

IMPLEMENTASI DATA MINING UNTUK KARTU INDONESIA SEHAT BAGI MASYARAKAT KURANG MAMPU MENGGUNAKAN METODE CLUSTERING PADA DINAS SOSIAL KOTA PALEMBANG

Wulan Lestari¹, Fatoni², Hutrianto³

^{1,2,3}Fakultas Ilmu Komputer, Universitas Bina Darma, Palembang, Indonesia

E-mail: ¹lestariwulan699@gmail.com, ²fatoni@binadarma.ac.id, ³hutrianto@binadarma.ac.id

Abstract

The Social Service Office of Palembang City is a government institution that is responsible for carrying out development in the field of social welfare, which covers all programs and activities that are intended to realize, foster and maintain, restore and develop social welfare carried out together as the responsibility of the government and society. In conducting data processing to determine the capable and underprivileged people as recipients of the Healthy Indonesia Card (KIS) with a process that takes a long time, given the large number of population data to be selected, especially in the city of Palembang. Then it can be raised into the research, namely "Implementation of Data Mining for Healthy Indonesia Cards for the Underprivileged Community Using the Clustering Method".

Keywords: Data Mining, Social Service, Clustering

Abstrak

Dinas Sosial Kota Palembang merupakan lembaga pemerintahan yang bertanggung jawab melaksanakan pembangunan di bidang kesejahteraan sosial, yang mencakup semua upaya program dan kegiatan yang ditunjukkan untuk mewujudkan, membina dan memelihara, memulihkan dan mengembangkan kesejahteraan sosial yang dilaksanakan bersama sebagai tanggung jawab pemerintah dan masyarakat. Dalam melakukan pengolahan data untuk menentukan masyarakat mampu dan kurang mampu sebagai penerima Kartu Indonesia Sehat (KIS) dengan proses yang membutuhkan waktu yang cukup lama, mengingat banyaknya jumlah data penduduk yang akan diseleksi, khususnya di wilayah Kota Palembang. Maka dapat diangkat ke dalam penelitian yaitu "Implementasi Data Mining untuk Kartu Indonesia Sehat Bagi Masyarakat Kurang Mampu Menggunakan Metode Clustering".

Kata Kunci: Data Mining, Dinas Sosial, Clustering

1. Pendahuluan

Dinas Sosial merupakan salah satu kantor pemerintahan yang ada di Palembang, yang terdiri dari beberapa bidang salah satunya perlindungan dan jaminan sosial yaitu mengenai Kartu Indonesia Sehat (KIS) merupakan kartu identitas peserta Jaminan Kesehatan SJSN (JKN) bagi penduduk Indonesia, khususnya fakir miskin dan tidak mampu serta iurannya dibayarkan oleh pemerintah, dengan pemakaian dapat dilakukan di mana saja, baik di klinik, puskesmas atau di rumah sakit mana pun yang ada di Indonesia.

Proses yang ada di Dinas Sosial saat ini dalam memberikan Kartu Indonesia Sehat, masyarakat membawa surat keterangan tidak mampu disertakan Kartu Keluarga, KTP dan akta kelahiran, kemudian melakukan pendaftaran dengan mengisi formulir sebagai peserta penerima KIS, selanjutnya tim survei yang akan ke rumah mengecek langsung kondisi ekonomi apakah layak atau tidak dimasukan ke KIS miskin.

Dikatakan layak apabila peserta KIS sesuai dengan kriteria terdiri dari jumlah penghasilan, pekerjaan, kondisi tempat tinggal, dan tidak terdaftar sebagai jaminan kesehatan BPJS. Jika dinyatakan layak, maka akan diajukan ke pusat sebagai peserta penerima Kartu Indonesia Sehat (KIS). Sedangkan pada proses pengolahan data penerima KIS pada Dinas Sosial dilakukan menggunakan Aplikasi *Microsoft Excel*, sehingga terdapat kerumitan dalam melakukan pengolahan data untuk menentukan masyarakat mampu dan kurang mampu sebagai penerima Kartu Indonesia Sehat (KIS) dengan proses yang membutuhkan waktu yang cukup lama, mengingat banyaknya jumlah data penduduk yang akan diseleksi, khususnya di wilayah Kota Palembang.

Berdasarkan permasalahan tersebut, maka data kependudukan perlu dikelompokkan terlebih dahulu kedalam kelompok berdasarkan kategori layak dan tidak layak sebagai penerima Kartu Indonesia Sehat (KIS) dengan pengolahan data dapat dilakukan dengan metode *cluster*. metode *cluster* adalah metode yang digunakan untuk mengelompokkan data kedalam satu atau lebih kelompok yang mempunyai karakter yang mirip. Sehingga dengan adanya pengelompokan ini akan diketahui penduduk yang ekonominya dinilai mampu dan tidak mampu sebagai penerima Kartu Indonesia Sehat (KIS).

2. Metodologi Penelitian

2.1. Metode Analisis Data

Metode yang digunakan oleh penulis adalah metode deskriptif yang merupakan suatu metode penelitian yang menggambarkan semua data atau keadaan obyek penelitian (seseorang, lembaga, masyarakat dan lain-lain) kemudian dianalisis dan dibandingkan berdasarkan kenyataan yang sedang berlangsung pada saat ini dan selanjutnya mencoba untuk memberikan pemecahan masalahnya. Penelitian deskriptif mempelajari masalah-masalah dalam masyarakat, baik itu menyangkut tata cara, situasi hubungan, sikap perilaku, cara pandang dan pengaruh-pengaruh dalam suatu kelompok masyarakat. (Widi, 2010).

2.2 Data Mining

Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007). Data mining, sering juga disebut sebagai *knowledge discovery in database (KDD)*. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007). Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu – ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing (Han, 2006). Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi (Kelulusan Mahasiswa, Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer, & Matematika, 2010). Data yang dibutuhkan dalam jumlah besar (Witten, 2005). Karakteristik data mining sebagai berikut

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi (Davies, 2004). Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata *data mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu data mining sebenarnya memiliki akar yang panjang dari

bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan database. Beberapa metode yang sering disebutsebut dalam literatur data mining antara lain clustering, classification, association rules mining, neural network, genetic algorithm dan lain-lain (Pramudiono, 2007).

2.3 Algoritma C-Means

Fuzzy C-means Clustering (FCM) atau 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 dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau *cluster* ditentukan oleh derajat keanggotaannya. Konsep dasar FCM, pertama kali adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap *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 menuju lokasi yang tepat dan nilai keanggotaan akan menuju ke *cluster* yang tepat. Perulangan ini didasari pada minimasi fungsi objektif dan maksimum iterasi yang telah ditentukan oleh *user*. Langkah-langkah dalam Algoritma *Fuzzy C-Means* adalah sebagai berikut : (Nurjanah, Farmadu, & Indriani, 2014)

1. Input Data
Input data yang akan di *cluster* X, berupa matriks berukuran $n \times m$ (n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, \dots, m$)
2. Tentukan :
 - a. Jumlah *cluster* = c ;
 - b. Pangkat / Pembobot = w ;
 - c. Maksimum iterasi = MaxIter ;
 - d. Error terkecil yang diharapkan = ξ
 - e. Fungsi obyektif awal = $P_0 = 0$;
 - f. Iterasi awal = $t = 1$;
3. Bangkitkan Bilangan *Random*
Bangkitkan bilangan *random* μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; sebagai elemen – elemen matriks partisi awal U.
Hitung jumlah setiap kolom (atribut) :
dengan $j = 1, 2, \dots, m$.
Hitung :
4. Hitung Pusat *Cluster* ke- k
Hitung pusat *cluster* ke- k : V_{kj} , dengan $k = 1, 2, \dots, c$; dan $j = 1, 2, \dots, m$.
5. Hitung fungsi objektif
Hitung fungsi objektif pada iterasi ke- t :
6. Hitung perubahan matrik
Hitung perubahan matrik partisi :
dengan $i = 1, 2, \dots, n$; dan $k = 1, 2, \dots, c$;
7. Cek Kondisi Berhenti
 - a. Jika ($|P_t - P_{t-1}| < \xi$) atau ($t > \text{MaxIter}$) maka berhenti;
 - b. Jika tidak: $t = t + 1$, ulangi langkah 4.

2.4 Waikato Environment for Knowledge Ananlysis (WEKA)

Weka adalah aplikasi data mining *open source* yang berbasis java. Aplikasi ini dikembangkan pada tahun 1994 dan pertama kali oleh sebuah universitas di selandia baru yang bernama universitas Waikato. aplikasi weka ini mulai menjadi aplikasi data mining *open source* yang sangat terkenal pada awal perkembangannya. Hal itu dikarenakan aplikasi weka memiliki kelebihan yang tidak dimiliki aplikasi data mining lainnya yaitu pada aplikasi weka terdapat banyak algoritma yang terdapat didalam aplikasi dan disertai juga machine learning, lalu juga dalam penggunaannya tidak terlalu rumit sehingga tidak menyulitkan penggunanya, dan ditambah dengan kelebihan lainnya bahwa algoritma-algoritma yang terdapat pada aplikasi weka selalu baru dan ter update, sehingga dengan beberapa kelebihan (Feri, dkk,

2010). Aplikasi weka tersebut digunakan banyak perusahaan dalam dunia bisnis untuk membantu dalam usaha bisnisnya, akademik pun juga tak mau ketinggalan untuk menggunakan aplikasi weka.

Aplikasi weka merupakan *software* yang terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling. Inti dari kekuatan pada aplikasi weka terletak pada algoritma yang makin lengkap dan canggih, namun walaupun begitu canggihnya aplikasi weka tersebut, letak keberhasilan data mining tetap ditentukan oleh manusia itu sendiri sebagai penggunaannya/user. Keberhasilan data mining itu berdasarkan pada pengumpulan data yang berkualitas tinggi, penggunaan model dan algoritma yang tepat. Sehingga secanggih serta sehebat apapun aplikasi data mining, tanpa kemampuan sang penggunaannya untuk menerapkannya maka tidak akan menghasilkan data mining yang tepat dan bermanfaat.

3. Hasil dan Pembahasan

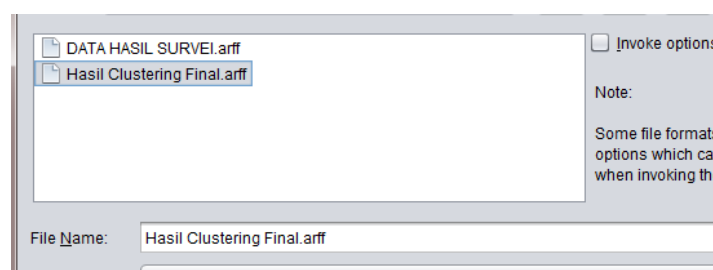
3.1. Pengujian Dengan WEKA

Setelah data absensi telah melalui proses KDD dan telah di simpan dalam format *csv* maka selanjutnya adalah Pengujian menggunakan *WEKA* untuk mendapatkan hasil



Gambar 1 Homepage WEKA

Gambar 1 di tampilkan halaman utama aplikasi *Weka*. Pada halaman awal ini bisa di lihat beberapa tools yang bisa di gunakan untuk pemrosesan data mining. Semua *tools* memiliki fungsi masing-masing.

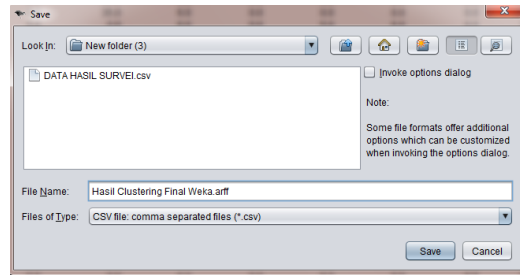


Gambar 2 Memilih file hasil clustering

Gambar 2 merupakan pengelompokkan data penduduk sesuai dengan *cluster* masing – masing, dari sini sudah bisa didapatkan jika penduduk yang masuk pada *cluster* 0 merupakan penduduk yang paling masuk kedalam aspek penduduk miskin di banding *cluster* lainnya yaitu *cluster* 1, *cluster* 2, dan *cluster* 3.

3.2 Pembahasan Memasukan Data

Setelah data sampel telah selesai diolah tahap selanjutnya adalah Pengujian menggunakan *WEKA*. Pada aplikasi *weka* pilih *Open file* dan pilih jenis *file*, dalam hal ini data yang diunakan berjenis *csv*.



Gambar 3 Open File Csv

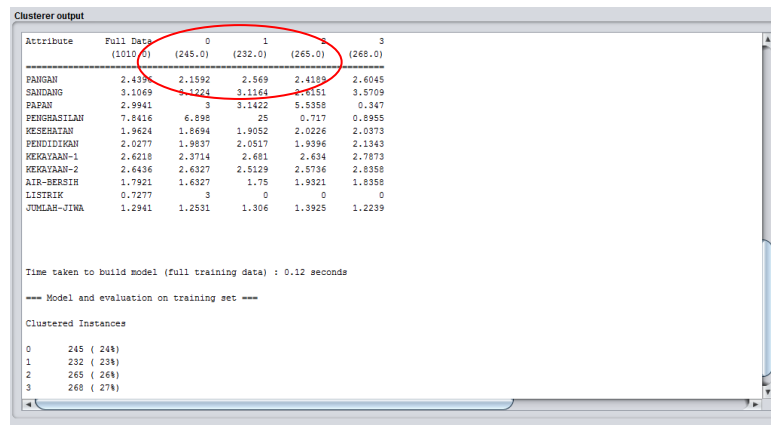
Gambar 3 di tampilkan satu jendela untuk memilih file yang akan di olah untuk data mining, dalam hal ini data yang akan di proses untuk data mining adalah data penduduk miskin.

Cluster	Jumlah Penduduk	Keterangan
Cluster 0 (24%)	245 Penduduk	Sangat Miskin
Cluster 1 (23%)	232 Penduduk	Hampir Sangat Miskin
Cluster 2 (26%)	265 Penduduk	Miskin
Cluster 3 (27%)	268 Penduduk	Mendekati Miskin

Gambar 4 Keterangan hasilclustering

3.3. Pembahasan Pemilihan Rumus

Pada tahap selanjutnya adalah penentuan metode yang akan dipakai dalam aplikasi *weka* tersebut. Pada aplikasi *weka* peneliti menggunakan rumus *C-Means* sesuai dengan rumus yang telah ditentukan pada penelitian ini.



Gambar 5 Proses Hasil Clustering

3.4. Hasil Proses Data

- Selanjutnya Untuk *Attribute* Pangan, dengan jumlah klasifikasi keseluruhan data yaitu 2.4395, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu untuk *Cluster* 0 = 2.1592, *Cluster* 1 = 2.569, *Cluster* 2 = 2.4189, *Cluster* 3 = 2.6045.
- Selanjutnya Untuk *Attribute* Sandang, dengan jumlah klasifikasi keseluruhan data yaitu 3.1069, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu untuk *Cluster* 0 = 3.1224, *Cluster* 1 = 3.1164, *Cluster* 2 = 2.6151, *Cluster* 3 =3.5109.
- Selanjutnya Untuk *Attribute* Papan, dengan jumlah klasifikasi keseluruhan data yaitu 2.9941, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu untuk *Cluster* 0 = 3, *Cluster* 1 = 3.1422, *Cluster* 2 = 5.5358, *Cluster* 3 =0.347.
- Selanjutnya Untuk *Attribute* Penghasilan, dengan jumlah klasifikasi keseluruhan data yaitu 7.8416, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu untuk *Cluster* 0 = 6.898, *Cluster* 1 = 25, *Cluster* 2 = 0.717, *Cluster* 3 =0.8955.

- e) Selanjutnya Untuk *Attribute* Kesehatan, dengan jumlah klasifikasi keseluruhan data yaitu 1.9624, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu untuk *Cluster* 0 = 1.8694, *Cluster* 1 = 1.9052, *Cluster* 2 = 2.0226, *Cluster* 3 = 2.0373.
- f) Selanjutnya Untuk *Attribute* Pendidikan, dengan jumlah klasifikasi keseluruhan data yaitu 2.0277, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 1.9837, *Cluster* 1 = 2.0517, *Cluster* 2 = 1,9396, *Cluster* 3 = 2.1343.
- g) Selanjutnya Untuk *Attribute* Kekayaan 1, dengan jumlah klasifikasi keseluruhan data yaitu 2.6218, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 2.3714, *Cluster* 1 = 2.681, *Cluster* 2 = 2.634, *Cluster* 3 = 2.7873.
- h) Selanjutnya Untuk *Attribute* Kekayaan 2, dengan jumlah klasifikasi keseluruhan data yaitu 2.6436, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 2.6327, *Cluster* 1 = 2.5129, *Cluster* 2 = 2.5736, *Cluster* 3 = 2.8358.
- i) Selanjutnya Untuk *Attribute* Air Bersih, dengan jumlah klasifikasi keseluruhan data yaitu 1.7921, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 1.6327, *Cluster* 1 = 1.75, *Cluster* 2 = 1.9321, *Cluster* 3 = 1.8358.
- j) Selanjutnya Untuk *Attribute* Listrik, dengan jumlah klasifikasi keseluruhan data yaitu 0.7277, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 3, *Cluster* 1 = 0, *Cluster* 2 = 0, *Cluster* 3 = 0.
- k) Selanjutnya Untuk *Attribute* Jumlah Jiwa, dengan jumlah klasifikasi keseluruhan data yaitu 1.2941, Sedangkan untuk posisi *Cluster* masing – masing cluster memiliki nilai presentase pengelompokkan yaitu *Cluster* 0 = 1.2531, *Cluster* 1 = 1.306, *Cluster* 2 = 1.3925, *Cluster* 3 = 1.2239.

4. Kesimpulan

Berdasarkan dari penelitian yang telah dilaksanakan dan sudah diuraikan dalam penerapan *data mining* dari data penduduk kurang mampu tepatnya pada tahun 2019 sebanyak kurang lebih 1000 penduduk, untuk menentukan penduduk yang benar benar kurang mampu, maka penulis dapat menarik kesimpulan sebagai berikut:

1. Penerapan *data mining* dengan teknik *CLUSTERING* dan metode *C-Means* yang dilakukan dapat menghasilkan sebuah informasi mengenai data penduduk miskin dengan 4 cluster, yang dibagi menjadi 4 kriteria kelompok penduduk yaitu :
 - a. Sangat Miskin
 - b. Hampir Sangat Miskin
 - c. Miskin
 - d. Mendekati Miskin.
2. Dalam penerapan *data mining* ini dapat memberikan informasi data penduduk yang benar-benar kurang mampu dan diharapkan mendapatkan bantuan dari pemerintah, dan suatu keputusan atau suatu pertimbangan bagi pemerintah untuk memeberikan bantuan yang tepat sasaran..

Daftar Pustaka

- Anggraeni, H. D., Saputra, R., & Noranita, B. (2016). Aplikasi Data Mining Analisis Data Transaksi Penjualan Obat Menggunakan Algoritma Apriori (Studi Kasus di Apotek Setya Sehat Semarang). *Jurnal Masyarakat Informatika*, 4(7), 1–8. <https://doi.org/10.14710/jmasif.4.7.1-8>
- Kelulusan Mahasiswa, T., Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer, D., & Matematika, J. (2010). aplikasi data mining untuk menampilkan informasi nuqson masykur huda j2f005280 program studi teknik informatika. Retrieved from http://eprints.undip.ac.id/23168/1/TA_NUXON_J2F005280.pdf
- Nurjanah, Farmadu, A., & Indriani, F. (2014). Implementasi Metode Fuzzy C-Means Pada, 01(01), 23–32.
- Aminulloh, S, G. 2016. “Penerapan Data Mining Pada Penyewaan Film Di Atika Menggunakan Metode Association Rules.”. Artikel Skripsi Universitas Nusantara PGRI Kediri