



Aplikasi Metode *Fuzzy Kernel K-Medoids* untuk Klasifikasi Kanker berdasarkan Konsentrasi Logam di dalam Darah

(Application of Fuzzy Kernel K-Medoids Method for Cancer Classification based on Metal Concentration in Blood)

ZUHERMAN RUSTAM^{1*}, ZUHELMI AZIZ²

¹Departemen Matematika-FMIPA Universitas Indonesia, Kampus UI Depok 16464.

²Fakultas Farmasi, Universitas Pancasila, Jl. Lenteng Agung, Jagakarsa, Jakarta 12640.

Diterima 2 September 2011, Disetujui 10 September 2011

Abstrak: Teknik klasifikasi telah diaplikasikan secara luas didalam bidang medis. Salah satunya adalah untuk klasifikasi kanker. Akurasi teknik ini sangat tinggi tergantung pada tipe data yang diproses (apakah data dapat atau tidak dapat dipisahkan secara linear) dan fungsi disimiliritas yang digunakan. Untuk mengatasi kedua hambatan tersebut dan meningkatkan akurasi teknik klasifikasi dibentuk suatu metode yang dinamakan *Fuzzy Kernel K-Medoids* (FKKM). Metode ini dapat digunakan untuk data yang dapat dipisahkan secara linear maupun tidak. Berdasarkan hasil penelitian terhadap konsentrasi logam Zn, Ba, Mg, Ca, Cu, dan Se dalam darah, didalam mendiagnosis penyakit kanker, FKKM memberikan hasil yang lebih baik dibandingkan dengan metode *Support Vector Machines*.

Kata kunci: klasifikasi, *fuzzy*, *kernel*, *medoids*, kanker.

Abstract: Classification technique has already been applied widely in the medical data. One of its applications is for classification of cancer. The accuracy of this technique highly depends on the type of data to be processed (whether the data are separable or non-separable) and the dissimilarity function used. To surmount those hindrances and to improve the accuracy of classification therefore a method named Fuzzy Kernel K-Medoids (FKKM). The method can be used for separable or non separable of data. Based on the research on the concentration data of Zn, Ba, Mg, Ca, Cu, and Se in blood in order to diagnose cancer, FKKM gives better result than the Support Vector Machines Method. This paper will discuss an application of the FKKM method on the concentration data of Zn, Ba, Mg, Ca, Cu, and Se in blood samples and compared with the Support Vector Machines Method for the diagnosis of cancer. Results showed that the FKKM method produced a better result than the Support Vector Machines Method.

Keywords: classification, fuzzy, kernel, medoids, cancer.

PENDAHULUAN

KLASIFIKASI dan *Clustering* merupakan salah satu teknik untuk memisahkan sejumlah data menjadi beberapa kelas atau kelompok sehingga data yang memiliki katakteristik serupa dikumpulkan dalam satu kelompok. Teknik klasifikasi telah digunakan untuk: klasifikasi leukemia^(1,2), tingkat keganasan *brain cancer*^(3,4,5), *protein fold*⁽⁶⁾, *HIV protease cleavage sites*⁽⁷⁾, jenis *G-protein coupled receptors*⁽⁸⁾, dan *protein secondary structure*^(9,10,11).

* Penulis korespondensi, Hp. 08129294176
e-mail: rustam@ui.ac.id

Diantara sekian banyak teknik klasifikasi yang telah digunakan, adalah untuk mendiagnosis kanker berdasarkan konsentrasi zat kimiawi dalam darah antara lain berdasarkan konsentrasi kandungan logam seng (Zn), barium (Ba), magnesium (Mg), kalsium (Ca, tembaga (Cu), dan selenium (Se). Salah satu teknik yang telah digunakan adalah *Support Vector Machines* (SVM) yang dilakukan oleh Ivanciuc⁽¹²⁾. Ivanciuc menggunakan data konsentrasi Zn, Ba, Mg, Ca, Cu, dan Se dalam darah yang diperoleh dari hasil pengukuran dengan *Inductive Couple Plasma* (ICP)⁽¹²⁾. Tingkat akurasi klasifikasi yang dihasilkan teknik SVM ini berkisar antara 51%-91%. Dengan menggunakan data yang digunakan oleh Ivanciuc



tersebut, akan dilakukan klasifikasi menggunakan metode *Fuzzy Kernel K-Medoids* (FKKM)⁽¹³⁾. FKKM merupakan perluasan dari metode *Fuzzy K-Medoids*⁽¹⁴⁾.

BAHAN DAN METODE

BAHAN. Bahan yang digunakan adalah data yang digunakan oleh Ivanciuc⁽¹²⁾. Tabel 1 menunjukkan konsentrasi Zn, Ba, Mg, Ca, Cu, dan Se dalam darah. Data tersebut diambil dari 74 pasien yang terdiri dari 42 pasien yang tidak mengidap kanker (Tabel 1A, kelas 1) dan 32 pasien yang mengidap kanker (Tabel 1B, kelas 2). Proses klasifikasi dilakukan dalam dua tahap yaitu Tahap *Training* dan Tahap Pengujian. Pada Tahap *Training* dibentuk model klasifikasi dengan menggunakan p% data dari masing-masing kelas, dan pada Tahap Pengujian dilakukan uji akurasi klasifikasi dengan menggunakan data yang tidak digunakan dalam Tahap *training*.

METODE. Setiap data ke i pada Tabel 1A dan 1B, dinotasikan dengan vektor $x_i = (Zn, Ba, Mg, Ca, Cu, Se)$. Data terdiri dari 2 kelas, dari masing-masing kelas dipilih p % data, data ini dinamakan data *training*. Sebuah data dari masing-masing kelas dipilih secara acak ditetapkan sebagai pusat *cluster* dan dinotasikan dengan v_1 dan v_2 secara keseluruhan. Pusat *cluster* akan selalu diperbaharui sampai tidak lagi mengalami perubahan secara signifikan. Tahap ini dinamakan juga tahap *training* atau tahap pembentukan model. Untuk memperbaharui pusat *cluster* dilakukan dengan cara sebagai berikut :

1. Tetapkan fungsi $K(x, y) = \exp(-\frac{\|x-y\|^2}{\sigma^2})$ dimana x dan y adalah vektor data dan σ bilangan positif.
2. Tetapkan masing-masing pusat *cluster* v_j , $j = 1, 2$ dimana j adalah kelas data.
3. Untuk setiap vektor data x_i , $i = 1, 2, \dots, n$: hitung $u_{ij} = \frac{d_{ij}^2}{\sum_{m=1}^c d_{im}^2}$, dimana $d_{ij}^2 = K(x_i, x_i) - 2K(x_i, v_j) + K(v_j, v_j)$.
4. Setiap pusat *cluster* v_j diganti dengan vektor data x , jika vektor data x tersebut membuat nilai $\sum_{i=1}^n u_{ij} d_{ij}^2$ minimum.
5. Ulangi langkah 3 dan 4 sampai masing-masing pusat *cluster* v_j tidak lagi berubah secara signifikan (selisih antara v_j saat ini dengan pada iterasi sebelumnya lebih kecil dari 0.05).

Setelah pusat *cluster* dan dihasilkan, dilakukan tahap pengujian untuk $(100 - p)\%$ dari data lainnya yang terdapat di masing-masing kelas. Data yang digunakan dalam tahap pengujian dinamakan data *testing*. Tahap pengujian dilakukan dengan cara sebagai berikut :

1. Pilih data *testing* x_i dari kelas ke j , $j = 1, 2$.
2. Hitung jarak dari data x_i ke pusat *cluster* v_1 dan v_2 dengan menggunakan formula :
 - i. Untuk Kelas 1: $d_{i1}^2 = K(x_i, x_i) - 2K(x_i, v_1) + K(v_1, v_1)$.
 - ii. Untuk Kelas 2: $d_{i2}^2 = K(x_i, x_i) - 2K(x_i, v_2) + K(v_2, v_2)$.
3. Pilih nilai minimum diantara kedua nilai d_{i1} dan d_{i2} .
4. Jika data x_i dari kelas ke 1 dan nilai minimum adalah d_{i1} , maka data x_i dikatakan telah diklasifikasikan dengan tepat. Hal yang sama dilakukan juga untuk data yang berada di kelas kedua.

Akurasi klasifikasi dihitung menggunakan persamaan, dimana t adalah banyaknya data yang tepat diklasifikasikan dan n adalah banyaknya data yang digunakan dalam tahap pengujian. Proses diatas diulang sebanyak 50 kali, untuk didapat nilai akurasi rata-rata untuk data *training* p%.

HASIL DAN PEMBAHASAN

Metode *Fuzzy Kernel K-Medoids* (FKKM) diatas digunakan untuk klasifikasi kanker berdasarkan konsentrasi Zn, Ba, Mg, Ca, Cu, dan Se dalam darah pasien kanker dan pasien normal pada Tabel 1A dan 1B. Percobaan dilakukan untuk data *training* sebanyak $p = 10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%$, dan 90% (dari seluruh data pada Tabel 1A dan 1B). Hasil lengkap percobaan tercantum pada Tabel 2. Pada Tabel 2, dicantumkan juga waktu komputasi rata-rata yang dibutuhkan mulai dari proses *training* sampai proses *testing* selesai dilakukan.

Dari Tabel 2 terlihat bahwa akurasi tertinggi, yaitu 100%, diperoleh jika menggunakan 90% data *training*. Hal ini menunjukkan bahwa dengan hanya mengandalkan data yang tersedia, *Fuzzy Kernel K-Medoids* mampu melakukan klasifikasi Kanker dan Non Kanker. Untuk pasien baru, dilakukan pengambilan data konsentrasi Zn, Ba, Mg, Ca, Cu, dan Se dalam darah. Data yang terkumpul tersebut langsung diolah pada tahap pengujian, tanpa harus melalui tahap *training*. FKKM akan menentukan apakah pasien tersebut mengidap kanker atau tidak dengan cara: data dinyatakan sebagai vektor x , kemudian ditentukan jarak dari x ke v_1 (d_{i1}) dan jarak dari x ke v_2 (d_{i2}) lalu hasil penentuan jarak tersebut dipilih nilai minimum. Jika jarak x ke v_1 yang minimum artinya pasien termasuk di dalam kelas 1 (pasien normal), tetapi jika jarak minimum adalah jarak dari x ke v_2 maka pasien termasuk di dalam kelas 2 (pasien penderita kanker).

Tabel 1A. Konsentrasi ($\mu\text{g/mL}$) Zn, Ba, Mg, Ca, Cu, dan Se untuk pasien normal (Kelas 1)⁽¹²⁾.

No	Zn	Ba	Mg	Ca	Cu	Se	Kelas
1	1.91	0.224	27.3	74.6	2.60	0.052	1
2	0.67	0.175	18.6	65.6	1.33	0.074	1
3	1.13	0.148	16.6	63.5	0.94	0.038	1
4	0.88	0.145	20.1	59.4	1.37	0.045	1
5	0.54	0.033	16.1	49.6	1.46	0.049	1
6	1.03	0.052	16.4	42.5	1.59	0.042	1
7	0.92	0.039	16.8	64.8	1.54	0.043	1
8	0.76	0.042	16.4	54.0	1.69	0.085	1
9	1.61	0.078	16.0	49.9	1.18	0.055	1
10	1.56	0.044	10.2	57.2	1.35	0.049	1
11	0.84	0.051	16.5	48.2	1.05	0.055	1
12	0.70	0.051	14.2	41.0	0.64	0.031	1
13	0.73	0.024	15.2	36.0	1.14	0.069	1
14	0.69	0.048	18.6	44.9	1.91	0.079	1
15	1.01	0.031	17.8	46.9	0.75	0.099	1
16	0.83	0.049	18.4	34.4	0.86	0.126	1
17	0.30	0.002	6.5	15.3	0.43	0.074	1
18	0.61	0.037	19.4	49.4	2.03	0.055	1
19	0.53	0.032	17.3	45.1	0.85	0.037	1
20	0.51	0.026	18.6	54.8	1.21	0.022	1
21	2.40	0.046	15.8	53.0	1.20	0.065	1
22	0.52	0.031	19.6	41.0	0.75	0.051	1
23	0.35	0.008	17.7	36.8	1.10	0.040	1
24	0.56	0.028	19.5	43.7	1.06	0.069	1
25	0.32	0.024	11.1	30.5	0.40	0.081	1
26	0.75	0.035	20.2	50.7	0.94	0.081	1
27	1.98	0.036	17.5	53.6	0.57	0.074	1
28	0.22	0.046	9.9	35.5	0.45	0.059	1
29	0.33	0.018	13.6	34.9	0.66	0.061	1
30	0.97	0.036	17.8	48.3	0.72	0.047	1
31	0.78	0.027	18.3	46.9	0.49	0.075	1
32	0.32	0.028	10.8	41.2	0.66	0.034	1
33	0.48	0.024	20.9	49.5	1.20	0.125	1
34	0.54	0.033	16.1	51.2	1.17	0.061	1
35	0.58	0.029	15.5	44.8	2.74	0.046	1
36	0.66	0.026	16.4	39.8	1.08	0.068	1
37	0.69	0.046	14.0	47.4	1.07	0.058	1
38	1.32	0.041	18.0	49.8	0.43	0.056	1
39	0.27	0.036	16.0	45.0	1.32	0.047	1
40	0.41	0.050	19.9	56.5	1.35	0.056	1
41	0.47	0.035	12.3	40.1	1.73	0.057	1
42	1.90	0.030	15.7	43.0	1.44	0.039	1

Tabel 1B. Konsentrasi ($\mu\text{g/mL}$) Zn, Ba, Mg, Ca, Cu, dan Se untuk pasien penderita Kanker (kelas 2)⁽¹²⁾.

No	Zn	Ba	Mg	Ca	Cu	Se	Kelas
1	0.65	0.005	19.9	78.4	0.58	0.088	2
2	0.63	0.012	20.7	81.4	1.02	0.066	2
3	0.52	0.032	19.4	74.1	0.68	0.059	2
4	0.66	0.007	23.7	86.5	1.01	0.070	2
5	0.64	0.023	20.4	78.4	0.94	0.073	2
6	0.67	0.026	20.2	85.6	1.09	0.071	2
7	0.67	0.022	19.4	85.1	0.84	0.052	2
8	0.67	0.006	19.6	76.7	0.85	0.081	2
9	0.73	0.013	17.8	74.7	0.84	0.074	2
10	0.51	0.010	16.4	77.2	0.88	0.084	2
11	0.54	0.017	18.6	74.7	1.14	0.081	2
12	0.70	0.009	21.6	78.8	0.97	0.071	2
13	0.41	0.013	17.4	60.1	0.69	0.075	2
14	0.55	0.017	20.8	71.2	0.98	0.083	2
15	0.58	0.012	21.7	71.4	0.74	0.068	2
16	0.46	0.007	18.2	68.3	0.81	0.096	2
17	0.44	0.035	21.1	71.6	1.31	0.057	2
18	0.54	0.013	22.5	79.5	0.86	0.075	2
19	0.48	0.006	18.3	71.9	0.76	0.046	2
20	0.49	0.034	17.7	68.9	0.73	0.088	2
21	0.47	0.021	15.2	66.3	1.00	0.067	2
22	0.45	0.163	16.9	65.6	0.80	0.067	2
23	0.49	0.008	15.6	63.0	0.74	0.072	2
24	0.43	0.143	15.3	57.0	0.83	0.049	2
25	1.76	0.243	12.5	52.1	0.64	0.082	2
26	0.70	0.008	13.8	63.8	0.84	0.052	2
27	2.20	0.067	15.6	65.8	0.96	0.066	2
28	0.58	0.032	9.2	41.8	0.98	0.087	2
29	1.09	0.010	10.8	42.3	0.60	0.076	2
30	0.55	0.201	14.3	60.0	0.57	0.065	2
31	0.60	0.228	18.8	72.6	0.99	0.069	2
32	1.08	0.238	20.2	76.4	0.95	0.060	2

Tabel 2. Akurasi rata-rata dan waktu komputasi rata-rata klasifikasi Kanker dan Non Kanker menggunakan FKKM berdasarkan konsentrasi Zn, Ba, Mg, Ca, Cu, dan Se dalam darah.

Data Training (%)	Akurasi (%)	Waktu Komputasi (detik)
10	90.8	0.31
20	89.7	0.16
30	90.2	0.21
40	93.2	0.26
50	94.6	0.35
60	89.3	0.48
70	90.5	0.59
80	92.9	0.7
90	100	0.84

SIMPULAN

Berdasarkan data konsentrasi Zn, Ba, Mg, Ca, Se dalam darah pasien dapat disimpulkan bahwa *Fuzzy Kernel K-Medoid* dapat digunakan untuk menetapkan apakah seseorang menderita kanker atau tidak. Metode yang dibahas ini merupakan alat bantu untuk dokter dan ahli kesehatan lainnya.

DAFTAR PUSTAKA

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek., Mesirov JP, *et al.* Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*.1999. 286: 531-37.
2. Sumati V, Singh L, Kumar S. Application of volutionary fuzzy neural network in predicting cancer subgroups. XXXII National Systems ConferenceNSC. 2008. 37-42.
3. Fikri A, Rustam Z, Pandelaki J Brain cancer (Astrocytoma) clustering menggunakan fuzzy C-Means. Prosiding Seminar Nasional Matematika 2010. Depok,4 Februari 2010: 271-7.
4. Wibowo AP , Rustam Z, Pandelaki J. Clustering brain cancer menggunakan possibilistik c-means. Prosiding Seminar Nasional Matematika 2010,Depok. 2010: 289-92.
5. Krismanti A, Rustam Z, Pandelaki J. Aplikasi spherical k-means pada pengklasifikasian brain cancer. Prosiding Seminar Nasional Matematika 2010,Depok. 2010:293-7.
6. Dehzangi A, Phon-Amnuaisuk S, and Dehzangi O. Using random forest for protein fold prediction problem: An empirical study. *J Information Science and Engineering*. 2010. (26): 1941-56.
7. Andonie R, Fabry-Asztalos L, Abdul-Wahid C.B. Abdul-Wahid S, Barker GI, Magill LC. Fuzzy ARTMAP prediction of biological activities for potential HIV-1 protease inhibitors using a small molecular data set. *EEE/ACM Transactions on Computational Biology and Bioinformatics*. 2011. 8(1): 80-93.
8. Bhasin M and Raghava GPS. GPCR pred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors, *Nucleic Acids Research*. 2004. (32): 383-9.
9. Cai CZ, Han LY, Ji ZL, Chen X and Chen YZ. SVM-Prot:Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res*. 2003. (31): 3692-7.
10. Cai CZ, Han LY, Ji ZL. and Chen YZ. Enzyme family classification by support vector machines. *Proteins*. 2004. (55): 66-76.
11. Patrizi G, Cifarelli C, Losacco V, and Patrizi G. Secondary structure classification of isoform protein markers in oncology, *Mathematical Approaches to Polymer Sequence Analysis and Related Problems* 2011. 7-67.
12. Ivanciuc O. Support vector mchines for cancer diagnosis from the Blood Concentration of Zn, Ba, Mg, Ca, Cu, and Se. *Internet Electronic Journal of Molecular Design*. 2002. (I): 418-27.
13. Rustam Z . Algoritma Fuzzy Kernel K-Medoids untuk klasifikasi data multikelas. Prosiding Seminar Nasional Teknologi Informasi 2010. Universitas Tarumanagara Jakarta. 2010. 75-9.
14. Krishnapuram R, Joshi A, and Yi L. A fuzzy relative of the medoids algorithm with application to web document and snippet clustering. *EEE International Fuzzy Systems Conference*, Seoul, Korea. 1999. 1281-6.