Jurnal IPTEK Bagi Masyarakat, Vol. 5, No. 1, (2025), 115-127 e-ISSN: 2807-7253
Doi: https://doi.org/10.55537/j-ibm.v5i1.1248 p-ISSN: 2807-7261

Sales Prediction of Kembar Fruit Salad Homemade Products Based on Transaction Data Using K-Nearest Neighbor

(Prediksi Penjualan Salad Buah Kembar Berbasis Produk Homemade Berdasarkan Data Transaksi Menggunakan K-Nearest Neighbor)

Anisa Rahman [a,1,*], Muhammad Siddik Hasibuan [a,2], Aidil Halim Lubis [a,3]

[a] *Universitas Islam Negeri Sumatera Utara, Medan, Indonesia*

E-mail: [1]anisasehati2@gmail.com; [2]muhammadsiddik@uinsu.ac.id; [3]aidilhalimlubis@uinsu.ac.id;

*Corresponding Author.
E-mail address: anisasehati2@gmail.com (A. Rahman).

Received: Juny 21, 2025 | Revised: August 31, 2025 | Accepted: August 31, 2025

Abstract: *K-Nearest Neighbor is a method used to classify data with the closest distance or what is called a lazy learning technique. Accurate sales predictions allow owners to plan raw material stock requirements more efficiently, thereby reducing the risk of overstocking (which can cause waste) or understocking (which has the potential to reduce revenue). This research will implement the KNN, K-Nearest Neighbor algorithm to show the extent to which the KNN method can produce accurate predictions in the context of sales. The stages of the method carried out in this study by determining the value of K, calculating the square of the euclid distance (query instance) of each object against the given training data, then sorting the objects into groups that have the smallest euclid distance, using the class label Y (nearest neighbor classification), using the k-nearest neighbor category that is the majority then the calculated query instance value can be predicted. This research produces a jupyter notebook-based sales prediction model that can be used to predict Twin Fruit Salad products based on variations. The dataset used consists of 112 data divided into 89 training data and 23 testing data. With 91% accuracy results, it contributes to increasing the turnover of twin fruit salad and can provide considerations regarding stock availability based on the amount sold.*
Keywords: *k-nearest neighbor; sales prediction; product classification; market trends.*

Abstrak: *K-Nearest Neighbor adalah metode yang digunakan untuk melakukan klasifikasi data yang jaraknya paling dekat atau yang disebut dengan teknik lazy learning. Prediksi penjualan yang akurat memungkinkan owner untuk merencanakan kebutuhan stok bahan baku secara lebih efisien, sehingga dapat mengurangi resiko kelebihan stok (yang dapat menyebabkan pemborosan) atau kekurangan stok (yang berpotensi menurunkan pendapatan). Pada penelitian ini akan mengimplementasikan algoritma KNN, K-Nearest Neighbor guna memperlihatkan sejauh mana metode KNN bisa menghasilkan prediksi yang akurat dalam konteks penjualan. Tahapan metode yang dilakukan pada penelitian ini dengan cara menentukan nilai K, menghitung kuadrat jarak euclid (query instance) masing-masing objek terhadap training data yang diberikan, kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil, menggunakan label class Y (klasifikasi nearest neighbor), dengan menggunakan kategori k-nearest neighbor yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung. Penelitian ini menghasilkan model prediksi penjualan berbasis jupyter notebook dapat digunakan untuk memprediksi produk Salad Buah Kembar berdasarkan variasi. Dataset yang digunakan terdiri dari 112 data yang dibagi menjadi 89 data training dan 23 data testing. Dengan hasil akurasi sebesar 91% memberikan konstribusi terhadap peningkatan omset Salad Buah Kembar serta bisa memberikan pertimbangan gambaran penyediaan stok berdasarkan jumlah yang terjual.*
Kata kunci: *k-nearest neighbor; prediksi penjualan; klasifikasi produk; tren pasar.*

Pendahuluan

Salah satu produk yang kini cukup populer di kalangan masyarakat, terutama di kota-kota besar, adalah Salad Buah. Produk ini diminati karena dianggap sebagai makanan sehat, segar, dan praktis (Adhi Putra, 2021). Namun, seperti produk makanan lainnya, salad buah memiliki karakteristik khusus, yakni mudah rusak dan memiliki masa simpan yang singkat. Oleh karena itu, memprediksi penjualan secara akurat sangat penting agar produsen tidak mengalami kerugian akibat kelebihan produksi atau kehilangan peluang karena kekurangan stok. Permasalahan yang sering dialami adalah pihak perusahaan masih mengalami kesulitan untuk memprediksi permintaan konsumen (Akbar, 2024). Salad buah mengandung bahan segar seperti buah dan krim yang mudah basi jika tidak disimpan dengan baik dan benar. Keunggulan dari Salad Buah Kembar adalah mayones atau krim diproduksi sendiri *(homemade)* tidak menggunakan bahan pengawet membuat mayones tidak memiliki daya tahan lama, sehingga saat produk tidak terjual seluruhnya mayones (krim) juga produk harus dibuang, yang pada akhirnya menimbulkan pemborosan bahan baku. Ketidakpastian permintaan juga berdampak pada ketidakefisienan dalam proses kerja (Ayuni & Fitrianah, 2020).

Dengan menerapkan teknik *data mining* di data transaksi penjualan, dapat digunakan untuk mengetahui *dessert* mana saja yang paling laris terjual, mengalami peningkatan atau penurunan (Darmawan et al., 2018). Hal ini bisa membantu dalam mendapatkan info baru yang bisa digunakan untuk mengatur persediaan barang serta memperbaiki proses pemesanan produk agar lebih efisien. Metode KNN memiliki keunggulan karena sederhana, mudah diimplementasikan, dan cukup efektif untuk berbagai kasus prediksi berbasis data historis (Lianda & Atmaja, 2021). Dalam konteks prediksi penjualan salad buah, metode ini dapat digunakan untuk mengidentifikasi pola penjualan berdasarkan variabel-variabel tertentu, seperti hari dalam seminggu, jumlah penjualan sebelumnya, atau kondisi cuaca. Dengan menggunakan KNN, pelaku usaha diharapkan dapat membuat estimasi jumlah produk yang perlu disiapkan setiap harinya, sehingga produksi menjadi lebih efisien dan tepat sasaran (Dewi et al., 2022).

Penelitian ini menerapkan algoritma K-Nearest Neighbor (KNN), yang merupakan bagian dari algoritma pembelajaran terawasi (*supervised learning*), di mana data baru diklasifikasikan berdasarkan mayoritas kategori dari data tetangga terdekatnya KNN. Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi. dengan memanfaatkan algoritma KNN (Elison et al., 2020). Dalam KNN terdapat metode klasifikasi yang fungsinya adalah menentukan kategori berdasarkan mayoritas kategori, dengan mencari kelompok objek pada data latih yang memiliki jarak paling dekat dengan objek pada data baru atau data pengujian. Penjualan merupakan syarat mutlak keberlangsungan suatu usaha, karena dengan penjualan maka akan didapatkan keuntungan. Semakin tinggi penjualan maka keuntungan yang akan didapat semakin besar. Untuk mencapai tujuan ini maka sangat diperlukan usaha-usaha agar konsumen mempunyai daya tarik dan sifat loyal dalam berbelanja di suatu unit usaha. Metode KNN memiliki keunggulan karena sederhana, mudah di implementasikan, dan cukup efektif untuk berbagai kasus prediksi berbasis data historis (Erdiansyah et al., 2022). Dalam konteks prediksi penjualan salad buah, metode ini dapat digunakan untuk mengidentifikasi pola penjualan berdasarkan variabel-variabel tertentu, seperti hari dalam seminggu, jumlah penjualan sebelumnya, atau kondisi cuaca. Dengan menggunakan KNN, pelaku usaha diharapkan dapat membuat estimasi jumlah produk yang perlu disiapkan setiap harinya, sehingga produksi menjadi lebih efisien dan tepat sasaran (Fauzi, 2020).

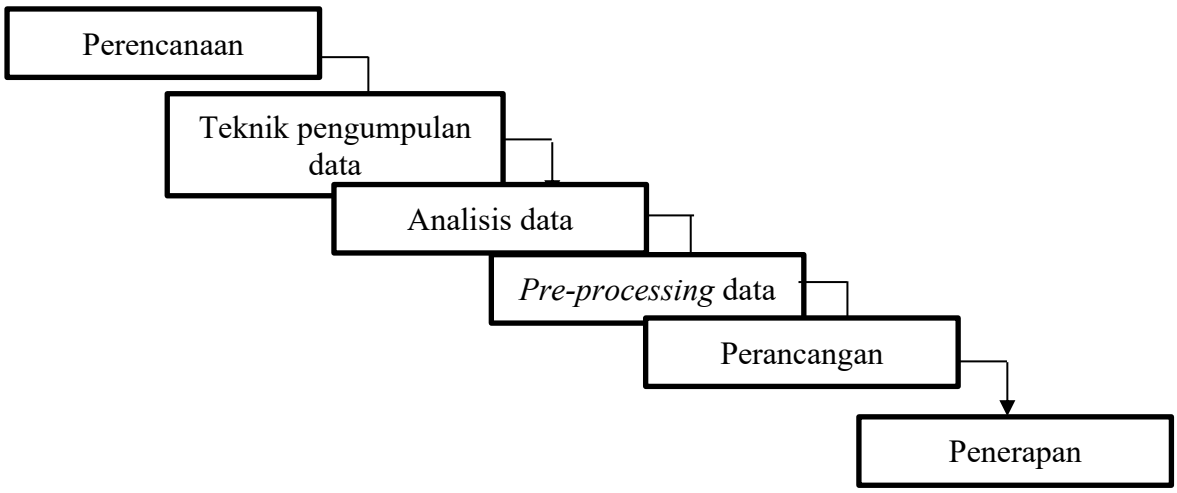
K-Nearest Neighbor (K-NN) dalam penelitian (Mardiyyah et al., 2024) adalah salah satu algoritma klasifikasi dalam *supervised learning* yang dikenal sederhana dan cukup populer di bidang *machine learning*. Metode ini telah banyak diterapkan dalam berbagai bentuk prediksi data, seperti dalam pengenalan aktivitas manusia pada pembelajaran daring, serta sebagai teknik klasifikasi untuk data terkait penyakit kanker. Diantara penelitian yang dimaksud

yaitu: 1) penelitian (Alfani W.P.R. et al., 2021) K-NN yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. 2) (Anisa & Andri, 2020) yang menggunakan algoritma K-NN dengan proses klasifikasi yang akan menghasilkan nilai akurasi sesuai dengan nilai *k* yang digunakan pada saat pengolahan klasifikasi data.

Tujuan dari dilaksanakan penelitian ini adalah untuk mengetahui kategori penjualan produk di outlet Salad Buah Kembar, dengan memprediksi produk Salad Buah Kembar agar stok persediaaan produk di outlet Salad Buah tersebut tidak cepat habis, serta menghasilkan model KNN dalam memprediksi produk Salad Buah Kembar berdasarkan variasi (Hayami et al., 2021).

Metode

Pada tahap ini proses penelitian ini dilakukan dalam beberapa tahapan yaitu studi literatur, teknik pengumpulan data, *pre-processing* data, perancangan, pengujian atau penerapan.



Gambar 1. Kerangka Penelitian penerapan.

Pada Gambar 1. ada beberapa tahapan yaitu studi literatur, teknik pengumpulan data, *pre-processing* data, perancangan dan penerapan pengujian atau dalam tahap perencanaan hal yang dilakukan adalah menentukan topik, menentukan objek penelitian, perumusan masalah, penentuan judul dan penentuan tujuan.

Teknik Pengumpulan Data

Adapun teknik dari pengumpulan data yaitu dengan melakukan observasi, wawancara dan studi pustaka:

1. Observasi
   Yaitu dengan pengumpulan data dengan pengamatan secara langsung dalam proses Prediksi Penjualan Produk Melalui Eksplorasi Data Dengan Metode *K-Nearest Neighbor* dengan segala aspek yang berhubungan langsung dengan penelitian.
2. Wawancara
   Yaitu dengan melakukan kegiatan wawancara untuk mencari informasi dengan menggunakan tanya jawab dengan pihak yang berwenang mengenai Penjualan Produk Salad Buah Terlaris. Data yang digunakan untuk dilakukan Metode *K-Nearest Neighbor*.
3. Studi Pustaka
   Melakukan studi pustaka mengenai teori-teori dan konsep yang berhubungan dengan penelitian, seperti teori mengenai keamanan data dengan menggunakan Metode *K-Nearest Neighbor*. Referensi yang digunakan penelitian adalah buku, jurnal ilmiah *online* dan situs.

KKN (K-Nearest Neighbor)

KNN (*K-Nearest Neighbor*) termasuk algoritma *supervised learning,* yang mana hasil dari *query instance* baru, diklasifikasikan berdasarkan mayoritas dari kategori pada *K-Nearest Neighbor* (Kafil, 2020). Kelas yang paling banyak muncul, yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan *training* data. Metode *K-Nearest Neighbor* ini merupakan metode algoritma *machine learning* yang sangat sederhana dalam implementasinya (Asyrofi & Asyrofi, 2023). Dengan diterapkannya algoritma *K-Nearest Neighbor* dapat mempermudah penjualan produk dengan mengambil data objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut (Dewi et al., 2022).

Rumus Jarak *Euclidiean*

$(x,y) = \sqrt{\sum_{i=1}^{n}(x_{i-}y_i)^2}$

Dist ((x,y),(a,b)) = $\sqrt{(x-a)^2 + (y-b)^2}$

Algoritma akan menemukan *K*-tetangga terdekat dari titik data untuk nilai *K* yang diberikan dan akan menetapkan kelas ke titik data tersebut dengan menempatkan kelas yang memiliki jumlah data terbanyak dari semua kelas tetangga *K*. Input *x* yang ditetapkan ke kelas dengan probabilitas terbesar.

Pre-Processing Data

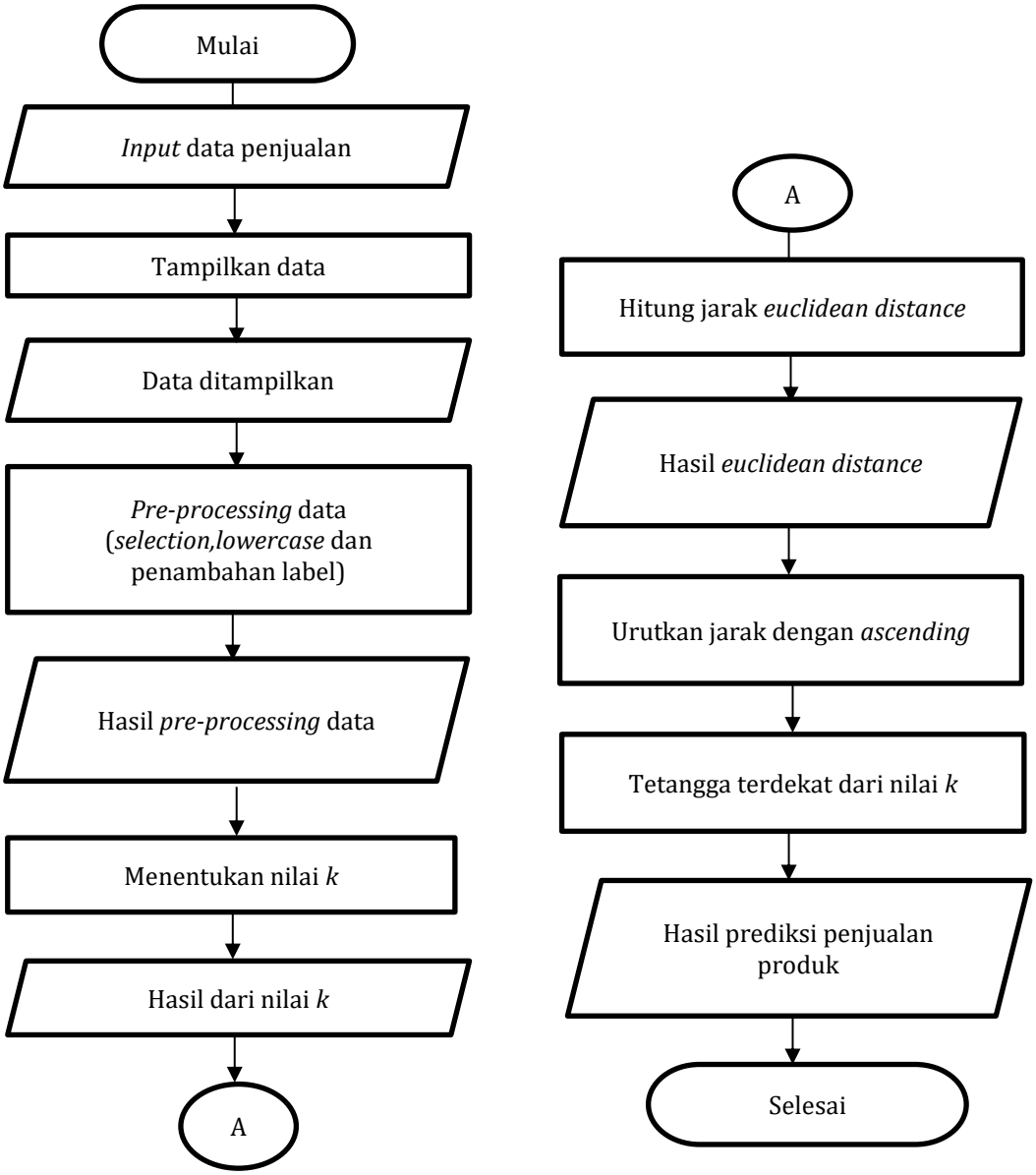
Setelah mengumpulkan data dan analisis kebutuhan, selanjutnya dilakukan tahap *pre-processing* data melalui *Jupyter Notebook* dengan Python dimana terdiri dari beberapa tahapan sebagai berikut:

1. Data *Selection,* pada tahap ini data penjualan produk salad buah bulan september-november 2024 diseleksi menjadi beberapa atribut saja, semula ada No, Tanggal Transaksi, Judul Produk, Status Pemesanan, Harga Dasar, Nama Pelanggan dan kuantitas, menjadi hanya Nama Produk, Harga dan kuantitas produk (Ma'arif, 2020).
2. Lowercase, berfungsi untuk mengubah seluruh huruf yang diblok menjadi huruf kecil. Huruf besar disebut juga *uppercase* atau kapital *letters.* Huruf kapital digunakan sebagai unsur pertama kata pada awal kalimat (National Cancer Institute, 2020). Pada tahap ini semua data set pada bagian judul produk di ubah menjadi huruf kecil.
3. Data *Cleaning,* pada tahap ini dilakukan pembersihan data duplikasi, dan missing value atau nilai yang hilang pada data (Nugraha et al., 2020). Langkah awal yang dilakukan untuk *cleaning* data adalah proses pembersihan data duplikasi.
4. *Transformation*, proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat utuk melakukan proses *mining* dengan cara melakukan peringkasan (agregasi).
5. Penambahan label, pada tahap ini menambahkan baris baru yaitu label untuk menjadi nilai aktual (Ritonga & Muhandhis, 2021). Nilai label diambil dari data produk salad buah 3 bulan terakhir di tahun 2024.

Perancangan

Setelah *pre-processing* dilakukan tahapan proses perancangan sistem diantaranya yaitu mengidentifikasi kebutuhan aplikasi, fungsi aplikasi, serta merancang *flowchart* pada aplikasi (Setiyani et al., 2020). Perancangan sistem adalah sebuah kegiatan merancang dan menentukan cara mengolah sistem informasi dari hasil analisa sistem sehingga dapat memenuhi kebutuhan dari pengguna termasuk aktivitas-aktivitas proses. Dalam penelitian ini, atribut yang dipilih seperti jenis produk dan jumlah order yang terjual, dan target penjualan bulan selanjutnya dihitung berdasarkan rata-rata penjualan bulan September

hingga November (Syahril et al., 2020). Total dari dataset yang digunakan kemudian data tersebut nantinya digunakan untuk data latih dan data uji. Selanjutnya menghitung jarak terkecil berdasarkan mayoritas nilai. Nilai *k* yang digunakan yaitu *k* =5. Adapun *flowchart* yang sudah dirancang seperti dibawah ini.



Gambar 2. *Flowchart* Algoritma

Pada Gambar 2. Flowchartnya berdasarkan rata-rata penjualan bulan September hingga November dan nilai *k* yang digunakan yaitu *k* =5. Penerapan algoritma *K-Nearest Neighbor* dilakukan di *Jupyter Notebook* dengan bahasa Python menggunakan data penjualan sebuah perusahaan salad buah untuk mengetahui penjualan produk terlaris salad buah tahun selanjutnya berdasarkan variabel yang telah ditentukan.

Hasil

Analisis Data

Pada penelitian ini menggunakan data dengan periode 3 bulan terakhir diantaranya September, Oktober dan November 2024 (Ziarahah et al., 2023). Penelitian ini akan menggunakan algoritma KNN dalam melakukan prediksi penjualan produk salad buah.

Pre-processing Data

1. Data *selection*

Data yang digunakan dalam penelitian ini yaitu data penjualan produk Salad Buah berdasarkan penjualan 3 bulan terakhir dari bulan September, Oktober dan November 2024 yang berasal dari Toko Salad Buah Kembar (Hamidi et al., 2023). Kemudian data tersebut diseleksi terlebih dahulu dengan cara menyeleksi kolom atribut yang peneliti perlukan saja dan akan digunakan untuk diolah dalam memprediksi penjualan salad buah terlaris. Adapun atribut yang terdapat dalam dataset berupa No, Nomor Pesanan, Tanggal Transaksi, Judul Produk, Status Pemesanan, Status Pemenuhan, Harga Dasar, Nama Pelanggan dan Kuantitas.

Tabel 1. Dataset Salad Buah

No	Nomor Pesanan	Tanggal Transaksi	Judul Produk	Status Pemesanan	Status Pemenuhan	Harga Produk	Nama Pelanggan	Kuantitas
1	#22-15475	12/06/20 24 14:34	BLACK SALAD LARGE	Completed	Fulfilled	44000	aira	1
2	#22-15474	12/06/20 24 12:04	RUJAK KU'INI	Completed	Fulfilled	25000		1
...	.....	.....	.....	.....	.....	.....	.....	.....
4052	F-2588221691	09/07/20 24 16:26	Salad original large			39000		1
4053	F-2587941404	09/07/20 24 12:48	Ximilu medium			25000		2

Data tersebut diseleksi dan menjadi beberapa atribut yang peneliti perlukan saja yaitu Judul Produk, Harga Dasar dan Kuantitas terjual dalam 3 bulan terakhir yaitu September, Oktober dan November.

Tabel 2. Hasil Seleksi Data

No	Judul Produk	Harga Produk	Kuantitas
1	BLACK SALAD LARGE	44000	1
2	RUJAK KUINI	25000	1
3	SALAD ORIGINAL X-TRA SMALL + MILSHAKE TARO	21000	1
4	SANDWICH FRUITS (SANDO)	17000	1
5	ES TELER	27000	1
.....	.....	.....	.....
4051	Asinan mix medium	28000	1
4052	Salad original large	29000	1
4053	Ximilu medium	35000	2

2. *Lowercase*

Lowercase adalah karakter huruf yang ditulis dengan huruf kecil. *Lowercase* berfungsi untuk mengubah seluruh huruf yang diblok menjadi huruf kecil. Huruf besar disebut juga uppercase atau kapital *letters*.

Tabel 3. Hasil Lowercase

No	Judul Produk	Harga Produk	Kuantitas
1	black salad large	44000	1
2	rujak ku'ini	25000	1
3	salad original x-tra small + milshake taro	21000	1
4	sandwich fruits (sando)	17000	1
5	es teler	27000	1
.....	.....	.....	.....
4051	asinan mix medium	28000	1
4052	salad original large	29000	1
4053	ximilu medium	35000	2

3. Data *Cleaning*

Pada tahap ini dilakukan pembersihan data duplikasi, dan *missing value* atau nilai yang hilang pada data. Salah satu permasalahan dalam kualitas data adalah data hilang (*missing value*). *Missing value* adalah suatu permasalahan dimana pada bagian data terdapat data yang tidak lengkap atau hilang.

```
Total Dataset   : 4053
[10]: Judul Produk    0
      Harga Dasar     0
      Kuantitas       0
      dtype: int64
```

Gambar 3. Missing Value.

Pada Gambar 3. Selanjutnya peneliti mengecek nilai yang kosong atau dinamakan dengan *missing value*, total dataset 4053 artinya tidak terdapat nilai yang kosong dari keseluruhan dataset yang ada.

```
Total dataset                          :  4053
Dataset bersih dari missing value      :  4053
```

Gambar 4. Rincian Missing Value

Pada Gambar 4. Hasil dari pembersihan data kosong yang dilakukan di *Jupyter Notebook* dengan bahasa Python yaitu dataset bersih dari *missing value*. Total dataset yang bersih dari *missing value* sebanyak 4053 atau hampir seluruh dataset bersih dari *missing value* (Ben et al., 2025). Fungsi dari pembersihan data-data kosong ini agar proses *training* dapat dilakukan. Selanjutnya akan dilakukan proses duplikasi data, duplikasi data dalam data *mining* dapat terjadi karena dua hal yaitu adanya *record* yang berulang dan adanya perbedaan identifikasi antara entitas yang sama dalam dunia nyata. Adanya duplikasi data pada dataset dapat mempengaruhi kualitas performa data *mining.*

Tabel 4. Pembersihan Duplikasi Data

No	Judul Produk	Harga Produk	Kuantitas
1	salad original x-tra small + milshake greentea	21000	11
2	black salad toping coklat small	29000	1
3	black salad x-tra small	17000	2
4	salad buah toping mix coklat + greentea xtra large 750 ml	79000	1
5	salad buah toping mix keju + coklat large 550 ml	59000	2
.....	.....	.....	.....
110	ximilu medium	25000	226
111	ximilu small	20000	72
112	ximilu x-tra large	40000	76

4. Pemberian Label (*labelling*)

Pada tahap ini menambahkan baris baru yaitu label untuk menjadi nilai aktual. Nilai label diambil dari berapa banyak jumlah terjual dan kemudian dikategorikan sebagai cukup laris, laris dan terlaris.

Tabel 5. Tabel Kategori

Jumlah Terjual/Perhari	Kategori
1-3	Cukup Laris
4-7	Laris
>8	Terlaris

Data dari kategori tersebut juga di dapat oleh peneliti melalui wawancara langsung dengan *owner* Salad Buah Kembar bahwa jika jumlah terjual salad minimal berjumlah 1 cup perhari maka sudah bisa dikategorikan sebagai kategori cukup laris.

Tabel 6. Penambahan Label

No	Judul Produk	Harga Produk	Kuantitas	Label
1	salad original x-tra small + milshake greentea	21000	11	Terlaris
2	black salad toping coklat small	29000	1	Cukup Laris
3	black salad x-tra small	17000	2	Cukup Laris
4	salad buah toping mix coklat + greentea xtra large 750 ml	79000	1	Cukup Laris
5	salad buah toping mix keju + coklat large 550 ml	59000	2	Cukup Laris
.....	.....	.....	.....	.....
110	ximilu medium	25000	226	Terlaris
111	ximilu small	20000	72	Terlaris
112	ximilu x-tra large	40000	76	Terlaris

Data Mining

Proses ini merupakan tahapan unutk mencari kecocokan dari pola dan informasi menarik dalam data yang terpilih dengan menggunakan metode *K-Nearest Neighbor* berdasarkan proses *pre-processing* secara keseluruhan, karena jumlah dataset yang cukup banyak dan tentunya akan mempengaruhi proses perhitungan manual menjadi panjang, maka penulis hanya memakai 10 data *training* dan 5 data *testing* yang akan dijadikan sebagai proses perhitungan manual.

Tabel 7. Data Training

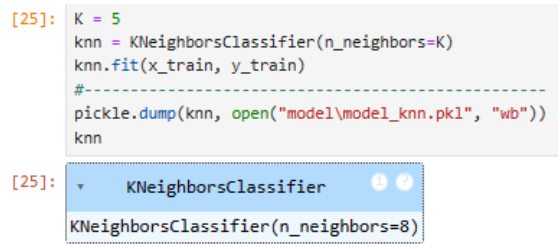
No	Judul Produk	Harga Produk	Kuantitas	Label
1	salad buah toping mix keju + greentea large	56000	1	Terlaris
2	salad original x-tra large	49000	430	Cukup Laris
3	salad buah toping mix coklat + greentea x- tra large	75000	1	Cukup Laris
4	black salad medium	39000	12	Cukup Laris
5	salad original toping green tea large	39000	40	Cukup Laris
....	.....	.....	.....	.....
87	salad original x-tra small + milkshake taro	21000	20	Terlaris
88	mayo taro - top keju	7000	10	Terlaris
89	smoothies king avocado	25000	13	Terlaris

Tabel 8. Data Testing

No	Judul Produk	Harga Produk	Kuantitas	Label
1	lemon tea	15000	9	?
2	salad buah toping mix keju + coklat x-tra large	75000	5	?
3	salad buah toping mix keju + coklat large 550 ml	59000	2	?
4	mayo greentea - top greentea	7000	1	?
5	lumpia ayam	30000	13	?
.....	.....	.....	.....	.....
21	fruit jelly ball small	22000	29	?
22	asinan jambu small	21000	9	?
23	extra almond	9000	3	?

Implementasi Algoritma K-Nearest Neighbor

Setelah dilakukannya penjumlahan pada data *training* dan data *testing*, selanjutnya data dapat diproses dengan KNN. *Training K-Nearest Neighbor* dilakukan di *Jupyter Notebook* menggunakan pemograman python 3. Proses KNN dimulai dengan *k* = 5 dan *import K-Nearest Neighbor* yang mendefinisikan jumlah tetangga terdekat (*K*) yang akan digunakan dalam algoritma *K-Nearest Neighbor* (KNN). Sedangkan knn.fit (x_train, y_train) merupakan melatih model KNN menggunakan data latih (x_train) dan labelnya (y_train).



Gambar 5. Training K-Nearest Neighbor.

Pada Gambar 5. Proses pelatihan ini membuat model yang dapat memprediksi kelas data baru dengan membandingkan jaraknya terhadap data latih.

1. Evaluasi Model

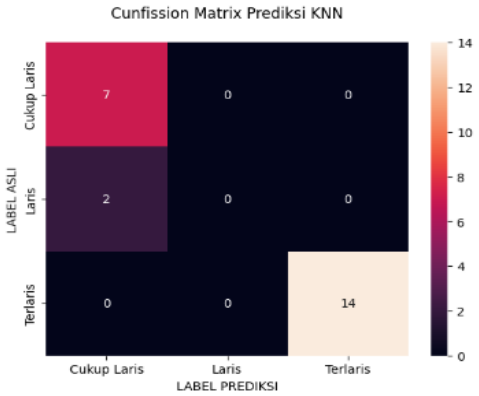
Berdasarkan model yang telah dikembangkan dari 80% data *training* dan 20% data *testing* menggunakan algoritma *K-Nearest Neighbor* dengan penggunaan nilai *K*=5. Dapat dievaluasi bahwasannya mendapatkan hasil prediksi sebesar 0,78%, Recall 100%, dan f1-Sore 0,88%. Hasil ini dapat dijadikan pertimbangan untuk mengkategorikan jika model

yang dikembangkan ini tergolong sangat baik dalam memprediksi data.

```
Akurasi Prediksi : 0.9130434782608695

[[ 7  0  0]
 [ 2  0  0]
 [ 0  0 14]]

              precision    recall  f1-score   support

 Cukup Laris       0.78      1.00      0.88         7
       Laris       0.00      0.00      0.00         2
    Terlaris       1.00      1.00      1.00        14

    accuracy                           0.91        23
   macro avg       0.59      0.67      0.62        23
weighted avg       0.85      0.91      0.88        23
```

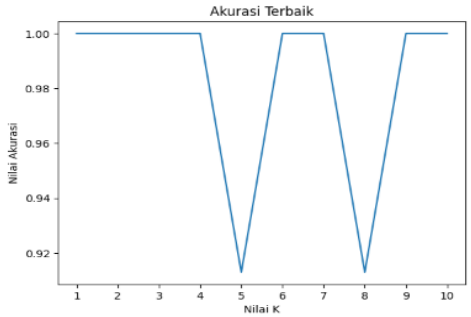
Gambar 6. Tampilan Akurasi Prediksi.

Pada gambar 6. Dapat dievaluasi bahwasannya mendapatkan hasil akurasi prediksi sebesar 0,91%, Recall 100%, dan f1-Sore 0,88%. Selanjutnya, untuk mempermudah dalam pemahaman confsusion matrix, peneliti membuat kembali confussion matrix ke dalam bentuk gambar yang rinci beserta penjelasan label yang berhasil diprediksi, berdasarkan confussion matrix data yang berhasil diprediksi ke dalam label Cukup Laris sebesar 4 data, ke dalam label Laris 0 data, dan ke dalam label Terlaris 10 data.



Gambar 7. *Cunfission Matrix Prediksi KNN.*

Pada Gambar 7.Berhasil diprediksi ke dalam label Cukup Laris sebesar 4 data, ke dalam label Laris 0 data, dan ke dalam label Terlaris 10 data. Selanjutnya, peneliti menampilkan tingkat akurasi berdasarkan penggunaan nilai *K*, adapun percobaan nilai *K* peneliti mulai dari *K*=1 sampai *K*=10, yang mana setiap penggunaan nilai *K* tersebut menghasilkan akurasi yang berbeda. Dengan demikian, peneliti bisa mengevaluasi dari nilai *K*=1 sampai *K*=10 berapa akurasi yang terbaik di dapatkan, tujuan dilakukannya cara ini adalah peneliti ingin mengetahui penggunaan nilai *K* terbaik yang menghasilkan akurasi yang tinggi.



Gambar 8. Tampilan Akurasi Terbaik.

Pada Gambar 8.Akurasinya cenderung sangat tinggi 100% untuk sebagian besar nilai k. Namun, ada dua titik di mana akurasi turun cukup rendah, yaitu sekitar K = 5 dan K = 8, dengan akurasi sekitar 0.91%. Menunjukkan hubungan antara nilai K dan nilai akurasi dalam sebuah model pembelajaran mesin yang berkaitan dengan algoritma K-Nearest Neighbors (K-NN). Sumbu x adalah nilai k yang menunjukkan nilai parameter k dalam algoritma K-NN, yang bernilai dari 1 hingga 10 sedangkan Sumbu y adalah nilai akurasi yang menunjukkan tingkat akurasi model berdasarkan masing-masing nilai k. Akurasinya cenderung sangat tinggi 100% untuk sebagian besar nilai k. Namun, ada dua titik di mana akurasi turun cukup rendah, yaitu sekitar K = 5 dan K = 8, dengan akurasi sekitar 0.91%. Arti dari bentuk grafik menyerupai huruf "W" adalah beberapa nilai k menghasilkan performa buruk sementara lainnya sangat optimal.

Diskusi

Pada penelitian sebelumnya yang dilakukan oleh (Dewi et al., 2022) dengan menggunakan metode KNN untuk penerapan data *mining* prediksi penjualan produk terlaris, objek yang diteliti pada penelitian ini adalah Toko UD Andar, dapat disimpulkan bahwa perhitungan dengan teknik data *mining* dan algoritma *k-nearest neighbor* di dapatkan hasil prediksi dengan nilai akurasi yang tinggi, dengan menerapkan metode k nearest neighbor kedalam sebuah sistem aplikasi maka dapat membantu UD tersebut dalam menyelesaikan masalah yang dihadapi, sehingga teknik data mining dan metode algoritma k nearest neighbor ini dapat diimplementasikan untuk memprediksi penjualan produk terlaris pada UD Andar.

Dan juga pada penelitian terdahulu dari (Alfani W.P.R. et al., 2021) penelitian ini menggunakan metode *K-Nearest Neighbor* (K-NN) untuk memprediksi penjualan produk Unilever di sebuah toko retail semi grosir. Berdasarkan hasil perhitungan data *mining* menggunakan teknik klasifikasi dan algoritma K-NN, didapatkan hasil prediksi penjualan produk berdasarkan nilai akurasi tertinggi dan terendah. Nilai akurasi tertinggi terhadap klasifikasi penjualan produk sebesar 86,66%. Sedangkan nilai akurasi terendah terhadap klasifikasi penjualan produk sebesar 40%. Data penjualan dari tahun 2017-2019 digunakan untuk melakukan prediksi, dengan demikian metode data *mining* dan algoritma K-NN ini dapat diimplementasikan untuk memprediksi penjualan produk Unilever di Toko Rizky Barokah Nganjuk.

Perbedaan penelitian sebelumnya dengan penelitian ini ialah berbeda tempat penelitian, produk, waktu, dan juga data yang digunakan. Penelitian ini yang akan diprediksi ialah ketersediaan stok penjualan salad menggunakan data penjualan 3 bulan terakhir di tahun 2024 yaitu September, Oktober dan November yang mana penelitian ini akan memprediksi penjualan produk salad buah melalui eksplorasi data dengan metode *K-Nearest Neighbor.* Pengujian yang dilakukan yaitu dengan jumlah 112 data, pembagian data 80% (89) data *training* dan 20% (23) data *testing*. Mengindikasikan bahwa model yang telah dibangun memiliki 91% tingkat akurasi, yang menandakan keberhasilan pemodelan dalam melakukan prediksi penjualan produk Salad Buah Kembar.

Kesimpulan

Berdasarkan dari penelitian yang dilaksanakan, diperoleh kesimpulannya bahwa sistem prediksi produk Salad Buah Kembar dalam melakukan prediksi dengan menggunakan algoritma KNN dapat membantu perusahaan untuk meningkatkan penjualan dan menambah stok serta mengetahui produk Salad Buah yang paling banyak di beli dan mampu meminimalisir terjadinya kerugian pada perusahaan Salad Buah Kembar. Model prediksi penjualan produk ini bisa memberikan pertimbangan serta gambaran penyediaan stok berdasarkan jumlah terjual. Pengujian yang dilakukan menghasilkan akurasi sebesar 91% yang memberikan konstribusi terhadap peningkatan omset Salad Buah Kembar serta bisa memberikan pertimbangan gambaran penyediaan stok berdasarkan jumlah yang terjual.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Universitas Islam Negeri Sumatera Utara, Medan, Indonesia, atas dukungan akademik dan fasilitas penelitian yang diberikan selama proses penelitian ini. Ucapan terima kasih juga disampaikan kepada Outlet Salad Buah Kembar sebagai lokasi penelitian yang telah menyediakan data dan informasi berharga sehingga penelitian ini dapat diselesaikan dengan baik.

Pernyataan Konflik Kepentingan

Penulis menyatakan bahwa tidak ada konflik kepentingan dalam kegiatan dan penulisan artikel ini.

Daftar Pustaka

Adhi Putra, A. D. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *8*(2), 636–646. https://doi.org/10.35957/jatisi.v8i2.962

Akbar, F. M. N. (2024). Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air. *Jurnal Tekno Kompak*, *18*(1), 28. https://doi.org/10.33365/jtk.v18i1.3241

Alfani W.P.R., A., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *6*(1), 155–160. https://doi.org/10.29100/jipi.v6i1.1910

Anisa, C., & Andri. (2020). Penerapan Algoritma k-Nearest Neighbor untuk Prediksi Penjualan Obat pada Apotek Kimia Farma Atmo Palembang. *Bina Darma Conference on Computer Science*, 199–208.

Asyrofi, R. R., & Asyrofi, R. (2023). Implementasi Aplikasi Jupyter Notebook Sebagai Analisis Kreteria Plagiasi Dengan Teknik Simantik. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 8(2), 627-637.

Ayuni, G. N., & Fitrianah, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika*, *14*(2), 79–86.

Ben, A., Amri, Y., Fnaiech, A., & Sahli, H. (2025). Journal of Electronic Science and Technology Automated ECG arrhythmia classi fi cation using hybrid CNN-SVM architectures. *Journal of Electronic Science and Technology*, *23*(3), 100316. https://doi.org/10.1016/j.jnlest.2025.100316

Darmawan, A., Kustian, N., & Rahayu, W. (2018). Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, *2*(3), 299. https://doi.org/10.30998/string.v2i3.2439

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 639–648.

https://doi.org/10.47065/bits.v3i4.1408

Elison, M. H., Asrianto, R., & Aryanto. (2020). Prediksi Penjualan Papan Bunga Menggunakan Metode Double Exponential Smoothing. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *2*(3), 45–56. https://doi.org/10.52005/jursistekni.v2i3.60

Erdiansyah, U., Irmansyah Lubis, A., & Erwansyah, K. (2022). Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil. *Jurnal Media Informatika Budidarma*, *6*(1), 208. https://doi.org/10.30865/mib.v6i1.3373

Fauzi, J. R. (2020). Algoritma Dan Flowchart Dalam Menyelesaikan Suatu Masalah Disusun Oleh Universitas Janabadra Yogyakarta 2020. *Jurnal Teknik Informatika*, *20330044*, 4–6.

Hamidi, A. A., Robertson, B., & Ilow, J. (2023). A new approach for ECG artifact detection using fine-KNN new approach for ECG artifact detection using classification and wavelet scattering features in vital health classification and wavelet scattering features in vital health applications applications. *Procedia Computer Science*, *224*, 60–67. https://doi.org/10.1016/j.procs.2023.09.011

Hayami, R., Sunanto, & Oktaviandi, I. (2021). Penerapan Metode Single Exponential Smoothing Pada Prediksi Penjualan Bed Sheet. *Jurnal CoSciTech (Computer Science and Information Technology)*, *2*(1), 32–39. https://doi.org/10.37859/coscitech.v2i1.2184

Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors. *Jurnal Mahasiswa Teknik Informatika (JATI)*, *3*(2), 59–66.

Lianda, D., & Atmaja, N. S. (2021). Prediksi Data Buku Favorit Menggunakan Metode Naïve Bayes (Studi Kasus: Universitas Dehasen Bengkulu). *Pseudocode*, *8*(1), 27–37. https://doi.org/10.33369/pseudocode.8.1.27-37

Ma'arif, A. (2020). Buku Ajar Pemrograman Lanjut Bahasa Pemrograman Python Oleh : Alfian Ma ' Arif. *Universitas Ahmad Dahlan*, 62.

Mardiyyah, N. W., Rahaningsih, N., & Ali, I. (2024). Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial. *Jurnal Mahasiswa Teknik Informatika*, *8*(2), 1491–1499.

National Cancer Institute. (2020). *Pseudocode – Definition (v1)*. Qeios. https://doi.org/10.32388/tf77dy

Nugraha, A.R.D, Auliasari, K., & Agus Pranoto, Y. (2020). IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus : BFI Finance Surabaya). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *4*(2), 14–20. https://doi.org/10.36040/jati.v4i2.2656

Ritonga, A. S., & Muhandhis, I. (2021). Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (Pca). *Edutic - Scientific Journal of Informatics Education*, *7*(2). https://doi.org/10.21107/edutic.v7i2.9247

Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review. *Faktor Exacta*, *13*(1), 35. https://doi.org/10.30998/faktorexacta.v13i1.5548

Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, *3*(1), 118. https://doi.org/10.53513/jsk.v3i1.202

Ziarahah, L. I., & Anwar, R. (2023). Akad Mudharabah Dan Relevansinya Dengan Tafsir Qur'an Surah an-Nisa Ayat 29 Tentang Larangan Mencari Harta Dengan Cara Yang Bathil. *Equality: Journal of Islamic Law (EJIL)*, *1*(1), 26-38. https://doi.org/10.15575/ejil.v1i1.480