Penerapan Data Mining Dalam Mengelompokkan Data Pengangguran Terbuka Menurut Provinsi Menggunakan Algoritma K-Means

Mochamad Wahyudi[1] , Lise Pujiastuti[2], Solikhun[3]
[1] Program Studi S1 Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta,*
[2]Program Studi S1 Sistem Informasi, STMIK Antar Bangsa, Tangerang, Indonesia
[3]Program Studi D3 Manajemen Informatika, AMIK Tunas Bangsa, Pematangsiantar,*
[1]wahyudi@bsi.ac.id, [2] lise.pujiastuti@gmail.com, [3] solikhun@amiktunasbangsa.ac.id*

Abstract

Open unemployment is a workforce that has absolutely no job. This unemployment occurs because the labor force has not gotten a job even though it has been trying optimally or because of laziness to find work or lazy to work. The government must be able to suppress open unemployment so that poverty alleviation programs can be achieved. The government needs accurate information about unemployment grouping data in Indonesia by province. This study aims to classify open unemployment data in Indonesia by province. Data is taken from the Indonesian Central Statistics Agency in the form of open unemployment data by province from 2015 to 2019. This grouping uses the K-Means Clustering method. Data is grouped based on 2 clusters. The results of this study are that there are 12 provinces in the high cluster and 22 provinces in the low cluster

Keywords: *Data Mining, K-Means, Open unemployment.*

1. Introduction

Pengangguran terbuka adalah angkatan kerja yang sama sekali tidak mempunyai pekerjaan. Pengangguran ini terjadi karena angkatan kerja tersebut belum mendapat pekerjaan padahal telah berusaha secara maksimal atau dikarenakan faktor malas mencari pekerjaan atau malas bekerja. Masalah pengangguran adalah masalah yang serius di negara kita. Pemerintah memiliki kewajiban untuk menekan angka pengangguran terbuka untuk mengentaskan kemiskinan di Indonesia. Untuk menangani masalah pengangguran terbuka di Indonesia, pemerintah membutuhkan informasi yang valid mengenai pengangguran terbuka. Belum ada informasi pengelompokkan data pengangguran terbuka berdasarkan provinsi. Berdasarkan permasalahan diatas maka diperlukan sebuah metode untuk mengelompokkan data pengangguran terbuka berdasarkan provinsi sehingga pemerintah dapat mengetahui provinsi mana berada di cluster tinggi tingkat penganggurannya dan provinsi mana yang berada di cluster rendah. Untuk mengelompokkan data pengangguran terbuka di Indonesia berdasarkan provinsi di perlukan sebuah teknik data mining menggunakan metode K-Means Clustering.

2. Metode Penelitian

2.1. *Data Mining*

Data mining merupakan proses iteratif dan interaktif untuk menemukan pola-pola atau model baru yang shahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang[1][2].

ISSN: 2686-0260

2.2. *K-Means*

Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (centroid) terdekat dengan data. Tujuan dari K-Means adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik centroid[3][4]**.**

2.3. *Clustering*

Menurut Baskoro *cluster* atau klusterisasi adalah salah satu alat bantu pada *data mining* yang bertujuan mengelompokkan objek-objek ke dalam *cluster* - *cluster*. *Cluster* adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode *Data mining*, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. *Cluster* berbeda dari klasifikasi karena *cluster* tidak memiliki variabel target. Tujuan *cluster* bukan untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel taget [5]. Metode K-Meansdapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek kedalam klaster tertentu berdasarkan rataan terdekat. Dalam prosedur pembentukan K-Means Cluster terdapat langkah-langkah yang dapat
dilakukan, antara lain:
1. Tentukan banyaknya klaster (k) yang akan dibentuk.
2. Bangkitkan k centroidawal (rata-rata setiap klaster).
3. Hitung jarak antara setiap objek dengan setiap centroid dan masukan objek tersebut ke dalam klaster yang sesuai berdasarkan jarak terdekat.
4. Tentukan centroid dari klaster yang baru.
5. Ulangi langkah 3 dan 4 sampai tidak ada lagi pemindahan objek antarklaster[6].

2.4. *Rapid Miner*

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofman dari *Institute of Technologi Blanchardstown dan Ralf Klinkenberg* dari *rapid-i.com* dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java di bawah lisensi *GNU Public Licence* dan *Rapid Miner* dapat dijalankan di sistem operasi manapun. Dengan menggunakan *Rapid Miner*, tidak dibutuhkan kemampuan koding khusus, karena semua fasilits sudah disediakan. *Rapid Miner* dikhususkan untuk penggunaan *data mining*. Model yang disediakan juga cukup banyak dan lengkap, seperti *Model Bayesian, Modelling, Tree Induction, Neural Network* dan lain-lain [7].

3. Hasil Dan Pembahasan

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses algoritma *k-means clustering* pada data pengangguran terbuka berdasarkan provinsi dengan menggunakan sebuah konsep *data mining*.

1. Menentukan jumlah data yang akan di *cluster*, dimana sampel data pengangguran terbuka yang akan digunakan dalam proses *clustering* yaitu data data pengangguran terbuka tahun 2015 sampai degan tahun 2019 dengan jumlah data sebanyak 34 provinsi.

Tabel 2 Tabel Data Pengangguran Terbuka Menurut Provinsi

Provinsi	Total 2015	Total 2016	Total 2017	Total 2018	Total 2019	Rata-Rata
Aceh	17,66	15,7	13,96	12,91	11,73	14,39
Sumatera Utara	13,1	12,33	12,01	11,15	10,97	11,91
Sumatera Barat	12,88	10,9	11,38	11,1	10,62	11,38
Riau	14,55	13,37	11,98	11,92	11,54	12,67
Jambi	7,07	8,66	7,54	7,51	7,81	7,72
Sumatera Selatan	11,1	8,25	8,19	8,25	8,47	8,85
Bengkulu	8,12	7,14	6,55	6,21	5,89	6,78
Lampung	8,58	9,16	8,76	8,39	7,99	8,58
Kep. Bangka Belitung	9,64	8,77	8,24	7,29	7,01	8,19
Kep. Riau	15,25	16,72	13,6	13,55	13,32	14,49
Dki Jakarta	15,59	11,89	12,5	11,58	11,35	12,58
Jawa Barat	17,12	17,46	16,71	16,33	15,72	16,67
Jawa Tengah	10,3	8,83	8,72	8,74	8,71	9,06
Di Yogyakarta	8,14	5,53	5,86	6,41	6	6,39
Jawa Timur	8,78	8,35	8,1	7,84	7,75	8,16
Banten	18,13	16,87	17,03	16,29	15,69	16,80
Bali	3,36	4,01	2,76	2,23	2,71	3,01
Nusa Tenggara Barat	10,67	7,6	7,18	7,1	6,69	7,85
Nusa Tenggara Timur	6,95	6,84	6,48	5,99	6,45	6,54
Kalimantan Barat	9,93	8,81	8,58	8,41	8,59	8,86
Kalimantan Tengah	7,68	8,49	7,36	7,19	7,43	7,63
Kalimantan Selatan	9,75	9,08	8,3	8,36	7,81	8,66
Kalimantan Timur	14,67	16,81	15,46	13,5	12,75	14,64
Kalimantan Utara	11,47	9,15	10,71	9,9	10,2	10,29
Sulawesi Utara	17,72	14	13,3	12,95	11,62	13,92
Sulawesi Tengah	7,09	6,75	6,78	6,62	6,69	6,79
Sulawesi Selatan	11,76	9,91	10,38	10,73	10,39	10,63
Sulawesi Tenggara	9,17	6,5	6,44	6,05	6,55	6,94
Gorontalo	7,71	6,64	7,93	7,65	7,53	7,49
Sulawesi Barat	5,16	6,05	6,19	5,61	4,63	5,53
Maluku	16,65	14,03	17,06	14,65	13,99	15,28
Maluku Utara	11,61	7,44	10,15	9,42	10,06	9,74
Papua Barat	12,69	13,19	14,01	11,97	11,52	12,68
Papua	7,71	6,32	7,58	6,11	7,07	6,96

2. Menetukan nilai Centroid yang ditentukan secara manual atau acak yang diambil dari Rata-Rata data.

Tabel2. Nilai Centroid

C1	**Maximum**	**16,80**
C2	**Average**	**3,01**

3. Menghitung jarak dari centroid

Untuk menghitung jarak antara titik centroid dengan titik tiap objek menggunakan Euclidion Distance.

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2}$$

Maka perhitungan untuk jarak dari *Centroid* ke-1 adalah sebagai berikut :

D (1.1) = $\sqrt{(16{,}80 - 14{,}39)^2}$ = 2,41

D (1.2) = $\sqrt{(16{,}80 - 11{,}91)^2}$ = 4,89

D (1.3) = $\sqrt{(16{,}80 - 11{,}39)^2}$ = 5,43

Dan seterusnya sampai dengan $D_{1.33}$. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut :

D (2.1) = $\sqrt{(3{,}01 - 14{,}39)^2}$ = 11,38

D (2.2) = $\sqrt{(3{,}01 - 11{,}91)^2}$ = 4,89

D (2.3) = $\sqrt{(3{,}01 - 11{,}39)^2}$ = 5,43

Dan seterusnya sampai dengan $D_{2.33}$. Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari ketiga *Centroid*. Tabel Jarak dari *Centroid* adalah sebagai berikut :

Tabel 4. Jarak *Centroid* Iterasi ke-1

C1	C2	Jarak Centroid
2,41	11,38	2,41
4,89	8,90	4,89
5,43	8,36	5,43
4,13	9,66	4,13
9,08	4,70	4,70
7,95	5,84	5,84
10,02	3,77	3,77
8,23	5,56	5,56
8,61	5,18	5,18
2,31	11,47	2,31
4,22	9,57	4,22
0,13	13,65	0,13
7,74	6,05	6,05
10,41	3,37	3,37
8,64	5,15	5,15
0,00	13,79	0,00
13,79	0,00	0,00
8,95	4,83	4,83
10,26	3,53	3,53
7,94	5,85	5,85

9,17	4,62	4,62
8,14	5,65	5,65
2,16	11,62	2,16
6,52	7,27	6,52
2,88	10,90	2,88
10,02	3,77	3,77
6,17	7,62	6,17
9,86	3,93	3,93
9,31	4,48	4,48
11,27	2,51	2,51
1,53	12,26	1,53
7,07	6,72	6,72
4,13	9,66	4,13
9,84	3,94	3,94
2,41	11,38	2,41
4,89	8,90	4,89
5,43	8,36	5,43

4. Menentukan *Cluster* atau Pengelompokan

Dalam menentukan *Cluster* dengan mencari nilai *Cluster* berdasarkan nilai minimal dari nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai minimal pada Iterasi 1. Berikut tabel *Cluster* pada Iterasi 1 sebagai berikut :

Tabel 5 *Cluster* Iterasi ke-1

C1	C2
1	
1	
1	
1	
	1
	1
	1
	1
	1
1	
1	
1	
	1
	1
	1
1	
	1
	1
	1
	1
	1
	1

1	
1	
1	
	1
1	
	1
	1
	1
1	
	1
1	
	1

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. Untuk mencari nilai Centroid selanjutnya dengan menggunakan Centroid baru pada Iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada Cluster di tabel diatas. Adapun Centroid baru untuk mencari Cluster selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada Cluster tersebut kemudian memmbagikannya sebanyak jumlah nilai. *Data* Centroid *baru Iterasi ke-1 adalah sebagai berikut :*

Tabel 6 *Centroid* Baru Iterasi Ke-1

C1	**13,45**
C2	**7,49**

Dengan menggunakan langkah – langkah yang sama seperti sebelumnya untuk menentukan Jarak dari *Centroid* dengan menggunakan *Centroid* baru Iterasi ke-1, maka berikut hasil Jarak dari *Centroid* :

Tabel 7. Jarak *Centroid* Iterasi ke-2

C1	C2	Jarak Centroid
0,70	6,77	0,70
1,78	4,29	1,78
2,32	3,76	2,32
1,02	5,05	1,02
5,98	0,10	0,10
4,84	1,23	1,23
6,91	0,84	0,84
5,12	0,96	0,96
5,50	0,57	0,57
0,79	6,87	0,79
1,11	4,96	1,11
2,97	9,05	2,97
4,63	1,44	1,44
7,31	1,23	1,23
5,53	0,54	0,54

3,11	9,18	3,11
10,68	4,61	4,61
5,85	0,23	0,23
7,15	1,08	1,08
4,83	1,24	1,24
6,06	0,01	0,01
5,03	1,04	1,04
0,94	7,02	0,94
3,41	2,67	2,67
0,22	6,30	0,22
6,91	0,83	0,83
3,06	3,01	3,01
6,75	0,68	0,68
6,20	0,13	0,13
8,17	2,09	2,09
1,58	7,66	1,58
3,96	2,12	2,12
1,02	5,06	1,02
6,74	0,66	0,66
0,70	6,77	0,70
1,78	4,29	1,78
2,32	3,76	2,32

Dari tabel Jarak *Centroid* diatas, maka *Cluster* atau pengelompokkan Iterasi ke-2 adalah sebagai berikut :

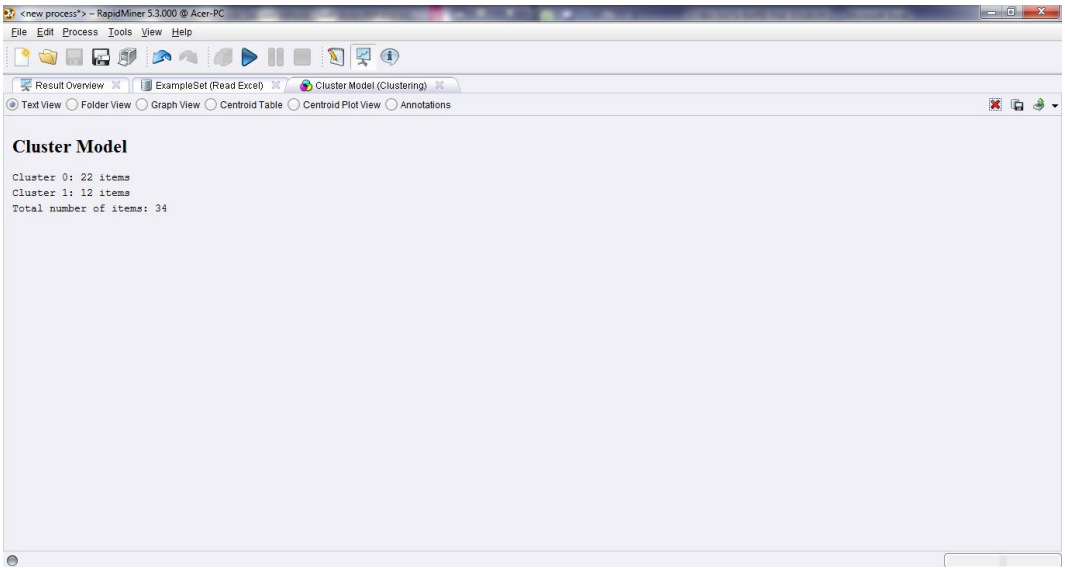
Tabel 8 *Cluster* Iterasi ke-2

C1	C2
1	
1	
1	
1	
	1
	1
	1
	1
	1
1	
1	
1	
	1
	1
	1
1	
	1
	1
	1
	1
	1

	1
1	
	1
1	
	1
1	
	1
	1
	1
1	
	1
1	
	1

Kemudian dilanjutkan ke iterasi selanjutnya sampai nilai *cluster* sama atau tidak berubah pada *cluster* Iterasi selanjutnya maka perhitungan dihentikan.

Untuk mendapatkan hasil pengelompokkan pada tahap selanjutnya dilakukan pengolahan data dengan menggunakan aplikasi *rapidminer* . Tahap ini akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan tools rapidminer. Dapat dilihat pada gambar 1:

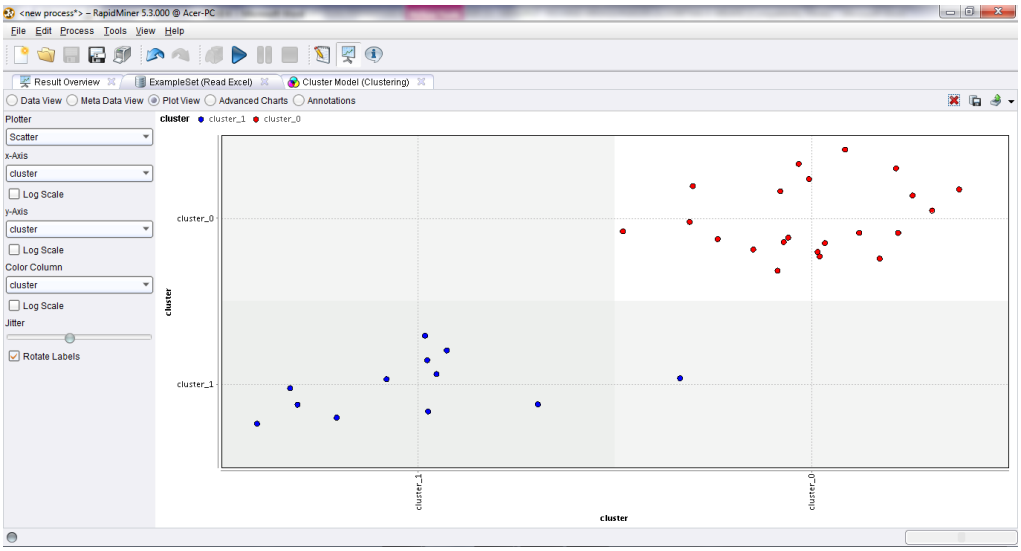


Gambar 1. Nilai *Cluster* Model

Keterangan :

1. Jumlah Cluster 0 (Rendah) berjumlah 22 Items
2. Jumlah Cluster 1 (Tinggi) berjumlah 12 Items
   Jumlah keseluruhan items adalah 34

Sehingga dapat diketahui hasil pengelompokkan rapidminer dapat dilihat pada gambar sebagai berikut :



Gambar 2. Hasil Pengelompokan

4. Kesimpulan

Hasil akhir dari penelitian yang menggunakan data sebanyak 34 provinsi ini, dapat disimpulkan bahwa telah didapatkan masing-masing nilai *cluster* yakni :

1. *Cluster* tinggi (C1) dengan jumlah sebanyak 12 Provinsi yaitu : 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.
2. *Cluster* rendah (C2) dengan jumlah sebanyak 22 Provinsi selain dari *cluster* tinggi.
3. Proses pemberentihan iterasi pada pengujian yang dilakukan pada penelitian ini yaitu terjadi pada iterasi ke 3.
4. Nilai hasil akurasi yang dilakukan dengan perhitungan manual dan dengan aplikasi *rapidminer* bernilai sama.

5. Referensi

[1] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018.

[2] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam," *Y. Darmi, A. Setiawan*, vol. 12, no. 2, pp. 148–157, 2016.

[3] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016.

[4] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.

[5] D. Retno and S. Mayangsari, "PENGELOMPOKKAN JUMLAH DESA / KELURAHAN YANG MEMILIKI DENGAN MENGGUNAKAN METODE K-MEANS CLUSTER," vol. 3, pp. 370–377, 2019.

[6] F. Ramdhani and A. Hoyyi, "Pengelompokan Provinsi Di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster," *J. Gaussian*, vol. 4, no. 4, pp. 875–884, 2015.

[7] S. Kasus, U. Dehasen, S. Haryati, A. Sudarsono, and E. Suryana, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4 . 5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.