

PENERAPAN METODE K-MEANS CLUSTERING PADA DATA TINGKAT PENGANGGURAN TERBUKA TAHUN 2016-2018 DAN 2019-2021

Sita Muharni¹ dan Sigit Andriyanto²

^{1,2}STMIK Dharma Wacana

Jalan Kenanga No. 3 Mulyojati Metro Barat Kota Metro

e-mail : sitamuharni@dharmawacana.ac.id¹, sigitandriyanto@dharmawacana.ac.id²

ABSTRACT

The problem of unemployment impacts poverty, crime, and inequality in living standards. The government must anticipate this impact through various government policies. Knowledge plays a vital role in supporting decision-making and formulating government policies related to unemployment. Several researchers have been mining data to gain new knowledge from Indonesia's Open Unemployment Rate (TPT) data. The available data needs to be continuously mined to gain new knowledge. This study aims to mine Indonesian TPT data from 2016-to 2021. More specifically, this study looks at changes in the 2016-2018 TPT data cluster compared to the 2019-2021 TPT data cluster. The data mining method is clustering analysis using the k-means algorithm. Research result; based on k-means clustering analysis using TPT data 2016-2018 dan 2019-2021, only Riau province rose to cluster 1 (low TPT), and only West Sumatra province fell to cluster 2 (High TPT).

Keywords— clustering analysis, data mining, k-mean, unemployment

ABSTRAK

Masalah pengangguran berdampak pada kemiskinan, kriminalitas, dan ketimpangan taraf hidup. Pemerintah harus mengantisipasi dampak tersebut melalui berbagai kebijakan pemerintah. Pengetahuan memainkan peran penting dalam mendukung pengambilan keputusan dan perumusan kebijakan pemerintah terkait dengan pengangguran. Beberapa peneliti telah menggali data untuk mendapatkan pengetahuan baru dari data Tingkat Pengangguran Terbuka (TPT) Indonesia. Data yang tersedia perlu terus digali untuk mendapatkan pengetahuan baru. Penelitian ini bertujuan untuk menambang data TPT Indonesia dari tahun 2016 hingga tahun 2021. Secara lebih spesifik penelitian ini melihat perubahan klaster data TPT tahun 2016-2018 dibandingkan dengan klaster data TPT tahun 2019-2021. Metode analisis clustering dalam penelitian ini menggunakan algoritma k-means. Hasil penelitian; berdasarkan analisis k-means clustering terhadap data TPT 2016-2018 dan 2019-2021, hanya Provinsi Riau yang naik ke cluster 1 (TPT rendah), dan hanya provinsi Sumatera Barat yang turun ke cluster 2 (TPT Tinggi).

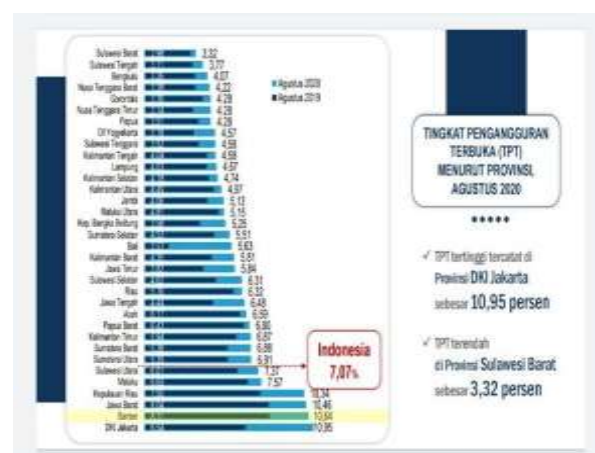
Kata Kunci— Analisa Clustering, Data Mining, K-Means, Pengangguran

I. 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 persentase jumlah tenaga kerja yang tidak diimbangi dengan pertumbuhan kesempatan kerja [1], 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 [2]. Tingginya angka pengangguran di suatu negara menandakan terjadi ketimpangan dalam neraca ketenagakerjaan. 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 [3]. 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.



Gambar 1. Persentase Tingkat Pengangguran Terbuka (TPT) Di Indonesia
(Sumber: *Badan Pusat Statistik (BPS,2020)*)

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 penganggutan terbuka tertinggi dan terendah di seluruh provinsi di Indonesia.

Gambar 1 diatas menjelaskan bahwa TPT tertinggi 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 diatas memperlihatkan perbedaan TPT di masing-masing provinsi dalam rentang satu tahun bergerak sangat dinamis. Angka TPT provinsi tertinggi dan terendah berbeda tiap tahunnya.

Data TPT tidak hanya berisi informasi angka-angka yang menunjukkan tingkat pengangguran 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 penambangun ini adalah informasi baru yang belum ditemukan sebelumnya. Definisi yang lebih sederhana data mining dikemukakan oleh [4], merupakan proses ekstraksi informasi atau pola yang penting atau menarik dari data yang berasal 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 kompleks ilmu komputer dan statistik. *Clustering* bekerja dengan mengelompokkan titik-titik data dalam dua kelompok atau lebih, dimana titik-titik data didalam kelompok yang sama lebih mirip satu sama lain disbanding dengan kelompok data lainnya [5]. Analisa *cluster* yang sering digunakan dalam penelitian adalah algoritma *k-means*. Keunggulan *k-means* mampu menghasilkan *cluster* yang optimum dengan konvergensi yang cepat [6].

Beberapa penelitian terdahulu menggunakan *k-means clustering* menemukan *insight* dari data TPT di Indonesia diantaranya; klasterisasi tingkat pengangguran di Indonesia berdasarkan data TPT tahun 2014-2019 [7]. Hasil penelitian ini menemukan 13 provinsi berada di *cluster* tertinggi. Sebelumnya penelitian klasterisasi tingkat pengangguran di wilayah pulau Jawa dan Bali menggunakan data TPT 2014-2019 dilakukan oleh [8]. *Cluster* data tingkat pengangguran dalam penelitian ini dibagi menjadi tiga *cluster*. Penelitian lainnya dengan metode yang sama dilakukan di provinsi Sulawesi Selatan oleh [9],

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 pengangguran rendah 21 wilayah kota/kabupaten. Penelitian lain juga menggunakan *k-means clustering* dengan variabel TPT, ditambah dengan variabel lainnya melakukan *clustering* jumlah penduduk Sumatera Barat berdasarkan angkatan kerja[10]

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

Penelitian ini bertujuan menggali lebih dalam data TPT dengan menganalisa lebih dalam hasil penelitian [7]. Penelitian tersebut hanya melakukan analisa cluster pada data TPT 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 kluster TPT terendah dan tertinggi, diharapkan dapat memberikan pengetahuan baru perubahan provinsi yang berhasil naik ke *cluster* yang memiliki TPT rendah atau provinsi yang turun ke *cluster* yang memiliki TPT tinggi

II. METODE PENELITIAN

Data yang digunakan adalah data sekunder Tingkat Pengangguran Terbuka (TPT) bulan agustus 2016-2019 dari 34 Provinsi dari Badan Pusat Statistik [11] Indonesia. Sebelum dilakukan proses penambahan, data dibagi menjadi 2 bagian; Pertama data TPT 2016-2018 dengan atribut A2016, A2017 dan A2018, kedua data TPT 2019-2021 dengan atribut A2019, A2020 dan A2021. Atribut A2016 artinya TPT bulan Agustus tahun 2016.

Sebelum menjalankan teknik *cluster*, terlebih dulu ditentukan jumlah *cluster* optimal. Salah satu cara menentukan jumlah *cluster* optimal menggunakan kriteria statistik *Within Sum of Square (WSS)*. WSS adalah salah satu kriteria untuk menghitung keragaman data dalam *cluster* yang terbentuk [12]. 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. Nilai WSS dalam penelitian di cari menggunakan aplikasi R Studio versi 2022.02. Klasterisasi data menggunakan algoritma *k-means* dalam paket clustering pada aplikasi R Studio.

Algoritma *k-means clustering* merupakan salah satu dari metode pengelompokkan data non-hirarki. Algoritma *k-means* mengelompokkan objek dengan terlebih dahulu mengidentifikasi 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* 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 1:

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

Dimana $D_1(X_1, X_2)$ adalah jarak diantara data ke-*i* dan data ke-*j*, X_{2j} adalah koordinat 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 2:

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

V_{ij} 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.

III. 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

Struktur data set TPT1 yang lebih rinci dapat dilihat 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 2), atribut prov 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 4.31 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 ...
```

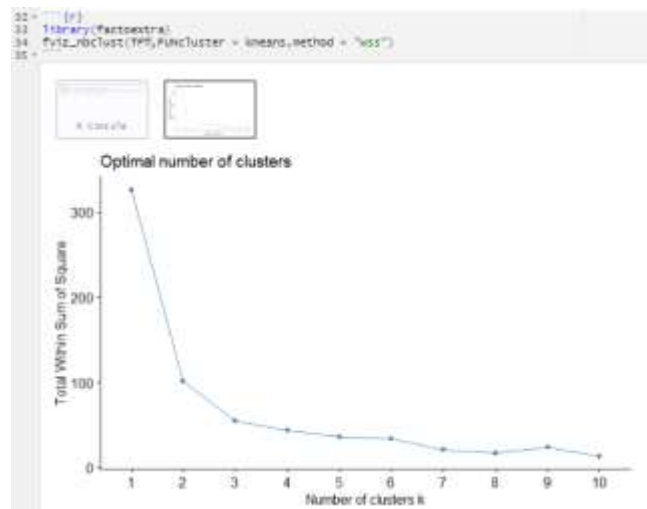
Gambar 2. Dataset TPT1

2. Penentuan Jumlah Cluster

Kriteria WSS dipakai untuk menentukan jumlah 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 3).



Gambar 3. Jumlah kluster optimal 2

3. Penerapan *k-means clustering*

K-means clustering diterapkan menggunakan *library(cluster)* bawaan dari 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 4 merupakan hasil penerapan *k-means clustering* data TPT 1. Tanda panah merupakan posisi data provinsi Sumatera Barat.


```

45 an.cluster$cluster
46 [1] 2 2 1 2 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2 1
47 {r}
48 table(an.cluster$cluster)
49 
    1  2
   23 11

```

Gambar 4 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 clustering* sama dengan tahap sebelumnya. Hasil penerapan algoritma *k-means clustering* dapat dilihat di gambar 5. Hasil *k-means clustering* 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 clustering* 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$cluster
85 [1] 2 2 2 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2 1
86 {r}
87 table(am.cluster$cluster)
88 
    1  2
   23 11

```

Gambar 5 Hasil klasterisasi data TPT2

4. Interpretasi cluster yang terbentuk

Selanjutnya dilakukan interpretasi cluster yang terbentuk pada dua data TPT berdasarkan nilai *centroid*. Pada

gambar 5a adalah nilai centroid data TPT1 dan TPT2. Nilai centroid pada cluster 1 data TPT 1 pada tahun 2016, 2017 dan 2018 sebesar 3.85, 4.06 dan 3.9

menunjukkan *cluster* 1 merupakan *cluster* dengan Tingkat Pengangguran Terbuka yang kecil, sedangkan pada gambar 5b, *cluster* 2 memiliki 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 Pengangguran Terbuka yang kecil sedangkan *cluster* 2 merupakan Tingkat pengangguran 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)

Gambar 5. 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 pengangguran 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 pandemi 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 sedikit berbeda dengan penelitian [7] sebelumnya, menyatakan 21 provinsi berada pada *cluster* dengan nilai TPT rendah sedangkan 13 provinsi berada pada *cluster* dengan nilai TPT tinggi. Perbedaan ini terjadi karena pada penelitian ini data dibagi menjadi 2 bagian dan data TPT yang digunakan memiliki rentang waktu (2016-2021) sedangkan penelitian [7] sebelumnya data tidak dibagi menjadi 2 bagian dan data TPT yang digunakan lebih banyak (2014-2019).

5. SIMPULAN

Penelitian ini menerapkan data mining menggunakan metode *clustering* untuk menganalisa perubahan *cluster* TPT 2016-2018 dan 2019-2021 di Indonesia. Dari hasil penelitian didapatkan pengetahuan baru hanya satu provinsi yang naik *cluster* 1 yaitu Provinsi Riau, Pada tahun 2016-2018 provinsi ini masuk cluster provinsi dengan tingkat pengangguran yang tinggi, pada tahun 2019-2021 provinsi Riau naik peringkat menjadi provinsi yang memiliki tingkat pengangguran yang rendah. Sebaliknya Provinsi Sumatera Barat turun dari *cluster* 1 sebagai provinsi yang memiliki angka tingkat pengangguran rendah menjadi provinsi yang masuk kategori provinsi dengan angka pengangguran yang tinggi pada tahun 2019-2021. Analisa cluster dengan membagi data TPT dalam dua kelompok data menghasilkan pengetahuan baru yaitu provinsi yang naik cluster dan provinsi yang turun cluster.

Studi lebih lanjut dapat dilakukan pemerintahan di tingkat provinsi mengevaluasi dan membuat kebijakan-kebijakan berkaitan dengan penurunan angka pengangguran 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

[9] dan [10]. Metode-metode *data mining* lainnya seperti model prediksi dan model klasifikasi juga dapat digunakan untuk menganalisa tingkat pengangguran di masing-masing provinsi. Diharapkan dengan analisa yang lebih beragam pemerintah daerah dapat memprediksi dan mengklasifikasi masalah-masalah baru akibat tingginya tingkat pengangguran di satu wilayah seperti; menurunnya tingkat kesejahteraan, produktivitas dan pendapatan masyarakat.

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] K. Johan, P. A. N. B. Marwoto, and D. Pratiwi, "Analisis Pengaruh Pertumbuhan Ekonomi, Inflasi Dan Investasi Terhadap Pengangguran Di Indonesia," *Ilm. Progresif Manaj. Bisnis*, vol. 13, no. 2, pp. 20–32, 2016.
- [3] 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.
- [4] 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.
- [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] 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.
- [7] 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.
- [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] 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.
- [10] 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.
- [11] BPS, “Tingkat Pengangguran Terbuka Menurut Provinsi,” 2022.
- [12] S. Muharni, S. Andriyanto, and D. Naista, “Implementasi Dempster Shafer Untuk Mendiagnosa Gangguan Kehamilan Pada Ibu,” *J. Inform.*, vol. 21, no. 2, pp. 146–160, 2021, doi: 10.30873/ji.v21i2.3004.