

Komparasi Klaster Pengangguran Terbuka di Indonesia Sebelum dan Saat Pandemi Covid-19 Menggunakan K-Mean Clustering

Rayhan Maliqi¹, Kursehi Falgenti²

^{1,2}Universitas Nusa Mandiri, Indonesia

Article Info

Article history:

Received Nov 18, 2022

Revised Feb 28, 2023

Accepted Mar 11, 2023

Keywords:

Data Mining

K-means Clustering

Unemployment

Covid-19

ABSTRACT

Open unemployment is a productive workforce of secondary education and higher education graduates who have not worked at all. The Indonesian Statistics Bureau (BPS) routinely publishes provincial open unemployment (TPT) data per semester. Many studies have analyzed TPT data by exploring new knowledge using data mining methods with cluster analysis techniques. Researchers have investigated the TPT data cluster under normal conditions before the COVID-19 pandemic. This study aims to find new knowledge from TPT data by comparing the TPT data cluster analysis results before the pandemic (2018-2019) with TPT data during the pandemic (2020-2021). Data mining techniques used are cluster analysis and the k-mean algorithm. The cluster analysis results are regional clusters with low and high unemployment rates before and during COVID-19. In addition, another finding is the movement from high to down clusters. Other interesting results are Covid-19 has the most impact on high unemployment in Aceh, North Sumatra, West Sumatra, Riau Islands, DKI Jakarta, West Java, Banten, East Kalimantan, North Sulawesi, Maluku, and West Papua. Anticipatory steps of the high open unemployment rates in 11 provinces, local government could design the program or policy that could support the BLT policy by the central government.

Copyright © 2023 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Kursehi Falgenti,

Prodi Sistem Informari Fakultas Teknologi Informasi, Universitas Nusa Mandiri,

Jl. Damai No. 8, Warung Jati Barat, Ragunan, Pasar Minggu, Jakarta Selatan.

Email: falgenti.kfe@nusamandiri.ac.id

1. PENDAHULUAN

Pengangguran terbuka merupakan golongan tenaga kerja produktif yang tidak bekerja sama sekali, mereka terdiri dari generasi muda yang baru menyelesaikan pendidikan menengah dan tinggi [1]. Tingkat pertumbuhan penciptaan lapangan kerja yang rendah menyebabkan tingginya angka pengangguran terbuka. Bila pemerintah membiarkan angka pengangguran terbuka tinggi, tanpa mengambil langkah antisipasi akan menimbulkan berbagai masalah, seperti; penurunan tingkat kesejahteraan, produktivitas dan pendapatan masyarakat. Penurunan produktivitas dan pendapatan masyarakat ini memicu timbulnya masalah-masalah kemiskinan, kriminalitas, dan ketimpangan standar hidup [2]. Pemerintah melalui Badan Pusat Statistik (BPS) Indonesia tiap tahun menerbitkan data angka Tingkat Pengangguran Terbuka (TPT) masing-masing provinsi pada bulan Februari dan bulan Agustus. Melalui data TPT ini, publik dapat mengetahui perubahan angka tingkat pengangguran terbuka di masing-masing provinsi. Data TPT sangat dinamis, data provinsi yang memiliki tingkat pengangguran terbuka tertinggi selalu berubah. Data TPT tidak hanya berisi informasi angka-angka tingkat pengangguran, bila digali lebih dalam kita dapat mengetahui pola pada data TPT. Pola baru yang ditemukan bisa menjadi sebuah pengetahuan baru untuk membantu pemerintah dan pihak terkait menyusun kebijakan dan program mengurangi tingkat pengangguran terbuka terutama di daerah.

Penambangan data (*data mining*) merupakan salah satu cara menemukan pengetahuan baru dari data. Definisi data mining secara sederhana adalah proses mengekstrak pola penting atau informasi yang menarik dari dalam kumpulan data [3]. Penambangan data terdiri dari *descriptive mining* dan *predictive mining*. *Descriptive mining* sangat bergantung pada statistik deskriptif, pendekatan kubus data atau *On-Line Analytical Processing* (OLAP), dan pendekatan induksi berorientasi atribut. Sedangkan *predictive mining* adalah menemukan pola yang berguna dalam data untuk membuat prediksi nontrivial pada data baru. Banyak metode dalam *data mining*, beberapa metode yang dapat digunakan perusahaan diantaranya; *Artificial Neural Networks*, *Cluster Analysis*, *Decision Trees*, *Rule Induction Methods*, *Genetic Algorithms*, *Ant Colony Optimization*, *Tabu Search*, *Support Vector Machines*, *Bayes* dll [4].

Metode *cluster analysis* merupakan metode terpercaya dan telah dipakai memecahkan banyak permasalahan kompleks untuk mengelompokkan data, diaplikasikan pada ilmu komputer dan statistik. Metode *cluster analysis* bekerja dengan membentuk kelompok dari titik-titik data numerik yang memiliki kemiripan yang tinggi dibanding dengan titik-titik data dikelompok data lainnya[5]. Metode ini merupakan cara paling alami mengelompokkan dataset dan cocok dipakai pada data tanpa label dengan distribusi multidimensi sebagai alternatif klasifikasi yang membutuhkan pelabelan data [6]. Karakteristik algoritma *clustering* dapat dibagi menjadi tiga. Pertama, partisi atau hirarki. Partisi membagi data observasi menjadi group sederhana sedangkan hirarki membagi data observasi menjadi group dan sub-group. Kedua, *hard clustering* atau *soft clustering*. *Hard clustering* setiap pengamatannya milik sebuah kelas tunggal sedangkan *soft clustering* dimana setiap pengamatan kemungkinan dimiliki oleh beberapa kelas. Ketiga, *centroid-based clustering* atau *density-based clustering*. *centroid-based clustering* bekerja dengan melihat jarak antara observasi dengan titik pusat *cluster* dan *density-based clustering* dengan melihat kerapatan lokal di sekitar observasi.

Peneliti banyak menggunakan algoritma *k-means* untuk pekerjaan mengelompokkan data. Cara kerja *k-mean* dengan menentukan jumlah awal kelompok dan *k-means* akan memilih pola *K* sebagai titik awal centroid secara acak. Algoritma *k-means* memproses secara berulang-ulang untuk mendapatkan basis data kluster. Jumlah iterasi untuk mencapai kluster *centroid* akan dipengaruhi oleh jumlah cluster yang akan dibantu. Kesederhanaan dan efisiensi merupakan keunggulan *k-mean* [6] dengan teknik yang sederhana dapat menghasilkan kluster yang optimum [7]. Namun demikian algoritma *k-mean* clustering memiliki beberapa kekurangan, diantaranya; 1) jumlah kluster harus di input ke *k-mean*, 2) selalu menghasilkan kelompok data walaupun pada data observasi tidak terdapat kelompok data. Bila kita meminta *k-mean* menghasilkan dua kelompok data dari data observasi, *k-mean* akan membagi data observasi menjadi dua kelompok data walaupun data observasi tersebut memiliki kesamaan. 3) *k-mean* tidak bekerja maksimal bila data tidak terpisah secara linier, 4) kinerja *k-mean* tidak akan maksimal bila kluster memiliki skala berbeda, bentuk berbeda atau jumlah data yang diobservasi tidak seimbang[6].

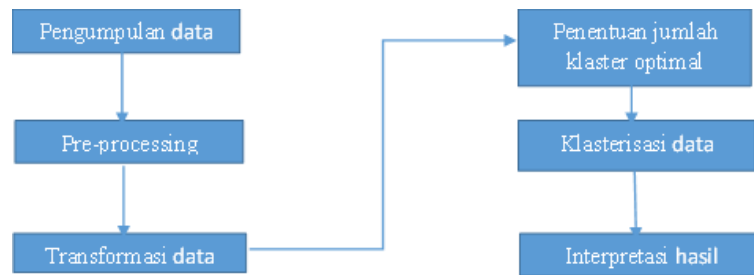
Beberapa peneliti telah menggali data TPT dan menemukan pengetahuan baru menggunakan *clustering analysis* dan algoritma *k-means*. Sembiring dkk [8] melakukan klusterisasi tingkat pengangguran di pulau Jawa dan pulau Bali menggunakan data TPT 2014-2019, penelitian ini menghasilkan tiga kluster data pengangguran. Penelitian lainnya oleh Tanjung [9] dengan melakukan klusterisasi tingkat pengangguran di seluruh provinsi di Indonesia menjadi dua kluster berdasarkan data TPT tahun 2014-2019. Pembagian data menjadi dua kluster ini bertujuan menemukan kelompok provinsi kluster tinggi dan kelompok provinsi kluster rendah dalam rentang waktu 9 tahun. Penelitian lainnya dengan metode yang sama oleh Akramunnisa & Fajriani di provinsi Sulawesi Selatan[10], penelitian ini menggabungkan data tingkat pengangguran terbuka (TPT) dengan variabel upah minimum kabupaten/kota (UMK) dan variabel laju pertumbuhan Indeks Pembangunan Manusia (IPM). Hasil penelitian ini adalah kluster kabupaten/kota dengan tingkat pengangguran tinggi sebanyak 3 kota/kabupaten sedangkan kluster wilayah kota/kabupaten tingkat pengangguran rendah sebanyak 21 kota/kabupaten. Penelitian [11] menggabungkan variabel TPT dengan variabel lain melakukan *clustering* jumlah angkatan kerja di Provinsi Sumatera Barat.

Penelitian analisis kluster data TPT diatas dilakukan saat kondisi perekonomian dalam keadaan normal. Diakhir tahun 2019 dunia mengalami *economic shock* [12] menyebabkan laju pertumbuhan ekonomi turun secara tiba-tiba karena wabah covid-19. Selain menyebabkan kematian, covid-19 mengakibatkan 8,8 juta orang kehilangan pekerjaan di tahun 2020 di Amerika Serikat [13]. Tingkat pengangguran naik menjadi 7.1 di sektor pariwisata dan pendidikan tahun 2020 di Australia[14]. Ekonomi Vietnam turun 7.1% di bulan agustus 2020[15], 100.000 orang kehilangan pekerjaan di Portugal [16]. Pada tri-semester ketiga tahun 2020, terdapat 13,7 juta orang kehilangan pekerjaan (14,6% dari penduduk usia 14 tahun ke atas) di Brazil, bertambah 3 juta orang pada Januari 2021. Sedangkan di Indonesia jumlah pengangguran akibat covid-19 29.2 Juta orang pada bulan Agustus 2020, turun menjadi 19.1 Juta orang pada bulan February 2021[17]. Berdasarkan data BPS, tingkat pengangguran terbuka Indonesia pada tahun 2018 sebelum pandemi covid-19 sekitar 5,3%, pada tahun 2019 sedikit menurun menjadi 5,23%. Pada awal pandemi tahun 2020 naik drastis menjadi 7,07% [18]. Kenaikan drastis angka pengangguran ditingkat nasional perlu dianalisa lebih lanjut pada tingkat provinsi.

Penelitian ini bertujuan menggali pengetahuan baru dari data TPT dengan melakukan komparasi hasil penambangan data TPT sebelum pandemi (2018-2019) dengan data TPT saat pandemi (2020-2021) menggunakan algoritma *k-mean*. Hasil klasterisasi dua kelompok data set TPT adalah pengetahuan lebih detail tentang dampak covid-19 pada masing-masing provinsi di Indonesia. Klasterisasi dua data set TPT menghasilkan kelompok provinsi pada klaster tinggi dan dan klaster rendah. Hasil klasterisasi data set TPT sebelum dan saat covid adalah pengetahuan baru mengenai provinsi mana saja yang berpindah klaster pada tahun sebelum adanya covid-19 dan saat adanya covid-19, pengetahuan baru selanjutnya adalah provinsi yang selalu berada pada klaster data rendah sebelum dan saat covid1919.

2. METODE

Penelitian ini menerapkan *data mining* untuk menambang data TPT. *Data Mining* atau bisa disebut juga *Knowledge Discovery in Database* (KDD) merupakan suatu proses penggunaan data yang bertujuan menemukan hubungan dari sekumpulan data dengan ukuran besar [19]. *Data Mining* berisi proses mencari trend atau pola yang diinginkan pada data dalam jumlah besar untuk membantu pengambilan keputusan pada waktu mendatang [20]. Metode *data mining* untuk menemukan pola pada data TPT adalah analisis klaster menggunakan algoritma *k-mean*. Tahapan analisis klaster data TPT terdiri dari 6 tahapan:



Gambar 1. Tahapan penelitian

1) Pengumpulan data

Data yang diklaster adalah data TPT sebelum pandemi bulan Februari dan Agustus tahun 2018 dan tahun 2019 selanjutnya dikomparasi dengan data TPT setelah pandemi bulan Februari dan Agustus 2020 dan tahun 2021. Sumber data TPT dari BPS [21].

Tabel 1. Data TPT per semester provinsi 2018 -2019 dan 2020-2021

Provinsi	A1	A2	A3	A4		B1	B2	B3	B4
Aceh	6.54	6.34	5.48	6.17	Aceh	5.40	6.59	6.30	6.30
Sumatera utara	5.61	5.55	5.57	5.39	Sumatera utara	4.71	6.91	6.01	6.33
Sumatera barat	5.68	5.66	5.38	5.38	Sumatera barat	5.25	6.88	6.67	6.52
Riau	5.55	5.98	5.36	5.76	Riau	4.92	6.32	4.96	4.42
Jambi	3.56	3.73	3.52	4.06	Jambi	4.26	5.13	4.76	5.09
Sumatera selatan	4.08	4.27	4.02	4.53	Sumatera selatan	3.90	5.51	5.17	4.98
Bengkulu	2.63	3.35	2.41	3.26	Bengkulu	3.08	4.07	3.72	3.65
Lampung	4.32	4.04	3.95	4.03	Lampung	4.26	4.67	4.54	4.69
Kep. Bangka belitung	3.59	3.61	3.32	3.58	Kep. Bangka belitung	3.35	5.25	5.04	5.03
Kep. Riau	7.30	8.04	7.02	7.50	Kep. Riau	5.98	10.34	10.12	9.91
.....
...
Maluku	7.07	6.95	6.61	6.69	Maluku	6.71	7.57	6.73	6.93
Maluku utara	4.56	4.63	4.96	4.81	Maluku utara	4.09	5.15	5.06	4.71
Papua barat	6.27	6.45	5.81	6.43	Papua barat	6.78	6.80	6.18	5.84
Papua	2.75	3.00	3.22	3.51	Papua	3.42	4.28	3.77	3.33

2) *Pre-processing*

Persiapan awal adalah membersihkan data untuk memastikan data set siap digunakan untuk *clustering*. Pekerjaan *pre-proccsing* diantaranya; membersihkan data dengan menghilangkan atribut nama provinsi yang tidak digunakan dalam analisis cluster. Selanjutnya memberi nama baru pada masing-masing atribut. Atribut yang diawali huruf A adalah data TPT sebelum pandemi, sedangkan atribut yang diawali huruf B adalah data TPT setelah pandemi. Penjelasan masing-masing atribut: atribut A1 = data TPT bulan February 2018, A2 = data TPT bulan Agustus 2018, A3 = data TPT bulan February 2019 dan A4 = data TPT bulan agustus 2019. B1 = data TPT bulan Februari 2020, B2 = data TPT bulan Agustus 2020.

3) Penentuan jumlah kluster optimal menggunakan teknik elbow

Menentukan titik optimal jumlah kluster menggunakan analisis sensitivitas WSS dengan metoda *elbow* (Thorndike 1953). Metode *elbow* bekerja dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Hasil persentase yang berbeda dari setiap nilai kluster dapat ditunjukkan dalam grafik sebagai sumber informasinya. Nilai kluster terbaik diperoleh dengan melihat sudut dalam grafik dan besarnya penurunannya. Nilai kluster pertama dengan nilai kluster kedua, nilai kluster kedua dan ketiga dan seterusnya dibandingkan. Nilai kluster yang memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai kluster tersebut adalah yang optimal.

4) Clustering data dengan k-mean

Menurut [1] *clustering* atau bisa disebut *segmentation*. Metode ini mengidentifikasi kelompok yang ada pada sebuah kasus dengan didasarkan kepada kelompok dengan atribut yang mirip satu sama lain. Cara kerjanya adalah dengan memisah beberapa kelompok data yang memiliki ciri khas masing-masing, dimana objeknya bisa berupa manusia, peristiwa dan lainnya yang didistribusikan ke dalam sebuah kelompok sehingga menjadi beberapa tingkatan yang akan saling terhubung diantara kluster, lemah dan kuatnya di antara anggota kluster yang berbeda akan terlihat pada anggota kluster yang sama.

Pada iterasi pertama, titik yang paling utama setiap kluster ditetapkan secara bebas [9]. Lalu dihitung antara tiap titik utama pada kluster dengan jarak data.

Langkah yang ada pada metode K-Means *Clustering* diantaranya :

- a) Menentukan jumlah kluster (k) yang ingin dibentuk seperlunya atau menggunakan metode analisis kluster terbaik.
- b) Nilai tengah atau *centroid* pada awal iterasi ditentukan secara *random*.
- c) Hitung jarak antara data dan setiap *centroid* menggunakan rumus *Euclidean Distance* untuk menghasilkan jarak terdekat antara setiap data dan *centroid*:

$$D_1(p, c)_n = \sqrt{\sum_{i=0}^n (p_i - c_i)^2}$$

p = data

c = *centroid*

i = iterasi

n = jumlah data

- d) Kelompokan data yang ada pada kluster dengan jarak yang paling pendek
- e) Menghitung untuk pusat kluster yang baru
- f) Langkah-langkah ke-dua hingga ke-empat dilakukan secara berulang sampai tidak ada lagi data yang berpindah ke kluster lain.

5) interpretasi hasil klusterisasi

menjelaskan hasil klusterisasi sesuai dengan tujuan penelitian melakukan komparasi kluster data TPT sebelum pandemi dan saat pandemi.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

Data TPT dari web BPS [21] dibagi menjadi dua data set. Data set TPT sebelum pandemi covid-19 di simpan dalam file excel terpisah dengan data TPT saat pandemi. Tahap pertama melakukan klusterisasi data TPT sebelum pandemi, setelah selesai, selanjutnya klusterisasi data TPT setelah pandemi.

Mengimport data TPT dengan jenis data numeric kedalam Rstudio TPT dengan perintah sebagai berikut:

```
TPT <- readxl::read_excel(path = "F:/Skripsi/Data/Buat Rstudio/TPT1819.xlsx", col_types =
c("numeric", "numeric", "numeric", "numeric"))
```

3.2 Pre-processing data

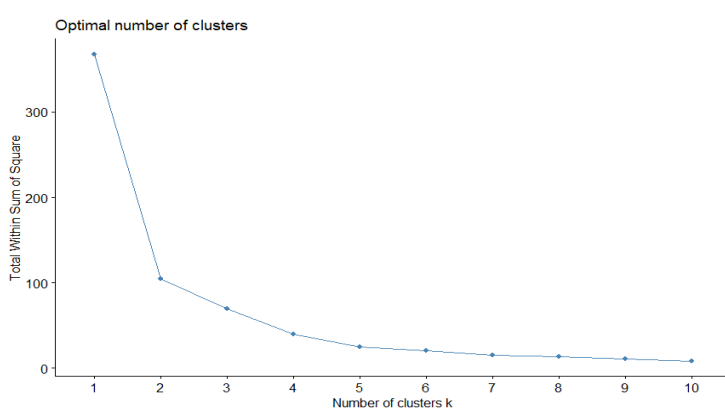
Pre-processing dimulai dengan menghilangkan atribut nama provinsi karena tidak dipakai dalam analisis kluster. Atribut A1, A2 adalah data TPT bulan Februari dan bulan Agustus tahun 2018, atribut A3 dan A4 adalah data TPT bulan February dan Agustus tahun 2019. Data TPT saat pandemi B1 dan B2 untuk data TPT bulan Februari 2020 dan Agustus 2020, B3 dan B4 adalah data TPT pada bulan February tahun 2021 dan Agustus 2021. Saat pre-processing semua data yang akan dikluster sudah dalam bentuk numerik. *Scaling* data tidak perlu dilakukan karena data sudah dalam jenis yang sama yaitu data presentase.

3.3 Menentukan klaster optimal

Jumlah klaster optimal diperoleh menggunakan metode wss dan teknik *elbow*. Dalam RStudio metode WSS dapat ditemukan pada paket `fviz_nbclust`.

```
fviz_nbclust(TPT, kmeans, method="wss")
```

Hasil analisis sensitivitas WSS dengan metoda *elbow* pada gambar 2. Dari pilihan 1 sampai 10 klaster, terlihat penurunan paling besar pada klaster 2. Patahan siku (*elbow*) juga terlihat pada klaster 2 dan klaster 4, tapi penurunan diklaster 4 tidak terlalu signifikan. Dengan demikian jumlah klaster optimal adalah 2



Gambar 2. Hasil jumlah klasster optimal

3.4 Klasterisasi menggunakan K-mean

Berikutnya adalah menjalankan perintah algoritma k-means untuk menghasilkan 2 pada data kelompok data TPT sebelum pandemi tahun 2018-2019, hasil klasterisasi akan disimpan dengan nama objek `am.Cluster` (gambar 3).

```
set.seed(1)
am.cluster <- kmeans(TPT, 2)
```

Gambar 3. Perintah algoritma K-Means untuk 2 Klaster

Untuk melihat hasil klasterisasi dari perintah sebelumnya menggunakan perintah `am.Cluster$Cluster` (gambar 4)

```
$cluster
[1] 2 2 2 2 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1
```

Gambar 4 Hasil Klasterisasi dari TPT tahun 2018-2019

Angka 1 menunjukkan provinsi yang berada di klaster C1 dengan jumlah 20 provinsi, sedangkan angka dua menunjukkan provinsi-provinsi yang berada pada kalaster C2 dengan jumlah 14 provinsi.

Selanjutnya untuk mengetahui centroid dari C1 dan C2 menggunakan perintah `am.cluster$centers`

```
> am.cluster$centers
      A1      A2      A3      A4
1 3.299000 3.596000 3.258000 3.657500
2 6.291429 6.527857 6.058571 6.219286
```

Gambar 5 centroid pada kelompok data pertama

Bisa dilihat jika C1 merupakan kelompok dari provinsi-provinsi dengan nilai tingkat pengangguran yang rendah dan C2 adalah kelompok dari provinsi yang memiliki nilai pengangguran tinggi (gambar 5). Kondisi ini terjadi sebelum pandemi covid-19 tahun 2018-2019.

Pada data set TPT2, yaitu data TPT saat pandemi tahun 2020-2021 perintah mengimport data TPT2 sama dengan data TPT sebelum pandemi covid-19.

```
TPT2 <- readxl::read_excel(path = "F:/Skripsi/Data/Buat Rstudio/TPT2021.xlsx", col_types = c("numeric",
"numeric", "numeric", "numeric"))
```

Setelah datanya dibaca Rstudio proses klasterisasi bisa dimulai dengan perintah seperti gambar 6.

```
set.seed(1)
am.cluster2 <- kmeans(TPT2, 2)
```

Gambar 6. Perintah klasterisasi data TPT2 menjadi dua cluster

Hasil dari pengolahan data TPT 2 adalah

```
$cluster
[1] 1 1 1 2 2 2 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2
```

Gambar 7. Hasil Klasterisasi TPT 2

Hasil klasterisasi data TPT2 yaitu data tahun 2020-2021 menghasilkan klaster 1 (C1) sebanyak 11 Provinsi sedangkan klaster 2 (C2) sebanyak 23 provinsi (gambar 7).

Centroid pada masing-masing klaster dihasilkan dengan menggunakan perintah `am.cluster2$centers`

```
> am.cluster2$centers
      B1      B2      B3      B4
1 6.158182 8.307273 7.503636 7.547273
2 3.651304 4.946087 4.548261 4.509130
```

Gambar 8. Nilai Tengah pada klasterisasi kelompok 2

Berdasarkan nilai centroidnya, C1 merupakan klaster provinsi dengan angka pengangguran yang tinggi sedangkan C2 merupakan klaster provinsi dengan angka pengangguran yang rendah (gambar 8). Berbeda dengan hasil klaster data TPT sebelum pandemi covid-19, dimana C1 merupakan dengan tingkat pengangguran rendah

3.5 Interpretasi hasil

Setelah menerapkan *data mining* dengan metode *K-Means Clustering* menggunakan data TPT sebelum pandemi tahun 2018-2019 provinsi di Indonesia, menghasilkan provinsi yang memiliki angka pengangguran tinggi (C2) ada sebanyak 14 provinsi, terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Riau, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Maluku dan Papua Barat. Sedangkan 20 provinsi dengan angka pengangguran rendah (C1) terdiri dari; Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, dan Papua

Kondisi di saat pandemi covid-19 yaitu pada tahun 2020-2021, sebelas provinsi memiliki angka pengangguran tinggi (C1) terdiri dari; Aceh, Sumatera Barat, Sumatera Utara, Kep. Riau, DKI Jakarta, Jawa

Barat, Banten, Kalimantan Timur, Sulawesi Utara, Maluku, Papua Barat. Sedangkan provinsi dengan tingkat pengangguran rendah (C2) terdiri dari 23 provinsi, diantaranya; Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, NTT, NTB Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara dan Papua.

Dari data diatas dapat dilihat terjadi perpindahan dari kluster tinggi ke kluster rendah. Provinsi sebelum covid-19 dengan tingkat pengangguran tinggi kemudian saat covid-19 berpindah menjadi kluster rendah adalah provinsi Riau, Kalimantan Utara, dan Sulawesi Selatan. Perpindahan ini bukan berarti tingkat pengangguran rendah di provinsi-provinsi diatas, tapi nilai centroid yang naik pada kluster TPT rendah dan TPT tinggi pada saat pandemi, kenaikan ini terjadi pada semua data TPT bulan February dan Agustus 2020 dan Bulan February dan Agustus Tahun 2021. Kenaikan tertinggi pada data TPT agustus 2022 dari 6.5 di tahun 2020 menjadi 8.3 di tahun 2021 (gambar 8).

```

> am.cluster$centers
      A1      A2      A3      A4
1 3.299000 3.596000 3.258000 3.657500
2 6.291429 6.527857 6.058571 6.219286

> am.cluster2$centers
      B1      B2      B3      B4
1 6.158182 8.307273 7.503636 7.547273
2 3.651304 4.946087 4.548261 4.509130

```

Gambar 8.

Perbandingan nilai tengah dari 2 kelompok data

Dari nilai centroid dua kelompok data bisa dilihat perbedaan yang cukup signifikan tingkat pengangguran terbuka sebelum pandemi covid-19 dan setelah covid-19. Nilai centroid yang sangat tinggi pada saat pandemi mulai terlihat pada bulan Agustus 2020 ini menunjukkan covid-19 sangat mempengaruhi meningkatnya angka pengangguran terbuka di Indonesia. Saat pandemi nilai centroid tertinggi juga pada bulan agustus 2020 sebesar 8.3 %.

4 PENUTUP

Hasil klasterisasi menggunakan k-mean clustering menunjukkan bahwa pandemi *Covid-19* sangat mempengaruhi kenaikan tingkat pengangguran terbuka pada provinsi di Indonesia. Dampak peningkatan angka pengangguran ini sangat dirasakan di provinsi yang sebelum pandemi covid-19 memiliki tingkat pengangguran yang tinggi yaitu Aceh, Sumatera Utara, Sumatera Barat, Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Sulawesi Utara, Maluku, dan Papua Barat. di saat covid 19 propinsi-provinsi tersebut tetap berada di kluster provinsi dengan angka pengangguran yang tinggi.

Dengan menganalisa data TPT sebelum dan saat covid-19 menggunakan teknik *clustering*, penelitian ini mendapatkan pengetahuan baru 11 provinsi yang paling berdampak terhadap covid-19 terhadap peningkatan angka pengangguran. Pengetahuan-pengetahuan dari hasil penambangan data TPT dapat membantu pemerintah provinsi pengambilan kebijakan. Pemerintah pusat telah mengantisipasi dengan mengeluarkan kartu pra-kerja memberikan bantuaun hidup kepada tenaga kerja yang terampak covid. Diharapkan pemerintah daerah di 11 provinsi yang memilki angka TPT tinggi sebelum dan saat covid-19 mendukung kebijakan pemerintah pusat mngantidipasi tingginya angka pengangguran dengan merancang program-program yang dapat memberi kesempatan bekerja bagi warganya yang tepat menghindari dampak lebih lanjut dari pengangguran seperti kemiskinan, kriminalitas, dan ketimpangan standar hidup pasca pandemi.

DAFTAR PUSTAKA

- [1] W. A. Amrullah, N. Istiyani, and F. Muslihatinningsih, "Analisis Determinan Tingkat Pengangguran Terbuka di Pulau Jawa Tahun 2007-2016," *e-Journal Ekon. Bisnis dan Akunt.*, vol. 6, no. 1, p. 43, 2019, doi: 10.19184/ejeba.v6i1.11074.
- [2] M. Wardiansyah, Yulmardi, and Z. Bahri, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat pengangguran (Studi kasus provinsi-provinsi Se-Sumatra)," *e-Jurnal Ekon. Sumberd. dan Lingkung. Vol.*, vol. 5, no. 1, pp. 13–18, 2016.
- [3] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014, doi: 10.21831/jpv.v4i2.2547.
- [4] T. W. Liao and E. Triantaphyllou, *Data Mining of Enterprise Data: Recent Advances in Algorithms and Applications*. World Scientific Publishing Co. Pte. Ltd., 2007.
- [5] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi:

- 10.25077/teknosi.v5i1.2019.17-24.
- [6] R. Garcia-Dias, S. Vieira, W. H. Lopez Pinaya, and A. Mechelli, "Clustering analysis," in *Machine Learning: Methods and Applications to Brain Disorders*, 2019, pp. 227–247.
- [7] T. H. Sardar and Z. Ansari, "An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm," *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 200–209, 2018, doi: 10.1016/j.fcij.2018.03.003.
- [8] F. Sembiring, S. B. Rizqi, M. A. Aziz, and D. Firmansyah, "Analisis Pemetaan Tingkat Pengangguran Di Pulau Jawa Dan Bali Dengan Metode K-Means," vol. 4, no. 1, 2019.
- [9] F. A. Tanjung, A. P. Windarto, and M. Fauzan, "Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 6, no. 1, p. 61, 2021, doi: 10.30645/jurasik.v6i1.271.
- [10] A.- Akramunnisa and F. Fajriani, "K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan," *J. Varian*, vol. 3, no. 2, pp. 103–112, 2020, doi: 10.30812/varian.v3i2.652.
- [11] D. Safira, M. Mustakim, E. D. Lestari, M. Iffa, and S. Annisa, "Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, p. 26, 2020, doi: 10.24014/rmsi.v6i1.8682.
- [12] W. Kawohl, "COVID-19, unemployment, and suicide," *The Lancet Psychiatry*, vol. 7, no. 5, pp. 389–390, 2020, doi: 10.1016/S2215-0366(20)30141-3.
- [13] S. M. Smith, "Unemployment rises in 2020, as the country battles the COVID-19 pandemic," *Mon. Labor Rev.*, vol. 2021, pp. 1–45, 2021, [Online]. Available: https://api.elsevier.com/content/abstract/scopus_id/85111248433.
- [14] H. S. Munawar, "Effects of COVID-19 on the Australian economy: Insights into the mobility and unemployment rates in education and tourism sectors," *Sustain.*, vol. 13, no. 20, 2021, doi: 10.3390/su132011300.
- [15] P. H. Nguyen, "Assessing the Unemployment Problem Using A Grey MCDM Model under COVID-19 Impacts: A Case Analysis from Vietnam," *J. Asian Financ. Econ. Bus.*, vol. 7, no. 12, pp. 53–62, 2020, doi: 10.13106/JAFEB.2020.VOL7.NO12.053.
- [16] F. Almeida, "The effects of COVID-19 on job security and unemployment in Portugal," *Int. J. Sociol. Soc. Policy*, vol. 40, no. 9, pp. 995–1003, 2020, doi: 10.1108/IJSSP-07-2020-0291.
- [17] R. T. Angita, R. Rinofah, and P. P. Sari, "Dampak Covid - 19 Terhadap Tingkat Pengangguran Di Indonesia," *J. Manag. Accounting, Econ. Bus.*, vol. 02, no. 01, pp. 486–491, 2021.
- [18] BPS, "Keadaan Ketenagakerjaan Indonesia Agustus 2021," *Ber. Resmi Stat. No. 84/11/Th. XXIV.*, vol. 2, 2021.
- [19] P. N. Harahap, "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT. Arma Anugerah Abadi Cabang Sei Rampah)," *Matics*, vol. 11, no. 2, pp. 46–50, 2019.
- [20] E. D. Sikumbang, "Penerapan data mining penjualan sepatu menggunakan metode algoritma apriori," *J. Tek. Komput. AMIK BSI*, vol. 4, no. 1, pp. 156–161, 2018.
- [21] BPS, "Tingkat Pengangguran Terbuka Menurut Provinsi," 2022. [Online]. Available: <https://www.bps.go.id/indicator/6/543/1/tingkat-pengangguran-terbuka-menurut-provinsi.html>.