Jurnal Informatika dan Tekonologi Komputer
E-ISSN: 2809-9230
P-ISSN: 2809-9249

Penerapan Algoritma K-Means Untuk Menganalisis Berdasarkan Jenis Dan Penyebab Kematian Di Jawa Barat

Variant Prakasa Wibowo [1], Safrizal [2*]

[1] Universitas Pembangunan Jaya, Tangerang Selatan, Banten; e-mail : variant.prakasawibowo@student.upj.ac.id
[2] Universitas Pembangunan Jaya, Tangerang Selatan, Banten; e-mail : safrizal.abdurrahman@upj.ac.id
* Corresponding Author : Safrizal

Abstract: West Java Province is one of the regions in Indonesia with fluctuating mortality rates each year. The high number of deaths in various areas, with diverse characteristics of death types and causes, encourages the need for in-depth analysis to map the pattern of mortality cases. This study aims to analyze mortality data in West Java Province based on the type of death, cause of death, and number of cases using the K-Means Clustering method. The data used in this study is secondary data obtained from the official Open Data Jawa Barat portal for the period 2019 to 2023, with a total of 506 mortality records that have undergone pre-processing and data transformation stages. The analysis process was carried out through clustering using the K-Means algorithm, with the optimal number of clusters determined by the Elbow Method, and the clustering quality evaluated using the Silhouette Score. The results showed that the mortality data could be grouped into three main clusters, each with different characteristics in terms of the number of deaths, types, and causes. Cluster 0 consists of data with a high number of deaths, mostly occurring in urban areas. Cluster 1 contains data with moderate numbers and a variety of causes, while Cluster 2 represents areas with low mortality rates spread across several districts. The obtained Silhouette Score value of 0.52 indicates that the quality of cluster separation is quite good. In addition to clustering results, data visualization showed that Dengue Fever has been the dominant cause of death in the last three years, with Cianjur Regency recorded as the region with the highest mortality cases. This study is expected to serve as a reference for formulating public health policies and mortality prevention strategies in West Java Province.

Keywords: West Java Province; Clustering; K-Means; Metode Elbow; Data mining

Received: May 26, 2025
Revised: June 1, 2025
Accepted: July 5, 2025
Published: July 8, 2025
Curr. Ver.: July 8, 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

Abstrak: Provinsi Jawa Barat adalah salah satu wilayah di Indonesia dengan kematian yang angka nya tidak stabil di setiap tahun. Tingginya jumlah kematian yang ada di berbagai daerah dengan karakteristik jenis dan penyebab yang beragam, mendorong keinginan analisa lebih lanjut yang mendalam untuk memetakan pola kasus kematian. Penelitian ini bertujuan untuk menganalisa data kematian di Provinsi Jawa Barat berdasarkan jenis kematian, penyebab kematian, dan jumlah kasus menggunakan metode K-Means Clustering**.** Data dalam penelitian ini berupa data sekunder yang didapat dari portal resmi Open Data Jawa Barat periode 2019 hingga 2023, dengan total 506 data kematian yang sudah melalui proses pre-processing dan transformasi data. Proses analisa dilakukan dengan tahapan clustering menggunakan algoritma K-Means, dengan jumlah klaster optimal ditentukan menggunakan Metode Elbow, dan evaluasi kualitas clustering menggunakan Silhouette Score. Hasil penelitian menunjukkan data kematian bisa dikelompokkan ke dalam tiga klaster utama, masing-masing dengan karakteristik jumlah kematian, jenis, dan penyebab yang berbeda. Klaster 0 terdiri dari data dengan jumlah kasus tinggi yang banyak terjadi di wilayah perkotaan, Klaster 1 berisi data dengan jumlah sedang dan variasi penyebab yang beragam, serta Klaster 2 dengan jumlah kematian rendah yang tersebar merata di beberapa kabupaten. Nilai Silhouette Score yang diperoleh sebesar 0.52 yang artinya kualitas pemisah antar klaster cukup baik. Selain hasil clustering, visualisasi data menunjukkan bahwa Demam Berdarah

menjadi penyebab kematian yang mendominasi dalam tiga tahun terakhir, dan Kabupaten Cianjur tercatat sebagai wilayah dengan angka kematian tertinggi. Hasil penelitian ini diharapkan bisa menjadi refrensi dasar perumusan kebijakan kesehatan dan upaya pencegahan kematian di Provinsi Jawa Barat.

Kata kunci: Provinsi Jawa Barat; Clustering; K-Means; Metode Elbow; Data mining

1. Pendahuluan

Kematian adalah peristiwa yang pasti di hadapi oleh manusia. Kematian terjadi karena di pengaruhi beberapa faktor, bisa dari lingkungan di sekitar, bencana alam, kecelakaan, penyakit menular dan tidak menular, atau bisa juga karena kelainan saat baru lahir. Dengan memahami data penyebab kematian yang terjadi, pemerintah dan lembaga kesehatan bisa melakukan pencegahan dan analisa kesehatan yang lebih tepat sasaran[1].

Dari data yang di peroleh dari open data Jawa barat, tercatat angka kematian sebanyak 4013 kasus dalam periode tahun 2019 hingga 2023. Hal ini bisa terjadi karena Provinsi Jawa Barat adalah daerah yang bisa di katakan mempunyai penduduk terbanyak di Indonesia. Data penelitian ini di kumpulkan menjadi satu dataset kematian di Provinsi Jawa Barat. Mengamati potensi pola pada dataset, penelitian ini akan menggunakan pengelompokan dengan metode clustering[2].

Clustering adalah metode yang ada pada machine learning yang dapat mengembangkan informasi data berdasarkan kesamaan dan perbedaan yang ada di antara data tersebut. Clustering di tujukan agar bisa membantu mencari struktur tersembunyi dalam data. Hal yang lain juga bisa untuk mengetahui fitur dataset yang berkaitan dengan data kematian di Provinsi Jawa Barat[3], [4.]

Clustering adalah bagian dari metode proses data mining dengan melakukan penggalian data dengan menagandalkan kemiripan antara tipe data pada dataset. Dengan bantuan metode clustering, data bisa dikelompokkan tanpa membutuhkan informasi sebelumnya tentang kategori atau kelompok yang diharapkan [5], [6].

Data mining bisa di artikan sebagai tindakan proses mencari, menemukan pola atau informasi dari sekelompok data yang jumlah datanya banyak. Dalam data mining bisa menggunakan kecerdasan buatan, statistika, dan teknik matematika. Dengan adanya data mining, data kematian yang hanya disimpan dalam bentuk angka atau tabel bisa dianalisa lebih mendalam, bisa menghasilkan informasi baru yang bermanfaat dalam pengambilan keputusan[7], [8].

Kontribusi yang di berikan pada penelitian, di harapkan bisa memberi referensi tambahan untuk pihak yang membutuhkan dalam melihat pola kematian di Provinsi Jawa Barat. Bersumber dari dataset yang ada, kontribusi pada penelitian adalah Mengidentifikasi jenis kematian dan wilayah dengan jumlah kematian tertinggi, yang dapat dijadikan dasar untuk pertimbangan dalam pencegahan kematian di tahun tahun berikutnya.

Penentuan jumlah klaster dengan bantuan metode elbow dan silhoutte score memberikan kepastian klaster yang paling efektif yang digunakan pada K-Means[9]. Pemilihan nilai *klaster* dalam algoritma K-Means memiliki keakuratan dalam memisahkan karakteristik data. Nilai *klaster* yang optimal akan membentuk klaster yang saling terpisah secara jelas. Dengan visualisasi interaktif, pola-pola kematian di berbagai wilayah kabupaten yang ada di Jawa Barat bisa teridentifikasi lebih jelas, yang nantinya bermanfaat untuk mendukung kebijakan kesehatan[10], [11].

Tujuan dari penelitian ini adalah menganalisa perubahan dan tren kematian yang terjadi di Provinsi Jawa Barat dalam lima tahun terakhir. Penelitian ini juga di harapkan bisa untuk memudahkan tim medis dan pemerintah daerah Jawa Barat untuk menganalisa pencegahan penyakit berdasarkan pola kematian yang terjadi. Hasil analisis ini diharapkan juga dapat

menjadi bahan pertimbangan bagi Pharmaceutical Analyst dalam mengembangkan jenis obat yang sesuai dengan kebutuhan kesehatan di setiap daerah di provinsi Jawa Barat, sehingga penanganan penyakit dapat dilakukan secara lebih tepat sasaran.

Penulisan artikel di penelitian ini berurutan agar pembaca bisa mengerti dengan jelas isi dari artikel pada penelitian ini. Urutan bacaan di mulai dari pendahuluan, selanjutnya ada metode yang di gunakan agar pembaca bisa mengulangi eksperimen yang sudah peneliti lakukan, selanjutnya ada sumber pustaka dan hasil pembahasan penelitian yang berisi temuan pada penelitian, dan bagian terakhir adalah berisi kesimpulan[12].

2. Kajian Pustaka atau Penelitian Terkait

Penelitian sebelumnya oleh Agus Nursikuwagus, Suherman, Hendry Gunawan, Ilham Alamsyah (2024) ini mengkaji pengelompokan penyebab kematian di Indonesia dari tahun 2000 hingga 2020 dengan bantuan metode K-Means Clustering. Dataset terdiri dari 649 instans yang mencakup data dari 34 provinsi, dengan kategori kematian seperti bencana alam, bencana non alam, dan bencana sosial.

Proses penelitian dilakukan melalui tahapan data mining, antara lain: Data preprocessing: cleansing data, mengatasi missing value, dan transformasi data diskrit ke numerik.

Pemodelan: menggunakan algoritma K-Means, dengan penentuan jumlah klaster optimal menggunakan metode elbow curve (hasil terbaik pada K=3).

Visualisasi: scatter plot, box plot, dan confusion matrix untuk menganalisis karakteristik masing-masing klaster

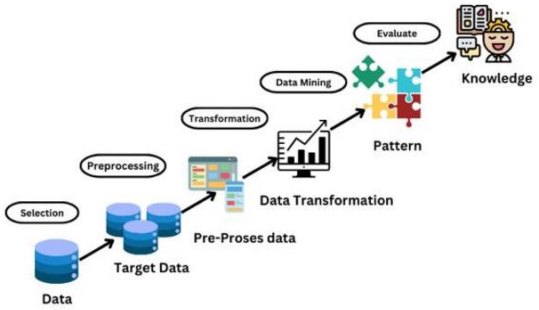
Hasil dari penelitian sebelumnya:

a. Klaster 1 (C1): kematian akibat bencana alam,
b. Klaster 2 (C2): bencana non alam dan penyakit (memiliki tingkat kematian tertinggi),
c. Klaster 3 (C3): bencana sosial (tingkat kematian terendah).

Meskipun proses clustering berhasil mengelompokkan data dengan cukup baik, namun klasifikasi kematian akibat bencana sosial belum bisa diprediksi dengan akurat, menunjukkan tantangan dalam menghadapi imbalance class dan perlunya teknik resampling pada studi lanjutan.[12]

3. Metode yang Diusulkan

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD) sebagai pendekatan dalam proses analisis data kematian di Provinsi Jawa Barat. Menurut peneliti metode KDD ini tahapannya sesuai dengan proses dalam data mining, mulai dari pengumpulan data hingga hasil[13], [14].



Gambar 1. Proses KDD (sumber: binus,2021)

3.1 Data Selection

Pada tahapan data selection, peneliti menyeleksi data yang sudah di ambil dari open data Provinsi Jawa Barat. Hasil seleksi akan di gunakan pada proses data mining, memilih data data yang di butuhkan untuk proses selanjutnya, kemudian di jadikan satu berkas excel, terpisah dari basis data yang di temukan di open data Provinsi Jawa Barat[15], [16].

3.2 Pre-processing (Cleaning)

Pada data yang sudah di seleksi pada tahap sebelumnya, di temukan data yang tidak lengkap dan banyak data dengan jumlah 0. Pada tahap ini peneliti menghapus variabel data yang berjumlah 0. Atribut data yang tidak relevan ini akan menjadi penghambat data mining, tentu bisa mempengaruhi akurasi dari data mining kedepannya[17], [18].

3.3 Data Transformation

Disini dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga akan menemukan kualitas hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu[19], [20].

3.4 Data Mining

Data yang sudah melalui proses pre-processing dan data transformation kemudian diolah menggunakan algoritma K-Means Clustering untuk melakukan pengelompokan data kematian di Provinsi Jawa Barat. Proses transformasi ini bertujuan untuk mengubah variabel kategorikal menjadi nilai numerik dan memastikan skala data seimbang agar algoritma dapat memproses data dengan optimal. Variabel kategorikal seperti jenis kematian dan penyebab kematian diubah jadi data numerik menggunakan metode Label Encoding, setiap kategori diubah menjadi bilangan angka unik tanpa mengubah arti data.

Setelah diproses encoding, dilakukan proses standarisasi data menggunakan StandardScaler untuk variabel numerik, yaitu penyebab_encoded, jenis_encoded, dan jumlah_kematian. Proses ini bertujuan agar semua variabel memiliki skala yang sama, yaitu rata-rata 0 dan standar deviasi 1, sehingga tidak ada variabel yang mendominasi dalam proses perhitungan jarak antar data.

Jumlah klaster yang digunakan dalam proses clustering ditentukan terlebih dahulu menggunakan Metode Elbow, yaitu metode yang mengukur nilai Within-Cluster Sum of Square (WCSS) untuk berbagai jumlah klaster. Jumlah klaster optimal dipilih berdasarkan titik tekuk (elbow point) pada grafik, di mana nilai WCSS mulai menurun secara signifikan.

Proses clustering kemudian dilakukan terhadap variabel numerik hasil transformasi menggunakan algoritma K-Means Clustering. Algoritma ini beroperasi dengan cara menentukan sejumlah klaster awal, menghitung jarak antara tiap data ke pusat klaster terdekat, dan memperbarui posisi centroid secara iteratif hingga posisi klaster stabil. Hasil clustering disimpan dalam kolom khusus untuk masing-masing data.

Untuk memudahkan visualisasi hasil clustering, diproses menggunakan Principal Component Analysis (PCA), dengan dua komponen utama (PC1 dan PC2) sebagai sumbu visualisasi dua dimensi. Visualisasi hasil clustering ditampilkan dalam bentuk scatter plot PCA 2D, di mana masing-masing data ditandai dengan warna berbeda sesuai label klaster yang terbentuk.

3.5 Interpretation (Evaluation)

Tahap Interpretation (Evaluation) dalam penelitian ini bertujuan untuk mengevaluasi kualitas hasil clustering yang telah diperoleh serta menilai seberapa baik data dapat dikelompokkan ke dalam klaster yang terbentuk. Proses interpretasi dilakukan dengan cara menganalisis hasil clustering yang divisualisasikan dalam bentuk scatter plot dua dimensi berbasis Principal Component Analysis (PCA). Visualisasi ini bertujuan untuk memudahkan pengamatan persebaran data antar klaster secara visual.

Selain itu, dilakukan penghitungan Silhouette Score untuk mengukur tingkat kesesuaian data dalam setiap klaster. Silhouette Score digunakan untuk menilai seberapa dekat data dalam satu klaster dengan data lain di klaster yang sama, dibandingkan dengan data di klaster lain. Nilai skor ini berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa data berada di klaster yang tepat, sedangkan nilai mendekati 0 atau negatif menunjukkan bahwa data kurang tepat berada dalam klaster tersebut.

Interpretasi juga dilakukan terhadap hasil visualisasi klasterisasi yang dihasilkan, untuk mengidentifikasi pola-pola distribusi kematian di Provinsi Jawa Barat berdasarkan jumlah kasus, jenis kematian, dan penyebab kematian. Analisis ini membantu dalam memahami karakteristik masing-masing klaster yang terbentuk serta distribusi geografisnya di wilayah kabupaten/kota.

3.7. Knowlage

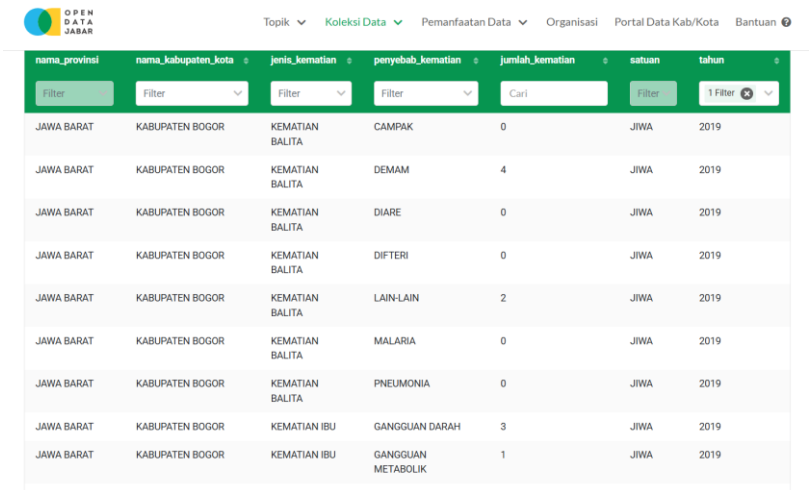
Knowlage ini bagian akhir metode KDD, yaitu penemuan informasi dari penelitian ini[21]. Peneliti berharap bisa membantu peneliti lain dalam melakukan penelitian tren kematian di tahun tahun selanjutnya di tiap daerah khusus nya di Provinsi Jawa Barat, sehingga pengembangan obat atau program kesehatan dapat lebih tepat sasaran.

4. Hasil dan Pembahasan

Dalam penerapan data mining pada penelitian ini akan di analisa menggunakan tahapan Knowledge Discovert in Database (KDD), sebelumnya data mentah yang ada di berbagai daerah di Provinsi Jawa Barat akan di kumpulkan menjadi jadikan satu file menggunakan tools excel.

4.1. Data Selection

Tahap Data Selection ini langkah awal dalam penelitian, Peneliti mengambil data sekunder yang diakses melalui website Open Data Jawa Barat sebagai bahan analisa**.** (opendata.jabarprov.go.id/id). Dataset yang diambil berisi informasi tentang jumlah kematian di Provinsi Jawa Barat berdasarkan jenis kematian, penyebab kematian, kabupaten/kota, dan tahun kejadian dalam kurun waktu 2019 hingga 2023. Pengambilan data dilakukan oleh peneliti dengan mengunduh dataset melalui fitur download yang tersedia di halaman website Open Data. Data yang telah diunduh, selanjutnya bisa disimpan dalam format Microsoft Excel (.xlsx) untuk diproses dan dianalisa lebih lanjut pada tahapan berikutnya.



OPEN DATA JABAR — Topik | Koleksi Data | Pemanfaatan Data | Organisasi | Portal Data Kab/Kota | Bantuan

nama_provinsi	nama_kabupaten_kota	jenis_kematian	penyebab_kematian	jumlah_kematian	satuan	tahun
Filter	Filter	Filter	Filter	Cari	Filter	1 Filter
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	CAMPAK	0	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	DEMAM	4	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	DIARE	0	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	DIFTERI	0	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	LAIN-LAIN	2	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	MALARIA	0	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN BALITA	PNEUMONIA	0	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN IBU	GANGGUAN DARAH	3	JIWA	2019
JAWA BARAT	KABUPATEN BOGOR	KEMATIAN IBU	GANGGUAN METABOLIK	1	JIWA	2019

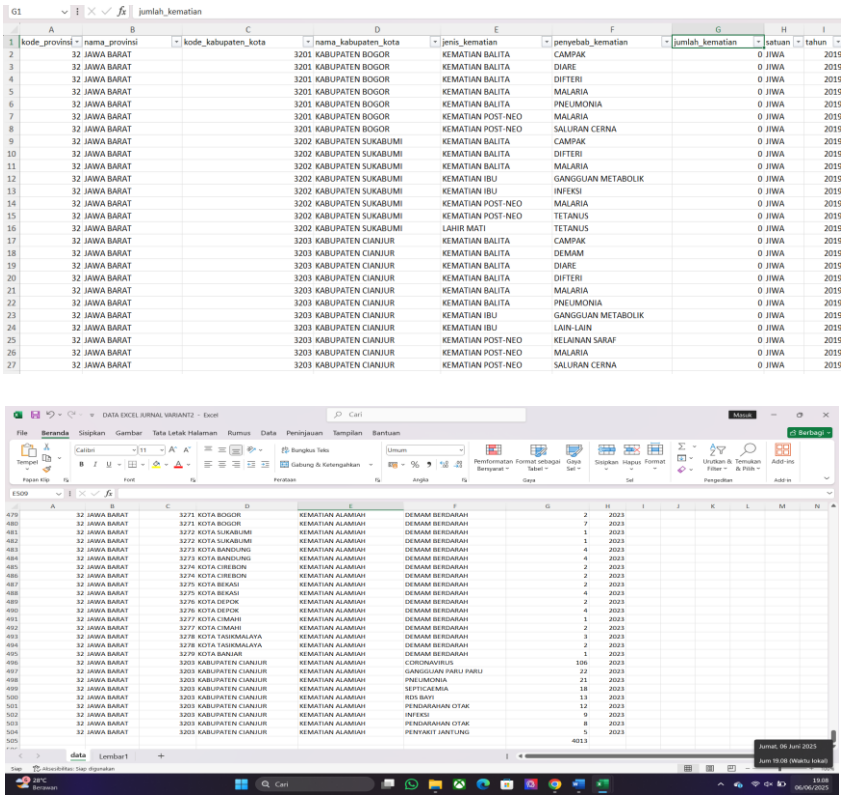
Gambar 2. Seleksi Dataset

4.2. Pre-processing (Cleaning)

Tahapan selanjutnya adalah melakukan proses Pre-processing (Cleaning) atau pembersihan data. Tahap ini penting untuk memastikan data yang digunakan dalam analisia dalam kondisi bersih, lengkap, dan sesuai kebutuhan penelitian. Hasil pembersihan data, dengan cara menghapus data yang memiliki nilai jumlah kematian = 0. Peneliti melakukan pemfilteran

dengan fitur yang ada di Microsoft excel, hal ini dilakukan karena data tersebut tidak memberikan kontribusi terhadap proses clustering dan dapat memengaruhi hasil analisis. Dengan menghapus data nol, hanya data kematian yang benar-benar terjadi saja yang dianalisis, sehingga hasil clustering menjadi lebih akurat dan relevan.

Hasil Pre-processing dari dataset yang peneliti temukan, diperoleh total sebanyak 506 data kematian yang tersebar di seluruh kabupaten/kota di Jawa Barat, dengan total 4.013 kasus kematian dalam kurun waktu 2019 hingga 2023. Dataset bersih ini kemudian digunakan dalam proses transformasi data pada tahap berikutnya.

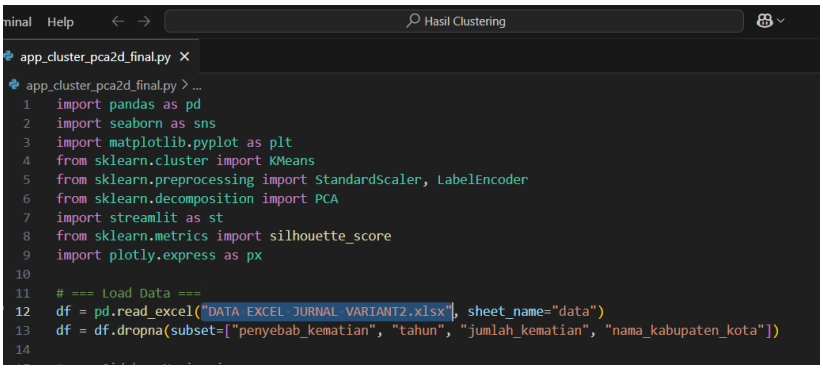


kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_kematian	penyebab_kematian	jumlah_kematian	satuan	tahun
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN BALITA	CAMPAK	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN BALITA	DIARE	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN BALITA	DIFTERI	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN BALITA	MALARIA	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN BALITA	PNEUMONIA	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN POST-NEO	MALARIA	0	JIWA	2019
32	JAWA BARAT	3201	KABUPATEN BOGOR	KEMATIAN POST-NEO	SALURAN CERNA	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN BALITA	CAMPAK	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN BALITA	DIFTERI	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN BALITA	MALARIA	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN IBU	GANGGUAN METABOLIK	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN IBU	INFEKSI	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN POST-NEO	MALARIA	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	KEMATIAN POST-NEO	TETANUS	0	JIWA	2019
32	JAWA BARAT	3202	KABUPATEN SUKABUMI	LAHIR MATI	TETANUS	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	CAMPAK	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	DEMAM	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	DIARE	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	DIFTERI	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	MALARIA	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN BALITA	PNEUMONIA	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN IBU	GANGGUAN METABOLIK	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN IBU	LAIN-LAIN	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN POST-NEO	KELAINAN SARAF	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN POST-NEO	MALARIA	0	JIWA	2019
32	JAWA BARAT	3203	KABUPATEN CIANJUR	KEMATIAN POST-NEO	SALURAN CERNA	0	JIWA	2019

Gambar 3. Hasil Penghapusan dan Perbaikan Data Tidak Terisi

4.3. Data Tranformation

Setelah data dibersihkan, data yang sudah siap, dimasukkan ke dalam aplikasi Visual Studio Code menggunakan bahasa Python. Pada bagian ini, proses pembacaan data dilakukan menggunakan perintah pd.read_excel() dari library pandas.



```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.decomposition import PCA
import streamlit as st
from sklearn.metrics import silhouette_score
import plotly.express as px

# === Load Data ===
df = pd.read_excel("DATA EXCEL JURNAL VARIANT2.xlsx", sheet_name="data")
df = df.dropna(subset=["penyebab_kematian", "tahun", "jumlah_kematian", "nama_kabupaten_kota"])
```

Gambar 4. Proses Load data Menggunakan Pandas

Gambar 4 juga menunjukan proses pembersihan tambahan untuk memastikan tidak ada data kosong di kolom-kolom penting yang akan digunakan dalam proses clustering. Hal ini dilakukan dengan perintah dropna() untuk menghapus baris yang memiliki nilai kosong pada kolom yang dipilih oleh peneliti yaitu penyebab_kematian, tahun, jumlah_kematian, dan nama_kabupaten_kota.

```
# Encode kategorikal
le_penyebab = LabelEncoder()
le_jenis = LabelEncoder()
df["penyebab_encoded"] = le_penyebab.fit_transform(df["penyebab_kematian"])
df["jenis_encoded"] = le_jenis.fit_transform(df["jenis_kematian"])

# Standarisasi
scaler = StandardScaler()
X = scaler.fit_transform(df[["penyebab_encoded", "jenis_encoded", "jumlah_kematian"]])
```

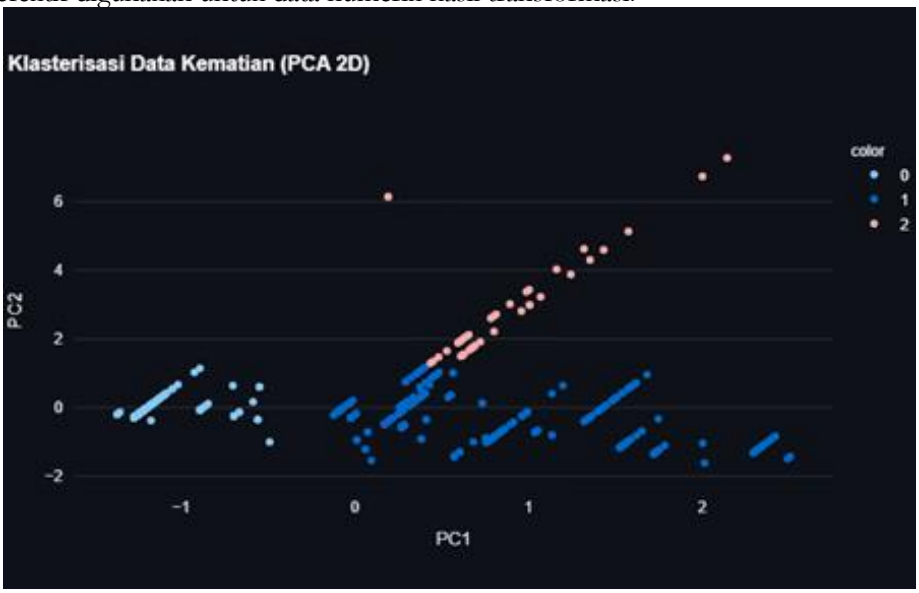
Gambar 5. Proses Label Endcoding

Gambar 5 dilakukan Label Encoding terhadap variabel jenis kematian dan penyebab kematian menggunakan fungsi LabelEncoder dari library Scikit-Learn. Proses ini mengubah setiap kategori menjadi bilangan bulat unik tanpa mengubah arti.

Selanjutnya, seluruh variabel numerik yang akan digunakan dalam proses clustering, yaitu penyebab_encoded, jenis_encoded, dan jumlah_kematian, dilakukan proses standarisasi menggunakan metode StandardScaler. Berfungsi untuk menyamakan skala nilai antar variabel agar tidak ada variabel yang mendominasi hasil clustering.

4.4. Data Mining

Tahap Data Mining merupakan inti dari proses analisis data pada penelitian ini. Pada tahap ini dilakukan proses clustering untuk mengelompokkan data kematian di Provinsi Jawa Barat berdasarkan karakteristik kemiripan antar data. Algoritma yang digunakan adalah K-Means Clustering, karena bisa membagi data ke jumlah klaster yang ditentukan sebelumnya dan efektif digunakan untuk data numerik hasil transformasi.



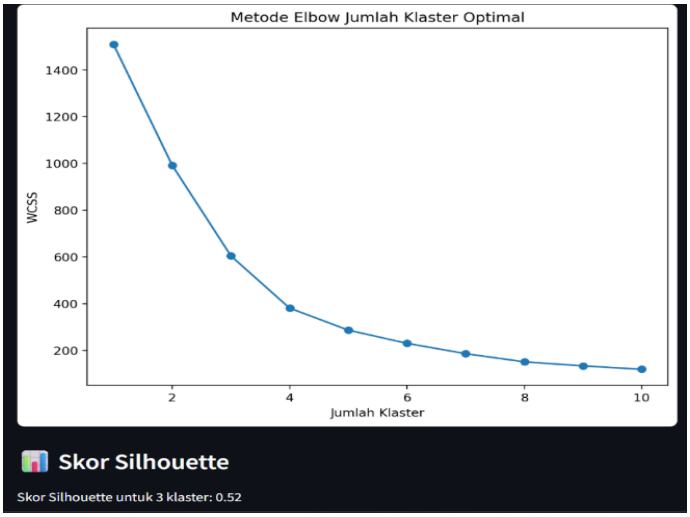
Gambar 6. Hasil Visualisasi Clustering

Hasil clustering divisualisasikan menggunakan Plotly Scatter Plot 2D yang ada pada gambar 6, dengan sumbu PC1 dan PC2 hasil reduksi dimensi PCA. Data pada scatter plot diwarnai berdasarkan label klaster yang terbentuk, sehingga memudahkan dalam melihat persebaran data antar klaster. Data diatur untuk menampilkan informasi tambahan (tahun, nama kabupaten/kota, jenis kematian, penyebab kematian, dan jumlah kematian) saat kursor diarahkan ke titik data yang di inginkan.

Hasil visualisasi juga menunjukkan data kematian bisa dikelompokkan ke tiga klaster dengan persebaran yang cukup terpisah. Masing-masing klaster memiliki karakteristik yang berbeda, baik dari jumlah kasus kematian, jenis kematian, penyebab kematian, maupun persebaran wilayah kabupaten/kota di Jawa Barat. Penjelasan mengenai beberapa klaster ini yaitu klaster 0 adalah kematian relatif tinggi dan didominasi oleh beberapa penyebab kematian tertentu yang terjadi di wilayah perkotaan dengan populasi padat, klaster 1 adalah kematian jumlah sedang, yang tersebar di berbagai kabupaten/kota, dengan variasi yang lebih beragam di banding klaster lain, klaster 2 adalah kematian dengan kasus rendah, tersebar di beberapa wilayah dengan karakter lebih merata dan penyebab nya umum.

4.5. Interpretation (Evaluation)

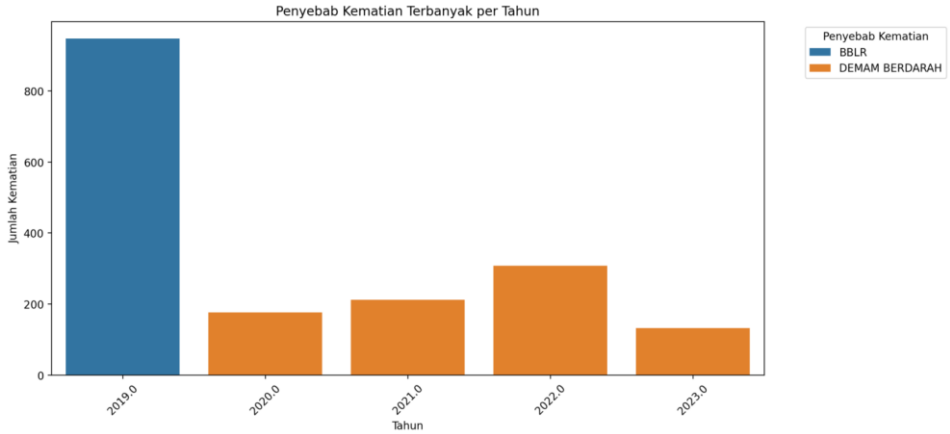
Gambar 6 adalah hasil uji menunjukkan bahwa jumlah klaster optimal berada di angka 3 klaster, ditandai dengan mulai melandainya nilai Within-Cluster Sum of Square (WCSS) pada titik tersebut. Untuk mengevaluasi kualitas klaster, dilakukan penghitungan Silhouette Score, yang menunjukkan nilai sebesar 0.52, tandanya pemisahan antar klaster cukup baik dan data dalam masing-masing klaster memiliki kemiripan karakteristik yang tinggi.



Gambar 6. Hasil uji Metode Elbow

4.6. Knowlage

Penelitian ini juga menghasilkan visualisasi pendukung yang bisa memberikan informasi tambahan mengenai kematian di Provinsi Jawa Barat. Berikut penjelasan tiap visualisasi yang diperoleh:

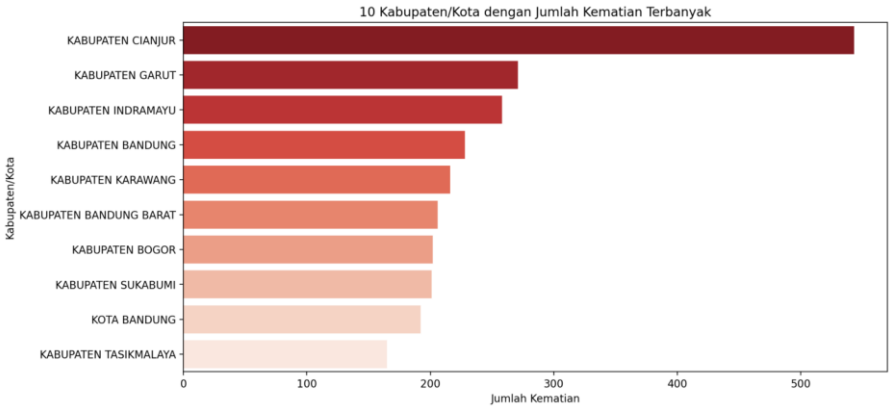


Gambar 7. Visualisasi Penyebab Kematian Terbanyak per Tahun

Gambar 7 menampilkan penyebab kematian paling banyak di Jawa Barat berdasarkan tahun. Hasil menunjukkan:

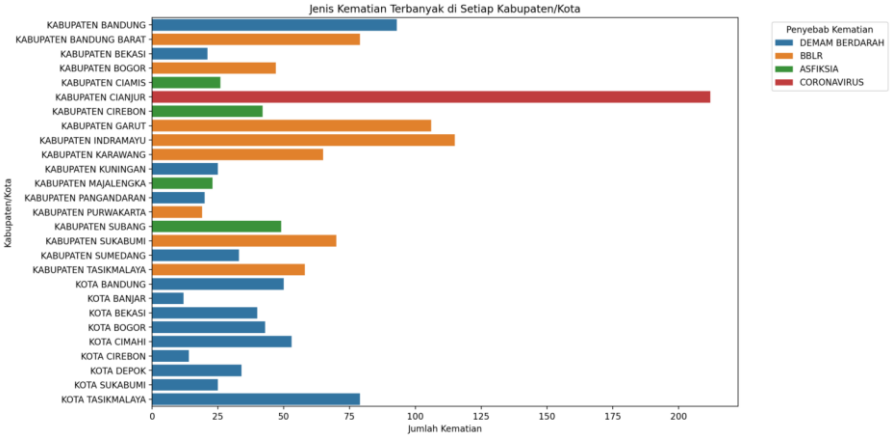
a. Pada tahun 2019, BBLR menjadi penyebab kematian tertinggi.
b. Mulai tahun 2020 hingga 2023, penyebab kematian Demam Berdarah menjadi dominan.

Dari grafik bisa disimpulkan terjadi perubahan pola penyebab kematian dari waktu ke waktu. Hal yang bisa menjadi dasar bagi dinas kesehatan untuk menjalankan evaluasi program penanganan penyakit menular seperti Demam Berdarah, yang mulai menjadi masalah utama dalam beberapa tahun terakhir.



Gambar 8. Visualisasi 10 Kabupaten/Kota dengan Jumlah Kematian Terbanyak

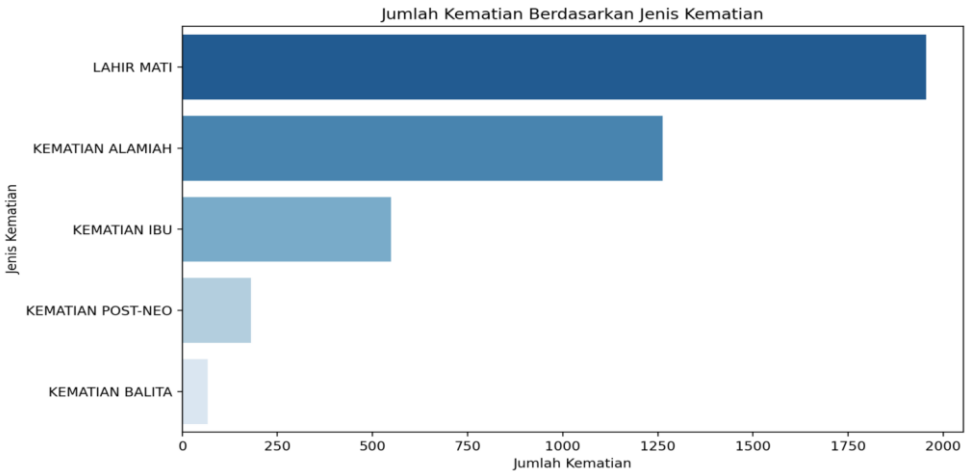
Gambar 8 ini menunjukkan 10 kabupaten/kota di Jawa Barat dengan jumlah kematian terbanyak. Visualisasi menunjukan Kabupaten Cianjur ada diposisi pertama dengan jumlah kematian tertinggi, diikuti oleh Kabupaten Garut dan Kabupaten Indramayu. Kota besar seperti Kota Bandung berada di posisi lebih bawah. Data ini memberikan gambaran bahwa angka kematian tinggi tidak selalu terjadi di daerah perkotaan saja, ternyata juga di beberapa wilayah kabupaten. Informasi ini dapat dijadikan dasar bagi pemerintah untuk menyesuaikan program kesehatan masyarakat di daerah terutama kabupaten.



Gambar 9. Visualisasi Jenis Kematian Terbanyak di Setiap Kabupaten/Kota

Gambar 9 memberi informasi jenis kematian paling dominan di masing-masing kabupaten/kota di Jawa Barat. Dari gambar tersebut dapat dilihat bahwa:

a. BBLR (Berat Badan Lahir Rendah) mendominasi di banyak kabupaten/kota seperti Kabupaten Garut, Indramayu, dan Karawang.
b. Penyebab kematian Coronavirus mendominasi di Kabupaten Cianjur.
c. Sementara di wilayah perkotaan seperti Kota Bandung, Kota Cimahi, dan Kota Bekasi lebih didominasi oleh Demam Berdarah.



Gambar 10. Visualisasi Jumlah kematian Berdasarkan Jenis kematian

Gambar 10 memberi informasi jumlah kematian berdasarkan jenis kematian di Provinsi Jawa Barat selama tahun 2019-2023. Bisa dilihat jenis kematian Lahir Mati menjadi kasus terbanyak, diikuti oleh kematian alami dan kematian ibu. Kematian Post-neonatal dan kematian Balita relatif lebih sedikit.

6. Kesimpulan

Dari data yang sudah peneliti temukan, bisa di simpulkan bahwa hasil dari hasil penelitian dapat dikelompokan ke dalam tiga klaster utama menggunakan algoritma K-means Clustering. Setiap klaster memiliki karakter yang berbeda, dari segi jumlah kasus, jenis kematian, dan penyebab kematian yang mendominasi. Hasil evaluasi menggunakan Silhouette Score, mendapat nilai sebesar 0.52 yang artinya pemisah antar klaster sudah cukup baik. Hasil analisa menggunakan Metode Elbow yang ada dalam penelitian ini juga menunjukan jumlah klaster optimalnya adalah 3 klaster. Visualisasi hasil clustering dan beberapa grafik, peneliti menemukan bahwa jenis kematian lahir mati merupakan kasus terbanyak, karena di visualisasi menunjukan di tahun 2019 kasus BBLR sangat mendominasi dan angkanya tinggi sekali, disusul oleh kematian alami dan kematian ibu. Dari sisi penyebab, Demam Berdarah menjadi kematian paling tinggi dan mendominasi di tahun 2020-2023. Berdasarkan persebaran wilayah, Kabupaten Cianjur, Kabupaten Garut, dan Kabupaten Indramayu merupakan tiga daerah dengan jumlah kasus kematian tertinggi dalam penelitian ini. Peneliti berharap hasil temuan ini bisa menjadi refrensi dan bahan pertimbangan bagi pemerintah daerah, dinas kesehatan, dan pihak terkait dalam menentukan prioritas kebijakan, program pencegahan, dan penanganan kematian yang lebih tepat sasaran di masing-masing wilayah di Provinsi Jawa Barat.

Daftar Pustaka

[1] C. Della Evania, W. Wiyanti, and C. Author, "Perbandingan Metode Holt's Linier dan GPT-4 untuk Peramalan Jumlah Kasus Kematian Diabetes Melitus di Indonesia," 2024. [Online]. Available: http://www.ojs.unm.ac.id/jmathcos

[2] M. Rizqi Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *Jurnal Teknologi Ilmu Komputer*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.

[3] E. Dwiguna and A. Bahtiar, "PENERAPAN DATA MINING UNTUK MENENTUKAN PENERIMA BANTUAN BLT MENGGUNAKAN METODE CLUSTERING K-MEANS PADA DESA PAMULIHAN," 2024.

[4] D. Setra Perdana, "PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN KETAHANAN PANGAN," vol. 3, no. 2, pp. 67–66.

[5] Wenny, "Bulletin of Information System Research (BIOS) Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokkan Tingkat Ekonomi Masyarakat," 2024, [Online]. Available: https://journal.grahamitra.id/index.php/bios

[6] S. Syahidatul Helma *et al.*, "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K-Means," 2019.

[7] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.

[8] A. Octaviani and P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *ANUVA*, vol. 4, no. 2, pp. 223–230, 2020.

[9] R. Kurniawan, S. Defit, and S. Sumijan, "Prediksi Tingkat Kerugian Peternak Akibat Penyakit pada Sapi Menggunakan Algoritma K-Means Clustering," *Jurnal Informasi dan Teknologi*, Sep. 2020, doi: 10.37034/jidt.v3i1.87.

[10] S. D. Oleh and A. Allaam, "Prediksi Churn Konsumen Menggunakan Algoritma Random Forest dengan Fuzzy C-Means untuk Meningkatkan Produktivitas Penjualan Bisnis," 2023.

[11] D. Sebagai *et al.*, "PENGELOMPOKAN PRODUK BERDASARKAN DATA PERSEDIAAN BARANG MENGGUNAKAN METODE ELBOW DAN K-MEDOID TUGAS AKHIR."

[12] A. Nursikuwagus, Suherman, and I. Alamsyah, "KLUSTERISASI PENYEBAB KEMATIAN DI INDONESIA DENGAN PENERAPAN ALGORITMA K-MEANS," *JSiI (Jurnal Sistem Informasi)*, vol. 11, no. 2, pp. 56–63, Sep. 2024, doi: 10.30656/jsii.v11i2.9276.

[13] F. Alghifari and D. Juardi, "Fauzan Alghifari Penerapan Data Mining Pada Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes."

[14] T. Lidia Putri and R. Danar Dana, "PENERAPAN DATA MINING PADA CLUSTERING DATA HARGA RUMAH DKI JAKARTA MENGGUNAKAN ALGORITMAK-MEANS," 2024. [Online]. Available: https://www.rumah123.com/jual/dki-jakarta/rumah/.

[15] D. A. Manalu and G. Gunadi, "IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI," *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.

[16] R. Putra Primanda, A. Alwi, and D. Mustikasari, "url : http://studentjournal.umpo.ac.id/index.php/komputek DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," 2021. [Online]. Available: http://studentjournal.umpo.ac.id/index.php/komputek

[17] M. Anggraini, R. Ayuning Tyas, I. Ana Sulasiyah, and Q. Aini, "IMPLEMENTASI ALGORITMA NAÏVE BAYES DALAM PENENTUAN RATING BUKU." [Online]. Available: www.kaggle.com

[18] D. A. Fakhri, S. Defit, and Sumijan, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering," *Jurnal Informasi dan Teknologi*, pp. 160–166, Sep. 2021, doi: 10.37034/jidt.v3i3.137.

[19] J. Hasil Penelitian dan Karya Ilmiah, M. Rizki, D. Devrika, F. Surayya Lubis, and I. Hadiyul Umam, "Aplikasi Data Mining dalam Penentuan Layout Swalayan dengan Menggunakan Metode MBA," 2019.

[20] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.

[21] G. Zahra, N. Fadhilah, R. A. Saputra, and A. H. Wibowo, "DETEKSI TINGKAT GANGGUAN KECEMASAN MENGGUNAKAN METODE RANDOM FOREST Anxiety Disorder Level Detection Using Random Forest Method", [Online]. Available: http://jurnal.umt.ac.id/index.php/jt/index