Informatics and Computer Engineering Journal
Volume 3 No. 2 Agustus 2023

Implementasi Bahasa Phyton Untuk Clusterisasi Data Penjualan Menggunakan Metode K-Means

Corie Mei Hellyana[1]
[1,]Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Indonesia
e-mail: [1]corie.cma@bsi.ac.id

Artikel Info : Diterima : 0**7**-0**8**-**2023** | Direvisi : 00-00-0000 | Disetujui : **10-08-2023**

Abstrak - Dalam sebuah bisnis atau perdagangan, proses untuk menjaga stok barang merupakan salah satu cara untuk memberikan kepuasan pelanggan. Untuk memenuhi hal tersebut, pemilik bisnis harus dapat menganalisa barang atau produk mana yang laku, cukup laku dan kurang laku dipasaran, bukan hal mudah apabila bisnis tersebut memiliki ratusan bahkan ribuan produk yang dijual setiap bulannya. Permasalahan tersebut dapat dipecahkan dengan membuat clusterisasi produk penjualan dengan menggunakan metode K-Means. Dalam proses kalsterisasi menggunakan bahasa pemrograman Phyton yang berorintasi objek yang menghasilkan 3 cluster, dimana 1 cluster untuk produk yang laku, 2 cluster berikutnya untuk produk yang cukup laku dan kurang laku. Hasil clusterisasi ini nantinya dapat digunakan oleh pemilik bisnis dalam menjaga stok barang dan menetapkan strategi penjualannya.

Kata Kunci : clustering, data mining, penjualan

Abstracts **-** *In a business or trade, the process of keeping stock of goods is one way to provide customer satisfaction. To fulfill this, business owners must be able to analyze which goods or products are selling well, selling well and not selling well in the market, not an easy thing if the business has hundreds or even thousands of products sold every month. This problem can be solved by clustering sales products using the K-Means method. In the clustering process using the Python programming language which is object oriented which produces 3 clusters, where 1 cluster is for products that sell well, the next 2 clusters for products that are selling well and not selling well. The results of this clusterization can later be used by business owners in maintaining inventory and establishing sales strategies.*

Keywords: clustering, data mining, sales

PENDAHULUAN

Dalam dunia bisnis/perdagangan saat ini menuntut berbagai kalangan untuk dapat mengembangkan bisnis agar mampu bertahan dalam persaingan dunia bisnis/perdagangan. Untuk mencapai hal tersebut dibutuhkan peningkatan kualitas produk, penambahan jenis produk dan pengurangan biaya operasional perusahaan. Dari sisi manajemen, dibutuhkan sistem informasi yang akan membantu dalam memberikan keputusan yang tepat bagi keberlangsungan perusahaan. Maka dari itu, penting bagi perusahaan untuk menganalisa ketersediaan stok barang. Salah satu yang dilakukan terkait dengan stok barang adalah dengan menganalisa data transaksi yang ada pada minimarket dengan aplikasi data mining, dimana data yang didapat bersumber dari laman kaggle.com.

Data mining merupakan proses penggalian data yang tersembunyi dari suatu database. Data mining juga disebut sebagai Knowledge In Database (KDD), dimana meruapakan suatu kegiatan yang mengumpulkan data lampau kemudian mencari pola hubungan terhadap data yang besar. Terdapat banyak algoritma yang digunakan untuk menemukan pola transaksi penjualan, salah satunya yaitu K-Means Clustering (Adani et al., 2019).

Proses clustering sendiri merupakan suatu proses pengelompokkan data berdasarkan atas prinsip kesaman kelas serta mengurangi kesamaan antar kelas. Tingkat keakuratan dalam memprediksi penjualan mempunyai dampak yang besar terhadap penjualan. Hasil perkiraan penjualan menggunakan metode k-means

dapat menjadi penghubung antara penawaran dan permintaan yang banyak sehingga mampu mengurangi biaya dan mempertahankan jumlah stok barang (Method et al., 2019).

Dalam penelitian yang berjudul Penerapan Data Mining Pada Penjualan Produk menggunaka Meode K-Means Clustering (studi Kasus Toko Sepatu Kakikaki) pada tahun 2022 oleh (Kristianto & Rudianto, 2022) yang membahas tentang penjualan produk sepatu dan sandal. Hasil dari penelitiannya adalah bahwa produk yang diminati masyarakat saat ini yaitu model sepatu fashion dan sandal dengan hasil Laku = 3, dan barang yang kurang diminati oleh masyarakat adalah produk Warrior dengan hasil kueang = 5 serta 7 produk sisanya memiliki hasil yang cukup.

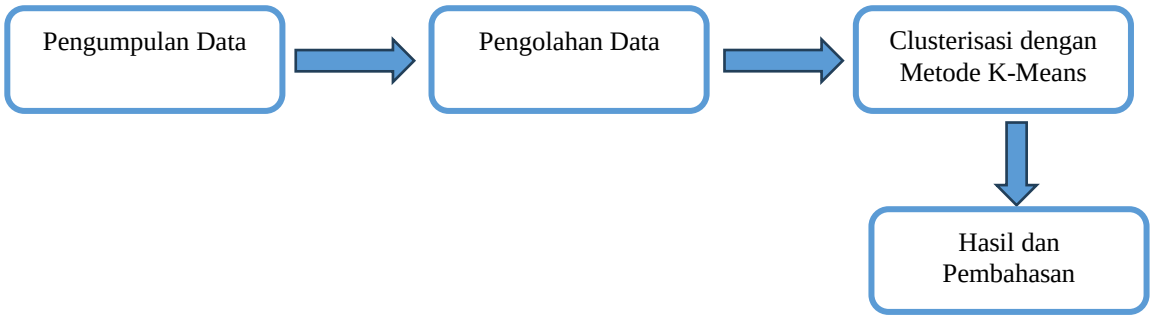
Penelitian dengan judul Clustering History Data Penjualan Menggunakan Algoritma K-Means, pada tahun 2021 oleh (Putra et al., 2021) menghasilkan penelitian dimana dari 23 item data penjualan dari tahun 2019-2020 dengan kategori barang laptop, komputer dan printer menghasilkan 3 kelompok cluster. Dari 3 cluster yang terbentuk dengan titik centoid awal yaitu C1 (800, 580,200), C2 (500, 450, 300) dan C3 (100, 190, 95) maka proses akhir berhenti pada iterasi ke 4 yang mendapatkan data pada cluster 1 atau sangat diminati memiliki 5 barang, cluster 2 atau diminati memiliki 4 barang dan cluster 3 atau kurang diminati memiliki 14 barang.

Klasterisasi merupakan salah satu klasifikasi tak terpandu (unsupervised classification). Ciri khas klasifikasi tak terpandu adalah pada data latih tidak tersedia target atau label yang berisi informasi kelas tiap-tiap tuple. Penentuan suatu tuple masuk klaster mana adalah dengan menggunakan jarak. Biasanya menggunakan Euclidean dan Manhattan (Herlawati et al., 2020). Dalam melakukan penelitian ini penulis menggunakan teknik clustering dengan algoritma k-means.

Bahasa pemrograman yang digunakan dalam melakukan proses clustering ini adalah phyton, kemudian data yang digunakan dalam penelitian ini menggunakan data clusterpenjualan.csv yang diperoleh dari laman kaggle.com. Phyton merupakan salah satu script bahasa pemrograman yang berbasis objek. Bahasa pemrograman ini dapat dijalankan dalam beberapa platform perangkat lunak dengan melalui berbagai sistem operasi.selain itu bahasa pemrograman phyton mendukung adanya library-library yang didalamnya menyediakan fungsi analisis data dan fiungsi machie learning, data processing serta visualisasi data (Manalu et al., 2022).

Scikit-learn meruapakan suatu library analisis data yang bersifat open source dan merupakan standar emas untuk Machine Learning (ML) dalam lingkungan Phyton. Library ini meliputi berbagai metode algoritma data mining, termasuk didalamnya terdapat algoritma untuk klasifikasi, regresi dan clustering (In & Activestate, 2023).

METODE PENELITIAN



Gambar 1. Tahapan penelitian

Langkah-langkah yang digunakan untuk membentuk cluster dengan menggunakan metode K-Means adalah (Nissa, 2023):

1. Memilih K buah titik yang akan dijadikan centroid secara acak.
2. Melakukan pengelompokkan data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster merupakan titik centroid yang telah dipilih sebelumnya.
3. Memperbaharui nilai titik centroid
4. Mengulangi langkah 2 dan 3 atau bahkan lebih sampai dengan nilai dari centroid tidak berubah kembali.

Dalam proses pengelompokkan data kedalam suatu cluster dilakukan dengan menggunakan cara menghitung jarak terdekat dari suatu data ke titik centroid. Dalam perhitungannya digunakan rumus:

$$d(x_i, x_j) = (|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \cdots + |x_{in} - x_{jn}|^g)^{1/g}$$

Dengan :

g = 1, untuk menghitung jarak manhattan

g = 2, untuk menghitung jarak Euclidean
g = ∞, untuk menghitung jarak Chebycey
x_1, x_2 adalah dua buah data yang akan dihitung jaraknya
p = dimensi dari sebuah data

Sedangkan untuk memperbaharui nilai dari centroid digunakan rumus:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Dimana:
μ_k= titik centroid dari kluster ke=K
N_k = banyaknya data pada cluster ke-K
x_q =data ke-q pada cluster ke-K

HASIL DAN PEMBAHASAN

1. Tahap pengumpulan Data

   Dataset penjualan yang digunakan dalam penelitian ini diambil dari laman kaggel.com.

kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata
5846 unique values	[null] 100% 5G;10;10;1.0000 0% Other (23) 0%	5845 total values	5847 total values	5847 total values
2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	1.0000
2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	1.0000
AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	1.1000
RIBUT KILOAN	RIBUT KILOAN	77	83	1.0779
7916248823	MINYAK TAWON FF	8	9	1.1250
7916248830	MINYAK TAWON EE	3	3	1.0000
7916248854	MINYAK TAWON CC	2	2	1.0000
870	KERTAS JEPUN 1 PACK	1	1	1.0000
1602	LEM KOREA ATRICO BOTOL	20	26	1.3000
7916248847	MINYAK GOSOK TAWON DD	16	17	1.0625

   Sumber: kaggle.com

   Gambar 2. Data Transaksi Penjualan

2. Tahap Pengolahan Data

   Dalam proses pengolahan dilakukan beberapa tahapan, diantaranya:

   a. Meng-import library phyton yang digunakan untuk kebutuhan dataframe, visualisasi dan clustering. Langkah-langkahnya sebagai berikut:

```
from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
```

   b. Langkah selanjutnya yaitu meng-input dataset. Dataset yang digunakan pada penelitian ini adalah

dataset clusterpenjualan.csv. Berikut langkah-langkahnya:

```
df = pd.read_csv('/kaggle/input/clsuterpenjualan/barang_keluar_imam.csv',
delimiter=';',skiprows=0, low_memory=False)
df.head()
```

3. Tahap K-Means Clustering

Analisis clustering dilakukan terhadap data penjualan yang didapat dari lamam kaggle.com. Proses dan hasil analisis K-Means dengan phyton untuk data penjualan sebagai berikut:

```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]


plt.scatter(df1.jumlah_transaksi, df1['total_penjualan'], color='green')
plt.scatter(df2.jumlah_transaksi, df2['total_penjualan'], color='red')
plt.scatter(df3.jumlah_transaksi, df3['total_penjualan'], color='blue')

plt.xlabel('Jumlah Transaki')
plt.ylabel('Total Penjualan')
plt.legend()
```

In [1]:

```
from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
```

```
/opt/conda/lib/python3.10/site-packages/scipy/__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is requir
ed for this version of SciPy (detected version 1.23.5
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

In [2]:

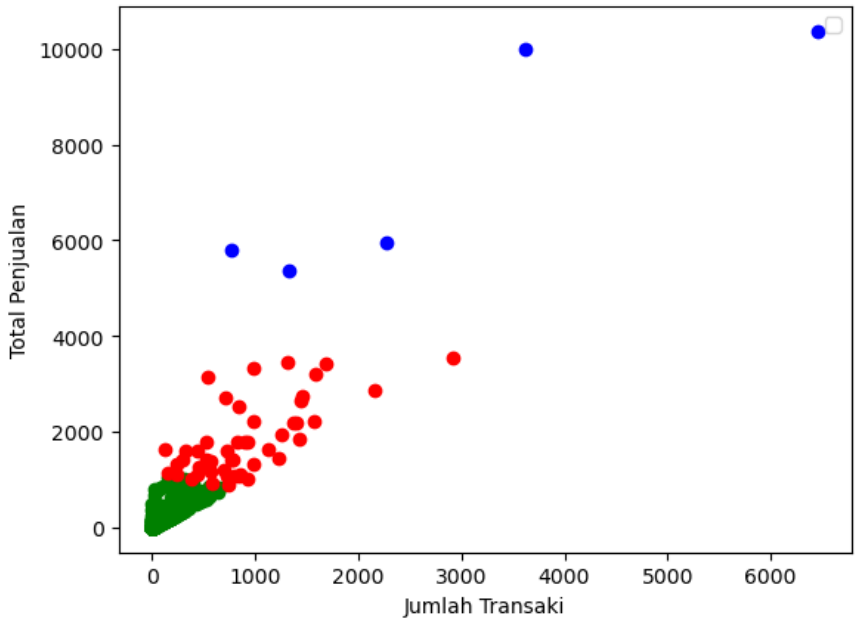
```
df = pd.read_csv('/kaggle/input/clsuterpenjualan/barang_keluar_imam.csv', delimiter=';',skiprows=0, low_memory=False)
df.head()
```

Out[2]:

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	1.0000
1	2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	1.0000
2	AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	1.1000
3	RIBUT KILOAN	RIBUT KILOAN	77	83	1.0779
4	7916248823	MINYAK TAWON EE	8	9	1.1250

Sumber: Hasil penelitian

Gambar 3. code import library dan output data



Sumber : Hasil penelitian

Gambar 4. Cluster data penjualan

Untuk dapat membuat cluster setelah dilakukan penentuan centroid data:

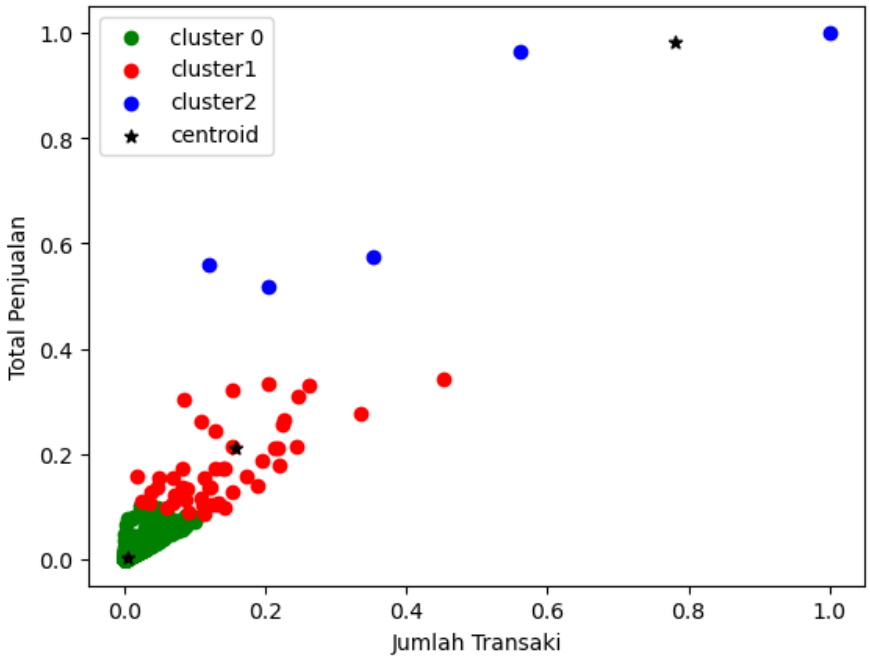
```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]


plt.scatter(df1.jumlah_transaksi, df1['total_penjualan'], color='green', label = 'cluster 0')
plt.scatter(df2.jumlah_transaksi, df2['total_penjualan'], color='red', label = 'cluster1')
plt.scatter(df3.jumlah_transaksi, df3['total_penjualan'], color='blue', label = 'cluster2')

plt.xlabel('Jumlah Transaki')
plt.ylabel('Total Penjualan')
plt.legend()

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,1], color='black', marker='*', label='centroid')
plt.legend()
```

Tabel 1. Prediksi cluster

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	1	1	1.0000	0
1	2 TANG MELATI 1 RENCENG (ISI 10)	2 TANG MELATI 1 RENCENG (ISI 10)	1	1	1.0000	0

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
2	AQUA 1500 ML 1 DUS	AQUA 1500 ML 1 DUS	10	11	1.1000	0
3	RIBUT KILOAN	RIBUT KILOAN	77	83	1.0779	0
4	7916248823	MINYAK TAWON FF	8	9	1.1250	0

Sumber : Hasil Penelitian



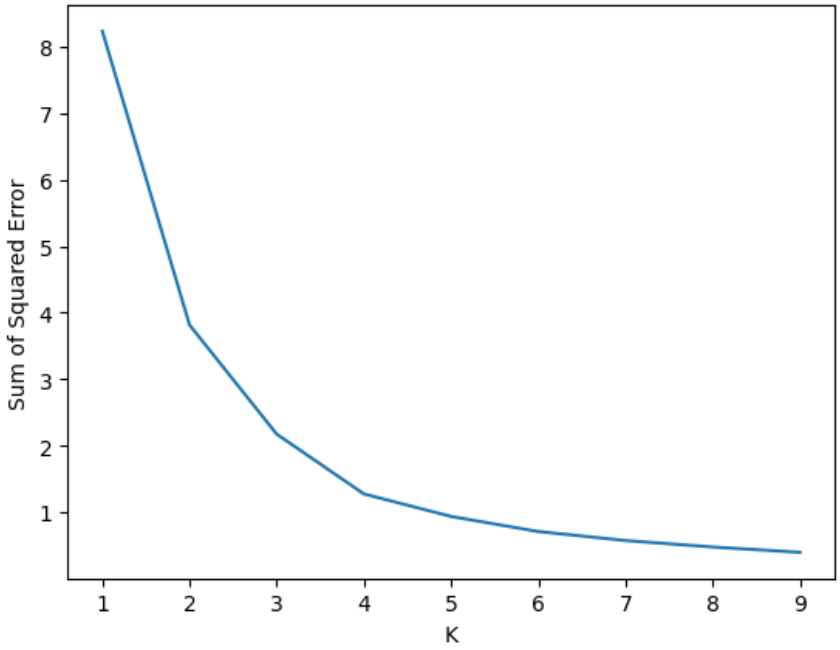
Sumber : Hasil penelitian

Gambar 4. sebaran cluster data penjualan

Berdasarkan gambar diatas:

a. Warna bintang hitam menunjukkan titik pusat atau lokasi cluaster yang disebut dengan centroid
b. Warna hijau menunjukkan cluster 0 dan menunjukkan data penjualan terendah
c. Warna merah menunjukkan cluster 1 dan menggambarkan data penjualan sedang
d. Warna biru menunjukkan cluster 2 dan menggambarkan data penjualan tertinggi

Data akan dapat dibaca menggunakan library ‘import pandas ad pd’ dan dapat dibaca menggunakan function pd.csv (clusterpenjualan). Dalam penelitian ini cluster yang dibuat dan didapatkan sebanyak 3 cluster dengan menggunakan metode elbow, dengan menggunakan library ‘import matplotlib.pyplot as plt’ dan menggunakan function plt.plot() serta plt.show(). Proses perhitungan K-Means Clustering dengan data yang telah bersih dan dibagi menjadi 3 cluster. K-Means clustering dilakukan dengan menggunakan library ‘from sklearn.cluster import Kmeans’ dan menggunakan function Kmeans (n_cluster=3) dan kmeans.fit(clusterpenjualan).



Sumber : Hasil penelitian

Gambar 5. Metode Elbow dalam penentuan jumlah cluster

Setelah memperoleh hasil K-Means Clustering, dilakukan visualisasi data menggunakan media tabel sebagai hasil kesimpulah pembahasan.

Tabel 2. Hasil Clusterisasi

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
0	2 TANG BLACK TEA 1 RENCENG (ISI 10)	2 TANG BLACK TEA 1 RENCENG (ISI 10)	0.000000	0.000000	1.0000	0
4922	8998866181068	CIPTADENT MAXI FRESH MINT 75G	0.015489	0.011195	1.1584	0
4921	8998866108799	ZINC ACTIVE FRESH RENCENG	0.013631	0.009265	1.0899	0
4920	8998866108324	MAMA LIME 220ML	0.020446	0.014476	1.1353	0
4919	8998866108317	EMERON HJB ANTI DANDRUF RENCENG	0.001394	0.000869	1.0000	0

	kode_barang	nama_barang	jumlah_transaksi	total_penjualan	rata_rata	cluster
...	...	...	...	...	...	...
6552	GULA PASIR 1 KG	GULA PASIR 1 KG	0.353005	0.572862	2.6039	2
4940	8998866200301	SEDAP GORENG	0.205235	0.517661	4.0460	2
4941	8998866200318	SEDAP AYAM BAWANG	0.119269	0.558483	7.5071	2
6477	DINGIN	DINGIN	1.000000	1.000000	1.6049	2
6553	GULA PASIR 1/2 KG	GULA PASIR 1/2 KG	0.560409	0.964196	2.7610	2

KESIMPULAN

Dari hasil penelitian *K-Means Clustering* untuk pengelompokan minat konsumen pada produk di minimarket yang telah diuraikan disimpulkan bahwa penerapan algoritma *K-Means Clustering* pada data penjualan produk di minimarket, menghasilkan sebuah informasi mengenai data pengelompokan minat konsumen tertinggi dan terendah. Sehingga data dijadikan rujukan bagi menajemen untuk mengatur stok barang agar toko tidak mengecewakan pelaggan karena barang yang ingin di beli tidak tersedia.

REFERENSI

Adani, N. F., Boy, A. F., Kom, S., Kom, M., Syahputra, R., Kom, S., & Kom, M. (2019). *Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan*. *x*, 1–11.

Herlawati, H., Bhayangkara, U., Raya, J., & Handayanto, R. T. (2020). *Penggunaan Matlab dan Python dalam Klasterisasi Data*. *May*. https://doi.org/10.31599/jki.v20i1.85

In, S., & Activestate, W. H. Y. (2023). *What is Scikit-Learn in Python? Ml*, 1–14.

Kristianto, W. W., & Rudianto, C. (2022). *Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K- Means Clustering (Studi Kasus Toko Sepatu Kakikaki)*. *5*, 90–98.

Manalu, D. A., Gunadi, G., & Informatika, T. (2022). *IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI*. *8*(1), 45–54.

Method, M., Indriyani, F., & Irfiani, E. (2019). *Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K-*. *7*(November), 109–113.

Nissa, N. K. (2023). *Unsupervised Learning : K-means Clustering using Python (Case : Online Retail Dataset)*.

Putra, Y. D., Sudarma, M., Bagus, I., & Swamardika, A. (2021). *Cluster ing History Data Penjualan Menggunakan*. *20*(2).