



PENERAPAN DATA MINING UNTUK PREDIKSI PRODUK TERLARIS MENGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN) (STUDI KASUS : KOPI ANANDA)

Rezi Maharani^{1*}, Joni Karman², Novi Lestari¹

^{1,2}Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau

³Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau

e-mail: *¹Rezimaharani03@gmail.com, ²joni_karman@univbinainsan.ac.id,
³Novilestari@univbinainsan.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui hasil implementasi data mining algoritma *K-Nearest Neighbor* (K-NN) di kopi ananda kota lubuklinggau. Seiring makin banyaknya data dari hasil aktifitas manusia yang di bantu teknologi informasi, maka di perlukan sistem untuk dapat menghasilkan informasi dan pengetahuan yang terkandung dalam limbah data tersebut. Dari tumpukan data tersebut dapat di olah lah secara terkomputeri sasi sehingga memudahkan para produsen untuk memproduksi kopi yang paling di minati konsumen itu sendiri. Dengan melakukan data mining prediksi menggunakan algoritma *K-Nearest Neighbor* yang di lakukan peneliti mendapatkan nilai akurasi yang di dapat adalah 100% di peroleh dengan nilai K=4 mendapatkan precision 1.00, recall 1.00, dan score 2 sehingga mendapatkan nilai akurasi 100%, di lihat dari grafik penjualan juga bahwa kopi hitam lah yang memiliki nilai tertinggi dan dapat di lihat juga kopi yang kurang laris kopi pinang muda. Pola yang di temukan ini dapat di gunakan sebagai penjunjung keputusan menejer dalam mengelola aktifitas penjualan.

Kata Kunci : data mining, Prediksi, data penjualan, *K-Nearest Neighbor*

Abstract

This research aims to determine the results of implementing the data mining algorithm K-Nearest Neighbor (K-NN) at Ananda Coffee, Lubuklinggau City. As more and more data from the results of human activities are assisted by information technology, a system is needed to be able to produce the information and knowledge contained in this wealth of data. From this pile of data, it can be processed in a computerized manner, making it easier for producers to produce the coffee that consumers themselves are most interested in. By carrying out data mining predictions using algorithms K-Nearest Neighbor What the researchers did was get an accuracy value of 100% with a value of K=4 precision 1.00, recall 1.00, and score 2 to get an accuracy value of 100%, looking at the sales graph, it is black coffee that has the highest value and you can also see that the less popular coffee is young areca nut coffee. The patterns found can be used to support managers' decisions in managing sales activities.

Keywords : data mining, Prediction, sales data, *K-Nearest Neighbor*



I. PENDAHULUAN

Indonesia merupakan pengeksport kopi terbesar keempat di dunia (Chu Zhang, 2018). Bagi negara Indonesia hal tersebut hanya berfokus ke kuantitas saja pada kenyataannya kualitas kopi Indonesia sudah diakui dari berbagai pihak di beragam negara. Sebagai komoditas andalan ekspor, kopi Indonesia sudah mampu dijadikan sebagai salah satu devisa negara. Hal ini dapat dilihat dari pembangunan perkebunan kopi yang semakin hari semakin meningkat. Dalam hal ini pembangunan perkebunan komoditas kopi didominasi oleh perkebunan rakyat lebih dari 90% dan sisanya dari perkebunan besar swasta atau negara sebesar 10% (Informatika & Komputer, 2022). Kopi Ananda adalah produsen pembuatan kopi yang sudah dikenal luas masyarakat Kota Lubuklinggau sebagai Kopi Bubuk yang berasal dari Lubuklinggau (info wong linggau, 2023), karena cita rasa kopi yang harum dan nikmat serta diolah dari biji kopi pilihan sehingga Kopi Bubuk Ananda mendapatkan tempat di hati konsumennya. Dalam kiprah bisnisnya yang sudah memasuki 23 tahun (Berdiri sejak tahun 2000) kini Kopi Bubuk Ananda telah banyak mengalami transformasi. Beberapa inovasi yang telah dilakukan dengan meracik kopi dengan varian lainnya sehingga tercipta brand baru seperti Kopi Pinang Muda, Kopi Kulit Manggis, Kopi Pasak Bumi, Kopi Durian, Kopi Jahe, Kopi Hitam. Di kopi ananda memiliki 31 orang karyawan (Themebeez, 2022). Penghasilan yang diperolehnya + 15.000.000.00 rupiah perbulannya. Biasanya produk yang diproduksi perharinya + 50 pcs sehingga perbulannya kadang sekitar + 400 pcs. Dari banyaknya produk yang terjual maka diperlukan proses untuk mengetahui produk yang laris dan tidak laris menggunakan data

mining. Data mining adalah proses menggali sebuah informasi yang sangat dibutuhkan saat menelusuri data penjualan dan permintaan kopi terbaik oleh rakyat Indonesia. Dalam hal ini data mining sangat berpengaruh dalam proses menggali sebuah data dari suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual (Yolanda & Fahmi, 2021). Sebuah istilah untuk penambangan data Temukan pengetahuan tersembunyi di database. Kemudian. Akan tetapi untuk menentukan produk telaris ini dibutuhkan sebuah algoritma *K-Nearest Neighbor* agar mempermudah Kopi Ananda untuk mengetahui produk-produk mana saja dan varian-varian mana saja yang banyak diminati masyarakat Indonesia. Dengan menggunakan Algoritma *K-Nearest Neighbor* (KNN) dalam melakukan klasifikasi terhadap objek maka dibutuhkan jarak terdekat dengan objek tersebut (Yulia et al., 2018). Algoritma *K-Nearest Neighbor* (K-NN) merupakan metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain, metode K-NN merupakan metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi (Yunita, n.d.). Tujuan dari penerapan algoritma *K-Nearest Neighbor* dapat mengklasifikasikan objek baru berdasarkan atribut dan training sample (Putra & Putri, n.d.).

II. METODOLOGI PENELITIAN

Pada penelitian ini Menggunakan jenis metode kuantitatif dimana metode jenis ini membantu para peneliti memecahkan masalah secara logis untuk menghasilkan data statistik di setiap langkah, menampilkan prosedur matematika untuk membantu menemukan solusi masalah dan menyajikan informasi



dalam bentuk digital/angka (Silvilestari, 2021)

1. Algoritma *k-nearest neighbor*

Penulisan menggunakan metode pengujian *Euclidean* yang terdapat pada metode *k-nearest neighbor*. Berikut adalah rumus menghitung jarak dengan *Euclidean* seperti di bawah ini (Dewi & Rahayu, 2022):

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2 + (x_i - y)^2 + \dots + \dots = \dots}$$

Keterangan Rumus :

X1 : nilai yang ada data training

X2 : nilai yang ada data testing

K merupakan dimensi atribut

Langkah-Langkah perhitungan *K-Nearest Neighbor* (KNN) Sebagai berikut:

1. Menentukan peramet K (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak *Euclidean* objek terhadap data training yang di berikan
3. Selanjutnya mengeluarkan hasil no 2 secara *ascending* (berurutan dari nilai tinggi kerendah)
4. Mengumpulkan kategori Y (Klasifikasi *Nearest Neighbor* berdasarkan nilai k) dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksi objek yang baru.(k)
5. pada algoritma *K-Nearest Neighbor* adalah banyaknya tetangga terdekat yang digunakan sebagai titik untuk melakukan klasifikasi pada data atau objek baru. Untuk menghitung jarak antar objek data pada algoritma *K-*

Tabel 1 Penjualan

Bln	Nama Produk	Kuantitas produk	Kuantitas Terjual	Kategori
1	Kopi Hitam ukuran 200 g	14560	10720	Laris
2	Kopi	9640	5770	Laris

	Durian ukuran 200 g			
3	Kopi Pinang Muda ukuran 200 g	6740	3955	Tidak Laris
4	Kopi Jahe ukuran 200 g	7630	5002	Laris
5	Kopi Kulit Manggis ukuran 200 g	8490	4389	Tidak Laris
6	Kopi Pasak Bumi ukuran 200 g	7950	5430	Laris

Berdasarkan transformasi data tersebut, langkah selanjutnya yaitu melakukan normalisasi data dengan menggunakan normalisasi min max. dengan menggunakan normalisasi min max ini maka hasil akurasi menjadi lebih tepat dibandingkan tanpa menggunakan normalisasi. Berikut rumus pada normalisasi min max yaitu :

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Keterangan :

X' : nilai skala baru

X : nilai data asli sebelum dilakukan normalisasi

Mim : nilai terendah dari data

Max : nilai tertinggi dari data

Berikut penyelesaian perhitungannya yang dibuat dalam bentuk tabel:

Nilai mim untuk kualitas produk : 6740

Nilai max untuk kuantitas produk : 14560

Nilai mim untuk kualitas terjual : 3955

Nilai mix untuk kuantitas terjual : 10720



Tabel 2. Data Normalisasi Min Max

No	Min Max Kuantitas Produk	Hasil	Min Max Kuantitas Penjualan	Hasil
1	14560-6740/14560-6740=	1	10720-3955/10720-3955 =	1
2	9640-6740/14560-6740=	0,371	1150-3955/10720-3955 =	0,2684
3	6740-6740/14560-6740=	0	3955-3955/10720-3955 =	0
4	7630-6740/14560-6740=	0,114	5002-3955/10720-3955 =	0,155
5	84900-6740/14560-6740=	0,22	4389-3955/10720-3955 =	0,064
6	7950-6740/14560-6740=	0,15	5430-3955/10720-3955 =	0,218

Proses penghitungan K tetangga terdekat adalah sebagai berikut:

1. Tentukan parameter K = jumlah tetangga terdekat.
2. Menghitung jarak antara data uji dengan seluruh data latih pada tahap transformasi menggunakan perhitungan jarak Euclidean Euclidean digunakan untuk menghitung jarak Euclidean

Tabel 3. Hasil dari Jarak Euclidean Distance

No	Nama Produk	Kuantitas produk	Kuantitas Terjual	Jarak Euclidean
1	Kopi Hitam ukuran 200 g	2550	2100	$\sqrt{(1 - 0,371)^2 + (1 - 0,064)^2} = 1,503$

2	Kopi Durian ukuran 200 g	2130	1150	$\sqrt{(0,371 - 0,3)^2 + (0,2684 - 0,064)^2} = 0,0418$
3	Kopi Pinang Muda ukuran 200 g	2340	910	$\sqrt{(0 - 0,371)^2 + (0 - 0,064)^2} = 0,3751$
4	Kopi Jahe ukuran 200 g	2220	1520	$\sqrt{(0,114 - 0,3)^2 + (0,155 - 0,064)^2} = 0,2653$
5	Kopi Kulit Manggis ukuran 200 g	1630	709	$\sqrt{(0,224 - 0,3)^2 + (0,064 - 0,064)^2} = 0,147$
6	Kopi Pasak Bumi ukuran 200 g	2060	1240	$\sqrt{(0,155 - 0,3)^2 + (0,218 - 0,064)^2} = 0,2397$

3. Mengurutkan hasil secara ascending(dari nilai tinggi ke rendah)

Tabel 4. Urutan Ascending

No	Jarak Euclidean	Rangking
1	$\sqrt{(1 - 0,371)^2 + (1 - 0,064)^2} = 1,503$	6
2	$\sqrt{(0,371 - 0,371)^2 + (0,2684 - 0,064)^2} = 0,0418$	2
3	$\sqrt{(0 - 0,371)^2 + (0 - 0,064)^2} = 0,3751$	5
4	$\sqrt{(0,114 - 0,371)^2 + (0,155 - 0,064)^2} = 0,2653$	4

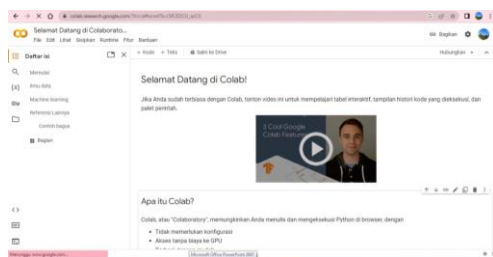
5	$\sqrt{(0,224 - 0,371)^2}$ $+ (0,064 - 0,064)^2$ $= 0,147$	1
6	$\sqrt{(0,155 - 0,371)^2}$ $+ (0,218 - 0,064)^2$ $= 0,2397$	3

Berdasarkan renting di atas kita bisa melihat bahwa bahwa renting 1 Kopi Pinang Muda adalah renting terkecil dan 6 kopi hitam renting terbesar sehingga “kopi hitam” produk terlaris. Produk yang paling di minati konsumen.

III. HASIL DAN PEMBAHASAN

1. Hasil

Dalam penelitian ini, implementasi dilakukan dengan menggunakan Python dan *Google Colab*. Untuk memulai program Python di *Google Colab*, diperlukan *library*. *Library* adalah kumpulan kode yang dapat digunakan berulang kali dalam program yang berbeda. Saat melakukan analisis sentimen dengan Python, beberapa *library* yang harus di pasang adalah *csv*, *pandas*, *numpy*, *sklearn*. pemasangan *library* dapat dilakukan dengan menggunakan perintah "*!pip install*". Setelah terinstal, *library* dapat diimpor dengan menggunakan perintah "*import*".



Gambar 1. Tampilan python google colab

1. Tahap Pengumpulan Data

Dalam penelitian ini, penulis mengumpulkan data secara langsung dari pihak pemilik kopi ananda penulis diberikan data penjualan selama 3 tahun

dari tahun 2020 – 2022, selanjutnya hasil data yang sudah dibersihkan diunduh dan di simpan dalam bentuk file *csv* (*comma delimited*) untuk dapat digunakan pada tahap selanjutnya.

2. Tahap Pengolahan Data

a) Tahap Pengimputan Data

Data yang sudah di simpan dalam bentuk *csv* sekarang di import ke *google colab*.

```
[ ] import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, classification_report

[ ] data = pd.read_csv('/content/DATA PENJUALAN ANANDA 3TH - Sheet1 (1).csv')

print(data.head())
print(data.describe())
```

	Nama Barang	Stok	harga	Terjual
0	Kopi Hitam Ukuran 200 g	14560	35.0	10720
1	Kopi Durian Ukuran 200 g	9640	40.0	5770
2	Kopi Pinang muda ukuran 200 g	6740	35.0	3955
3	Kopi Jahe Ukuran 200 g	7630	35.0	5002
4	Kopi Kulit manggis ukuran 200g	8490	35.0	4389

	Stok	harga	Terjual
count	6.000000	6.000000	6.000000
mean	9168.333333	35.833333	5877.666667
std	2810.703945	2.041241	2463.321227
min	6740.000000	35.000000	3955.000000
25%	7710.000000	35.000000	4542.250000
50%	8220.000000	35.000000	5216.000000
75%	9352.500000	35.000000	5685.000000
max	14560.000000	40.000000	10720.000000

Gambar 2. Import Data Dalam Bentuk Csv

b) Pembagian Data Training dan Data Testing

Data Training adalah Kumpulan data yang digunakan untuk melatih model atau algoritma pembelajaran mesin. Data ini memberikan model informasi tentang pola dalam data dan memungkinkan model mempelajari dan beradaptasi dengan pola tersebut. Sedangkan data testing adalah kumpulan data yang digunakan untuk menguji performa model setelah melatih model menggunakan data pelatihan. Data pengujian digunakan untuk mengukur seberapa baik model yang dilatih dapat membuat prediksi atau melakukan tugas yang diberikan padanya.

Tabel 5. Data Training

Stok	Harga
14560	35.000
9640	40.000
6740	35.000
7630	35.000
8490	35.000

7950	35.000
------	--------

Tabel 6 Data Testing

Terjual
10720
5770
3955
5002
4389
5430

```
[4] X = data.iloc[:, [1, 2]].values
     y = data.iloc[:, -1].values

print(X)

[[14560.  35.]
 [ 9640.  40.]
 [ 6740.  35.]
 [ 7630.  35.]
 [ 8490.  35.]
 [ 7950.  35.]]

[7] print(y)

[10720  5770  3955  5002  4389  5430]
```

Gambar 3. Pembagian data Training dan Data Testing

2. Fitur

Pada tahap pembagian fitur yaitu di sini memiliki 2 fitur yaitu 'Stok' dan 'harga'.

```
features = data[['Stok', 'harga',]]
```

Gambar 4 Pembagian fitur

3. Normalisasi Data

Pada tahap normalisasi ini Proses mengubah skala fitur atau variabel dalam suatu dataset sehingga nilainya berada

```
[11] from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)

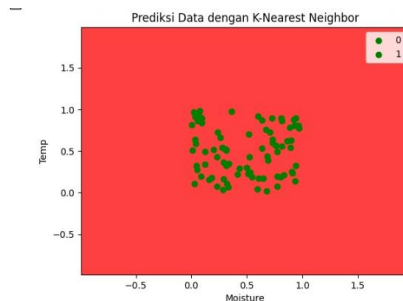
[12] print(X_train)
     print(X_test)
     print(y_train)
     print(y_test)

Stok harga
5 7950 35.0
2 6740 35.0
4 8490 35.0
3 7630 35.0
Stok harga
0 14560 35.0
1 9640 40.0
5 1
2 1
4 1
3 1
Name: Terjual, dtype: int64
0 1
1 1
Name: Terjual, dtype: int64
```

Gambar 5. Normalisasi Data

4. Metode K-Nearest Neighbors (KNN)

Pada tahap penerapan metode K-Nearest Neighbors (KNN) Nilai k yang di gunakan adalah 4 sehingga mendapatkan precision 1.00, recall 1.00, dan score 2 sehingga mendapatkan nilai akurasi 100%.



Gambar 6. Grafik K-Nearest Neighbors

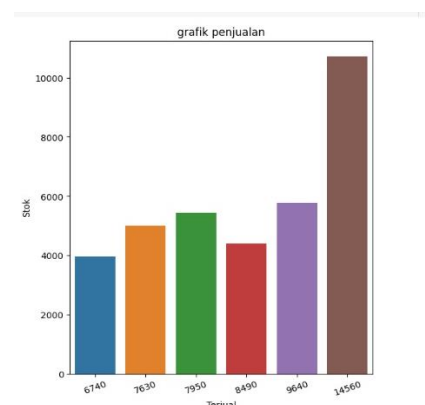
5. Prediksi

Pada tahap prediksi laris atau tidak laris didapatkan bahwa produk 1 di prediksikan laris . Dan juga di predisi ini juga memiliki grafik yang menentukan produk mana yang paling laris, pada grafik tersebut dapat dilihat bahwa grafik berwarna coklat itu adalah produk kopi hitam, warna ungu kopi durian, warna merah kopi kulit manggis, warna hijau kopi pasak bumi, warna kuning kopi jahe, warna biru pinang muda.

```
[ ] new_data = pd.DataFrame({'Stok': [50], 'harga': [100]})
     new_data_scaled = scaler.transform(new_data)
     prediction = knn_model.predict(new_data_scaled)
     print(f'Prediction: {"Laris" if prediction[0] == 1 else "Tidak Laris"}')

Prediction: Laris
```

Gambar 7. Prediksi



Gambar 8. Grafik Penjualan



2. Pembahasan

Penelitian ini dilakukan untuk mengetahui prediksi produkterlaris di kopi ananda menggunakan metode *K-Nearest Neighbors* (KNN). Untuk tahap pertama dalam penelitian ini adalah dengan melakukan observasi, dokumentasi dan wawancara sehingga mendapatkan data penjualan dari 2020 – 2022, setelah data penjualan di kumpulkan lalu di lakukan pembersihan, tahap selanjutnya melakukan *preprocessing* data atau mempersiapkan data untuk di gunakan. Adapun tahap – tahap dalam *preprocessing* seperti melakukan *import* data menentiukan data treinin dan data testing, pembagian fitur, normalisasi data, penerapan Metode *K-Nearest Neighbors* (KNN), eveluasi model, dan prediksi. Data ulasan yang sudah melalui tahap *preprocessing* disimpan dalam format file CSV (*Comma Delimited*). Hasil dari *preprocessing* didapatkan penerapan metode *K-Nearest Neighbors* (KNN). Nilai K yang di gunakan adalah 4 sehingga mendapatkan precision 1.00, recall 1.00, dan score 2 sehingga mendapatkan nilai akurasi 100%, langkah selanjutnya tahap prediksi laris atau tidak laris didapatkan bahwa produk 1 di prediksi laris . Dan juga di predisi ini juga memiliki grafik yang menentukan produk mana yang paling laris, pada grafik tersebut dapat dilihat bahwa grafik bewarna coklat itu adalah produk kopi hitam, warna ungu kopi durian, warna merah kopi kulit manggis, warna hijau kopi pasak bumi, warna kuning kopi jahe, warna biru pinang muda. Dan dapat di lihat bahwa kopi hitam adalah produk terlaris.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan tentang data mining prediksi produk terlaris menggunakan metode *K-Nearest Neighbors*

(KNN) studi kasus kopi ananda, dapat disimpulkan bahwa Metode *K-Nearest Neighbors* (KNN) dapat digunakan untuk memprediksi produk terlaris. Dapat dibuktikan dengan nilai akurasi yang di dapat adalah 100% di peroleh dengan nilai K=4 mendapatkan precision 1.00, recall 1.00, dan score 2 sehingga mendapatkan nilai akurasi 100%, di lihat dari grafik penjualan juga bahwa kopi hitam lah yang memiliki nilai tertinggi dan dapat di lihat juga kopi yang kurang laris kopi pinang muda.

V. SARAN

Untuk meningkatkan hasil pada penelitian selanjutnya berikut adalah saran yang dapat diterapkan untuk mendapatkan hasil yang lebih maksimal. Untuk peneliti selanjutnya di sarankan menggunakan metode klasifikasi lain, penerapan algoritma *K-Nearest Neighbors* di python sangat praktis manun perlu dilakukan perbandingan dengan algortma lainnya

VI. DAFTAR PUSTAKA

- Chu Zhang, F. L. &Yong H. (2018). *Identifikasi varietas biji kopi menggunakan pencitraan hyperspectral: pengaruh metode preprocessing dan analisis spektra pixel-bijaksana*. Laporan Ilmiah. <https://www.nature.com/articles/s41598-018-20270-y>
- Dewi, S. P., & Rahayu, E. (2022). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor*. 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- info wong linggau. (2023). *Kopi Ananda terpilih sebagai stand terbaik di Linggau Expoo 2023*. <https://infowonglinggau.com/kopi-ananda-terpilih-sebagai-stand-terbaik->



- di-linggau-expoo-2023/
Informatika, P. S., & Komputer, F. I.
(2022). *Aplikasi Prediksi Penjualan Kopi dengan Metode Single Exponential Smoothing untuk Mengetahui Produk Kopi Terlaris*. 8(200), 0–4.
- Putra, M. Y., & Putri, D. I. (n.d.). *Pemanfaatan Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI*. 16(2), 176–187.
- Silvilestari, S. (2021). Data Mining Menggunakan Algoritma K-Nearest Neighbor Dalam Menentukan Kredit Macet Barang Elektronik. *Jurnal Media Informatika Budidarma*, 5(3), 1063.
<https://doi.org/10.30865/mib.v5i3.3100>
- Themebeez. (2022). *Kopi ANANDA Kini Dengan Slogan Baru “Belum Ke Linggau Kalau Belum Ngopi ANANDA.”*<https://www.inilahkito.com/kopi-ananda-kini-dengan-slogan-baru-belum-ke-linggau-kalau-belum-minum-kopi-ananda/>
- Yolanda, I., & Fahmi, H. (2021). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti Terlaris Pada PT . Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor*. 3(3), 9–15.
- Yulia, B., Fahik, L., Djahi, B. S., Rumlaklak, N. D., & Komputer, J. I. (2018). Data Mining Untuk Klasifikasi Status Gizi Desa Di Kabupaten Malaka Menggunakan Metode K-Nearest Neighbor. *J-Icon*, 6(1), 1–7.
<https://ejurnal.undana.ac.id/jicon/article/view/348>
- Yunita, F. (n.d.). *DIABETES MELLITUS MENGGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN)*. 223–230.