

Perbandingan Kinerja Algoritma *Decision Tree* dan *Random Forest* untuk Klasifikasi Nutrisi pada Makanan Cepat Saji

Nuurul Izzati Yaman¹, Ayu Ratna Juwita², Santi Arum Puspita Lestari³, Sutan Faisal⁴
^{1,2,3,4}Universitas Buana Perjuangan Karawang, Indonesia

**email*: if20.nuurulyaman@mhs.ubpkarawang.ac.id

Info Artikel

Dikirim: 20 Mei 2024
Diterima: 4 Desember 2024
Diterbitkan: 4 Desember 2024

Kata kunci:

Decision Tree;
Gaya Hidup;
Klasifikasi;
Makanan;
Random Forest.

ABSTRAK

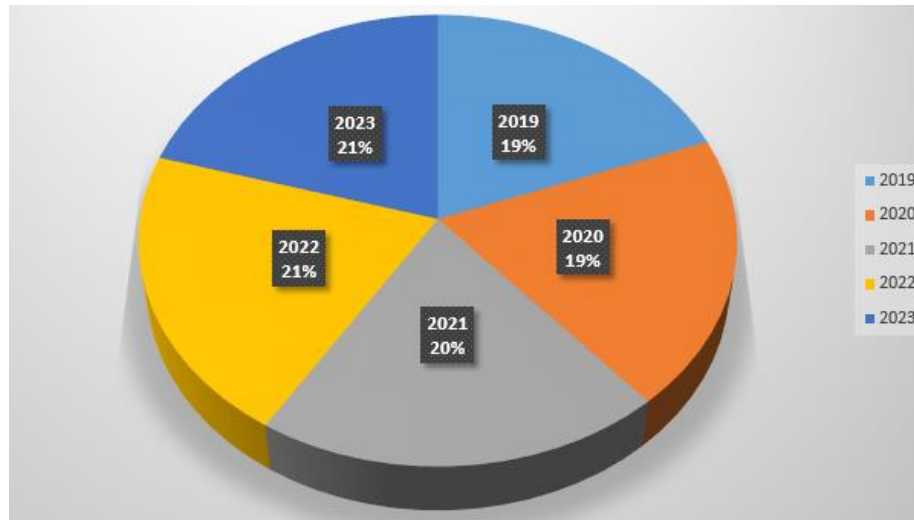
Makanan cepat saji telah menjadi bagian penting dari gaya hidup modern yang sibuk, makanan cepat saji lebih digemari karena membuat makan menjadi mudah dan nyaman. Anak muda zaman sekarang sangat menyukai makanan instan. namun, konsumsi makanan instan yang berlebihan dapat memicu berbagai masalah kesehatan, termasuk pola makan yang obsesif. Hal ini menimbulkan kebutuhan untuk mengembangkan metode analisis yang lebih akurat untuk mengklasifikasikan data nutrisi makanan cepat saji, tujuan klasifikasi adalah untuk memperoleh model pohon keputusan yang dapat digunakan untuk mengantisipasi dan memperhatikan bagaimana variable pada data yang berhubungan satu sama lain. Dalam membandingkan kinerja Algoritma *Decision Tree* dan *Random Forest* dalam memproses data nutrisi makanan cepat saji ditemukan bahwa semua variabel memiliki korelasi. Hasil implementasi ditemukan bahwa kedua model memiliki kemampuan yang luar biasa. kinerja Algoritma *Decision Tree* dan *Random Forest* pada dataset yang sama, *Random Forest* mengungguli *Decision Tree* dengan nilai akurasi 66.67%, sedangkan *Decision Tree* hanya mencapai 55.56%, menunjukkan bahwa *Random Forest* mampu memberikan prediksi yang lebih akurat untuk kelas data uji. Selain itu, karakteristik kelompok *Random Forest*, di mana beberapa pohon keputusan digabungkan, memberikan keunggulan dalam menangani kompleksitas data dan meningkatkan generalisasi model. Hasil ini menunjukkan bahwa pembelajaran kelompok dapat meningkatkan kinerja dan keandalan prediksi dalam membangun model klasifikasi, terutama dalam kasus dataset yang kompleks.

1. PENDAHULUAN

Masalah gizi dalam pembangunan kependudukan masih merupakan masalah utama dalam tatanan kependudukan global [1]. Akibatnya, masalah ini menjadi salah satu poin penting dalam *Millennium Development Goals* (MDGs). Kesehatan adalah hal paling penting bagi manusia dimana datangnya penyakit melalui pola makan [2]. Kesehatan sangat penting karena pola makan dapat menyebabkan penyakit, terutama bagi individu yang memiliki pekerjaan yang sangat padat serta tidak mempunyai waktu untuk sarapan dan lebih memilih makanan cepat saji yang mudah diakses di banyak restaurant atau kafe. Makanan cepat saji dikenal dengan istilah *junk food* diartikan sebagai makanan sampah atau makanan tidak bergizi yang dianggap tidak mengandung nutrisi penting untuk tubuh [3].

Makanan cepat saji telah menjadi bagian penting dari gaya hidup modern yang sibuk, karena mereka membuat makan dengan mudah dan nyaman. Remaja saat ini sangat menyukai makanan cepat saji. Namun,

mengonsumsi makanan cepat saji dalam jumlah yang berlebihan dapat menyebabkan berbagai masalah kesehatan, salah satunya adalah obesitas. Makanan cepat saji juga sering dikaitkan dengan kurangnya nutrisi. Ini karena orang seringkali tidak memiliki banyak waktu untuk melakukan rutinitas sehari-hari, yang membuat mereka kurang memperhatikan nutrisi dalam makanan yang mereka konsumsi. Remaja harus dididik dan didorong untuk mengambil tanggung jawab untuk memilih kudapan yang sehat, makanan yang mengandung zat gizi sangat penting untuk tubuh tumbuh dan berkembang [4].



Gambar 1. Rata-rata harian konsumsi protein Per Kapita (Tidak termasuk konsumsi protein makanan jadi) Gram.

Sumber: <https://www.bps.go.id/id>

Menurut data dari Badan Pusat Statistik Indonesia, rata-rata konsumsi protein per orang meningkat secara signifikan dari tahun 2019 hingga 2023 [4]. Fenomena ini dapat dianggap sebagai tanda bahwa perubahan signifikan dalam pola konsumsi masyarakat. Perubahan ini dapat disebabkan oleh sejumlah variabel, termasuk peningkatan kesadaran tentang gizi, dinamika ekonomi yang lebih baik dan perubahan dalam preferensi masyarakat tentang gaya hidup dan makanan. Menurut data Riskesdas sebagai besar orang Indonesia yang berusia lebih dari sepuluh tahun mengonsumsi makanan manis lebih dari sekali setiap hari 53.1%, Makanan berlemak, berkolesterol, dan gorengan lebih dari satu kali setiap hari 40.7%, dan hampir empat dari lima orang Indonesia [5].

Hal ini menimbulkan kebutuhan untuk mengembangkan metode analisis yang lebih akurat untuk mengklasifikasikan data nutrisi makanan cepat saji. klasifikasi adalah suatu mode untuk menemukan pola yang dapat memisahkan kelas data yang satu dengan yang dan menentukan objek yang termasuk dalam kategori tertentu dengan melihat bagaimana kelakuan dan karakteristik kelompok yang telah didefinisikan [6]. Tujuan klasifikasi adalah untuk mendapatkan model pohon keputusan yang dapat digunakan untuk membuat prediksi dan melihat bagaimana variabel dalam data berhubungan satu sama lain [7]. Berbagai algoritma klasifikasi digunakan untuk meningkatkan efisiensi dan akurasi pengelolaan data nutrisi makanan cepat saji. Salah satu algoritma yang digunakan adalah algoritma *Decision Tree* dan *Random Forest*.

Menurut Shafarindu, *Decision tree* adalah teknik yang cukup populer untuk klasifikasi dan prediksi karena menggunakan struktur pohon untuk menentukan urutan keputusan dan konsekuensi mereka [8]. *Decision tree* sering disebut struktur yang mirip dengan *flowchart* di mana masing-masing node mewakili nilai atribut, masing-masing cabang mewakili hasil pengujian, dan masing-masing daun mewakili kelas atau distribusi kelas [9]. Sedangkan, *Random Forest* merupakan salah satu metode yang digunakan untuk klasifikasi dan regresi, *Random forest* mampu menerjemahkan seperti yang dibentuk dari sekumpulan pohon keputusan juga [10]. *Random Forest* memiliki beberapa kelebihan, termasuk kemampuan untuk meningkatkan akurasi jika terdapat

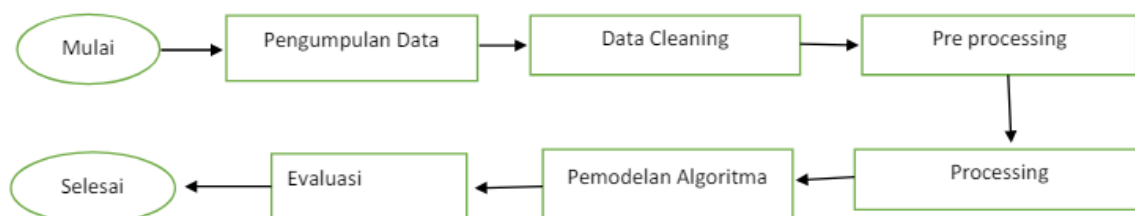
data yang hilang dan melawan outliers, serta efisiensi penyimpanan data. Selain itu, *Random Forest* memiliki proses seleksi fitur yang memungkinkannya mengambil fitur terpenting, yang dapat meningkatkan kinerja model klasifikasi. Karena proses pilih fitur, *Random Forest* pasti dapat beroperasi secara efektif dengan data besar dengan parameter yang kompleks [11].

Menurut Hady Suryono (2021) *Random Forest* membuat banyak pohon dengan banyak iterasi, membuatnya mirip dengan hutan. Namun, *Random Forest* mengubah algoritmanya sehingga hasil prediksi dari semua subpohon memiliki korelasi yang kecil. Keputusan tentang klasifikasi dibuat berdasarkan suara yang paling banyak diberikan untuk setiap pohon. Jika prediksi submodel independen, penggabungan prediksi dari berbagai model dalam ansambel akan berhasil [12].

Perbandingan kinerja algoritma *Decision Tree* dan *Random Forest* dapat memberikan informasi yang berharga untuk meningkatkan pemahaman tentang kinerja algoritma *Decision Tree* dan *Random Forest* dalam mengklasifikasi nutrisi dalam makanan saji. Oleh karena itu, Tujuan dari penelitian ini adalah untuk melihat bagaimana algoritma *Decision Tree* dan *Random Forest* bekerja dalam analisis nutrisi pada makanan cepat saji. Oleh karena itu, kami dapat menemukan model yang paling cocok untuk memberikan informasi yang relevan dan akurat tentang kualitas nutrisi dari berbagai pilihan menu makanan cepat saji. Diharapkan hasil penelitian ini akan memberikan dasar ilmiah untuk pembuatan program edukasi kesehatan dan perencanaan strategis bagi produsen makanan cepat saji untuk meningkatkan kualitas nutrisi produk mereka dan mendorong remaja untuk mengikuti gaya hidup sehat.

2. METODE PENELITIAN

Penelitian ini menerapkan algoritma *Decision Tree* dan *Random Forest* untuk melihat hasil akurasi terbaik tentang menganalisis nutrisi pada makanan instan. Penelitian ini mencakup beberapa tahapan, diantaranya sebagai berikut:



Gambar 1: Alur Prosedur Penelitian

2.1 Pengumpulan Data

Penelitian ini diawali dengan pengumpulan data dari melalui studi literature dan *platform* publik yaitu *kaggle.com*. Studi literature dilakukan untuk mencari landasan teori dan sumber referensi buku atau jurnal tentang teknik data mining dan digunakan untuk memeriksa penelitian terdahulu, metode yang akan digunakan dan tinjau literature. Dan pengambilan data pada situs *Kaggle.com*. Dataset yang diperoleh berjumlah 1147 dengan jumlah kolom yaitu 14 yang terdiri dari *Company, Item, Calories from, Total, Saturated, Trans Fat, Cholesterol, Sodium, Carbs, Fiber, Sugars, Protein, Weight*.

2.2 Data Cleaning

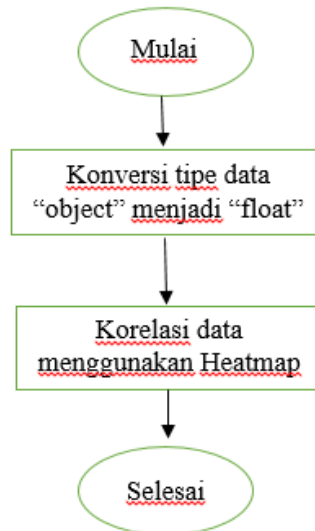
Tahapan ini merupakan tahapan untuk memastikan kebersihan, konsistensi, dan kualitas data sehingga dapat digunakan dan memberikan hasil akurat dalam analisis selanjutnya. Langkah-langkah dalam proses *Data cleaning* diantaranya sebagai berikut:

- 1) Menampilkan jumlah *missing value*

- 2) Menghapus data yang mengandung *missing value*
- 3) Menampilkan jumlah baris sebelum dan sesudah melakukan Data Cleaning

2.3 Pre Processing

Tahapan ini merupakan tahapan pengolahan data jika data tersebut memiliki permasalahan, proses yang dilakukan diantaranya sebagai berikut:



Gambar 2: Flowchart Pre processing

2.4 Processing

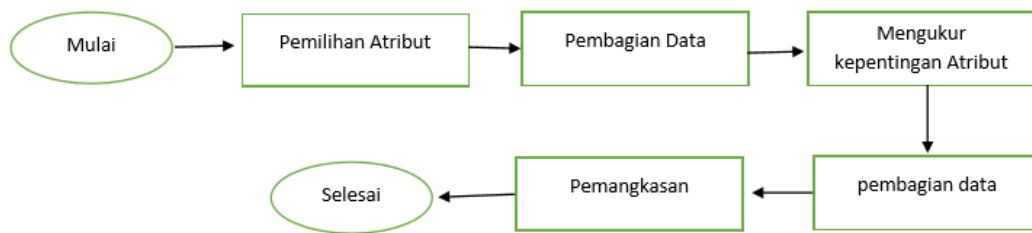
Tahapan ini merupakan tahap dimana data yang telah dikumpulkan untuk tujuan analisis lebih lanjut, langkah-langkah dalam proses ini diantaranya sebagai berikut:

- 1) *Balance dan Imbalance*
- 2) *Principal Component Analysis (PCA)*
PCA digunakan untuk mengurangi ukuran dan membantu dalam mempercepat waktu komputasi dan mengurangi kompleksitas perhitungan.
- 3) Membagi data menjadi data pelatihan dan pengujian

2.5 Pemodelan Algoritma

- 1) *Decision Tree*

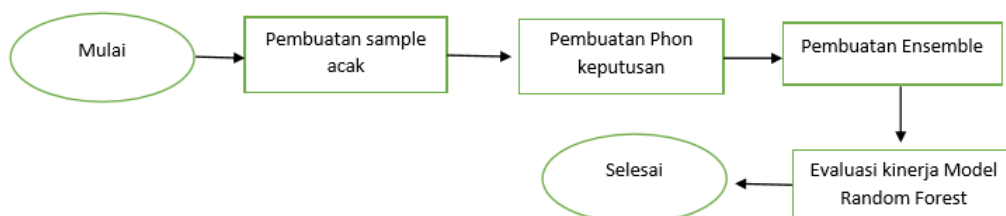
Setiap simpul (node) pohon mewakili sifat yang telah diuji dalam struktur pohon yang dikenal sebagai *Decision Tree*. Setiap cabang berfungsi sebagai pembagian hasil uji, dan node daun, atau daun, berfungsi sebagai representasi kelompok kelas tertentu. Dalam sebuah hierarki keputusan, tingkat node tertinggi adalah akar, atau akar, yang biasanya merupakan sifat yang memiliki dampak besar pada kelas tertentu. Untuk menyelesaikan masalahnya, *Decision Tree* biasanya menggunakan strategi pencarian dari bawah ke atas. Dalam metode klasifikasi nilai atribut, jalur dari node akar (akar) ke node akhir (bunga) baru akan dipelajari, setelah itu kelas baru akan ditentukan [13]. Berikut tahapan-tahapan dalam pemodelan algoritma *Decision Tree*:

Gambar 2. Tahapan Algoritma *Decision Tree*

- a. **Pemilihan Atribut**
Algoritma memilih atribut terbaik untuk membagi data pada setiap langkah pembangunan pohon. Atribut yang dipilih harus memiliki kemampuan terbaik untuk memisahkan data menjadi kelas yang homogen.
- b. **Pembagian data**
Data dibagi menjadi beberapa subset berdasarkan nilai atribut setelah atribut terbaik dipilih. Setiap simpul pohon mengalami proses ini.
- c. **Mengukur Kepentingan Atribut**
Setelah pembagian, masing-masing atribut dinilai berdasarkan kriteria tertentu, seperti Gain Informasi atau Gini Impurity. Atribut yang memberikan pembagian yang paling homogen, atau yang paling baik, akan memiliki nilai kepentingan yang tinggi.
- d. Untuk setiap simpul dalam pohon, proses pemilihan atribut dan pembagian data dilakukan secara rekursif hingga tercapai kondisi berhenti. Kondisi berhenti dapat berupa mencapai tingkat kedalaman maksimum, jumlah sampel minimum di simpul, atau ketika tidak ada atribut yang dapat membagi data dengan lebih baik.
- e. **Pemangkasan**
Setelah pohon dibuat, proses pemangkasan, juga dikenal sebagai pruning, dilakukan untuk menghilangkan simpul yang tidak penting atau menyebabkan overfitting pada data uji.

2) *Random Forest*

Satu set data latihan dikumpulkan dan dimasukkan ke dalam suatu pohon untuk memulai proses latihan. Dalam sebuah node, pemilihan fitur akan dibagi dan diambil secara acak. *Bagging* memilih sampel berulang kali, menggunakan penggantian [14]. Dalam proses perancangan model algoritma *Random Forest*, model algoritma *Random Forest* dibangun dengan menggunakan pohon keputusan "n" untuk menemukan nilai "n" terbaik. Selanjutnya, algoritma *Random Forest* menghasilkan hasil optimum [15]. Berikut merupakan table *Pseudocode* algoritma *Random Forest* [16]. Untuk tugas klasifikasi dan regresi, Algoritma Hutan Acak adalah teknik pembelajaran kelompok yang didasarkan pada gagasan membuat sejumlah besar pohon keputusan secara acak.

Gambar 3: Tahapan Algoritma *Random Forest*

2.6 Evaluasi

Pada tahapan evaluasi ini Untuk memahami sejauh mana algoritma dapat membantu menghasilkan nilai klasifikasi yang akurat dan berguna, digunakan juga parameter seperti akurasi, presisi, recall, dan F1-Score yang akan membantu mengukur kinerja model dalam menentukan model mana yang paling akurat dalam mengklasifikasikan nutrisi makanan cepat saji, berikut adalah rumus yang dapat dilihat dibawah ini :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan:

TP = True Positif (memprediksi positif dan itu benar).

TN = True Negatif (memprediksi negatif dan itu benar).

FP = False Positif (memprediksi positif dan itu salah).

FN = False Negatif (memprediksi negatif dan itu salah)

3. HASIL DAN PEMBAHASAN

Pada bagian ini membahas hasil dari pengujian yang dilakukan pada dataset *fastfood* menggunakan algoritma *Decision tree* dan *Random Forest*. Adapun proses yang dilakukan selama pengujian, sebagai berikut:

3.1 Pengumpulan Data

Pengumpulan data dari dataset nutrisi makanan cepat saji atau *Fastfood* pada situs *Kaggle.com* diperoleh dataset yang berjumlah 1147 dengan 14 kolom yang terdiri dari *Company, Item, Calories, Calories from, Total, Saturated, Trans Fat, Cholesterol, Sodium, Carbs, Fiber, Sugars, Protein, Weight*. Setelah tahapan pengumpulan data dilakukan dilanjutkan dengan tahapan *import* data. Tahapan *import* data merupakan tahapan pertama dalam proses analisis data dan melibatkan pemuatan dari berbagai sumber ke dalam lingkungan analisis data seperti *python*.

3.2 Data Cleaning

Pada tahapan ini yaitu melakukan pembersihan data, berikut tahapan dalam proses pembersihan data:

- 1) Menampilkan jumlah *missing value*

Pada tahapan pembersihan data diawali dengan melihat jumlah *missing value* dari dataset yang digunakan.

```

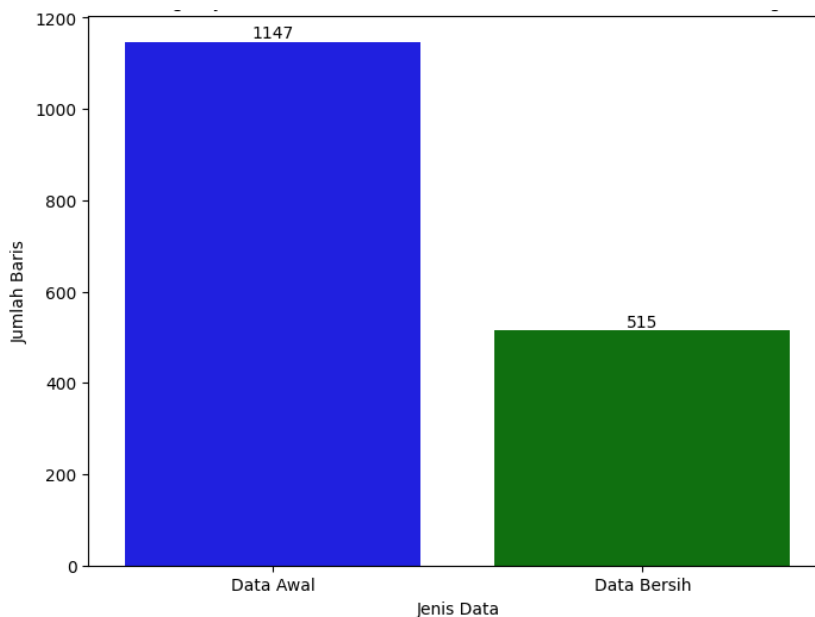
Jumlah Missing Value per Kolom:
Company                0
Item                   0
Calories               0
Calories from\Fat     505
Total Fat\N(g)        56
Saturated Fat\N(g)    56
Trans Fat\N(g)        56
Cholesterol\N(mg)     0
Sodium \N(mg)         0
Carbs\N(g)            56
Fiber\N(g)            56
Sugars\N(g)           0
Protein\N(g)          56
Weight Watchers\NPts  260
dtype: int64
    
```

2) Menghapus data yang mengandung *missing value*

Tabel 1: Dataset setelah menghapus Missing Value

Company	Item	Calories	Calo ries from Fat	Total Fat (g)	Saturat ed Fat (g)	Trans Fat (g)	Choles terol (mg)	Sodi um (mg)	Carbs (g)	Fiber (g)	Sugars (g)	Protein (g)	Weight Watchers Points
McDonald's	Hamburger	250	80	9	3.5	0.5	25	520	31	2	6	12	247.5
McDonald's	Cheeseburger	300	110	12	6	0.5	40	750	33	2	6	15	297
McDonald's	Double Cheeseburger	440	210	23	11	1.5	80	1150	34	2	7	25	433
McDonald's	McDouble	390	170	19	8	1	65	920	33	2	7	22	383
McDonald's	Quarter Pounder® with Cheese	510	230	26	12	1.5	90	1190	40	3	9	29	502
...
Burger King	BK® Café Mocha Frappe- 16 fl oz	400	90	10	7	0	40	270	68	1	58	9	456
Burger King	BK® Café Mocha Frappe- 20 fl oz	510	110	13	8	0	50	350	87	1	71	12	577
Burger King	BK® Café Caramel Frappe- 12 fl oz	300	80	8	5	0	35	210	50	0	41	7	339
Burger King	BK® Café Caramel Frappe- 16 fl oz	400	90	10	7	0	40	280	68	0	56	9	454
Burger King	BK® Café Caramel Frappe- 20 fl oz	500	110	12	8	0	50	350	86	0	71	11	568

3) Menampilkan jumlah baris sebelum dan sesudah melakukan Data Cleaning



Gambar 4: Barplot Data Awal dan Data Bersih

Dari Gambar 4 di atas menunjukkan bahwa *Barplot* data awal menampilkan distribusi atau frekuensi dari data kategorikal sebelum proses pembersihan atau pengolahan dilakukan yang belum rapi atau memiliki beberapa masalah seperti data yang hilang, duplikat atau kesalahan entri. dan *Barplot* data bersih menunjukkan distribusi atau frekuensi dari data setelah proses pembersihan atau pengolahan data dilakukan.

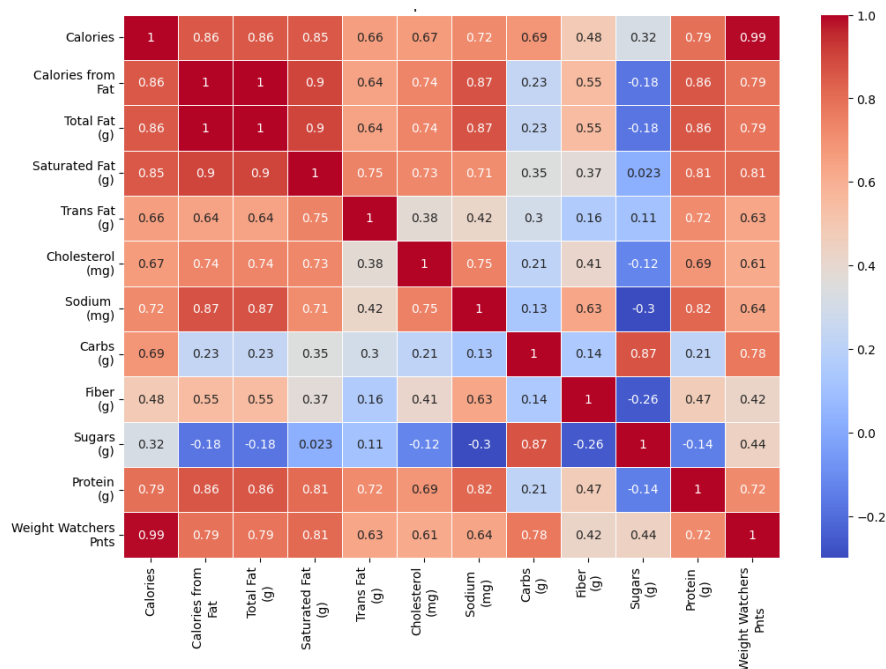
3.3 Pre Processing

Tahapan ini merupakan proses pengolahan data yang diperlukan jika terdapat permasalahan dalam dataset. Berikut adalah tahapan-tahapan *pre processing*.

- 1) Konversi type data “Object” menjadi “Float”
 - a. Membaca Dataset
 - b. Memeriksa dataset untuk mengetahui tipe data
 - c. Mengidentifikasi kolom-kolom yang memiliki type data “object”
 - d. Mengonversi kolom-kolom dengan type data “object” menjadi “float”
 - e. Konversi berhasil

Dalam pengolahan data, tipe data "objek" dapat merujuk pada tipe data yang biasanya menyimpan nilai non-angka, seperti string, teks, atau kombinasi beberapa jenis data. Misalnya, kolom dalam dataframe dapat berisi teks seperti nama, alamat, atau kategori. Namun, tipe data "float" adalah jenis data numerik yang digunakan untuk menyimpan nilai bilangan real. Dalam representasi komputer, floats diwakili sebagai angka pecahan dengan titik desimal, yang memungkinkan penggunaan nilai bulat dan non-bulat. Dalam analisis data, misalnya, kolom dengan tipe data "float" dapat berisi nilai-nilai seperti angka pecahan yang menunjukkan pengukuran, persentase, atau jenis nilai numerik lainnya. Konversi tipe data dari “object” menjadi “float” adalah salah satu tahapan yang biasa dilakukan dalam *pre processing* data sebelum di analisis lebih lanjut. Banyak algoritma pembelajaran mesin membutuhkan input dalam bentuk *numeric*, seperti algoritma klasifikasi. Dengan mengonversi tipe data “object” menjadi “float”, dapat membuat menjadi lebih sesuai untuk digunakan dengan algoritma-algoritma ini.

- 2) Korelasi data menggunakan Heatmap



Gambar 7: Heatmap Korelasi Nutrisi Makanan Cepat Saji

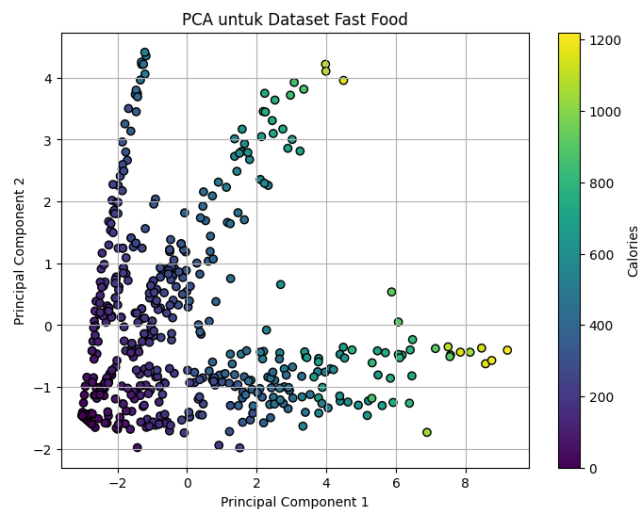
Dari hasil visualisasi heatmap diatas menunjukkan rentan nilai tertentu. Skala warna bergerak dari warna terang ke warna gelap untuk menunjukkan nilai yang lebih rendah. Dan warna-warna di antara keduanya menunjukkan rentan nilai di antara keduanya. Setiap sel dari heatmap diatas berisi nilai numeric yang mewakili kekuatan korelasi antara faktor-faktor dalam dataset. Nilai mendekati 1 menunjukkan korelasi positif yang kuat, sementara nilai yang mendekati -1 menunjukkan korelasi negative yang kuat. Garis diagonal dengan nilai 1.00 menunjukkan bahwa setiap faktor berkorelasi sempurna dengan dirinya sendiri. Gambar diatas menarik karena secara visual merangkum data nutrisi yang kompleks dan menunjukkan bagaimana komponen-komponen diatas berhubungan satu sama lain pada makanan cepat saji. Misalnya terdapat korelasi positif yang tinggi antara lemak total dan kalori atau lemak jenuh dan lemak trans.

3.4 Processing

1) Balance & Inbalance

Jumlah sampel dari setiap nilai yang ditampilkan pada kolom "Kalori" data set; misalnya, tiga sampel memiliki nilai kalori 740, dua sampel memiliki nilai 560, dan seterusnya. Ini adalah ringkasan jumlah kali masing-masing nilai muncul dalam dataset. Di antara nilai-nilai dalam kolom "Calories", distribusi sampel relatif seragam atau seimbang, artinya tidak ada satu nilai yang secara signifikan lebih dominan atau lebih jarang muncul daripada yang lain, menurut pernyataan "Dataset seimbang". Karena tidak ada bias yang signifikan terhadap kelas tertentu, hal ini dapat bermanfaat untuk melatih model.

2) Principal Component Analysis (PCA)



Gambar 3. PCA Fastfood

PCA merupakan metode statistic yang digunakan untuk mengurangi dimensi dari dataset yang kompleks, sementara mempertahankan sebanyak mungkin informasi yang mungkin. Ini berguna dalam menganalisis dataset besar dengan banyak fitur atau variabel. Untuk dataset nutrisi makanan cepat saji, PCA membantu dalam mengidentifikasi pola-pola yang tersembunyi dalam informasi tersebut. Misalnya memiliki dataset dengan banyak variabel yang mewakili nutrisi. Dengan mengurangi dimensi, PCA membantu menyederhanakan interpretasi informasi, yang dapat membantu dalam menganalisis lebih lanjut seperti klasifikasi makanan berdasarkan nilai nutrisi, mengidentifikasi pola makanan yang umum, atau bahkan dalam mengembangkan rekomendasi eat less yang lebih baik.

3) Data Training dan Data Testing

Jumlah Data Training : 403

Jumlah Data Uji : 112

Dalam konteks pengujian kinerja dua algoritma, data ini digunakan untuk membandingkan performa dari dua algoritma yang berbeda dalam mengklasifikasikan nutrisi makanan cepat saji.

3.5 Pemodelan Algoritma

1) *Decision Tree*

Dalam data mining dan pembelajaran mesin, algoritma *Decision Tree* adalah metode pembelajaran yang populer dan mudah digunakan. Dengan membagi data ke dalam subset yang semakin kecil berdasarkan fitur tertentu algoritma ini membangun sebuah pohon keputusan. Setiap simpul dalam pohon tersebut mewakili keputusan yang didasarkan pada fitur yang diuji dan setiap subdivisi dari pohon tersebut mewakili peluang hasil dari uji fitur tersebut. Decision tree adalah model prediktif yang menggunakan struktur berbentuk pohon untuk melakukan pengambilan keputusan berdasarkan serangkaian aturan logis. Dengan cara yang mirip dengan proses pengambilan keputusan manusia, model ini menunjukkan hubungan antara fitur input dan target output. Model membuat keputusan berdasarkan nilai fitur tertentu pada setiap simpul (node) dalam pohon keputusan. Setiap simpul mewakili pertanyaan atau tes tentang fitur tertentu, dan cabang-cabang dari simpul mewakili kemungkinan hasil dari tes tersebut. Akhirnya, daun-daun (daun) pohon berfungsi sebagai prediksi atau label model untuk data yang dimasukkan.

Tabel 2: Laporan Klasifikasi Algoritma Decision Tree

Label data	Precision	Recall	F1-Score	Support
120.0	0.00	0.00	0.00	1
190.0	0.00	0.00	0.00	1
210.0	0.00	0.00	0.00	0
310.0	0.00	0.00	0.00	1
540.0	0.00	0.00	0.00	0
560.0	0.00	0.00	0.00	1
630.0	0.00	0.00	0.00	0
660.0	0.00	0.00	0.00	1
740.0	0.00	0.00	0.00	0
860.0	0.00	0.00	0.00	1
1110.0	0.50	1.00	0.67	1
1130.0	0.00	0.00	0.00	1
1150.0	0.00	0.00	0.00	1
1190.0	0.00	0.00	0.00	0

2) *Random Forest*

Random Forest adalah salah satu algoritma ensemble yang paling populer dalam pembelajaran mesin, terutama untuk tugas klasifikasi dan regresi. Ini membangun beberapa pohon keputusan (decision trees) selama pelatihan dan menggabungkan prediksi dari setiap pohon untuk meningkatkan kinerja model secara keseluruhan. Random Forest didasarkan pada pohon keputusan. Setiap pohon di Random Forest adalah model pembelajaran independen yang mengambil keputusan berdasarkan serangkaian aturan yang diperoleh dari data. Pohon-pohon ini dibangun secara independen dengan menggunakan subset data yang diambil secara acak (bootstrap samples) dan subset fitur (random subsets of features).

Tabel 3: Sampel Acak menggunakan Algoritma Random Forest

	Calories	Total	Saturated	Trans	Cholesterol	Sodium	Carbs	Fiber	Sugars	Protein
Calories from Fat	Fat (g)	Fat (g)	Fat (g)	(mg)	(mg)	(g)	(g)	(g)	(g)	(g)
1110.0	240.0	26.0	16.0	2.0	100.0	350.0	194.0	0.0	168.0	25.0
660.0	360.0	40.0	12.0	1.5	90.0	980.0	49.0	2.0	11.0	28.0

	Calories	Total	Saturated	Trans	Cholesterol	Sodium	Carbs	Fiber	Sugars	Protein
Calories	from Fat	Fat (g)	Fat (g)	Fat (g)	(mg)	(mg)	(g)	(g)	(g)	(g)
560.0	240.0	27.0	9.0	0.5	260.0	1300.0	56.0	3.0	7.0	24.0
1150.0	710.0	79.0	31.0	3.5	240.0	2150.0	49.0	2.0	10.0	61.0
860.0	220.0	24.0	15.0	1.5	75.0	260.0	144.0	0.0	124.0	18.0

Tabel ini menunjukkan nilai untuk setiap variabel yang terkait dengan kumpulan data makanan cepat saji. Setiap baris menunjukkan satu sampel makanan cepat saji dengan nilai yang berkaitan dengan kalori, lemak, kolesterol, sodium, karbohidrat, serat, gula, dan protein.

3.6 Evaluasi

Salah satu metode klasifikasi yang paling sederhana dan mudah dipahami adalah Algoritma Decision Tree. Menggunakan berbagai metrik untuk menilai kinerja model dalam mengklasifikasikan data. Metri pertama yang kami pertimbangkan adalah akurasi. Dengan akurasi sebesar 55.56%, model Decision Tree menunjukkan prediksi kelas yang berhasil untuk sekitar setengah dari data uji yang digunakan. Namun, untuk mendapatkan pemahaman yang lebih baik, nilai akurasi ini harus dipertimbangkan bersama dengan metrik lainnya. Sebaliknya, model Random Forest adalah pendekatan ensemble yang didasarkan pada Decision Tree. Dalam evaluasi kinerjanya, ditemukan bahwa model Random Forest memiliki akurasi yang sedikit lebih tinggi, mencapai 66.67%. Ini menunjukkan bahwa, secara keseluruhan, model Random Forest memiliki kinerja yang lebih baik dalam memprediksi kelas data uji yang digunakan. Sifat ensemble dari algoritma ini, di mana ia menggunakan banyak data uji, dapat memberikan hasil yang lebih baik.

4. KESIMPULAN

Dengan mempertimbangkan penelitian yang telah dilakukan, dapat disimpulkan bahwa Dalam evaluasi kinerja Algoritma Decision Tree dan Random Forest pada dataset yang sama, Random Forest mengungguli Decision Tree dengan nilai akurasi 66.67%, sedangkan Decision Tree hanya mencapai 55.56%, menunjukkan bahwa Random Forest mampu memberikan prediksi yang lebih akurat untuk kelas data uji.

Selain itu, karakteristik kelompok Random Forest, di mana beberapa pohon keputusan digabungkan, memberikan keunggulan dalam menangani kompleksitas data dan meningkatkan generalisasi model. Hasil ini menunjukkan bahwa pembelajaran kelompok dapat meningkatkan kinerja dan keandalan prediksi dalam membangun model klasifikasi, terutama dalam kasus dataset yang kompleks.

Sebaliknya, evaluasi kinerja tidak hanya bergantung pada keakuratan. Untuk mendapatkan pemahaman yang lebih baik tentang kinerja model, metrik tambahan seperti ketepatan, recall, dan skor F1 juga harus dipertimbangkan. Untuk prediksi kelas dalam dataset fastfood, Random Forest akhirnya menjadi pilihan terbaik.

REFERENSI

- [1] Marini, H. S. Tambunan, Z. A. Siregar, A. P. Windarto, and F. Rizki, "Penerapan Data Mining Klasifikasi Gizi Bayi Dengan Algoritma Decision Tree C4.5," *ZAHRA Buletin Big Data*, vol. 1, no. 2, pp. 88–96, 2022.
- [2] A. P. Wahyu, H. Heryono, M. B. Chaniago, and D. Hamdani, "Smart Canteen : Perilaku Mengatur Pola Makan Dengan Membaca Nilai Nutrisi (Conventional Deep Learning Neural Network)," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 6, no. 2, pp. 115–121, 2020, doi: 10.33197/jitter.vol6.iss2.2020.356.
- [3] I. Pamela, "Fast Food Consumption Behavior in Adolescent and ITS Impact for Health," *Jurnal IKESMA*, vol. 14, no. 2, pp. 144–153, 2018.

- [4] B. P. Statistik, "Rata-rata harian konsumsi protein," 2023.
- [5] V. Cyrilla *et al.*, "Artikel Penelitian Hubungan Kebiasaan Mengonsumsi Makanan Cepat Saji dengan Indeks Massa Tubuh pada Mahasiswa Fakultas Kedokteran Ukrida Angkatan 2016 Periode September 2017 Relationship Between Habit of Fast Food Consumption Against Body Mass Index in St," no. September, 2017.
- [6] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
- [7] S. Alam, A. Siregar, and A. Juwita, "Penerapan Algoritme C4. 5 untuk Klasifikasi Kasus COVID-19," ... , *Technology and ...*, vol. III, 2022.
- [8] A. W. Wardhana, E. Patimah, A. I. Shafarindu, Y. M. Siahaan, B. V. Haekal, and D. S. Prasvita, "Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree," *Senamika*, vol. 2, no. 1, pp. 660–667, 2021.
- [9] S. Supangat, A. R. Amna, and T. Rahmawati, "Implementasi Decision Tree C4.5 Untuk Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun," *Teknika*, vol. 7, no. 2, pp. 73–78, 2018, doi: 10.34148/teknika.v7i2.90.
- [10] M. Mia, A. F. N. Masruriyah, and A. R. Pratama, "The Utilization of Decision Tree Algorithm In Order to Predict Heart Disease," *Jurnal Sisfotek Global*, vol. 12, no. 2, p. 138, 2022, doi: 10.38101/sisfotek.v12i2.551.
- [11] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis : Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [12] H. Suryono, H. Kuswanto, and N. Iriawan, "Rice phenology classification based on random forest algorithm for data imbalance using Google Earth engine," *Procedia Computer Science*, vol. 197, no. 2021, pp. 668–676, 2021, doi: 10.1016/j.procs.2021.12.201.
- [13] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
- [14] A. Hidayanti, A. M. Siregar, S. A. P. Lestari, and Y. C. Cahyana, "Model Analisis Kasus Covid-19 Di Indonesia Menggunakan Algoritma Regresi Linier Dan Random Forest," *Petir*, vol. 15, no. 1, pp. 91–101, 2021, doi: 10.33322/petir.v15i1.1487.
- [15] A. Samosir, M. S. Hasibuan, W. E. Justino, and T. Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung," *Prosiding Seminar Nasional Darmajaya*, vol. 1, no. 0, pp. 214–222, 2021.
- [16] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS : InDonEsiA journal Information System*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealis.v6i2.3036.