

# MODEL ANALISIS *CLUSTERING K-MEANS* UNTUK PENGELOMPOKKAN PENDAPATAN DAERAH BERDASARKAN JENIS PEKERJAAN

Muhammad Faisal\*<sup>1</sup>, Janu Ilham Saputro<sup>2</sup>, Wiranti Sri Utami<sup>3</sup>, Suharmanto<sup>4</sup>

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi Universitas Raharja

<sup>2</sup>Program Studi Sistem Komputer, Fakultas Sains dan Teknologi Universitas Raharja

<sup>3</sup>Program Studi Teknik Informatika, Fakultas Teknik Universitas Cendekia Abditama

<sup>4</sup>Program Studi Sistem Informasi, Fakultas Teknologi Informasi Universitas Graha Ganesha

E-mail: muhammad.faisal@raharja.info\*<sup>1</sup>, janu@raharja.info<sup>2</sup>,

wirantisutami@uca.ac.id<sup>3</sup>, suharmanto.java@gmail.com<sup>4</sup>

## Abstrak

Pendapatan merupakan suatu penghasilan yang diterima oleh individu, Perusahaan, organisasi, bahkan Daerah yang diterima dari hasil kerja, usaha, aktifitas jual beli, dan layanan jasa kurun waktu tertentu. Penghasilan atau pendapatan disetiap daerah memiliki nilai bervariasi hal tersebut disebabkan oleh bentuk usaha yang dijalankan baik dalam sektor Industri, Jasa, maupun Pertanian. Pendapatan dari sektor Industri, Jasa, dan Pertanian akan memberikan dampak terhadap perkembangan ekonomi disuatu Daerah. Dalam penelitian ini penulis akan melakukan pengelompokan atau *cluster* dengan memanfaatkan metode K-terhadap data Pendapatan didasari dari sektor Industri, Jasa, dan Pertanian yang diperoleh dari *dataset* Badan Pusat Statistik (BPS) Pendapatan Daerah pada Bulan Februari 2025. Algoritma *Clustering K-Means* akan mengelompokkan pendapatan pada tiap Provinsi berdasarkan sektor pekerjaan. Penggunaan algoritma ini menghasilkan 2 cluster dengan jumlah *cluster* terendah berada pada *cluster* 2 dengan rata-rata pendapatan pada Bidang Pertanian 1570, Industri 1559, dan jasa 1994. Kemudian *cluster* tertinggi berada pada *cluster* 1 dengan rata-rata pendapatan jenis usaha sektor Pertanian 2084, pendapatan jenis usaha sektor industri 2491, dan pendapatan jenis usaha sektor jasa 2684. Dengan menerapkan Algoritma *cluster K-Means* dapat diketahui sektor mana saja yang memperoleh pendapatan terbesar.

**Kata Kunci**—Pendapatan, Sektor, Cluster, K-Means

## Abstract

*Income is an income received by individuals, companies, organizations, and even regions received from work, business, buying and selling activities, and service services over a certain period. Income or revenue in each region has a varying value, this is due to the form of business carried out, whether in the Industrial, Service, or Agricultural sectors. Income from the Industrial, Service, and Agricultural sectors will have an impact on economic development in a region. In this study, the author will conduct grouping or clustering using the K-method on Income data based on the Industrial, Service, and Agricultural sectors obtained from the Central Statistics Agency (BPS) Regional Income dataset in February 2025. The K-Means Clustering Algorithm will group income in each Province based on the work sector. The use of this algorithm produces 2 clusters with the lowest number of clusters in cluster 2 with an average income in the Agriculture sector of 1570, Industry 1559, and services 1994. Then the highest cluster is in cluster 1 with an average income of the Agriculture sector business type of 2084, income of the Industry sector business type of 2491, and income of the Service sector business type of 2684. By applying the K-Means cluster algorithm, it can be seen which sectors obtain the largest income.*

**Keywords**— Income, Sector, Cluster, K-Means

## 1. PENDAHULUAN

Memiliki penghasilan atau pendapatan baik pada bidang wiraswasta maupun pekerja individu merupakan harapan setiap orang, karena dengan memiliki penghasilan tetap seseorang dapat menjalankan roda perekonomian dengan aman baik untuk kebutuhan pokok maupun kebutuhan lainnya. Pendapatan merupakan suatu hasil yang bersifat kotor yang diperoleh dari suatu entitas didalamnya terdapat komponen pendapatan dari suatu individu maupun organisasi [1]. Jenis pendapatan terdapat 3 sektor yang merupakan konsep awal dari pengukuran pendapatan secara nasional yang bersumber dari sektor Tani, *Industry*, dan bidang Jasa [2]. Pada bidang Pertanian menjadi sektor yang paling diperhatikan secara khusus sebagai pusat Pembangunan daerah, dan akan berkesinambungan terhadap tata Kelola dan fungsi dari perolehan berkaitan dengan hasil pangan [3]. Sektor industri merupakan sektor ekonomi yang luas dan membentuk bagian besar yang disebut dengan sektor sekunder, sektor industri Adalah segmen ekonomi yang didalamnya terdapat proses pembuatan, pengiriman, dan produksi [4]. Pada sektor jasa menjadi penopang pertumbuhan ekonomi yang menghasilkan nilai dibawah maksimum, *Bright Institute* membuat kesimpulan sektor pendapatan dari bidang jasa mewakili 49% tenaga kerja dan memberikan sumbangsih 42% dari Produk Domestik Bruto [5]. Setiap Pemerintah daerah atau Provinsi mempunyai tanggung jawab dan kewajiban yang akan digunakan pada bidang yang menjadi perhatian khusus terhadap Sektor Tani, *Industry*, dan Bidang Jasa [6].

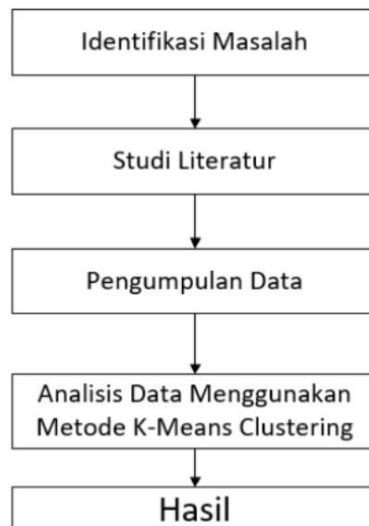
Untuk mengetahui pendapatan pada tiap daerah atau Provinsi dari beberapa sektor seperti Pertanian, Industri, dan Jasa maka penulis akan melakukan penelitian terkait pendapatan daerah atau Provinsi dari sektor tersebut. Penelitian ini akan menggunakan *Data mining* dan algoritma *cluster* K-Means. Dengan dilakukan penelitian ini diharapkan dapat dimanfaatkan oleh Pihak terkait atau Pemerintah Daerah dalam memantau pendapatan Masyarakat atau warga dari berbagai profesi seperti sektor Pertanian, Industri, dan Jasa. Penelitian terhadap sekelompok atau individu yang memiliki suatu profesi dapat dilakukan dengan menggunakan Teknik *Data mining*, dan pemanfaatan Algoritma *Cluster* K-Means. Algoritma *Cluster* K-Means Adalah metode non- hierarki yang mengambil separuh dari komponen *dataset* yang akan dijadikan *center cluster* awal [7].

Beberapa penelitian sebelumnya yang menggunakan algoritma *Cluster* K-Means yang dilakukan oleh Darmawan, dkk tentang Algoritma *Cluster* K-Means dalam prediksi penyakit hewan, Penelitian tersebut membuat informasi mengenai jumlah kasus penyakit pada suatu daerah yang dapat diakses secara luas dengan pemanfaatan teknologi informasi [8]. Kemudian terdapat penelitian yang dilakukan oleh Fatkhudin, dkk yang melakukan penelitian mengenai pengelompokan mahasiswa yang Menyusun atau menyelesaikan tugas akhir menggunakan *data mining*, Penelitian ini memanfaatkan metode *Cluster* K-Means sebagai metode yang digunakan melakukan Analisa *cluster* pembagian tugas akhir, algoritma ini digunakan karena paling efisien dan sederhana [9]. Kemudian terdapat penelitian yang dilakukan oleh Faisal dan Sita, yang meneliti tentang Universitas terbaik menggunakan *cluster* K-Means, setiap Universitas mengakui bahwa mereka Adalah yang terbaik namun terdapat beberapa kategori untuk mengetahui bahwa Universitas tersebut merupakan Universitas terbaik diantaranya dari sumber *World rank*, *Institution*, *Country*, dan *Score* penelitian ini menggunakan metode *Cluster* K-Means untuk uji coba data [10]. Berdasarkan referensi yang di lakukan dapat di ketahui bahwa algoritma *cluster* K-Means dapat diandalkan untuk Analisa pengelompokan data dan oleh karena itu penulis menggunakan Algoritma K-Means dalam Pengelompokan Pendapatan berdasarkan jenis Pekerjaan.

Pemodelan *Data mining* pada penelitian ini akan menggunakan algoritma *Cluster* K-Means dalam mengelompokkan pendapatan Daerah berdasarkan Sektor disetiap Provinsi Indonesia. Pendapatan daerah disetiap Provinsi berbeda-beda tergantung dari peraturan daerah yang berkaitan dengan UMR atau UMK. Dengan menerapkan penelitian menggunakan metode *cluster* K-Means dapat menghasilkan informasi berharga bagi Pihak terkait ataupun bagi para Peneliti lainnya yang akan melakukan Penelitian dibidang Algoritma *Cluster* K-Means.

## 2. METODE PENELITIAN

Penulis akan menjelaskan Langkah-langkah penelitian dalam bentuk kerangka pemikiran. Bentuk dari kerangka pemikiran adalah perumusan atau penjabaran yang menjelaskan tahapan dan proses penelitian yang sedang berjalan. Kerangka pemikiran tersebut dapat dilihat pada gambar dibawah ini.



**Gambar 1.** Kerangka Pemikiran

Berdasarkan gambar diatas penelitian ini akan melalui beberapa tahapan, diantaranya Identifikasi masalah yaitu melakukan identifikasi terhadap Masalah yang ada berdasarkan sumber atau data yang diterima. Kemudian ditahap selanjutnya Studi literatur yaitu penulis melakukan studi literatur *review* terhadap penelitian sebelumnya yang berkaitan dengan penelitian yang akan penulis bahas. Selanjutnya Pengumpulan data Penulis mencari data yang diperlukan dari sumber Badan Pusat Statistik (BPS). Proses selanjutnya Analisis Data Menggunakan Metode K-Means *Clustering* yaitu penulis melakukan Analisa dari beberapa tahapan menggunakan *cluster* K-Means. Dan yang terakhir Adalah hasil yaitu penulis memberikan informasi hasil terkait penelitian yang telah dilakukan.

Tahapan berikutnya Adalah Menyusun Analisa yang akan dilalui dari penelitian ini, Tahap tersebut diantaranya:

### 2.1. *Data Mining*

*Data mining* Adalah menggali informasi yang logis dalam Upaya menemukan informasi yang bermanfaat, setelah pola dan informasi telah ditemukan maka dapat digunakan sebagai media dalam pengambilan Keputusan [11]. *Data mining* Adalah proses untuk mengungkapkan hubungan yang tidak terduga yang terdapat dalam suatu data yang digunakan untuk mengungkap hubungan yang didasari dari diagram *scatterplots*, *scatterplots* Adalah proses pengujian linieritas koefisien korelasi yang memetakan titik berpasangan dalam grafik [12].

### 2.2. *Clustering* K-Means

*Cluster* merupakan proses yang digunakan untuk membagi data kedalam beberapa kelompok yang memiliki karakter dan menyerupai suatu jenis data dengan data lainnya, Algoritma ini memiliki dua jenis pengelompokkan yaitu hirarki dan non-hirarki [13]. Algoritma *Cluster* K-Means memerlukan beberapa Langkah dalam prosesnya, Langkah pertama Adalah menginisialisasi nilai pusat K dimana nilai K sama dengan jumlah *Cluster* yang akan dipilih untuk Kumpulan data tertentu [14].

2.3. Metode *Elbow*

Penelitian ini Penulis melakukan Analisa Data berdasarkan sumber Badan Pusat Statistik (BPS) yang telah terbentuk dalam table *dataset*, kemudian penulis akan melakukan Analisa pengelompokkan pendapatan Daerah dengan metode *Cluster K-Means* dan menggunakan metode *Elbow*. Algoritma *K-Means* melakukan tugasnya dengan meminimalkan *Within-cluster sum of square* (WCSS) [15]. Rumus *Euclidean distance*, dapat dilihat dibawah ini.

$$d = (b_i, a_t) = \sqrt{\sum_j^l = 1 (b_{ij} - a_{tj})^2}$$

Keterangan:

$d$  = Jarak nilai data & nilai *centroid*

$b_i$  = nilai data,  $i = 1,2,3, \dots, n$  = Jumlah data

$a_t$  = nilai *centroid cluster*,  $t = 1,2,3, \dots, k$  = jumlah *cluster*  $l$  = jumlah atribut

Kemudian penulis juga menggunakan rumus *Euclidean distance*, penggunaan metode *elbow* digunakan untuk mengamati nilai perbandingan antara jumlah *cluster* membentuk struktur siku pada suatu titik, semakin besar jumlah *cluster*  $k$  maka nilai (*Sum of Square Error*) akan semakin kecil Metode *elbow* memiliki kelebihan dan kekurangan, kelebihannya antara lain Sederhana, Penerapan luas, dan Teori yang jelas. Kekurangannya Adalah Subjektivitas, ambigu, dan memerlukan Perhitungan berulang. Berikut ini adala Rumus metode *elbow*.

$$SSE = \sum_k^k = 1 \sum x_1 \in sk ||x_1 - c_k||_2^2$$

3. HASIL DAN PEMBAHASAN

Penelitian yang akan dilakukan ini menggunakan *dataset* yang diperoleh dari Badan Pusat Statistic (BPS) dan menggunakan Bahasa Pemrograman R, *tools* yang digunakan Adalah *Posit Cloud*. Data yang diperoleh Adalah *dataset* Pendapatan daerah dari sektor Pertanian, Industri, dan Jasa, data tersebut akan dikelompokkan menggunakan Algoritma *Cluster K-Means*. Tabel dibawah ini merupakan *dataset* dari Pendapatan Daerah atau Provinsi dari sektor Pertanian, Industri, dan Jasa.

**Tabel 1.** *Dataset* Pendapatan Daerah Sektor Pertanian, Industri, dan Jasa

Provinsi	Pertanian	Industri	Jasa
Aceh	2301,5	1790	2163,7
Sumatera Utara	1157,2	1728,5	2140,5
Sumatera Barat	1326,1	1569,3	2147,3
Riau	2091,8	2399,7	2861
Jambi	2143,3	2489,8	2326,9
Sumatera Selatan	2182	2128,1	1930,3
Bengkulu	2483,3	0	2171,2
Lampung	2034,7	1905,9	1922,6

...	...	...	...
Maluku Utara	1642,6	2183,8	2254,5
Papua Barat	1642,2	1598,9	1841,2
Papua Barat Daya	1677,4	3596,2	2844,4
Papua	1918,6	2121,7	2961,2
Papua Selatan	2526,3	2781,2	2938,2
Papua Tengah	1821,5	3497,8	2972,3
Papua Pegunungan	1827,9	0	3307,5

Langkah berikutnya *dataset* yang telah didapatkan tersebut akan di *import* kedalam *tools Posit Cloud*. Hal ini bertujuan untuk dilakukan Analisa menggunakan *Cluster K-Means*. Gambar tersebut dapat dilihat dibawah ini.

Provinsi	Pertanian	Industri	Jasa
1 ACEH	2301.5	1790	2163.7
2 SUMATERA UTARA	1157.2	1728.5	2140.5
3 SUMATERA BARAT	1326.1	1569.3	2147.3
4 RIAU	2091.8	2399.7	2861
5 JAMBI	2143.3	2489.8	2326.9
6 SUMATERA SELATAN	2182	2128.1	1930.3
7 BENGKULU	2483.3	0	2171.2
8 LAMPUNG	2034.7	1905.9	1922.6
9 KEP BANGKA BELITUNG	2521.3	2495.2	2213.5
10 KEP RIAU	2126.1	1834.9	2666.3
11 DKI JAKARTA	2141.1	3146.4	3031.3
12 JAWA BARAT	1472	1465.5	2154.2
13 JAWA TENGAH	968.5	1420.9	1687.1
14 DI YOGYAKARTA	785.4	1175.3	1775.8
15 JAWA TIMUR	1206.4	2019.9	1669.7
16 BANTEN	1672.2	1730.2	2418.7
17 BALI	1787.2	1508.1	3080
18 NUSA TENGGARA BARAT	1364.9	1252.9	1508.4
19 NUSA TENGGARA TIMUR	1360.9	955.4	1304.5
20 KALIMANTAN BARAT	1654.1	2134.5	2535.9
21 KALIMANTAN TENGAH	1747.9	2815.9	2853.7
22 KALIMANTAN SELATAN	1836.4	1996.7	2103.9
23 KALIMANTAN TIMUR	2246.1	3104	3206.1
24 KALIMANTAN UTARA	2366.3	2392.7	2399.2

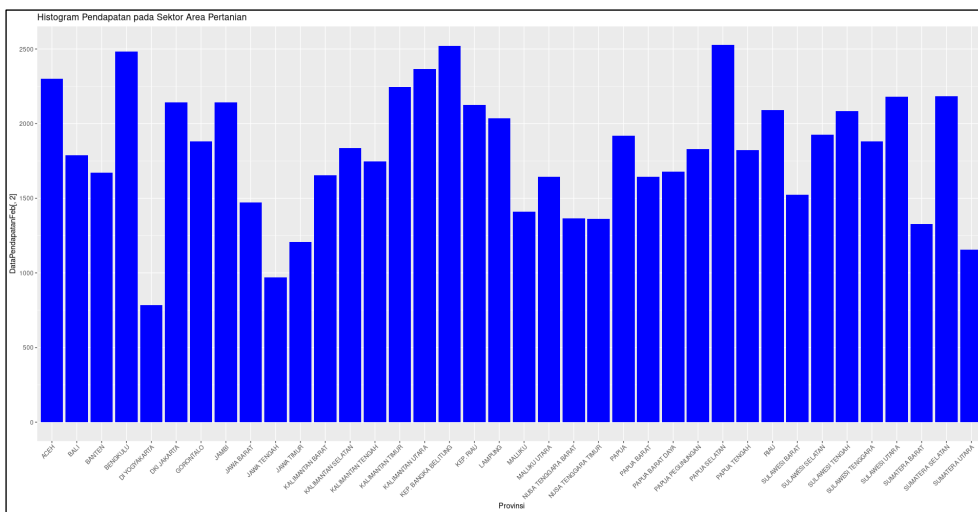
**Gambar 2.** *Dataset* Pendapatan Daerah yang telah di *import* ke *Posit Cloud*

Proses selanjutnya penulis akan mengambil data dari variabel numerik dan akan diberi nama index, kemudian akan menampilkan struktur objek index menggunakan fungsi khusus dari Bahasa R. Proses tersebut dapat diperhatikan pada gambar dibawah ini.

```
> index <- DataPendapatanFeb[,2:4]
> rownames(index) <- DataPendapatanFeb$Provinsi[1:38]
> str(index)
data.frame: 38 obs. of 3 variables:
 $ Pertanian: num 2302 1157 1326 2092 2143 ...
 $ Industri : num 1790 1728 1569 2409 2490 ...
 $ Jasa     : num 2164 2140 2147 2861 2327 ...
```

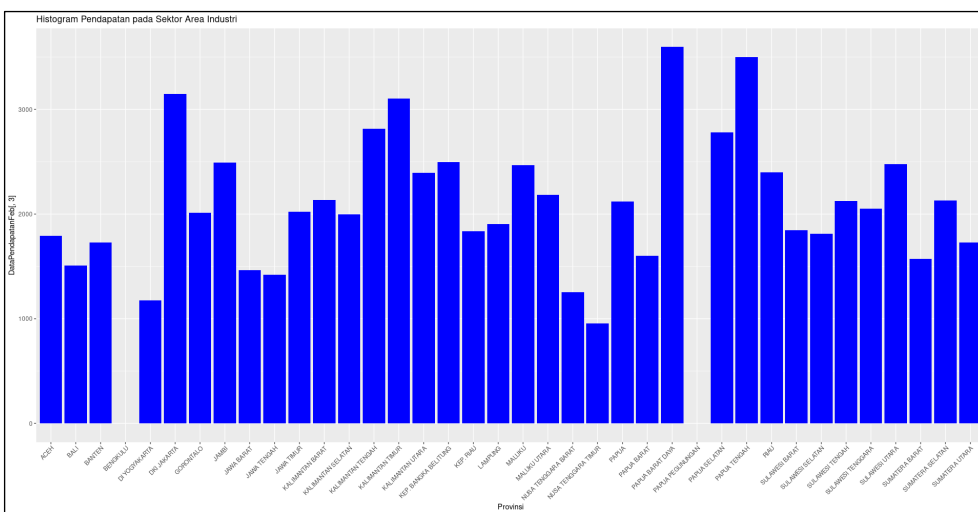
**Gambar 3.** Mengambil data numerik dari *dataset* Pendapatan Daerah

Proses membuat data dengan tampilan diagram histogram pada masing-masing sektor pendapatan daerah, gambar tersebut dapat diperhatikan dibawah ini.



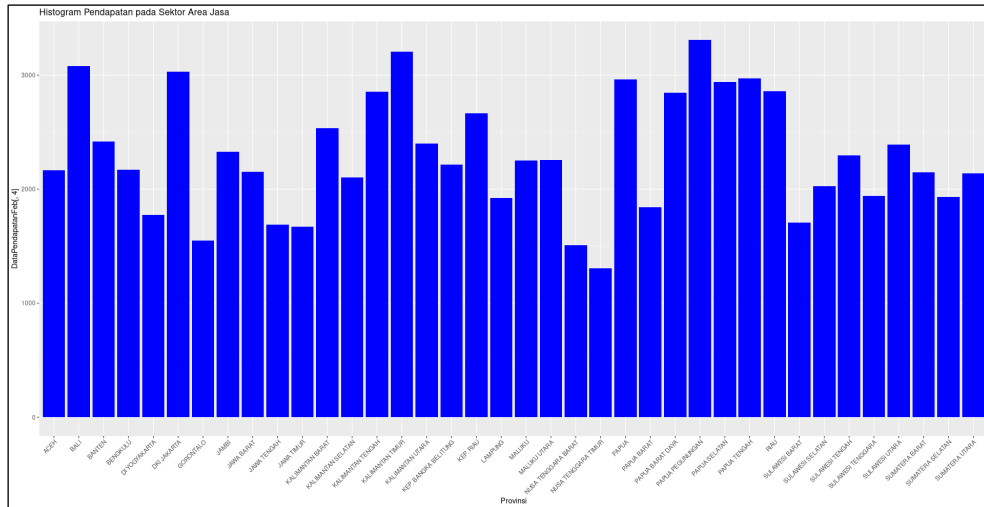
**Gambar 4.** Diagram histogram Pendapatan pada Sektor Pertanian

Pada Gambar visualisasi diatas dapat diperhatikan Sektor Pertanian dari Setiap Kategori memiliki Tingkat yang tinggi. Maka dapat dikatakan Sektor Pertanian merupakan Sektor yang menyumbang Pendapata Daerah cukup tinggi.



**Gambar 5.** Diagram histogram Pendapatan pada Sektor Industri

Pada gambar visualisasi diatas dapat diperhatikan hampir semua Provinsi memiliki Pendapatan pada Sektor Industri cukup tinggi juga kecuali pada Provinsi Bengkulu dan Papua Pegunungan.



**Gambar 6.** Diagram histogram Pendapatan pada Sektor Jasa

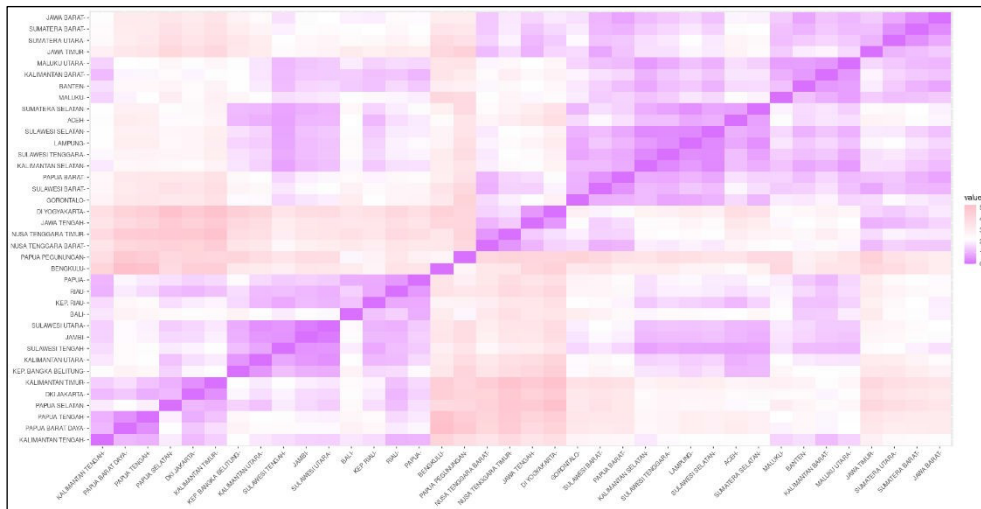
Pada gambar visualisasi diatas Sektor Jasa juga menyumbang Pendapatan yang cukup tinggi pada setiap Provinsi.

Selanjutnya penulis akan menghitung nilai korelasi masing-masing variabel menggunakan metode *Pearson*. Proses tersebut dapat dilihat pada gambar dibawah ini.

```
> korelasi <- cor(Data[,2:4], method = 'pearson')
> korelasi
      Pertanian Industri   Jasa
Pertanian 1.0000000 0.2751649 0.4385977
Industri   0.2751649 1.0000000 0.3894515
Jasa       0.4385977 0.3894515 1.0000000
>
```

**Gambar 7.** Nilai korelasi antar variabel menggunakan metode *Pearson*

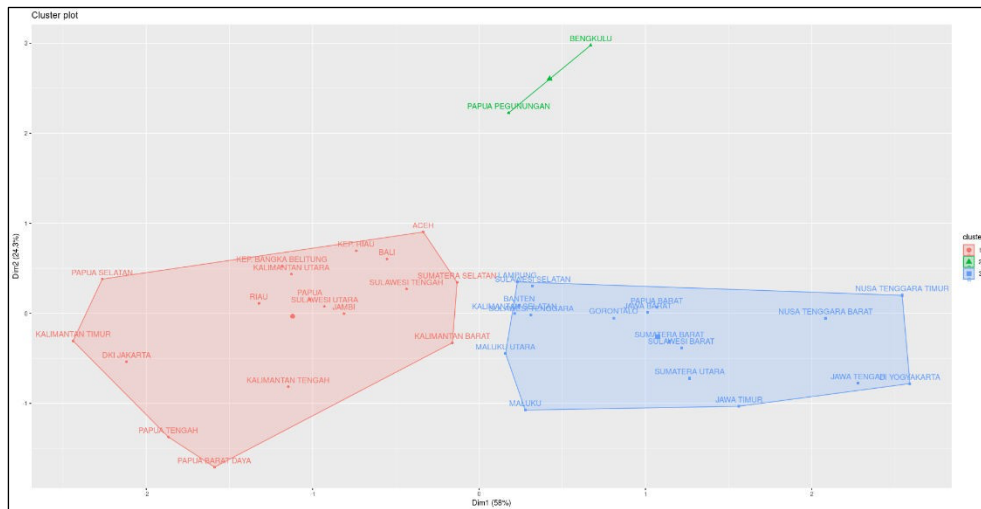
Pada proses selanjutnya penulis akan menghitung jarak antara observasi atau titik-titik dalam data yang telah distandarisasi dan selanjutnya akan dibuat kedalam bentuk visualisasi matriks jarak tersebut. Penulis akan membuat parameter jarak yang akan ditampilkan dalam bentuk plot gradien dan akan menyesuaikan skema warna dalam visualisasi jarak. Tampilan visualisasi tersebut dapat dilihat pada gambar dibawah ini.



Gambar 8. Visualisasi jarak matriks

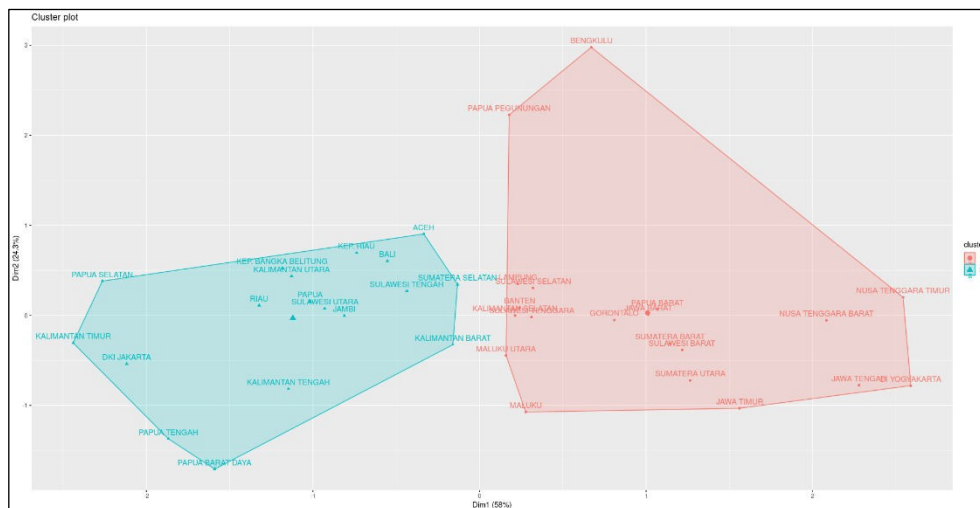
3.2. Cluster K-Means

Proses menentukan Analisa *cluster* K-Means penulis akan 2 *cluster* dan 3 *cluster*, proses pertama penulis akan menggunakan 3 *cluster* untuk mengatur secara acak nilai tertentu. Proses selanjutnya penulis akan melakukan perhitungan *Cluster* K-Means, dan akan menampilkan visualisasi dalam bentuk *cluster plot*, proses tersebut dapat dilihat pada gambar dibawah ini.



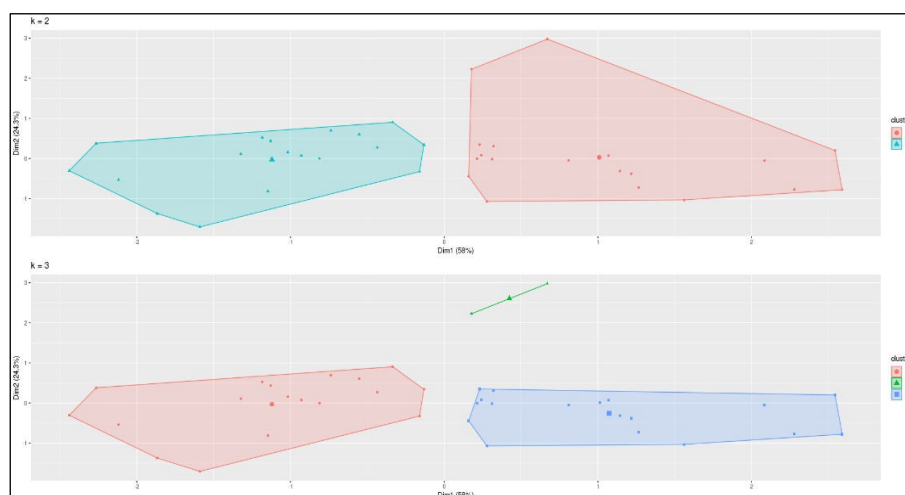
Gambar 9. Visualisasi Cluster plot dengan Cluster K-Means 3 cluster

Setelah melakukan *cluster plot* dengan 3 *cluster* penulis kemudian membuat *cluster* K-Means dengan 2 *cluster*. Tampilan tersebut dapat dilihat pada gambar dibawah ini.



**Gambar 10.** Visualisasi *Cluster plot* dengan *Cluster K-Means 2 cluster*

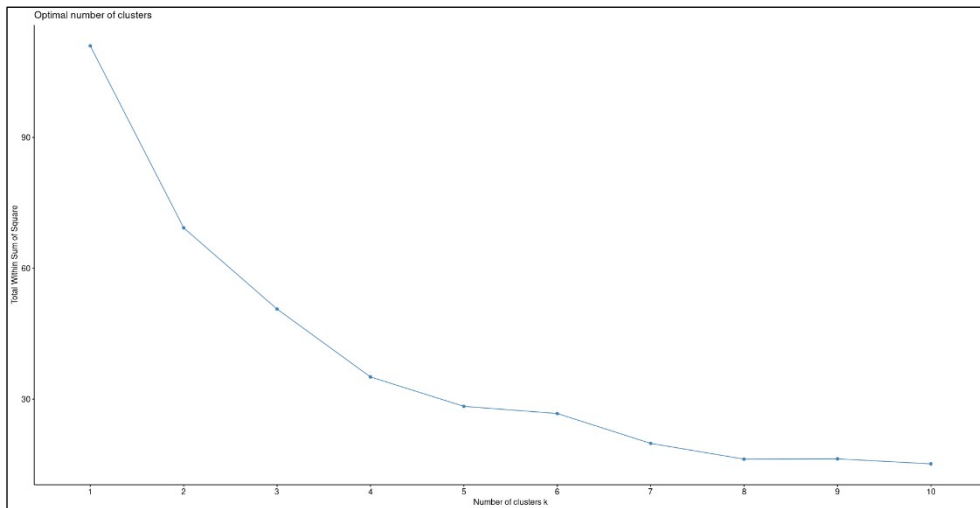
Setelah membuat tampilan visualisasi *cluster plot* dengan 3 *cluster* dan 2 *cluster* berikutnya penulis akan membuat tampilan perbandingan antara *cluster plot* 3 dengan 2, Tampilan tersebut dapat dilihat pada gambar dibawah ini.



**Gambar 11.** Tampilan perbandingan *cluster plot* 3 dengan 2.

Dari tampilan perbandingan *cluster plot* diatas dapat dilihat terdapat perbedaan sedikit, Dimana pada *cluster plot* dengan 3 *cluster* Provinsi Bengkulu dengan Provinsi Papua Pegunungan berada pada *cluster* 2. Sedangkan *cluster plot* 2 Provinsi Bengkulu dengan Provinsi Papua Pegunungan masuk kedalam *cluster* 1.

Pada tahap berikutnya penulis akan menentukan jumlah *cluster* optimal menggunakan metode *elbow*, metode *elbow* merupakan metode *Within-Cluster Sum of Squares* (WCSS) yang digunakan untuk mengukur sejauh mana data yang berada pada tiap *cluster* tersebar dari pusat *cluster*. Umumnya metode ini digunakan untuk menentukan jumlah *cluster* dan mengurangi jumlah variasi data secara signifikan. Proses tersebut dapat dilihat pada gambar dibawah ini.



Gambar 12. Tampilan Grafik menggunakan metode *elbow*

Pada tampilan grafik gambar diatas terdapat penurunan secara signifikan pada titik 2, selanjutnya terjadi penurunan Kembali pada titik berikutnya, penulis menyimpulkan bahwa proses ini akan menggunakan 2 cluster pada.

Proses selanjutnya penulis akan melakukan Analisa akhir dari Cluster K-Means dengan jumlah cluster sebanyak 2 cluster. Proses tersebut dapat dilihat pada gambar dibawah ini.

```

K-means clustering with 2 clusters of sizes 18, 20

Cluster means:
  Pertanian  Industri      Jasa
1  0.6353723  0.6392244  0.6676908
2 -0.5718350 -0.5753020 -0.6009217

Clustering vector:
      ACEH      SUMATERA UTARA      SUMATERA BARAT      RIAU      JAMBI      SUMATERA SELATAN
1         1         2         2         1         1         1
BENGKULU         2      LAMPUNG KEP. BANGKA BELITUNG      KEP. RIAU      DKI JAKARTA      JAWA BARAT
2         2         2         1         1         1         2
JAWA TENGAH         2      DI YOGYAKARTA      JAWA TIMUR      BANTEN      BALI      NUSA TENGGARA BARAT
2         2         2         2         2         1         2
NUSA TENGGARA TIMUR      KALIMANTAN BARAT      KALIMANTAN TENGAH      KALIMANTAN SELATAN      KALIMANTAN TIMUR      KALIMANTAN UTARA
2         2         1         1         1         1         1
SULAWESI UTARA         1      SULAWESI TENGAH      SULAWESI SELATAN      SULAWESI TENGGARA      GORONTALO      SULAWESI BARAT
1         1         1         2         2         2         2
MALUKU         2      MALUKU UTARA      PAPUA BARAT      PAPUA BARAT DAYA      PAPUA      PAPUA SELATAN
2         2         2         1         1         1         1
PAPUA TENGAH         1      PAPUA PEGUNUNGAN
1         2
    
```

Gambar 11. Analisa cluster K-Means Pendapatan Daerah dengan 2 cluster

Proses berikutnya akan dilakukan inisialisasi variabel dan kemudian akan dihitung nilai rata-rata atau *mean* dari beberapa variabel cluster yang telah ditentukan menggunakan metode K-Means. Proses tersebut dapat dilihat pada gambar dibawah ini.

```

> mean.cluster
# A tibble: 2 x 5
  Cluster Mean_Pertanian Mean_Industri Mean_Jasa Total_Mean
  <int>      <dbl>      <dbl>      <dbl>      <dbl>
1     1         2084.         2491.         2648.         7223.
2     2         1570.         1559.         1994.         5123.
    
```

Gambar 12. Nilai rata-rata antar variabel

Proses terakhir penulis akan menentukan table akhir sebagai Kesimpulan dari perhitungan *Cluster K-Means*. Tampilan table tersebut dapat dilihat pada gambar dibawah ini.

Cluster	Mean_Pertanian	Mean_Industri	Mean_Jasa	Total_Mean	Provinsi
1	2084	2491	2648	7223	ACEH, SUMATERA UTARA, SUMATERA BARAT, RIAU, JAMBI, SUMATERA SELATAN, BENGKULU, LAMPUNG, KEP BANGKA BELITUNG, KEP RIAU, DKI JAKARTA, JAWA BARAT, JAWA TENGAH, DI YOGYAKARTA, JAWA TIMUR, BANTEN, BALI, NUSA TENGGARA BARAT, NUSA TENGGARA TIMUR, KALIMANTAN BARAT, KALIMANTAN TENGAH, KALIMANTAN SELATAN, KALIMANTAN TIMUR, KALIMANTAN UTARA, SULAWESI UTARA, SULAWESI TENGAH, SULAWESI SELATAN, SULAWESI TENGGARA, GORONTALO, SULAWESI BARAT, MALUKU, MALUKU UTARA, PAPUA BARAT, PAPUA BARAT DAYA, PAPUA, PAPUA SELATAN, PAPUA TENGAH, PAPUA PEGUNUNGAN
2	1570	1559	1994	5123	ACEH, SUMATERA UTARA, SUMATERA BARAT, RIAU, JAMBI, SUMATERA SELATAN, BENGKULU, LAMPUNG, KEP BANGKA BELITUNG, KEP RIAU, DKI JAKARTA, JAWA BARAT, JAWA TENGAH, DI YOGYAKARTA, JAWA TIMUR, BANTEN, BALI, NUSA TENGGARA BARAT, NUSA TENGGARA TIMUR, KALIMANTAN BARAT, KALIMANTAN TENGAH, KALIMANTAN SELATAN, KALIMANTAN TIMUR, KALIMANTAN UTARA, SULAWESI UTARA, SULAWESI TENGAH, SULAWESI SELATAN, SULAWESI TENGGARA, GORONTALO, SULAWESI BARAT, MALUKU, MALUKU UTARA, PAPUA BARAT, PAPUA BARAT DAYA, PAPUA, PAPUA SELATAN, PAPUA TENGAH, PAPUA PEGUNUNGAN

Gambar 13. Tabel Analisa *cluster K-Means*

#### 4. KESIMPULAN

Penggunaan Algoritma *cluster K-Means* dalam penelitian ini untuk mengelompokkan data pendapata Daerah atau Provinsi pada setiap sektor pekerjaan diharapkan dapat menjadi pusat informasi tentang pendapatan dari sektor Tani, *Industry*, dan Bidang Jasa. Dapat diperhatikan bahwa perbandingan Pendapatan pada sektor Tani *cluster 1* adalah 2028 dan *cluster 2* adalah 1570, kemudian pada pendapatan sektor industri *cluster 1* adalah 2941 dan *cluster 2* adalah 1559, sedangkan untuk sektor pendapatan jasa *cluster 1* adalah 2648 dan *cluster 2* adalah 1994. Dapat disimpulkan bahwa sektor Jasa merupakan sektor pendapat yang paling tinggi.

#### 5. SARAN

Penulis berharap penelitian ini dapat memberikan informasi mengenai pendapata tiap Daerah atau Provinsi dan dijadikan sebagai landasan atau Upaya dalam meningkatkan sektor pendapatan dibidang lainnya.

## DAFTAR PUSTAKA

- [1] A. Hayes, "Pendapatan: Definisi, Rumus, Perhitungan, dan Contoh," Investopedia, 31 Mei 2025. [Online]. Available: <https://www.investopedia.com/terms/r/revenue.asp>. [Accessed 4 September 2025].
- [2] C. Afriza, M. Solekhah, R. F. Ahmad and M. L. Rohmi, "Pendapatan Nasional 3 Sektor Dan 4 Sektor," *PAJAMKEU : Pajak dan Manajemen Keuangan*, vol. 1, no. 3, pp. 45-62, 2024.
- [3] M. Batubara and M. M. Pane, "Pengaruh Pertanian terhadap Pendapatan Nasional," *Jurnal Penelitian Ekonomi Akuntansi (JENSI)*, vol. 7, no. 1, pp. 74-81, 2023.
- [4] D. Januaji, "Apa Itu Sektor Industri? Berikut Empat Kategori Ekonomi dan Contoh Utamanya," accurate, 19 Januari 2022. [Online]. Available: <https://accurate.id/bisnis-ukm/sektor-industri/>. [Accessed 8 September 2025].
- [5] H. R. Putra, "Sektor Jasa RI Didominasi Pekerjaan Informal, Ekonom Ini Sayangkan Nilai Tambah yang Rendah," Tempo, 2 Oktober 2024. [Online]. Available: <https://www.tempo.co/ekonomi/sektor-jasa-ri-didominasi-pekerjaan-informal-ekonom-ini-sayangkan-nilai-tambah-yang-rendah-3531>. [Accessed 8 September 2025].
- [6] N. I. Khomali, "PENERAPAN DATA MINING DALAM MENEMUKAN POLA DATA ANGGARAN PENDAPATAN BELANJA DAERAH KOTA MALANG METODE FP-GROWTH," *Jurnal Informatika*, vol. 23, no. 2, pp. 214-228, 2023.
- [7] M. Wahyudi, M. R. Saragih and S. , Data Mining, Jakarta: Yayasan Kita Menulis, 2020. [8] S. A. D. Darmawan and K. , "PENERAPAN METODE K-MEANSCLUSTERING DAN SIMPLE MOVING AVERAGE UNTUK MEMPREDIKSI JENIS PENYAKIT DI PROVINSI JAWA TIMUR," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 1, no. 4, pp. 877-886, 2024.
- [9] A. Fatkhudin, A. Khambali, F. A. Artanto, M. and N. A. P. Zade, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)," *Jurnal Minfo Polgan*, vol. 12, no. 2, pp. 777-783, 2023.
- [10] F. Dikarya and S. Muhami, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN UNIVERSITAS TERBAIK DI DUNIA," *Jurnal Informatika*, vol. 22, no. 2, pp. 124-131, 2022.
- [11] M. Arhami and M. Nasir, Data mining Algoritma dan Implementasi, Yogyakarta: Andi, 2020.
- [12] N. Purwati, "Data Mining," in *Terminologi Data Mining*, Banyumas, CV.ZT Corpora, 2021, pp. 2-23.
- [13] S. F. Mandang and B. N. Sari, "Penerapan K-Means Cluster pada Daerah Penggunaan Teknologi di Indonesia," *Journal of Information System*, vol. 6, no. 1, pp. 131-138, 2021.
- [14] V. Winland, "Apa itu k-means clustering?," IBM, 26 Juni 2024. [Online]. Available: <https://www.ibm.com/id-id/think/topics/k-means-clustering>. [Accessed 16 September 2025].
- [15] A. Rajsya, P. and A. R. Manga, "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," *LINIEAR*, vol. 1, no. 4, pp. 395-403, 2024.