# Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering

Rayhan Maliki[1], Kursehi Falgenti[2], Sinta Priani[3], Fajrul Fithri[4], Muhammad Suherman[5], Dwi Satria Nugraha[6]

[1,2,3,4,5,6]Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur -13620, Indonesia

e-mail: [1]rayhan.maliqia7x@gmail.com, [2]falgenti.kfe@nusamandiri.ac.id, sintapriani15@gmail.com, fajrul.faf@gmail.com, muhammadsuherman.bsi@gmail.com, dsatriaprilio@gmail.com

Artikel Info : Diterima : 17-04-2022 | Direvisi : 12-07-2022 | Disetujui : 21-07-2022

**Abstrak** - Masalah pengangguran berdampak pada kemiskinan, kriminalitas, dan ketimpangan taraf hidup. Pemerintah perlu mengantisipasi dampak tersebut melalui berbagai kebijakan. Pengetahuan memainkan peran penting dalam mendukung pengambilan keputusan dan perumusan kebijakan terkait dengan pengangguran. Beberapa peneliti telah menambang data untuk mendapatkan pengetahuan baru dari data Tingkat Pengangguran Terbuka (TPT) Indonesia. Eksplorasi data secara kontiniu berpeluang mendapatkan pengetahuan baru. Penelitian ini bertujuan untuk menambang data TPT Indonesia dari tahun 2016 hingga tahun 2021. Hal ini secara lebih spesifik untuk melihat perubahan cluster data TPT tahun 2016-2018 dengan cluster data TPT tahun 2019-2021. Penelitian ini merupakan penelitian analisis clustering menggunakan algoritma *k-means*. Tahapan penelitian ini meliputi: 1) pengumpulan data, 2) *pre-pocessing*, 3) penentuan jumlah cluster optimal, 4) clustering data dengan *k-mean*, dan 5) interpretasi hasil klasterisasi. Berdasarkan analisis *k-means clustering* pada data TPT 2016-2018 dan 2019-2021, terdapat 23 provinsi pada cluster 1 (TPT rendah) dan 11 provinsi di cluster 2 (TPT tinggi). Hanya Provinsi Riau yang naik ke *cluster* 1 (TPT rendah), dan hanya provinsi Sumatera Barat yang turun ke *cluster* 2 (TPT Tinggi). Hasil penelitian bukan hanya menghasilkan jumlah provinsi di cluster tinggi dan rendah tapi juga mengetahui provinsi yang pindah ke cluster yang lebih tinggi atau sebaliknya pada masing-masing data TPT.

Kata Kunci : Analisa clustering, data mining, k-means, pengangguran

***Abstracts*** *- The problem of unemployment has an impact on poverty, crime, and inequality in living standards. The government needs to anticipate these impacts through various policies. Knowledge is essential in supporting decision-making and policy formulation related to unemployment. Researchers have been mining data to gain new knowledge from Indonesia's Open Unemployment Rate (TPT) data. Continuous exploration of data has the opportunity to gain new knowledge. This study aims to mine Indonesian TPT data from 2016 to 2021. More specifically, look at changes in the 2016-2018 TPT data cluster with the 2019-2021 TPT data cluster. This research is a clustering analysis research using the k-means algorithm. The stages of clustering research consist of 1) data collection, 2) pre-processing, 3) determination of the optimal number of clusters, 4) data clustering with the k-mean, and 5) interpretation of the results of clustering. Based on the k-means clustering analysis of 2016-2018 and 2019-2021 TPT data, there are 23 provinces in cluster 1 (low TPT) and 11 provinces in cluster 2 (high TPT). Only Riau Province rose to cluster 1 (low TPT), and only West Sumatra province dropped to cluster 2 (High TPT). The study's results not only yielded the number of provinces in the high and low clusters but also found out which provinces moved to a higher cluster or vice versa in each TPT data.*

*Keywords : clustering analysis, data mining, k-mean, unemployment*

## PENDAHULUAN

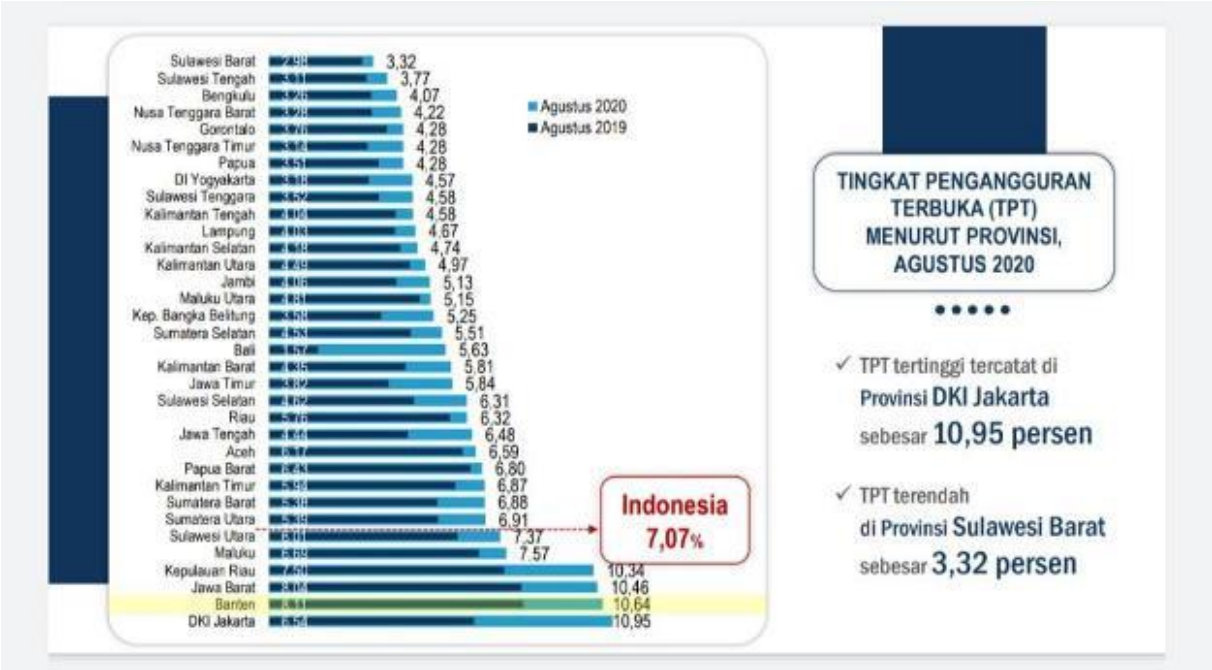
Seseorang atau kelompok orang yang sedang tidak memiliki pekerjaan atau tidak memiliki penghasilan tetap dianggap sebagai pengangguran. Sedangkan angkatan kerja yang sedang berusaha mencari pekerjaan dan mempersiapkan suatu usaha bisa dikategorikan sebagai pengangguran terbuka. Pengangguran terjadi karena banyaknya jumlah tenaga kerja yang tidak diimbangi dengan pertumbuhan kesempatan kerja (Amrullah et al., 2019), sebagai akibat dari minimnya lapangan pekerjaan di sebuah perusahaan maupun tempat usaha di suatu daerah.

Kondisi sosial ekonomi suatu negara dapat diketahui dari banyak angka pengangguran (Johan et al., 2016). Tingginya angka pengangguran di suatu negara menandakan terjadi ketimpangan dalam neraca ketenagakerjaan.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Masalah ini berkaitan erat dengan pertumbuhan ekonomi suatu daerah. Hubungan pertumbuhan ekonomi dan pengangguran bersifat positif dan negatif. Pertumbuhan ekonomi melalui GDP bersifat positif, namun tidak diikuti oleh peningkatan kapasitas produksi menyebabkan pengangguran tetap meningkat (Wardiansyah et al., 2016). Rendahnya tingkat pertumbuhan penciptaan lapangan kerja untuk menampung tenaga kerja menyebabkan munculnya masalah baru di suatu daerah seperti; menurunnya tingkat kesejahteraan, produktivitas dan pendapatan masyarakat, sehingga berdampak pada kemiskinan, kriminalitas, dan ketimpangan standar hidup. Oleh karena itu penting untuk menggali pengetahuan lebih dalam dari data pengangguran sebagai dasar pengambilan keputusan dan penentuan kebijakan oleh pemerintah mengantisipasi munculnya masalah baru akibat pengangguran di suatu daerah.




Sumber: Badan Pusat Statistik (BPS,2020)

**Gambar 1**. Persentase Tingkat Pengangguran Terbuka (TPT) Di Indonesia

Badan Pusat statistik (BPS) merilis data Tingkat Pengangguran Terbuka (TPT) pada bulan February dan bulan Agustus tiap tahun. Dari grafik tingkat pengangguran terbuka menurut provinsi di Indonesia dari bulan Agustus 2019 dan bulan Agustus 2020 (Gambar 1) dapat diketahui provinsi yang memiliki tingkat pengangguran terbuka tertinggi dan terendah di seluruh provinsi di Indonesia. Gambar 1 di atas menjelaskan bahwa TPT tertingi pada bulan agustus tahun 2020 terdapat pada provinsi DKI Jakarta Sebesar 10,95 persen dan TPT terendah terdapat pada provinsi Sulawesi Barat sebesar 3.32%. Sedangkan data TPT tertinggi pada bulan Agustus tahun 2019 di Provinsi Banten sebesar 10,34 persen dan terendah di Provinsi Bali sebesar 1.57 %. Grafik di atas memperlihatkan perbedaan TPT di masing-masing provinsi dalam rentang satu tahun bergerak sangat dinamis. Angka TPT provinsi tertinggi dan terendah berbeda tiap tahun.

Data TPT tidak hanya berisi informasi angka-angka yang menunjukkan tingkat pengagguran di daerah. Banyak pengetahuan yang bisa ditambang *(mining)* dari data tingkat pengangguran tersebut. Penambangan data *(data mining)* merupakan disiplin ilmu untuk menemukan pengetahuan baru *(insight)* dari data. *Data mining* merupakan kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi hasil penambangan ini adalah informasi baru yang belum ditemukan sebelumnya. Definisi yang lebih sederhana *data mining* dikemukan oleh Susanto & Sudiyatno (2014), merupakan proses ekstraksi informasi atau pola yang penting atau menarik dari data yang berasl dari database yang besar. Proses menambang data juga dikenal dengan nama *Knowledge Discovery* in *Databases* (KDD).

*Clustering* merupakan salah satu metode *data mining* terpercaya dan menjadi instrumen yang valid memecahkan masalah komplek ilmu komputer dan statistik. *Clustering* bekerja dengan mengelompokkan titik-titik data dalam dua kelompok atau lebih, dimana titik-titik data di dalam kelompok yang sama lebih mirip satu sama lain dibanding dengan kelompok data lainnya (Gustientiedina et al., 2019). Analisa *cluster* yang sering digunakan dalam penelitian adalah algoritma *k-means*. Keunggulan *k-means* mampu menghasilkan *cluster* yang optimum dengan konvergensi yang cepat (Sardar & Ansari, 2018).

Beberapa penelitian terdahulu menggunakan *k-means clustering* menemukan *insight* dari data TPT di Indonesia, diantaranya; klasterisasi tingkat pengangguran berdasarkan data TPT tahun 2014-2019 (Tanjung et al.,

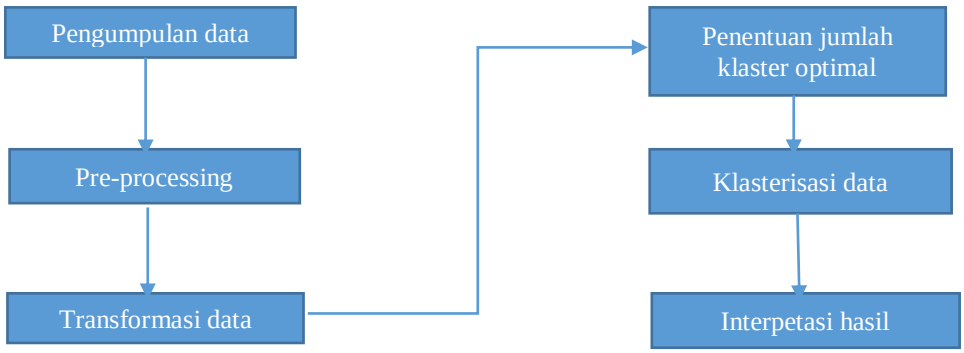
2021). Penelitian yang dilakukan Tanjung et al., (2021) hanya bertujuan menemukan kelompok provinsi berada di *cluster* tinggi dan kelompok provinsi yang berada dalam *cluster* rendah dalam rentang waktu 9 tahun. Penelitian yang telah dilakukan (Tanjung et al., 2021) menarik untuk dikaji lebih lanjut. Penelitian yang lebih spesifik dengan mengklasterisasi dua kelompok data TPT dengan rentang waktu masing-masing tiga tahun. Dari komparasi hasil dua kelompok data TPT diharapkan memperoleh pengetahuan baru. Sebelumnya penelitian klasterisasi tingkat pengangguran di wilayah pulau Jawa dan Bali menggunakan data TPT 2014-2019 dilakukan oleh Sembiring et al., (2019). *Cluster* data tingkat pengangguran dalam penelitian ini dibagi menjadi tiga *cluster*. Penelitian lainnya dengan metode yang sama dilakukan di provinsi Sulawesi Selatan oleh Akramunnisa & Fajriani (2020), penelitian ini tidak hanya menggunakan data tingkat pengangguran terbuka (TPT) untuk analisa *cluster*, tapi juga upah minimum kabupaten/kota (UMK) dan laju pertumbuhan indeks pembangunan manusia (IPM), penelitian ini menemukan 2 kluster kabupaten/kota dengan tingkat pengangguran tinggi sebanyak 3 kota/kabupaten sedangkan tingkat pengguran rendah 21 wilayah kota/kabupaten. Penelitian lain juga menggunakan *k-means clustering* dengan variabel TPT, ditambah dengan variabel lainya melakukan *clustering* jumlah penduduk Sumatera Barat berdasarkan angkatan kerja(Safira et al., 2020)

Dari penelitian di atas, para peneliti telah menemukan pengetahuan baru setelah melakukan penambangan data menggunakan metode analisis *cluster*. Pengetahuan tersebut berupa pemetaan tingkat pengguran berdasarkan *cluster* data TPT di tingkat nasional, regional maupun di tingkat provinsi. Penambangan data masih dapat dilakukan untuk untuk menemukan pengetahuan lainnya pada data TPT.

Penelitian ini bertujuan menggali lebih dalam data TPT dengan menganalisa lebih lanjut hasil penelitian Tanjung et al., (2021). Penelitian tersebut hanya melakukan analisa *cluster* pada data TPR 2014-2019, pada penelitian ini analisa *cluster* dilakukan pada dua kelompok data TPT berbeda, pertama kelompok data TPT tahun 2016-2018 kedua kelompok data TPT tahun 2019-2021. Analisa *cluster* pada penelitian ini tidak hanya mengetahui jumlah provinsi yang berada pada cluster TPT terendah dan tertinggi tapi juga perpindahan antar klaster dapat diidentifikasi. Hasil klasterisasi memberikan pengetahuan baru perubahan provinsi yang berhasil naik ke *cluster* yang memiliki TPT rendah atau provinsi yang turun ke *cluster* yang memiliki TPT tinggi.

## METODE PENELITIAN

Tahapan penelitian *clustering* data ini terdiri dari 6 tahap sebagaimana ditunjukan pada Gambar 2. Keenam tahapan tersebut adalah sebagi berikut:




Sumber: Penelitian (2022)

Gambar 2. Tahapan Penelitian

1. Pengumpulan data
Tahap pertama dimulai dengan pengumpulan data TPT. Data TPT yang dipakai adalah data TPT terbaru dari 34 provinsi yang dikeluarkan BPS Indonesia pada semester kedua atau bulan agustus tahun 2016-2021 (BPS, 2022).

2. Pre-processing
Persiapan awal pekerjaan yang dilakukan adalah membersihkan data dengan menghilangkan atribut nama provinsi yang tidak digunakan dalam analisis cluster. Selanjutnya untuk memudahkan identifikasi, masing-masing atribut diberi nama baru, yaitu: atribut A2016, A2017 dan A2018, A2019, A2020 dan A2021. Attribut A2016 artinya TPT bulan Agustus tahun 2016.

3. Transformasi data
Sebelum dilakukan proses penambangan, data dibagi menjadi 2 bagian; data TPT 2016-2018 dengan attribut

A2016,A2017 dan A2018 dan TPT 2019-2021 dengan atribut A2019, A2020 dan A2021.

4. Menentukan jumlah *cluster* optimal
Sebelum menjalankan teknik *cluster*, terlebih dulu ditentukan jumlah *cluster* optimal. Penentuan cluster optimal menggunakan analisis sensitivitas WSS dengan metoda elbow (Thorndike, 1953). WSS adalah salah satu kriteria untuk menghitung keragamaan data dalam *cluster* yang terbentuk. Semakin kecil keragaman data dalam *cluster* yang terbentuk menunjukkan bahwa *cluster* yang terbentuk sudah sesuai. Melalui kriteria WSS, kita dapat membandingkan jumlah *cluster* yang optimal untuk menganalisa data.

5. Klasterisasi dengan metode K-mean
Algoritma *k-means clustering* merupakan salah satu dari metode pengelompokkan data non-hirarki. Algoritma k-means mengelompokkan objek dengan terlebih dahulu mengindentifikasi data yang akan di *cluster*. *K-means* adalah algoritma yang sederhana dapat diterapkan pada data dengan jumlah kecil maupun besar. Pada iterasi pertama, titik utama setiap *cluster* ditetapkan secara bebas. Lalu dihitung antar jarak data dengan tiap titik utama pada *cluster*. Penjelasan cara kerja algoritma *K-mean* seperti dijelaskan dalm penelitian (Pradnyana & Permana, 2018) sebagai berikut:

a) Tentukan jumlah *cluster* (k), tetapkan pada pusat *cluster* sembarang.
*b)* Hitung jarak antara setiap data ke dalam *cluster* dengan jarak yang paling pendek dengan menggunakan persamaan ukuran jarak *Euclidean distance*
dengan persamaan:

$$D_I(X_1, X_2) = \|x_2 - x_1\| = \sqrt{\sum_{j=1}^{p}\{x_{2j} = x_{1j}\}^2} \quad (1)$$

Dimana $D_1(X_1, X_2)$ adalah jarak diantara data ke-i dan data ke-j

$X_{2j}$ adalah kordinat data $x_2$ pada dimensi j
$X_{1j}$ adalah kordinat data $x_1$ pada dimensi j dan P dimensi data

c) Kelompokkan data ke dalam *cluster* yang dengan jarak yang paling pendek dengan menggunakan rumus pada persamaan:

$$V_{ij} = \frac{\sum_{k=1}^{Ni} x_{kj}}{Ni} \quad (2)$$

$V_{lj}$ adalah data *cluster* ke – i kolom j
$X_{kj}$ adalah data ke – k kolom ke – j
$N_i$ banyaknya anggota *cluster* ke- i

d) Hitung pada pusat cluster yang baru menggunakan persamaan (1)
Ulangi langkah a) sampai dengan d) hingga tidak terjadi lagi perpindahan data pada *cluster* yang berbeda.

6. Interpretasi hasil
Hasil dari proses klasterisasi kemudian diinterpetasikan sesuai dengan tujuan penelitian untuk melihat perubahan klaster data TPT tahun 2016-2018 dibandingkan dengan klaster data TPT tahun 2019-2021.

Seluruh tahapan penelitian tersebut didukung oleh aplikasi R Studio versi 2022.02. Klasterisasi data menggunakan algoritma *k-means* dalam paket library clustering sedangkan cara menentukan jumlah cluster optimal dengan menganalisa patahan pada grafik hasil analisis WSS. Analisis WSS terdapat dalam library factoextra pada aplikasi R Studio.

## HASIL DAN PEMBAHASAN

### 1. Analisis Data

Tabel 1 adalah data set TPT 2016-2021 yang telah dibagi menjadi dua; data TPT1 (2016-2018) dan data TPT2 (2019-2021). Masing masing bagian terdiri data TPT pada bulan agustus pada masing-masing tahun. Data set terdiri dari data TPT dari seluruh provinsi di Indonesia 3 kolom mewakili tahun dan 34 baris mewakili data TPT masing-masing provinsi. Data TPT ini sudah dalam format numerik dan seragam karena itu tidak diperlukan proses standarisasi data.

Tabel 1 Data TPT 1 dan TPT2

| TPT 1 | | | | TPT 2 | | | |
|---|---|---|---|---|---|---|---|
| Prov | A2016 | A2017 | A2018 | Prov | A2019 | A2020 | A2021 |
| Aceh | 7,57 | 6,57 | 6,34 | Aceh | 6,17 | 6,59 | 6,30 |
| Sumatera Utara | 5,84 | 5,60 | 5,55 | Sumatera Utara | 5,39 | 6,91 | 6,33 |
| Sumatera Barat | 5,09 | 5,58 | 5,66 | Sumatera Barat | 5,38 | 6,88 | 6,52 |
| Riau | 5,09 | 6,22 | 5,98 | Riau | 5,76 | 6,32 | 4,42 |
| Jambi | 4,00 | 3,87 | 5,98 | Jambi | 4,06 | 5,13 | 5,09 |
| Sumatera Selatan | 4,31 | 4,39 | 4,27 | Sumatera Selatan | 4,53 | 5,51 | 4,98 |
| Bengkulu | 3,30 | 3,74 | 3,35 | Bengkulu | 3,26 | 4,07 | 3,65 |
| Lampung | 4,62 | 4,33 | 4,04 | Lampung | 4,03 | 4,67 | 4,69 |
| Sulawesi Tenggara | 2,72 | 3,30 | 3,19 | Sulawesi Tenggara | 3,52 | 4,58 | 3,92 |
| Gorontalo | 2,76 | 4,28 | 3,70 | Gorontalo | 3,76 | 4,28 | 3,01 |
| Maluku | 7,05 | 9,29 | 6,95 | Maluku | 6,69 | 7,57 | 6,93 |
| … | … | … | … | … | … | … | … |
| … | … | …. | … | … | … | … | … |
| Papua Barat | 7,46 | 6,49 | 6,45 | Papua Barat | 6,43 | 6,80 | 5,84 |
| Papua | 3,35 | 3,62 | 3,00 | Papua | 3,51 | 4,28 | 3,33 |

Sumber: Penelitian (2022)

Struktur data set TPT1 yang lebih rinci dapat dilihat menggunakan R Studio, terdiri dari matrix 34 X 4 dan memiliki 4 atribut, tapi hanya 3 atribut yang akan dipakai untuk analisa *cluster*, yaitu A2016, A2017 dan A2018 (gambar 3), atribut provinsi dihapus dari data set karena tidak berhubungan dengan tujuan penelitian.

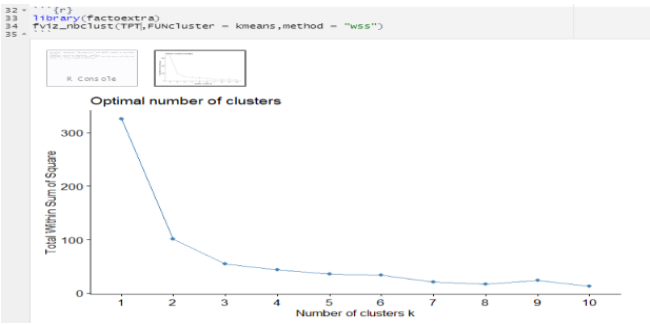
```
tibble [34 x 4] (S3: tbl_df/tbl/data.frame)
 $ Prov : chr [1:34] "ACEH" "SUMATERA UTARA" "SUMATERA BARAT" "RIAU" ...
 $ A2016: num [1:34] 7.57 5.84 5.09 5.09 4 4.31 3.3 4.62 2.6 7.69 ...
 $ A2017: num [1:34] 6.57 5.6 5.58 6.22 3.87 4.39 3.74 4.33 3.78 7.16 ...
 $ A2018: num [1:34] 6.34 5.55 5.66 5.98 5.98 4.27 3.35 4.04 3.61 8.04 ...
```

Sumber: Penelitian (2022)

Gambar 3. Data set TPT1

## 2. Penentuan Jumlah Cluster Optimal

Kriteria WSS dipakai untuk menentukan jumah kluster optimal. Dalam R Studio, fungsi fviz_nbclust dari package factoextra digunakan menjalankan WSS untuk memilih jumlah cluster optimal. Cara menentukan banyaknya cluster yang terbentuk dengan melihat patahan siku (*elbow*) pada kurva WSS. Berdasarkan kriteria WSS jumlah cluster optimal yang dihasilkan dari data TPT adalah 2 cluster (gambar 4).




Sumber: Penelitian (2022)

Gambar 4 Jumlah kluster optimal 2

### 3. Penerapan *k-means clustering*

*K-means clustering* diterapkan dengan menggunakan *library(cluster)* yang telah tersedia pada RStudio. Proses klusterisasi dimulai dengan menentukan nilai *centroid* (pusat cluster) awal secara random dari data TPT1. Setelah nilai *centroid* awal ditentukan, selanjutnya menghitung jarak masing-masing data terhadap *centroid*. Perhitungan jarak menggunakan *Euclidean Distence*. Data yang memiliki jarak terkecil dengan *centroid* membentuk cluster. Jumlah *cluster* optimal sebelumnya sudah ditentukan ditentukan 2 *cluster* menggunakan WSS. Jarak data ke masing-masing *centroid* menentukan dicluster mana data tersebut bergabung.

Setelah proses penentuan *cluster* tahap pertama selesai, dilanjutkan dengan iterasi kedua dengan cara yang sama tapi posisi *centroid* berbeda. Pada iterasi kedua ini terjadi perpindahan data dari satu cluster ke cluster lainnya berdasarkan jarak data dengan *centroid* di iterasi kedua. Iterasi dilanjutkan sampai tidak lagi terjadi perpindahan posisi data, menandakan dua cluster data telah terbentuk berdasarkan jarak rata-rata terhadap *centroid*. Gambar 5 merupakan hasil penerapan k-means clustering data TPT 1. Tanda panah merupakan posisi data provinsi Sumatera Barat

```
45 an.cluster$cluster
46 ```

 [1] 2 2 1 2 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2 1

47 ```{r}
48 table(an.cluster$cluster)
49 ```

 1  2
23 11
```

Sumber: Penelitian (2022)

Gambar 5 Hasil *cluster* data TPT1 dengan k-mean clustering

Hasil klasterisasi terdapat 23 provinsi pada *cluster* 1 dan 11 provinsi di *cluster* 2. Empat data pertama menunjukkan Provinsi Aceh di *cluster* 2, Sumatra Utara di *cluster* 2, Sumatera barat di *cluster* 1 dan Riau di *cluster* 2. Selanjutnya dilakukan penerapan *k-mean clustering* pada data TPT2 (2019-2021). Langkah-langkah penerapan algoritma *k-mean clusering* sama dengan tahap sebelumnya. Hasil penerapan algoritma *k-means clustering* dapat dilihat di gambar 6. Hasil *k-means clusteiring* pada data TPT2 menunjukkan jumlah data pada *cluster* 1 dan *cluster* 2 sama banyak dengan hasil *k-mean clustering* pada data TPT1 yaitu 23 provinsi pada *cluster* 1 dan 11 provinsi pada *cluster* 2. Tapi terjadi perubahan anggota *cluster*. Hasil *k-mean clustering* pada data TPT1 provinsi Sumatera Barat berada di *cluster* 1. *K-mean clusteiring* pada data TPT2 posisi Sumatra Barat berubah menjadi *cluster* 2. Sebaliknya provinsi Riau pada *k-means clustering* data TPT1 di *cluster* 2, kemudian pindah ke *cluster* 1. Perubahan kluster tidak ditemukan pada provinsi lain.

```
83 ```{r}
84 am.cluster$cluster
85 ```

 [1] 2 2 2 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2 1

86
87 ```{r}
88 table(am.cluster$cluster)
89 ```

 1  2
23 11
```

Sumber: Penelitian (2022)

Gambar 6 Hasil klasterisasi data TPT2

### 4. Interpretasi cluster yang terbentuk

Selanjutnya dilakukan interpretasi cluster yang terbentuk pada dua data TPT berdasarkan nilai *centroid*. Pada gambar 7a adalah nilai *centroid* data TPT1 dan TPT2. Nilai *centroid* pada *cluster* 1 data TPT 1 pada tahun 2016, 2017dan 2018 sebesar 3.85, 4.06 dan 3.9 menunjukkan *cluster* 1 merupakan *cluster* dengan Tingkat Pengangguran Terbuka yang kecil, sedangkan pada gambar 7b, *cluster* 2 memilki nilai *centroid* 7.16, 7.28 dan 6.88 menunjukkan *cluster* 2 merupakan cluster dengan tingkat pengangguran yang besar. Data TPT2 juga menunjukkan *cluster* 1 merupakan cluster dengan nilai Tingkat Pengagguran Terbuka yang kecil sedangkan *cluster* 2 merupakan Tingkat pengguran yang besar.

```
51 ```{r}
52 an.cluster$centers
53 ```

  A2016    A2017    A2018
1  3.85 4.061304 3.907826
2  7.16 7.278182 6.880000
```

(a)

```
92 ```{r}
93 am.cluster$centers
94 ```

     A2019    A2020    A2021
1 3.826957 4.946087 4.509130
2 6.563636 8.307273 7.547273
```

(b)

Sumber: Penelitian (2022)

Gambar 7 nilai *centroid* data TPT 1 dan data TPT2

Dari nilai *centroid* dapat diketahui dari data pada *cluster* 2, tahun 2020 merupakan tingkat pengangguran terbuka tertinggi yaitu 8.3. Keadaan ini berkaitan dengan masa pandemi yang melanda Indonesia menyebabkan banyak warga negara yang kehilangan pekerjaan. Di tahun 2021 tingkat penganguran masih tinggi, diindikasikan pandemi covid masih mempengaruhi angka tingkat pengangguran terbuka di Indonesia. Nilai *centroid cluster* 1 di tahun 2020 dan 2021 masing-masing 4.9 dan 4.5 lebih besar dibandingkan nilai *centroid cluster* 1 di tahun 2017 dan 2018 masing-masing 4 dan 3.9, hal ini juga menunjukkan angka pengangguran naik di seluruh provinsi di Indonesia selama masa pendemi ini.

Hasil *cluster* data setelah dipisah menjadi dua bagian menunjukkan terdapat 23 provinsi yang memiliki nilai TPT rendah dan 11 provinsi memiliki nilai TPT tinggi. Hasil penelitian ini berbeda dengan penelitian (Tanjung et al., 2021) sebelumnya, menyatakan 21 provinsi berada pada *cluster* dengan nilai TPT rendah sedangkan 13 provinsi berada pada kaster dengan nilai TPT tinggi. Perbedaan ini terjadi karena pada penlitian ini data dibagi menjadi 2 bagian dan data TPT yang digunakan memiliki rentang waktu (2016-2021) sedangkan penelitian (Tanjung et al., 2021) sebelumnya data tidak dibagi menjadi 2 bagian dan data TPT yang digunakan lebih banyak (2014-2019).

## KESIMPULAN

Penelitian ini menerapkan *data mining* menggunakan metode *clustering* untuk menganalisa perubahan *cluster* TPT 2016-2018 dan TPT 2019-2021 di Indonesia. Dengan melakukan penambangan data pada dua data TPT dalam rentang waktu yang berbeda dapat digali lebih banyak informasi. Selain mendapatkan jumlah provinsi pada cluster tingkat pengagguran rendah dan cluster tinggi pada dua data TPT berbeda, penambangan data ini menghasilkan pengetahuan baru yaitu perpindahan dua anggota cluster atau provinsi yang naik cluster dan turun cluster. Provinsi Riau pada tahun 2016-2018 masuk cluster 2, yaitu cluster provinsi dengan tingkat pengagguran yang tinggi, pada tahun 2019-2021 provinsi Riau naik peringkat menjadi provinsi yang memiliki tingkat pengguran yang rendah. Sebaliknya tahun 2019-2021 Sumatera Barat turun dari *cluster* 1 sebagai provinsi yang memiliki angka tingkat pengguran rendah menjadi provinsi yang masuk kategori provinsi dengan angka pengangguran yang tinggi. Temuan lainnya adalah tingginya *centroid* pada data TPT 2 (2019-2021) menandakan tingginya tingkat pengguran di seluruh provinsi di Indonesia. Tahun 2020 sampai 2022 ini adalah masa pandemi covid-19. Temuan ini akan diteliti lebih lanjut dengan melihat perbadingan tingkat pengguran sebelum dan saat pandemi Covid-19.

Provinsi yang konsisten berada pada cluster 2 dengan tingkat pengguran yang tinggi. Pemerintah daerah dapat mengevaluasi dan membuat kebijakan-kebijakan berkaitan dengan penurunan angka pengguran di wilayahnya. Studi lebih lanjut juga dapat dilakukan dengan menambahkan atribut-atribut lain yang berkorelasi untuk menganalisa *cluster* tingkat kemiskinan di masing-masing provinsi seperti studi yang dilakukan oleh Akramunnisa & Fajriani, (2020) dan (Safira et al., 2020). Metode-metode *data mining* lainnya seperti model prediksi dan model klasifikasi juga dapat digunakan untuk mengalisa tingkat pengangguran di masing-masing provinsi. Diharapkan dengan analisa yang lebih beragam pemerintah daerah dapat memprediksi dan mengklasifikasi masalah-masalah baru akibat tingginya tingkat pengganguran di satu wilayah seperti; menurunnya tingkat kesejahteraan, produktivitas dan pendapatan masyarakat.

## REFERENSI

Akramunnisa, A.-, & Fajriani, F. (2020). K-Means Clustering Analysis pada PersebaranTingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan. *Jurnal Varian*, *3*(2), 103–112. https://doi.org/10.30812/varian.v3i2.652

Amrullah, W. A., Istiyani, N., & Muslihatinningsih, F. (2019). Analisis Determinan Tingkat Pengangguran Terbuka di Pulau Jawa Tahun 2007-2016. *E-Journal Ekonomi Bisnis Dan Akuntansi*, *6*(1), 43. https://doi.org/10.19184/ejeba.v6i1.11074

BPS. (2022). *Tingkat Pengangguran Terbuka Menurut Provinsi*. https://www.bps.go.id/indicator/6/543/1/tingkat-pengangguran-terbuka-menurut-provinsi.html

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Johan, K., Marwoto, P. A. N. B., & Pratiwi, D. (2016). *INFLASI DAN INVESTASI TERHADAP PENGANGGURAN DI INDONESIA*. *13*(November), 20–32.

Pradnyana, G. A., & Permana, A. A. J. (2018). Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran. *JUTI: Jurnal Ilmiah Teknologi Informasi.[Online]*, *16*(1), 59–68.

Safira, D., Mustakim, M., Lestari, E. D., Iffa, M., & Annisa, S. (2020). Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *6*(1), 26. https://doi.org/10.24014/rmsi.v6i1.8682

Sardar, T. H., & Ansari, Z. (2018). An analysis of MapReduce efficiency in document clustering using parallel K-means algorithm. *Future Computing and Informatics Journal*, *3*(2), 200–209. https://doi.org/10.1016/j.fcij.2018.03.003

Sembiring, F., Rizqi, S. B., Aziz, M. A., & Firmansyah, D. (2019). *Analisis Pemetaan Tingkat Pengangguran Di Pulau Jawa Dan Bali Dengan Metode K-Means*. *4*(1).

Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi*, *4*(2), 222–231. https://doi.org/10.21831/jpv.v4i2.2547

Tanjung, F. A., Windarto, A. P., & Fauzan, M. (2021). Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, *6*(1), 61. https://doi.org/10.30645/jurasik.v6i1.271

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*, 267–276.

Wardiansyah, M., Yulmardi, & Bahri, Z. (2016). Analisis Faktor-Faktor Yang Mempengaruhi Tingkat pengangguran (Studi kasus provinsi-provinsi Se-Sumatra). *E-Jurnal Ekonomi Sumberdaya Dan Lingkungan Vol.*, *5*(1), 13–18.