

# Analisis Klaster Tingkat Pengangguran di Provinsi Kalimantan Timur Menggunakan Algoritma *K-Means*

Dimaz Harits<sup>1\*</sup>, Ari Andriyas Puji<sup>2</sup>, Marulan Andivas<sup>1</sup>, Dedi Dermawan<sup>2</sup>, Eaden Ahmed Thoriq<sup>1</sup>

<sup>1</sup>Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Balikpapan  
Jl. Pupuk Raya, Gn. Bahagia, Balikpapan, Kalimantan Timur, 76114

<sup>2</sup>Program Studi Teknik Industri Fakultas Teknik, Universitas Muhammadiyah Riau  
Jl. KH. Ahmad Dahlan No.88, Tampan, Riau, 28124  
E-mail:dimaz.harits@uniba-bpn.ac.id\*

## Abstract

*East Kalimantan is the province with the second-highest Regional Gross Domestic Product in Indonesia after Jakarta. The amount of GRDP in East Kalimantan does not match the open unemployment rate. The questions is, why does the Province with the 2nd highest GRDP have the 7th worst Labor Force Participation Rate? This study tries to answer from the level of cluster analysis. The grouping process can provide an initial perspective on the characteristics of the Open Unemployment Rate in East Kalimantan. The results of the study show that there are two clusters of unemployment rates with quite different characteristics. Cluster 1 has a fairly high open unemployment rate, with work participation below cluster 2. Interestingly, even though cluster 1's TPT and TPAK are inferior to cluster 2, their GRDP, HDI, and PMDN values far exceed cluster 2. This finding is quite interesting because the high and low GRDP is not in line with the Unemployment Rate, contrary to Okun's Theory.*

**Keywords:** *Clustering, Data Mining, K-Means Algotirhm, Unemployment Rate*

## Abstrak

Kalimantan Timur merupakan Provinsi dengan Pendapatan Domestik Bruto Regional tertinggi kedua di Indonesia setelah Jakarta. Besarnya PDRB Kalimantan Timur ternyata tidak sesuai dengan Tingkat Pengangguran Terbukanya. Hal ini tentu mengundang pertanyaan, mengapa Provinsi dengan PDRB tertinggi ke-2 justru Tingkat Partisipasi Angkatan Kerjanya terburuk ke-7. Penelitian ini mencoba menjawab dari tataran analisis klaster. Proses pengelompokanya dapat memberi sudut pandang awal tentang karakteristik Tingkat Pengangguran Terbuka di Kalimantan Timur. Hasil penelitian menunjukkan terdapat dua klaster Tingkat pengangguran dengan karakteristik cukup berbeda. Klaster 1 memiliki tingkat pengangguran terbuka yang cukup tinggi, dengan partisipasi kerja dibawah klaster 2. Menariknya, walaupun TPT dan TPAK klaster 1 kalah dengan klaster 2, nilai PDRB, IPM dan PMDN mereka jauh melampaui klaster 2. Temuan ini cukup menarik, sebab tinggi rendahnya PDRB tidak sejalan dengan Tingkat Pengangguran, bertolak belakang dengan Teori Okun.

**Kata kunci:** Algoritma K-Means, Data Mining, Klasterisasi, Tingkat Pengangguran

## 1. Pendahuluan

Kalimantan Timur merupakan Provinsi dengan Pendapatan Domestik Bruto Regional tertinggi kedua di Indonesia setelah Jakarta [1]. Tidak hanya itu, tepat ditengah-tengah Kalimantan Timur, pusat Pemerintahan Republik Indonesia akan berpindah. Samarinda dan

Balikpapan kini menjadi penyangga sekaligus gerbang menuju Ibukota Nusantara. Samarinda sebagai Pusat Pemerintahan Kalimantan Timur menjadi penyangga segala proses transisi otoritas Nusantara. Sementara itu Balikpapan, siap menjadi pusat bisnis dan gerbang investasi.

Keberadaan Ibukota Nusantara sudah selayaknya mendatangkan nilai positif bagi

Kalimantan Timur. Selain sebagai lahan basah tenaga kerja lokal, keberadaan IKN diharapkan mampu mendatangkan investasi besar bagi Kalimantan Timur.

Besarnya PDRB Kalimantan Timur ternyata tidak sesuai dengan Tingkat Pengangguran Terbukanya. Menurut laporan Badan Pusat Statistik Tingkat Pengangguran Terbuka di Kalimantan Timur sebesar 6.83% dengan Tingkat Partisipasi Angkatan Kerja sebesar 65.49%, terburuk ke-7 dari seluruh Provinsi di Indonesia [2].

Hal ini tentu mengundang pertanyaan, mengapa Provinsi dengan PDRB tertinggi ke-2 justru Tingkat Partisipasi Angkatan Kerjanya terburuk ke-7. Apakah Hukum Okun [3] yang menyatakan hubungan negatif antara Tingkat Pengangguran Terbuka dengan PDRB tidak berlaku? Benarkan semakin besar PDRB maka tingkat pengangguran semakin kecil?

Penelitian ini mencoba menjawab dari tataran analisis kluster. Analisis ini setidaknya akan menjadi dasar pada penelitian-penelitian yang lebih spesifik menjawab pertanyaan-pertanyaan diatas.

Analisis Kluster memungkinkan kita melihat ciri dari setiap elemen uji yang ada [4]. Proses pengelompokannya dapat memberi sudut pandang awal tentang karakteristik Tingkat Pengangguran Terbuka di Kalimantan Timur khususnya untuk masing-masing Kabupaten.

Beberapa penelitian telah melakukan hal serupa untuk kasus yang berbeda. Sarbaini dkk [5] melakukan analisis kluster menggunakan metode *Fuzzy Clustering* pada Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja, PDRB dan IPM di Provinsi Riau. Ia mengelompokkan kluster berdasarkan 2 kategori, yaitu Tingkat Pengangguran Tinggi dan Tingkat Pengangguran Rendah. Akan tetapi, penelitian ini berhenti pada proses Klusterisasi, tanpa mengupas dengan rinci karakteristik masing-masing Kluster.

Studi Anggraini [6] menganalisis Pengangguran Terbuka pada Provinsi Jawa Tengah dengan mempertimbangkan variable Tingkat Pendidikan dan realisasi PMDN Jawa Tengah sebagai acuan analisis.

Berbeda dari keduanya, Sahnoun dan Abdennadher [7] berusaha memodelkan keterkaitan antara Pembiayaan Kebijakan Pasar Tenaga Kerja negara-negara OECD dengan tingkat Penganggurannya. Uniknya, besaran tingkat Pembiayaan Kebijakan Pasar Tenaga Kerja berpengaruh signifikan mengurangi tingkat pengangguran.

Berbeda dengan studi-studi sebelumnya, penelitian ini melibatkan empat variabel untuk mencirikan karakteristik tingkat pengangguran di

Kalimantan Timur. Variabel-variabel tersebut adalah Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja, Pendapatan Domestik Regional Bruto, Indeks Pembangunan Manusia dan realisasi Penanaman Modal Dalam Negeri Provinsi Kalimantan Timur tahun 2021.

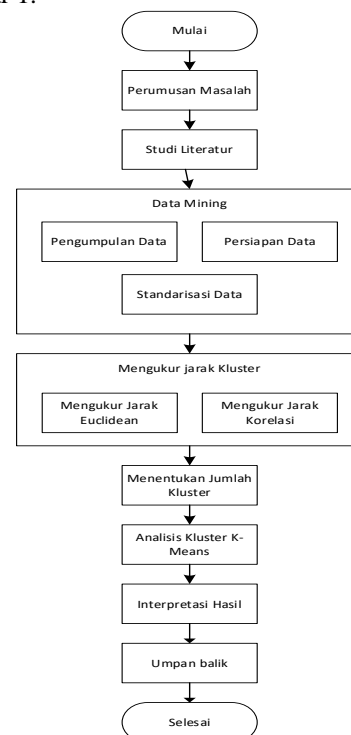
Masing-masing data merupakan data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Kalimantan Timur dan dapat diakses oleh khalayak umum.

Metode yang digunakan adalah *K-Means Clustering* menggunakan *R Programming*. *K-Means Clustering* berbasis *R Programming* dipilih karena praktis, mudah, serta dilengkapi fitur visualisasi data yang baik.

## 2. Metodologi

Untuk mencapai tujuan penelitian, dibutuhkan metodologi yang tepat dan akurat. Rangkaian metodologi penelitian ini secara umum terdiri dari proses perumusan masalah, studi literatur, pengumpulan dan pengolahan data, serta penarikan kesimpulan dari hasil pengolahan yang diperoleh.

Secara umum, garis besar metodologi penelitian ini dapat dilihat pada diagram alir Gambar 1.



Gambar 1. Diagram Alir Penelitian

Masing-masing data diperoleh dari laporan resmi yang dikeluarkan oleh Badan Pusat Statistik Provinsi Kalimantan Timur melalui laman

kalim.bps.go.id. Hasil olahan data yang akan digunakan dapat dili

Data kemudian dioleh agar siap untuk digunakan. Adapun rincian data yang digunakan sebagai masukan proses analisis kluster *K-Means* dapat dilihat pada Tabel 1.

**Tabel 1.**  
Data TPT, TPAK, PDRB, IPM, PMDN Provinsi Kalimantan Timur tahun 2021

No	Kabupaten	TPT (%)	TPAK (%)	PDRB (Juta Rupiah)	IPM (%)	PMDN (Juta Rupiah)
1	Paser	5.70	65.34	53995918.57	72.93	567121.50
2	Kutai Barat	5.14	72.53	32092829.61	72.07	1437227.90
3	Kutai Kartanegara	5.66	64.46	177318198.79	74.06	3319550.00
4	Kutai Timur	5.35	67.97	135378777.31	73.81	5279054.00
5	Berau	5.82	65.55	43024098.02	75.20	445264.40
6	Penajam Paser Utara	2.95	66.85	10087318.41	72.01	404886.00
7	Mahakam Ulu	3.14	74.83	2972544.04	67.95	575498.90
8	Balikpapan	8.94	64.29	111318497.73	80.71	16773016.60
9	Samarinda	8.16	64.29	71153222.80	80.76	585533.30
10	Bontang	9.92	66.63	57949557.05	80.59	910229.60

Sumber: BPS Kalimantan Timur (diolah)

Setelah data berhasil dikumpulkan dan disiapkan, rangkaian pengolahan data ini beralih ke proses standarisasi data. Standarisasi data merupakan proses konversi masing-masing data ke dalam skala yang sama sehingga bisa diolah dalam proses klusterisasi.

Kemudian data hasil standarisasi tersebut dihitung jarak korelasi dan euclidean. Proses perhitungan jarak ini dilakukan sebagai gambaran awal klasifikasi antar elemen yang diuji.

Langkah selanjutnya ialah proses penentuan jumlah kluster dan analisis kluster *K-Means*. Setelah hasil analisis diperoleh, proses pengambilan kesimpulan kemudian dilakukan.

### 3. Hasil dan Pembahasan

Penjabaran hasil dan pembahasan terdiri atas tiga sub bab yaitu hasil dan pembahasan proses Data Mining, Pengukuran Jarak Kluster dan Analisis Kluster *K-Means*. Adapun rincian pembahasan masing-masing sub bab sebagai berikut.

#### 3.1. Data Mining

Proses *data mining* meliputi proses pengumpulan data, persiapan data dan standarisasi data. Data yang akan digunakan dapat dilihat pada Tabel.1. Adapun hasil standarisasi data dapat dilihat pada Gambar.2.

	TPT	TPAK	PDRB	IPM	PMDN
[1,]	-0.1645433	-0.5335813	-0.27776167	-0.47712123	-0.48432604
[2,]	-0.4083111	1.4501051	-0.66942903	-0.67448740	-0.31320108
[3,]	-0.1819553	-0.7763690	1.92746663	-0.21779127	0.05699766
[4,]	-0.3168982	0.1920230	1.17751294	-0.27516515	0.44237587
[5,]	-0.1123073	-0.4756433	-0.47395793	0.04383365	-0.50829184
[6,]	-1.3616174	-0.1169796	-1.06292789	-0.68825713	-0.51623311
[7,]	-1.2789105	2.0846640	-1.19015309	-1.62000902	-0.48267845
[8,]	1.2458277	-0.8232712	0.74727106	1.30835408	2.70290848
[9,]	0.9062939	-0.8232712	0.02904236	1.31982886	-0.48070497
[10,]	1.6724214	-0.1776765	-0.20706337	1.28081461	-0.41684652
attr(,"scaled:center")					
	TPT	TPAK	PDRB	IPM	PMDN
	6.078	67.274	69529096.233	75.009	3029738.220
attr(,"scaled:scale")					
	TPT	TPAK	PDRB	IPM	PMDN

**Gambar 2.** Hasil Standarisasi Data

#### 3.2. Mengukur Jarak Kluster

Pengukuran jarak kluster melibatkan dua ukuran, yaitu jarak korelasi dan euclidean. Hasil pengukuran jarak masing-masing dapat dilihat pada Gambar 3 dan Gambar 4

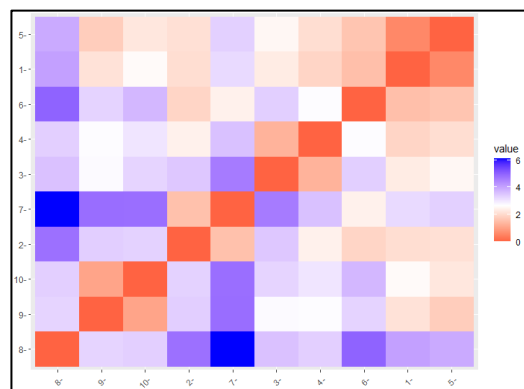
	1	2	3	4	5	6	7	8	9	10
1	0.0	2.1	2.3	1.9	0.6	1.5	3.2	4.1	2.1	2.6
2	2.1	0.0	3.5	2.4	2.1	1.9	1.5	4.8	3.4	3.3
3	2.3	3.5	0.0	1.3	2.5	3.4	4.6	3.6	2.7	3.3
4	1.9	2.4	1.3	0.0	2.1	2.7	3.6	3.4	2.7	3.0
5	0.6	2.1	2.5	2.1	0.0	1.6	3.3	3.9	1.7	2.2
6	1.5	1.9	3.4	2.7	1.6	0.0	2.4	5.0	3.3	3.7
7	3.2	1.5	4.6	3.6	3.3	2.4	0.0	6.1	4.8	4.8
8	4.1	4.8	3.6	3.4	3.9	5.0	6.1	0.0	3.3	3.4
9	2.1	3.4	2.7	2.7	1.7	3.3	4.8	3.3	0.0	1.0
10	2.6	3.3	3.3	3.0	2.2	3.7	4.8	3.4	1.0	0.0

**Gambar 3.** Hasil Perhitungan Jarak Euclidean

	1	2	3	4	5	6	7	8	9	10
1	0.0	1.5	0.5	1.0	0.8	2.0	1.6	0.9	0.6	0.5
2	1.5	0.0	1.6	1.1	1.4	0.3	0.0	1.7	1.7	1.4
3	0.5	1.6	0.0	0.2	1.3	1.5	1.5	0.8	1.0	1.3
4	1.0	1.1	0.2	0.0	1.8	1.0	0.9	1.0	1.5	1.8
5	0.8	1.4	1.3	1.8	0.0	1.4	1.6	0.9	0.1	0.1
6	2.0	0.3	1.5	1.0	1.4	0.0	0.2	1.3	1.6	1.6
7	1.6	0.0	1.5	0.9	1.6	0.2	0.0	1.7	1.8	1.5
8	0.9	1.7	0.8	1.0	0.9	1.3	1.7	0.0	0.7	0.9
9	0.6	1.7	1.0	1.5	0.1	1.6	1.8	0.7	0.0	0.1
10	0.5	1.4	1.3	1.8	0.1	1.6	1.5	0.9	0.1	0.0

**Gambar 4.** Hasil Pengukuran Jarak Korelasi

Untuk memudahkan proses pembacaan, hasil perhitungan jarak disajikan dalam bentuk visual korelasi pada Gambar 5.



**Gambar 5.** Visualisasi Korelasi antar Elemen

Angka pada sumbu x dan y Gambar.5 mengindikasikan urutan elemen data. Angka 1

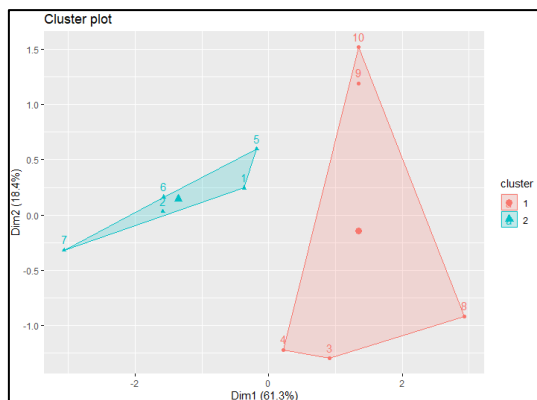
adalah Kabupaten Paser, angka 2 Kabupaten Kutai Barat dan seterusnya.

Indikator warna adalah nilai korelasi silang antar elemen yang tertera pada legenda grafik Gambar 5. Merah mendakan tingkat kecocokan tinggi, dan Biru menandakan tingkat kecocokan rendah. Contohnya, kecocokan karakteristik Tingkat Pengangguran 7 (Kabupaten Mahakam Ulu) dan 8 (Kota Balikpapan) berwarna biru tua, dengan kata lain kecocokannya sangatlah rendah.

### 3.3. Analisis Kluster K-Means

Sebelum melakukan analisis Kluster *K-Means*, maka jumlah kluster harus ditentukan lebih dulu. Proses penentuan jumlah kluster bisa berdasarkan studi literatur atau proses perhitungan menggunakan beberapa metode misalnya metode *Elbow*.

Penelitian ini menggunakan 2 kluster berdasarkan studi serupa yang dilakukan oleh Sarbaini, dkk [5]. Adapun hasil klusterisasi karakteristik Tingkat Pengangguran di Kalimantan Timur dapat dilihat pada Gambar 6.



**Gambar 6.** Hasil Klusterisasi Tingkat Pengangguran di Kalimantan Timur

Berdasarkan Gambar 6 diperoleh 2 kelompok kluster dengan karakteristik masing-masing. Tiadanya irisan dan jarak yg cukup jauh menunjukkan karakteristik Tingkat pengangguran pada masing-masing kluster signifikan berbeda.

Anggota kluster 1 adalah elemen 3 (Kutai Kertanegara), elemen 4 (Kutai Timur), elemen 8 (Balikpapan), elemen 9 (Samarinda), dan elemen 10 (Bontang).

Adapun anggota Kluster ke 2 adalah elemen 1 (Paser), elemen 2 (Kutai Barat), elemen 5 (Berau), elemen 6 (Penajam Paser Utara), dan elemen 7 (Mahakam Ulu).

Untuk memperoleh deskripsi karakteristik masing-masing kluster, nilai rata-rata variabel uji kluster dimunculkan. Adapun nilai rata-rata variabel uji masing-masing kluster dapat dilihat pada Gambar 7.

```
# A tibble: 2 × 6
```

Cluster	TPT	TPAK	PDRB	IPM	PMDN
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	7.61	65.5	110623651.	78.0	5373477.
2	4.55	69.0	28434542.	72.0	686000.

**Gambar 7.** Karakteristik Masing-Masing Cluster.

Tingkat Pengangguran Terbuka (TPT) pada kluster 1 terbilang tinggi, berada diangka rata-rata 7.61%, dengan rata-rata Tingkat Partisipasi Angkatan Kerja (TPAK) sebesar 65.5%. Rata-rata Pendapatan Domestik Regional Bruto (PDRB) terbilang sangat tinggi menyentuh angka 110 triliun rupiah dengan indeks pembangunan manusia sebesar 78.0%, nilai yang cukup tinggi. Rata-rata nilai realisasi Penanaman Modal Dalam Negeri cukup tinggi, sebesar 5,3 Triliun Rupiah.

Berbeda dari kluster 1, Tingkat Pengangguran Terbuka (TPT) pada kluster 2 terbilang rendah, berada diangka rata-rata 4.55%, dengan rata-rata Tingkat Partisipasi Angkatan Kerja (TPAK) melampaui kluster 1 sebesar 69.0. Akan tetapi, rata-rata Pendapatan Domestik Regional Daerahnya jauh lebih kecil dibandingkan kluster 1 sebesar 28 triliun pertahun dengan indeks pembangunan manusia terpaut 6 point dibawah kluster 1 sebesar 72%. Rata-rata nilai realisasi Penanaman Modal Dalam Negeri jauh tertinggal dengan kluster 1 sebesar 686 milyar Rupiah.

## 4. Simpulan

Tingkat pengangguran tahun 2021 berdasarkan TPT, TPAK, PDRB, IPM dan PMDN di Kalimantan Timur terbagi menjadi dua karakter kluster. Kluster 1 memiliki tingkat pengangguran terbuka yang cukup tinggi, dengan partisipasi kerja dibawah kluster 2. Menariknya, walaupun TPT dan TPAK kluster 1 di bawah kluster 2, nilai PDRB, IPM dan PMDN mereka jauh melampaui kluster 2. Temuan ini cukup menarik, sebab tinggi rendahnya PDRB tidak sejalan dengan Tingkat Pengangguran bertolak belakang dengan Teori Okun.

Temuan ini diharapkan menjadi acuan kebijakan pemerintah daerah. Bahwa solusi tingkat pengangguran dan kemajuan daerah tidak bisa dipukul rata. Masing-masing daerah memiliki karakteristik tingkat pengangguran yang berbeda, maka diperlukan kebijakan yang berbeda pula.

Temuan signifikansi positif antara Tingkat pengangguran dengan PDRB juga menjadi hal yang harus digaris bawahi. Diperlukan studi lanjutan untuk memahami fenomena dan disparitas yang terjadi.

**Daftar Pustaka**

- [1] BPS, "Produk Domestik Regional Bruto Kabupaten/Kota di Indonesia 2017-2021," Jakarta, 2022. [Online]. Available: <https://www.bps.go.id/publication/2022/06/07/59cf6a8a96b61d09c6d65260/produk-domestik-regional-bruto-kabupaten-kota-di-indonesia-2017-2021.html>
- [2] BPS, "Keadaan Ketenagakerjaan Indonesia Agustus 2021," Jakarta, 2022.
- [3] N. G. Mankiw, "The Savers – Spenders Theory of Fiscal Policy," *Am. Econ. Rev.*, vol. 90, no. 2, pp. 120–125, 2000, doi: 10.1257/aer.90.2.120.
- [4] T. Ullman, C. Hennig, and A.-L. Boulesteix, "Validation of cluster analysis results on validation data: A systematic framework," *WIREs Data Min. Knowl. Discov.*, vol. 12, no. 3, pp. 1–9, 2021, doi: <https://doi.org/10.1002/widm.1444>.
- [5] Sarbaini, W. Saputri, Nazaruddin, and F. Muttakin, "Cluster Analysis Menggunakan Algoritma Fuzzy K-Means Untuk Tingkat Pengangguran Di Provinsi Riau," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 1, no. 2, pp. 78–84, 2022.
- [6] Z. A. Anggraini, "Analisis Pengangguran Terbuka Di Provinsi Jawa Tengah Tahun 2015-2019," *J. Ilmu Ekon. JIE*, vol. 5, no. 4, pp. 712–722, 2021, doi: 10.22219/jie.v5i04.17820.
- [7] M. Sahnoun and C. Abdennadher, "A simultaneous-equation model of active labour market policies and change in unemployment rate: evidence from OECD countries," *Policy Stud.*, vol. 43, no. 1, pp. 3–20, 2022, doi: 10.1080/01442872.2020.1754384.