IMPLEMENTASI PENGELOMPOKAN REALISASI BELANJA MENGGUNAKAN ALGORITMA K-MEANS DI PROVINSI DKI JAKARTA

Sri Widyastuti[1], Irfan Ali[2]
[1] Komputerisasi Akuntansi, STMIK IKMI Cirebon
[2] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon, Jawa Barat 45131
irfanaali0.0@gmail.com

ABSTRAK

Dalam era globalisasi dan perkembangan teknologi informasi yang cepat, pengolahan dan analisis data menjadi krusial dalam pengambilan keputusan di berbagai sektor, termasuk sektor publik. Penelitian ini menghadapi masalah kompleksitas data realisasi belanja, terdiri dari banyak unit kerja dengan karakteristik yang beragam. Untuk mengatasi hal ini, dilakukan pengelompokan data menggunakan pendekatan K-Means Clustering menggunakan bahasa pemrograman Python. Data dibagi ke dalam kelompok cluster berbeda menggunakan metode K-Means Clustering. Penelitian ini menerapkan metode tersebut pada realisasi belanja di Provinsi DKI Jakarta tahun 2020. Hasil pengelompokan menunjukkan tiga cluster, dengan cluster 0 memiliki unit kerja terbanyak (436 unit kerja), cluster 1 memiliki unit kerja paling sedikit (18 unit kerja), dan cluster 2 memiliki 69 unit kerja. Unit kerja pada cluster 1 belum mencapai target pengeluaran yang ditetapkan. Identifikasi terhadap cluster tinggi (503 data) dan rendah (20 data) mengindikasikan bahwa perhatian khusus diperlukan pada cluster rendah untuk mencapai target yang ditetapkan pemerintah.

Kata kunci : *Globalisasi, K-Means Clustering, Pengolahan Data, Teknologi Informasi.*

1. PENDAHULUAN

Anggaran diperlukan untuk menjalankan pemerintahan dan mencapai tujuan yang telah ditetapkan. Menurut Undang-Undang Nomor 17 Tahun 2003 tentang Keuangan Negara, pengelolaan anggaran termasuk dalam pengelolaan keuangan negara. Salah satu aturannya adalah pembuatan rencana keuangan tahunan yang dikenal sebagai anggaran pendapatan dan belanja negara (APBN). [1]

Dalam metode K-Means Clustering terdapat beberapa konsep dan definisi yang relevan. Data realisasi belanja merupakan data yang menggambarkan jumlah belanja yang telah di realisasi atau dilaksanakan pada suatu unit kerja di Provinsi DKI Jakarta pada tahun 2020. Unit kerja merupakan unit atau bagian yang melakukan aktivitas atau pekerjaan tertentu dalam suatu organisasi di Provinsi DKI Jakarta. Tujuan dari penggunaan metode ini adalah untuk mengelompokan data menjadi k cluster eksklusif. Algoritma ini bekerja dengan cara meminimalkan jarak antara titik data ke pusat cluster yang terdekat. Metode K-Means Clustering adalah untuk menentukan jumlah cluster (k) yang diinginkan.[2]. Dalam penelitian ini, menggunakan bahasa pemograman Python untuk mengimplementasikan metode K-Means Clustering. [3]. Penelitian ini melakukan analisis terhadap data realisasi belanja di Provinsi DKI Jakarta pada tahun 2020, dengan menggunakan metode K-Means Clustering.

Penelitian yang dilakukan oleh Irna Yuniarfi dan Saifullah dalam jurnal informasi dan komputer, tahun 2021 yang berjudul "PENERAPAN ALGORITMA K-MEANS UNTUK MENGELOMPOKAN USIA CALON PENERIMA VAKSIN DI KAB. NGAWI" didasarkan pada dasar yang telah dijelaskan diatas, Algoritma K-Means adalah teknik yang menggunakan sistem partisi untuk melakukan pengelompokan data, proses pemodelan ini tanpa bantuan atau pengawasan dari pihak mana pun. Secara ilmiah, K-means adalah salah satu metode penambangan data terpenting. Metode K-Means Clustering merupakan metode pengelompokan non-hierarchical ini bertujuan untuk mengelompokan objek mulai dari identifikasi data yang akan di-cluster.

Dalam penelitian tersebut, analisis data realisasi belanja ini dilakukan dengan membagi unit kerja di Provinsi DKI Jakarta menjadi beberapa klaster berdasarkan karakteristiknya. Setelah dilakukan Clustering menggunakan metode K-Means Clustering, ditemukan beberapa klaster yang mempunyai karakteristik serupa dalam realisasi belanja.

Data yang mendukung pentingnya penelitian ini dilakukan berasal dari fakta yang dijumpai dilokasi penelitian serta sumber-sumber jurnal terkait. Tujuan dari data yang disajikan dalam bentuk tabel adalah untuk melakukan analisis pengelompokan realisasi belanja dari unit kerja di Provinsi DKI Jakarta tahun 2020 dengan menggunakan metode K-Means Clustering.

Tabel 1. Dataset Awal

Q1	Q2	Q3	Q4	Q5
Kepulauan Seribu	SUKU DINAS PENDIDIKAN KABUAPTEN	9162663820	6485470130	70.8
Kepulauan seribu	SUKU DINAS KESEHATAN KABUPATEN ADMINISTRASI KEPULAUAN SERIBU	2300654566	1390878177	60.5
Kepulauan seribu	PUSAT KESEHATAN MASYARAKAT KECAMATAN KEP. SERIBU UTARA	16862193080	14982477547	88.9
…	…	…	…	…
Jakarta utara	KELURAHAN KALIBARU	12479056839	11631393228	93.2

Dalam tabel yang disajikan, Q1 mengacu pada kota atau kabupaten, Q2 mencerminkan unit kerja, Q3 menunjukkan target yang ditetapkan, Q4 mencatat realisasi atau pencapaian aktual, dan Q5 menggambarkan persentase capaian terhadap target yang telah ditetapkan. Informasi yang tersedia dalam tabel tersebut memberikan gambaran singkat tentang kinerja dan pencapaian setiap unit kerja di berbagai kota atau kabupaten, dengan persentase mencerminkan sejauh mana target telah tercapai. Dilihat dari tabel 1 diatas, maka dijelaskan sebagai berikut :

1. Kota_kab, berisi nama kota atau kabupaten di Provinsi DKI Jakarta.
2. Unit kerja, berisi nama unit kerja di Provinsi DKI Jakarta.
3. Target, berisi pencapaian target yang diharapkan.
4. Realisasi, berisi jumlah uang yang telah digunakan atau terealisasi dalam penyelenggara kegiatan unit kerja tersebut.
5. Persentase, berisi nilai persentase dari target dan realisasi.

Realisasi belanja di Provinsi DKI Jakarta tahun 2020 dikelompokan dalam penelitian ini menggunakan metode K-Means Clustering.[4]

2. TINJAUAN PUSTAKA

2.1. Landasan Teori

2.1.1. Data Mining

Data mining adalah proses ekstrasi informasi bermanfaat dari kumpulan data dengan tujuan menemukan pola atau hubungan yang menarik.[5]. Tujuan dari data mining adalah untuk menemukan pola atau hubungan dalam data yang dapat membantu pengambilan keputusan lebih baik. Proses ini menggunakan teknik seperti clustering, regresi, dan klasifikasi.

Berikut perbedaan antara teknik clustering, regresi dan klasifikasi adalah sebagai berikut :

a) Clustering
   Termasuk jenis unsupervised learning dan melibatkan mengelompokan data ke dalam kelompok yang berbeda satu sama lain dan memiliki ciri yang sebanding. Tidak ada label yang ditentukan dalam clustering. Beberapa algoritma clustering populer meliputi K-Means, DBSCAN, dan Hierarchical clustering. Clustering adalah metode pengelompokan data yang memisahkan data menjadi beberapa kelompok tertentu yang diinginkan.[6]
b) Regresi
   Regreasi adalah metode analisis data yang digunakan untuk menentukan hubungan antara dua variabel atau lebih. Tujuan dari regresi adalah untuk menyederhanakan hubungan antara variabel-variabel tertentu. Tujuan regresi adalah untuk memperkirakan nilai variabel tertentu berdasarkan nilai variabel lain yang terkait dengannya. [6]
c) Klasifikasi
   Klasifikasi adalah termasuk kategori supervised learning dan mencakup mengelompokkan data ke dalam kelompok dengan label atau kategori. Dalam klasifikasi, data sudah memiliki label atau kategori yang diberikan sebelumnya. Beberapa algoritma yang populer untuk klasifikasi meliputi Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes Classifier, dan lain-lain.

Secara singkat, regresi digunakan untuk menyederhanakan hubungan antara variabel, klasifikasi digunakan untuk mengelompokkan data berdasarkan label yang sudah ada, dan clustering digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan seberapa mirip data itu sendiri. Data mining dibagi menjadi beberapa tahap, menurut peneliti [7]

Tahapan data mining adalah rangkaia proses penambangan (mining) data terdiri dari beberapa tahap, seperti :

1. Pembersihan data
   untuk menghilangkan data yang tidak konsisten dan suara (noise).
2. Integrasi data
   untuk menggabungkan data dari berbagai sumber.
3. Transformasi data
   untuk mengubah data menjadi bentuk yang sesuai untuk penambangan (mining).
4. Aplikasi teknik penambangan (mining)
5. Evaluasi pola yang ditemukan
   untuk menemukan informasi yang menarik atau bernilai.

Presentasi pengetahuan dengan teknik evaluasi.

2.1.2. K-Means Clustering

K-Means Clustering adalah metode pembelajaran mesin otonom yang mengelompokan data menjadi k kelompok eksklusif. Algoritma K-means digunakan untuk mengelompokan data dengan

membaginya kedalam grup yang berbeda-beda, didasarkan pada kesamaan karakteristiknya, dan digunakan untuk mengelompokan data menjadi beberapa kelompok [8].

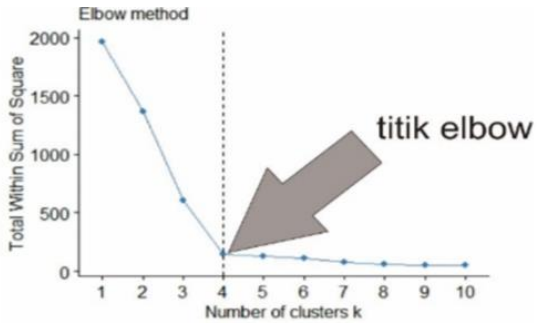
2.1.3. Metode Elbow

Metode Elbow melihat hasil presentase perbandingan di antara banyaknya cluster yang akan menghasilkan bentuk siku pada suatu titik.[9] Teknik elbow adalah metode yang digunakan untuk menentukan jumlah cluster yang optimal dalam algoritma K-Means. Teknik ini didasarkan pada plot SSE (Sum of Squared Errors) terhadap jumlah cluster. SSE adalah jumlah kuadrat jarak antara setiap titik data dengan pusat cluster terdekatnya. Plot SSE terhadap jumlah cluster akan membentuk kurva yang menurun, dimana SSE akan semakin kecil seiring dengan penambahan jumlah cluster.

Rumus teknik Elbow untuk menghitung jumlah cluster optimal adalah sebagai berikut :

1. Hitung SSE (Sum of Squared Errors) untuk setiap cluster (k).
2. Temukan titik elbow (elbow point) di antara kurva SSE menurun dan kemudian menyatakan.
3. Jumlah cluster pada titik elbow adalah jumlah cluster optimal.

Penelitian yang dilakukan oleh (Ayu et al., 2019) semakin kecil nilai SSE, semakin baik kualitas cluster. Berikut contoh grafiknya :



Gambar 1. Penentuan jumlah cluster terbaik dengan 330 data menggunakan elbow

2.1.4. Python

Menurut penelitian yang dilakukan oleh [10] Python adalah bahasa pemograman tinggi yang dipahami dengan cepat karena memiliki otomatisasi manajemen memori. Studi tersebut menemukan bahwa Python adalah bahasa pemograman tinggi yang dapat melakukan eksekusi direktif (interpretatif) menggunakan pemograman berorientasi objek dan semantik dinamis untuk meningkatkan keterbacaan sysntax.

Tujuan penerapan clustering pada Jupyter Notebook dengan menggunakan Python adalah untuk membuat proses data yang besar menjadi lebih mudah, dan library yang dimiliki membuat pengolahan dan visualisasi data lebih mudah.[1]

3. METODE PENELITIAN

3.1. Sumber Data

Dalam melakukan penelitian ini peneliti akan dihadapkan pada suatu data, pada penelitian ini untuk mengetahui seberapa efektif nya data yang akan digunakan untuk clustering. Peneliti mengambil realisasi belanja di Provinsi DKI Jakarta pada tahun 2020 dari website resmi Open Data Jakarta. Data ini beralamat di Jl. Medan Merdeka Selatan No. 8-9 Blok H Lt. 13, Jakarta, Indonesia. Peneliti mengumulkan data melalui link internet https://data.jakarta.go.id pada hari Rabu, 06 Desember 2023 pada pukul 20.18 WIB. Organisasi yang bertanggung jawab pada data ini yaitu Badan Pengelola Keuangan Daerah atau Badan yang mengurus keuangan daerah di DKI Jakarta.

3.2. Teknik Pengumpulan Data

Beberapa jenis pengumpulan data adalah wawancara, observasi, kusioner, dan studi dokumentasi. Selain itu, metode ini menunjukkan cara data digunakan melalui pengambilan, wawancara, pengamatan, dan tes dokumentasi.

3.2.1. Studi Dokumentasi

Pembuatan rencana keuangan tahunan seperti anggaran pendapatan dan belanja negara (APBN) adalah salah satu dari banyak aspek yang dapat dilakukan dalam studi dokumentasi terkait dengan pengelolaan anggaran realisasi pemerintah. Selain itu, Open Data Jakarta, sebagai inisiatif dari Pemprov DKI Jakarta, menyediakan berbagai data dari berbagai unit pemerintah Provinsi DKI Jakarta.

3.2.2. Studi Literatur

Dalam penggunaan metode K-Means Clustering untuk mengimplementasikan realisasi belanja di Provinsi DKI Jakarta. Bahwa metode K-Means merupakan metode pengelompokan non-hierarchical yang bertujuan untuk mengelompokan objek mulai dari identifikasi data yang akan di-cluster.

3.3. Tahapan Perancangan

Tahapan perancangan ini merupakan tahapan atau langkah awal dan persiapan dalam menyelesaikan penelitian, yang meliputi perencanaan kegiatan, pemilihan topik, dan merumuskan masalah yang akan diteliti. Berikut adalah tahapan perancangan proses analisis K-Means Clustering pada Python:

1. Import Library
2. Import Dataset
3. Data Preprocessing
4. Features Scalling
5. K-Means Clustering
6. Visualisasi Hasil K-Means Clustering

4. HASIL DAN PEMBAHASAN

4.1. Menerapkan metode K-Means Clustering menggunakan bahasa pemograman Python pada realisasi belanja unit kerja di Provinsi DKI Jakarta 2020

a) Impor library

Informasi tentang realisasi belanja di Provinsi DKI Jakarta tahun 2020 disimpan dalam file data CSV. Impor library menggunakan pandas, sklearn, dan library visual seperti matplotlib dan seaborn.

```
# library
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
```

Gambar 2. Impor library

b) Impor Dataset

Memuat data CSV dalam DataFrame menggunakan pandas. Serta, menampilkan 5 baris pada dataset.

```
# Import Dataset
df = pd.read_csv('/content/data-realisasi-belanja-menurut-unit-kerja-di-provinsi-dki-jakarta-tahun-2020.csv')
df.head()
```

	kota_kab	unit_kerja	target	realisasi	persentase
0	kepulauan seribu	SUKU DINAS PENDIDIKAN KABUPATEN	9.162664e+09	6.485470e+09	70.8
1	kepulauan seribu	SUKU DINAS KESEHATAN KABUPATEN ADMINISTRASI KE...	2.300655e+09	1.390878e+09	60.5
2	kepulauan seribu	PUSAT KESEHATAN MASYARAKAT KECAMATAN KEP. SERI...	1.686219e+10	1.498248e+10	88.9
3	kepulauan seribu	PUSAT KESEHATAN MASYARAKAT KECAMATAN KEP. SERI...	1.697437e+10	1.621588e+10	95.5
4	kepulauan seribu	SUKU DINAS SUMBER DAYA AIR KABUPATEN ADMINISTR...	5.958359e+10	5.806352e+10	97.4

Gambar 3. Hasil impor dataset

c) Data Preprocessing

Pada dataset realisasi belanja menurut unit kerja di Provinsi DKI Jakarta tahun 2020, tidak ada data kosong dan penghapusan data.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 523 entries, 0 to 522
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   kota_kab    523 non-null    object
 1   unit_kerja  523 non-null    object
 2   target      523 non-null    float64
 3   realisasi   523 non-null    float64
 4   persentase  523 non-null    float64
dtypes: float64(3), object(2)
memory usage: 20.6+ KB
```

Gambar 4. Informasi data

d) Features Scalling

Proses normalisasi atau transformasi data diperlukan untuk memastikan skala yang serupa, skala yang digunakan untuk normalisasi data adalah MinMaxScaler.

Tabel 2. Hasil MinMaxScaler

Cluster	target	realisasi
0	0.03959023	0.02990603
1	0.00974771	0.00620392
2	0.07307507	0.06943754

e) K-Means Clustering

Peneliti menggunakan tiga cluster, yakni cluster 0, cluster 1, dan cluster 2. Untuk melakukan pengelompokan pada atribut target dan realisasi. Penggunaan tiga cluster ini bertujuan untuk analisis dan evaluasi yang lebih mendalam terhadap data. Dibawah ini hasil centroid pada tiga cluster :

Tabel 3. Hasil centroid

Cluster	Target	Realisasi
0	0.03420079	0.03488608
1	0.63187446	0.61831973
2	0.1784636	0.17370444

Setelah mendapatkan hasil centroid, maka berikut hasil pembagian data berdasarkan tiga cluster, yaitu Cluster 0, Cluster 1, dan Cluster 2 :

Tabel 4. Hasil cluster 0

Q1	Q2	Q3	Q4	Q5	Q6
Kepulauan seribu	SUKU DINAS PENDIDIKAN KABUPATEN	0.039590	0.029906	70.8	0
Jakarta pusat	KELURAHAN MANGGA DUA SELATAN	0.043719	0.045729	97.8	0
…	…	…	…	…	…
Jakarta pusat	KELURAHAN KEBON KOSONG	0.043105	0.044693	96.9	0
JUMLAH UNIT KERJA CLUSTER 0 ADALAH 436					

Tabel 5. Hasil cluster 1

Q1	Q2	Q3	Q4	Q5	Q6
Jakarta selatan	SUDIN PEMBERDAYAAN, PERLINDUNGAN ANAK DAN PENGENDALIAN PENDUDUK	0.438989	0.469626	99.7	1
Jakarta timur	SUDIN PEMBERDAYAAN, PERLINDUNGAN ANAK DAN PENGENDALIAN PENDUDUK	0.682527	0.730161	100.0	1
…	…	…	…	…	…
Jakarta timur	SUKU DINAS LINGKUNGAN HIDUP KOTA ADMINISTRASI JAKARTA TIMUR	0.721667	0.748770	96.9	1
JUMLAH UNIT KERJA CLUSTER 1 ADALAH 18					

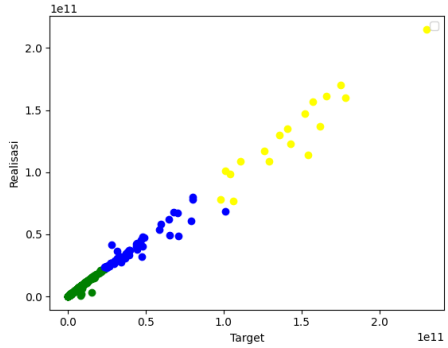
Tabel 6. Hasil cluster 2

Q1	Q2	Q3	Q4	Q5	Q6
Jakarta selatan	Pusat Kesehatan Masyarakat Kecamatan Setia Budi	0.131675	0.129583	92.0	2
Jakarta selatan	Pusat Kesehatan Masyarakat Keccamatan Tebet	0.142771	0.146161	95.7	2
…	…	…	…	…	…
Jakarta selatan	Pusat Kesehatan Masyarakat Kecamatan Pancoran	0.132384	0.131262	92.7	2
JUMLAH UNIT KERJA CLUSTER 2 ADALAH 69					

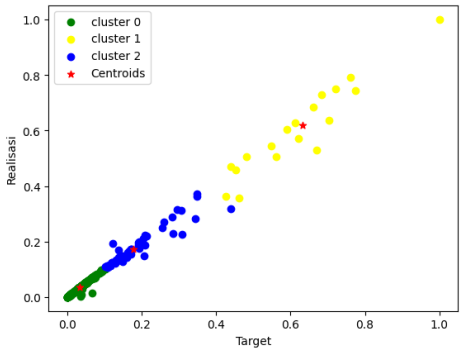
Tabel tersebut menyajikan informasi mengenai kota atau kabupaten (Q1), unit kerja (Q2), target (Q3), realisasi (Q4), persentase pencapaian (Q5), dan cluster (Q6). Data kota atau kabupaten (Q1) mengidentifikasi lokasi geografis dari unit kerja (Q2) yang terlibat. Unit kerja (Q2) merupakan entitas atau divisi yang dianalisis dalam mencapai target (Q3). Target (Q3) adalah tujuan yang ditetapkan untuk dicapai oleh unit kerja, sedangkan realisasi (Q4) mencerminkan pencapaian aktual terhadap target tersebut. Persentase (Q5) menggambarkan sejauh mana target telah tercapai, sementara cluster (Q6) memberikan klasifikasi atau pengelompokan berdasarkan karakteristik atau kriteria tertentu. Keseluruhan, tabel ini menyediakan gambaran singkat tentang performa unit kerja dalam mencapai target di berbagai kota atau kabupaten, dengan memperhitungkan persentase pencapaian dan pengelompokan dalam cluster tertentu.

f) Visualisasi Hasil K-Means Clustering

Visualisasi hasil K-Means dengan scatterplot untuk melihat sebaran unit kerja dalam tiap cluster dan centroid.



Gambar 5. Hasil persebaran cluster



Gambar 6. Hasil persebaran centroid

4.2. Menentukan Jumlah Cluster Optimal pada K-Means Clustering menggunakan Metode Elbow dan SSE

4.2.1. Inisialisasi jumlah cluster optimal

Untuk menentukan jumlah cluster yang diinginkan, mulailah dengan menginisialisasi jumlah cluster optimal. Karena jumlah cluster yang tepat mempengaruhi hasil dari algoritma K-Means, penentuan jumlah cluster yang tepat sangat penting. Peneliti menggunakan kode berikut untuk menentukan jumlah cluster yang optimal :

```
# pengelompokan menggunakan metode Elbow dengan SSE
k_rng = range(1, 10)
sse = []

for k in k_rng:
  kmeans = KMeans(n_clusters=k)
  kmeans.fit(df[['target', 'realisasi']])
  sse.append(kmeans.inertia_)
```

Gambar 7. Kode inisialisasi jumlah cluster optimal

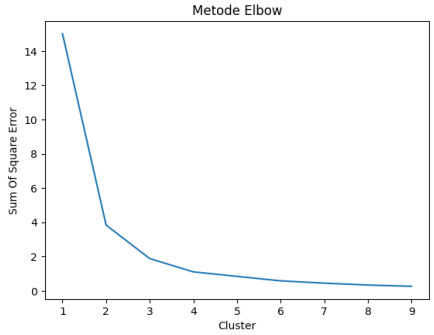
4.2.2. Hasil nilai SSE

Dalam algoritma clustering K-Means, nilai Sum Of Square Errors (SSE) adalah penjumlahan dari seluruh jarak masing-masing data dengan titik pusat clusternya. Dengan demikian, SSE dapat membantu dalam menentukan jumlah klaster yang optimal.

Tabel 7. Hasil nilai SSE

Cluster	Nilai SSE (Sum of Square Errors)
1	15.003979862837099
2	3.835809912711171
3	1.8818840676710162
4	1.1064687880917778
5	0.8443250277889012
6	0.582873023633755
7	0.44832308594021486
8	0.33904150837533764
9	0.2626653061270848

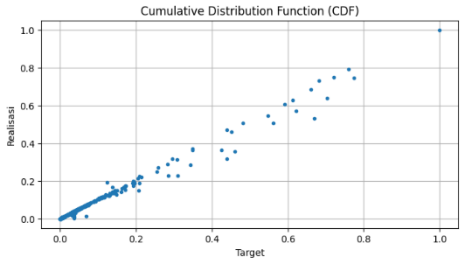
1. Visualisasi jumlah cluster optimal
   Visualisasi ini melibatkan pembuatan grafik yang menujukkan perubahan dalam nilai SSE dengan jumlah cluster (k). Berikut visualisasinya :



Gambar 8. Visualisasi hasil cluster optimal

4.3. Menampilkan visual CDF pada K-Means Clustering

Menampilkan visual Cumulative Distribution Function (CDF) pada hasil clustering menggunakan algoritma K-Means pada kolom target dan realisasi.



Gambar 9. Hasil visual CDF

4.4. Mengidentifikasi pola dan karakteristik kelompok unit kerja berdasarkan realisasi belanja pada cluster tinggi dan rendah

Pada tahap ini, dilakukan identifikasi pola dan karakteristik kelompok unit kerja berdasarkan realisasi belanja pada cluster tinggi dan rendah. Melibatkan analisis cluster untuk mengelompokan unit kerja berdasarkan pola belanja, kemudian mengidentifikasi karakteristik cluster tinggi dan rendah. Berikut langkah-langkah untuk mengidentifikasi pola dan karakteristik kelompok unit kerja berdasarkan realisasi belanja pada cluster tinggi dan rendah

4.5. Identifikasi unit kerja berdasarkan cluster tinggi dan rendah

Identifikasi unit kerja berdasarkan cluser tinggi dan cluster rendah adalah analisis data yang bertujuan untuk mengelompokan unit kerja kedalam dua kategori utama, yaitu cluster tinggi dan cluster rendah. Proses ini melibatkan teknik clustering untuk mengidentifikasi pola atau kesamaan di antara unit kerja dan memisahkan kedalam kelompok-kelompok yang serupa.

Tabel 8. Hasil Cluster Tinggi

Q1	Q2	Q3	Q4	Q5	Q6	Q7
Kepulauan seribu	PUSAT KESEHATAN MASYARAKAT KECAMATAN KEP. SERIBU UTARA	0.073075	0.069438	88.9	0	Tinggi
Kepulauan seribu	PUSAT KESEHATAN MASYARAKAT KECAMATAN KEP. SERIBU SELATAN	0.073563	0.075176	95.5	0	Tinggi
…	…	…	…	…	…	…
Jakarta utara	KELURAHAN KALIBARU	0.054013	0.053847	93.2	0	Tinggi
JUMLAH CLUSTER TINGGI = 503						

Tabel 9. Hasil Cluster Rendah

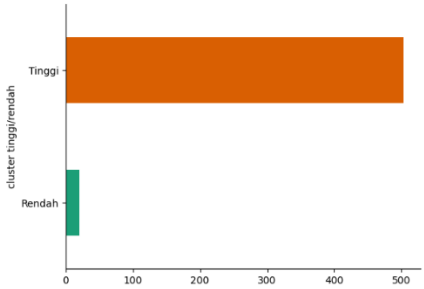
Q1	Q2	Q3	Q4	Q5	Q6	Q7
Kepulauan seribu	SUKU DINAS KABUPATEN	0.039590	0.029906	70.8	0	Rendah
Kepulauan seribu	SUKU DINAS KESEHATAN KABUPATEN ADMINISTRASI KEPULAUAN SERIBU	0.009748	0.006204	60.5	0	Rendah
Kepulauan seribu	SUKU DINAS PERUMAHAN RAKYAT DAN KAWASAN PERMUKIMAN KABUPATEN	0.004824	0.002608	52.9	0	Rendah
…	…	…	…	…	…	…
Jakarta utara	SUKU BADAN PENDAPATAN DAERAH KOTA JAKARTA UTARA DAN KABUPATEN	0.000461	0.000303	74.2	0	Rendah
JUMLAH CLUSTER RENDAH = 20						

Dalam tabel tersebut, Q1 merujuk pada kota atau kabupaten yang menjadi fokus, sedangkan Q2 mengidentifikasi unit kerja yang terlibat dalam pelaksanaan tugas. Q3 mencerminkan target yang ditetapkan untuk mencapai sukses, sementara Q4 merepresentasikan realisasi atau pencapaian aktual dalam pelaksanaan proyek atau tugas tersebut. Q5 memberikan informasi persentase pencapaian terhadap target yang telah ditetapkan. Q6 mencatat cluster atau kelompok tertentu yang mungkin memengaruhi hasil. Terakhir, Q7 menunjukkan apakah cluster tersebut termasuk dalam kategori tinggi atau rendah, memberikan gambaran mengenai tingkat kesulitan atau keberhasilan yang dapat diantisipasi.

4.6. Visualisasi hasil cluster tinggi dan rendah

Visualisasi hasil cluster tinggi dan rendah adalah proses untuk memvisualisasikan hasil clustering pada

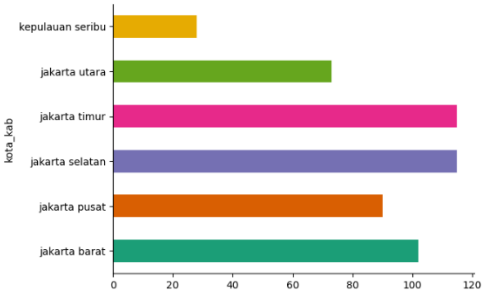
data pengeluaran pemerintah (realisasi) pemerintah di Provinsi DKI Jakarta berdasarkan cluster tinggi dan rendah. Visualisasi ini dapat membantu pemahaman data. Berikut hasil visualisasinya :



Gambar 10. Hasil visualisasi cluster tinggi dan rendah

4.7. Visualisasi Jumlah Data berdasarkan Kota atau Kabupaten

Visualisasi ini memberikan gambaran tentang sebaran atau kontribusi data untuk setiap kota atau kabupaten di Provinsi DKI Jakarta tahun 2020. Berikut hasil visualisasinya :



Gambar 11. Hasil visualisasi persebaran Kota atau Kabupaten

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa penggunaan metode K-Means Clustering dalam menganalisis realisasi belanja di Provinsi DKI Jakarta pada tahun 2020 menghasilkan tiga cluster, yaitu cluster 0, 1, dan 2. Cluster 0, yang terdiri dari 436 unit kerja, berhasil mencapai target pengeluaran, sementara cluster 1 dengan 18 unit kerja dan cluster 2 dengan 69 unit kerja masih memerlukan peningkatan untuk mencapai target pengeluaran tahun 2020. Dengan menerapkan teknik clustering, pemerintah dapat mengidentifikasi unit kerja yang membutuhkan perhatian khusus, menganalisis pola, dan karakteristik kelompok unit kerja berdasarkan realisasi belanja pada cluster tinggi dan rendah. Pengelolaan yang tepat antara realisasi dan nilai persentase dapat disesuaikan guna mencapai target yang telah ditetapkan. Untuk mendapatkan pemahaman yang lebih komprehensif terkait implementasi pola pengeluaran realisasi belanja di Provinsi DKI Jakarta, penelitian mendatang dapat mempertimbangkan penggunaan teknik clustering tambahan dan melakukan analisis tambahan terhadap perubahan pola pengeluaran dari tahun ke tahun.

DAFTAR PUSTAKA

[1] I. D. Murti Suyoto, T. Rachmadi, and L. T. Parulian, “MENENTUKAN CLUSTER YANG TEPAT DENGAN K-MEANS DALAM RANGKA MENGUKUR EFEKTIVITAS PELAKSANAAN ANGGARAN PADA KEMENTERIAN AGRARIA DAN TATA RUANG/BADAN PERTANAHAN,” *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 13–22, Jun. 2022, doi: 10.37365/jti.v8i1.126.

[2] 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.

[3] M. R. Wardhana, A. Triayudi, and N. Hayati, “Analisis Faktor Calon Nasabah PT. Bank Central Asia dalam Pembuatan Rekening Online menggunakan Metode K-Means Clustering Studi Kasus Wisma Asia BCA,” *Jurnal Teknologi Informasi dan Komunikasi)*, vol. 6, no. 1, p. 2022, 2022, doi: 10.35870/jti.

[4] I. Y. Ulya, S. Anwar, and U. Zakia, “E-JURNAL EKONOMI DAN BISNIS UNIVERSITAS UDAYANA”, [Online]. Available: https://ojs.unud.ac.id/index.php/EEB/index

[5] A. Winarta and W. J. Kurniawan, “OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON,” *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1, 2021.

[6] J. Hutagalung and F. Sonata, “Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 1187, Jul. 2021, doi: 10.30865/mib.v5i3.3113.

[7] D. Marlina and M. Bakri, “Penerapan data mining untuk memprediksi transaksi nasabah dengan algoritma c4.5,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, 2021.

[8] P. Alkhairi and A. P. Windarto, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara*. [Online]. Available: https://seminar-id.com/semnas-sainteks2019.html

[9] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, “Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Penelitian,” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 16, no. 1, p. 13, Mar. 2021, doi: 10.30872/jim.v16i1.5182.

[10] M. Karunia Rahmadhika and A. M. Thantawi, “Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade .”