

## **LAPORAN AKHIR**

### **“KLASIFIKASI NUTRITION-SCORE PADA MINUMAN BERDASARKAN KANDUNGAN GIZI”**



#### **KELOMPOK: SD-A1 - I (09):**

- |                              |             |
|------------------------------|-------------|
| 1. BRAHMANTYA FIKRI SETYA P. | (164221034) |
| 2. CINDASHANY PRIJANTIKA     | (164221057) |
| 3. ANDREAS HENDRA HERWANTO   | (164221064) |
| 4. MOCHAMMAD NAUFAL IZAH R.  | (164221087) |
| 5. AMIRA NAILA ZANIRA        | (164221100) |

**TEAM-BASED PROJECT  
MATA KULIAH DATA MINING I  
PROGRAM STUDI TEKNOLOGI SAINS DATA  
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN  
UNIVERSITAS AIRLANGGA  
2024**

## DAFTAR ISI

<b>BAB 1.....</b>	<b>6</b>
<b>PENDAHULUAN.....</b>	<b>6</b>
1.1 Latar Belakang.....	6
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	7
<b>BAB 2.....</b>	<b>8</b>
<b>TINJAUAN PUSTAKA.....</b>	<b>8</b>
2.1 Nutrition Score.....	8
2.1.1 Klasifikasi Nutrition-score.....	8
2.2 Decision Tree.....	9
2.3 Random Forest.....	9
<b>BAB 3.....</b>	<b>10</b>
<b>METODOLOGI PENELITIAN.....</b>	<b>10</b>
3.1 Pengumpulan Data.....	10
3.1.1 Teknik Pengambilan Sampel.....	10
3.2 Variabel Penelitian.....	10
3.3 Algoritma Machine Learning Klasifikasi.....	11
3.3.1 Algoritma Random Forest.....	11
3.3.2 Algoritma Decision Tree.....	12
3.4 Model Evaluation.....	13
3.4.1 Confusion Matrix.....	13
3.4.2 Metrics Evaluation.....	13
<b>BAB 4.....</b>	<b>15</b>
<b>HASIL DAN PEMBAHASAN.....</b>	<b>15</b>
4.1 Insight Exploratory Data Analysis (EDA).....	15
4.1.1 Data Shape.....	15
4.1.2 Data Description.....	15
4.1.3 Heatmap Korelasi Pearson.....	16
4.1.4 Barplot Variabel Y.....	16
4.1.5 Boxplot Variabel X.....	17
4.2 Pre Processing.....	20

4.2.1 Cleaning Data.....	20
4.2.2 Missing Value.....	21
4.2.3 Encoding Variabel.....	22
4.2.4 Handling Outlier.....	22
4.2.5 Scaling Data.....	23
4.3 Modeling dan Evaluation.....	23
4.3.1 Data Partition.....	23
4.3.2 Hyperparameter Tuning.....	23
4.3.3 Model Evaluation.....	24
4.3.3.1 Random Forest Model.....	24
4.3.3.2 Decision Tree Model.....	25
4.3.4 Feature Importance.....	26
4.3.4.1 Random Forest Model.....	26
4.3.4.2 Decision Tree Model.....	27
<b>BAB 5.....</b>	<b>28</b>
<b>KESIMPULAN DAN SARAN.....</b>	<b>28</b>
5.1 Kesimpulan.....	28
5.2 Saran.....	28
<b>DAFTAR PUSTAKA.....</b>	<b>29</b>
<b>LAMPIRAN.....</b>	<b>30</b>

## DAFTAR GAMBAR

Gambar 4.1 Data Description.....	14
Gambar 4.2 Heatmap Korelasi Pearson.....	15
Gambar 4.3 Barplot Variabel Y.....	15
Gambar 4.4 Boxplot Variabel Energy.....	16
Gambar 4.5 Boxplot Variabel Fat.....	16
Gambar 4.6 Boxplot Variabel Saturated Fat.....	16
Gambar 4.7 Boxplot Variabel Carbohydrates.....	17
Gambar 4.8 Boxplot Variabel Sugar.....	17
Gambar 4.9 Boxplot Variabel Fiber.....	17
Gambar 4.10 Boxplot Variabel Proteins.....	18
Gambar 4.11 Boxplot Variabel Salt.....	18
Gambar 4.12 Boxplot Variabel Fruits,vegetables and legumes.....	18
Gambar 4.13 Deteksi Simbol pada Setiap Kolom.....	19
Gambar 4.14 Cek Missing Value.....	20
Gambar 4.15 Encoding Variabel.....	21
Gambar 4.16. Scaling Data.....	22
Gambar 4.17 Confusion Matrix.....	23
Gambar 4.18 Confusion Matrix.....	24
Gambar 4.19 Feature Importance Decision Tree.....	25
Gambar 4.20 Feature Importance Random Forest.....	26

## **DAFTAR TABEL**

Tabel 3.1 Deskripsi Variabel.....	10
-----------------------------------	----

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan industri minuman yang pesat telah meningkatkan variasi produk minuman yang tersedia di pasaran. Banyaknya variasi minuman ini menawarkan pilihan yang beragam bagi konsumen, namun juga menimbulkan tantangan tersendiri dalam menentukan minuman yang sehat dan bergizi. Sehingga, untuk membantu konsumen memilih produk yang lebih sehat, diperlukannya sistem penilaian gizi seperti Nutri-Score.

Nutri-Score adalah sebuah label yang diberikan pada produk-produk makanan dan minuman berdasarkan kandungan gizinya. Penilaian gizi pada Nutri-Score memberikan label A (kualitas gizi terbaik) hingga E (kualitas gizi terburuk) berdasarkan kandungan energi, gula, lemak jenuh, garam, serat, dan protein dalam produk. Dengan adanya label ini, dapat mempermudah konsumen dalam memilih produk minuman yang sehat dan tidak sehat.

Penelitian ini bertujuan untuk mengembangkan sebuah model *machine learning* untuk membantu penerapan sistem Nutri-Score pada berbagai minuman di Indonesia berdasarkan informasi label kemasan. Analisis dilakukan dengan menggunakan model klasifikasi Random Forest dan Decision Tree untuk mengkategorikan minuman sesuai dengan skor Nutri-Score yang dihitung. Melalui metode klasifikasi ini, model akan dievaluasi, sehingga mampu mengkategorikan nutri-score produk dengan lebih akurat.

Hasil penelitian diharapkan memberikan wawasan lebih dalam tentang kualitas gizi minuman yang ada di pasaran dan mendorong penerapan label gizi yang lebih informatif. Selain itu, hasil ini juga dapat menjadi dasar bagi kebijakan kesehatan masyarakat untuk mengatur label gizi pada produk minuman, dengan tujuan meningkatkan kualitas kesehatan masyarakat secara keseluruhan.

### **1.2 Rumusan Masalah**

1. Bagaimana klasifikasi minuman berdasarkan Nutri-Score?
2. Bagaimana efektivitas metode klasifikasi dengan menggunakan model Random Forest dan Decision Tree dalam mengkategorikan minuman berdasarkan Nutri-Score?

### **1.3 Tujuan Penelitian**

- a) Tujuan Umum  
Mengembangkan sebuah model *machine learning* untuk membantu penerapan sistem Nutri-Score pada berbagai minuman di Indonesia berdasarkan informasi label kemasan
- b) Tujuan Khusus
  1. Mengklasifikasikan Minuman Berdasarkan Nutri-Score dengan menggunakan model Random Forest dan Decision Tree

2. Mengevaluasi model yang telah dibangun.

#### **1.4 Manfaat Penelitian**

1. Meningkatkan kesadaran konsumen tentang pentingnya memperhatikan kandungan gizi dalam minuman yang mereka konsumsi.
2. Temuan dari penelitian ini dapat menjadi dasar bagi pembuat kebijakan dalam merumuskan regulasi yang lebih ketat mengenai label gizi pada produk minuman.
3. Meningkatkan kesadaran masyarakat tentang pentingnya kandungan gizi.

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1 Nutrition Score**

Label informasi nilai gizi yang biasa dikenal sebagai *Nutrition Information* atau *Nutrition Fact* memuat daftar kandungan zat gizi pangan sesuai format yang telah ditetapkan (BPOM, 2018). Tabel informasi nilai gizi telah diatur pada Peraturan Nomor 22 Tahun 2019 oleh Badan Pengawasan Obat dan Makanan (BPOM) tentang Informasi Nilai Gizi pada label pangan olahan. Pada peraturan tersebut diberikan informasi mengenai jumlah sajian per kemasan, jenis dan jumlah kandungan zat gizi, takaran saji, jenis dan jumlah kandungan zat non gizi, dan persentase AKG (Angka Kecukupan Gizi).

*Nutrition Score* merupakan sistem penilaian dengan memberikan skor pada produk makanan dan minuman berdasarkan kandungan gizinya kemudian di klasifikasi berdasarkan abjad A,B,C,D, dan E. Metode penentuan *Nutrition Score* didasari oleh beberapa faktor komponen gizi seperti energi, lemak, lemak jenuh, karbohidrat, gula, serat, protein, dan garam. Informasi ini penting bagi konsumen agar dapat membuat keputusan yang lebih baik mengenai asupan nutrisi harian mereka.

##### **2.1.1 Klasifikasi Nutrition-score**

1. **Label A** : Label ini menunjukkan produk memiliki kualitas nutrisi tinggi. Produk ini rendah energi, lemak jenuh, gula, dan garam, tetapi kaya serat, protein, buah-buahan, sayuran, kacang-kacangan, dan biji-bijian.
2. **Label B** : Produk dengan klasifikasi ini juga memiliki kualitas nutrisi yang baik. Mereka mungkin memiliki sedikit lebih tinggi tingkat energi, lemak jenuh, gula, atau garam dibandingkan label A tetapi masih merupakan pilihan yang wajar untuk diet sehat.
3. **Label C** : Ini menunjukkan kualitas nutrisi sedang. Produk-produk ini mungkin memiliki tingkat energi, lemak jenuh, gula, dan garam yang lebih tinggi dari label A dan B. Meskipun masih bisa menjadi bagian dari diet seimbang jika dikonsumsi dengan bijak, disarankan untuk memprioritaskan produk dengan Nutri-score yang lebih tinggi.
4. **Label D** : Produk yang diklasifikasikan sebagai D memiliki kualitas nutrisi yang lebih rendah. Mereka mungkin memiliki tingkat energi, lemak jenuh, gula, dan garam yang lebih tinggi, membuat mereka pilihan yang kurang diinginkan untuk diet sehat. Konsumsi harus dibatasi.
5. **Label E** : Label ini menunjukkan kualitas nutrisi terendah. Produk-produk di label E memiliki tingkat energi, lemak jenuh, gula, dan garam yang tinggi, dan mungkin kurang memiliki komponen-komponen bermanfaat seperti serat, protein, buah-buahan, sayuran, kacang-kacangan, dan biji-bijian. Disarankan



untuk membatasi konsumsi produk-produk ini karena mereka tidak berkontribusi pada diet sehat.

## **2.2 Decision Tree**

Menurut Kusrini dan Luthfi (2017), Decision tree atau pohon keputusan merupakan metode klasifikasi yang sangat kuat dan populer. Decision tree dapat mengubah data yang sangat besar menjadi sebuah pohon keputusan yang mempresentasikan aturan. Decision tree merupakan teknik klasifikasi yang digunakan untuk melakukan klasifikasi terhadap sekumpulan objek atau data dengan representasi struktur pohon, dimana setiap node internal mewakili atribut, setiap cabang mewakili keputusan berdasarkan atribut, dan setiap daun node mewakili hasil akhir. Pohon keputusan umumnya digunakan untuk eksplorasi pola pada sekumpulan data serta menemukan hubungan antara sejumlah variabel input dengan variabel output.

## **2.3 Random Forest**

Menurut Breiman, L (2001) Random Forest adalah metode ensemble learning untuk klasifikasi dan regresi yang terdiri dari sejumlah besar pohon keputusan yang dilatih secara independen. Setiap pohon dalam hutan dibangun menggunakan subset acak dari data pelatihan, dan pemilihan fitur yang dipertimbangkan untuk setiap split juga dilakukan secara acak. Hasil akhir dari Random Forest ditentukan melalui agregasi (misalnya voting mayoritas untuk klasifikasi atau rata-rata untuk regresi) dari hasil masing-masing pohon individu.

Random forest merupakan alat yang efektif dalam memprediksi. Karena menerapkan prinsip Law of Large Numbers maka jarang terjadi overfitting pada model. Selain itu, kita juga dapat melihat kemampuan sebuah model Random Forest dalam memprediksi dari kekuatan masing-masing prediktor individu dan korelasinya. (Breiman 2001)

## **BAB 3**

### **METODOLOGI PENELITIAN**

#### **3.1 Pengumpulan Data**

Pengumpulan data merupakan tahapan penting dalam sebuah penelitian, dimana data yang relevan dikumpulkan untuk dianalisis lebih lanjut. Pada penelitian ini, kami menggunakan data primer yang diperoleh langsung dari objek tanpa melalui pihak ketiga.

##### **3.1.1 Teknik Pengambilan Sampel**

Penelitian ini menggunakan data yang diekstraksi melalui proses scrapping dari website Open Food Facts. Kami menggunakan script scraping pada Python untuk mengotomatisasi pengambilan data dari situs web yang telah diidentifikasi. Dengan menggabungkan API Selenium dan BeautifulSoup, kami dapat melakukan otomatisasi browser web, parsing dan ekstraksi data dari HTML. (Witten, Frank & Hall, 2011).

Dengan menggunakan teknik pengambilan sampel ini, kami dapat mengumpulkan dataset yang cukup besar dan beragam yaitu sebanyak 449 data. Hal ini dapat memungkinkan kami untuk melakukan analisis lebih mendalam untuk mengklasifikasikan score nutrisi pada minuman berdasarkan kandungan gizinya.

#### **3.2 Variabel Penelitian**

Untuk mengetahui klasifikasi dari Nutri-Score berdasarkan kandungan gizinya, kita perlu mengetahui variabel yang akan digunakan sebagai berikut:

Nama Variabel	Tipe Data	Deskripsi Variabel
Nama Minuman	Kategorik	Nama produk minuman yang diambil dari web.
Nutri Score	Kategorik	Skor nutrisi yang diberikan berdasarkan kandungan gizi.
Energy (Per 100 ml)	Numerik	Jumlah energi yang terkandung dalam produk ditunjukkan dengan jumlah kalori yang tersedia dari konsumsi produk.
Fat (Per 100 ml)	Numerik	Jumlah lemak total yang terkandung dalam produk.
Saturated Fat (Per 100 ml)	Numerik	Kandungan lemak jenuh dalam produk per 100 ml yang diukur dalam gram.

Carbohydrates (Per 100 ml)	Numerik	Jumlah total karbohidrat yang terkandung dalam produk, termasuk gula alami dan tambahan serta pati.
Sugars (Per 100 ml)	Numerik	Kandungan jumlah gula dalam produk per 100 ml, termasuk gula alami dan gula tambahan selama proses pembuatan.
Fiber (Per 100 ml)	Numerik	Jumlah serat makanan yang terkandung dalam produk per 100 ml yang diukur dalam gram.
Proteins (Per 100 ml)	Numerik	Jumlah protein yang terkandung dalam produk sebagai nutrisi penting untuk perbaikan jaringan tubuh.
Salt (Per 100 ml)	Numerik	Jumlah garam yang terkandung dalam produk per 100 ml yang diukur dalam gram.
Fruits, Vegetables and Legumes (Per 100 ml in %)	Numerik	Persentase kandungan buah-buahan, sayuran, dan kacang-kacangan dalam produk per 100 ml yang menunjukkan seberapa besar proporsi bahan alami dalam produk.

Tabel 3.1 Deskripsi Variabel

### 3.3 Algoritma Machine Learning Klasifikasi

Dalam proses data mining, algoritma machine learning klasifikasi memiliki peran penting dalam mengelola dan menganalisis volume informasi yang besar. Algoritma klasifikasi bertujuan untuk memprediksi kategori atau label kelas dari data berdasarkan pola dan pengetahuan dari data training. Dalam kasus ini kami menggunakan dua metode klasifikasi yaitu Random Forest dan Decision Tree.

#### 3.3.1 Algoritma Random Forest

Random Forest merupakan salah satu algoritma machine learning yang digunakan untuk melakukan klasifikasi atau regresi. Algoritma ini bekerja dengan menggabungkan beberapa pohon keputusan yang dibangun dari sampel yang dipilih secara acak. Metode ini terdiri dari root node, internal node dan leaf node. Oleh karena itu, Random Forest dapat meningkatkan akurasi prediksi dan mengurangi overfitting pada decision tree individu. Algoritma Random Forest tidak memiliki rumus matematis yang pasti, tetapi menggunakan konsep ensemble learning dengan langkah-langkah sebagai berikut.

- Menentukan jumlah pohon (k) yang akan dibangun dari total fitur (m), dimana k kurang dari m.
- Mengambil sampel acak sebanyak N dari dataset untuk setiap pohon menggunakan bootstrap sampling.
- Pada setiap pohon, dilakukan pengambilan subset prediktor secara acak dari m fitur, dimana jumlah fitur yang dipilih biasanya adalah  $\sqrt{m}$  untuk klasifikasi dan  $m/3$  untuk regresi .
- Mengulangi proses pada langkah kedua dan ketiga sebanyak k kali untuk membangun k pohon keputusan.
- Menggabungkan hasil prediksi dari setiap pohon yang diperoleh dengan cara voting terbanyak untuk hasil klasifikasi atau dengan menghitung rata-rata untuk hasil regresi dari seluruh pohon yang dibangun.

### 3.3.2 Algoritma Decision Tree

Decision Tree adalah metode kecerdasan buatan berbentuk struktur pohon dengan konsep flowchart yang digunakan untuk membantu membuat keputusan atau menyelesaikan tugas yang berkaitan dengan regresi dan klasifikasi. Decision Tree terdiri dari kumpulan node yang dihubungkan oleh cabang. Setiap node internal mewakili atribut dan cabang menggambarkan hasil pengujian atau nilai input atribut sedangkan leaf node mewakili kelas atau distribusi kelas.

Decision Tree berguna untuk menggabungkan pola, pengetahuan atau informasi ke dalam sebuah bentuk pohon keputusan yang dapat dipahami. Dalam prosesnya, Decision Tree mampu menguraikan data kompleks dan menyederhanakan ke dalam bentuk yang lebih mudah untuk pengambilan keputusan. Secara matematis, pemilihan atribut untuk setiap node keputusan didasarkan pada metrik seperti Information Gain dan Gini Index.

#### 1. Information Gain

$$IG(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot Entropy$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi atribut A

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

$|S|$  = Jumlah kasus dalam S

- Perhitungan nilai entropi

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

## 2. Gini Index

$$Gini(S) = 1 - \sum_{i=1}^c p_i^2$$

Keterangan:

S = Himpunan Kasus

c = Jumlah Kelas

$p_i$  = Proporsi elemen dalam kelas ke-i dalam himpunan S

## 3.4 Model Evaluation

### 3.4.1 Confusion Matrix

Confusion Matrix adalah suatu alat yang penting untuk melakukan evaluasi model klasifikasi yang digunakan. Confusion Matrix memberikan informasi secara rinci dan mendalam mengenai hasil model prediksi dibandingkan dengan hasil yang sebenarnya. Berikut adalah elemen-elemen dari confusion matrix:

1. True Positives (TP): Jumlah instance yang benar benar diprediksi sebagai positif
2. True Negatives (TN): Jumlah instance yang benar benar diprediksi sebagai negatif.
3. False Positives (FP): Jumlah instance yang salah diprediksi sebagai positif.
4. False Negatives (FN): Jumlah instance yang salah diprediksi sebagai negatif.

Dengan menganalisis elemen-elemen di atas, didapatkan sebuah wawasan mengenai jenis kesalahan yang terjadi pada model dan mendapatkan informasi seberapa baik model tersebut dalam proses klasifikasi.

### 3.4.2 Metrics Evaluation

Untuk lebih mengetahui mengenai evaluasi model, ada beberapa metrik utama yang dihitung berdasarkan confusion matrix. Metrik-metrik ini memberikan penilaian dalam bentuk kuantitatif terhadap kinerja model yang digunakan. Metrik yang digunakan sebagai berikut:

1. Akurasi (Accuracy): Rasio instance yang diprediksi dengan benar terhadap total instance. Akurasi adalah ukuran efektivitas keseluruhan model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP}$$

2. Presisi (Precision): Rasio instance positif yang berhasil diprediksi dengan benar terhadap total prediksi positif. Menunjukkan akurasi prediksi positif.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall: Rasio instance positif yang diprediksi dengan benar terhadap instance positif yang sebenarnya, hal ini mengukur kemampuan model untuk mengidentifikasi instance positif.

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

4. F1-Score: Rata-rata dari presisi dan recall. Ini memberikan sebuah metrik tunggal yang menyeimbangkan antara presisi dan recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Presisi + Recall}$$

Metrik-metrik di atas akan memberikan pandangan yang lebih komprehensif terhadap kinerja dari model yang digunakan, memberikan informasi mengenai kelemahan dan kekuatan dari model sehingga dapat diketahui area yang perlu diperbaiki. Dengan menggunakan metrik-metrik ini, kita dapat membuat sebuah keputusan yang berdasar untuk mengoptimalkan dan memperbaiki model. (Muhamad dan Rizal, 2023).

## BAB 4

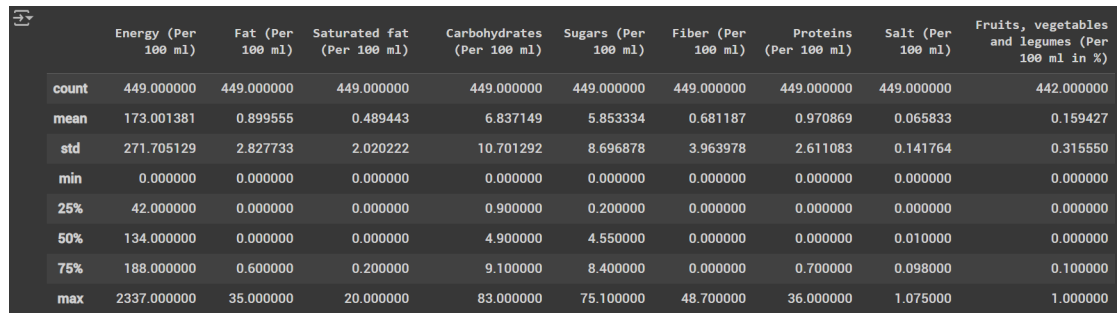
### HASIL DAN PEMBAHASAN

#### 4.1 Insight Exploratory Data Analysis (EDA)

##### 4.1.1 Data Shape

Data shape merupakan informasi mengenai dimensi dari dataset, dimana angka pertama (sebelum koma) merujuk pada jumlah baris dan angka kedua (setelah koma) merujuk pada jumlah kolom. Dataset yang digunakan dalam penelitian ini memiliki *shape* (449,11), ini artinya terdapat 449 baris dan 11 kolom.

##### 4.1.2 Data Description

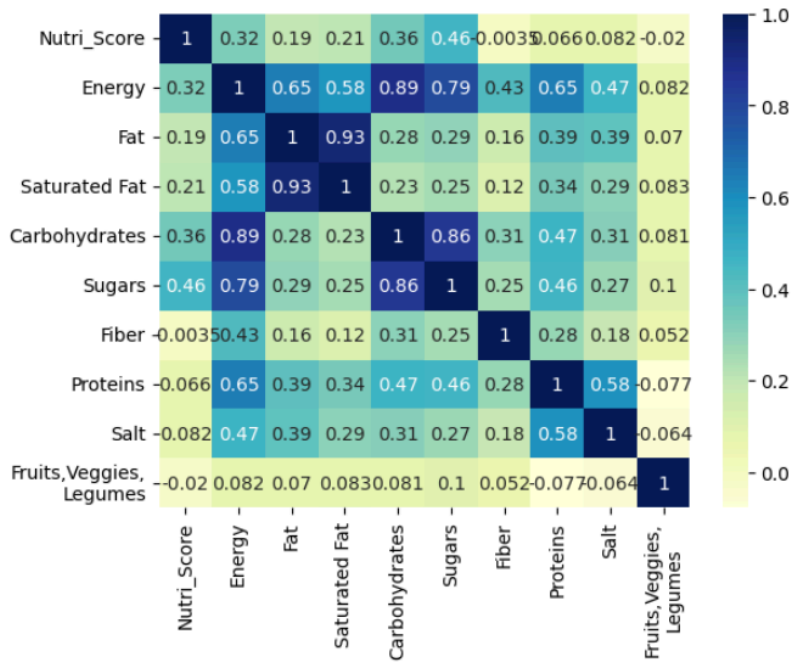


	Energy (Per 100 ml)	Fat (Per 100 ml)	Saturated fat (Per 100 ml)	Carbohydrates (Per 100 ml)	Sugars (Per 100 ml)	Fiber (Per 100 ml)	Proteins (Per 100 ml)	Salt (Per 100 ml)	Fruits, vegetables and legumes (Per 100 ml in %)
count	449.000000	449.000000	449.000000	449.000000	449.000000	449.000000	449.000000	449.000000	442.000000
mean	173.001381	0.899555	0.489443	6.837149	5.853334	0.681187	0.970869	0.065833	0.159427
std	271.705129	2.827733	2.020222	10.701292	8.696878	3.963978	2.611083	0.141764	0.315550
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	42.000000	0.000000	0.000000	0.900000	0.200000	0.000000	0.000000	0.000000	0.000000
50%	134.000000	0.000000	0.000000	4.900000	4.550000	0.000000	0.000000	0.010000	0.000000
75%	188.000000	0.600000	0.200000	9.100000	8.400000	0.000000	0.700000	0.098000	0.100000
max	2337.000000	35.000000	20.000000	83.000000	75.100000	48.700000	36.000000	1.075000	1.000000

Gambar 4.1 Data Description

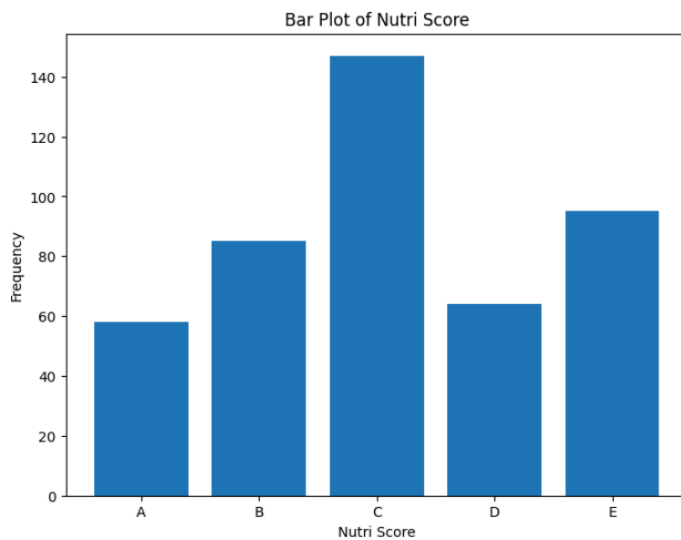
Data description berfungsi untuk memberikan ringkasan statistik dari variabel numerik pada dataset. Ringkasan statistik ini meliputi jumlah data (count), nilai unique, rata-rata, standar deviasi (std), nilai minimum (min), kuartil (25%, 50%, 75%), dan nilai maksimum (max). Informasi ini dapat memberikan wawasan awal terkait pola dan distribusi data, sehingga peneliti dapat memilih langkah analisis yang tepat dalam mengolah data.

### 4.1.3 Heatmap Korelasi Pearson



Gambar 4.2 Heatmap Korelasi Pearson

### 4.1.4 Barplot Variabel Y



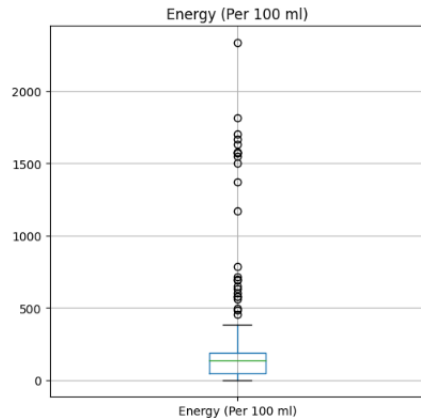
Gambar 4.3 Barplot Variabel Y

Heatmap korelasi menunjukkan hubungan linear antara variabel-variabel pada dataset. Nilai korelasi ini berada dalam rentang -1 hingga 1. Nilai korelasi positif menunjukkan hubungan yang positif, sedangkan korelasi negatif menunjukkan hubungan yang negatif. Selain itu, warna yang lebih gelap menunjukkan korelasi yang lebih kuat, baik positif maupun negatif.

Gambar 4.3 menunjukkan bar plot frekuensi dari *Nutri-Score*. Berdasarkan barplot ini, dapat dilihat bahwa *Nutri-Score* C memiliki frekuensi tertinggi, diikuti oleh *Nutri-Score* E, *Nutri-Score* B, *Nutri-Score* D, dan frekuensi yang terendah terdapat pada *Nutri-Score* A.

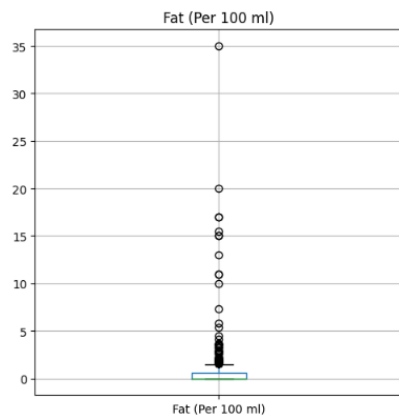


#### 4.1.5 Boxplot Variabel X



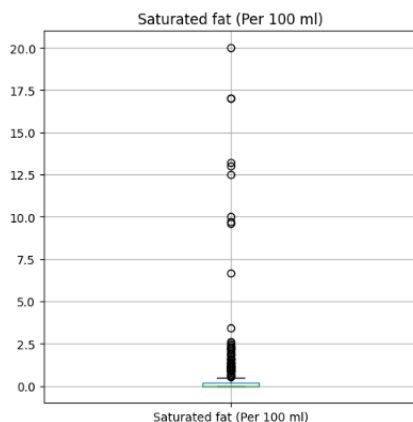
Gambar 4.4 Boxplot Variabel Energy

Berdasarkan Gambar 4.4, dapat disimpulkan bahwa variabel Energy memiliki right-skewed distribution. Boxplot energy menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan energi yang rendah, yaitu di bawah 250 g/100ml. Namun, masih terdapat beberapa minuman yang memiliki kandungan energi yang tinggi, yaitu berkisar antara 500 g/100ml hingga lebih dari 2000 g/100 ml.



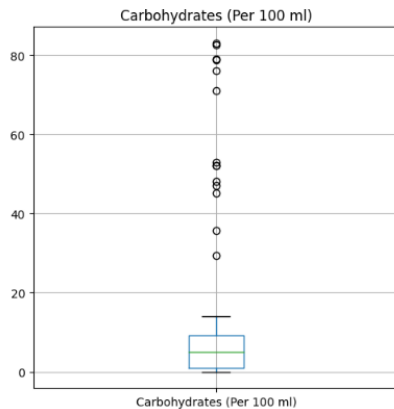
Gambar 4.5 Boxplot Variabel Fat

Berdasarkan Gambar 4.5, dapat simpulkan bahwa variabel Fat memiliki right-skewed distribution. Boxplot Fat menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan lemak total yang sangat rendah, yaitu di bawah 2.5 g/100ml. Akan tetapi, masih terdapat beberapa minuman yang memiliki kandungan lemak yang tinggi, yaitu berkisar antara 5 g/100ml hingga yang tertinggi 35 g/100 ml.



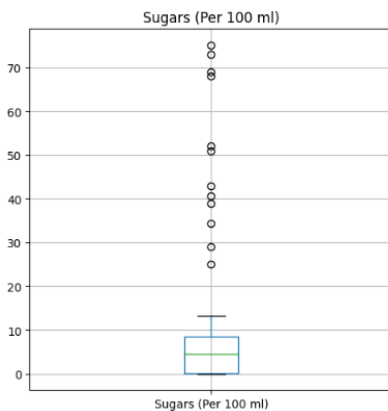
Gambar 4.6 Boxplot Variabel Saturated Fat

Berdasarkan Gambar 4.6, dapat simpulkan bahwa variabel Saturated Fat memiliki right-skewed distribution. Boxplot Saturated Fat menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan lemak jenuh yang sangat rendah, yaitu di bawah 2 g/100ml. Namun, masih terdapat cukup banyak minuman yang memiliki kandungan lemak jenuh yang tinggi, yaitu berkisar antara 2.5 g/100ml hingga yang tertinggi 20 g/100ml.



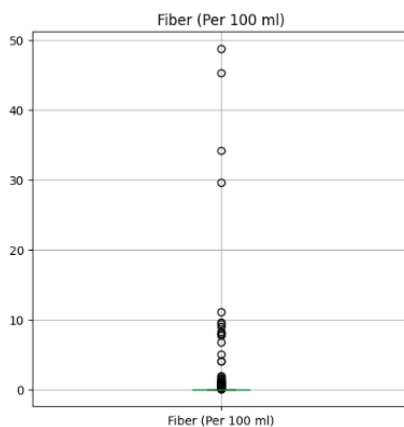
Gambar 4.7 Boxplot Variabel Carbohydrates

Berdasarkan Gambar 4.7, dapat disimpulkan bahwa variabel Carbohydrates memiliki right-skewed distribution. Boxplot Carbohydrates menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan karbohidrat yang rendah, yaitu berkisar antara 0 - 15 g/100ml. Namun, masih terdapat beberapa minuman yang memiliki kandungan karbohidrat yang tinggi, yaitu berkisar antara 20 g/100ml hingga lebih dari 80 g/100ml.



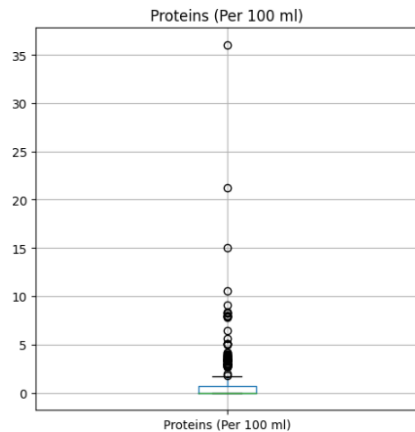
Gambar 4.8 Boxplot Variabel Sugar

Berdasarkan Gambar 4.8, dapat disimpulkan bahwa variabel Sugar memiliki right-skewed distribution. Boxplot Sugar menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan gula yang rendah, yaitu di antara 0 - 15 g/100ml. Akan tetapi, masih terdapat beberapa minuman yang memiliki kandungan gula yang tinggi, yaitu berkisar antara 20 g/100ml hingga lebih dari 70 g/100 ml

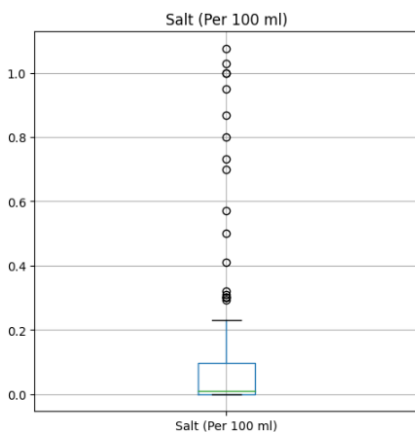


Gambar 4.9 Boxplot Variabel Fiber

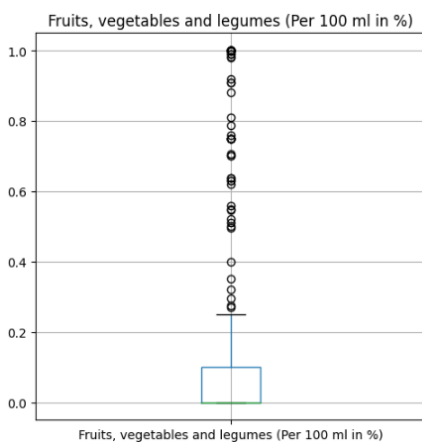
Berdasarkan Gambar 4.9, dapat disimpulkan bahwa variabel Fiber memiliki right-skewed distribution. Boxplot Fiber menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan serat yang sangat rendah, yaitu kurang dari 5 g/100ml. Selain itu, masih terdapat cukup banyak minuman yang memiliki kandungan serat yang tinggi, yaitu berkisar antara 10 g/100ml hingga 50 g/100ml.



Gambar 4.10 Boxplot Variabel Proteins



Gambar 4.11 Boxplot Variabel Salt



Gambar 4.12 Boxplot Variabel Fruits,vegetables and legumes

Berdasarkan Gambar 4.10, dapat disimpulkan bahwa variabel Proteins memiliki right-skewed distribution. Boxplot Proteins menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan protein yang rendah, yaitu kurang dari 5 g/100ml. Selain itu, masih terdapat beberapa minuman yang memiliki kandungan protein yang tinggi, yaitu berkisar antara 5 g/100ml hingga lebih dari 35 g/100 ml

Berdasarkan Gambar 4.11, dapat disimpulkan bahwa variabel Salt memiliki right-skewed distribution. Boxplot Salt menunjukkan bahwa sebagian besar sampel minuman memiliki kandungan garam yang sangat rendah, yaitu berkisar antara 0 - 0.2 g/100ml. Namun, masih terdapat cukup banyak minuman yang memiliki kandungan garam yang tinggi, yaitu berkisar antara 0.4 g/100ml hingga lebih dari 1g/100ml.

Berdasarkan Gambar 4.12, dapat disimpulkan bahwa variabel Fruits,vegetables and legumes memiliki right-skewed distribution. Boxplot ini menunjukkan bahwa sebagian besar sampel minuman memiliki persentase kandungan buah, sayur, dan kacang yang sangat rendah, yaitu berkisar antara 0% - 20%. Selain itu, masih terdapat beberapa minuman yang memiliki persentase kandungan buah, sayur, dan kacang yang tinggi, yaitu berkisar antara 40% hingga lebih dari 100%.

## 4.2 Pre Processing

### 4.2.1 Cleaning Data

Pada tahap cleaning data, beberapa langkah penting dilakukan untuk memastikan data bersih dan siap untuk analisis. Pada kolom Nutri\_Score, teks "Nutri-Score" dan spasi dihapus. Kolom Energy (Per 100 ml) dibersihkan dengan menghapus koma dan spasi yang tidak diperlukan. Simbol yang tidak diinginkan seperti '<', '?', 'g', dan '-' pada berbagai kolom dihapus, dan nilai '?' serta '-' diganti dengan nol. Kolom Fruits, vegetables and legumes (Per 100 ml in %) diubah dari persentase menjadi desimal, dan semua kolom numerik dikonversi menjadi tipe data float. Gambar di bawah ini menampilkan inkonsistensi data pada setiap kolom.

	Nama_Kolom	Symbols_Detected
0	Nutri_Score	A, B, C, D, E
1	Energy (Per 100 ml)	
2	Fat (Per 100 ml)	<, g
3	Saturated fat (Per 100 ml)	<, ?, g
4	Carbohydrates (Per 100 ml)	<, g
5	Sugars (Per 100 ml)	<, ?, g
6	Fiber (Per 100 ml)	-, <, ?, g
7	Proteins (Per 100 ml)	<, g
8	Salt (Per 100 ml)	<, ?, g
9	Fruits, vegetables and legumes (Per 100 ml in %)	%, ?, a, n

Gambar 4.13 Deteksi Simbol pada Setiap Kolom

Hasil dari proses cleaning data menunjukkan bahwa dataset telah dibersihkan dari simbol yang tidak diinginkan, nilai tertentu telah diubah menjadi nol, dan persentase telah dikonversi menjadi desimal. Kolom numerik juga telah dikonversi menjadi tipe data float, sehingga data siap untuk tahap analisis lebih lanjut.

#### 4.2.2 Missing Value

Gambar di bawah ini menampilkan ringkasan mengenai *missing value* untuk setiap variabel yang ada pada dataset. Dapat dilihat bahwa variabel Nutri Score, Energy, Fat, Saturated Fat, Carbohydrates, Sugars, Fiber, Protein, dan Salt tidak memiliki nilai yang hilang. Namun, pada variabel Fruits, vegetables and legumes terdapat 7 nilai yang hilang atau sebesar 2% dari total data yang ada.

	Missing Values	Missing Values (%)	Data Type
Nutri_Score	0	0.0	object
Energy (Per 100 ml)	0	0.0	float64
Fat (Per 100 ml)	0	0.0	float64
Saturated fat (Per 100 ml)	0	0.0	float64
Carbohydrates (Per 100 ml)	0	0.0	float64
Sugars (Per 100 ml)	0	0.0	float64
Fiber (Per 100 ml)	0	0.0	float64
Proteins (Per 100 ml)	0	0.0	float64
Salt (Per 100 ml)	0	0.0	float64
Fruits, vegetables and legumes (Per 100 ml in %)	7	2.0	float64

Gambar 4.14 Cek Missing Value

Temuan di atas menunjukkan bahwa secara garis besar, dataset yang dikumpulkan cukup lengkap, hanya sebagian kecil (2%) yang hilang pada variabel Fruits, vegetables and legumes. Hal itu menunjukkan bahwa perlu dilakukannya imputasi data untuk mengisi data yang hilang tersebut.

Untuk mengatasi adanya kekosongan data, dilakukannya imputasi data menggunakan nilai 0. Hal itu dipilih karena asumsi bahwa dari produk tersebut tidak memiliki kandungan Fruits, vegetables and legumes dan juga untuk menghindari adanya bias serta dengan pertimbangan bahwa data yang hilang hanya sebagian kecil sebesar 2%.

### 4.2.3 *Encoding Variabel*

Pada tahap encoding label ini, kami akan mengubah nilai kategori dalam kolom Nutri\_Score menjadi nilai numerik menggunakan metode label encoding. Metode ini mengganti setiap kategori dengan angka unik, yang penting untuk analisis statistik dan machine learning. Menggunakan LabelEncoder dari scikit-learn, nilai kategori dalam kolom Nutri\_Score diubah menjadi angka 0 hingga 4. Berikut adalah hasil encoding label

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
df['Nutri_Score'] = le.fit_transform(df['Nutri_Score'])

for label, class_ in enumerate(le.classes_):
    print(f"Class {label}: {class_}")

Class 0: A
Class 1: B
Class 2: C
Class 3: D
Class 4: E
```

Gambar 4.15 Encoding Variabel

Hasil dari proses encoding menunjukkan bahwa kategori dalam kolom Nutri\_Score telah berhasil diubah menjadi nilai numerik, yang memudahkan proses analisis lebih lanjut dan penerapan algoritma machine learning. Dataset yang telah di-encode disimpan dalam variabel df2 untuk memastikan kelangsungan pre-processing dan analisis berikutnya.

### 4.2.4 *Handling Outlier*

Pada tahap handling outlier ini, kami mengidentifikasi dan menangani outlier dalam dataset. Gambar di bawah ini menunjukkan distribusi dari beberapa variabel dalam dataset melalui box plot. Outlier terlihat pada beberapa variabel seperti Energy, Fat, Saturated fat, Carbohydrates, Sugars, Fiber, Proteins, Salt, dan Fruits, vegetables and legumes.

Untuk menangani outlier, kami menggunakan metode winsorizing, di mana nilai outlier diubah menjadi NaN jika mereka berada di luar batas yang ditentukan. Setelah itu, kami menggunakan IterativeImputer dari scikit-learn untuk mengisi nilai-nilai yang hilang menggunakan regresi linier. Proses ini mengiterasi beberapa kali untuk mengisi nilai-nilai yang hilang berdasarkan hubungan antar variabel dalam dataset.

Dengan demikian, kami berhasil menangani outlier dalam dataset, sehingga data menjadi lebih konsisten dan siap untuk tahap analisis selanjutnya.

#### 4.2.5 *Scaling Data*

Langkah terakhir dalam *preprocessing* adalah melakukan scaling data. Proses ini penting untuk memastikan bahwa semua variabel memiliki skala yang sama, sehingga tidak ada variabel yang mendominasi yang lain dalam analisis. Kami menggunakan *StandardScaler* dari *scikit-learn* untuk melakukan scaling.

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
scaled = scaler.fit_transform(df[df.columns[1:10]])
scaled_df = pd.DataFrame(scaled, columns = df.columns[1:10])
df = pd.concat([df.drop(columns = df.columns[1:10]), scaled_df], axis=1)
```

Gambar 4.16. Scaling Data

Data diskalakan menggunakan *StandardScaler*, yang mengubah nilai-nilai pada setiap kolom menjadi distribusi dengan rata-rata 0 dan standar deviasi 1. Dengan scaling data, semua variabel dalam dataset kini berada dalam skala yang sama, memudahkan proses analisis lebih lanjut dan memastikan bahwa hasil analisis tidak bias karena perbedaan skala antar variabel. Dataset yang telah diskalakan disimpan dalam variabel *df*, siap untuk analisis berikutnya.

### 4.3 Modeling dan Evaluation

#### 4.3.1 Data Partition

Kami menggunakan fungsi *train\_test\_split* dari *scikit-learn* untuk memisahkan data dengan rasio 70% untuk pelatihan dan 30% untuk pengujian. Variabel target yang dipisahkan adalah *Nutri\_Score*, sementara fitur lainnya digunakan sebagai variabel input.

#### 4.3.2 Hyperparameter Tuning

Pada tahap ini, kami mengoptimalkan parameter *max\_depth* dari model *Random Forest*. Kami mengevaluasi performa model dengan beberapa nilai *max\_depth* yaitu 3, 5, 7, dan 9. Dari hasil pengujian, terlihat bahwa semua nilai *max\_depth* menghasilkan akurasi yang sama, yaitu sekitar 96.3%.

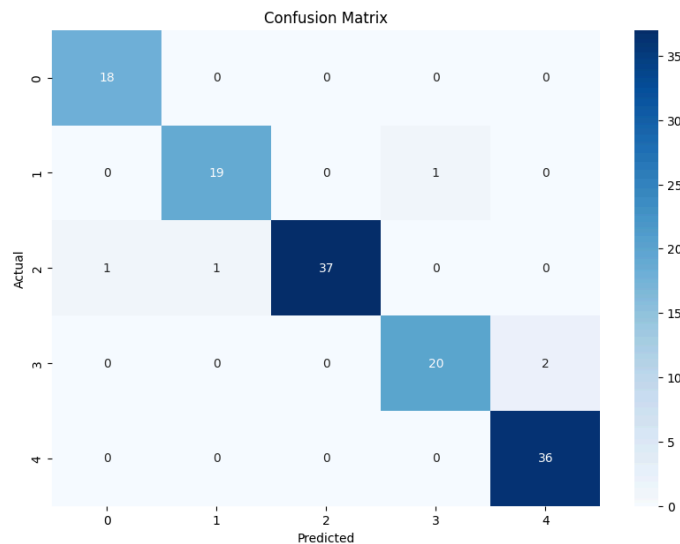
### 4.3.3 Model Evaluation

Kami menggunakan model Random Forest dan Decision Tree dengan parameter `max_depth 3`, yang memberikan hasil terbaik, untuk melakukan prediksi pada data uji. Hasil prediksi kemudian dievaluasi menggunakan matriks kebingungan (Confusion Matrix) dan metrik evaluasi lainnya seperti Akurasi, Presisi, Recall, dan F1-Score.

#### 4.3.3.1 Random Forest Model

##### Confusion Matrix

Gambar di bawah ini menunjukkan confusion matriks untuk model Random Forest yang digunakan.



Gambar 4.17 Confusion Matrix

##### Metrics Evaluation

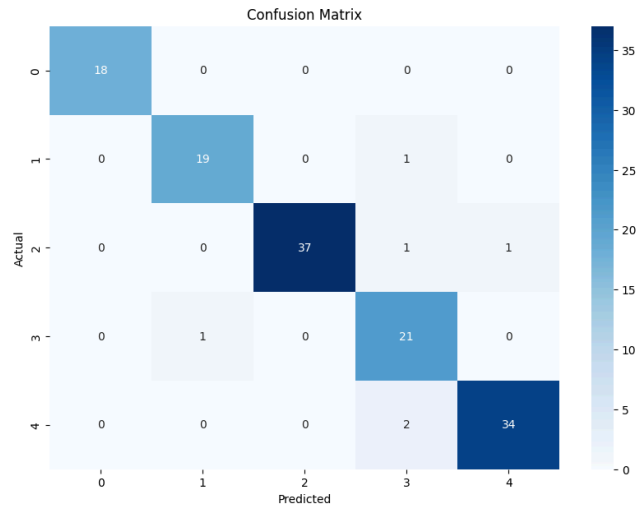
- Accuracy: Akurasi dari model ini adalah 96.3%, menunjukkan bahwa model mampu memprediksi dengan tepat 96.3% dari seluruh data uji.
- Precision: Presisi model adalah 0.9594, menunjukkan tingkat akurasi dari prediksi positif.
- Recall: Recall dari model adalah 0.9616, menunjukkan kemampuan model dalam menangkap semua kasus positif.
- F1-Score: F1-Score dari model adalah 0.9600, yang merupakan rata-rata harmonis dari presisi dan recall.



#### 4.3.3.2 Decision Tree Model

##### Confusion Matrix

Gambar di bawah ini menunjukkan confusion matriks untuk model Decision Tree yang digunakan.



Gambar 4.18 Confusion Matrix

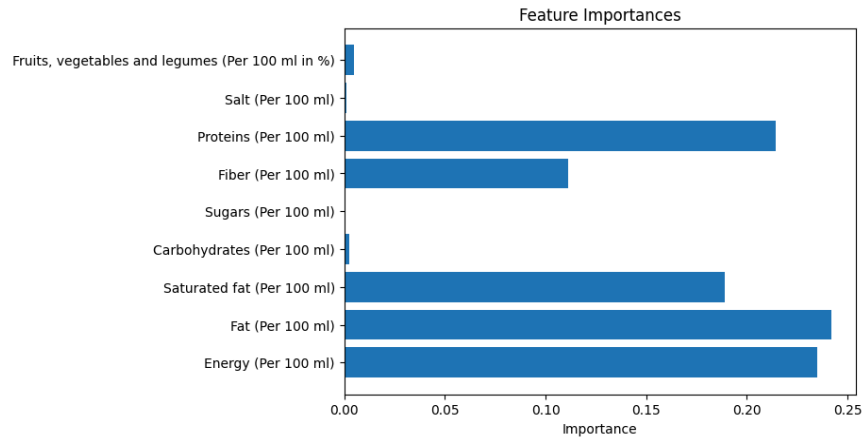
##### Metrics Evaluation

- Accuracy: Akurasi dari model ini adalah 95.56%, menunjukkan bahwa model mampu memprediksi dengan tepat 95.56% dari seluruh data uji.
- Precision: Presisi model adalah 0.9523, menunjukkan tingkat akurasi dari prediksi positif.
- Recall: Recall dari model adalah 0.9595, menunjukkan kemampuan model dalam menangkap semua kasus positif.
- F1-Score: F1-Score dari model adalah 0.9550, yang merupakan rata-rata harmonis dari presisi dan recall.

Secara keseluruhan, model Random Forest dan Decision Tree menunjukkan performa yang sangat baik dengan akurasi, presisi, recall, dan F1-Score yang tinggi. Model Random Forest sedikit lebih unggul dalam hal akurasi dan F1-Score dibandingkan dengan model Decision Tree.

## 4.3.4 Feature Importance

### 4.3.4.1 Random Forest Model



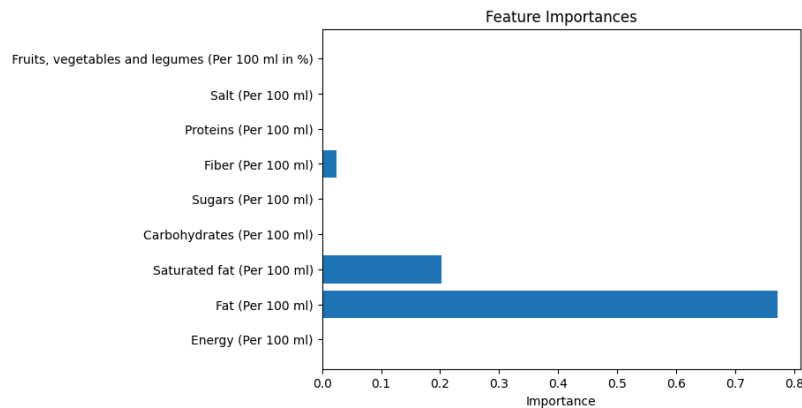
Gambar 4.20 Feature Importance Random Forest

1. Energy (Per 100 ml):
  - Memiliki tingkat kepentingan yang tinggi, menunjukkan bahwa energi per 100 ml adalah salah satu faktor paling signifikan dalam model Random Forest ini.
2. Fat (Per 100 ml):
  - Juga menunjukkan tingkat kepentingan yang tinggi, hampir setara dengan energi, yang berarti lemak per 100 ml merupakan prediktor penting dalam model.
3. Saturated fat (Per 100 ml):
  - Memiliki tingkat kepentingan yang signifikan, meskipun lebih rendah daripada fat dan energy. Ini menunjukkan bahwa lemak jenuh juga merupakan faktor penting dalam model.
4. Proteins (Per 100 ml):
  - Menunjukkan tingkat kepentingan yang cukup tinggi, yang berarti protein per 100 ml berperan penting dalam prediksi model.
5. Fiber (Per 100 ml):
  - Memiliki tingkat kepentingan yang sedang. Serat per 100 ml masih relevan dalam model, namun tidak sekuat fitur lainnya.
6. Sugars (Per 100 ml):
  - Menunjukkan tingkat kepentingan yang sangat rendah, menunjukkan bahwa gula per 100 ml kurang signifikan dalam prediksi model.

7. Carbohydrates (Per 100 ml):
  - Hampir tidak memiliki tingkat kepentingan, menunjukkan bahwa karbohidrat per 100 ml tidak banyak berkontribusi dalam model.
8. Salt (Per 100 ml):
  - Juga menunjukkan tingkat kepentingan yang sangat rendah, menunjukkan bahwa garam per 100 ml kurang signifikan dalam model.
9. Fruits, vegetables and legumes (Per 100 ml in %):
  - Memiliki tingkat kepentingan yang sangat rendah, hampir nol, menunjukkan bahwa persentase buah-buahan, sayuran, dan kacang-kacangan per 100 ml hampir tidak berpengaruh dalam model prediksi.

Dari diagram ini, kita dapat menyimpulkan bahwa dalam model Random Forest yang digunakan, fitur yang paling signifikan adalah Energy (Per 100 ml) dan Fat (Per 100 ml), diikuti oleh Saturated fat (Per 100 ml) dan Proteins (Per 100 ml). Fitur-fitur seperti Sugars, Carbohydrates, Salt, dan Fruits, vegetables and legumes memiliki pengaruh yang sangat kecil atau hampir tidak ada dalam model prediksi ini.

#### 4.3.4.2 Decision Tree Model



Gambar 4.19 Feature Importance Decision Tree

1. Fat (Per 100 ml):
  - Fitur ini memiliki tingkat kepentingan yang sangat tinggi, menunjukkan bahwa fat per 100 ml adalah prediktor yang paling signifikan dalam model Decision Tree ini.

2. Energy (Per 100 ml):
  - Menunjukkan tingkat kepentingan yang signifikan, namun lebih rendah dibandingkan fat. Ini menunjukkan bahwa energy per 100 ml juga merupakan faktor penting dalam model.
3. Saturated fat (Per 100 ml):
  - Memiliki tingkat kepentingan yang lebih rendah tetapi masih relevan, menunjukkan bahwa lemak jenuh juga berperan dalam model, meskipun tidak sebesar fat dan energy.
4. Carbohydrates (Per 100 ml):
  - Menunjukkan tingkat kepentingan yang kecil, menunjukkan bahwa karbohidrat per 100 ml memiliki pengaruh yang kecil dalam model.
5. Fiber (Per 100 ml):
  - Memiliki tingkat kepentingan yang sangat kecil, hampir tidak berpengaruh dalam prediksi model.
6. Fitur lainnya (Sugars, Proteins, Salt, Fruits, vegetables, and legumes):
  - Fitur-fitur ini memiliki tingkat kepentingan yang sangat rendah atau bahkan nol, menunjukkan bahwa mereka tidak signifikan dalam model prediksi ini.

Dari diagram ini, kita dapat menyimpulkan bahwa dalam model Decision Tree yang digunakan, fitur yang paling signifikan adalah Fat (Per 100 ml), diikuti oleh Energy (Per 100 ml) dan Saturated fat (Per 100 ml). Fitur-fitur seperti Carbohydrates, Fiber, Sugars, Proteins, Salt, dan Fruits, vegetables, and legumes memiliki pengaruh yang sangat kecil atau hampir tidak ada dalam model prediksi ini.

## **BAB 5**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian yang telah dilakukan, didapatkan kesimpulan sebagai berikut:

- Nutri-score terbagi menjadi lima kategori yaitu A, B, C, D, dan E, dimana kategori A merupakan produk minuman dengan kandungan gizi terbaik hingga E yang merupakan kualitas gizi terburuk.
- Model Random Forest dan Decision Tree menunjukkan performa yang sangat baik dalam mengklasifikasikan nutrition-score pada minuman. Model Random Forest memiliki akurasi 96.3%, sementara Decision Tree memiliki akurasi 95.56%, yang menunjukkan kemampuan kedua model tersebut dalam memprediksi dengan tepat sebagian besar data uji. Presisi dan recall untuk kedua model juga tinggi, dengan Random Forest memiliki presisi 0.9594 dan recall 0.9616, serta Decision Tree memiliki presisi 0.9523 dan recall 0.9595. Ini menunjukkan bahwa model mampu mengidentifikasi instance positif dengan baik dan mengurangi kesalahan dalam prediksi positif. F1-Score dari model Random Forest adalah 0.9600 dan Decision Tree adalah 0.9550, yang menunjukkan keseimbangan yang baik antara presisi dan recall, memberikan gambaran komprehensif tentang performa model. Penelitian ini juga menunjukkan bahwa penggunaan data gizi dalam klasifikasi nutrition-score dapat memberikan hasil yang akurat dan dapat diandalkan, serta pentingnya data yang lengkap dan preprocessing yang baik untuk mendapatkan hasil yang optimal.

#### **5.2 Saran**

Berdasarkan temuan dan kesimpulan tersebut, beberapa saran untuk penelitian selanjutnya dan implementasi praktis adalah sebagai berikut:


- Selain menggunakan model Random Forest dan Decision Tree, penelitian selanjutnya dapat mencoba algoritma lain seperti Gradient Boosting, AdaBoost, atau Neural Networks untuk melihat apakah ada peningkatan performa dalam klasifikasi.
- Melakukan validasi eksternal terkait dataset yang di dapat untuk meningkatkan keakuratan model.
- Hasil dari penelitian ini dapat diimplementasikan pada aplikasi nyata seperti aplikasi mobile yang dapat membantu konsumen dalam memilih minuman berdasarkan nilai gizi yang terklasifikasi dengan baik.

## DAFTAR PUSTAKA

- Astuti, D., Rofii, M. and Sudarsono, R., 2021. Komparasi Tingkat Akurasi Random Forest dan Decision Tree C4.5 Pada Klasifikasi Data Penyakit Infertilitas.
- Badan Pengawasan Obat dan Makanan (BPOM), 2018. Peraturan Nomor 22 Tahun 2019 tentang Informasi Nilai Gizi pada Label Pangan Olahan. Jakarta: BPOM.
- Breiman, L., 2001. 'Random forests', *Machine Learning*, 45(1), pp. 5-32.
- Kusrini, & Luthfi, E.T., 2017. *Algoritma Data Mining*. Yogyakarta: Andi.
- Muhamad, F. & Rizal, A.S., 2023. 'Klasifikasi dan Evaluasi Performa Model Random Forest untuk Prediksi Stroke', *JT : Jurnal Teknik*, 12(02), pp. 72-80.
- Perdana, A., 2021. Analisis Data Mining Algoritma Decision Tree Pada Prediksi Persediaan Obat (Studi Kasus: Apotek Franch Farma).
- Witten, I.H., Frank, E. & Hall, M.A., 2011. 'Data Mining: Practical Machine Learning Tools and Techniques. 3rd' ed. Burlington: Morgan Kaufmann.
- Szabo de Edelenyi, F. *et al.* (2019) 'Ability of the nutri-score front-of-pack nutrition label to discriminate the nutritional quality of foods in the German food market and consistency with nutritional recommendations', *Archives of Public Health*, 77(1). doi:10.1186/s13690-019-0357-x.
- Bandinkan kualitas nutrisi produk - produk makanan dengan Nutri-Score! [WWW Document], n.d. . Open Food Facts. URL <https://id.openfoodfacts.org/nutriscore>
- Vassaux, S., 2023. nutri score abcde [WWW Document]. Melba. URL <https://melba.io/en/ai/haccp/nutri-score-abcde>
- IterativeImputer [WWW Document], n.d. . scikit-learn. URL <https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html>

## LAMPIRAN


Kode Analisis (Google Collaboration):

 DM1\_SD-A1\_164221034\_164221057\_164221064\_164221087\_164221100.ipynb

Kode Scrapping (Google Collaboration):

 Scrapping\_OpenFoodFacts\_Kelompok9.ipynb

Dataset (Spreadsheet):

 beverages\_nutriscore\_dataset