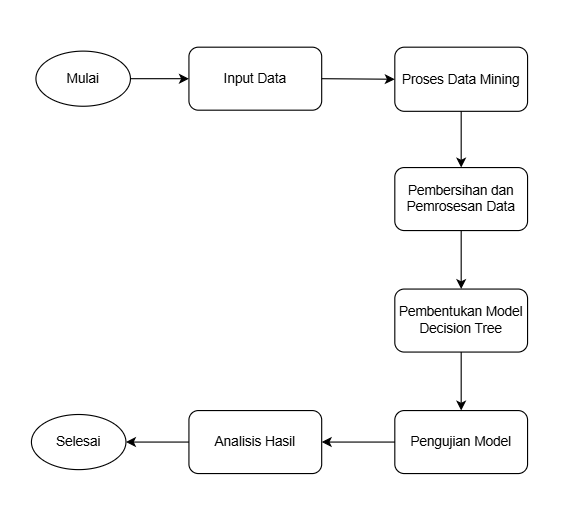
Jenis data yang digunakan dalam penelitian ini antara lain

1. Data Sekunder

Data ini didapatkan atau dikumpulkan oleh dari beberapa studi sebelumnya. Atau bisa dikatakan data yang diperoleh secara tidak langsung. Dalam penelitian ini diperoleh berdasarkan jurnal-jurnal dan buku guna mendapatkan landasan teori berupa pendapat-pendapat ataupun bahan lainnya yang penting untuk penelitian.

## **3.6. Pemodelan**

Penelitian ini akan menggunakan metode Algoritma *Decision Tree* sebagai pemodelan yang akan dilakukan untuk pengolahan data *check up* di rumah sakit. Dalam melakukan proses data mining, Teknik klasifikasi merupakan teknik data mining yang dipilih dalam penelitian ini. Pemodelan yang dilakukan bertujuan untuk mencari klasifikasi dari tiap data masukan. Sehingga dengan adanya hal tersebut dapat dijadikan bahan evaluasi dalam menentukan apakah pasien tersebut menderita *chronic kidney disease*. Untuk lebih jelasnya, Langkah-langkah pemodelan data mining menggunakan algoritma *Decision Tree* adalah sebagai berikut :



## **3.7. Pengujian**

Pelaksanaan pengujian pada penelitian ini dilakukan demi mengetahui hasil operasi yang dianalisa apakah memenuhi tujuan penelitian atau tidak. Pada penelitian ini menggunakan minimum  
 Pengujian pada penelitian dilakukan untuk mengetahui hasil perhitungan yang dianalisa apakah memenuhi kriteria atau tidak. Pada penelitian ini menggunakan minimum Support dan minimum Confidence sebagai alat tolak ukur serta rule yang diambil yaitu rule yang memiliki nilai tertinggi dari yang lainnya. Sehingga akan dilakukan proses pencarian nilai support dan nilai confidence pada masing masing item maupun pada kombinasi item. Support digunakan untuk mengukur nilai persentase gabungan antar item, sedangkan untuk Confidence digunakan untuk mengukur seberapa kuat hubungan antar item. jika nilai support dan nilai confidence tidak melebihi minimum support dan minimum confidence maka rule yang dihasilkan bukan digolongkan sebagai aturan asosiasi yang kuat. Terdapat juga nilai Lift, Conviction, dan Leverage yang dapat dijelaskan sebagai berikut.

· Nilai Lift merupakan nilai yang digunakan untuk menunjukan validasi dalam proses transaksi. Proses transaksi dinyatakan valid jika nilai lift yang dihasilkan menunjukan nilai lebih dari 1(satu).

· Nilai Conviction digunakan untuk mengukur tingkat implikasi dalam suatu aturan.

· Nilai Leverage yaitu nilai yang digunakan untuk mengukur tingkat keterkaitan antara item antecedent dan item consequence. Jika nilai leverage menunjukan nilai 0 maka dapat diartikan bahwa item antecedent dan item consequent tidak memiliki keterkaitan.

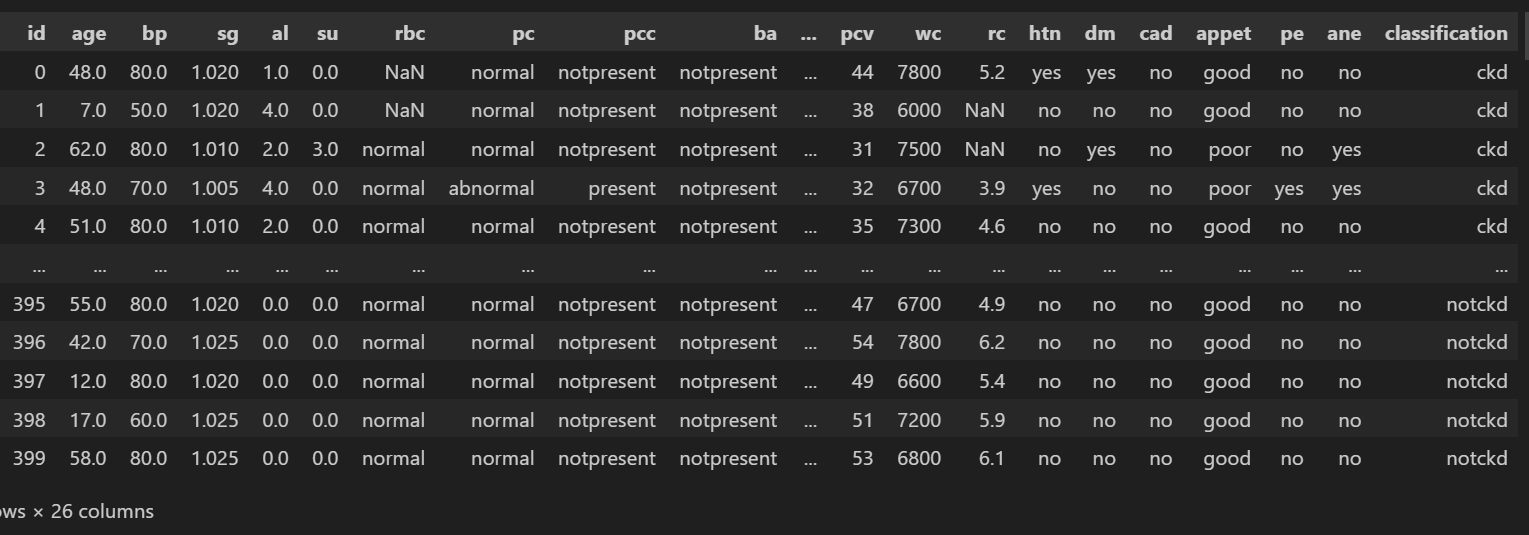
# 

# 

# **BAB IV**

# **HASIL DAN PEMBAHASAN**

## **4.1. Data Check Up**

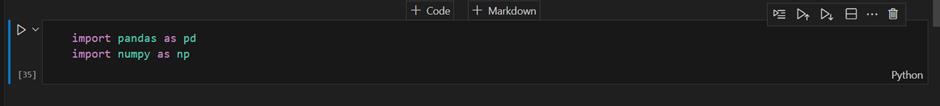


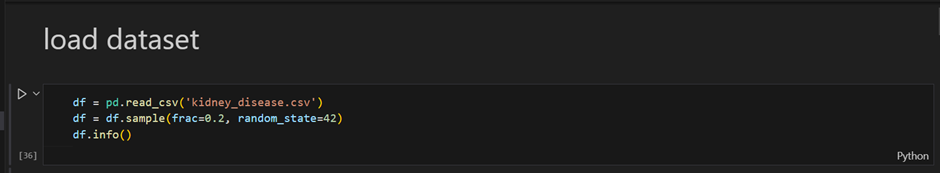
Gambar 4. 1 Data Penjualan

Pada gambar diatas, terdapat beberapa fitur seperti age, bp, sg, al, su, rbc hingga fitur ane yang menghasilkan *classification*. Setiap satu inputan seluruh fitur menghasilkan satu classification data sehingga jika terjadi 400 inputan maka terdapat 400 *classification* data pula. Dengan begitu data tersebut perlu dikelompokan menjadi satu file excel dataset yang siap diolah.

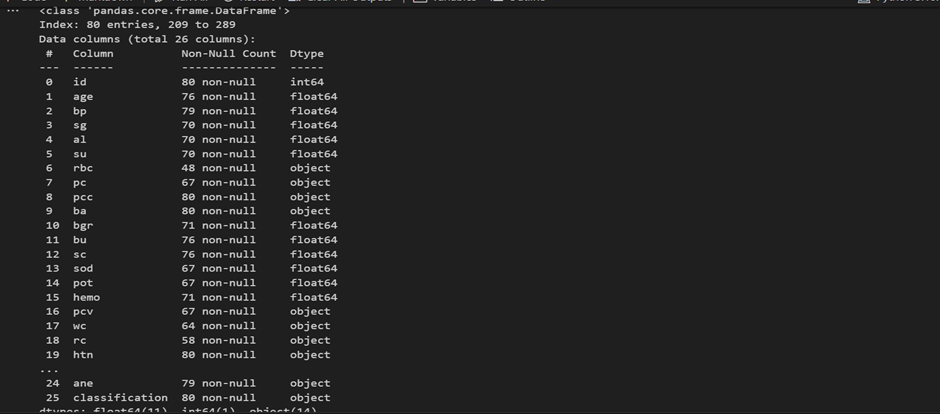
## **4.2. Implementasi**

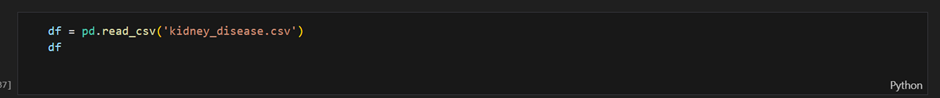
1. LOAD DATASET



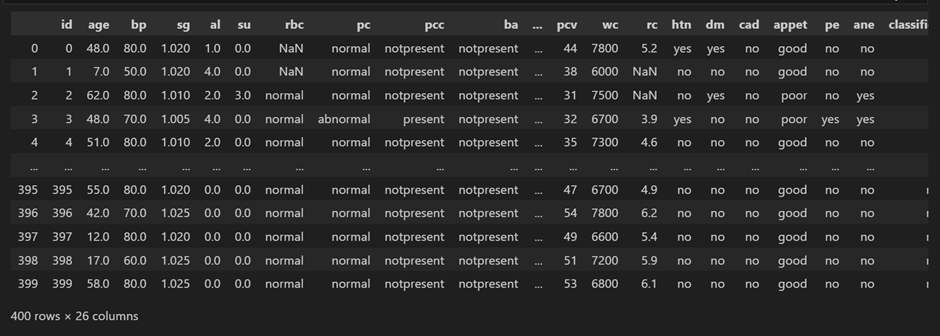


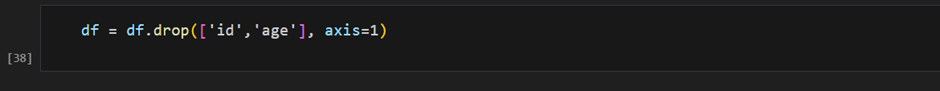
outputnya :

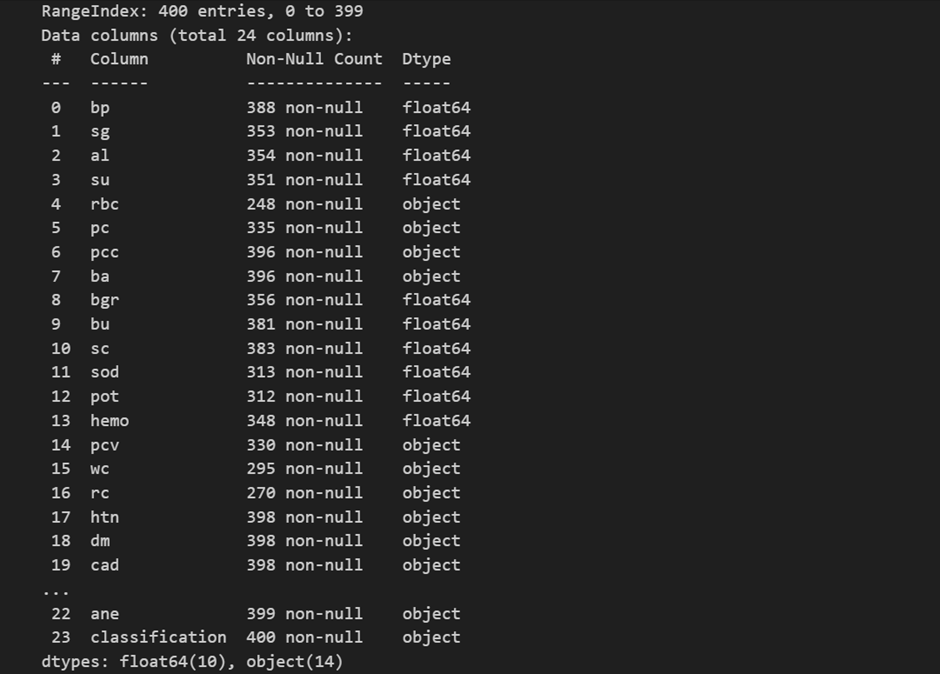


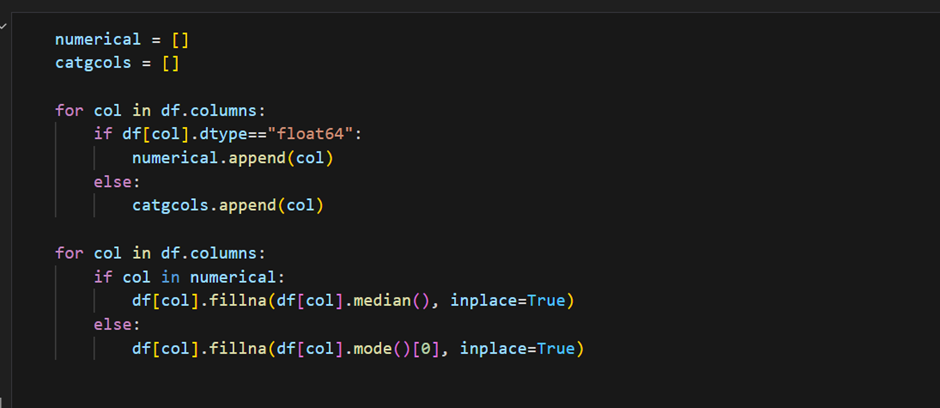


outputnya :

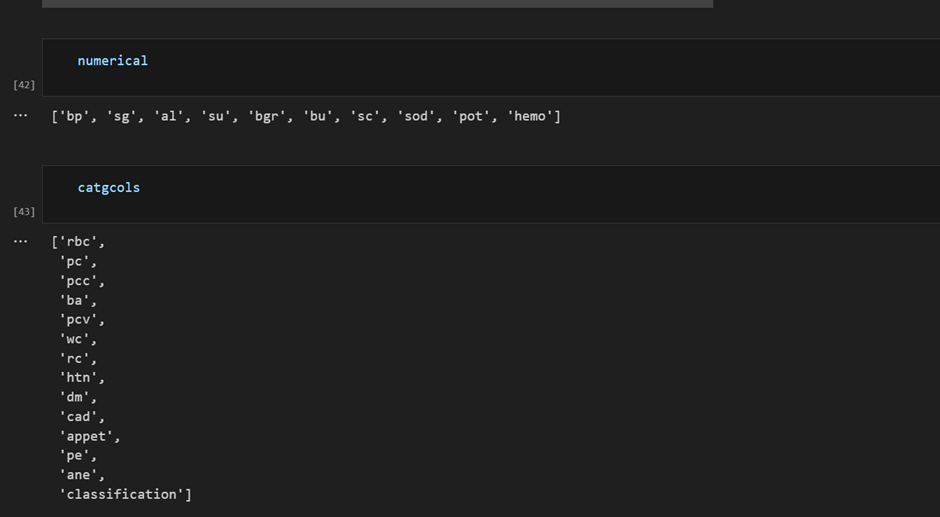


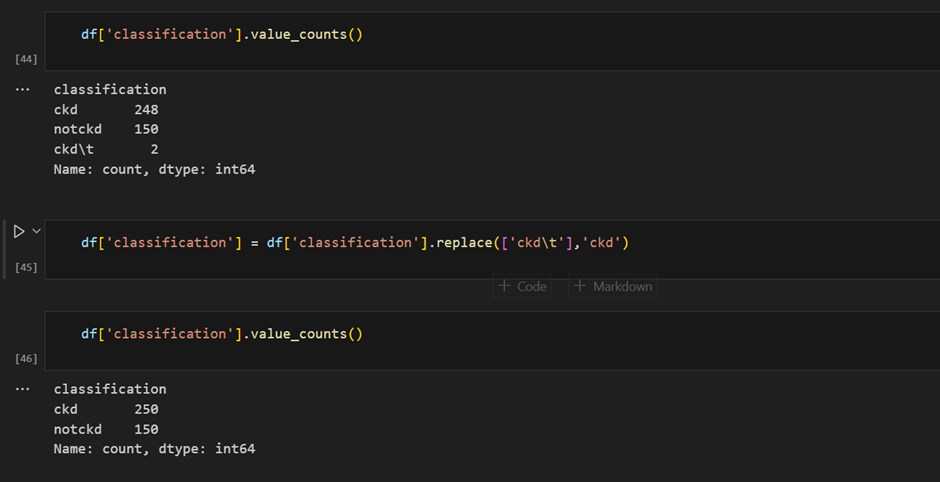


outputnya : 

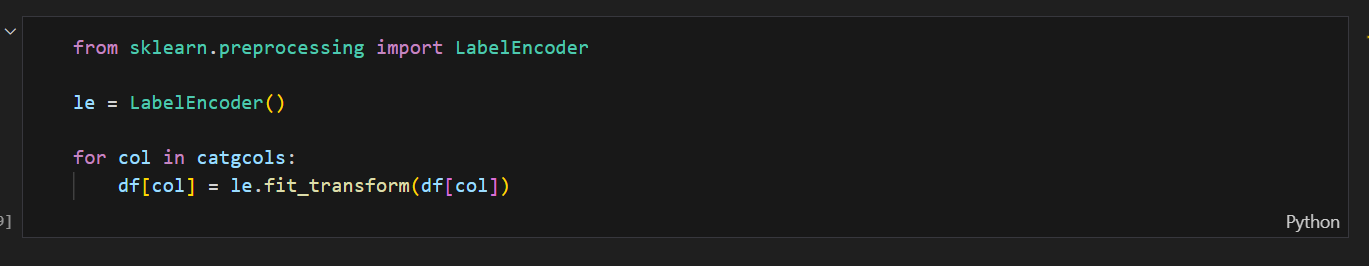


outputnya :

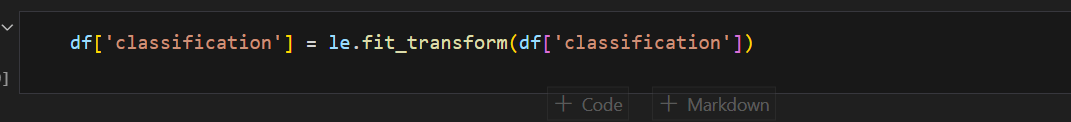




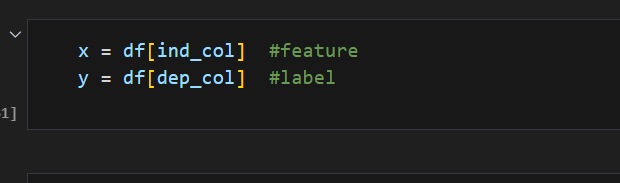
1. TRANSFORMASI DAN PREPROCCESING DATASET
2. s



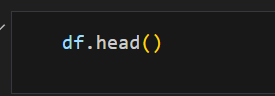
1. s



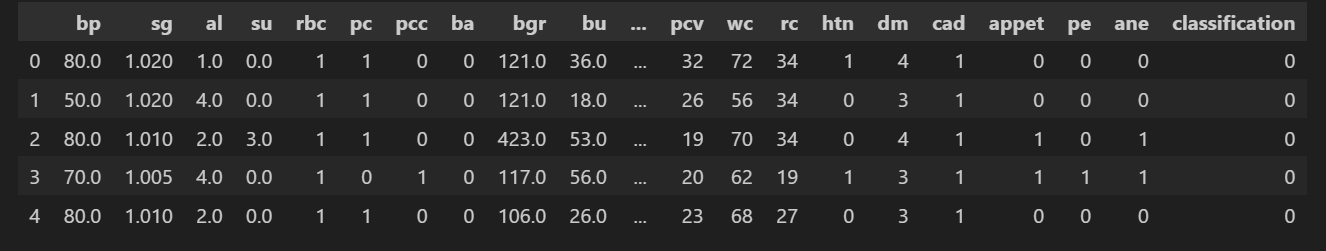
1. s



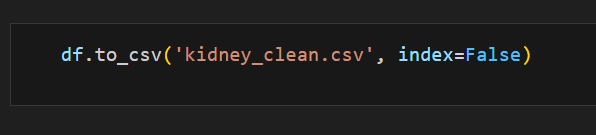
1. s



outputnya :



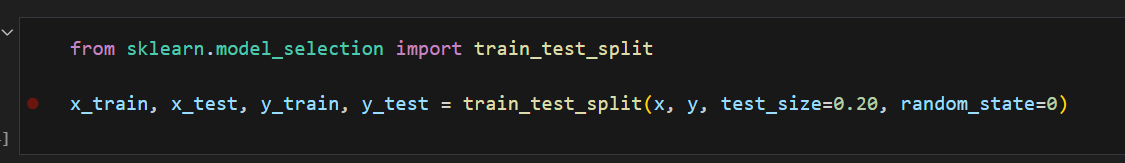
1. s



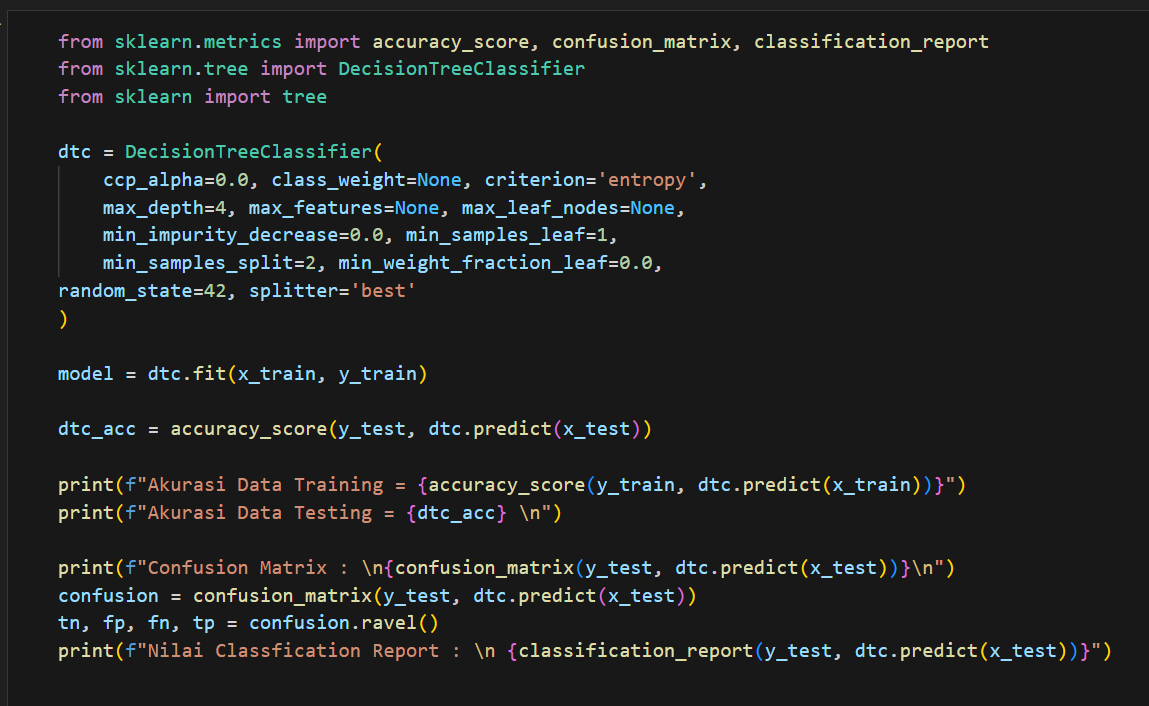
outputnya :



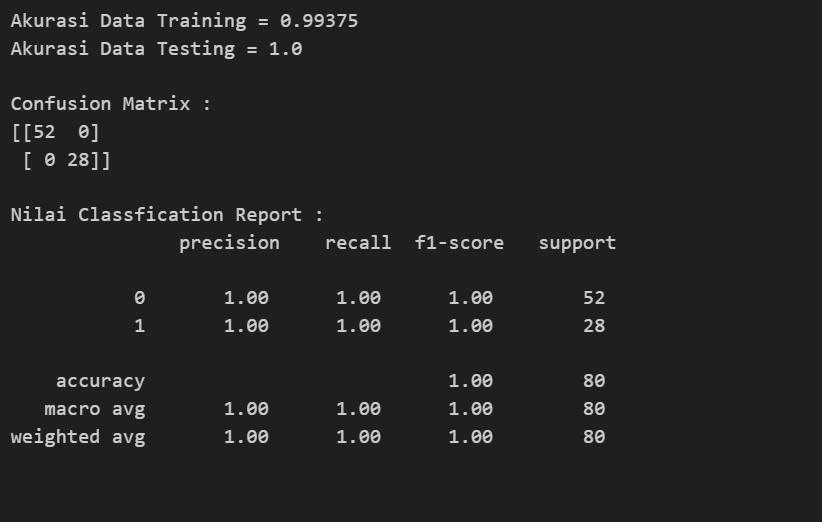
1. s



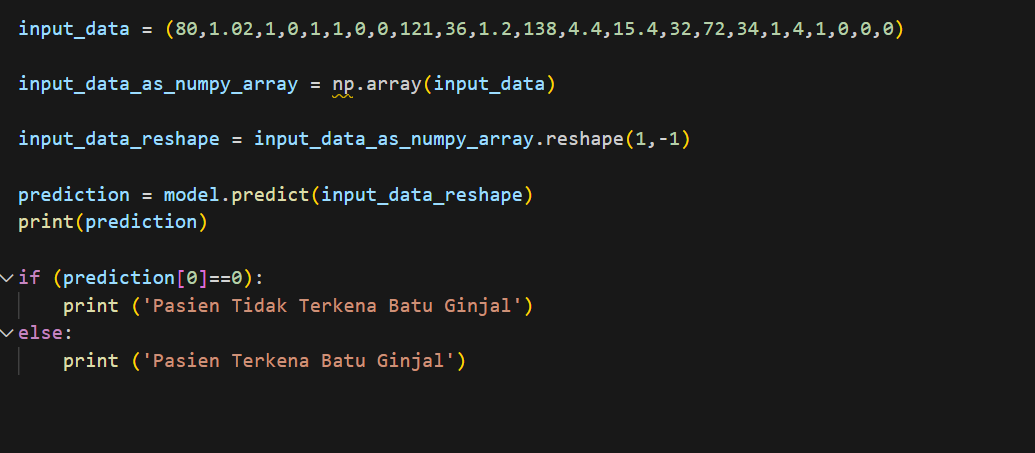
1. PEMBUATAN , PELATIHAN , DAN EVALUASI MODEL DENGAN DECISION TREE
2. s



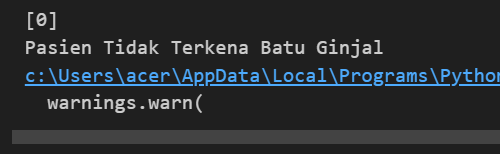
outputnya :



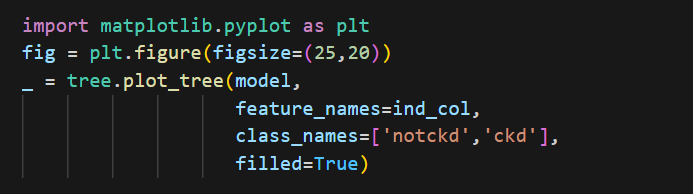
1. SIMULASI MODEL
2. d



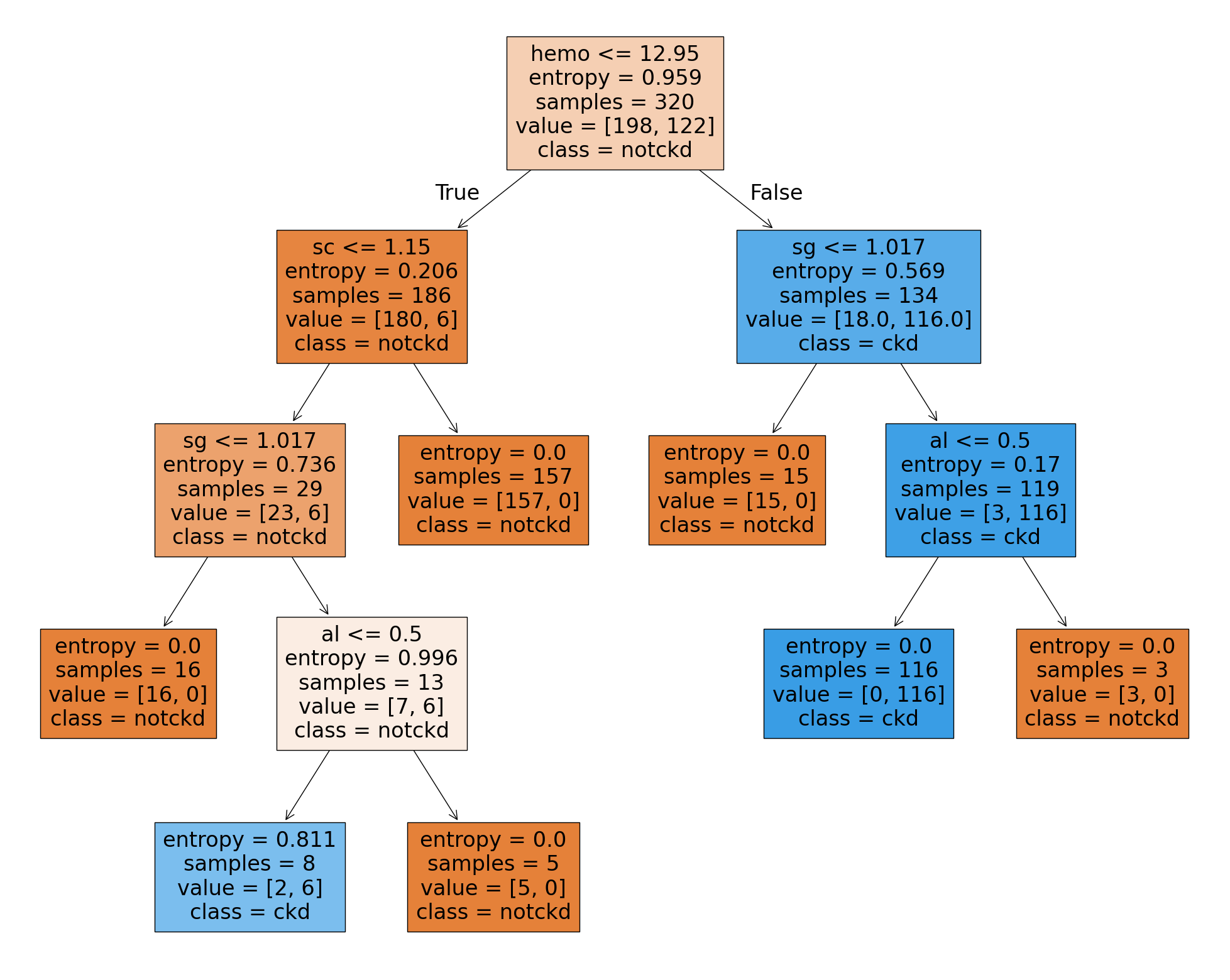
outputnya :



1. VISUALISASI MODEL DENGAN DECISION TREE
2. s



outputnya :



# **BAB V**

# **KESIMPULAN**

## **5.1 Kesimpulan**

## **5.2 Penelitian Selanjutnya**