

Identifikasi Spasial Calon Walikota Manado Tahun 2015 Menggunakan Algoritma Fuzzy C-Means

Winsy Weku¹, Charitas Fibriani²

¹Program Studi Matematika, FMIPA, UNSRAT Manado, winsy_weku@unsrat.ac.id

²Program Studi Ilmu Komputer, Fakultas Teknologi Informasi Universitas Kristen Satya Wacana Salatiga, charitas.fibriani@staff.uksw.edu

Abstrak

Analisis cluster bertujuan untuk mengidentifikasi kelompok-kelompok objek yang sama, dan membantu untuk menemukan distribusi pola dan korelasi menarik dalam set data yang besar. Fuzzy Clustering telah banyak dipelajari dan diterapkan dalam berbagai bidang utama dan validasi fuzzy clustering memainkan peran yang sangat penting dalam fuzzy clustering. Dalam penelitian ini akan dilakukan pengelompokan wilayah kaum pemilih untuk Calon Walikota Manado tahun 2015. Metode yang digunakan adalah algoritma Fuzzy C-Means. Dengan menggunakan 3 cluster dan 2 cluster didapatkan pengelompokan 28, 18 dan 40 untuk 3 cluster dan pengelompokan 46 dan 40 untuk 2 cluster. Hasil tersebut dapat dilihat dengan memperhatikan derajat keanggotaan fuzzy-nya.

Kata kunci: Calon Walikota, Algoritma Fuzzy C-Means, Geospasial

Spatial Identification of Manado's Candidate Major in 2015 Using Fuzzy C-Means Algorithm

Abstract

Cluster analysis aims to identify the groups to the same object, and helping to find distribution patterns and interesting correlation in large data sets. Fuzzy Clustering has been widely studied and applied in many key areas and validation of fuzzy clustering plays a very important role in the fuzzy clustering. In this research will be the regional grouping voters for Mayor Candidate Manado in 2015. The method used is Fuzzy C-Means algorithm. By using 3 clusters and 2 cluster grouping obtained 28, 18 and 40 for the three clusters and the grouping of 46 and 40 for the second cluster. The results can be seen by noting its fuzzy membership degree.

Keywords : Major Candidate, Fuzzy C-Means Algorithm, Geospatially

1. Pendahuluan

Clustering adalah salah satu pendekatan umum untuk pemodelan deskriptif dari sejumlah besar data, yang memungkinkan analisis untuk fokus pada representasi tingkat yang lebih tinggi dari data. Metode Clustering menganalisis dan mengeksplorasi dataset untuk mengasosiasikan objek dalam kelompok, sehingga benda-benda di setiap kelompok memiliki karakteristik yang sama. Karakteristik ini dapat dinyatakan dalam berbagai cara: misalnya, satu dapat menggambarkan objek dalam sebuah cluster sebagai penduduk yang dihasilkan oleh distribusi gabungan, atau sebagai sekumpulan objek yang meminimalkan jarak dari pusat massa dari kelompok.

Clustering melibatkan tugas membagi titik data ke dalam kelas homogen atau kelompok sehingga barang-barang di kelas yang sama yang semirip mungkin dan item dalam kelas yang berbeda adalah sebagai berbeda mungkin. Clustering juga dapat dianggap sebagai bentuk kompresi data, di mana sejumlah besar sampel diubah menjadi sejumlah kecil prototipe perwakilan atau cluster. Tergantung pada data dan aplikasi, jenis tindakan kesamaan dapat digunakan untuk mengidentifikasi kelas, di mana ukuran kesamaan mengontrol bagaimana cluster terbentuk. Beberapa contoh nilai-nilai yang dapat digunakan sebagai ukuran kesamaan antara jarak, konektivitas, dan intensitas.

Dalam statistik dan optimasi, masalah biasa untuk memaksimalkan atau meminimalkan fungsi, dan variabel dalam ruang tertentu. Seperti masalah optimasi ini mungkin menganggap

beberapa jenis, masing-masing dengan karakteristik sendiri, banyak teknik yang telah dikembangkan untuk menyelesaikannya. teknik ini sangat penting dalam *Data Mining* dan wilayah *Knowledge Discovery* karena dapat digunakan sebagai dasar untuk sebagian besar metode yang kompleks dan kuat.

2. Algoritma Fuzzy C-Means

Fuzzy c-means (FCM) adalah metode clustering yang memungkinkan satu bagian dari data milik dua atau lebih kelompok. Metode ini (yang dikembangkan oleh Dunn pada tahun 1973 dan ditingkatkan oleh Bezdek pada tahun 1981) sering digunakan dalam pengenalan pola. Hal ini didasarkan pada minimalisasi fungsi tujuan berikut [1][2][3][4][5][6]:

$$J_m(\mu, a) = \sum_{j=1}^n \sum_{i=1}^c (\mu_{ij})^m d^2(x_j, a_i) \tag{1}$$

Dimana bobot eksponen $1 < m < +\infty$ menyajikan tingkat kekaburan dan notasi $d(x_j, a_i)$ menunjukkan jarak (perbedaan) ukuran antara x_j titik data dan center cluster a_i . Partisi fuzzy dilakukan melalui optimasi berulang (meminimalkan) dari fungsi tujuan $J_m(\mu, a)$, dengan update keanggotaan μ_{ij} dan pusat cluster a_i oleh:

$$a_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \tag{2}$$

$$\mu_{ij} = \frac{d(x_j, a_i)^{-\frac{2}{m-2}}}{\sum_{k=1}^c d(x_j, a_k)^{-\frac{2}{m-2}}} \quad i=1,2,\dots,c; j=1,2,\dots,n \tag{3}$$

Berdasarkan urutan eksekusi untuk tahap s menggunakan stage $(s-1)$ menurut persamaan update (2) dan (3). Prosedur ini konvergen ke minimum lokal atau titik pelana J_m .

Algoritma ini terdiri dari langkah-langkah berikut:

1. *Tetapkan $2 \leq c \leq n$ dan tetapkan sembarang $\epsilon > 0$.
Berikan inisialisasi $\mu^{(0)} = (\mu_1^{(0)}, \dots, \mu_c^{(0)})$ dan misalkan $s=1$.*
2. *Hitung pusat klaster $a^{(s)}$ dengan $\mu^{(s-1)}$ dimana*

$$a_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m}$$
3. *Update $\mu^{(s)}$ dengan $a^{(s)}$ dimana*

$$\mu_{ij} = \frac{d(x_j, a_i)^{-\frac{2}{m-2}}}{\sum_{k=1}^c d(x_j, a_k)^{-\frac{2}{m-2}}}$$
4. *Bandingkan $\mu^{(s)}$ to $\mu^{(s-1)}$ in a convenient matrix form $\| \cdot \|$.
IF $\| \mu^{(s)} - \mu^{(s-1)} \| < \epsilon$, STOP
ELSE $s=s+1$ dan kembali ke langkah 2.*

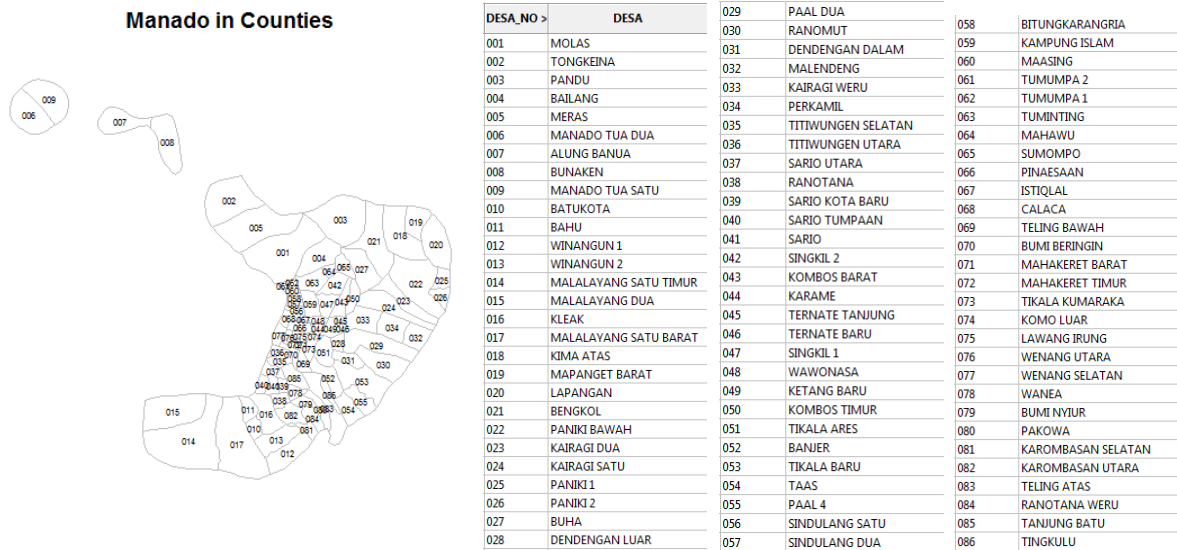
3. Pembahasan

Penelitian ini termasuk contoh analisis geospasial dari dunia nyata: geo-demografis. Ukuran cluster yang dipilih berdasarkan prosedur BIC. Sebagai alat komputasi yang dipilih adalah R, dimana paket R yang akan digunakan adalah e-1071 dikembangkan oleh David Meyer, tersedia dalam CRAN repositori. Tata Ruang penyajian data dilakukan dengan ArcView.

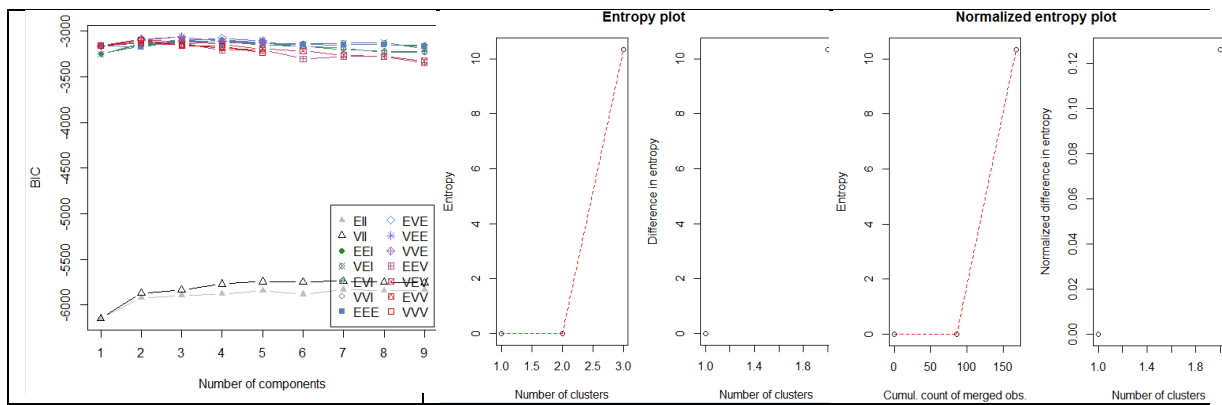
Contoh analisis data yang akan diamati dalam penelitian ini adalah geodemographic, karena data geodemographic menggunakan beberapa batas administratif dalam kota seperti pada contoh pada Gambar 1.

Peta administrasi kota Manado akan dilakukan analisis cluster untuk melihat mana daerah dikelompokkan secara terpisah dengan mempertimbangkan variabel yang ada, yaitu lokasi, bujur dan lintang, 1 pemilih, 2 pemilih dan 3 pemilih.

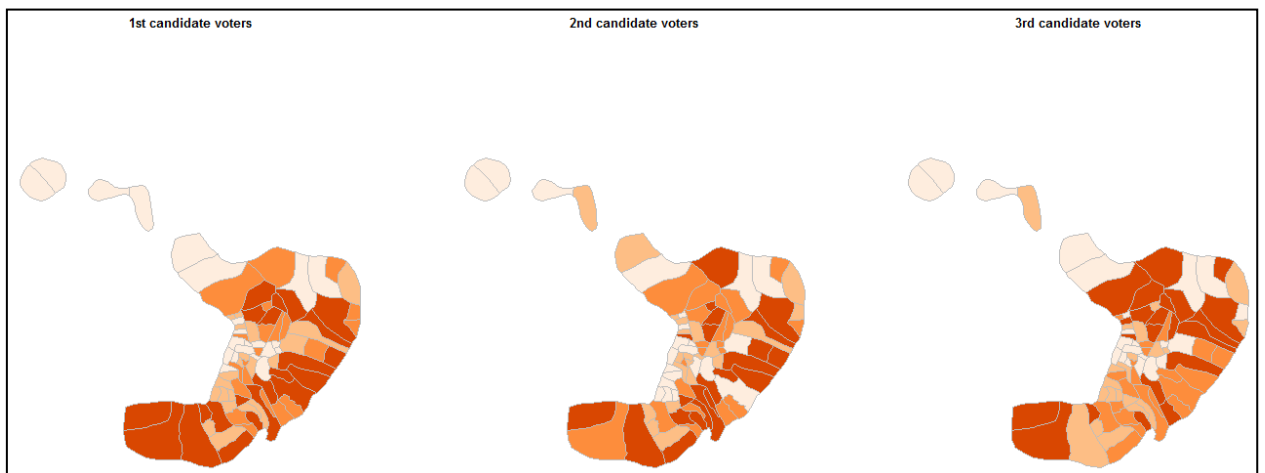
Perkiraan untuk mengetahui berapa banyak cluster yang akan digunakan, digunakan Bayesian Information Criteria (BIC). Menurut BIC dalam Gambar 3, perkiraan cluster optimal adalah 3clusters (perhatikan nilai BIC tertinggi dan jumlah entropi plot).



Gambar 1. Peta administrasi kotamadya Manado

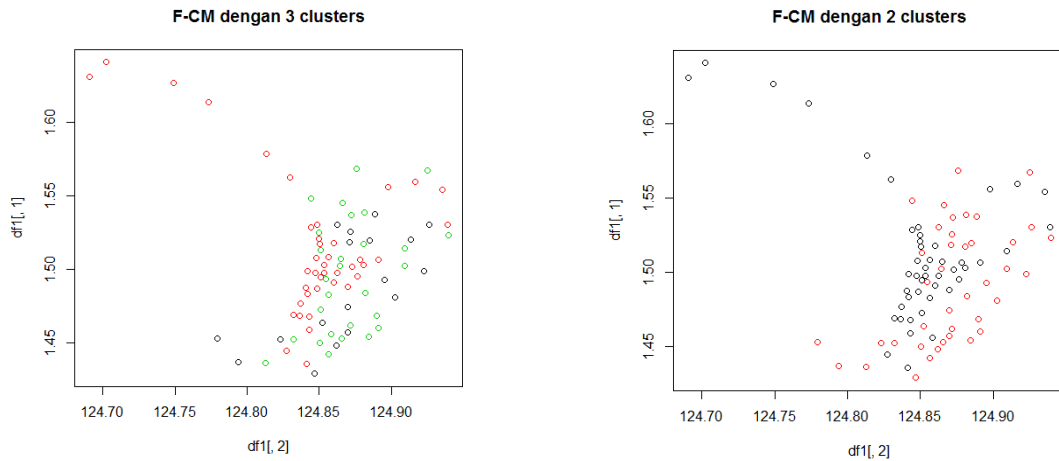


Gambar 2. Ukuran Kluster Optimal dan Entropinya berdasarkan BIC



Gambar 3. Keadaan awal para pemilih di Kotamadya Manado berdasarkan wilayah

Berdasarkan keadaan awal dari pemilih, distribusi pemilih didistribusikan baik berarti bahwa semua daerah memiliki pemilih 59anatic. Sementara ada daerah tertentu yang memiliki pemilih umum, misalnya dengan id 015, 022 dan 023 (Malalayang Dua, Paniki Bawah dan Kairagi Dua, masing-masing) berarti bahwa pemilih di tiga wilayah tersebut adalah pemilih tertinggi dan dibagi secara merata kepada semua tiga kandidat .



Gambar 4. Hasil Klaster di wilayah Kotamadya Manado menggunakan parameter $m=2$ untuk klaster $c=3$ dan $c=2$

Banyaknya kelompok untuk 3 cluster adalah 28, 18 dan 40 untuk masing-masing cluster. Sedangkan perbandingan keanggotaan untuk pengelompokkan dengan 2 cluster adalah 46 untuk kelompok pertama dan 40 untuk kelompok kedua.

Tabel 1. Derajat Keanggotaan dari 86 wilayah untuk menentukan posisi klaster

Derajat keanggotaan untuk 3 cluster				Derajat keanggotaan untuk 2 cluster		
0	0.068544467	0.028198993	0.90325654	0	0.933255063	0.066744937
1	0.078340115	0.033862004	0.88779788	1	0.923033832	0.076966168
2	0.106113647	0.048254584	0.84563177	2	0.898314557	0.101685443
3	0.181743398	0.043981453	0.77427515	3	0.944942888	0.055057112
4	0.053312508	0.018282741	0.92840475	4	0.960636232	0.039363768
5	0.832615763	0.126952684	0.04043155	5	0.099190706	0.900809294
6	0.054748440	0.021583460	0.92366810	6	0.946666184	0.053333816
7	0.088260415	0.038175862	0.87356372	7	0.915594274	0.084405726
8	0.087275077	0.038439052	0.87428587	8	0.914404424	0.085595576
9	0.962757168	0.022498243	0.01474459	9	0.217308020	0.782691980
10	0.722595897	0.149027770	0.12837633	10	0.345630776	0.654369224
11	0.081210785	0.022931164	0.89585805	11	0.979871587	0.020128413
12	0.632295009	0.298585169	0.06911982	12	0.098218332	0.901781668
13	0.092293147	0.880709451	0.02699740	13	0.114743980	0.885256020
14	0.157481769	0.792690176	0.04982806	14	0.142255215	0.857744785
15	0.450366715	0.233837268	0.31579602	15	0.527327960	0.472672040
16	0.346751997	0.066875122	0.58637288	16	0.894940793	0.105059207
17	0.430438617	0.142198582	0.42736280	17	0.700245349	0.299754651
18	0.083704799	0.025556111	0.89073909	18	0.972129697	0.027870303
19	0.034388669	0.011426317	0.95418501	19	0.973900985	0.026099015
20	0.471939259	0.411122221	0.11693852	20	0.192926161	0.807073839
21	0.097989373	0.043975244	0.85803538	21	0.905470932	0.094529068
22	0.074841938	0.032029841	0.89312822	22	0.926144742	0.073855258
23	0.496659070	0.121603588	0.38173734	23	0.688352402	0.311647598
24	0.073877327	0.020864475	0.90525820	24	0.976321583	0.023678417
25	0.031092553	0.011643088	0.95726436	25	0.964347478	0.035652522
26	0.030683594	0.011655334	0.95766107	26	0.965190275	0.034809725
27	0.923413167	0.037870887	0.03871595	27	0.339379618	0.660620382
28	0.033763577	0.012206351	0.95403007	28	0.965172761	0.034827239
29	0.118882592	0.846248079	0.03486933	29	0.117176909	0.882823091
30	0.102465382	0.028305199	0.86922942	30	0.974724011	0.025275989
31	0.097132222	0.035568922	0.86729886	31	0.929228672	0.070771328
32	0.058923955	0.017539240	0.92353680	32	0.981150850	0.018849150
33	0.009928648	0.003274770	0.98679658	33	0.990712371	0.009287629
34	0.061686751	0.025580301	0.91273295	34	0.935758348	0.064241652
35	0.221593348	0.055473346	0.72293331	35	0.915455086	0.084544914
36	0.709598519	0.118086194	0.17231529	36	0.477936741	0.522063259
37	0.387661593	0.080332724	0.53200568	37	0.853461274	0.146538726
38	0.452837083	0.481154169	0.06600875	38	0.067363378	0.932636622

Derajat keanggotaan untuk 3 cluster	Derajat keanggotaan untuk 2 cluster
39 0.674470765 0.088716688 0.23681255	39 0.648856927 0.351143073
40 0.560594963 0.382128101 0.05727694	40 0.047239503 0.952760497
41 0.285519618 0.619485794 0.09499459	41 0.177015536 0.822984464
42 0.016579402 0.005135904 0.97828469	42 0.996118934 0.003881066
43 0.708942952 0.080762996 0.21029405	43 0.641093472 0.358906528
44 0.099349254 0.034077826 0.86657292	44 0.949532038 0.050467962
45 0.033064919 0.011439083 0.95549600	45 0.978200550 0.021799450
46 0.069504219 0.023574117 0.90692166	46 0.964317026 0.035682974
47 0.421615052 0.487532732 0.09085222	47 0.125589020 0.874410980
48 0.343399852 0.075005004 0.58159514	48 0.876906681 0.123093319
49 0.761128673 0.139580291 0.09929104	49 0.277068402 0.722931598
50 0.081588348 0.898136431 0.02027522	50 0.084568214 0.915431786
51 0.091408197 0.891704128 0.01688767	51 0.048240512 0.951759488
52 0.350412243 0.556648537 0.09293922	52 0.155404927 0.844595073
53 0.593288175 0.314293224 0.09241860	53 0.135441884 0.864558116
54 0.312835436 0.542076452 0.14508811	54 0.269742205 0.730257795
55 0.319455495 0.065440298 0.61510421	55 0.903265713 0.096734287
56 0.845839460 0.077824996 0.07633554	56 0.330882119 0.669117881
57 0.131928533 0.043181283 0.82489018	57 0.932466837 0.067533163
58 0.211418463 0.046595686 0.74198585	58 0.945474880 0.054525120
59 0.325557050 0.628878797 0.04556415	59 0.036094878 0.963905122
60 0.068693633 0.023139517 0.90816685	60 0.965389624 0.034610376
61 0.323393890 0.094818656 0.58178745	61 0.833760845 0.166239155
62 0.327919527 0.569231722 0.10284875	62 0.185605610 0.814394390
63 0.356603530 0.608346043 0.03505043	63 0.004105313 0.995894687
64 0.920236892 0.042804725 0.03695838	64 0.286292564 0.713707436
65 0.362975527 0.549729631 0.08729484	65 0.145426718 0.854573282
66 0.203289575 0.060616688 0.73609374	66 0.894880605 0.105119395
67 0.357293000 0.093443479 0.54926352	67 0.810003314 0.189996686
68 0.314128362 0.078075364 0.60779627	68 0.873467253 0.126532747
69 0.028348610 0.009731031 0.96192036	69 0.979556356 0.020443644
70 0.672874364 0.077702627 0.24942301	70 0.689231342 0.310768658
71 0.407944871 0.381999447 0.21005568	71 0.343021892 0.656978108
72 0.844029860 0.080746194 0.07522395	72 0.319829528 0.680170472
73 0.879734428 0.080875967 0.03938960	73 0.164090551 0.835909449
74 0.260378183 0.629468230 0.11015359	74 0.219171756 0.780828244
75 0.830290309 0.121098985 0.04861071	75 0.132041667 0.867958333
76 0.273696376 0.688721496 0.03758213	76 0.026750330 0.973249670
77 0.054745252 0.021967996 0.92328675	77 0.946978401 0.053021599
78 0.766059165 0.198997191 0.03494364	78 0.024660570 0.975339430
79 0.179937325 0.763086604 0.05697607	79 0.146908690 0.853091310
80 0.602480108 0.300838326 0.09668157	80 0.151551455 0.848448545
81 0.557991699 0.364985418 0.07702288	81 0.106930903 0.893069097
82 0.145344013 0.808750997 0.04590499	82 0.135634731 0.864365269
83 0.828744273 0.135306298 0.03594943	83 0.070516690 0.929483310
84 0.930821713 0.048831557 0.02034673	84 0.117905413 0.882094587
85 0.476496658 0.235930515 0.28757283	85 0.495790834 0.504209166

4. Kesimpulan

Algoritma Fuzzy C-Means dapat digunakan untuk mengelompokkan data pemilih untuk calon walikota Manado pada tahun 2015. Hal ini berguna sebagai acuan pemetaan bagi kandidat yang lain untuk melihat wilayah mana saja yang memberikan respon yang sangat baik ataupun kurang baik sebagai bagian dari demokrasi. Hasil yang dapat dilihat adalah dengan menggunakan 3 cluster dan 2 cluster didapatkan pengelompokkan 28, 18 dan 40 untuk 3 cluster dan pengelompokkan 46 dan 40 untuk 2 cluster. Hasil tersebut dapat dilihat dengan memperhatikan derajat keanggotaan fuzzy-nya.

5. Pustaka

- [1] Yang, M. S., Fuzzy Clustering And Its Applications, Department of Applied Mathematic Chung Yuan Christian University ChungLi 32023, Taiwan URL: <http://www2.math.cycu.edu.tw/TEACHER/MSYANG/fuzzy-e/fuzzy.htm>
- [2] Cannon, R. L., Dave, J. V., Bezdek, J. C. 1986. Efficient Implementation Of The Fuzzy C-Means Clustering Algorithms, *IEEE Transactions On Pattern Analysis And Machine Intelligence*. Vol. Pami-8, No. 2, March 1986
- [3] Bezdek, J.C. 1974. Cluster validity with fuzzy sets, *J. Cybernet.* 3:58–73.
- [4] Bezdek, J.C. 1974. Numerical taxonomy with fuzzy sets, *J. Math. Biol.* 1:57–71.
- [5] Bezdek, J.C. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
- [6] Bezdek, J.C. 1998. Pattern Recognition in Handbook of Fuzzy Computation, IOP Publishing Ltd., Boston, NY.