

Penerapan Metode *K-Means* Pada Pengelompokan Pengangguran Di Indonesia

Fadhillah Azmi Tanjung¹, Agus Perdana Windarto², M Fauzan³

¹Mahasiswa Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar

^{2,3}STIKOM Tunas Bangsa, Pematangsiantar

¹fazmitanjung@gmail.com, ²agus.perdana@amiktunasbangsa.ac.id,

³m.fauzan@amiktunasbangsa.ac.id.

Abstract

Unemployment is a group of labor force who has not done an activity that generates money. Someone who is said to be unemployed can also be categorized as people who have not worked, people who are looking for work, or people who have worked but have not gotten productive results. The purpose of this study is to analyze the unemployment stay by province in Indonesia. This research data is sourced from the Central Statistics Agency in 2014 - 2019. This study uses data mining techniques, namely the K-means algorithm, the K-means method is a clustering method that functions to break the dataset into groups. The K-means method can be used for percentage unemployment data by province. Data will be divided or grouped into 2 Clusters, where Cluster 1 is the group of provinces with the highest potential for unemployment with the results of 13 provinces and Cluster 2 is the province with the lowest potential unemployment results which is 21 provinces. The results of this study are as a way to assist the government in expanding employment to develop and improve the economy in each province in Indonesia. It is hoped that this research can provide input to the government. In particular, the provinces with minimal employment opportunities in Indonesia have an impact on unemployment.

Keywords: Data Mining, Algoritma K-means, Pengangguran, RapidMiner, Clustering

Abstrak

Pengangguran adalah sebuah golongan angkatan kerja yang belum melakukan suatu kegiatan yang menghasilkan uang, Seseorang yang di katakan pengangguran juga dapat dikategorikan sebagai orang yang belum bekerja, orang yang sedang mencari pekerjaan, atau orang yang sudah bekerja namun belum mendapatkan hasil yang produktif. Tujuan dari penelitian ini adalah untuk menganalisa tinggal pengangguran menurut Provinsi yang ada di Indonesia. Data Penelitian ini bersumber dari Badan Pusat Statistika Tahun 2014 - 2019. Penelitian ini menggunakan teknik Data mining yakni Algoritma K-means, Metode K-means merupakan Metode Clustering yang berfungsi untuk memecahkan dataset menjadi berkelompok. Metode K-means dapat digunakan untuk data persentase pengangguran berdasarkan provinsi. Data akan dibagi atau dikelompokkan menjadi 2 Cluster, dimana Cluster 1 adalah kelompok provinsi dengan potensi tertinggi untuk pengangguran dengan hasil 13 provinsi dan Cluster 2 adalah provinsi dengan hasil pengangguran potensi terendah yaitu 21 provinsi. Hasil dari penelitian ini adalah sebagai cara untuk membantu pemerintah dalam memperluas lapangan pekerjaan untuk mengembangkan dan meningkatkan ekonomi di setiap provinsi di Indonesia. Diharapkan penelitian ini dapat memberikan masukan terhadap pemerintah. Khususnya provinsi – provinsi yang minim dalam lapangan pekerjaan di Indonesia yang berdampak pada pengangguran.

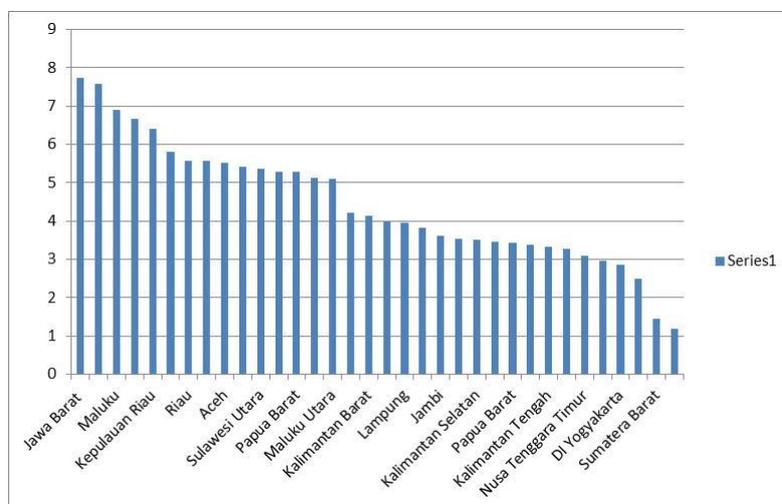
Kata kunci: Data Mining, K-means Algorithm, Unemployment, RapidMiner, Clustering

1. PENDAHULUAN

Pengangguran merupakan suatu situasi kondisi dimana seseorang yang termasuk dalam golongan angkatan kerja itu bisa di sebut juga dengan seorang yang mempunyai keinginan untuk mendapatkan sebuah pekerjaan akan tetapi sama sekali belum bisa mencapainya, “Pengangguran ialah keadaan seseorang

yang dimana orang tersebut ingin bekerja namun sama sekali tidak mendapatkan pekerjaan yang sesuai dengan bidangnya” [1]. “Pengangguran menjadi perbincangan yang sangat serius dan masalah utama yang dihadapi oleh Negara berkembang, termasuk Indonesia dalam hal minimnya ketenagakerjaan” [2]. Menurut [3] “Banyaknya jumlah penduduk di Indonesia dan minimnya lapangan pekerjaan yang ada menimbulkan permasalahan pengangguran di Indonesia menjadi sebuah permasalahan yang sangat sulit untuk diatasi oleh pemerintah, tidak cepatnya penanganan dari pemerintah dalam mempertegas masalah ini, dan membuat perekonomian di Indonesia semakin tidak bisa untuk di prediksi atau bisa dikatakan sangat terpuruk”. Tingkat Pengangguran Terbuka (TPT) mengalami penurunan sejak tahun 2016. Adapun Negara-Negara dengan pengangguran terbanyak di dunia seperti Lesetho, Palestina, Venezuela, Namibia, Yaman, Kepulauan Marshall, Kiribati, Djibouti, Senegal, Suriah, dan Indonesia. Dalam hal ini pemerintah harus berperan aktif dalam kepada penduduknya untuk mengurangi tingkat pengangguran yang semakin meningkat. “Menurut [4] di Negara Indonesia yang masuk kategori pengangguran sebesar 7,64 persend an pekerja paruh waktu sebesar 23,83 persend dalam setahun terakhir. orang yang berpendidikan atau lulusan universitas ada 11,71 juta yang tidak mempunyai pekerjaan, Indonesia juga termasuk dalam pengangguran Terbesar di dunia”.

Berikut ini adalah grafik tingkat pengangguran di seluruh provinsi Indonesia dapat dilihat pada gambar 1.



Gambar 1. Presentase Tingkat Pengangguran Di Indonesia
(Sumber : Badan Pusat Statistik (BPS), 2016)

Dari gambar 1. memperlihatkan bahwa hampir seluruh provinsi di Indonesia memiliki tingkat pengangguran tinggi. Provinsi Jawa Barat berada di provinsi tertinggi dengan tingkat pengangguran sebanyak 7,73% dan di Sumatera Barat berada di posisi terendah dengan Tingkat Pengangguran sebanyak 1,45%, Hal ini dikarenakan pertumbuhan penduduk yang semakin besar, sehingga angka pengangguran tinggi, minim nya lapangan pekerjaan terbuka, persaingan yang ketat dalam mencari pekerjaan. Pengangguran dapat menyebabkan dampak yang

sangat negatif bagi Indonesia dan mengakibatkan banyaknya tingkat kejahatan dan tidak mampu untuk bersaing dengan Negara lain jika Indonesia tidak dapat memperluas lapangan pekerjaan terbuka.

Berdasarkan uraian diatas, banyak cabang kecerdasan buatan dalam ilmu komputer yang dapat menyelesaikan suatu permasalahan tersebut. secara kompleks diantaranya sistem pendukung keputusan, sistem pakar, data mining dan lain sebagainya. Beberapa penelitian terdahulu tentang pengangguran: oleh Sirait & Marhaeni, 2013) “Berdasarkan koefisien β (beta) masing masing variable bebas yang didapat dari hasil pengolahan data dengan menggunakan aplikasi SPSS 17 for windows yang memiliki kontribusi besar terhadap variabel terikatnya adalah pertumbuhan ekonomi, yaitu sebesar 0,663% atau sebesar 66,3% sehingga pertumbuhan ekonomi memiliki pengaruh yang besar terhadap jumlah pengangguran di provinsi bali”. Menurut [6] “Berdasarkan perkembangan tingkat pengangguran tertinggi dialami Provinsi Bengkulu sebesar 5,86%, diikuti oleh Provinsi Aceh 2,3%, dan Provinsi Jambi sebesar 2.0%. dan yang terendah adalah Provinsi Sumsel sebesar -9.0%”. Dari penelitian terdahulu kelahmahannya adalah mereka hanya meneliti 1 Provinsi atau Sumatera saja. Disini penulis memakai sebuah “metode K-Means *clustering* yang artinya K-Means *clustering* merupakan yang digunakan untuk pengelompokkan secara partisi yang menggunakan data yang berbeda-beda algoritma K-means *clustering* proses penggunaannya tergantung pada data dan hasil akhirnya” [7]. “Kelebihan metode ini adalah mudah untuk diimplementasikan dan dijalankan, waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat, mudah untuk diadaptasi, umum digunakan”. Diharapkan Penelitian ini dapat memberikan informasi kepada pemerintah tentang data Pengelompokan Tingkat Pengangguran di Indonesia yang berdampak pada menurunnya perekonomian Negara, berdampak pada ketidakstabilan politik, ataupun dapat dilihat dari segi mental masyarakat di Indonesia..

2. METODOLOGI PENELITIAN

2.1. Data Mining

Menurut [8] menjelaskan bahwa “Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan trend dengan menambang sejumlah *repository* data dalam jumlah besar, dan menggunakan teknologi pengenalan pola seperti statistic dan teknik matematika”. Data mining disebut juga dengan *Knowledge discovery in database* (KKD) ataupun pengenalan pola. Data mining dapat dibagi menjadi empat kelompok, yaitu model prediksi (*prediction modelling*), analisis kelompok (*Cluster analysis*), analisis asosiasi (*association analysis*) dan deteksi anomaly (*anomaly detection*).

2.2. Clustering

Menurut [9] “*Clustering* merupakan suatu pengelompokkan data mining yang bertujuan untuk melakukan sebuah klasifikasi ataupun memprediksi nilai dari variabel target yang mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi berkelompok-kelompok yang memiliki kemiripan karakteristik



sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen dari pada objek yang berada pada kelompok yang berbeda”. Setiap himpunan adalah *cluster* Setiap himpunan bagian adalah *cluster*, sehingga objek yang didalam *cluster* mirip satu sama dengan yang lainnya dan memiliki perbedaan dengan objek dari *cluster* yang lain.

2.3. Metode K-Means

Menurut [10] “metode k-means merupakan metode pengelompokan yang berkaitan dengan metode k-medoids yang melakukan partisi set sebuah data ke sejumlah *cluster* yang sudah ditetapkan langsung di awal”. Metode k-means sangat sederhana untuk di jalankan dan diterapkan, cukup cepat, mudah di pergunakan, dan mudah beradaptasi umum digunakan dalam berbagai aplikasi kecil hingga menengah. Secara ilmiah, k-means menjadi salah satu metode yang paling penting dalam penggunaan data mining. Metode k-means *clustering* merupakan salah satu dari pengelompokan metode nonhirarki yang tujuannya mengelompokkan objek yang diawali dengan mengidentifikasi data yang akan di *cluster*. Pada pertama kali iterasi, titik utama setiap cluster yang ditetapkan secara bebas. Lalu dihitung antar jarak data dengan tiap titik utama pada *cluster*. Dan untuk melakukan perhitungan jarak data ke-*i* pada pusat cluster ke-*j* yang diberi nama ($D(1_2)$).

Langkah – langkah metode *k-means clustering* diantaranya:

- 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 $D_i(x_1, x_2)$ 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 :

$$V_{ij} = \frac{\sum_{k=1}^{N_i} X_{kj}}{N_i} \quad (2.2)$$

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)
- e) Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Data

Proses analisis data dapat dilakukan setelah adanya pengumpulan data. Dalam melakukan penelitian ini, penulis melakukan analisis data statistik deskriptif. Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang diperoleh tidak dari sumbernya langsung, melainkan sudah dikumpulkan oleh pihak lain dan sudah diolah serta memiliki keterkaitan dengan permasalahan yang diteliti.

Tabel 1. Data Persentase Pengangguran Menurut Provinsi
(Sumber: Badan Pusat Statistik)

No	Provinsi	Tahun					
		2014	2015	2016	2017	2018	2019
1.	Aceh	7,88	8,83	7,85	6,98	6,45	5,53
2.	Sumatera Utara	6,09	6,55	6,17	6,00	5,57	5,56
3.	Sumatera Barat	6,41	6,44	5,45	5,69	5,55	5,29
4.	Riau	5,78	7,27	6,69	5,99	5,96	5,57
5.	Jambi	3,79	3,54	4,33	3,77	3,76	3,62
6.	Sumatera Selatan	4,40	5,55	4,12	4,09	4,12	3,99
7.	Bengkulu	2,55	4,06	3,57	3,28	3,11	2,50
8.	Lampung	4,93	4,29	4,58	4,38	4,20	3,96
9.	Kepulauan Bangka Belitung	3,91	4,82	4,38	4,12	3,63	3,39
...
34.	Papua	3,46	5,71	3,16	3,79	3,05	3,42

Tabel 1. menunjukkan data yang telah diperoleh dari Badan Pusat Statistik Nasional mengenai data Pengangguran yang terdiri dari 34 provinsi dari tahun 2014-2019. Pada tahapan ini pengolahan data pengangguran menurut provinsi di Indonesia menggunakan metode *K-means*. Berikut adalah data pengangguran menurut provinsi di Indonesia yang diambil penulis melalui situs Badan Pusat Statistik Nasional. Proses analisa data dapat dilakukan setelah adanya pengumpulan data yang bersifat valid. Dalam melakukan penelitian ini, penulis melakukan analisa data statistik yang bersifat sekunder yang dimana data di peroleh tidak dari sumbernya langsung, melainkan sudah dikumpulkan dan diolah secara terinci yang dimana data berkaitan dengan permasalahan yang sedang diteliti.

3.2. Pengolahan Data Manual *K-means*

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut adalah uraian perhitungan manual proses *clustering* hasil data pengangguran dengan menggunakan sebuah metode *K-means*. Proses *clustering* dimulai dengan menentukan sebuah data yang akan di *cluster*. Dalam hal ini variabel data yang akan digunakan adalah dari tahun 2014 – 2019, dan dari penelitian ini penulis



mengambil nilai rata - rata dari setiap provinsi mulai dari tahun 2014 - 2019 sebagai data yang akan ditentukan nilai *cluster* nya. Berikut adalah langkah - langkah penyelesaian yang akan dilakukan penulis dalam mengelompokkan hasil data pengangguran berdasarkan provinsi dengan menggunakan metode *K-means*:

a) Menentukan Jumlah Data Yang Akan di *Cluster*

Dimana sampel data yang digunakan dalam proses *clustering* adalah data yang berasal dari Badan Pusat Statistik (BPS) pada tahun 2014 dan 2019 (6 tahun) dengan jumlah data sebanyak 34 provinsi. Untuk melakukan proses perhitungan terutama kita mencari nilai rata-rata dari data yang akan diolah, berikut cara untuk memperoleh rata - rata:

$$R1 = 7,88 + 8,83 + 7,85 + 6,98 + 6,45 + 5,53 / 6 = \mathbf{7,25}$$

$$R2 = 6,09 + 6,55 + 6,17 + 6,00 + 5,57 + 5,56 / 6 = \mathbf{5,99}$$

$$R3 = 6,41 + 6,44 + 5,45 + 5,69 + 5,55 + 5,29 / 6 = \mathbf{5,81}$$

$$R4 = 5,78 + 7,27 + 6,69 + 5,99 + 5,96 + 5,57 / 6 = \mathbf{6,21}$$

$$R5 = 3,79 + 3,54 + 4,33 + 3,77 + 3,76 + 3,62 / 6 = \mathbf{3,80}$$

Lakukan proses berikut sampai dengan R34.
Berikut hasil nilai rata-rata yang telah di dapat:

Tabel 2. Nilai Hasil Rata - Rata Data Pengangguran Menurut Provinsi

No	Provinsi	Rata Rata
1	Aceh	7,25
2	Sumatera Utara	5,99
3	Sumatera Barat	5,81
4	Riau	6,21
5	Jambi	3,80
...
34	Papua	3,77

Tabel 3. Parameter

Parameter	Nilai	Keterangan
<i>Cluster</i>	2	(Max/Min)
<i>Data</i>	34	Provinsi
<i>Atribut</i>	6	(Tahun)

- b) Menetapkan nilai k jumlah *cluster* pengangguran sebanyak 2 *cluster* (k-2). *Cluster* yang dibentuk yaitu *cluster* tertinggi dan *cluster* terendah.
- c) Menentukan nilai *centroid* (pusat *cluster*) awal yang telah ditentukan secara random berdasarkan nilai rata-rata data yang telah didapatkan. *Cluster* tertinggi (C1) diperoleh dari nilai maximum, *cluster* terendah (C2) diperoleh dari nilai minimum.
Berikut adalah nilai *centroid* data awal pada iterasi 1 :



Tabel 4. *Centroid* Data Awal (Iterasi1)

<i>Cluster</i>	Hasil
C1 (Maximum)	8,54
C2 (Minimum)	1,50

d) Menghitung Jarak Setiap Data Terhadap *Centroid* (pusat *cluster*)

Setelah data nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster* dengan menggunakan rumus yang perhitungannya dapat kita lihat sebagai berikut:

Dilakukan perhitungan jarak terhadap data pengangguran dengan titik pusat (*centroid*) pada *cluter* pertama:

$$D_{(1.1)} = \sqrt{(8,54 - 7,25)^2} = 1,28$$

$$D_{(1.2)} = \sqrt{(8,54 - 5,99)^2} = 2,55$$

$$D_{(1.3)} = \sqrt{(8,54 - 5,81)^2} = 2,73$$

$$D_{(1.4)} = \sqrt{(8,54 - 6,21)^2} = 2,33$$

$$D_{(1.5)} = \sqrt{(8,54 - 3,80)^2} = 4,73$$

Lakukan proses tersebut sampai dengan D (1.34)

Selanjutnya perhitungan jarak pengangguran dengan *centroid* dengan *cluster* kedua:

$$D_{(2.1)} = \sqrt{(1,50 - 7,25)^2} = 5,75$$

$$D_{(2.2)} = \sqrt{(1,50 - 5,99)^2} = 4,49$$

$$D_{(2.3)} = \sqrt{(1,50 - 5,81)^2} = 4,31$$

$$D_{(2.4)} = \sqrt{(1,50 - 6,21)^2} = 4,71$$

$$D_{(2.5)} = \sqrt{(1,50 - 3,80)^2} = 2,30$$

Lakukan proses tersebut sampai dengan D (2.34)

Berikut tabel 5 adalah hasil dari Perhitungan jarak data dengan titik pusat *cluster* pada iterasi 1 menggunakan *Eulidean Distence*:

Tabel 5. Jarak *Centroid* (Iterasi 1)

C1	C2	Jarak Terpendek
1,28	5,75	1,28
2,55	4,49	2,55
2,73	4,31	2,73
2,33	4,71	2,33
4,73	2,30	2,30
...
4,77	2,27	2,27



- e) Menentukan Posisi *Cluster* atau Pengelompokan
Dalam menentukan posisi *cluster* masing-masing data Pengangguran berdasarkan jarak minimum data terhadap pusat *cluster*. Data yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan tabel 6 posisi data dengan tiap *cluster* pada iterasi 1 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada *cluster*:

Tabel 6. *Cluster* (Iterasi 1)

C1	C2
1	
1	
1	
1	
	1
...	...
	1
12	22

Berikut ini adalah hasil perhitungan jarak data pada setiap titik pusat *cluster* pada iterasi 1:

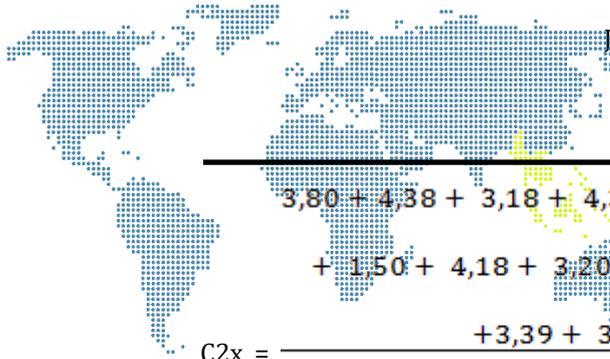
Tabel 7. Hasil *Cluster* (Iterasi 1)

<i>Cluster</i>	Hasil
C1 (Maximum)	12
C2 (Minimum)	22

Selanjutnya dalam metode *K-means*, perhitungan berhenti apabila *cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *cluster* pada iterasi selanjutnya sampai nilai iterasinya sama.

- f) Lalu Lakukan dengan menghitung *centroid* baru atau titik pusat baru menggunakan hasil dari setiap anggota pada masing – masing *cluster*. Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru pada iterasi 1 dengan menjumlahkan nilai sesuai yang tertera pada *cluster* yang terdapat pada tabel 6. Adapun *centroid* baru untuk mencari *cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membagikannya sebanyak jumlah nilai. Berikut contoh perhitungan titik pusat *cluster* baru pada iterasi ke dua (x) :

$$C1x = \frac{7,25 + 5,99 + 5,81 + 6,21 + 6,99 + 6,68 + 8,35 + 7,50 + 6,96 + 5,37 + 7,77 + 5,93}{12} = 6,73$$



$$C2x = \frac{3,80 + 4,38 + 3,18 + 4,39 + 4,04 + 4,68 + 3,10 + 4,08 + 8,54 + 1,50 + 4,18 + 3,20 + 4,22 + 3,61 + 4,19 + 4,40 + 3,41 + 3,39 + 3,62 + 2,46 + 4,98 + 3,77}{22} = 3,96$$

g) Selanjutnya dilakukan kembali langkah ke 4 sampai 6. Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data pengangguran tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data masih berubah maka proses iterasi berlanjut pada iterasi berikutnya. Berikut tabel hasil *cluster* iterasi 2 :

Tabel 8. Centroid Baru (Iterasi2)

<i>Cluster</i>	Hasil
C1 (Maximum)	6,73
C2 (Minimum)	3,96

Setelah *cluster* iterasi dua didapat. Selanjutnya melakukan sebuah perhitungan nilai *centroid* baru dengan nilai rata-rata data sebagai berikut:

$$D_{(1.1)} = \sqrt{(6,73 - 7,25)^2} = 0,52$$

$$D_{(1.2)} = \sqrt{(6,73 - 5,99)^2} = 0,74$$

$$D_{(1.3)} = \sqrt{(6,73 - 5,81)^2} = 0,93$$

$$D_{(1.4)} = \sqrt{(6,73 - 6,21)^2} = 0,52$$

$$D_{(1.5)} = \sqrt{(6,73 - 3,80)^2} = 2,93$$

Lakukan proses tersebut sampai dengan D (1.34) Selanjutnya perhitungan jarak pengangguran dengan *centroidcluster* kedua:

$$D_{(2.1)} = \sqrt{(3,96 - 7,25)^2} = 3,30$$

$$D_{(2.2)} = \sqrt{(3,96 - 5,99)^2} = 2,03$$

$$D_{(2.3)} = \sqrt{(3,96 - 5,81)^2} = 1,85$$

$$D_{(2.4)} = \sqrt{(3,96 - 6,21)^2} = 2,25$$

$$D_{(2.5)} = \sqrt{(3,96 - 3,80)^2} = 0,16$$

Lakukan proses tersebut sampai dengan D (2.34)

Berikut tabel 9 adalah hasil dari Perhitungan jarak data dengan titik pusat *cluster* pada iterasi 2 menggunakan *Eulidean Distence*:



Tabel 9. Jarak *Centroid Cluster* (Iterasi 2)

C1	C2	Jarak Terpendek
0,52	3,30	0,52
0,74	2,03	0,74
0,93	1,85	0,93
0,52	2,25	0,52
2,93	0,16	0,16
...
2,97	0,19	0,19

Data yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan tabel 10 posisi data dengan tiap *cluster* pada iterasi 2 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada *cluster*. Selanjutnya menghitung jarak data dengan masing-masing *centroid*:

Tabel 10. *Cluster* (Iterasi 2)

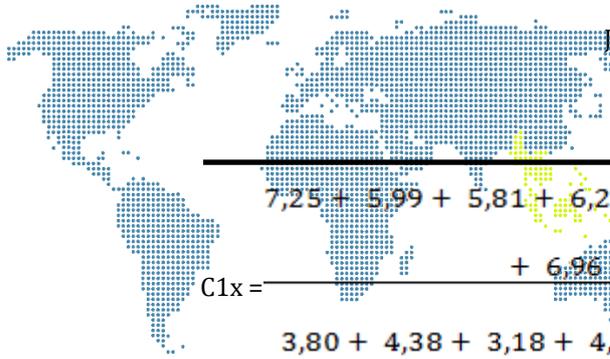
C1	C2
1	
1	
1	
1	
	1
...	...
	1
13	21

Berikut ini adalah hasil perhitungan jarak data pada setiap titik pusat *cluster* pada iterasi 2:

Tabel 11. Hasil *Cluster* (Iterasi 2)

<i>Cluster</i>	Hasil
C1 (Maximum)	13
C2 (Minimum)	21

Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru pada iterasi 2 dengan menjumlahkan nilai sesuai yang tertera pada *cluster* yang terdapat pada tabel 10. Selanjutnya menghitung titik pusat baru menggunakan hasil dari setiap anggota pada masing - masing *cluster*. Berikut contoh perhitungan titik pusat *cluster* baru pada iterasi ke dua (x):



$$C1x = \frac{7,25 + 5,99 + 5,81 + 6,21 + 6,99 + 6,68 + 8,35 + 8,54 + 7,50 + 6,96 + 5,37 + 7,77 + 5,93}{13} = 6,87$$

$$C2x = \frac{3,80 + 4,38 + 3,18 + 4,39 + 4,04 + 4,68 + 3,10 + 4,08 + 1,50 + 4,18 + 3,20 + 4,22 + 3,61 + 4,19 + 4,40 + 3,41 + 3,39 + 3,62 + 2,46 + 4,98 + 3,77}{21} = 3,74$$

Maka, *centroid* data baru iterasi 2 adalah sebagai berikut:

Tabel 12. Centroid Baru (Iterasi 3)

<i>Cluster</i>	Hasil
C1 (Maximum)	6,87
C2 (Minimum)	3,74

Setelah *cluster* iterasi 2 didapat, Selanjutnya melakukan sebuah perhitungan nilai *centroid* baru dengan nilai rata-rata data sebagai berikut:

$$D_{(1.1)} = \sqrt{(6,87 - 7,25)^2} = 0,38$$

$$D_{(1.2)} = \sqrt{(6,87 - 5,99)^2} = 0,88$$

$$D_{(1.3)} = \sqrt{(6,87 - 5,81)^2} = 1,07$$

$$D_{(1.4)} = \sqrt{(6,87 - 6,21)^2} = 0,66$$

$$D_{(1.5)} = \sqrt{(6,87 - 3,80)^2} = 3,07$$

Lakukan proses tersebut sampai dengan D (1.34)

Selanjutnya perhitungan jarak pengangguran dengan *centroid cluster* kedua:

$$D_{(2.1)} = \sqrt{(3,74 - 7,25)^2} = 3,51$$

$$D_{(2.2)} = \sqrt{(3,74 - 5,99)^2} = 2,25$$

$$D_{(2.3)} = \sqrt{(3,74 - 5,81)^2} = 2,07$$

$$D_{(2.4)} = \sqrt{(3,74 - 6,21)^2} = 2,47$$

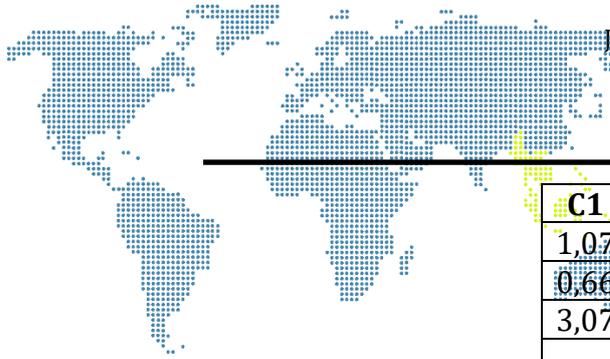
$$D_{(2.5)} = \sqrt{(3,74 - 3,80)^2} = 0,06$$

Lakukan proses tersebut sampai dengan D (2.34)

Berikut tabel 13 adalah hasil perhitungan jarak data dengan titik pusat *cluster* pada iterasi 3 menggunakan *Eulidean Distance*.

Tabel 13. Jarak Centroid (Iterasi 3)

C1	C2	Jarak Terpendek
0,38	3,51	0,38
0,88	2,25	0,88



C1	C2	Jarak Terpendek
1,07	2,07	1,07
0,66	2,47	0,66
3,07	0,06	0,06
...
3,11	0,03	0,03

Data yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan tabel 14 posisi data dengan tiap *cluster* pada iterasi 3 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada *cluster*.

Tabel 14. *Cluster* (Iterasi 3)

C1	C2
1	
1	
1	
1	
	1
...	...
	1
13	21

Berikut ini adalah hasil perhitungan jarak data pada setiap titik pusat *cluster* pada iterasi 3:

Tabel 15. Hasil *Cluster* (Iterasi 3)

<i>Cluster</i>	Hasil
C1 (Maximum)	13
C2 (Minimum)	21

Berikut di dapatkan sebuah hasil akhir, pada iterasi ke 2 dan iterasi ke 3 di dapatkan hasil yang sama kedua iterasi tersebut bernilai C1=13, C2= 21 Sehingga posisi *cluster* pada data tidak mengalami perubahan sehingga proses iterasi berhenti. Berdasarkan posisi *cluster* masing-masing data pengangguran dan nilai *cluster* hasil iterasi 3.

33. Hasil Percobaan

Pada tahap ini penelitian menggunakan perangkat lunak *Rapidminer* 5.3. untuk pengujian data yang telah di analisis. Berikut penjelasan pengujian dengan menggunakan software *rapidminer* 5.3. keterkaitan dari hasil yang didapat antara perhitungan manual algoritma dengan hasil yang ditampilkan oleh *tools rapidminer* dan dapat disimpulkan bahwa hasil perhitungan manual dari metode

K-means dan dari aplikasi *software Microsoft Excel 2010* hasilnya sama dengan dan dapat *input* ke dalam aplikasi *Rapidminer*.

Row No	Provinsi	id	cluster	2014.0	2015.0	2016.0	2017.0	2018.0	2019.0
1	Aceh	1	cluster_0	7.893	8.929	7.852	6.976	6.453	6.534
2	Sumatera Ut. 2		cluster_0	6.088	6.549	6.165	6.000	5.575	5.561
3	Sumatera Bt. 3		cluster_0	6.413	6.440	5.449	5.694	5.551	5.294
4	Riau	4	cluster_0	5.775	7.275	6.686	5.993	5.958	5.573
5	Jambi	5	cluster_1	3.790	3.536	4.331	3.770	3.757	3.623
6	Sumatera Si. 6		cluster_1	4.402	5.553	4.123	4.095	4.123	3.987
7	Bengkulu	7	cluster_1	2.547	4.059	3.568	3.275	3.107	2.503
8	Lampung	8	cluster_1	4.934	4.286	4.562	4.377	4.196	3.959
9	Kepulauan R. 9		cluster_1	3.907	4.620	4.363	4.119	3.531	3.368
10	Kepulauan F. 10		cluster_0	5.976	7.625	8.362	6.797	6.777	6.405
11	DKI Jakarta	11	cluster_0	9.156	7.796	5.943	6.252	5.792	5.132
12	Jawa Barat	12	cluster_0	8.556	8.559	8.728	8.351	8.163	7.727
13	Jawa Tengah	13	cluster_1	5.564	5.150	4.415	4.363	4.369	4.216
14	DI Yogyakarta	14	cluster_1	2.748	4.072	2.767	2.931	3.206	2.858
15	Jawa Timur	15	cluster_1	4.102	4.392	4.177	4.051	3.922	3.827
16	Banten	16	cluster_0	9.471	9.065	8.435	8.514	8.143	7.584
17	Bali	17	cluster_1	1.637	1.679	2.003	1.382	1.114	1.192

Gambar 1. Tampilan Data Perhitungan *Tools Rapidminer*

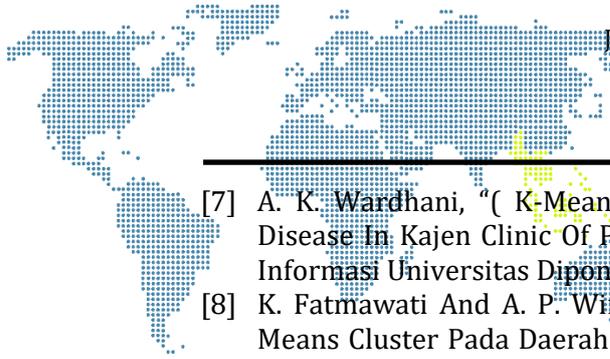
4. SIMPULAN

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa:

- Penerapan data mining dengan metode *k-means clustering* dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah www.bps.go.id dengan subjek data pengangguran berdasarkan provinsi (2014 - 2019). Jumlah *record* yang digunakan sebanyak 34 provinsi dengan menghasilkan 2 *cluster* yakni *cluster* tinggi sebanyak 13 provinsi dan *cluster* rendah sebanyak 21 provinsi.
- Berdasarkan hasil pengujian *k-means* untuk kasus persentase data pengangguran menggunakan *tools Rapidminer 5.3* diperoleh hasil yang sama dengan analisis perhitungan metode dimana diperoleh 13 provinsi dengan *cluster* tinggi yang menjadi pusat perhatian bagi pemerintah untuk memperluas lapangan pekerjaan untuk pengangguran pada provinsi tersebut.

DAFTAR PUSTAKA

- R. Franita, "Analisa Pengangguran Di Indonesia," Vol. 1, 2016.
- D. P. Herniwati And R. Handayani, "Pengangguran Terbuka Di Provinsi Jawa Tengah," *Diponegoro J. Econ.*, Vol. 1, P. 159, 2019.
- Imsar, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka Di Indonesia Periode 1989-2016," *Hum. Falah*, Vol. 5, No. 1, Pp. 144-164, 2018.
- Badan Pusat Statistik, "Keadaan Ketenagakerjaan Indonesia Februari 2018," *Badan Pus. Stat.*, No. 42/05/Th. Xxi, Pp. 1-16, 2018, Doi: No. 74/11/35/Th.Xvi, 5 November 2018.
- N. Sirait And A. A. I. N. Marhaeni, "Analisis Beberapa Faktor Yang Berpengaruh Terhadap Jumlah Pengangguran Kabupaten / Kota Di Provinsi Bali," *E-Jurnal Ekon. Univeritas Udayana*, Vol. 2, No. 2, Pp. 108-118, 2013.
- M. Wardiansyah, Yulmardi, And Z. Bahri, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran (Studi Kasus Provinsi-Provinsi Se-Sumatera)," *E-Jurnal Ekon. Sumberd. Dan Lingkungan*, Vol. 5, No. 1, Pp. 13-18, 2016.



- [7] A. K. Wardhani, "(K-Means Algorithm Implementation For Clustering Of Patients Disease In Kajen Clinic Of Pekalongan) Anindya Khrisna Wardhani Magister Sistem Informasi Universitas Diponegoro," *J. Transform.*, Vol. 14, No. 1, Pp. 30-37, 2016.
- [8] K. Fatmawati And A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, Vol. 3, No. 2, P. 173, 2018, Doi: 10.24114/Cess.V3i2.9661.
- [9] T. Wibowo, "Penerapan Data Mining Pemilihan Siswa Kelas Unggulan Dengan Metode K-Means Clustering Di Smp N 02 Tasikmadu," 2018.
- [10] M. A. W. K. Murti, "Plagiat Merupakan Tindakan Tidak Terpuji Penerapan Metode K-Means Clustering Untuk Mengelompokkan Potensi Produksi Buah – Buah Di Provinsi Daerah Istimewa Yogyakarta," 2017.