Pendekatan Fuzzy Clustering untuk Mengetahui Potensi Usaha mikro di Surabaya

Muhibbuddin, Muhammad Isa Irawan, Subchan
Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111
e-mail : mii@its.ac.id

Abstrak-Surabaya memiliki peran yang besar dalam menciptakan lapangan pekerjaan untuk warga Surabaya dan daerah lain sekitar Surabaya. Salah satunya melalui kegiatan Industri. di Jawa Timur sendiri, sebesar 54,43 persen berasal dari UMKM, sehingga UMKM berperan penting dalam perbaikan ekonomi daerah dan dapat membantu pemerintah dalam pengentasan pengangguran. Ada beberapa macam Usaha Mikro dengan omzet dan tenaga kerja yang berbeda-beda. Sebanyak 1288 Usaha Mikro yang ada di Surabaya, sehingga dilakukan pengelompokan Usaha Mikro di Kecamatan Tegalsari sebagai perwakilan usaha mikro di Surabaya, sebanyak 115 usaha mikro yang telah terdaftar di Dinas Perindustrian dan Perdagangan Kota Surabaya. Dengan tujuan untuk mengetahui perkembangan kelompok usaha mikro sesuai dengan omset dan tenaga kerja yang dimiliki.

Dalam penelitian ini menggunakan metode clustering hierarki (single linkage, complete linkage, average linkage) dan non hierarki (fuzzy c-means). Setelah dilakukan pengelompokan dengan keempat metode, kemudian mendapatkan pengelompokan terbaik dengan kriteria SSE (sum square error) .

Kata Kunci— ***Industri, Usaha Mikro, Metode Clustering***

I. PENDAHULUAN

Surabaya dikenal sebagai kota metropolitan yang termasuk didalamnya aktivitas ekonomi yang berkembang secara cepat. Perusahaan-perusahaan besar yang dimiliki oleh pemerintah maupun swasta banyak ditemui di Kota Surabaya. Sehingga kegiatan perekonomian berkembang cepat dan memicu perkembangan ekonomi daerah lain di sekitar Kota Surabaya [1].

Industri di Jawa Timur 54,34 persen berasal dari UMKM dan mampu menampung 98 persen tenaga kerja Industri Jawa Timur 54,34 persen berasal dari UMKM dan mampu menampung 98 persen tenaga kerja. Sehingga selain mensejahterakan pelaku UMKM, juga membantu pemerintah dalam pengentasan pengangguran. UMKM atau yang lebih dikenal Usaha Mikro Kecil dan Menengah adalah merupakan sebuah klasifikasi dari kapasitas suatu usaha dari mulai mikro, kecil, dan menengah. UMKM dapat menjadi satu faktor percepatan pembangunan kota. Sehingga Pemerintah kota Surabaya melakukan berbagai upaya untuk memaksimalkan peran pemberdayaan masyarakat dalam pembangunan kota. Salah satunya dengan menggelar pelatihan bagi Kader Pemberdayaan Masyarakat. KPM mengalami perkembangan yang signifikan. Mulanya, jumlah KPM tahun 2009 hanya 18 orang. Kemudian meningkat menjadi 181 orang pada 2010 dan 287 orang pada 2011. Sedangkan tahun 2012, jumlah KPM melonjak menjadi 466 orang.

Sehingga perkembangan UMKM di Surabaya dapat meningkat setiap tahunnya. Disamping itu, pada setiap usaha dalam perkembangannya akan memiliki tingkat perkembangan yang berbeda antara usaha yang satu dengan yang lainnya. Sehingga perlu dianalisa kelompok usaha yang sudah berkembang maupun yang belum.

Setiap usaha mikro dalam perkembangannya dapat dilihat dari jumlah omzet hasil kegiatan usahanya dan jumlah tenaga kerja yang terlibat didalamnya. Dalam analisis statistik, omzet dan tenaga kerja dinyatakan sebagai variabel. Analisis ini bertujuan untuk mengelompokkan objek-objek berdasarkan karakteristik diantara objek-objek tersebut, sebagai hasilnya akan terbentuk kelompok-kelompok dengan ciri khas tiap kelompok. Dari penelitian ini akan didapatkan hasil kelompok usaha mikro berdasarkan omzet dan tenaga kerja.

Berdasarkan latar belakang tersebut dalam tugas akhir ini penulis mengambil judul "Pendekatan Fuzzy Clustering untuk Mengetahui Potensi UMKM di Surabaya".

II. TINJAUAN PUSTAKA

A. Analisis Cluster

Analisis cluster adalah suatu alat untuk mengelompokkan sejumlah *n* objek berdasarkan *p* variabel yang secara relatif mempunyai kesamaan karakteristik diantara objek-objek tersebut. Objek tersebut akan dikelompokkan ke dalam satu atau lebih kelompok (cluster) sehingga objek–objek yang berada dalam satu *cluster* akan mempunyai kemiripan atau kesamaan karakter. Adapun ciri-ciri *cluster* yang baik adalah :

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu cluster (*within-cluster*).
2. Heterogenitas (perbedaan) yang tinggi antar cluster yang satu dengan cluster yang lainnya (*between-cluster*).

Dalam pengelompokan, terdapat dua macam metode, yaitu metode hierarki dan nonhierarki, dimana metode hierarki mengelompokkan suatu pengamatan secara bertahap, sedangkan pada metode non hierarki dilakukan dengan melakukan partisi pada ruang sampel.

B. Pengelompokan Hierarki

Metode hierarki merupakan metode *clustering* yang hasil pengelompokannya disajikan secara bertingkat atau

berjenjang dari $n, (n-1), \ldots, 1$ kelompok digabungkan ditampilkan dalam bentuk diagram yang disebut *dendrogram*.. Fungsi jarak yang seringkali digunakan adalah *Euclidean*, dimana didefinisikan sebagai jarak antara observasi ke-i dan ke-k. Rumus jarak euclid dari objek ke-*i* menuju objek ke-*h* dirumuskan pada persamaan (2.1).

$$d(x,y) = \sqrt[p]{\sum_{i=1}^{n}(x_i - y_i)^p} \; ; p > 0 \qquad (2.1)$$

a. *Single Linkage* (Pautan Tunggal/*Nearest Neighbor*)

Single linkage membentuk kelompok-kelompok dari individu dengan menggabungkan jarak terdekat atau terkecil. *Single linkage* dirumuskan pada persamaan (2.2).

$$d_{(ih)g} = min\{d_{ig}, d_{hg}\} \qquad (2.2)$$

dengan $d_{(ih)g}$ jarak objek *ih* dengan objek *g*, d_{ig} jarak terpendek objek *i* dengan objek *g*, d_{hg} adalah jarak terpendek objek *h* dengan objek *g*.

b. *Complete Linkage* (Pautan Lengkap/*Furthest Neighbor*)

Complete linkage mengelompokkan semua objek dalam cluster yang berada paling jauh satu sama lainnya. *Complete linkage* dirumuskan pada persamaan (2.3)

$$d_{(ih)g} = min\{d_{ig}, d_{hg}\} \qquad (2.3)$$

dengan $d_{(ih)g}$ jarak objek pada kelompok (*ih*) dengan objek *g*, d_{ig} jarak terjauh objek *i* dengan objek *g*, d_{hg} adalah jarak terjauh objek *h* dengan objek *g*.

c. *Average Linkage* (Pautan Rataan/*Between Method*)

Metode ini memperlakukan jarak antara dua cluster sebagai jarak rata-rata antara semua pasangan individu. Untuk memperoleh penggabungan (aglomerasi) objek satu dengan lainnya dirumuskan sebagai berikut :

$$d_{(ih)g} = \frac{\sum_i \sum_g d_{ig}}{N_{(ih)} N_g} \qquad (2.4)$$

dengan $d_{(ih)g}$ jarak antara objek g pada kelompok (ih) dan objek g. $N_{(ih)}$ adalah banyaknya anggota dalam kelompok (ih), sedangkan N_g banyaknya anggota pada kelompok g.

B. Metode Non Hierarki (Non Hierarchical Clustering)

Berlawanan dengan metode hierarki, prosedur pengelompokan non hierarki ini tidak dilakukan secara bertahap. Metode yang tergolong pengelompokan non hierarki diantaranya adalah fuzzy *c-means*.

Fuzzy c means adalah salah satu bagian dari metode K-means, karena harus menentukan jumlah titik cluster terlebih dulu untuk menentukan anggota-anggota cluster. Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster, dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menujui lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif (Gelley,2000).

$$J(U, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_c) = \sum_{j=1}^{n} \sum_{i=1}^{C} (\mu_{ij})^m (d_{ij})^2 \qquad (2.5)$$

dimana u_{ij} diantara 0 dan 1, c_i adalah pusat cluster *i* , $d_{ij} = \|\boldsymbol{c}_i - \boldsymbol{x}_j\|$ adalah jarak Euclidean antara pusat cluster ke-*i* dan data ke-*j*.

$$\boldsymbol{c}_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m \boldsymbol{x}_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \qquad (2.6)$$

$$\mathrm{u}_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \qquad (2.7)$$

Algoritma pengelompokan Fuzzy C-means diberikan sebagai berikut:

1. Menentukan
   - Jumlah kelompok yang akan dibentuk $(2 \leq c \leq n)$
   - Weighting exponent $(m = 2)$
   - Nilai treshold (ε)
   - Fungsi objektif awal = 0
   - Iterasi awal, t=1
2. Inisialisasi matrik fungsi keanggotaan awal U^0
3. Menghitung fuzzy *cluster* center (**c**) dengan peramaan(2.6)
4. Update anggota matriks (U) dengan persamaan (2.7)
5. Membandingkan nilai keanggotaan dalam matriks U, jika
   $|U^{t+1} - U^t| < \varepsilon$, makasudah konvergen. Jika
   $|U^{t+1} - U^t| < \varepsilon$, kembali ke langkah 3.

C. Internal Cluster Dispersion Rate (icdrate)

Terdapat beberapa kriteria dalam menilai kebaikan dalam pengelompokan. Kebaikan dalam cluster adalah homogenitas yang tinggi antar anggota dalam satu cluster (*within-cluster*) dan heterogenitas yang tinggi antar cluster (*between-cluster*). Intinya adalah untuk menilai homogenitas dalam kelompok dan heterogenitas antar kelompok. Perbandingan metode pengelompokan dapat diukur dengan menghitung rata-rata persebaran *internal cluster* terhadap partisi secara keseluruhan. Metode ini sering digunakan dalam menaksir akurasi dari algoritma pengelompokan. Semakin kecil nilai *icdrate*, semakin baik hasil pengelompokannya. Perhitungan *internal cluster dispersion rate* (*icdrate*) sebagai berikut :

$$icdrate = 1 - \frac{SSB}{SST} = 1 - \frac{(SST - SSW)}{SST} = 1 - R^2 \ldots \qquad (2.6)$$

dengan :

$$R^2 = \frac{(SST - SSW)}{SST}$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^{c} \sum_{k=1}^{p} \left(x_{ij}{}^k - \overline{x^k}\right)^2$$

SST : total jumlah dari kuadrat jarak terhadap rata-rata keseluruhan

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^{c} \sum_{k=1}^{p} \left(x_{ij}{}^k - \overline{x_j^k}\right)^2$$

SSW : total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya

c : banyak kelompok

n_c : banyak data pada kelompok ke-*i*

p : banyak variabel

$x_{ij}{}^k$: sampel ke-i pada kelompok ke-j dan variabel ke-*k*

$\overline{x^k}$: rata-rata seluruh sampel pada variabel-*k*

$\overline{x_j^k}$: rata-rata sampel pada kelompok ke-*j* dan variabel ke-*k*

III. PEMBAHASAN

C. *Deskripsi Statistik*

Sebanyak 115 usaha mikro yang ada di kecamatan Tegalsari kota Surabaya, dimana merupakan kecamatan dengan usaha mikro terbanyak untuk mewakili kota Surabaya dengan jumlah seluruhnya sebanyak 1288 Siup Mikro. Setiap usaha mikro memiliki tenaga kerja dan omset yang nantinya akan dianalisa pengelompokannya dan dihitung besarnya pajak sebesar 1%, yang bisa diperoleh dari jumlah omzet UMKM.

D. *Hierarchical Clustering*

Metode pengelompokan hierarki dalam penelitian ini adalah penggabungan (*agglomerative*). Metode yang termasuk *hierarchical clustering* diantaranya *single linkage*, *complete linkage*, dan *average linkage.*

Sebelum menggunakan metode fuzzy c-means dengan menggunakan *n* titik pusat untuk memperoleh hasil cluster, maka perlu melihat kondisi klaster dari metode Hirarki yaitu dengan melihat perbedaan anggota dari masing-masing metode(single, average, dan complete) dengan menggunakan inputan $n \leq 2$ titik cluster.

a. Menggunakan 2 titik cluster

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.1 Perbandingan anggota kelompok metode hirarki dengan 2 titik cluster

Metode	Cluster 1	Cluster 2
Single linkage	111	4
Average linkage	111	4
Complete linkage	111	4

b. Menggunakan 3 titik cluster

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.2 Perbandingan anggota kelompok metode hirarki dengan 3 titik cluster

Metode	Cluster 1	Cluster 2	Cluster 3
Single linkage	101	10	4
Average linkage	101	10	4
Complete linkage	101	10	4

c. Menggunakan 4 titik cluster

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.3 Perbandingan anggota kelompok metode hirarki dengan 4 titik cluster

Metode	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Single linkage	101	10	3	1
Average linkage	101	10	3	1
Complete linkage	47	54	20	4

d. Menggunakan 5 titik cluster

Diperoleh kelompok cluster sebagai berikut:

Tabel 4.4 Perbandingan anggota kelompok metode hirarki dengan 5 titik cluster

Metode	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Single linkage	101	10	2	1	1
Average linkage	87	14	10	3	1
Complete linkage	47	54	10	3	1

Dari 4 tabel diatas, dapat diketahui bahwa cluster dengan 5 titik pusat memiliki jumlah masing-masing anggota cluster yang berbeda, sehingga pada proses pengelompokan selanjutnya akan menggunakan 5 titik pusat cluster.

4.2.1 Metode Single Linkage

Dari analisa sebelumnya, untuk proses cluster menggunakan 5 titik cluster, sehingga dalam *cluster observation* minitab untuk mendapatkan kelompok cluster adalah sebagai berikut :

>>Variabel or distance matrix : 'Tenaga Kerja'- 'omset tiap bulan'
>>Linkage methode = Single
>>Distance measure = Euclidean
>>Number of cluster = 5

Sehingga diperoleh kelompok cluster sebagai berikut :

a. Cluster pertama

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.5 Anggota kelompok pertama dari metode *Single Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Alexa	1	97
2	Alvin Cake	2	168
3	Aneka Kue Berkah	1	165
4	Aris Kroket	1	180
...	...	...	...
101	Zhifa Cake	1	153

b. Cluster kedua

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.6 Anggota kelompok kedua dari metode *Single Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Benyamin Donat	5	600
2	CV Bintang Dunia	6	460
3	DH 45	2	600
4	El-Be Jajanan	4	468
...	...	...	..
10	Wan Wan Siomay	4	496

c. Cluster ketiga

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.7 Anggota kelompok ketiga dari metode *Single Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Kue Mangkok	10	930
2	Onde-Onde	8	900

d. Cluster keempat

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.8 Anggota kelompok keempat dari metode *Single Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Lemper Rahma	8	1050

e. Cluster kelima

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.9 Anggota kelompok kelima dari metode *Single Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	PT Lien Tripilllar perkasa	13	780

4.2.2 Metode Average Linkage

Dari analisa sebelumnya, untuk proses cluster menggunakan 5 titik cluster, sehingga dalam *cluster observation* minitab untuk mendapatkan kelompok cluster adalah sebagai berikut :

>>Variabel or distance matrix : 'Tenaga Kerja'- 'omset tiap bulan'
>>Linkage methode = Average
>>Distance measure = Euclidean
>>Number of cluster = 5
Sehingga diperoleh kelompok cluster sebagai berikut :

a. Cluster pertama

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.10 Anggota kelompok pertama dari metode *Average Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	A-Bee Pudding	1	186
2	Alvin Cake	2	168
3	Ana Bakery	1	240
4	Ana Prima	1	234
...	.....	....	....
87	Zhifa Cake	1	153

b. Cluster kedua

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.11 Anggota kelompok kedua dari metode *Average Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Alexa	1	97
2	Fajar Baru Motor	1	55
3	Falah Kue Basah	1	105
4	Galleriez Handicraft	2	110
...	.....	....	....
14	UD Tunggal Jaya Motor	2	71

c. Cluster ketiga

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.12 Anggota kelompok ketiga dari metode *Average Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Benyamin Donat	5	600
2	CV Bintang Dunia	6	460
3	DH 45	2	600
4	El-Be Jajanan	4	468
...	...	...	...
10	Wan Wan Siomay	4	496

d. Cluster keempat

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.13 Anggota kelompok keempat dari metode *Average Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Kue Mangkok	10	930
2	Lemper Rahma	8	1050
3	Onde-Onde	8	900

e. Cluster kelima

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.14 Anggota kelompok kelima dari metode *Average Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	PT Lien Tripilllar perkasa	13	780

4.2.3 Complete Linkage

Dari analisa sebelumnya, untuk proses cluster menggunakan 5 titik cluster, sehingga dalam *cluster observation* minitab untuk mendapatkan kelompok cluster adalah sebagai berikut :

>>Variabel or distance matrix : 'Tenaga Kerja'- 'omset tiap bulan'
>>Linkage methode = Complete
>>Distance measure = Euclidean
>>Number of cluster = 5
Didapatkan kelompok cluster sebagai berikut :

a. Cluster pertama

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.15 Anggota kelompok pertama dari metode *Complete Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	A-Bee Pudding	1	186
2	Alexa	1	97
3	Alvin Cake	2	168
4	Aneka Kue Berkah Jaya	1	165
...	.....	....	....
47	Zhifa Cake	1	153

b. Cluster kedua

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.16 Anggota kelompok kedua dari metode *Complete Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Alexa	1	240
2	Ana Prima	1	234
3	Aneka Jajan Pasar	1	219
4	Aneka Putuh Ayu	2	225
...	...	....	...
54	Wingko Farel	3	294

c. Cluster ketiga

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.17 Anggota kelompok ketiga dari metode *Complete Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Benyamin Donat	5	600
2	CV Bintang Dunia	6	460
3	DH 45	2	600
4	El-Be Jajanan	4	468
...	...	...	...
10	Wan Wan Siomay	4	495

d. Cluster keempat

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.18 Anggota kelompok keempat dari metode *Complete Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Kue Mangkok	10	930
2	Lemper Rahma	8	1050
3	Onde-Onde	8	900

e. Cluster kelima

Diperoleh kelompok cluster sebagai berikut :

Tabel 4.19 Anggota kelompok kelima dari metode *Complete Linkage*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	PT Lien Tripilllar Perkasa	13	780

4.3 Non-Hierarchical Clustering

Metode pengelompokan non-hierarki dalam penelitian ini menggunakan *Fuzzy C- Means.*

4.3.1 Fuzzy C-Means

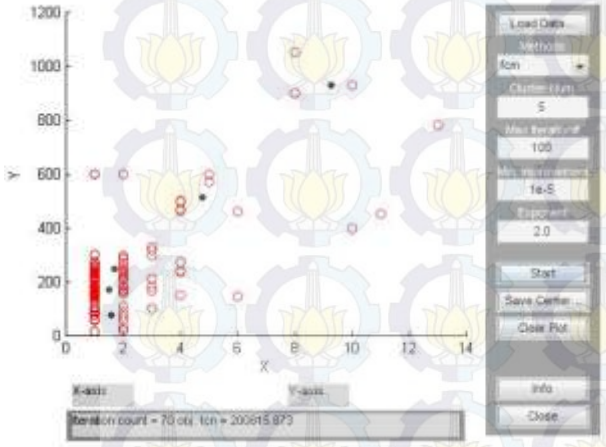
Dari analisa perbandingan jumlah titik cluster yang terjadi perbedaan anggota antara metode Hirarki diketahui jumlah titik cluster adalah 5 titik pusat, maka dalam command windows pada matlab untuk melihat titik pusat dengan jumlah titik cluster sebanyak 5 titik.

>> ketik "findcluster" pada command window matlab berikut adalah tampilan awal pada matlab :

Gambar 4.1 Interface fuzzy c-means

>> input "data" pada "Load Data"
>> isi "Cluster Num" dengan angka "5
>>isi "Eksponent" dengan angka 2 (Klawonn dan Höppner, 2001)

Sehingga didapatkan titik pusat sebagai berikut :



Gambar 4.2 Titik pusat cluster

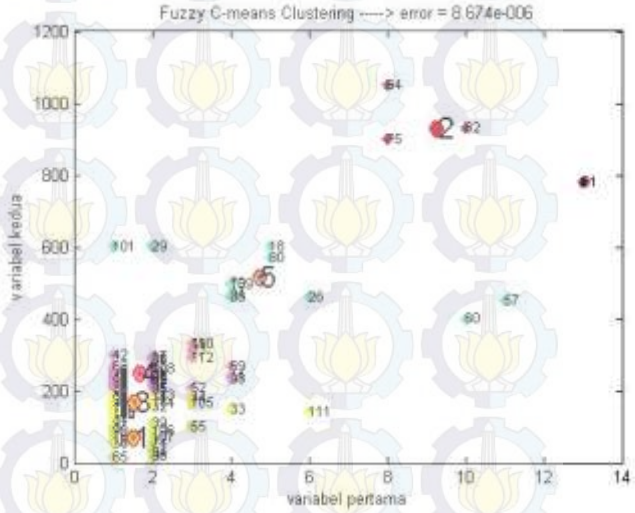
Didapatkan titik pusat :

$C_1 = (1.5357\ ,68.9907)$
$C_2 = (1.5652\ ,165.5085)$
$C_3 = (1.7076\ ,247.1605)$
$C_4 = (4.7556\ ,514.5992)$
$C_5 = (9.2839\ ,929.1598)$

Dengan menghitung jarak Euclidean pada persamaan(4.1) antara setiap data dengan masing-masing titik pusat :

$$d(1,C_1) = \sqrt{(1-1{,}5357)^2 + (186-68{,}9907)^2} = 117{,}01$$
$$d(1,C_2) = \sqrt{(1-1{,}5652)^2 + (186-165{,}5085)^2} = 20{,}5$$
$$d(1,C_3) = \sqrt{(1-1{,}7076)^2 + (186-247{,}1605)^2} = 61{,}16$$
$$d(1,C_4) = \sqrt{(1-4{,}7556)^2 + (186-514{,}5992)^2} = 328{,}62$$
$$d(1,C_5) = \sqrt{(1-9{,}2839)^2 + (186-929{,}1598)^2} = 743{,}21$$
$$d(2,C_1) = \sqrt{(1-1{,}5357)^2 + (97-68{,}9907)^2} = 28{,}01$$
$$d(2,C_2) = \sqrt{(1-1{,}5652)^2 + (97-165{,}5085)^2} = 68{,}51$$
$$d(2,C_3) = \sqrt{(1-1{,}7076)^2 + (97-247{,}1605)^2} = 150{,}16$$
$$d(2,C_4) = \sqrt{(1-4{,}7556)^2 + (97-514{,}5992)^2} = 417{,}61$$
$$d(2,C_5) = \sqrt{(1-9{,}2839)^2 + (97-929{,}1598)^2} = 832{,}20$$
$$\vdots \qquad\qquad \vdots$$
$$d(115,C_1) = \sqrt{(1-1{,}5357)^2 + (153-68{,}9907)^2} = 84{,}01$$
$$d(115,C_2) = \sqrt{(1-1{,}5652)^2 + (153-165{,}5085)^2} = 12{,}52$$
$$d(115,C_3) = \sqrt{(1-1{,}7076)^2 + (153-247{,}1605)^2} = 94{,}16$$
$$d(115,C_4) = \sqrt{(1-4{,}7556)^2 + (153-514{,}5992)^2} = 316{,}62$$
$$d(115,C_5) = \sqrt{(1-9{,}2839)^2 + (153-929{,}1598)^2} = 776{,}2$$

Berikut ini adalah simulasi anggota kelompok dari data ke titik pusat terdekat :



Gambar 4.3 Hasil pengelompokan fuzzy c-means

Dari gambar diatas dapat dilihat bahwa terdapat 5 titik pusat dengan masing-masing anggotanya. Satu kelompok cluster memiliki warna seragam.

a. Cluster pertama

Pada cluster pertama (C_1), dari perhitungan jarak *Euclidean* diperoleh anggota kelompok sebagai berikut :

Tabel 4.21 Anggota kelompok pertama dari metode *fuzzy c-means*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Alexa	1	97
2	Fajar Baru Motor	1	55
3	Falah Kue Basah	1	105
4	Galleriez Handicraft	2	110
...	...	...	...
13	UD Tunggal Jaya Motor	2	71

b. Cluster kedua

Pada cluster kedua (C_2), dari perhitungan jarak *Euclidean* diperoleh anggota kelompok sebagai berikut:

Tabel 4.22 Anggota kelompok kedua dari metode *fuzzy c-means*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	A Bee Pudding	1	186
2	Alvin Cake	1	97
3	Aneka Kue Berkah Jaya	1	165
4	Aris Kroket	1	180
...	....	....	....
33	Zhifa Cake	1	153

c. Cluster ketiga

Pada cluster ketiga (C_3), dari perhitungan jarak *Euclidean* diperoleh anggota kelompok sebagai berikut:

Tabel 4.23 Anggota kelompok ketiga dari metode *fuzzy c-means*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Ana Bakery	1	240
2	Ana Prima	1	234
3	Aneka Jajanan Pasar	1	219
4	Aneka Putuh Ayu	2	225
...	.....	.....	....
54	Wingko Farel	3	294

d. Cluster keempat

Pada cluster keempat (C_4), dari perhitungan jarak *Euclidean* diperoleh anggota kelompok sebagai berikut:

Tabel 4.24 Anggota kelompok keempat dari metode *fuzzy c-means*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Benyamin Donat	5	600
2	CV Bintang Dunia	6	460
3	DH 45	2	600
4	El-Be Jajanan	4	468
...	....	...	...
11	Wan Wan Siomay	4	495

e. Cluster Kelima

Pada cluster kelima (C_5), dari perhitungan jarak *Euclidean* diperoleh anggota kelompok sebagai berikut:

Tabel 4.25 Anggota kelompok kelima dari metode *fuzzy c-means*

No	Nama Perusahaan	Tenaga Kerja	Omzet
1	Kue Mangkok	10	930
2	Lemper Rahma	8	1050
3	Onde-Onde	8	900
4	PT Lien Tripilllar perkasa	13	780

4.4 Uji Homogenitas dan Pemilihan Metode Terbaik

Membandingkan antara hasil pengelompokan metode klastering dapat dilakukan dengan berbagai cara dan rumusan. Salah satunya adalah dengan menghitung performansi klaster dari hasil pengolahan data. Salah satu cara yang paling sering dipakai dan cukup sederhana adalah jumlah dari kesalahan kuadrat yakni SSE (Sum Squared Error) dengan rumusan sebagai berikut :

$$\boldsymbol{SSE} = \sum_{i=1}^{k} \sum_{x \in D_i} \|\boldsymbol{x} - \boldsymbol{m_i}\|$$

Keterangan :

$\boldsymbol{k}$ = jumlah klaster

$\boldsymbol{D_i}$ = nilai set data ke-*i*

$\boldsymbol{m_i}$ = mean (rata-rata) dari setiap klaster.

Untuk pemilihan metode terbaik dilakukan dengan metode icdrate (*internal cluster dispersion rate*)

Berikut hasil perhitungan nilai SSE untuk masing-masing metode. (Data selengkapnya ada pada lampiran E)

1. SSE Fuzzy C-Means Clustering

Tabel 4.26 Nilai SSE Fuzzy C-Means

No	Cluster	$\boldsymbol{m_i}$	SSE
1	Cluster pertama	69,2	360,99
2	Cluster kedua	161,36	517,76
3	Cluster ketiga	246,16	1243,41
4	Cluster keempat	509,57	663,4
5	Cluster kelima	915,05	299,94
Nilai SSE Fuzzy C-Means			3085,5

2. SSE Single Linkage

Tabel 4.27 Nilai SSE *single linkage*

No	Cluster	$\boldsymbol{m_i}$	SSE
1	Cluster pertama	194,61	5784,96
2	Cluster kedua	520,52	575,84
3	Cluster ketiga	915,04	30,02
4	Cluster ke-4	1050,03	0
5	Cluster kelima	780,11	0
Nilai SSE *single linkage*			6390,82

3. SSE Average Linkage

Tabel 4.28 Nilai SSE *Average Linkage*

No	Cluster	$\boldsymbol{m_i}$	SSE
1	Cluster pertama	214,79	3706,07
2	Cluster kedua	69,2	360,98
3	Cluster ketiga	520,52	575,85
4	Cluster ke-4	960,04	179,98
5	Cluster ke-5	780,98	0
Nilai SSE *Average Linkage*			4822,88

4. SSE Complete Linkage

Tabel 4.29 Nilai SSE Complete Linkage

No	Cluster	$\boldsymbol{m_i}$	SSE
1	Cluster pertama	1323,06	1814,46
2	Cluster kedua	245,53	1574,52
3	Cluster ketiga	520,52	575,85
4	Cluster ke-4	960,04	179,98
5	Cluster ke-5	780,11	0
Nilai SSE Complete Linkage			4144,81

Perbandingan nilai icdrate (*internal cluster dispersion rate*) dari masing-masing metode :

Tabel 4.30 Perbandingan nilai icdrate

No	Metode	SSB	SST	Nilai icdrate
1	Fuzzy C-Means	3235452	3397451	0.048
2	Single Linkage	2846889	3397451	0.162
3	Average Linkage	3167684	3397451	0.068
4	Complete Linkage	3090341	3397451	0.09

Dari perhitungan nilai SSE dan icdrate pada masing-masing metode didapat bahwa metode fuzzy c-means memiliki nilai kesalahan paling rendah dan memiliki nilai icdrate yang paling kecil daripada metode lainnya, sehingga metode fuzzy c-means adalah metode terbaik dalam penelitian yang memiliki tingkat homogenitas inter-*cluster* yang lebih tinggi dibandingkan dengan hasil klasterisasi dengan menggunakan metode *single linkage*, *average linkage* dan *complete linkage*, sehingga didapatkan anggota cluster terbaik yaitu :

a.) Cluster pertama sebanyak 13 anggota cluster, dengan rata-rata omzet sebesar Rp 1.900.000 sampai Rp 11.000.000 setiap bulan dengan banyak pekerja 1 sampai dengan 3 orang setiap perusahaan.
b.) Cluster kedua sebanyak 33 anggota cluster dengan rata-rata omzet sebesar Rp 3.000.000 sampai Rp 20.100.000 setiap bulan dengan banyak pekerja 1 sampai dengan 6 orang setiap perusahaan.
c.) Cluster ketiga sebanyak 54 anggota cluster dengan rata-rata omzet sebesar Rp 17.400.000 sampai Rp 40.000.000 setiap bulan dengan banyak pekerja 1 sampai dengan 10 orang setiap perusahaan.
d.) Cluster keempat sebanyak 11 anggota cluster dengan rata-rata omzet sebesar Rp 12.000.000 sampai Rp 60.000.000 setiap bulan dengan banyak pekerja 1 sampai dengan 11 orang setiap perusahaan.
e.) Cluster kelima sebanyak 4 anggota cluster dengan rata-rata omzet sebesar Rp 78.000.000 sampai Rp 105.000.0000 setiap bulan dengan banyak pekerja 8 sampai dengan 13 orang setiap perusahaan.

E. Pajak dan Jumlah Pekerja

Berdasarkan data yang telah didapatkan, jumlah pajak yang dapat diperoleh dari UMKM yang terdaftar dalam SIUP Mikro Dinas Peridustrian dan Perdagangan Kota Surabaya adalah :

a. Jumlah Omzet dari 115 UMKM selama 1 bulan
   Omzet Total = Rp 2.851.950.000
b. Pajak menurut PP nomor 46 tahun 2013
   Pajak = 1% × Omzet
   Pajak = 1% × Rp 2.851.950.000
   = Rp 28.519.500
   Maka pajak yang dapat diperoleh untuk setiap bulannya sebesar Rp 28.519.500
c. Jumlah pekerja yang dapat diserap dari kegiatan 115 UMKM yang berada di kecamatan Tegalsari adalah sebanyak 258 orang.

IV. KESIMPULAN

Dari hasil pembahasan, diperoleh kesimpulan sebagai berikut:

1. Berdasarkan perbandingan variasi anggota dari metode Hirarki (Single Linkage, Average Linkage, dan Complete Linkage) didapatkan 5 titik cluster.
2. Berdasarkan metode Fuzzy C-Means didapatkan titik pusat cluster :
   $C_1 = (1.5357\,, 68.9907)$
   $C_2 = (1.5652\,, 165.5085)$
   $C_3 = (1.7076\,, 247.1605)$
   $C_4 = (4.7556\,, 514.5992)$
   $C_5 = (9.2839\,, 929.1598)$
3. Berdasarkan data yang telah didapatkan, jumlah pajak yang dapat diperoleh dari UMKM yang terdaftar dalam SIUP Mikro Dinas Peridustrian dan Perdagangan Kota Surabaya adalah Rp 28.519.500 dan tenaga kerja yang dapat diserap dari kegiatan 115 UMKM yang berada di kecamatan Tegalsari adalah sebanyak 258 orang.

DAFTAR PUSTAKA

[1] Purwanto, A.D.(2013). "Fuzzy Clustering untuk Optimasi Perencanaan Unit Pembangkit Listrik di Jawa Timur*"*.Jurusan Matematika. Institut Teknologi Sepuluh Nopember. Surabaya.

[2] Lazulfa, I. (2012) "Analisis Cluster Kabupaten/Kota di Jawa Timur Berdasarkan Tingkat Pencemaran Udara". Jurusan Matematika. Institut Teknologi Sepuluh Nopember. Surabaya.

[3] Hartini,Entin.(2004). "Metode Clustering Hirarki". Risalah komputasi dalam Sains dan Teknologi Nuklir(XVI).Jakarta : BATAN.

[4] Mingoti, S.A & Lima, J.O.(2006). Comparing SOM neural Network with fuzzy c-Means, K- means and Traditional hierarchical clustering algorithms. European Journal of Operational Research 174: 1742-1759.

[5] Wakhidah, N. (2012). "Klasifikasi Data Multidimensi Menggunakan Subtractive Clustering Dan K-Nearest Neightbor". Fakultas Teknologi Informasi dan Komunikasi Universitas Semarang.

[6] BAPEDA JATIM (23 Juli 2013,16:03). "Koperasi dan UMKM jadi Modal Hebat. http://bappeda.jatimprov.go.id. (diakses tanggal 4 september 2013)

[7] Ardianto, A. F.(2012). "Pemkot Surabaya Gelar Pelatihan Pemberdayaan Masyarakat".[Online].Tersedia http:// www.beritajatim.com / detailnews / 2012-09-24. (diakses 04 september 2013).

[8] Mulyahjo. "Pengertian Usaha Mikro". http://mulyahjo.blogspot.com (diakses pada tanggal 4 oktober 2013)

[9] Ichsan (30 desember 2011). "Pengertian dan Kriteria UMKM". http://tunas63.wordpress.com (diakses pada tanggal 4 oktober 2013)

[10] Pravitasari, Anindiya Apriliyanti. (5 desember 2009). "Penentuan Banyak Kelompok dalam Fuzzy C-Means Cluster Berdasarkan Proporsi Eigen Value dari Matriks Similarity dan Indeks XB (Xie dan Beni). Matematika FMIPA. Universitas Padjajaran. Bandung.

[11] Ekawati, Ratna & Yulis, Nurul. 2013. "Klasifikasi Usaha Kecil dan Menengah (UKM) Sektor Industri Dengan Metode Fuzzy C-Means Clustering Wilayah Kota Cilegon". Teknik Industri. Universitas Sultan Ageng Tirtayasa. Banten.