

Penerapan Data Mining untuk *Clustering* Data Penduduk Miskin Menggunakan Metode *K-Means*

Parjito*¹, Permata²

^{1,2}Universitas Teknokrat Indonesia

e-mail: djito@teknokrat.ac.id

Abstract

Poverty that occurs in the community is a major problem that must be an important concern for the government in any country in various provincial, district and city governments, such as providing assistance to the poor in the form of allowances that can relieve the community and help the community, Community Health Insurance, Rice for the Poor (Raskin), Education Cost Assistance for Academic Achievement Improvement (BBP-PPA) Tuition Fee Assistance which is only intended for prospective students who are unable or poor (BIDIKMISI), and other programs that can help the poor, but in reality the problem of poverty cannot be overcome. Suka Bhakti Village carried out grouping of poor and non-poor communities based on basic determinant variables, namely the latest education of members or family heads, ownership of several family assets, total family income from both the head of the family and a combination of family members, the area of the building owned, the type of floor used, and the type of wall. The use of data mining with cluster techniques through the *K-Means* method can group community data to get clusters that should be able to receive social assistance from the government. The results of the validity calculations carried out, the results obtained for the number of cluster 1 are 130 data and for cluster 2 are 682. Testing this data is carried out using the *Davies-Bouldin Index (DBI)* method. The resulting number of suitable clusters is 2 clusters with a value of 0,1643 which means that the cluster is suitable for grouping community data.

Keyword: Social Assistance; Clustering; Data Mining; *K-Means*; Poor Society

Abstrak

Kemiskinan yang terjadi pada masyarakat merupakan masalah utama yang harus menjadi perhatian penting bagi pemerintah baik di negara manapun di berbagai pemerintahan provinsi, kabupaten dan kota, khususnya pada Desa Suka Bhakti Kecamatan Gedungaji Baru Kabupaten Tulang Bawang Provinsi Lampung. Berbagai cara telah dilakukan dalam penanggulangan kemiskinan yang dilakukan oleh pemerintah seperti halnya pemberian bantuan-bantuan kepada masyarakat miskin berupa tunjangan yang dapat meringankan masyarakat dan membantu masyarakat, Jaminan Kesehatan Masyarakat, Beras Miskin (Raskin), Bantuan Biaya Pendidikan Peningkatan Prestasi Akademik (BBP-PPA), Bantuan Biaya Pendidikan yang hanya ditujukan untuk calon mahasiswa yang tidak mampu atau miskin (BIDIKMISI), dan program-program lainnya yang dapat membantu masyarakat miskin. Tetapi pada kenyataannya masalah kemiskinan belum dapat ditanggulangi. Desa Suka Bhakti Kecamatan Gedungaji Baru Kabupaten Tulang Bawang Provinsi Lampung melakukan pengelompokan masyarakat yang miskin dan yang tidak miskin berdasarkan variabel penentu dasar yaitu pendidikan terakhir anggota atau kepala keluarga, kepemilikan beberapa asset keluarga, jumlah penghasilan keluarga baik dari kepala keluarga maupun gabungan dari anggota keluarga, luas bangunan yang dimiliki, jenis lantai yang digunakan, dan jenis dinding. Penggunaan data mining dengan teknik *cluster* melalui metode *K-Means* dapat mengelompokkan data masyarakat untuk mendapatkan *cluster* yang seharusnya dapat menerima bantuan sosial dari pihak pemerintah. Berdasarkan hasil perhitungan validitas yang dilakukan, maka didapatkan hasil untuk jumlah *cluster* 1 adalah sebanyak 130 data dan untuk *cluster* 2 berjumlah 682. Pengujian data ini dilakukan dengan menggunakan metode *Davies-Bouldin Index (DBI)* dihasilkan jumlah *cluster* yang cocok adalah 2 *cluster* dengan nilai 0,1643 yang berarti *cluster* tersebut cocok untuk digunakan untuk mengelompokkan data masyarakat.

Kata kunci: Bantuan Sosial; Clustering; Data Mining; *K-Means*; Masyarakat Miskin

1. Pendahuluan

Desa menurut PP NO 72 TAHUN 2005, adalah kesatuan masyarakat hukum yang memiliki batas-batas wilayah dan memiliki kewenangan untuk mengatur serta mengurus kepentingan masyarakat setempat yang diakui dan dihormati dalam sistem Negara Kesatuan Republik Indonesia. Sebagai kesatuan masyarakat yang terkecil, desa memiliki kewenangan untuk mengatur diri, mengembangkan diri dan menggali potensi yang dimiliki untuk kesejahteraan dan kemakmuran masyarakat dalam kerangka negara kesatuan Republik Indonesia. Namun desa juga dapat di tandai dengan kemiskinan dimana-mana, jumlah pengangguran yang tinggi, tingkat kecerdasan masyarakat masih rendah dan distribusi pendapatan tidak merata yang mengakibatkan terhambatnya laju pertumbuhan ekonomi [1].

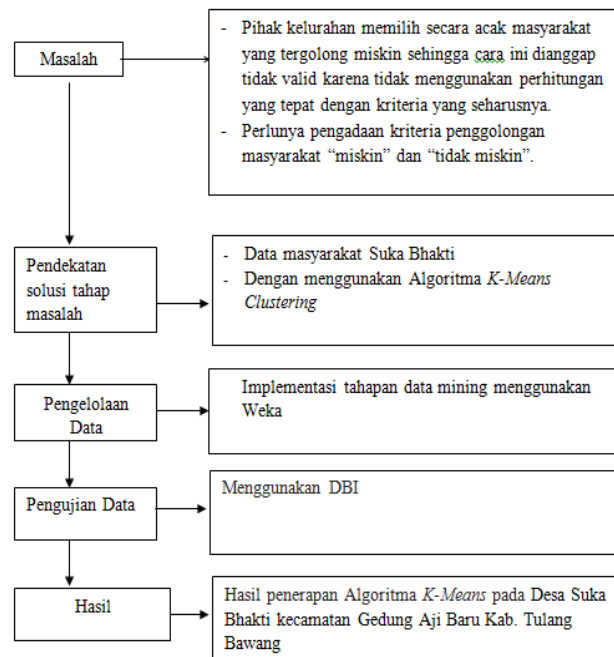
Desa Suka Bhakti Kecamatan Gedungaji Baru Kabupaten Tulang Bawang merupakan pemberian seorang perintis kampung Desa Suka Bhakti yaitu Suyono Bayu Haji. Nama Desa Suka Bhakti yang terdiri dari dua suku kata yaitu Suka artinya senang, dan Bhakti artinya taat atau patuh. Sehingga dapat disimpulkan bahwa Desa Suka Bhakti artinya adalah senang dan taat kepada Tuhan Yang Maha Esa serta pemerintah. Kantor kelurahan Desa Suka Bhakti yang melayani masyarakat secara langsung pastinya setiap hari melakukan kegiatan pengelolaan data kependudukan seperti jumlah data penduduk, data penduduk berkecukupan dan data penduduk miskin. Data kependudukan yang lengkap dan terbaru (*update*) merupakan data wajib yang harus dimiliki dan diperhatikan oleh pihak kelurahan, yang nantinya akan diberikan ke kantor kecamatan agar bisa digunakan sebagai alat kontrol mengetahui pertumbuhan dan perkembangan penduduk.

Kemiskinan yang terjadi pada masyarakat merupakan masalah utama yang menjadi perhatian pemerintah di negara manapun di berbagai pemerintahan provinsi, kabupaten dan kota khususnya Desa Suka Bhakti Kecamatan Gedungaji Baru Kabupaten Tulang Bawang. Berbagai cara penanggulangan kemiskinan telah dilakukan oleh pemerintah seperti pemberian bantuan-bantuan masyarakat miskin berupa tunjangan, Jaminan Kesehatan Masyarakat, Raskin, BPP-PPA, BIDIKMISI, dan program-program lainnya, tetapi yang pada kenyataannya masalah kemiskinan belum dapat ditanggulangi.

Kantor Desa Suka Bhakti masih belum mengetahui kelompok atau variabel-variabel apa saja yang sangat berpengaruh pada data kemiskinan dalam hal pengelolaan data penduduk miskin. Pada penelitian ini penulis ingin menerapkan data mining pada data Penduduk Desa Suka Bhakti menggunakan algoritma *K-Means Clustering* [2], [3], [4]. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database [5]. Data mining juga merupakan proses statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [6]. Data mining dapat berfungsi dalam berbagai hal, antara lain *characterization, discrimination, association, classification, clustering, outlier* dan *trend analysis* [7]. Dari beberapa fungsi tersebut, teknik yang digunakan untuk mengelompokkan data dengan karakteristik yang sama adalah *clustering* [8]. *K-Means clustering* sebagai salah satu metode data *clustering non-hirarki* dengan mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Kelompok atau *cluster* yang didapat merupakan pengetahuan atau informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan [9], [10].

2. Metode Penelitian

Sebelum melakukan proses data mining dengan baik maka perlu digali permasalahan yang ditemui. Sebuah upaya dilakukan dengan mewawancarai pihak-pihak yang terkait dalam permasalahan terutama perangkat desa dan instansi terkait. Berikut ini adalah skema dari tahapan penelitian yang dilakukan.



Gambar 1. Kerangka Penelitian

Gambar 1 menunjukkan tahapan - tahapan dari penelitian ini. Tahap pertama yang peneliti lakukan adalah dengan melakukan identifikasi masalah pada penelitian yang sedang dilakukan. Identifikasi masalah dilakukan agar dapat diketahui kendala-kendala dan permasalahan yang muncul untuk dapat dianalisa dan ditemukan solusi yang tepat [11]. Identifikasi masalah dapat dilakukan dengan wawancara dan observasi di lapangan [12]. Masalah yang terjadi pada Desa Suka Bhakti kecamatan Gedung Aji Baru Kab. Tulang Bawang yaitu dalam menentukan masyarakat yang tergolong miskin. Pihak kelurahan hanya melakukan pemilihan secara acak sehingga bantuan yang diberikan belum tepat pada sasaran. Adapun masalah di Desa Suka Bhakti kecamatan Gedung Aji Baru Kab. Tulang Bawang yaitu:

- a) Pihak kelurahan memilih secara acak masyarakat yang tergolong miskin sehingga cara ini dianggap tidak valid karena tidak menggunakan perhitungan yang tepat .
- b) Saat ini proses kerja di Desa Suka Bhakti kecamatan Gedung Aji Baru Kab. Tulang Bawang masih kurang efektif dalam menentukan masyarakat miskin dan tidak miskin.

Tahap kedua peneliti mengajukan usulan berupa analisis *clustering* data penduduk miskin menggunakan Algoritma *K-Means* untuk dapat mengetahui kelompok atau variabel-variabel apa saja yang sangat berpengaruh pada data kemiskinan.

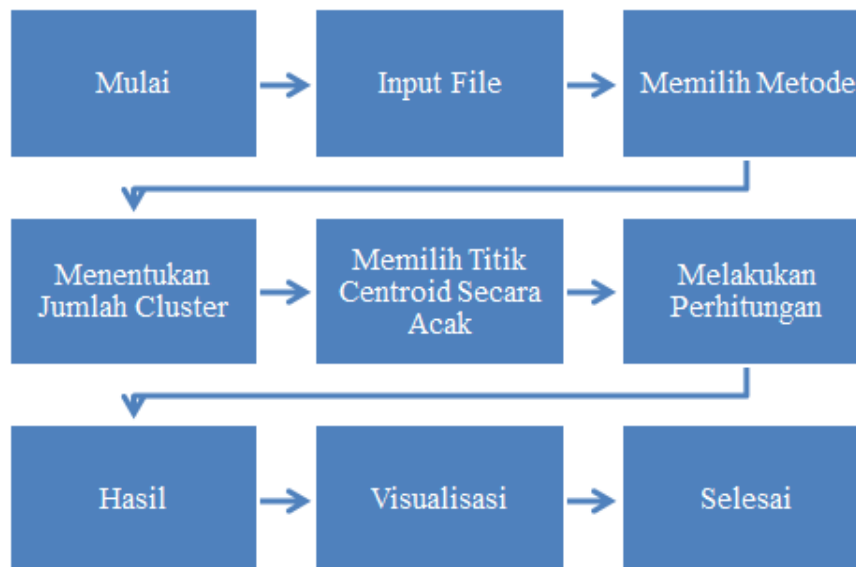
Setelah usulan diterima, peneliti melakukan analisis data. Dalam analisis menggunakan data mining, suatu hal yang sangat penting adalah data, karena data merupakan objek yang digunakan untuk penambangan dan pengujian [13]. Analisis data yang digunakan adalah Algoritma *K-Means* dimana tahapan-tahapannya terdiri dari:

- a) Menyiapkan data training, dimana data training ini didapat dari kelurahan Desa Suka Bhakti.
- b) Menentukan kriteria apa saja yang akan digunakan untuk menjadi parameter dalam menentukan masyarakat yang tergolong miskin dan tidak miskin.
- c) Menentukan jumlah *cluster* yang akan digunakan
- d) Menghitung masing-masing kriteria untuk pengelompokan *cluster*

Pengujian data merupakan tahapan terakhir yang dilakukan dalam penelitian, pada tahap ini data yang akan di uji telah divalidasi atau telah ditetapkan dan data tidak berubah nilainya. Pengujian data dilakukan menggunakan DBI untuk melihat seberapa baiknya *clustering* yang sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data. Hasil dari penelitian ini berupa penerapan Algoritma *K-Means* pada Desa Suka Bhakti kecamatan Gedung Aji Baru Kab. Tulang Bawang.

Alur proses penentuan *cluster* pada data dimulai dari menyeleksi data, pembersihan data, transformasi data menjadi angka dan pemilihan atribut yang akan digunakan kemudian menjadi

data *training* yang siap diolah. Berikut adalah alur dari proses penentuan *cluster* pada Desa Suka Bhakti Kecamatan Gedung Aji Baru Kab. Tulang Bawang yang ditunjukkan pada gambar 2.



Gambar 2. Alur Proses Menentukan Cluster

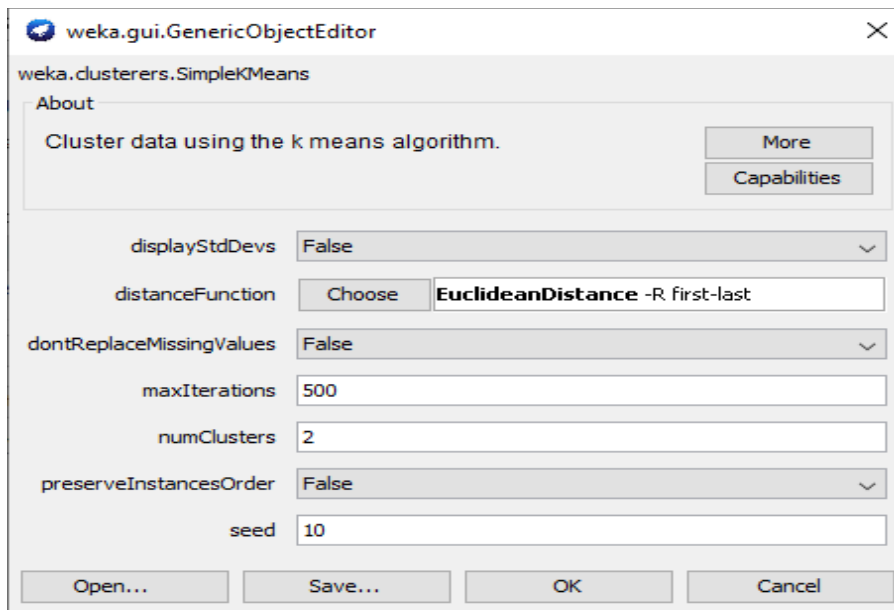
Alur proses pembentukan *cluster* dengan algoritma *K-Means* menggunakan *software* Weka, langkah-langkah yang dilakukan adalah sebagai berikut. Langkah pertama melakukan proses *input* data ke dalam Weka dengan memilih submenu *open file* pada menu *preprocess*. File yang akan di inputkan merupakan data dengan format *.csv. Langkah kedua adalah memilih algoritma *K-means*. Pada *software* Weka algoritma *K-Means* diberi nama dengan *SimpleKMeans*. Langkah ketiga, menentukan jumlah *cluster* yang sesuai dengan kebutuhan analisis. Selanjutnya *software* Weka akan secara otomatis memilih titik *centroid* secara acak. Kemudian *software* akan secara otomatis menghitung jarak setiap data ke titik *centroid*. Perhitungan akan berhenti jika seluruh data telah konsisten atau tidak berubah keanggotaan *cluster*-nya. Langkah terakhir setelah dilakukan perhitungan maka akan tampil hasil dan visualisasi dari keanggotaan *cluster*.

3. Hasil dan diskusi

Data yang digunakan dalam penelitian ini adalah data penduduk Desa Suka Bhakti Kecamatan Gedung Aji Baru Kab. Tulang dan data yang di peroleh sebanyak 812 dengan pengujian menggunakan aplikasi WEKA. WEKA merupakan *tools* data mining yang memiliki fitur pengujian lengkap dan mudah digunakan. Pengguna dapat melakukan *clustering* data dan menganalisa pola penyebaran data tanpa menuliskan kode pogram sama sekali. Dari *clustering* yang dilakukan diperoleh hasil yaitu *confosion matrix* 103 data yang menjadi anggota *cluster* 0 atau termasuk dalam kategori miskin dan 709 menjadi anggota *cluster* 1 atau termasuk dalam kategori tidak miskin, dengan pusat *cluster* 0 adalah 0.333333,1,1,0,0.5,0 dan *cluster* 1 adalah 0.666667,1,1,0,1,1. Dalam penelitian ini, proses iterasi berenti pada iterasi ke-4 dengan persentase 13% pada *cluster* 0 dan 83% pada *cluster* 1. *Cluster* 0 dilambangkan dengan warna biru sementara *cluster* 1 dengan warna merah. Terdapat angka 405.5 yang merupakan titik tengah dari jumlah data yang dihitung, sementara 812 merupakan jumlah seluruh data. Berikut merupakan visualisasi dari hasil *cluster*.

3.1. Penentuan Jumlah Cluster

Pengujian sistem antarmuka merupakan pengujian yang dilakukan dengan menggunakan layanan yang disediakan oleh suatu sistem [14]. Pada penelitian ini pengujian sistem antarmuka dilakukan dengan menggunakan aplikasi WEKA. Dalam tahapan ini peniliti memilih metode *simple K-Means* lalu memasukan jumlah *cluster* sesuai dengan kebutuhan penelitian, dalam penelitian ini jumlah *cluster* yang digunakan sebanyak 2 *cluster*. Berikut merupakan tampilan *input cluster* dengan cara *double* klik pada metode.

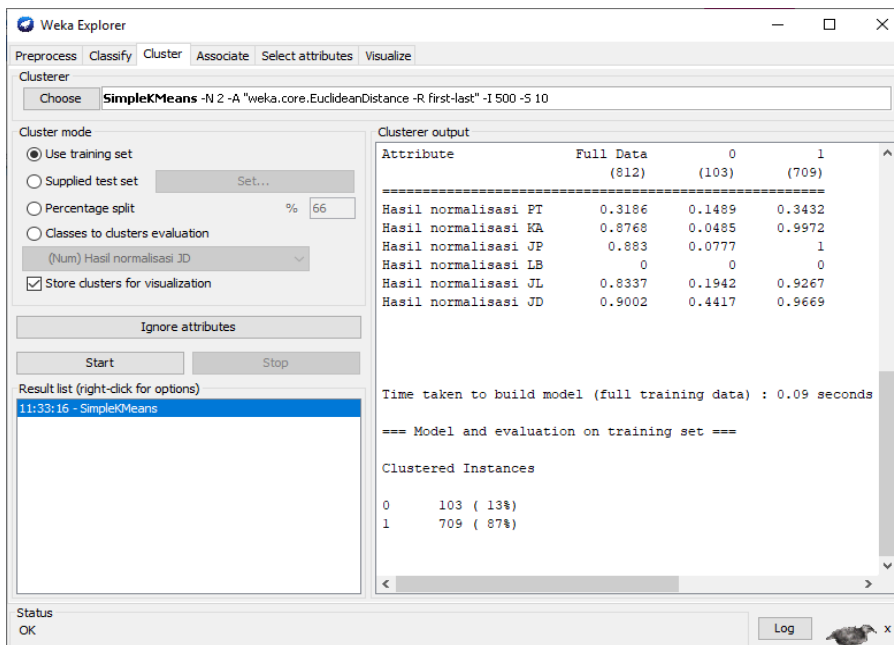


Gambar 3. Input Jumlah Cluster

Setelah melakukan pemilihan dan penentuan jumlah *cluster* langkah selanjutnya yaitu peneliti mengklik *button start* untuk melakukan proses *cluster*.

3.2. Perhitungan K-Means

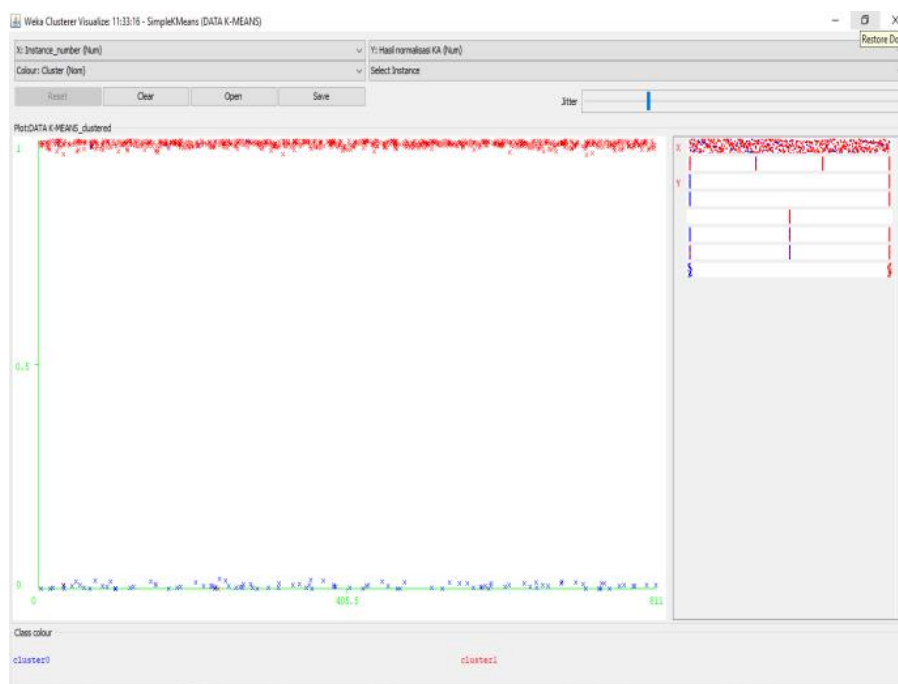
Pada proses perhitungan ini yang dilakukan adalah memilih metode *SimpleKMeans* serta menentukan jumlah *cluster* *laster* sebanyak 2 *cluster*. Secara otomatis Weka akan memilih titik *centroid* secara acak dengan nilai *cluster* 0 adalah 0.333333,1,1,0,0.5,0 dan *cluster* 1 adalah 0.666667,1,1,0,1,1. *Cluster* 0 dikategorikan sebagai masyarakat miskin sementara *cluster* 1 dikategorikan sebagai masyarakat tidak miskin. Setelah dilakukannya perhitungan secara otomatis maka dihasilkan *confusion matrix* 103 data yang menjadi anggota *cluster* 0 atau termasuk dalam kategori miskin dan 709 menjadi anggota *cluster* 1 atau termasuk dalam kategori tidak miskin. Dengan menggunakan iterasi sebanyak 4 kali. Berikut merupakan tampilan hasil pengujian pada aplikasi WEKA:



Gambar 4. Hasil Perhitungan K-Means

3.3. Hasil Pengujian

Data yang digunakan dalam penelitian ini adalah data Data penduduk Desa Suka Bhakti Kecamatan Gedung Aji Baru Kab. Tulang dan data yang di peroleh sebanyak 812 dengan pengujian menggunakan aplikasi WEKA. WEKA merupakan *tools* data mining yang memiliki fitur pengujian lengkap dan mudah digunakan. Pengguna dapat melakukan *clustering* data dan menganalisa pola penyebaran data tanpa menuliskan kode pogram sama sekali. Dari *clustering* yang dilakukan diperoleh hasil yaitu *confosion matrix* 103 data yang menjadi anggota *cluster* 0 atau termasuk dalam kategori miskin dan 709 menjadi anggota *cluster* 1 atau termasuk dalam kategori tidak miskin, dengan pusat *cluster* 0 adalah 0.333333,1,1,0,0.5,0 dan *cluster* 1 adalah 0.666667,1,1,0,1,1. Dalam penelitian ini, proses iterasi berhenti pada iterasi ke-4 dengan persentase 13% pada *cluster* 0 dan 83% pada *cluster* 1. *Cluster* 0 dilambangkan dengan warna biru sementara *cluster* 1 dengan warna merah. Terdapat angka 405.5 yang merupakan titik tengah dari jumlah data yang dihitung, sementara 812 merupakan jumlah seluruh data. Berikut merupakan visualisasi dari hasil *cluster*.



Gambar 5. Visualisasi keanggotaan cluster

3.4. Tingkat Akurasi Klasterisasi

Dari hasil perhitungan peneliti menggunakan perhitungan akurasi menggunakan DBI sebagai *validitas internal* yang dilakukan adalah untuk mengetahui seberapa baik *clustering* yang sudah dilakukan dengan cara menghitung kuantitas dan fitur turunan dari set data. Secara esensial, DBI menginginkan nilai terkecil (*non-negatif* ≥ 0) untuk menilai baiknya *cluster* yang didapat. Indeks tersebut dapat digunakan untuk pengambilan keputusan untuk memilih jumlah *cluster* yang paling cocok digunakan [9], [15], [16], [17].

Dari hasil perhitungan didapatkan nilai 0,164278 yang berarti *cluster* yang dievaluasi menghasilkan *cluster* yang baik. Perhitungan validitas yang dilakukan, maka didapatkan hasil untuk jumlah *cluster* 0 adalah 130 data dan untuk *cluster* 1 berjumlah 682. Dimana untuk label *cluster* 0 sebagai masyarakat “Miskin” dan untuk *cluster* 1 dinyatakan sebagai masyarakat yang tergolong “Tidak Miskin”. Dari hasil analisis dapat diambil kesimpulan bahwa ada 6 faktor penentu masyarakat Desa Suka Bhakti Kecamatan Gedungaji Baru adalah pendidikan terakhir kepala keluarga, kepemilikan aset, jumlah penghasilan, luas bangunan, jenis lantai, dan jenis dinding.

4. Kesimpulan

Berdasarkan hasil penelitian menggunakan metode *K-Means* yang telah dilakukan maka didapat hasil berupa kategori masyarakat miskin dan tidak miskin yang dapat dipergunakan sebagai salah satu alat yang digunakan untuk mengambil keputusan dalam menentukan

kelompok masyarakat yang mendapatkan bantuan dengan melihat *variable* Kepemilikan Aset. Hal ini terjadi karena dalam pembahasan yang dilakukan nilai variabel kepemilikan aset mempunyai pengaruh pembuatan *cluster* terbaik.

Berdasarkan hasil pengujian data training menggunakan Weka dapat disimpulkan dari 812 data training penduduk, 103 data digolongkan menjadi miskin dan 709 menjadi tidak miskin dengan iterasi sebanyak 4 kali. Persentase 13% pada kategori miskin dan 83% pada kategori tidak miskin dari hasil penerapan pada *tools* weka sedangkan pada validitas yang dilakukan menggunakan metode DBI dihasilkan jumlah *cluster* yang cocok adalah 2 *cluster* dengan nilai 0,1643 yang berarti *cluster* tersebut cocok untuk digunakan.

Referensi

- [1] W. Isdijoso, A. Suryahadi, and A. Akhmedi, "Penetapan Kriteria dan Variabel Pendataan Penduduk Miskin yang Komprehensif dalam Rangka Perlindungan Penduduk Miskin di Kabupaten/Kota," *Kertas Kerja SMERU*, (September), pp. 1–16, 2016.
- [2] A. N. Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput*, pp. 1–12, 2016.
- [3] S. Adinugroho and Y. A. Sari, *Implementasi Data Mining Menggunakan Weka*. Universitas Brawijaya Press, 2018.
- [4] J. Purba, *Pengelolaan Lingkungan Sosial*, 2nd ed. Jakarta: Yayasan Obor Indonesia, 2005.
- [5] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 25–34, 2020.
- [6] D. Iskandar and Y. K. Suprpto, "Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C 4.5 Dan Naïve Bayes," *Netw. Eng. Res. Oper.*, vol. 2, no. 1, 2016.
- [7] R. I. Borman, B. Priopradono, and A. R. Syah, "Klasifikasi Objek Kode Tangan pada Pengenalan Isyarat Alphabet Bahasa Isyarat Indonesia (Bisindo)," 2017.
- [8] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [9] I. M. R. Yuliawan, D. C. Khrisne, and P. A. Mertasana, "Penerapan Algoritma K-Means Clustering dalam Penentuan Nilai Huruf pada Permainan Susun Kata Bahasa Bali," *J. SPEKTRUM*, vol. 6, no. 3, pp. 87–93, 2019.
- [10] N. F. Kahar, "Penentuan Tingkat Kemiskinan dengan Metode Fuzzy C-Means di Kabupaten Bone Bolango," Universitas Negeri Gorontalo, 2015.
- [11] Y. Rahmanto, J. Alfian, D. Damayanti, and R. I. Borman, "Penerapan Algoritma Sequential Search pada Aplikasi Kamus Bahasa Ilmiah Tumbuhan," *J. Buana Inform.*, vol. 12, no. 1, pp. 21–30, 2021.
- [12] R. I. Borman, I. Yasin, M. A. P. Darma, I. Ahmad, Y. Fernando, and A. Ambarwari, "Pengembangan dan Pendampingan Sistem Informasi Pengolahan Pendapatan Jasa pada PT. DMS Konsultan Bandar Lampung," *J. Soc. Sci. Technol. Community Serv.*, vol. 1, no. 2, pp. 24–31, 2020.
- [13] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Pahlevi, and D. E. Kurniawan, "Classification of the Period Undergraduate Study Using Back-propagation Neural Network," in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.
- [14] R. D. Gunawan, T. Oktavia, and R. I. B. I. Borman, "Perancangan Sistem Informasi Beasiswa Program Indonesia Pintar (PIP) Berbasis Online (Tudi Kasus: SMA N 1 Kota Bumi)," *MIKROTIK J. Manaj. Inform.*, vol. 8, no. 1, pp. 43–54, 2018.
- [15] E. Karyadiputra, "Analisis Algoritma Naive Bayes Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial," *Technol. J. Ilm.*, vol. 7, no. 4, 2016.
- [16] E. Prasetyo, "Data mining mengolah data menjadi informasi menggunakan matlab," 2014.
- [17] A. N. Ulfah, "Analisis Kinerja Algoritma Fuzzy C-Means dan K-Means pada Data Kemiskinan," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 1, no. 2, pp. 139–148, 2015.