

Implementasi Algoritma K-Means Clustering Dalam Mengelompokkan Kepatuhan Wajib Pajak Bumi dan Bangunan Dengan Optimasi Elbow

Amri Muliawan Nur^{1*}, Hariman Bahtiar², Mila Agustiarini Jannah³

^{1,2,3}Program Studi Sistem Informasi, Universitas Hamzanwadi

*muliaamriga@gmail.com

Abstrak

Pajak Bumi dan Bangunan (PBB) merupakan salah satu sumber pendapatan asli daerah yang memiliki peran penting dalam mendukung pembangunan di berbagai daerah. Oleh karena itu tingkat kepatuhan pajak perlu ditingkatkan dengan berbagai strategi seperti sosialisasi maupun edukasi secara berkelanjutan untuk meningkatkan kesadaran atas pentingnya membayar pajak, selain itu pula peningkatan pelayanan yang lebih baik perlu dilakukan. Tujuan penelitian ini mengelompokkan tingkat kepatuhan wajib pajak PBB di Kecamatan Sakra menggunakan algoritma K-Means Clustering. Data yang digunakan dalam penelitian ini adalah data PBB Kecamatan Sakra tahun 2023, dengan jumlah dataset sebanyak 376 dengan melibatkan lima atribut utama, yaitu: luas bumi, luas bangunan, ketetapan PBB, status pembayaran, dan denda. Hasil yang diperoleh dari pengolahan menggunakan algoritma K-Means, memperoleh jumlah cluster yang sangat baik. Berikut adalah hasil cluster yang diperoleh : Cluster 1 tingkat kepatuhan tinggi, berjumlah 355 item. Cluster 2 tingkat kepatuhan sedang, berjumlah 18 item dan cluster 3 tingkat kepatuhan rendah, berjumlah 3 item. Hasil pengelompokan ini dapat menjadi acuan bagi pihak berwenang untuk merumuskan strategi peningkatan kepatuhan yang lebih terarah melalui edukasi dan pelayanan yang lebih baik kedepannya.

Kata kunci : Clustering , Data Mining, K-Means, Pajak Bumi dan Bangunan (PBB)

Abstract

Land and Building Tax (Pajak Bumi dan Bangunan, or PBB) is one of the primary sources of regional revenue that plays a significant role in supporting development across various regions. Therefore, efforts to improve tax compliance must be enhanced through various strategies, such as continuous socialization and education, to raise awareness of the importance of paying taxes. Additionally, improving the quality of services is essential. This study aims to classify the compliance levels of PBB taxpayers in Sakra District using the K-Means Clustering algorithm. The data used in this research is the 2023 PBB dataset for Sakra District, comprising 376 entries and involving five key attributes: land area, building area, PBB assessment, payment status, and penalties. The results obtained from processing using the K-Means algorithm indicate an optimal number of clusters, as follows: Cluster 1 represents a high compliance level, consisting of 355 items; Cluster 2 represents a moderate compliance level, consisting of 18 items; and Cluster 3 represents a low compliance level, consisting of 3 items. These clustering outcomes can serve as a reference for authorities in formulating more targeted strategies to enhance tax compliance through improved education and services in the future.

Keywords : Clustering , Data Mining, K-Means, Land and Building Tax (PBB).

1. Pendahuluan

Sebuah negara dikatakan maju jika mampu membangun wilayah dan masyarakatnya, baik secara fisik maupun non fisik. Pembangunan fisik berupa infrastruktur seperti jalan, jembatan,

gedung sekolah, dan lain sebagainya, sedangkan pembangunan non fisik berupa peningkatan layanan kesejahteraan masyarakat. Indonesia merupakan Negara yang kaya akan sumber daya alamnya. Oleh karena itu, dalam melaksanakan

pembangunan, sumber daya tersebut harus digunakan secara rasional dan efektif. Salah satu sektor penting dalam pembangunan adalah sektor perpajakan, yang menjadi sumber pendapatan bagi negara untuk membiayai berbagai layanan masyarakat dan proyek pembangunan.

Pajak Bumi dan Bangunan (PBB) adalah salah satu jenis pajak yang menjadi sumber pendapatan potensial dan berperan penting dalam mendukung pendapatan daerah^[1]. Dalam pembiayaan negara, pajak dianggap sebagai sumber dana yang paling potensial, akan tetapi dalam merealisasikannya negara masih kesulitan dalam pemungutan pajak. Hal ini disebabkan karena rendahnya tingkat kesadaran masyarakat dalam membayar pajak dan juga rendahnya kepercayaan masyarakat kepada pengelola administrasi pajak^[2].

Kecamatan Sakra merupakan salah satu kecamatan di Kabupaten Lombok Timur, Nusa Tenggara Barat yang memiliki potensi pajak yang cukup besar dan diketahui memiliki tingkat kepatuhan wajib pajak yang tinggi. Tingginya tingkat kepatuhan wajib pajak di Kecamatan Sakra merupakan hal yang menarik untuk diteliti. Hal ini karena beberapa daerah lain di Indonesia masih memiliki tingkat kepatuhan wajib pajak yang rendah. Oleh karena itu, perlu dilakukan analisa untuk mengelompokkan wajib pajak di Kecamatan Sakra Kabupaten Lombok Timur

Nusa Tenggara Barat berdasarkan tingkat kepatuhannya. Untuk mendukung analisa ini, perlu adanya analisa yang mendalam. Pendekatan data *mining* dapat digunakan untuk menganalisa data pajak di kecamatan Sakra.

Salah satu algoritma data *mining* yang dapat digunakan untuk mengelompokkan wajib pajak adalah *Algoritma K-Means Clustering*. Metode ini dapat digunakan untuk mengelompokkan data pajak berdasarkan kedekatan antar data yang memiliki ciri identik yang sama. Dengan memanfaatkan metode *K-Means Clustering*, setiap wajib pajak akan dibagi dalam 3 *Cluster*, yaitu tingkat kepatuhan membayar pajak tinggi, tingkat kepatuhan membayar pajak sedang, dan tingkat kepatuhan membayar pajak rendah. Dengan menggunakan metode *Clustering*, diharapkan dapat diperoleh informasi mengenai karakteristik wajib pajak di Kecamatan Sakra.

Dari uraian yang penulis paparkan di atas, maka penulis tertarik mengambil judul "Implementasi *Algoritma K-Means Clustering* Dalam Mengelompokkan Tingkat Kepatuhan Wajib Pajak Bumi dan Bangunan di Wilayah Sakra Kabupaten Lombok Timur".

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Untuk menunjang pembuatan artikel ini, penulis mengacu pada beberapa penelitian sebelumnya.

- Penelitian yang dilakukan oleh Amri Muliawan Nur, dkk. Yang berjudul "Penerapan *Algoritma K-Means Clustering* Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi". Berdasarkan hasil analisa yang dilakukan dengan menggunakan 350 data maka dapat ditarik kesimpulan bahwa metode *K-Means Clustering* dapat mengelompokkan data dengan baik ditandai dengan hasil cluster yang diperoleh yaitu sebesar 225 item *smartphone*, dengan kategori kelompok sangat direkomendasikan, 98 item *smartphone* direkomendasikan, dan 27 item *smartphone* kelompok kurang direkomendasikan. dengan. Pengelompokan ini diharapkan dapat membantu penjual untuk lebih mudah meningkatkan stok *smartphone* yang direkomendasikan sesuai dengan kebutuhan konsumen dalam hal harga dan spesifikasi^[3].
- Penelitian yang dilakukan oleh Suhartini dan Ria Yuliani dengan judul "Penerapan Data Mining untuk Meng-*Cluster* Data Penduduk Miskin Menggunakan *Algoritma K-Means* di Dusun Bagik Endep Sukamulia Timur" menunjukkan bahwa penentuan *Centroid* (titik pusat) pada tahap awal algoritma *K-Means* memiliki pengaruh signifikan terhadap hasil clustering. Dalam penelitian ini, pengujian dilakukan menggunakan 200 dataset. Berdasarkan hasil pengujian dengan menggunakan *RapidMiner* dan membentuk 3 cluster, ditemukan bahwa jumlah penduduk terbanyak berada pada Cluster ke-3, yang tergolong sebagai penduduk dengan tingkat ekonomi rendah (miskin)^[4].
- Penelitian yang dilakukan oleh Tita Puspita Sari, April Lia Hananto, Elfina Novalia, Tukino, dan Shofa Shofia Hilabi dengan judul "Implementasi *Algoritma K-Means* dalam Analisis Klasterisasi Penyebaran Penyakit HIV/AIDS" menunjukkan hasil analisis sebagai berikut: *Cluster 1* dengan kriteria *sangat tinggi* terdiri dari 1 data dan memperoleh hasil sebesar 3%. *Cluster 2* dengan kriteria *sedang* terdiri dari 2 data dan memperoleh hasil sebesar 7%. *Cluster 3* dengan kriteria *tinggi* terdiri dari 5 data dan memperoleh hasil sebesar 17%. *Cluster 0* dengan kriteria *rendah* terdiri dari 22 data dan memperoleh hasil sebesar 73%. Hasil ini menggambarkan sebaran klaster berdasarkan tingkat penyebaran penyakit HIV/AIDS dengan menggunakan metode algoritma *K-Means*^[5].
- Penelitian yang dilakukan oleh Andrian Hidayat, Nurhidayati, dan Amri Muliawan Nur dengan judul "Implementasi *Algoritma K-Means* Untuk Klasterisasi Peserta Keluarga Berencana Berdasarkan Tingkat Risiko Kehamilan Di Desa Pringgasela Selatan".

Berdasarkan hasil penelitian, dapat ditarik kesimpulan yaitu Tingkat risiko kehamilan di Desa Pringgasela Selatan membentuk dua cluster risiko, yaitu: *Cluster* 1 (Risiko Tinggi): Berjumlah 170 orang. *Cluster* 0 (Risiko Rendah): Berjumlah 230 orang. Hasil ini menggambarkan distribusi tingkat risiko kehamilan di kalangan peserta program Keluarga Berencana berdasarkan analisis klasterisasi menggunakan metode K-Means^[6].

- Penelitian yang dilakukan oleh Nurhidayati, Lola Mauliya, Suhartini berjudul “*Clustering Data Pasien Covid Berdasarkan Usia dan Gejala Menggunakan Algoritma K-Means*”. Dari penerapan data *mining* menggunakan metode *Clustering* dengan *Algoritma K-Means* didapatkan dua kluster yaitu *Cluster* usia dan gejala, yang merupakan tolak ukur pemerintah dalam memutus rantai penyebaran virus Covid-19 tahun 2020. Dari hasil pengolahan tersebut, didapatkan hasil pada *Cluster* 1 (Berdasarkan usia) berjumlah 866 items yang terdiri dari pasien dengan rentang usia 42–81 tahun, dengan jumlah kasus kematian sebanyak 50 orang. Sedangkan pada kluster 2 (Berdasarkan gejala) berjumlah 1566 items dengan rentang usia 1-41 tahun dengan kasus kematian 16 orang^[7].

2.2. Landasan Teori

1. Data Mining

Data *mining* adalah sebuah metode penambangan data yang menemukan pola pengetahuan dari data yang informasinya masih tersembunyi^{[8][9]}.

Proses data mining terdiri dari serangkaian tahapan yang dilakukan secara sistematis untuk menghasilkan pengetahuan baru yang bermanfaat. Berikut adalah penjelasan dari masing-masing tahap secara lebih rinci :

a. Seleksi Data

Pada tahap ini, data dipilih dari berbagai sumber dan ditentukan mana yang akan digunakan sebagai data target. Proses ini bertujuan memastikan hanya data yang relevan dan penting yang akan dianalisis.

b. Preprocessing Data

Tahap ini dilakukan untuk meningkatkan kualitas data sebelum dianalisis lebih lanjut. Proses ini meliputi pembersihan data, mengisi data yang kosong, menghapus inkonsistensi, dan memastikan data memiliki format yang sesuai agar dapat diproses dengan baik.

c. Transformasi Data

Data yang sudah melalui tahap preprocessing kemudian diubah ke bentuk yang lebih cocok untuk analisis. Ini bisa meliputi proses normalisasi, penggabungan data, pemilihan fitur yang relevan, atau metode lain agar analisis lebih efektif dan akurat.

d. Data Mining

Ini adalah tahap utama dalam data mining, di mana berbagai metode dan algoritma diterapkan untuk menemukan pola, hubungan, atau informasi yang tersembunyi dalam data. Teknik yang digunakan meliputi klasifikasi, clustering, regresi, atau analisis asosiasi untuk menggali wawasan yang dapat digunakan.

e. Interpretasi dan Evaluasi

Setelah pola atau informasi ditemukan melalui proses data mining, tahap ini dilakukan untuk mengevaluasi dan memahami hasil tersebut. Interpretasi bertujuan memahami makna dari pola yang ditemukan, sementara evaluasi memastikan hasil yang diperoleh akurat dan relevan sebelum diterapkan^[10].

2. Algoritma *K-Means*

K-Means adalah *algoritma* analisis kluster non-hierarkis yang digunakan untuk mengelompokkan data berdasarkan kedekatan antar data. Analisis kluster bertujuan mengelompokkan data berdasarkan karakteristik atau variabel tertentu^[11]. Sama halnya seperti metode pengelompokan data yang lain, tujuan *K-Means* adalah memaksimalkan kesamaan fitur dalam satu kelompok sambil meminimalkan perbedaan antar kelompok^[12].

3. Metode Clustering

Clustering adalah metode untuk membagi satu kumpulan data menjadi beberapa kelompok atau himpunan yang disebut sebagai *Cluster*^[13]. Anggota dalam setiap *Cluster* memiliki kesamaan karakteristik satu sama lain, tetapi memiliki perbedaan karakteristik dibandingkan dengan anggota dari *Cluster* lainnya^[14].

4. Pajak Bumi Dan Bangunan (PBB)

Pajak bumi dan bangunan (PBB) merupakan pajak pemerintah pusat dan digolongkan sebagai pajak langsung serta dipungut setiap tahun. Walaupun PBB merupakan pajak pusat tetapi dalam pengelolaan dilaksanakan oleh Direktorat Jenderal Pajak melalui Kantor Pelayanan Pajak Bumi dan Bangunan dan hasilnya dibagi dua yaitu 10% untuk pemerintah pusat dan 90% untuk pemerintah daerah ^{[15][16]}

3. Metode Penelitian

3.1 Metode Pengumpulan Data

1. Observasi

Teknik observasi yang digunakan untuk pengumpulan data melalui pengamatan langsung ke Kantor Pajak Bumi Dan Bangunan Kecamatan Sakra. Kegiatan ini dilakukan sebagai bagian dari riset untuk mengumpulkan informasi yang dibutuhkan dalam menyelesaikan penelitian ini.

2. Wawancara

Pengumpulan data dilakukan melalui komunikasi dan wawancara langsung dengan aparat yang

bertanggung jawab dalam pengelolaan pajak bumi dan bangunan (PBB). Mereka memiliki peran aktif dalam administrasi perpajakan dan memahami perilaku wajib pajak bumi dan bangunan di Kecamatan Sakra

3. Studi Pustaka

Metode pengumpulan data dilakukan oleh peneliti dengan cara mempelajari buku-buku yang relevan, serta melakukan pencarian referensi melalui internet (browsing) untuk mendapatkan artikel atau jurnal yang berkaitan dengan penelitian ini

3.2 Tahapan Penelitian

Berikut adalah beberapa tahapan yang dilakukan pada penelitian ini :



Gambar 1. Tahapan Penelitian

1. Tahapan Pengumpulan data

Tahap ini dilakukan untuk mengumpulkan dan memperoleh data yang sebenarnya. data yang digunakan adalah data pembayaran pajak bumi dan bangunan (PBB) Kecamatan Sakra tahun 2023 dengan jumlah data 376.

2. Pengolahan Data

Pada tahap ini dilakukan prose pengolahan data data pajak menggunakan algoritma *K-Means Clustering*, dengan tujuan untuk mengelompokkan pengguna wajib pajak yang memiliki kelompok data yang berkarateristik

sama. Untuk mengetahui kelompok data yang karateristik sama dapat dihitung dengan pendekatan perhitungan jarak menggunakan *Euclidean Distance*.

3. Analisa Data

Tahapan Analisa data dilakukan dengan menggunakan *Google Colab*. Tujuan dari Analisa ini untuk memperoleh cluster data pajak dengan 3 kategori yaitu kepatuhan wajib pajaknya tinggi, kepatuhan wajib pajaknya sedang, dan kepatuhan wajib pajaknya rendah.

3.3 Lokasi Penelitian

Adapun lokasi yang dijadikan sebagai tempat penelitian yaitu di Kantor pajak bumi dan bangunan (PBB) Kecamatan Sakra Kabupaten Lombok Timur Nusa Tenggara Barat

4. Hasil dan Pembahasan

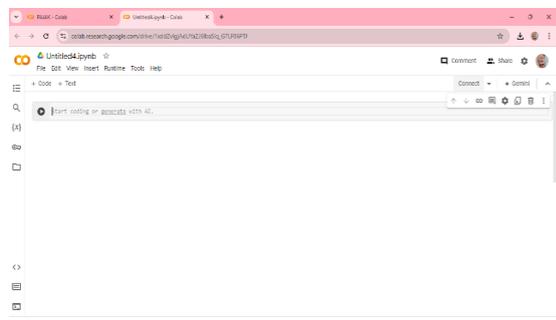
4.1 Pengolahan Data Pada *Google Colab*

Proses pengolahan data pajak di Kecamatan Sakra dengan melibatkan 376 dataset menggunakan *Google Colab* dilakukan dalam beberapa tahapan :

1. Menghubungkan *data* :

Tahap ini merupakan tahap awal pengolahan data. Sebelum memulai pengolahan *data*, pastikan terlebih dahulu bahwa *Google Colab* sudah terbuka dengan baik, serta pastikan *Google Colab* terhubung dengan jaringan internet agar dapat berfungsi dengan baik. Tahap ini dapat

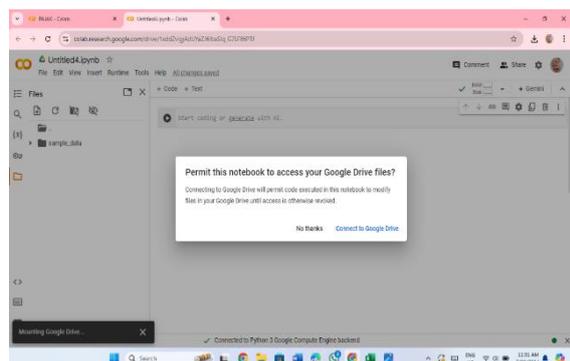
dilihat pada gambar di bawah



Gambar 2. Menghubungkan Data

2. Sinkronisasi Google Drive

Tahap ini merupakan Sinkronisasi *Google Drive* dengan *Google Colab*. Hal ini dilakukan untuk ekstraksi data di *Google Colab*.



Gambar 3. Sinkronisasi Google Colab

3. Memasukkan Data

Tahap ini merupakan tahap memasukkan data ke dalam *Google Colab*, Sebelum memasukkan *data*, terlebih dahulu ubah format *data Excel* ke dalam format CSV, hal ini dilakukan dengan tujuan agar data yang dimasukkan bisa terbaca dan ditampilkan saat dijalankan. Berikut ini adalah langkah-langkah untuk memasukkan *data* ke *Google Colab*:

- Mengimport Library yang Diperlukan

Tahap ini merupakan tahap mengimport *library* yang dibutuhkan untuk analisis data, *Clustering*, visualisasi, dan penyimpanan file, yang terdiri dari *import pandas*, *sklearn* dan *matplotlib*.

```
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
```

Gambar 4. Code Mengimpor Library

- Memuat Dataset

Tahap ini merupakan tahap memuat dataset dari file CSV dan menampilkan beberapa baris pertama untuk memastikan data dimuat dengan benar.

```
df = pd.read_csv('DATASET SKRIPSI.csv')
df.head(12)
```

	Id	Luas Bumi	Luas Bangunan	Ketetapan PBB	Sudah Bayar	Denda
0	1	4,078	126	26,718	26,718	0
1	2	237	36	15,000	15,000	0
2	3	3,550	48	15,000	0	0
3	4	7,128	117	15,000	0	0
4	5	940	88	15,000	0	0
5	6	225	225	591,413	591,413	0
6	7	544	54	15,000	0	0
7	8	988	346	174,610	174,610	0
8	9	2,028	168	15,000	15,000	0
9	10	3,996	75	15,000	15,000	0
10	11	7,438	75	15,000	15,000	0
11	12	2,283	42	15,000	0	0

Gambar 5. Code Memuat Dataset

4. Pra-Pemrosesan Data

Tahap selanjutnya adalah Pra-pemrosesan data (*data preprocessing*). Tahap ini merupakan tahap yang penting dalam analisis data dengan tujuan untuk mempersiapkan data mentah menjadi format yang lebih bersih dan lebih mudah dianalisis. Proses ini melibatkan beberapa tahapan untuk mengatasi masalah yang mungkin ada dalam data mentah, seperti data yang hilang,

format yang tidak konsisten, *redundansi* data atau *outlier*. Tahapan ini dapat dilihat pada gambar di bawah.

```
[7] df = df.drop(columns=['Id'])

[8] df['Luas Bumi'] = df['Luas Bumi'].str.replace(', ', '').astype(float)
df['Ketetapan PBB'] = df['Ketetapan PBB'].str.replace(', ', '').astype(float)
df['Sudah Bayar'] = df['Sudah Bayar'].str.replace(', ', '').astype(float)

df = df.dropna()
print(df.head(10))
```

	Luas Bumi	Luas Bangunan	Ketetapan PBB	Sudah Bayar	Denda
0	4078.0	126	26718.0	26718.0	0
1	237.0	36	15000.0	15000.0	0
2	3550.0	48	15000.0	0.0	0
3	7128.0	117	15000.0	0.0	0
4	940.0	88	15000.0	0.0	0
5	225.0	225	591413.0	591413.0	0
6	544.0	54	15000.0	0.0	0
7	989.0	346	174610.0	174610.0	0
8	2028.0	168	15000.0	15000.0	0
9	3995.0	75	15000.0	15000.0	0

Gambar 6. Code Preprocessing Data

5. Klusterisasi Awal

Tahap klusterisasi awal merupakan langkah pertama dalam proses klusterisasi *data*, di mana model klusterisasi diterapkan untuk pertama kalinya pada *dataset* untuk membentuk kluster *data*.

```
kmeans = KMeans(n_clusters=k, random_state=0)
kmeans.fit(df)
df['cluster'] = kmeans.labels_
print(df.head(10))
```

	Luas Bumi	Luas Bangunan	Ketetapan PBB	Sudah Bayar	Denda	Cluster
0	4078.0	126	26718.0	26718.0	0	0
1	237.0	36	15000.0	15000.0	0	0
2	3550.0	48	15000.0	0.0	0	0
3	7128.0	117	15000.0	0.0	0	0
4	940.0	88	15000.0	0.0	0	0
5	225.0	225	591413.0	591413.0	0	2
6	544.0	54	15000.0	0.0	0	0
7	989.0	346	174610.0	174610.0	0	0
8	2028.0	168	15000.0	15000.0	0	0
9	3995.0	75	15000.0	15000.0	0	0

Gambar 7. Code Melakukan Clustering Dengan K-Means

6. Analisis Elbow

Tahap Analisis *Elbow* adalah metode yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma klusterisasi, terutama pada metode K-Means. Metode ini berfungsi untuk mengidentifikasi titik di mana penambahan jumlah kluster tidak memberikan peningkatan

yang signifikan dalam menjelaskan variabilitas data.

```
k_list = range(1, 11)
sse = []
for k in k_list:
    kmeans = KMeans(n_clusters=k, random_state=0)
    sse.append(kmeans.inertia_)
```

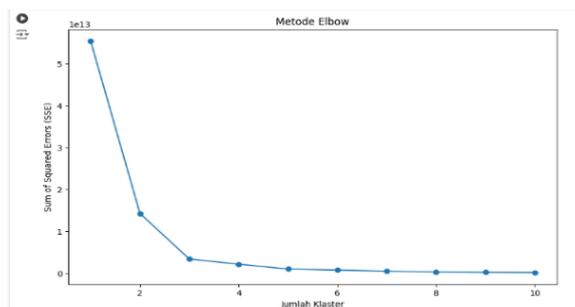
Gambar 8. Code Analisis Elbow

7. Plot SSE

Membuat plot yang menggambarkan hubungan antara nilai *SSE* (*Sum of Squared Errors*) dengan jumlah kluster (*k*) dilakukan untuk membantu mengidentifikasi titik "Elbow." Titik ini menunjukkan jumlah kluster yang optimal, di mana penurunan nilai *SSE* mulai melambat dan menandai batas optimal untuk jumlah kluster yang digunakan.

```
plt.figure(figsize=(10, 6))
plt.plot(k_list, sse, marker='o')
plt.xlabel('Jumlah Kluster')
plt.ylabel('Sum of Squared Errors (SSE)')
plt.title('Metode Elbow')
plt.show()
```

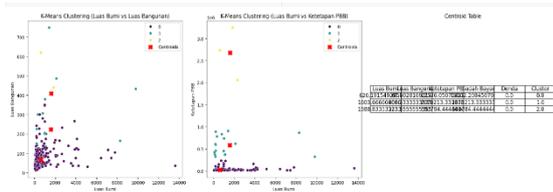
Gambar 9. Code Membuat Plot Dari Nilai SSE Terhadap Jumlah Kluster Untuk Menemukan Titik Elbow



Gambar 10. Titik Elbow

8. Visualisasi Titik Centroid

Tahap ini menggambarkan *data-data* yang dikelompokkan dalam kluster tertentu berdasarkan titik pusat/ centroid.



Gambar 11. Centroid

9. Cluster Model

Tahap ini menggambarkan *Cluster model* yang merupakan inti dari hasil proses *clustering*. Hasil ini masih dalam bentuk model umum, yang berarti belum dijelaskan secara rinci mengenai hasil dari proses *clustering* itu sendiri. Berikut adalah tampilan model kluster yang dihasilkan setelah dijalankan di Google Colab.

```
cluster_counts = df['cluster'].value_counts().sort_index()
print("Jumlah anggota dalam setiap cluster:")
print(cluster_counts)
```

```
Jumlah anggota dalam setiap cluster:
cluster
0      355
1       18
2         3
Name: count, dtype: int64
```

Gambar 12. Cluster Model

4.2 Pembahasan

Berdasarkan hasil analisis terhadap 376 data wajib pajak di Kecamatan Sakra menggunakan algoritma *K-Means*, maka diperoleh hasil *cluster* sebagai berikut : *Cluster 1* sebagai tingkat kepatuhan pembayaran pajak tinggi berjumlah

355 item, *Cluster 2* sebagai tingkat kepatuhan sedang berjumlah 18 item dan *cluster 3* sebagai tingkat kepatuhan rendah berjumlah 3 item. Hasil ini menandakan bahwa Algoritma *K-Means* merupakan salah satu algoritma *clustering* yang handal dalam mengelompokkan data, penentuan jarak menjadi faktor penentu *cluster* data yang dihasilkan. Dampak yang diperoleh dengan adanya hasil analisa ini, pihak pajak Kecamatan Sakra dapat memantau kepatuhan wajib pajak di kecamatan Sakra. Tindak lanjut dari hasil analisa tersebut dapat dijadikan sebagai langkah strategis untuk meningkatkan pembayar pajak. Adapun langkah strategis yang diambil untuk meningkatkan pembayaran pajak di kecamatan Sakra adalah dengan cara melakukan edukasi pada peserta wajib pajak mengenai pentingnya pajak bagi pembangunan Bangsa dan Negara, selain itu pula pelayanan pembayaran pajak di Kecamatan Sakra perlu ditingkatkan, guna melayani peserta wajib pajak dengan lebih baik, sehingga penerimaan pajak akan lebih meningkat.

Berikut ini adalah analisa tahapan dalam menentukan cluster data wajib pajak menggunakan Algoritma *K-Means Clustering* di kecamatan Sakra :

1. Siapkan dataset pajak yang sudah di *preprocessing* sebelumnya.
2. Penentuan jumlah *cluster* (k) data dilakukan dengan menggunakan metode Elbow, hal ini

dilakukan untuk memperoleh cluster optimal.

Dari hasil yang diperoleh didapatkan titik cluster terbaik yaitu sebanyak 3 cluster.

3. Tentukan titik pusat (centroid) awal secara acak. Dalam penelitian ini, titik pusat yang digunakan sebagai titik awal dapat dilihat pada tabel berikut :

Tabel 1. Centroid Awal

C0	4,078	126	26.718	26.718	0
C1	225	225	591.413	591.413	0
C2	620	620	2.732.800	2.732.800	0

4. Hitung jarak setiap data dari masing-masing *centroid* dengan menggunakan rumus korelasi *Euclidean Distance*.

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Proses perhitungan jarak data pertama ke masing-masing *centroid* awal. Pada tahap ini disebut dengan iterasi 1 :

$$d_1 = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

$$d_1 = \sqrt{\left[\begin{array}{l} (4078 - 4078)^2 + \\ (126 - 126)^2 + (26718 - 26718)^2 + \\ (26718 - 26718)^2 + (0 + 0)^2 \end{array} \right]}$$

$$d_1 = \sqrt{[(0) + (0) + (0) + (0) + (0)]}$$

$$d_1 = \sqrt{0} = 0$$

Dari hasil perhitungan diatas didapatkan hasil jarak data pertama dengan *centroid* pertama adalah 0.

Perhitungan jarak *Euclidean Distance* selanjutnya yaitu antara data perama dengan *centroid* kedua.

$$d_2 = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

$$d_2 = \sqrt{\left[\begin{array}{l} (4078 - 225)^2 + (126 - 225)^2 + \\ (26718 - 591413)^2 + \\ (26718 - 591413)^2 + (0 + 0)^2 \end{array} \right]}$$

$$d_2 = \sqrt{\left[\begin{array}{l} (14845609) + (9801) + \\ (318880443025) + \\ (318880443025) + (0) \end{array} \right]}$$

$$d_2 = \sqrt{637775741460}$$

$$d_2 = 798608,628$$

Hasil perhitungan jarak data pertama dengan *centroid* kedua adalah 798608,628.

Perhitungan jarak *Euclidean Distance* selanjutnya yaitu perhitungan jarak data pertama dengan *centroid* ketiga

$$d_3 = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$$

$$d_3 = \sqrt{\left[\begin{array}{l} (4078 - 620)^2 + \\ (126 - 620)^2 + (26718 - 2,732,800)^2 + \\ (26718 - 2,732,800)^2 + (0 + 0)^2 \end{array} \right]}$$

$$d_3 = \sqrt{\left[\begin{array}{l} (11995212164) + \\ (244036) + (7322921842724) + \\ (7322921842724) + (0) \end{array} \right]}$$

$$d_3 = \sqrt{14645844856848}$$

$$d_3 = 382697.459$$

Hasil perhitungan jarak data pertama dengan *centroid* ketiga adalah 382697.459.

Untuk perhitungan jarak data yang lain dapat dilakukan dengan cara yang sama.

5. Setelah diperoleh hitungan jarak antara data dengan *centroid*, langkah berikutnya adalah menghitung jarak minimum kemudian kelompokkan data berdasarkan jarak minimumnya.
6. Jika hasil analisis yang dilakukan secara matematis belum sama dengan hasil yang

diperoleh dari *Google Colab*, maka dilakukan proses iterasi lanjutan dengan menentukan titik *centroid* yang baru.

7. Setelah menentukan titik *centroid* baru untuk setiap klaster, lakukan iterasi berikutnya hingga tidak ada perubahan pada klaster dan semua data tetap berada pada klaster yang sama tanpa berpindah ke klaster lain

5. Kesimpulan

Penelitian ini menggunakan data pajak bumi dan bangunan Kecamatan Sakra tahun 2023 dengan jumlah sampel sebanyak 376 data dan membentuk 3 klaster ($K=3$). Hasil pengelompokan menggunakan algoritma K-Means Clustering membagi data ke dalam tiga kelompok, yaitu: Kelompok dengan tingkat kepatuhan tinggi, berjumlah 355 item, kelompok dengan tingkat kepatuhan sedang, berjumlah 18 item, kelompok dengan tingkat kepatuhan rendah, berjumlah 3 item. Hasil ini menunjukkan bahwa penentuan titik pusat cluster memiliki pengaruh signifikan terhadap jumlah iterasi yang dihitung selama proses clustering. Dengan penerapan metode ini, pihak pajak Kecamatan Sakra dapat dengan mudah memantau masyarakat yang kepatuhan wajib pajaknya tinggi, sedang, dan rendah

6. Daftar Pustaka

- [1] H. Pratiwi, M. Muhaimin, and W. O. Rayyani, "Kontribusi Pajak Bumi Dan Bangunan (Pbb) Dalam Meningkatkan Penerimaan Pajak Daerah," *Amnesty J. Ris. Perpajak.*, vol. 3, no. 1, pp. 24–30, 2020.
- [2] N. Wulandari and D. Wahyudi, "Pengaruh Pengetahuan Perpajakan, Sanksi Pajak, Kesadaran Wajib Pajak, dan Kualitas Pelayanan Pajak terhadap Kepatuhan Wajib Pajak dalam Membayar Pajak Bumi dan Bangunan di Desa Mranggen Kabupaten Demak," *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 14853–14870, 2022.
- [3] A. M. Nur, M. Saiful, H. Bahtiar, and M. T. Hidayat, "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 478–488, 2024.
- [4] S. Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021, doi: 10.29408/jit.v4i1.2986.
- [5] T. P. Sari, A. L. Hananto, E. Novalia, T. Tukino, and S. S. Hilabi, "Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit Hiv/Aids," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 104–114, 2023.
- [6] A. Hidayat and A. Muliawan Nur, "Implementasi Algoritma K-Means Untuk Klasterisasi Peserta Keluarga Berencana Berdasarkan Tingkat Risiko Kehamilan Di Desa Pringgasela Selatan," *Nopember*, vol. 1, no. 2, pp. 103–117, 2023.
- [7] N. Nurhidayati, L. Mauliya, and S. Suhartini, "Clustering Data Pasien Covid Berdasarkan Usia dan Gejala Menggunakan Algoritma K-Means," *Infotek*

- J. Inform. dan Teknol.*, vol. 6, no. 2, pp. 443–452, 2023, doi: 10.29408/jit.v6i2.17488.
- [8] A. M. Nur, M. F. Wazdi, B. Harianto, and M. F. Zaini, "Implementation of Naive Bayes Algorithm in Analyzing Acceptance of Poor Student Assistance," *J. Phys. Conf. Ser.*, vol. 1539, no. 1, 2020, doi: 10.1088/1742-6596/1539/1/012018.
- [9] Z. Setiawan *et al.*, *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia, 2023.
- [10] W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada $\text{\textcircled{L}}$ Lombok Vape On $\text{\textcircled{L}}$," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020.
- [11] O. Prastiwi and S. Mulyati, "PENERAPAN ALGORITME K-MEANS CLUSTERING DALAM PENGELOMPOKAN PENYAKIT PASIEN PADA UPTD PUSKESMAS WURYANTORO," in *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 2023, pp. 577–586.
- [12] N. K. Zuhail, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," in *Seminar Nasional Teknologi & Sains*, 2022, pp. 200–205.
- [13] M. Miranda, N. Rahaningsih, and R. D. Dana, "Analisis Clustering Data Anak Balita di Posyandu Kampung Sukarame Menggunakan Algoritma K-Means," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 136–141, 2024.
- [14] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.
- [15] J. F. Kolatung, "Analisis Tingkat Kepatuhan Wajib Pajak Bumi Dan Bangunan Di Kota Manado," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 9, no. 2, pp. 1006–1014, 2021.
- [16] A. Waruwu, M. Yetri, and F. Setiawan, "Implementasi Data Mining Dalam Mengelompokkan Data penduduk Kurang Mampu Menggunakan Metode K-Means Clustering," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 6, pp. 945–955, 2023.