

Clustering Self Organizing Maps (SOM) dengan perbaikan missing value menggunakan Naïve Bayes Imputation pada pengelompokan wilayah Penyandang Masalah Kesejahteraan Sosial (PMKS)

by Triando Hamonangan Saragih

Submission date: 07-Mar-2022 08:36PM (UTC+0700)

Submission ID: 1778385604

File name: Journal_Research_Noor_Hidayah.docx (197.68K)

Word count: 2661

Character count: 16618



2 Clustering Self Organizing Maps (SOM) dengan perbaikan missing value menggunakan Naïve Bayes Imputation pada pengelompokan wilayah Penyandang Masalah Kesejahteraan Sosial (PMKS)

Clustering Self Organizing Maps (SOM) with the improvement of missing value using Naïve Bayes Imputation on clustering regions with social welfare problems (PMKS)

Noor Hidayah, Muliadi^{*)}, Irwan Budiman, Dodon Turianto Nugrahadi, Rudy Herteno

Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat
Jl. Jend. A. Yani km.36, Kampus ULM, Banjarbaru, Indonesia 70714

Cara sitasi: N. Hidayah, M. Muliadi, I. Budiman, D. T. Nugrahadi, and R. Herteno, "Clustering Self Organizing Maps (SOM) dengan perbaikan missing value menggunakan Naïve Bayes Imputation pada pengelompokan wilayah Penyandang Masalah Kesejahteraan Sosial (PMKS)," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. x, pp. xx-xx, 2021. doi: 10.14710/jtsiskom.2022.xxxxx [Online].

Abstract – People with Social Welfare Problems (PMKS) are problems in community groups who have difficulties in carrying out social functions. The study was conducted to determine the characteristics of PMKS problems in the South Kalimantan region. How to find out these problems by doing clustering. The clustering method used is Self Organizing Maps (SOM) and filling in the blank Naïve Bayes Imputation (NBI) data which is compared to filling in the Statistical Method (Mean, Median, Mode). The steps taken are starting with filling in the blank data with NBI and Statistical Methods. Then the clustering process was carried out using SOM with the cluster results evaluated using the Davies Bouldin-Index (DBI). The results obtained are improvements in NBI occupying the best clustering results with the smallest value of 0.032 with the division of 2 clusters.

Keywords – persons with social welfare (PMKS); Self Organizing Maps (SOM); Naïve Bayes Imputation (NBI); Statistical Methods

Abstrak – Penyandang Masalah Kesejahteraan Sosial (PMKS) merupakan permasalahan pada kelompok masyarakat yang memiliki kesulitan dalam menjalankan fungsi sosial. Penelitian dilakukan untuk mengetahui bagaimana karakteristik permasalahan PMKS di wilayah Kalimantan Selatan. Cara mengetahui permasalahan tersebut dengan melakukan clustering. Metode clustering yang digunakan adalah Self Organizing Maps (SOM) dan pengisian data kosong Naive Bayes Imputation (NBI) yang mana dibandingkan dengan pengisian Metode Statistik (Mean, Median, Modus). Tahapan yang dilakukan dimulai dengan pengisian data kosong dengan NBI dan Metode Statistik. Kemudian dilakukan proses

clustering menggunakan SOM dengan hasil cluster dievaluasi menggunakan Davies Bouldin-Index (DBI). Hasil yang didapatkan adalah perbaikan NBI menempati hasil clustering terbaik dengan nilai terkecil yaitu 0.032 dengan pembagian 2 cluster.

Kata kunci – Penyandang Masalah Kesejahteraan Sosial (PMKS); Self Organizing Maps (SOM); Naïve Bayes Imputation (NBI); Metode Statistik

I. PENDAHULUAN

2
Di era globalisasi saat ini, permasalahan kesejahteraan sosial sering dijumpai di Indonesia. Hal tersebut disebabkan oleh perubahan kehidupan masyarakat yang dipicu oleh peningkatan kebutuhan hidup, persaingan hidup yang ketat, keterbatasan dan ketidakmampuan untuk beradaptasi. Peningkatan jenis dan kualitas dari masalah kesejahteraan semakin lama semakin kompleks, sehingga dalam melakukan penanganan harus dilakukan dengan koordinasi yang lebih [1]. Penyandang Masalah Kesejahteraan Sosial (PMKS) merupakan kelompok perseorang, keluarga dan/atau masyarakat yang tidak bisa melaksanakan fungsi sosialnya dikarenakan adanya suatu hambatan. Sehingga mereka hidup dengan kebutuhan jasmani maupun rohani yang tidak terpenuhi [2]. Tentunya setiap daerah memiliki perbedaan karakteristik ketika dikelompokkan. Salah satu cara yang dapat dilakukan untuk mengetahui perbedaan tersebut yaitu dengan melakukan pengelompokan wilayah berdasarkan fitur data dari PMKS agar kebijakan selanjutnya dapat tepat sasaran [3].

Clustering adalah proses untuk menempatkan sekumpulan record data kedalam satu himpunan atau kelompok yang disebut cluster, sehingga dalam satu cluster memiliki record data dengan karakteristik yang sama dan berbeda dengan cluster lainnya. Permasalahan dasar dari clustering adalah bagaimana membagi

^{*)} Penulis korespondensi (Muliadi)
Email: muliadi@ulm.ac.id

sekumpulan data yang memiliki kesamaan semirip mungkin ke dalam satu *cluster* [4].

Algoritma *Self Organizing Maps* (SOM) awal mulanya dikenalkan oleh Profesor Teuvo Kohonen tahun 1996 [5],[6] dengan teknik *training artificial neural network*, yang menggunakan basis *winner takes all*, meskipun berbasis ANN, SOM tidak menggunakan target kelas untuk setiap data [7] namun SOM memerlukan penentuan laju pembelajaran fungsi serta iterasi yang diinginkan dalam proses pengelompokannya [8]. Pada proses penerapan algoritma SOM perlu dilakukan validasi cluster untuk menentukan apakah suatu model cluster tersebut baik [9]. SOM memiliki sejumlah keunggulan yaitu tidak memerlukan asumsi tentang distribusi variabel, lebih mudah diimplementasikan dan mampu memecahkan masalah dengan kompleksitas yang tinggi, serta efektif dalam menangani data *noise* dan data *missing* [10].

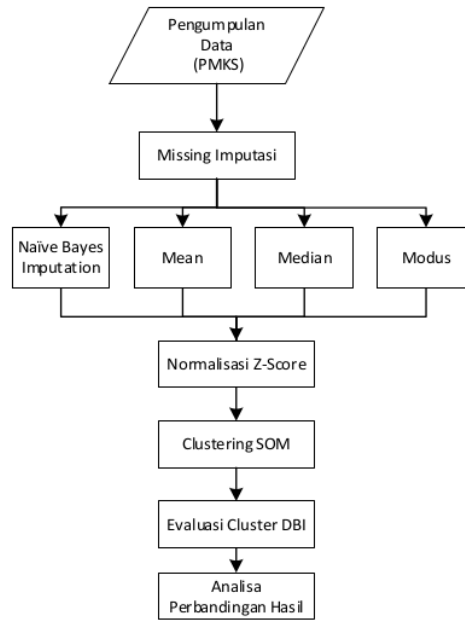
Kasus yang sangat sering terjadi dalam praktik adalah banyaknya nilai hilang atau kosong di dalam dataset yang disebut *missing value*. *Missing Value* merupakan masalah dalam menjaga suatu kualitas data. *Missing value* akan menjadi masalah ketika jumlah data dalam jumlah besar [11],[12], sejumlah penelitian menyelidiki bias dan efesian data akan berkurang jika diatas 50% [13]. Ada beberapa pendekatan statistik yang banyak digunakan untuk menangani nilai-nilai yang hilang dari kumpulan data seperti mean, median dan modus [14], selain menggunakan metode statistik dapat juga dilakukan dengan model imputasi prediksi menggunakan algoritma machine learning, seperti Naïve Bayes Imputation (NBI) untuk penanganan data yang hilang dengan menggunakan data yang lengkap untuk memprediksi nilai suatu variabel yang hilang pada data yang tidak lengkap. [15],[16].

Penelitian ini bertujuan untuk imputasi data PMKS menggunakan metode machine learning (NBI) dan metode statistik (Mean, Median dan Modus) yang mana keakuratan perbaikan data ini akan dilihat dari hasil perbandingan clustering menggunakan SOM yang mana hasil clustering tersebut akan di evaluasi menggunakan DBI. Selain itu penelitian ini bertujuan untuk menentukan prioritas wilayah yang perlu segera dilakukan penanganan berdasarkan urgensi dari setiap wilayah.

II. METODE PENELITIAN

Kerangka kerja penelitian ini ditunjukkan pada Gambar 1. Tahapan pertama yang dilakukan adalah melakukan pengumpulan dataset Penyandang Masalah Kesejahteraan Sosial (PMKS) yang diambil dari Laporan Rencana Strategis (RENSTRA) 2021 Dinas Sosial Provinsi Kalimantan Selatan yang terdiri dari 13 data dan 27 fitur. Fitur tersebut ditunjukkan pada Tabel 1.

Tahapan kedua adalah melakukan prediksi data kosong dengan menggunakan Metode Naïve Bayes Imputation dan Metode Statistik (Mean, Median dan Modus) sehingga akan dihasilkan 4 dataset dengan hasil imputasi yang berbeda.



Gambar 1. Kerangka Penelitian

Tabel 1. Fitur pada dataset Penyandang Masalah Kesejahteraan Sosial (PMKS)

No	Atribut
1	Anak balita terlantar
2	Anak terlantar
3	Anak yang berhadapan dengan hukum
4	Anak jalanan
5	Anak dengan kedistabilisan (ADK)
6	Anak yang menjadi korban kekerasan atau diperlakukan salah
7	Anak yang memerlukan perlindungan khusus
8	Lanjut usia terlantar
9	Penyandang disabilitas
10	Tuna susila
11	Gelandangan
12	Pengemis
13	Pemulung
14	Kelompok minoritas
15	Bekas warga binaan lembaga pemasyarakatan (BWBPL)
16	Orang dengan HIV/AIDS (ODHA)
17	Korban penyalahgunaan napza
18	Korban trafficking
19	Korban tindak kekerasan
20	Pekerja migran bermasalah sosial (PMBS)
21	Korban bencana alam
22	Korban bencana sosial
23	Perempuan rawan sosial ekonomi
24	Fakir miskin
25	Keluarga bermasalah sosial psikologi
26	Komunitas adat terpencil (KAT)
27	Rumah tidak layak huni



Adapun proses prediksi data hilang dengan NBI dilakukan dengan langkah-langkah berikut :

1. Hitung nilai Mean pada setiap fitur

$$\text{Mean } (\mu) = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

2. Hitung nilai Standar Deviasi dari setiap fitur

$$s = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}} \quad (2)$$

3. Hitung nilai NBI dari setiap data yang tersedia dari tiap fitur

$$P(x_k | C_i) = \frac{1}{\sigma_{ik} \sqrt{2\pi}} e^{-\frac{(x_k - \mu_{ik})^2}{2\sigma_{ik}^2}} \quad (3)$$

4. Gunakan perumusan logaritma untuk memperbesar nilai NBI setiap data yang didapat.

$$P = (-\log(P))^{-1} = \frac{1}{-\log(P)} \quad (4)$$

5. Totalkan nilai NBI yang diperoleh dari fitur yang sedang dihitung

$$\text{Total} = \sum_{i=1}^n p \quad (5)$$

6. Hasil total nilai NBI digunakan sebagai nilai prediksi data hilang pertama pada fitur yang dihitung
7. Ulangi proses dari awal untuk memprediksi data hilang berikutnya pada fitur
8. Perhitungan NBI pada langkah 3, data hasil prediksi sebelumnya tidak dimasukkan dalam perhitungan

Tahapan ketiga melakukan normalisasi atau penyamaan rentang pada dataset setelah dilakukan proses penanganan data kosong atau *missing value*. Proses normalisasi ini dilakukan pada dataset *missing value* yang telah di tangani menggunakan metode statistika yaitu Mean, Median dan Modus dan dataset yang telah di tangani oleh metode *Machine Learning* yaitu Naïve Bayes Imputation (NBI). Proses normalisasi dilakukan dengan menggunakan algoritma Z-Score.

$$\text{Nilai Baru} = \frac{\text{nilailama} - \text{mean}}{\text{stdev}} \quad (6)$$

Tahapan keempat melakukan proses clustering menggunakan Self Organizing Maps (SOM) dengan uji coba iterasi mulai dari iterasi 100 hingga 1000 pada setiap dataset. Algoritma SOM sendiri bekerja dimulai dari penentuan *learning rate*, ukuran ketetanggaan awal dan inialisasi vektor pada bobot awal. Secara sederhana algoritma SOM ditentukan 1 *winner* data dari seluruh data yang kemudian pusat cluster yang paling dekat dengan *winner* data bergerak mendekati titik *winner*. Setelah semua data berada diposisi yang baru dilakukan proses perhitungan jarak antara *winner* data dengan setiap pusat data yang kemudian proses diulangi sampai iterasi benar-benar mencapai hasil maksimal.

Adapun proses clustering menggunakan SOM dilakukan dengan langkah-langkah berikut :

1. Inialisasi bobot w_{ij} . Tentukan radius ketetanggaan (R), nilai *learning rate*, nilai faktor pengali *learning rate*, banyak *epoch*.
2. Melakukan perulangan tahapan sampai dengan jumlah iterasi dengan langkah 3-7.
3. Melakukan pada masing-masing data masukan x_i dengan langkah 4-6.
4. Perhitungan jarak seperti pada persamaan

$$d_j = \sum_{i=0}^n (w_{ij} - x_i)^2 \quad (7)$$

5. Mencari nilai yang minimum pada index sebanyak *neuron*.
6. Fungsi *update* bobot di sesuaikan dengan nilai radius dalam persamaan

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha [x_i - w_{ij}(\text{old})] \quad (8)$$

7. Perbaharui nilai laju pembelajaran

$$a(\text{new}) = \beta * a(\text{old}) \quad (9)$$

Langkah selanjutnya adalah dengan melakukan proses pengukuran kualitas cluster. Proses ini dilakukan untuk mengukur kualitas pada uji coba pembagian cluster dan menentukan mana pembagian dengan cluster terbaik. Tahap pengukuran kualitas cluster ini menggunakan *Davies Bouldin Index* (DBI) yang melakukan perhitungan dengan nilai DBI terkecil nantinya akan menjadi cluster terbaik.

Perumusan *Davies Bouldin Index* (DBI) dapat dilihat dalam langkah-langkah berikut :

1. Cari nilai rata-rata tiap cluster
2. Hitung variansi data dari masing-masing cluster

$$\text{var}(x) = \frac{1}{N_{maks}-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (10)$$

3. Cari R max dari langkah 2 dan 3

$$R_i = j=1, k, i \neq j R_{ij} \quad (11)$$

$$R_{ij} = \frac{\text{var}(c_i) + \text{var}(c_j)}{\|c_i - c_j\|} \quad (12)$$

4. Hitung nilai DBI

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (13)$$

5. Hasil akhir adalah nilai DBI dari cluster

Setelah dilakukan pengukuran hasil kualitas cluster menggunakan *Davies Bouldin Index* (DBI) tahap selanjutnya adalah melakukan analisa hasil pengaruh perbaikan *missing value* menggunakan Naïve Bayes Imputation, Mean, Median dan Modus untuk menentukan seberapa besar pengaruh perbaikan data kosong pada hasil cluster yang kemudian ditahapan ini

menentukan yang mana cluster dengan hasil terbaik pada pengelompokan wilayah Penyandang Masalah Kesejahteraan Sosial (PMKS) Kalimantan Selatan.

III. HASIL DAN PEMBAHASAN

Proses imputasi pada data PMKS Kalimantan Selatan terdapat 3 fitur yang tidak dapat diimputasi karena jumlah data hilang lebih dari 90% pada setiap fitur, jika dilakukan proses prediksi tentu akan menghasilkan nilai prediksi yang bias. Ketiga fitur yang tidak dapat diimputasi tersebut terdapat pada Tabel 2 sehingga data yang digunakan pada penelitian ini terdiri dari 13 data dan 24 fitur.

Clustering yang dilakukan menggunakan SOM menggunakan nilai alpha standar (0.05, 0.01) dengan dimensi 3 x 3, Topology Rectangular dan percobaan iterasi dari 100 hingga 1000.

Pengujian mulai dari iterasi 100 hingga 1000 ini di evaluasi langsung menggunakan DBI. Adapun hasil clustering SOM pada data PMKS dengan perbaikan data untuk keempat metode tersebut dinyatakan dalam Tabel 3 untuk metode NBI, Tabel 4 untuk Mean, Tabel 5 untuk Median dan Tabel 6 untuk Modus.

Dari hasil clustering tersebut akan dipilih iterasi yang menghasilkan nilai paling optimal. Pada perbaikan data menggunakan NBI paling optimal berada di iterasi 300, Mean dan Median di iterasi 700, dan Modus di iterasi 100. Berdasarkan evaluasi tersebut hasil clustering yang optimal pada masing-masing iterasi ditunjukkan pada Tabel 7. Berdasarkan perolehan evaluasi optimal pada Tabel 7 dapat dilihat bahwa evaluasi clustering terbaik diperoleh pada perbaikan data menggunakan NBI di dua cluster dengan nilai evaluasi DBI terkecil yaitu 0.032 yang mana persebaran wilayah dengan pembagian dua cluster ini dapat dilihat pada Gambar 2.

Pada penelitian ini dapat ditentukan pembagian cluster terbaik yang dapat diimplementasikan terhadap data Penyandang Masalah Kesejahteraan Sosial (PMKS) wilayah Kalimantan Selatan dapat dibagi menjadi 2 cluster dengan perbaikan data menggunakan NBI, dengan hasil pembagian cluster dapat dilihat pada Tabel 8. Pada Tabel 8 tersebut memberikan hasil cluster pertama yang terdiri dari 8 anggota dan cluster kedua terdiri dari 5 anggota.

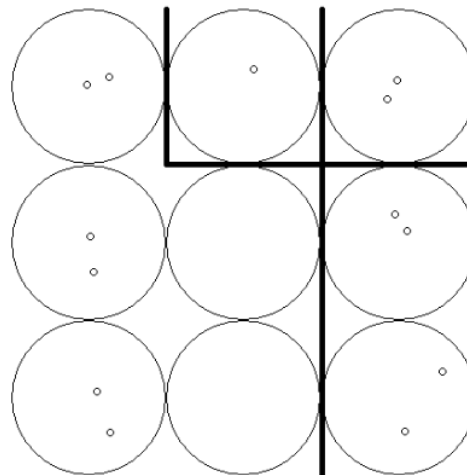
Berdasarkan hasil pembagian cluster tersebut dapat ditentukan tingkat prioritas wilayah melalui rata-rata setiap fitur berdasarkan pembagian cluster. Tingkat prioritas tersebut dapat dilihat pada Tabel 9 dari hasil rata-rata cluster yang diperoleh dapat terlihat tingkat rata-rata kasus dari setiap fitur selain itu didapatkan pula cluster kedua memiliki rata-rata lebih tinggi yang berarti wilayah pada cluster kedua memiliki tingkat prioritas tertinggi dibandingkan dengan cluster pertama.

Metode Naïve Bayes Imputation (NBI) berhasil melakukan proses imputasi data terbaik yang dapat dilihat dari nilai evaluasi cluster yang diperoleh, hal ini sesuai dengan proses imputasi lainnya yang menggunakan NBI [15], [16]. Selain itu hasil imputasi ini membuktikan bahwa NBI mampu melakukan proses

Tabel 2. Fitur pada dataset PMKS yang tidak dapat diimputasi

Kabupaten /Kota	Fitur		
	Korban Trafficking	Pekerja Migran Bermasalah Sosial	Rumah Tidak Layak Huni
Tanah Laut	(kosong)	(kosong)	(kosong)
Kotabaru	(kosong)	(kosong)	(kosong)
Banjar	(kosong)	4	(kosong)
Barito	(kosong)	(kosong)	(kosong)
Kuala Tapin	(kosong)	(kosong)	(kosong)
Hulu Sungai Selatan	1	(kosong)	(kosong)
Hulu Sungai Tengah	(kosong)	(kosong)	(kosong)
Hulu Sungai Utara	(kosong)	(kosong)	2818
Tabalong	(kosong)	(kosong)	(kosong)
Tanah Bumbu	(kosong)	(kosong)	(kosong)
Balangan	(kosong)	(kosong)	(kosong)
Banjarmasin	(kosong)	(kosong)	(kosong)
Banjarbaru	(kosong)	(kosong)	(kosong)

Clustering



Gambar 2. Rplot Visualisasi Clustering Data

imputasi lebih baik dari pada metode pendekatan statistik yang biasanya digunakan untuk menangani data yang hilang [14].

Kemampuan analisa cluster yang dilakukan oleh Algoritma *Self Organizing Maps* (SOM) dapat dilakukan



Tabel 3. Hasil evaluasi clustering SOM pengujian 100 hingga 1000 pada perbaikan data PMKS menggunakan NBI

Jumlah Cluster	Iterasi									
	100	200	300	400	500	600	700	800	900	1000
2	7.399	4.208	0.032	1.205	0.769	14.371	10.486	1.972	2.806	4.112
3	21.346	30.278	2.368	28.334	3.141	19.746	17.343	5.244	66.395	11.584
4	38.038	22.708	1.776	21.25	11.404	5.782	8.098	22.441	8.874	15.015
5	0.884	9.517	0.539	1.511	8.043	4.611	6.479	2.495	7.099	9.861

Tabel 4. Hasil evaluasi clustering SOM pengujian 100 hingga 1000 pada perbaikan data PMKS menggunakan Mean

Jumlah Cluster	Iterasi									
	100	200	300	400	500	600	700	800	900	1000
2	0.754	5.278	1.536	17.105	2.81	37.985	2.667	2.886	18.616	91.901
3	10.118	424.246	1.539	84.426	3.991	3.608	4.725	2.882	17.984	9.0423
4	7.588	6.937	0.924	0.304	4.515	1.255	0.798	0.243	0.244	586.371
5	6.071	0.937	0.739	0.729	4.019	1.004	0.148	0.149	0.587	0.543

Tabel 5. Hasil evaluasi clustering SOM pengujian 100 hingga 1000 pada perbaikan data PMKS menggunakan Median

Jumlah Cluster	Iterasi									
	100	200	300	400	500	600	700	800	900	1000
2	4.447	32.917	6.085	1.521	13.116	2.805	0.21	2.497	13.436	0.479
3	0.715	7.623	7.805	1.439	6.328	6.272	3.317	8.308	66.696	2.306
4	0.473	2.966	9.508	0.753	14.254	7.736	0.792	15.327	0.246	6.799
5	0.265	2.822	7.606	0.4	0.373	6.189	0.174	0.875	0.59	7.334

Tabel 6. Hasil evaluasi clustering SOM pengujian 100 hingga 1000 pada perbaikan data PMKS menggunakan Modus

Jumlah Cluster	Iterasi									
	100	200	300	400	500	600	700	800	900	1000
2	0.1261	163.476	74.19	8.755	9.415	12.117	1.433	1.706	1.356	0.179
3	1.776	10.026	2.759	7.15	6.613	30.966	1.323	7.89	1.125	5.011
4	0.686	8.934	18.385	0.341	10.758	5.763	0.86	11.46	0.531	3.758
5	0.198	7.147	25.972	1.848	0.59	1.67	1.186	1.033	0.781	0.16

Tabel 7. Evaluasi clustering optimal

No.	Jumlah Cluster	Evaluasi Cluster NBI Iterasi 300	Evaluasi Cluster Mean Iterasi 700	Evaluasi Cluster Median Iterasi 700	Evaluasi Cluster Modus Iterasi 100
1.	2	0.032	2.667	0.21	0.126
2.	3	2.368	4.725	3.317	1.776
3.	4	1.776	0.798	0.792	0.686
4.	5	0.539	0.148	0.174	0.198

Tabel 8. Anggota cluster dengan perbaikan data menggunakan NBI

Cluster	Jumlah Anggota	Anggota
1	8	Tanah Laut, Kota Baru, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong dan Tanah Bumbu
2	5	Banjar, Barito Kuala, Balangan, Banjarmasin dan Banjarbaru

dengan baik, meski dibebberapa percobaan pada iterasi tertentu hasil yang didapatkan kurang optimal, sehingga persebaran cluster di anggap kurang tepat. Maka dari itu perlu dilakukan pengujian iterasi untuk menemukan hasil paling optimal untuk dipilih sebagai persebaran terbaik. Kemampuan cluster SOM ini dapat mengatasi

permasalahan yang umumnya terjadi dalam proses cluster [4] selain itu SOM dapat membuktikan kemampuan clusternya sesuai dengan [7]-[10] dan dalam prosesnya dapat ditentukan tingkat prioritas penanganan berdasarkan rata-rata cluster yang diperoleh.



Tabel 9. Profilisasi cluster

Variabel	Rata-rata	
	Cluster 1	Cluster 2
Anak balita terlantar	84.58663308	81
Anak terlantar	922.875	331.2
Anak yang berhadapan dengan hukum	3.369992013	62.8973413
Anak jalanan	4.251800233	24.2020358
Anak dengan kedistabilan (ADK)	94.55864122	19.5327919
Anak yang menjadi korban kekerasan atau diperlakukan salah	2.613448301	11.7396748
Anak yang memerlukan perlindungan khusus	4.220958434	2.76385546
Lanjut usia terlantar	2419.125	1939
Penyandang disabilitas	962.5	1457.2
Tuna susila	37.64930233	73.694178
Gelandangan	3.659543958	30.2732453
Pengemis	16.69957177	64.4803817
Pemulung	7.009137553	93.740641
Kelompok minoritas	2.625853117	2.23865801
Bekas warga binaan lembaga pemasyarakatan (BWBPL)	100.9620889	121
Orang dengan HIV/AIDS (ODHA)	2.709017502	1.29938781
Korban penyalahgunaan napza	102.2124244	26.5124713
Korban tindak kekerasan	12.08195209	10.7237858
Korban bencana alam	1641.091818	1199.6989
Korban bencana sosial	173.1739529	2.39307397
Perempuan Rawan Sosial Ekonomi	874.75	1859.6
Fakir Miskin	7213.875	5415.2
Keluarga bermasalah sosial psikologi	69.49284652	1457.06208
Komunitas adat terpencil (KAT)	94.45712734	932.799101
TOTAL	14850.55111	15220.2516

IV. KESIMPULAN

Penggunaan metode *Naïve Bayes Imputation* (NBI) untuk penanganan data hilang menunjukkan keakuratan yang lebih baik dibandingkan dengan pendekatan metode statistik (Mean, Median, dan Modus) berdasarkan evaluasi clustering SOM yang diperoleh dari DBI yaitu sebesar 0.032. Proses clustering ini pula mendapatkan tingkat prioritas penanganan wilayah PMKS berdasarkan rata-rata cluster dengan cluster kedua mendapatkan prioritas pertama dibanding cluster satu.

DAFTAR PUSTAKA

[1] C. I. Basuki, A. Arwan, E. Muhammad, and A. Jonemaro, "Pengembangan Sistem Monitoring Penyandang Masalah Kesejahteraan Sosial

Berbasis Web (Studi Kasus : Dinas Sosial Kota Malang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 3, pp. 709–716, 2020.

[2] I. Hidayatin, S. Adinugroho, and C. Dewi, "Pengelompokan Wilayah berdasarkan Penyandang Masalah Kesejahteraan Sosial (PMKS) dengan Optimasi Algoritme K-Means menggunakan Self Organizing Map (SOM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 8, pp. 7524–7531, 2019, [Online]. Available: <http://jptiik.ub.ac.id/index.php/j-ptiik/article/download/5881/2852>.

[3] D. N. Firmansyah, S. Adinugroho, and B. Rahayudi, "Kesejahteraan Sosial Menggunakan Algoritme Self-Organizing Maps Dengan Perbaikan Missing Value K-Nearest Neighbors," *Jurnal Pengembangan Teknologi Informasi dan Ilmu*



- Komputer*, vol. 3, no. 7, pp. 7205–7213, 2019.
- [4] F. N. S. Damanik, "Implementasi Algoritma Self Organizing Mapping Dalam Pencarian Jodoh," *Jurnal Stiondo Profesional*, vol. VI, no. September, pp. 140–156, 2020.
- [5] A. F. Solikin, K. Kusriani, and F. W. Wibowo, "Analisis Cluster Data Interkomparasi Anak Timbangan dengan Algoritma Self Organizing Maps," *Jurnal Teknologi Informasi dan Sistem Informasi*, vol. 7, no. 2, pp. 435–448, 2021, doi: 10.28932/jutisi.v7i2.3698.
- [6] N. N. Halim and E. Widodo, "Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps," *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika dan Nilai Islami)*, vol. 1, no. 1, pp. 188–194, 2017, [Online]. Available: <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>.
- [7] N. Y. Kusrahman, I. Purnamasari, and L. M. Komputasi, "Optimasi Self-Organizing Map Menggunakan Particle Swarm Optimization untuk Mengelompokkan Desa / Kelurahan Tertinggal di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur (Studi Kasus : Data Potensi Desa Tahun 2018)," *Jurnal Eksponensial*, vol. 11, no. 2, pp. 139–144, 2020.
- [8] R. Umar, A. Fadlil, and R. R. Az-Zahra, "Self Organizing Maps(SOM) untuk Pengelompokkan Jurusan di SMK," *Khazanah Informatika. Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 2, p. 131, 2018, doi: 10.23917/khif.v4i2.7044.
- [9] I. IRWAN, A. Y. HASHARI, H. IHSAN, and A. ZAKI, "Penggunaan Self Organizing Map Dalam Pengelompokan Tingkat Kesejahteraan Masyarakat," *Jambura Journal Of Probability and Statistics*, vol. 1, no. 2, pp. 57–68, 2020, doi: 10.34312/jjps.v1i2.7266.
- [10] M. D. Kartikasari, "Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan," *Jambura Journal Of Mathematics*, vol. 3, no. 2, pp. 187–196, 2021, doi: 10.34312/jjom.v3i2.10942.
- [11] M. Hasyim and U. Y. Pasuruan, "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-NN Imputation Pada Algoritma C4.5," *Jurnal Resistor (Rekayasa Sistem Komputer)*, vol. 2, no. 2, 2019.
- [12] M. Gopal Krishna, N. Durgaprasad, S. Deepa Kanmani, G. Sravan Reddy, and D. Revanth Reddy, "Comparative analysis of different imputation techniques for handling missing dataset," *International Journal Of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 7, pp. 347–351, 2019.
- [13] P. Madley-Dowd, R. Hughes, K. Tilling, and J. Heron, "The proportion of missing data should not be used to guide decisions on multiple imputation," *Journal Clinical Epidemiology*, vol. 110, pp. 63–73, 2019, doi: 10.1016/j.jclinepi.2019.02.016.
- [14] S. I. Khan, A. Sayed, and L. Hoque, "SICE : an improved missing data imputation technique," *Journal Big Data*, vol. 7, pp. 7–37, 2020, doi: 10.1186/s40537-020-00313-w.
- [15] B. K. Khotimah and H. Suprajitno, "Modeling naïve bayes imputation classification for missing data," *IOP Conf. Series. Earth and Environmental Science*, 2019, doi: 10.1088/1755-1315/243/1/012111.
- [16] A. S. Arifianto, D. R. Hartadi, and A. N. N. ES, "Prediksi Missing Imputation Untuk Data Penyebaran Menggunakan Naïve Bayes," *Jurnal Teknologi Informasi dan Terapan*, vol. 3, no. 1, pp. 300–306, 2016.



©2021. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Clustering Self Organizing Maps (SOM) dengan perbaikan missing value menggunakan Naïve Bayes Imputation pada pengelompokan wilayah Penyandang Masalah Kesejahteraan Sosial (PMKS)

ORIGINALITY REPORT

19%

SIMILARITY INDEX

18%

INTERNET SOURCES

1%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	jtsiskom.undip.ac.id Internet Source	6%
2	repository.ub.ac.id Internet Source	5%
3	silinda.bekasikab.go.id Internet Source	3%
4	www.researchgate.net Internet Source	1%
5	download.garuda.ristekdikti.go.id Internet Source	1%
6	ejurnal.ung.ac.id Internet Source	1%
7	Moch. Lutfi, Mochamad Hasyim. "PENANGANAN DATA MISSING VALUE PADA KUALITAS PRODUKSI JAGUNG DENGAN MENGGUNAKAN METODE K-NN IMPUTATION	1%

PADA ALGORITMA C4.5", Jurnal RESISTOR (Rekayasa Sistem Komputer), 2019

Publication

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On