

PENENTUAN *CLUSTER* OPTIMUM PADA TINGKAT PENGANGGURAN DAN TINGKAT KEMISKINAN DI JAWA TIMUR MENGGUNAKAN ALGORITMA *FUZZY C-MEANS*

¹Nigati Triuspita (1510651077), ²Ari Eko Wardoyo, S.T, M.Kom, ³Agung Nilogiri, S.T, M.Kom

Jurusan Teknik Informatika

Fakultas Teknik

Universitas Muhammadiyah Jember

E-mail: nitritaoke@gmail.com

Abstrak-- Ada banyak metode yang digunakan dalam menyelesaikan *clustering* suatu data. Salah satunya adalah metode *Fuzzy C-Means* (FCM) yang merupakan salah satu metode yang dapat diandalkan dalam menyelesaikan masalah *clustering* di wilayah Jawa Timur. Penelitian ini bertujuan untuk menentukan *cluster* optimum di wilayah Jawa Timur yang dapat membantu pemerintah untuk mengetahui permasalahan dan membantu pengambilan kebijakan pada wilayah-wilayah kabupaten/kota di provinsi Jawa Timur. Proses penelitian menggunakan data dari badan pusat statistik yaitu tingkat pengangguran dan tingkat kemiskinan pada tahun 2010 sampai 2015. Pada penelitian ini digunakan *Davies Bouldin Index* (DBI) sebagai uji validasi *cluster* untuk penentuan *cluster* optimum. Data tingkat pengangguran dan tingkat kemiskinan di analisis menggunakan RStudio. Dari hasil perhitungan metode FCM dan juga penentuan *cluster* optimum didapat hasil pada 2 *cluster* dengan nilai DBI 1.275877, 3 *cluster* dengan nilai DBI 0.9936573, 4 *cluster* dengan nilai DBI 0.8736634. *Cluster* optimum terdapat pada 4 *cluster* dengan nilai yang minimum.

Kata kunci – *Fuzzy C-Means*, *Davies Bouldin Index*

1.1 Pendahuluan

Upaya menurunkan tingkat pengangguran dan menurunkan tingkat kemiskinan adalah sama pentingnya. Pengangguran dan kemiskinan adalah sebuah contoh permasalahan pembangunan dan sosial kemasyarakatan. Pembangunan sendiri merupakan aspek penting untuk melihat kinerja seberapa efektifnya penggunaan sumber daya yang ada sehingga lapangan kerja dapat menyerap angkatan kerja yang tersedia. Semakin meningkat produksi barang dan jasa maka semakin banyak juga angkatan kerja yang dibutuhkan maka pengangguran dan kemiskinan semakin menurun. Secara teori masyarakat yang mempunyai pekerjaan dan berpenghasilan yang dapat memenuhi kebutuhan hidupnya, maka tidak dikatakan miskin.

Jumlah pengangguran akan meningkat dan menurun setiap tahunnya sesuai dengan pertumbuhan penduduk. Menurut data BPS (Badan Pusat Statistik) di Jawa timur sendiri tingkat pengangguran pada tahun 2015 sekitar 4.47%, pada tahun 2016 menurun sekitar 4.21% dan pada tahun 2012 menurun sekitar 4%. Untuk tingkat kemiskinan pada tahun 2015 sekitar 12,34%, pada tahun 2016 menurun sekitar 12.05% dan pada tahun 2017 menurun sekitar 11.77%. Jika dilihat perkabupatennya pada tahun 2017 pengangguran tertinggi yaitu di Kota Malang sekitar 7.22% sedangkan kemiskinan tertinggi yaitu kabupaten Sampang sekitar 23.56% dibanding kabupaten/kota lainnya.

Pada penelitian sebelumnya sudah ada pengetahuan untuk mengetahui kelompok wilayah namun menggunakan

atribut atau variabel yang berbeda. Penelitian ini mengelompokkan wilayah berdasarkan tingkat pengangguran terbuka dan tingkat kemiskinan yang dapat membantu pemerintah untuk mengetahui permasalahan dan membantu pengambilan kebijakan pada wilayah-wilayah kabupaten/kota di provinsi Jawa Timur. Penentuan wilayah kota yang memiliki tingkat pengangguran yang tinggi sangat berpengaruh pada tingkat kemiskinan berdasarkan pengetahuan dari jumlah data Badan Pusat Statistik. Karena hal itu, untuk mendapatkan pengetahuan tersebut maka diperlukan suatu proses pengolahan data suatu teknik *data mining*. Teknik *data mining* yang akan digunakan pada penelitian ini adalah metode *Fuzzy C-Means*. Penelitian diharapkan dapat menghasilkan suatu informasi yang dapat meng*cluster* atau mengelompokkan sebuah wilayah kabupaten/kota di Jawa Timur

2.1 Tinjauan Pustaka

a) Jawa timur

Provinsi Jawa Timur merupakan wilayah Negara Republik Indonesia yang menempati wilayah paling timur Pulau Jawa setelah provinsi DKI Jakarta, Banten, Jawa Barat, Jawa Tengah dan D.I. Yogyakarta. Jawa Timur telah menjadi bagian Negara Republik Indonesia sejak awal kemerdekaan.

Secara geografis Jawa Timur berbatasan dengan Pulau Bali di sebelah timur, sebelah barat berbatasan dengan Provinsi Jawa Tengah, sebelah utara berbatasan dengan Laut Jawa, di ujung selatan berbatasan dengan Samudera Hindia.

Luas wilayah provinsi Jawa Timur yaitu 46.428,57 km secara administrative terbagi menjadi 38 kabupaten/kota, dengan rincian 29 kabupaten dan 9 kota

b) Pengangguran

Menurut Badan Pusat Statistik TPT (Tingkat Pengangguran Terbuka) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Pengangguran terbuka terdiri dari mereka yang tidak punya pekerjaan dan mencari pekerjaan, mereka yang tidak punya pekerjaan dan mempersiapkan usaha, mereka yang tidak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan, mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja.

c) Kemiskinan

Kemiskinan adalah suatu keadaan dimana seseorang tidak mampu memelihara dirinya sendiri sesuai dengan taraf kehidupan kelompok dan juga tidak mampu memanfaatkan tenaga mental, maupun fisiknya dalam kelompok tersebut

d) Data Mining

Data mining adalah istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk

mengekstraksi dan mengidentifikasi informasi dan pengetahuan yang terkait dari berbagai database besar.

Data Mining memiliki peranan penting dalam bidang ilmu dan teknologi, industri, cuaca, dan keuangan, karena *Data Mining* dapat mengolah data dalam skala besar. *Data Mining* memiliki beberapa metode yaitu *clustering*, *regresi*, *klasifikasi*, *seleksi variable*, dan *market basket analisis*.

e) *Clustering*

Clustering merupakan teknik menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik tertentu yang telah ditentukan. Dalam pengelompokan tersebut nilai label nya belum diketahui sehingga diharapkan setelah melakukan pengelompokan data dapat diketahui label dari data tersebut. Metode *clustering* juga sering disebut tahapan awal sebelum melakukan metode lain seperti klasifikasi.

f) *Fuzzy C-Means*

Fuzzy C-Means yaitu salah satu teknik pengelompokan data yang mana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotan.

Fuzzy C-means Clustering (FCM), dikenal juga sebagai *Fuzzy ISODATA*, merupakan salah satu metode *clustering* yang merupakan bagian dari metode *Hard K-Means*. FCM menggunakan model pengelompokan *fuzzy* sehingga data

yang didapat menjadi anggota dari semua cluster yang terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat data dalam suatu kelas atau cluster ditentukan oleh derajat keanggotaannya.

g) *Davies Bouldin Index*

Davies-bouldin index adalah salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi.

Pengelompokan, kohesi dapat diartikan sebagai jumlah dari kedekatan data terhadap centroid dari cluster yang diikuti. Sedangkan separasi didasarkan pada jarak antar centroid dari clusternya.

3.1 Metodologi

a) Metode Penelitian

Pada metode penelitian ini dilakukan pendekatan kuantitatif yang dimana analisisnya berupa data-data numeric (angka). Penelitian ini menggunakan metode Algoritma *Fuzzy C-Means*. Algoritma ini dipakai untuk mengelompokkan dua atau lebih *cluster* yang dipisahkan ke dalam kelompok yang berbeda-beda.

b) Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data pada Badan Pusat Statistik Jawa Timur terdiri dari 38 wilayah sebanyak 456 data. Pengumpulan data diawali dengan mengunjungi Badan Pusat Statistik Jember dan melakukan wawancara,

studi pustaka dari beberapa jurnal, literature, buku-buku, dan situs internet yang berkaitan dengan penelitian ini.

c) Data Set

Data set untuk penelitian ini adalah data tingkat pengangguran dan tingkat kemiskinan di Jawa Timur mulai dari tahun 2010 sampai dengan 2015. Dimana data tersebut akan di *cluster* menggunakan algoritma *Fuzzy C-Means*.

4.1 Hasil dan Pembahasan

1) Penerapan Algoritma *Fuzzy C-Means*.

Berikut adalah langkah-langkah algoritma *Fuzzy C-Means* :

a. Menentukan data yang akan di cluster

Data Ke	Kab/Kota	TPT 2010	TK 2010	TPT 2015	TK 2015
1	Kab. Trenggalek	2.15	15.98	2.46	13.39
2	Kab. Kediri	3.75	15.52	5.02	12.91
3	Kab. Lumajang	3.17	13.98	2.60	11.52
4	Kab. Banyuwangi	3.92	11.25	2.55	9.17
5	Kab. Probolinggo	2.02	25.22	2.51	20.82
6	Kab. Mojokerto	4.84	12.23	4.05	10.57
7	Kab. Bojonegoro	3.29	18.78	5.01	15.71
8	Kab. Sampang	1.77	32.47	2.51	25.69
9	Kota Malang	8.68	5.90	7.28	4.60
10	Kota Surabaya	6.84	7.07	7.01	5.82
.....
38	Kota Batu	5.55	5.11	4.29	4.71

b. Menentukan Parameter Awal

- Jumlah Cluster = c
- Pangkat = w
- Maksimum iterasi = MaxIter
- Error terkecil yang diharapkan = ξ
- Fungsi objektif awal = $P_0 = 0$
- Iterasi awal = $t = 1$

- c. Membangkitkan bilangan random (0-1).

Menghitung jumlah setiap kolom:

$$Q_i = \sum_{k=1}^c \mu_{ik}$$

dengan $j=1,2,\dots,n$.

Menghitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i}$$

- d. Menghitung pusat cluster

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

- e. Menghitung fungsi objektif pada iterasi ke-t

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

- f. Menghitung perubahan matriks partisi

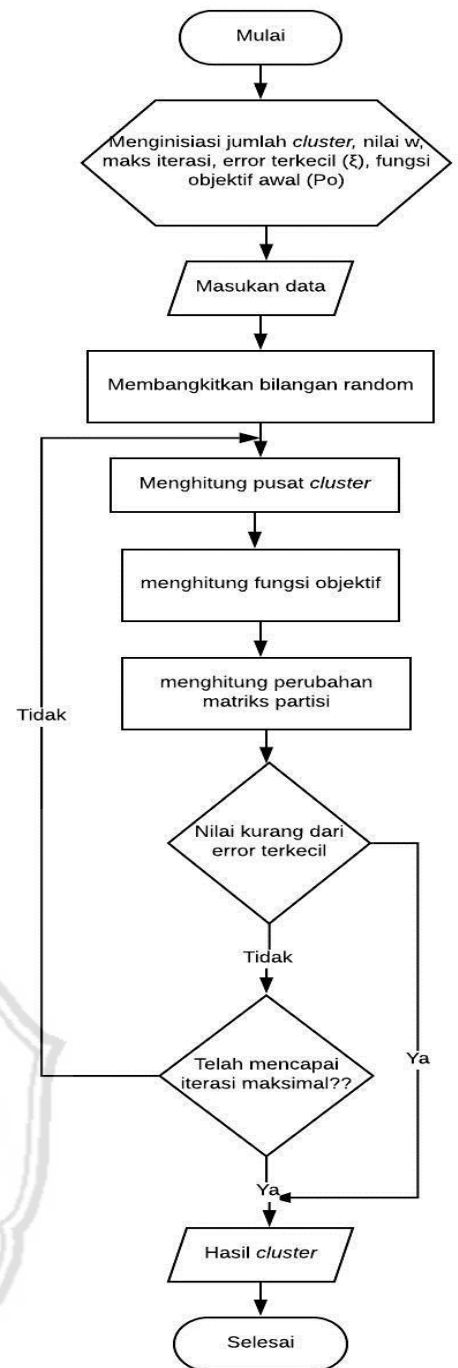
$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}$$

dengan: $i=1,2,\dots,n$; dan $k=1,2,\dots,c$

- g. Memeriksa kondisi berhenti

- jika : $t > \text{MaxIter}$ maka berhenti artinya jika hasil pengurangan dari fungsi objektif dan fungsi objektif awal kurang dari error terkecil atau iterasi awal lebih dari maksimal iterasi maka akan berhenti
- jika tidak : $t = t+1$, kembali ke langkah 4

Berikut penggambaran algoritma *C-Means* menggunakan flowchart:



- 2) Penentuan Jumlah *Cluster* Optimum

- Mengelompokkan data berdasarkan *cluster* yang di ikuti. Menghitung DBI pada *Fuzzy C-Means* menggunakan hasil dari pusat *cluster* yang kemudian dikelompokkan sesuai dengan hasil dari derajat keanggotaan
- Menentukan nilai perhitungan kerapatan (*density*) data pada *cluster* atau *Sum of Square Within (SSW)*

- c) Melakukan perhitungan untuk nilai separasi atau *Sum of Square Between* (SSB)
- d) Menghitung nilai rasio dan DBI

3) Hasil

Semakin kecil nilai *Davies Bouldin Index* yang diperoleh maka semakin baik *cluster* yang diperoleh dari pengelompokan. Maka untuk *cluster* optimum pada tingkat pengangguran dan tingkat kemiskinan di Jawa Timur menggunakan 4 *cluster*

Cluster	DBI
2	1,275877
3	0,9936573
4	0,8736634

Kab. Pasuruan, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kota Probolinggo , C2 terdapat 4 wilayah terdiri dari Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang, Kab. Sumenep, C3 terdapat 10 wilayah terdiri dari Kab. Pacitan, Kab. Trenggalek, Kab. Bondowoso, Kab. Situbondo, Kab. Ngawi, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Gresik dan C4 terdapat 9 wilayah terdiri dari Kab. Sidoarjo, Kab. Kediri, Kab. Blitar, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu.

5.1 Kesimpulan dan Saran

a) Kesimpulan

Dari hasil penelitian yang dilakukan, kesimpulan yang didapat bahwa:

1. Hasil dari penerapan metode *Fuzzy C-Means* berdasarkan tingkat pengangguran dan tingkat kemiskinan yang mengelompokan wilayah kabupaten/kota di Jawa Timur yang diperoleh *cluster* optimum dengan hasil 4 *cluster* berdasarkan nilai *davies bouldin index* 0,8736634.
2. Hasil pengelompokan dengan 4 *cluster* pada C1 terdapat 15 wilayah terdiri dari Kab. Ponorogo, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi,

b) Saran

Saran yang dapat dikembangkan pada penelitian ini adalah:

1. Penelitian ini dapat dikembangkan menggunakan metode clustering yang berbeda
2. Dapat juga menggunakan uji validasi cluster yang berbeda dan menambahkan jumlah cluster yang akan dihitung
3. Dapat juga meningkatkan jumlah atribut dan variabel pada data yang digunakan

6.1 Daftar Pustaka

[1] Badan Pusat Statistik Provinsi Jawa Timur. 2017. *Keadaan Angkatan Kerja Provinsi Jawa Timur 2017*. Jember : Badan Pusat Statistik.

- [2] Dinas Kesehatan Provinsi Jawa Timur. *Ringkasan Eksekutif, Data dan Informasi Kesehatan Provinsi Jawa Timur tahun 2013*.
- [3] Jain, A. K., Murthy, M. N. & Flynn, P. J. 1999. *Data Clustering: A Review* ACM Computing Surveys, Vol. 31, No. 3.
- [4] Prasetyo, eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Penerbit Andi Yogyakarta
- [5] Santoso, B. 2007. *Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.
- [6] Soekanto, Soerjono . 1982. *Sosiologi : Suatu Pengantar*. Rajawali Press.
- [7] Turban, E dkk. 2005. *Decision Support Systems and Intelligent Systems*. Yogyakarta : Andi Offset.

