

## OPTIMASI NILAI DAVIES BOULDIN INDEX PADA PROGRAM PENDAFTARAN TANAH SISTEMATIS LENGKAP (PTSL) MENGUNAKAN ALGORITMA K-MEANS DAN PCA

Muhammad Hilmy Naufan<sup>1)</sup>, Rudi Kurniawan<sup>2)</sup>, Tati Suprapti<sup>3)</sup>

<sup>1,2,3)</sup> Jurusan Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No. 10 B Majasem, Kec. Kesambi, Kota Cirebon, Jawa Barat, Indonesia

E-mail : <sup>1)</sup>[muhammadhilmynaufan28@gmail.com](mailto:muhammadhilmynaufan28@gmail.com), <sup>2)</sup>[rudi226ikmi@gmail.com](mailto:rudi226ikmi@gmail.com),

<sup>3)</sup>[tatisuprapti112004@gmail.com](mailto:tatisuprapti112004@gmail.com)

### ABSTRAK

Pada data program Pendaftaran Tanah Sistematis Lengkap (PTSL) ditemukan sebuah fenomena masalah terkait tidak adanya tindak lanjut dikarenakan masih terhambat dalam proses pengelompokan, oleh karena itu penelitian ini dengan mengembangkan solusi berbasis teknologi informasi berupa teknik *Clustering* yang mengimplementasikan algoritma *K-Means* dikombinasikan dengan *Principal Component Analysis* (PCA), pada akhirnya dapat mencapai sebuah tujuan optimasi yaitu proses mengoptimalkan sesuatu menjadi lebih baik dengan melihat hasil akurasi dari nilai *Davies Bouldin Index* (DBI) terendah yang mendekati angka 0. Metode penelitian mencakup pengumpulan dan pemrosesan data dari desa Bandorasawetan, Kecamatan Cilimus, Kabupaten Kuningan hingga tahap evaluasi dengan menggunakan *Davies Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa *Cluster* yang optimal dicapai pada  $K = 5$  dengan penggunaan PCA melalui pendekatan *Fixed Number* yang mengatur berdasarkan jumlah komponen utama yang ingin dipertahankan yaitu = 1, diketahui hasilnya mempertahankan berupa atribut NJOP Bangunan. Atribut ini memiliki distribusi yang lebih terpusat dalam satu *Cluster*, dalam arti memiliki pola yang konsisten dengan nilai DBI mencapai 0.049, memiliki kinerja lebih baik dibandingkan *K-Means* tanpa PCA dengan nilai DBI sebesar 0.466. Studi ini diharapkan dapat meningkatkan kualitas dalam pengelolaan data pertanahan agar menjadi lebih efektif.

**Kata kunci** : Pendaftaran Tanah Sistematis Lengkap, *Clustering*, *Davies Bouldin Index*

### ABSTRACT

*In the Complete Systematic Land Registration (PTSL) program data, a problem phenomenon was found related to the absence of follow-up due to still being hampered in the grouping process, therefore this study by developing an information technology-based solution in the form of Clustering techniques that implement the K-Means algorithm combined with Principal Component Analysis (PCA), can ultimately achieve an optimization goal, namely the process of optimizing something to be better by looking at the accuracy results of the lowest Davies Bouldin Index (DBI) value approaching 0. The research method includes data collection and processing from Bandorasawetan Village, Cilimus District, Kuningan Regency to the evaluation stage using the Davies Bouldin Index (DBI). The results showed that the optimal Cluster was achieved at  $K = 5$  with the use of PCA through the Fixed Number approach which regulates based on the number of main components to be maintained, namely = 1, it is known that the results maintain the NJOP Building attribute. This attribute has a more centralized distribution in one Cluster, in the sense that it has a consistent pattern with a DBI value reaching 0.049, has better performance than K-Means without PCA with a DBI value of 0.466. This study is expected to improve the quality of land data management to make it more effective.*

**Keywords** : Complete Systematic Land Registration, *Clustering*, *Davies Bouldin Index*

## 1. PENDAHULUAN

### 1.1. Latar Belakang Penelitian

Pendaftaran tanah merupakan kegiatan yang dilakukan oleh pemerintah dalam rangka menjamin kepastian dan perlindungan hukum terhadap hak-hak atas tanahnya [1]. Ini sebagai langkah strategis untuk mengurangi konflik agraria dan pertanahan, serta dapat meningkatkan pertumbuhan ekonomi, untuk mendukung tercapainya keadilan dalam kepemilikan lahan [2]. Seperti yang telah diketahui, program PTSL dapat menggunakan teknologi informasi untuk mempercepat pendaftaran dan pengumpulan data pertanahan [3]. Kemajuan teknologi informasi ini membawa dampak signifikan di bidang informatika terutama dalam data *Mining* dan analisis data, telah membuka peluang baru untuk menerapkan teknik *Clustering* yang merupakan proses pengelompokan berdasarkan sejumlah kesamaan [4]. Metode *Clustering* dengan menggunakan algoritma *K-Means* dapat membagi data ke dalam sejumlah *Cluster* yang ditentukan, data dalam satu kelompok memiliki karakteristik yang sama tetapi memiliki karakteristik yang berbeda dengan kelompok lainnya, namun algoritma ini hanya bekerja pada atribut numerik [5]. Agar hasil lebih optimal, maka sebelum proses *Clustering* dengan *K-Means* dilakukan, diperlukan *Principal Component Analysis* (PCA), yaitu teknik *Dimensionality Reduction* untuk menyederhanakan data dengan membantu dalam menentukan variabel-variabel paling signifikan dan berpengaruh, namun tetap mempertahankan sebagian besar informasi [6].

Permasalahan yang relevan untuk diteliti dalam konteks ini adalah terkait penerapan teknik pengelompokan, yang sering kali terhambat dalam menentukan jumlah *Cluster* yang optimal, terutama ketika berhadapan dengan data tanah yang kompleks dan banyak seperti pada data program PTSL. Dalam penelitian ini meskipun dengan PCA dapat menyederhanakan pemrosesan data, namun keseimbangan penggunaan *Dimensionality Reduction* dengan *Fixed Number* pada jumlah *Number of Components* menjadi pertimbangan untuk menerapkan PCA. Lalu membutuhkan penetapan

jumlah *Cluster* yang disebut nilai *K*, perlu diatur secara acak setiap kali proses dilakukan. Oleh karena itu, beberapa kali percobaan diperlukan untuk menemukan jumlah inisialisasi yang lebih baik [7]. Seiring dengan terus berkembangnya bidang informatika, sangat penting untuk meningkatkan sebuah pemahaman khususnya dalam menggunakan kombinasi *K-Means* dan *Principal Component Analysis* (PCA) pada data agraria dan pertanahan, hal ini menunjukkan bahwa penelitian lebih lanjut di bidang ini diperlukan untuk menghasilkan solusi yang lebih optimal.

Penelitian sebelumnya berfokus pada penerapan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data produksi pertanian di Kabupaten Bojonegoro, dan metode *K-Means Clustering* untuk pengelompokan wilayah berdasarkan hasil produksi padi sawah tahun 2017-2020. Hasilnya menunjukkan bahwa dengan penetapan pada jumlah *Number of Components* atau berfokus pada komponen utama dapat menghasilkan DBI paling optimal. dataset yang direduksi menjadi satu komponen utama dan dikelompokkan dalam tiga *Cluster* menghasilkan DBI paling optimal yaitu 0.4072. Secara teoritis, penelitian ini mendukung penerapan PCA dalam meningkatkan efisiensi dan kualitas *Clustering*, sementara secara praktis, hasil pengelompokan ini dapat membantu pemerintah dalam pengambilan keputusan terkait pengelolaan sektor pertanian [8].

Tujuan utama dari penelitian ini adalah untuk optimasi nilai *Davies Bouldin Index* (DBI) pada model pengelompokan data program Pendaftaran Tanah Sistematis Lengkap (PTSL), mengembangkan solusi berbasis teknologi informasi menggunakan algoritma *K-Means* dan *Principal Component Analysis* (PCA). Penelitian ini tidak hanya berkontribusi pada wacana akademis tetapi juga menawarkan solusi praktis bagi para pembuat kebijakan dan administrator pertanahan yang ingin meningkatkan hasil pendaftaran tanah.

### 1.2. Landasan Teori

#### a. Pendaftaran Tanah Sistematis Lengkap

Program Pendaftaran Tanah Sistematis Lengkap (PTSL) merupakan salah satu program Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional untuk mensertifikatkan tanah di seluruh Indonesia. Dengan terdaftarnya hak atas tanah berarti secara hukum diakui hubungan kepemilikan antara subjek dan objek secara administratif. Dalam mewujudkan Indonesia lengkap terdaftar, dimulai dengan melakukan perbaikan kualitas data pertanahan dan diproses melalui pendaftaran tanah sistematis lengkap (PTSL) yang pada akhirnya menuju pendaftaran desa lengkap [9].

b. *Data Mining*

*Data Mining* merupakan proses yang menggunakan teknik statistik, kecerdasan buatan, matematika, dan *Machine Learning* untuk mengolah serta mengidentifikasi informasi yang bisa digunakan sebagai pengetahuan yang berasal dari sebuah *Big data* [10].

c. *Clustering*

*Clustering* adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam *Subset*, sehingga dalam setiap data memiliki arti yang bermanfaat, dimana dalam *Cluster* terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan benda yang terdapat pada *Cluster* lainnya [11].

d. *K-Means*

*K-Means* merupakan algoritma *Clustering* yang berulang-ulang dengan menetapkan nilai *K* secara acak, dimana nilai tersebut menjadi pusat dari *Cluster* atau disebut sebagai *Centroid*, *Mean* atau *Means*. Algoritma *K-Means* dalam implementasinya sangat mudah, cepat, sederhana dan mudah beradaptasi serta mempunyai kemampuan yang besar dalam mengolah data cukup besar, adapun yang menjadi kelemahan dalam algoritma *K-Means* yaitu saat menentukan jumlah *Cluster* awal, karena bergantung pada inisial data yang diberikan [12].

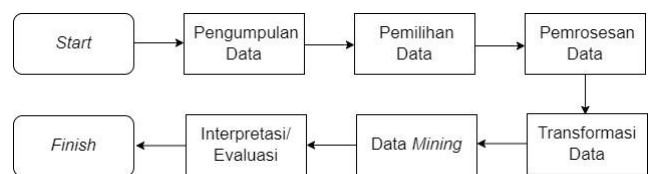
e. *Principal Component Analysis (PCA)*

*Principal Components Analysis (PCA)* adalah salah satu teknik yang memungkinkan representasi data berdimensi tinggi ke dalam bentuk dimensi rendah yang lebih mudah digunakan dengan menggunakan ketergantungan antara variabel tanpa kehilangan terlalu banyak informasi, ide utama dari analisis komponen utama adalah pengurangan dimensi, hal ini dapat dicapai dengan mentransformasi ke satu set variabel baru, yang disebut juga sebagai komponen utama (PC) [13].

## 2. METODE PENELITIAN

Dalam penelitian ini, metode yang digunakan adalah metode kuantitatif dengan pendekatan eksperimental, karena untuk meneliti terhadap suatu masalah yang sudah jelas tertera. Akan diuji untuk mengetahui hubungan sebab akibat dari variabel yang ada agar hasilnya bisa bersifat empiris. Metode penelitian kuantitatif merupakan penelitian yang memakai data yang berhubungan dengan angka-angka ditambahkan pada penekanan terhadap pengukuran hasil yang objektif disertai dengan analisis statistik menggunakan perangkat lunak yang sesuai untuk mengungkap hubungan atau perbedaan yang signifikan antar variabel [14].

Tahapan untuk metode atau teknik yang diusulkan divisualisasikan dalam Gambar 1 berikut :



**Gambar 1.** Tahapan Metode Penelitian

a. *Pengumpulan Data*

Dalam penelitian ini data yang digunakan merupakan teknik data sekunder, yang didapat secara observasi dari Pemerintah Desa Bandorasawetan, Kecamatan Cilimus, Kabupaten Kuningan, dan sudah bekerjasama dengan kantor Badan Pertanahan Nasional (BPN) Kabupaten Kuningan untuk mendapatkan data resmi dan informasi terkini tentang para pendaftar

Program Pendaftaran Tanah Sistematis Lengkap (PTSL) di desa tersebut. Dengan diambilnya data dari instansi tersebut dipastikan dapat dipertanggungjawabkan dengan jumlah total 775 data dan dikelompokkan dalam 17 atribut yaitu Nomor, Kabupaten, Kecamatan, Desa, Nomor Berkas, Nama, NIK, Alamat, NIB, Letak Tanah, Luas, Alas Hak, Tanggal Surat Pernyataan, No. SPT PBB, NJOP Tanah, NJOP Bangunan, dan Penggunaan Tanah.

b. Pemilihan Data

Pada tahap ini merupakan proses menganalisis data dari database dengan memilih data-data seperti apa saja yang kita butuhkan karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses data *Mining*. Proses seleksi dilakukan pada atribut data program Pendaftaran Tanah Sistematis Lengkap (PTSL) Di Bandorasawetan, Kecamatan Cilimus, Kabupaten Kuningan pada tahun 2024 sebanyak 17 atribut.

c. Pemrosesan Data

Dari data yang telah diseleksi sebelumnya akan dilihat apakah pada data tersebut terdapat data yang ganda, kosong dan *Missing*. Oleh karena itu dari data yang sudah dikumpulkan akan dibersihkan dari kesalahan atau ketidaklengkapan dengan melakukan proses *Cleaning* agar sesuai dengan format yang diinginkan.

d. Transformasi Data

Transformasi adalah tahap membuat data yang akan diproses menjadi sesuai dengan model ataupun algoritma yang ingin digunakan dalam tahap pengolahan data.

e. Data Mining

Pada tahap data *Mining*, data akan di proses menggunakan *Machine Learning*, dengan penerapan kombinasi *Principal Component Analysis* (PCA) dan algoritma *K-Means*.

f. Interpretasi/Evaluasi

Pada tahap interpretasi/evaluasi, data yang sudah diproses menggunakan *Machine Learning* akan diukur menggunakan operator kinerja untuk menghitung *Davies Bouldin Index* (DBI), dengan tujuan untuk memperoleh nilai K yang optimal.

### 3. HASIL DAN DISKUSI

Dalam pembahasan ini akan menguraikan proses optimasi nilai *Davies Bouldin Index* (DBI) pada model pengelompokkan data, yang akan mempengaruhi terbentuknya jumlah *Cluster* pada data program Pendaftaran Tanah Sistematis Lengkap (PTSL) menggunakan *Machine Learning* yaitu AI Studio 2024.0.1 dengan penerapan kombinasi algoritma *K-Means* dan *Principal Component Analysis* (PCA), lalu *Tools Microsoft Excel Home 2024* untuk proses penentuan atribut komponen utama.

#### 3.1. Pengumpulan Data

Data yang telah diperoleh pada program Pendaftaran Tanah Sistematis Lengkap (PTSL) di desa Bandorasawetan, Kecamatan Cilimus, Kabupaten Kuningan dapat dilihat secara singkat pada Tabel 1 berikut :

**Tabel 1.** Dataset Program PTSL 2024

No	No Berkas	Nama	...	Penggunaan Tanah
1.	9025	TAUFIK RACHMAT	...	Rumah Tinggal
2.	9026	MASKINAH	...	Rumah Tinggal
3.	9027	RASMIAH	...	Rumah Tinggal
4.	9028	HJ. AMINI	...	Rumah Tinggal
5.	9029	AHMAD SAROJI	...	Kebun
...	...	...	...	...
775	117296	PEMERIN TAH DESA BANDORA SAWETAN	...	Kebun

#### 3.2. Pemilihan Data

Pada tahap ini sebelum melakukan pemilihan data, maka perlu memasukkan operator “*Read Excel*” yang berfungsi untuk membaca data *Excel* pada program Pendaftaran Tanah Sistematis Lengkap (PTSL). Selanjutnya melakukan *Import* data. Operator *Read Excel* dapat dilihat pada Gambar 2 berikut :



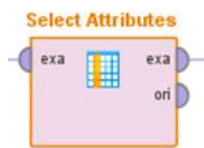
**Gambar 2.** Operator *Read Excel*

Pada Tabel 2 berikut merupakan parameter pada operator *Read Excel* untuk menampilkan sumber data utama dalam proses analisis :

**Tabel 2.** Parameter Operator *Read Excel*

No.	Parameter	Lokasi Penyimpanan
1.	<i>Excel File</i>	E:\PCA\DATASET-PTSL-2024-DESA BANDORASAWETAN.xlsx

Tahap berikutnya yaitu memasukkan operator “*Select Attributes*”, yang berguna untuk memilih atau menentukan atribut mana saja yang akan dipakai dalam analisis. Operator *Select Attributes* dapat dilihat pada Gambar 3 berikut :



**Gambar 3.** Operator *Select Attributes*

Namun, untuk memilih atribut tertentu dari dataset yang akan digunakan dalam analisis, perlu mengatur pada parameter *Select Attributes* dengan mengubah *Attributes filter type* menjadi *a Subset*. Dalam proses pengelompokkan dari total 17 atribut, yang digunakan hanya 10 atribut. Dapat dilihat pada Tabel 3 berikut :

**Tabel 3.** *Select Attributes*

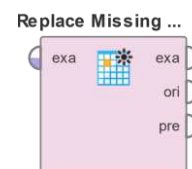
No.	<i>Attributes</i>	<i>Select Attributes</i>
1.	No	No Berkas
2.	Kecamatan	NIK
3.	Kabupaten	Alamat
4.	Desa	NIB
5.	Nama	Letak Tanah
6.	Alas Hak	Luas Tanah
7.	Tgl Surat Pernyataan	No SPT PBB
8.	-	NJOP Tanah
9.	-	NJOP Bangunan
10.	-	Penggunaan Tanah

Pada bagian *Select Attributes* terdapat 7 atribut yang tidak digunakan dalam analisis karena alasan tertentu.

- Atribut Nomor tidak diperlukan karena AI Studio 2024.0.1 sudah otomatis memberi nomor pada setiap baris data.
- Atribut Desa, Kabupaten, dan Kecamatan dianggap sebagai satu kesatuan yang mewakili wilayah administrasi, sehingga tidak perlu disertakan karena analisis lebih berfokus pada aspek lain.
- Atribut Tgl Surat Pernyataan dan Alas Hak tidak banyak memberikan variasi atau perbedaan yang berarti antara data, sehingga dianggap kurang relevan.
- Dengan mengabaikan atribut-atribut ini, analisis dapat lebih terfokus pada informasi yang benar-benar penting. Lalu terdapat 10 atribut yang akan digunakan karena atribut ini memiliki karakteristik yang beragam dan memberikan informasi yang lebih mendalam serta relevan untuk analisis.

### 3.3. Pemrosesan Data

Tahap Pemrosesan merupakan langkah yang dilakukan untuk membersihkan dan mempersiapkan data sebelum analisis dimulai. Dengan penggunaan operator “*Replace Missing Values*” dalam proses ini, dilakukan beberapa langkah penting seperti menangani data yang kosong, menghapus data yang duplikat, dan mengisi atau menangani nilai-nilai yang hilang agar formatnya sesuai dengan standar yang diperlukan. Operator *Replace Missing Values* dapat dilihat pada Gambar 4 berikut :



**Gambar 4.** Operator *Replace Missing Values*

Parameter pada operator *Replace Missing Values* menggunakan pengaturan standar tanpa ada perubahan. Untuk hasil yang diperoleh seperti yang ditampilkan pada Tabel 4 berikut :

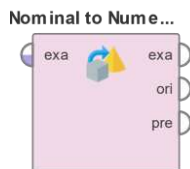
**Tabel 4.** Hasil *Missing Values*

No.	Atribut	Type Atribut	Missing Values
1.	No Berkas	Integer	0
2.	NIK	Polynomial	0
3.	Alamat	Polynomial	0
4.	NIB	Integer	0
5.	Letak Tanah	Polynomial	0
6.	Luas	Integer	0
7.	No SPT PBB	Polynomial	0
8.	NJOP Tanah	Integer	0
9.	NJOP Bangunan	Integer	0
10.	Penggunaan Tanah	Polynomial	0

Pada data program Pendaftaran Tanah Sistematis Lengkap (PTSL) tidak menunjukkan adanya perbedaan terkait data yang kosong, ganda, ataupun *Missing*. Karena data tersebut sudah memiliki kualitas yang baik sehingga data sudah siap untuk dianalisis.

### 3.4. Transformasi Data

Dalam tahap ini transformasi data akan merubah atribut yang telah dipilih ke dalam bentuk yang cocok untuk digunakan, diperlukan untuk mengatasi setiap atribut yang terdapat pada data program Pendaftaran Tanah Sistematis Lengkap (PTSL) yang bersifat numerik dan non-numerik. Pada pengelompokan menggunakan algoritma *K-Means* hanya bekerja pada atribut numerik, oleh karena itu memerlukan operator “*Nominal to Numerical*”, sehingga data program PTSL sudah sesuai dengan format yang diinginkan. Operator *Nominal to Numerical* dapat dilihat pada Gambar 5 berikut :



**Gambar 5.** Operator *Nominal to Numerical*

Pada operator *Nominal to Numerical* hanya mengubah pada parameter *Coding Type* menjadi *Unique Integers* untuk mengonversi data kategorikal (nominal) menjadi nilai numerik yang unik. Hal ini berguna dalam analisis data

agar dapat bekerja secara optimal. Berikut pada Tabel 5 ini menampilkan hasil awal dari atribut data yang masih dalam bentuk kategorikal saja dan belum diubah menjadi atribut numerik :

**Tabel 5.** Atribut Kategorikal

NIK	Letak Tanah	No. SPT PBB	Penggunaan Tanah
NIK. 3208131	DSN	32.10.160. 002.013-	Rumah
5079400 04	PON 2	0043.0	Tinggal
NIK. 3208135	DSN	32.10.160. 002.013-	Rumah
4127300 08	PON 1	0004.0	Tinggal
NIK. 3172014	DSN	32.10.160. 002.013-	Rumah
4046900 04	PON 1	0044.0	Tinggal
NIK. 3208132	DSN	32.10.160. 002.013-	Kebun
7037200 01	PON 1	0002.0	
...	...	...	...
NIK. 3208135	DSN	32.10.160. 002.027-	Kebun
3018200 07	WAGE	0001.0	

Lalu Tabel 6 berikut menampilkan hasil yang semula berupa tipe atribut kategorikal, telah diubah menjadi atribut numerik yaitu NIK, Letak tanah, No. SPT PBB dan Penggunaan tanah :

**Tabel 6.** Hasil Operator *Nominal to Numerical*

NIK	Letak Tanah	No. SPT PBB	Penggunaan Tanah
0	0	0	0
1	1	1	0
2	1	2	0
3	1	3	0
4	1	4	1
5	2	5	0
...	...	...	...
625	49	691	1

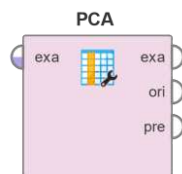
- Jumlah data pada kolom NIK dan No. SPT PBB diketahui tidak mencapai angka 775 disebabkan oleh adanya pemohon yang sama pada kolom NIK, di mana satu pemohon

dapat mengajukan lebih dari satu bidang tanah.

- Selain itu, kolom No. SPT PBB juga menunjukkan adanya data yang sama, karena satu nomor SPT PBB dapat digunakan untuk beberapa bidang tanah.

### 3.5. Data Mining

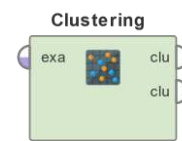
Sebelum tahap *Clustering* dilakukan, maka operator *Principal Component Analysis* (PCA) diperlukan untuk optimasi nilai *Davies Bouldin Index* (DBI) pada model pengelompokan data agar menjadi lebih optimal. Operator PCA dapat dilihat Gambar 6 berikut :



**Gambar 6.** Operator PCA

Penggunaan *Principal Component Analysis* (PCA) dalam *Clustering* berpengaruh karena dapat mereduksi dimensi data tanpa kehilangan informasi penting, sehingga menghindari masalah *Curse of Dimensionality* yang memiliki arti yaitu menurunnya kinerja algoritma sehingga sulit untuk menemukan pola yang bermakna. PCA juga menghilangkan redundansi dengan mengubah variabel asli menjadi komponen utama yang saling tidak berkorelasi, sehingga algoritma *K-Means* dapat bekerja lebih optimal. Dengan menyaring variabel yang paling berpengaruh, hasil pengelompokan menjadi lebih akurat, stabil, dan lebih mudah divisualisasikan.

Parameter dalam PCA dengan *Dimensionality Reduction* yang digunakan yaitu *Fixed Number* dengan *Number of Components* = 1. Setelah operator PCA diterapkan, maka tahap berikutnya memasukkan operator *Clustering* yang berfungsi untuk mengelompokkan data. Penelitian ini menggunakan algoritma *K-Means* dengan menggunakan *Cluster* yang ingin dibentuk yaitu  $K=5$ . Operator *Clustering (K-Means)* dapat dilihat pada Gambar 7 berikut :



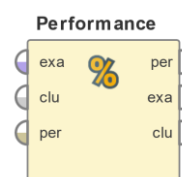
**Gambar 7.** Operator *Clustering (K-Means)*

Parameter pada operator *Clustering (K-Means)* dimana diketahui  $K=5$ , *Max run*=10 kali putaran, *Type Measure*=*Numerical Measures* dengan *Euclidean Distance* dan *Max Optimization Steps* = 100. Hasil operator *Clustering (K-Means)* yang dikombinasikan dengan PCA menghasilkan *Cluster Model*, dengan total 5 *Cluster* yang terbentuk, seperti yang ditampilkan pada Tabel 7 berikut :

**Tabel 7.** *Cluster Model*

No.	<i>Cluster Model</i>	
1.	<i>Cluster 0</i>	518 Items
2.	<i>Cluster 1</i>	108 Items
3.	<i>Cluster 2</i>	5 Items
4.	<i>Cluster 3</i>	124 Items
5.	<i>Cluster 4</i>	20 Items

Setelah data dikelompokkan menggunakan algoritma *K-Means* yang dikombinasikan dengan PCA, langkah berikutnya adalah memvalidasi hasil klaster yang terbentuk. Salah satu metode yang biasa digunakan untuk ini adalah dengan mengukur performa jarak antar klaster (*Cluster Distance Performance*), yang membantu mengevaluasi seberapa baik klaster yang terbentuk. Pada tahap evaluasi ini, digunakan model *Performance* untuk menghitung nilai *Davies Bouldin Index* (DBI) pada setiap proses yang dilakukan secara berulang, agar dapat menentukan nilai  $K$  yang paling optimal dalam proses klasterisasi dengan melihat hasil nilai DBI yang terendah dalam arti angka yang mendekati 0, menunjukkan klaster yang lebih baik. Operator *Performance* dapat dilihat pada Gambar 8 berikut :



**Gambar 8.** Operator *Performance*

Pada parameter *Cluster Distance Performance*, hanya mengubah *Main Criterion* menjadi *Davies Bouldin* dan klik untuk memberikan tanda (✓) pada *Maximize*, agar hasil *Average* dan DBI yang diperoleh tidak bernilai negatif.

Dari penelitian ini, nilai *Davies Bouldin Index* yang diperoleh paling mendekati 0 adalah 0.049. Nilai ini dicapai dengan menggunakan pendekatan *Fixed Number* dengan jumlah *Number of Components* = 1, dan  $K = 5$ . Berikut hasil *Operator Performance* dapat dilihat pada Tabel 8 berikut :

**Tabel 8.** Hasil *Operator Performance*

<i>Performance Vector</i>	
<i>Avg. within centroid distance</i>	161124432.675
<i>Avg. within centroid distance_cluster_0</i>	32205665.625
<i>Avg. within centroid distance_cluster_1</i>	1001248986.434
<i>Avg. within centroid distance_cluster_2</i>	64425.800
<i>Avg. within centroid distance_cluster_3</i>	395654.261
<i>Avg. within centroid distance_cluster_4</i>	231336.867
<i>Davies Bouldin</i>	0.049

- *Avg. Within Centroid Distance* adalah rata-rata jarak semua data dalam seluruh *Cluster* ke pusatnya masing-masing.
- *Avg. Within Centroid Distance\_Cluster* adalah rata-rata jarak semua data dalam satu *Cluster* tertentu ke pusatnya. Nilai ini menunjukkan seberapa rapat atau tersebar data dalam *Cluster*.

### 3.6. Interpretasi/Evaluasi

Pada tahap ini sebanyak 6 kali percobaan dilakukan dengan nilai  $K$  yang berbeda untuk menentukan nilai yang paling optimal. Nilai  $K$  yang optimal adalah yang menghasilkan nilai DBI mendekati 0. Dari rangkaian percobaan ini,  $K = 5$ . Dengan jumlah *Number of Component* = 1, yaitu mempertahankan 1 komponen utama yang memberikan hasil terbaik dengan nilai DBI mencapai 0.049 dibanding tanpa menggunakan PCA menghasilkan nilai DBI lebih besar yaitu 0.466. Hasil percobaan ini dapat ditampilkan

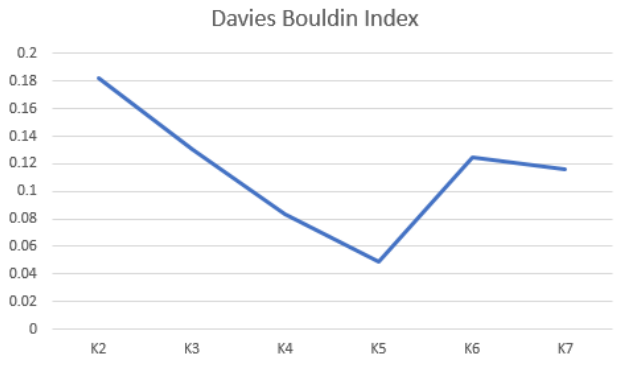
pada Tabel 9 berikut dengan pendekatan *Fixed Number* yang menampilkan nilai DBI untuk setiap percobaan dengan PCA dan tanpa penggunaan PCA :

**Tabel 9.** Hasil DBI (PCA dan Tanpa PCA)

<i>Fixed Number (Number of Components)</i>	Nilai $K$	Hasil DBI + PCA	Nilai DBI Tanpa PCA
1	2	0.182	0.225
	3	0.130	0.253
	4	0.083	0.545
	5	0.049	0.466
	6	0.125	0.949
	7	0.116	0.621
	2	2	0.214
3		0.215	0.253
4		0.170	0.545
5		0.308	0.466
6		0.343	0.949
	7	0.357	0.621

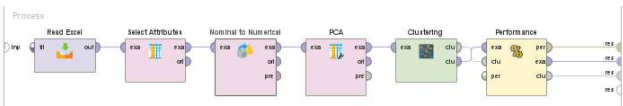
PCA secara signifikan meningkatkan kinerja pengelompokan karena mengurangi dimensi data dengan tetap mempertahankan informasi yang paling relevan, sehingga algoritma *K-Means* dapat bekerja lebih efisien. Dengan menghilangkan variabel yang memiliki korelasi tinggi, PCA membantu membentuk kluster yang lebih terpisah dan stabil sehingga perhitungan jarak antar titik dalam *K-Means* menjadi lebih akurat. Hal ini terbukti dalam penelitian ini, di mana penggunaan PCA menurunkan nilai *Davies Bouldin Index* (DBI) dari 0.466 menjadi 0.049, menunjukkan bahwa kluster yang terbentuk lebih optimal.

Gambar 9 berikut menampilkan hasil dari metode *Elbow* yang digunakan untuk menentukan jumlah kluster optimal dengan melihat perubahan nilai *Davies Bouldin Index* (DBI). Pada grafik *Elbow*, titik optimal ditandai dengan perubahan signifikan dalam kemiringan kurva, yang menunjukkan bahwa menambah jumlah kluster setelah titik ini tidak lagi memberikan perbaikan yang signifikan dan membuktikan bahwa  $K=5$  merupakan hasil yang optimal :



**Gambar 9. Metode Elbow**

Berikut pada Gambar 10 ini menunjukkan seluruh rangkaian langkah dalam model proses yang dijalankan di AI Studio 2024.0.1 mulai dari pemrosesan data awal hingga analisis akhir, untuk memastikan hasil yang optimal dalam pengelompokan data :



**Gambar 10. Model Seluruh Rangkaian di AI Studio 2024.0.1**

Pada Tahap ini mengidentifikasi dari total 5 kluster yang terbentuk, manakah *Cluster* yang terbaik dengan menghitung selisih setiap rata-rata yang dihasilkan dari *Cluster 0* sampai dengan *Cluster 4*, dapat diketahui hasil yang mendekati angka 0 disebut sebagai *Cluster* terbaik. Dibuktikan pada Tabel 10 berikut :

**Tabel 10. Cluster Terbaik**

K = 5	C0	C1	C2	C3	C4
Avg. within centroid distance_cluster	32205	10012	6442	39565	23133
Avg. within centroid distance	665.	48986	5.800	4.261	6.867
Selisih	12891	84012	16106	16072	16089
	876	455	000	877	309
	7.1	3.8	6.9	8.4	5.8

Pemilihan *Cluster 0* sebagai yang terbaik murni berdasarkan nilai selisih dari *Avg. within centroid distance\_cluster* dan *Avg. within centroid*

*distance*, *Cluster 0* ditetapkan sebagai yang terbaik karena memiliki jarak terdekat dengan komponen utama (PC1) dengan nilai 128918767.1, menunjukkan struktur yang kompak dan menunjukkan bahwa data dalam *Cluster* tersebut mencerminkan fitur atau karakteristik utama yang memiliki kemiripan yang lebih tinggi dari data keseluruhan dengan total 518 items.

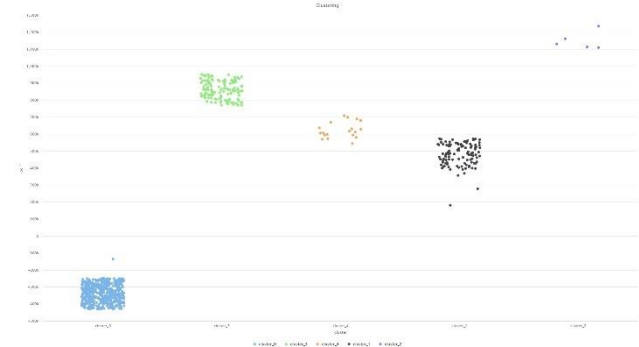
Gambar 11 berikut ini merupakan hasil *Scatter Plot* saat penggunaan algoritma *K-Means* saja dengan penggunaan  $K=5$  :



**Gambar 11. Hasil Scatter Plot K-Means**

- Pada scatter plot tanpa PCA, kluster tampak lebih menyebar dan beberapa titik data masih berdekatan satu sama lain, menunjukkan kemungkinan adanya tumpang tindih antar kluster.
- Tanpa PCA, semua informasi digunakan, yang bisa membuat pemisahan kelompok menjadi kurang optimal jika ada fitur yang memiliki korelasi tinggi.

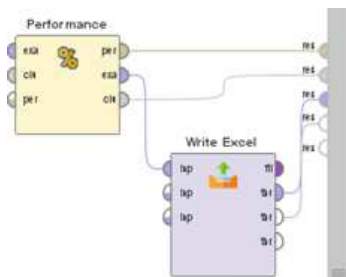
Gambar 12 berikut ini merupakan hasil *Scatter Plot* antara penggunaan algoritma *K-Means* dan *Principal Component Analysis (PCA)* dengan  $K=5$ , dan *Number of Components = 1* :



**Gambar 12.** Hasil *Scatter Plot K-Means* Dengan PCA

- PCA membantu menyaring informasi yang paling penting dari data, sehingga pola kelompok lebih mudah dikenali.
- Gambar tersebut membuktikan bahwa *Cluster 0* memiliki tingkat kemiripan karakteristik data yang lebih tinggi. Secara keseluruhan, *Scatter Plot* ini menunjukkan perbedaan yang jelas di antara setiap *Cluster*, yang mengindikasikan adanya variasi karakteristik antar kelompok dan dapat membantu dalam analisis data lebih lanjut.

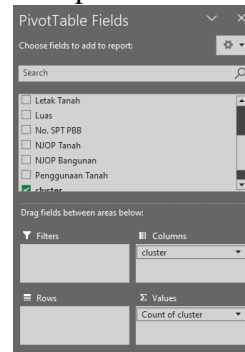
Lalu Pada tahap akhir adalah proses untuk menentukan satu komponen utama dalam atribut yang dipertahankan dengan menggunakan operator “*Write Excel*” untuk mengekspor jumlah *Cluster* yang dihasilkan dalam data. Penggunaan operator *Write Excel* dapat pada Gambar 13 berikut :



**Gambar 13.** Operator *Write Excel*

Pada parameter *Write Excel*, hanya perlu menentukan lokasi penyimpanan hasil ekspor proses dalam bentuk file berformat .xlsx. Data hasil *Cluster* nantinya akan disalin dan digabungkan dengan data *Excel* dari program PTSL. Dalam karakteristik data, terdapat satu komponen utama yang dipertahankan menggunakan operator PCA dengan pendekatan *Fixed Number*. Untuk menentukan atribut yang dipertahankan, dapat menggunakan *Pivot Table* yang tersedia pada menu *Insert* di *Microsoft Excel*. Dengan memilih semua data atribut yang ada, kemudian masukkan atribut ke dalam *Pivot Table Fields*. Atur *Columns* dan *Values* yang dapat diisi menggunakan atribut *Cluster*, sementara *Rows* diisi dengan melakukan *Drag*

pada atribut yang terdapat pada data. *Pivot Table Fields* dapat dilihat pada Gambar 14 berikut :



**Gambar 14.** *Pivot Table Fields*

*Rows* yang diisi dengan atribut data luas yang tertuju pada luas tanah, menunjukkan variasi data yang cukup beragam, dengan luas yang tersebar di berbagai *Cluster* dalam arti tidak terkonsentrasi hanya pada satu *Cluster*. Oleh karena itu, atribut luas tidak dapat dianggap sebagai satu komponen utama yang dicari. Hasilnya ditampilkan pada Tabel 11 berikut :

**Tabel 11.** *Rows* Atribut Luas

Luas Tanah	C0	C1	C2	C3	C4
31	1	-	-	-	-
32	1	-	-	-	-
33	3	-	-	-	-
34	-	-	-	1	-
50	1	1	-	-	-
...	...	...	...	...	...
13900	1	-	-	-	-

*Rows* yang diisi dengan atribut data letak tanah menunjukkan bahwa terdapat beberapa blok dan dusun tertentu yang tersebar di berbagai *Cluster* dalam arti tidak terpusat pada satu *Cluster* saja. Hal ini menunjukkan bahwa atribut letak tanah bukanlah satu komponen utama yang dicari. Hasilnya ditampilkan pada Tabel 12 berikut :

**Tabel 12.** *Rows* Atribut Letak Tanah

Letak Tanah	C0	C1	C2	C3	C4
BL					
CIGOWEK 1	1	-	-	-	-
BL					
CIKENDAN	-	-	-	1	-

BL					
CIPENDEUY	1	-	-	-	-
BL	-	1	-	-	-
LAME					
DSN	1	1	-	-	-
CIKONDANG					
...	...	...	...	...	...
PGR	11	1	-	-	-
CIHAMBAR					

Rows yang diisi dengan atribut data penggunaan tanah menunjukkan bahwa jenis-jenis penggunaan tanah tersebar di berbagai Cluster dalam arti tidak terpusat pada satu Cluster tertentu. Dengan demikian, atribut penggunaan tanah tidak dapat dianggap sebagai satu komponen utama yang dicari. Hasilnya ditampilkan pada Tabel 13 berikut :

**Tabel 13.** Rows Atribut Penggunaan Tanah

Penggunaan Tanah	C0	C1	C2	C3	C4
Bangunan SMA					
Pertiwi	1	-	-	-	-
Kantor Desa	1	-	-	-	-
Kebun	160	7	-	6	2
Kolam	2	2	-	-	-
...	...	...	...	...	...
Taman					
Bandorasawetan	1	-	-	-	-

Rows yang diisi dengan atribut data NJOP tanah menunjukkan bahwa nilai jual objek pajak tanah tersebar di berbagai Cluster dalam arti tidak terpusat pada satu Cluster tertentu. Oleh karena itu, atribut NJOP tanah tidak dapat dianggap sebagai satu komponen utama yang dicari. Hasilnya ditampilkan pada Tabel 14 berikut :

**Tabel 14.** Rows Atribut NJOP Tanah

NJOP Tanah	C0	C1	C2	C3	C4
0	9	-	-	-	-
14000	1	-	-	-	-
27000	2	-	-	-	2
36000	8	1	-	-	3
48000	4	2	-	-	-
...	...	...	...	...	...
160000	16	11	-	4	3

Selanjutnya, Rows yang diisi dengan atribut data NJOP bangunan menunjukkan bahwa

nilai jual objek pajak bangunan memiliki pola yang lebih konsisten atau homogen di antara data, tidak tersebar, dan cenderung terpusat dalam satu Cluster. Oleh karena itu, atribut NJOP bangunan dapat dianggap sebagai satu komponen utama yang dicari. Hasilnya ditampilkan pada Tabel 15 berikut :

**Tabel 15.** Rows Atribut NJOP Bangunan

NJOP Bangunan	C0	C1	C2	C3	C4
0	517	-	-	-	-
128000	1	-	-	-	-
595000	-	1	-	-	-
700000	-	4	-	-	-
823000	-	103	-	-	-
968000	-	-	-	-	20
1200000	-	-	-	124	-
1516000	-	-	5	-	-

#### 4. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa kombinasi algoritma *K-Means* dan *Principal Component Analysis* (PCA) dapat mengoptimasi nilai *Davies Bouldin Index* (DBI) sebesar 0.049 pada model pengelompokan data program Pendaftaran Tanah Sistematis Lengkap (PTSL), dengan nilai kluster optimal yaitu  $K = 5$  dengan menggunakan *Number of Components* = 1, yang mempertahankan atribut NJOP Bangunan sebagai komponen utama. Cluster terbaik ditetapkan pada Cluster 0 yang memiliki kedekatan dengan komponen utama (PC1) dengan nilai sebesar 128918767.1. Jadi dapat disimpulkan bahwa karakteristik data dari Cluster 0 yang memiliki nilai NJOP Bangunan sebesar 0 dan 128.000, Cluster 1 yang memiliki nilai NJOP Bangunan sebesar 595.000, 700.000 dan 823.000, Cluster 2 yang memiliki nilai NJOP Bangunan sebesar 1.516.000, Cluster 3 yang memiliki nilai NJOP Bangunan sebesar 1.200.000, dan Cluster 4 yang memiliki nilai NJOP Bangunan sebesar 968.000. Untuk penelitian selanjutnya disarankan mengeksplorasi metode *Clustering* lain seperti *DBSCAN* atau algoritma berbasis hierarki untuk membandingkan hasil yang diperoleh dan memberikan alternatif pendekatan yang lebih optimal.

## 5. DAFTAR PUSTAKA

- [1] I. K. Ayu, “Kepastian Hukum Penguasaan Tanah Melalui Pendaftaran Tanah Sistematis Lengkap di Kota Batu,” *Mimb. Huk. - Fak. Huk. Univ. Gadjah Mada*, vol. 31, no. 3, pp. 338–351, Jan. 2020, doi: 10.22146/jmh.41560.
- [2] R. B. Kumalasari, “Legal Protection for Village Communities Against Land Rights in the Construction of a National Strategy Project Through a Complete Systematic Land Registration System (Study of Pasuruan District Land Registration),” *Int. J. Multicult. Multireligious Underst.*, vol. 8, no. 7, p. 48, 2021, doi: 10.18415/ijmmu.v8i7.2879.
- [3] M. M. Ariwibowo, Suharno, and Wahyuni, “Efektivitas Pemanfaatan Aplikasi Locus Gis dan Mapit Gis Untuk Pengumpulan Data Pendaftaran Tanah,” *J. Tunas Agrar.*, vol. 3, no. 1, pp. 116–144, 2020.
- [4] Y. Arie Wijaya, D. Achmad Kurniady, E. Setyanto, W. Sanur Tarihoran, D. Rusmana, and R. Rahim, “Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities,” *TEM J.*, vol. 10, no. 3, pp. 1099–1103, Aug. 2021, doi: 10.18421/TEM103-13.
- [5] N. A. Khairani and E. Sutoyo, “Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province,” *Int. J. Adv. Data Inf. Syst.*, vol. 1, no. 1, pp. 9–16, Apr. 2020, doi: 10.25008/ijadis.v1i1.13.
- [6] S. Dewi and M. A. I. Pakereng, “IMPLEMENTASI PRINCIPAL COMPONENT ANALYSIS PADA K-MEANS UNTUK KLASTERISASI TINGKAT PENDIDIKAN PENDUDUK KABUPATEN SEMARANG,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 4, pp. 1186–1195, Nov. 2023, doi: 10.29100/jupi.v8i4.4101.
- [7] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means Algorithm: A Comprehensive Survey and Performance Evaluation,” *Electronics*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.
- [8] D. Hedyati and I. M. Suartana, “Penerapan Principal Component Analysis ( PCA ) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi,” *JIEET (Journal Inf. Eng. Educ. Technol.)*, vol. 5, no. 2, pp. 49–54, 2021.
- [9] N. Rahmawati, “Pendaftaran Tanah Berbasis Desa Lengkap,” *Tunas Agrar.*, vol. 5, no. 2, pp. 127–141, Apr. 2022, doi: 10.31292/jta.v5i2.177.
- [10] J. W. P. Putra, E. A. Suganda, and I. Purnamasari, “Penerapan RapidMiner dengan Metode K-Means dalam Penentuan Kluster Gangguan Jaringan WIFI Provider PT.XYZ di Daerah Karawang,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 31–35, 2022.
- [11] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, Dec. 2020, doi: 10.25105/jti.v10i3.8428.
- [12] J. Nasir, “Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means,” *J. SIMETRIS*, vol. 11, no. 2, pp. 1–13, 2020.
- [13] A. Siregar, A. Buono, and K. Priandana, “Perbandingan Algoritma K-Means dan Fuzzy C-Means untuk Clustering Citra Daun Melon,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1503–1510, Dec. 2022, doi: 10.47065/bits.v4i3.2534.
- [14] M. Y. Balaka, *Metode penelitian Kuantitatif*, vol. 1. 2022.