



APPLICATION OF THE CLUSTERING METHOD FOR GROUPING DATA BASED ON FAMILY SOCIO-ECONOMIC STATUS

Erika Hidayati

Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer KAPUTAMA
Jl. Veteran No. 4A-9A, Binjai, Sumatera Utara, Indonesia

e-mail : erikahidayati412@gmail.com

Received: June 17, 2023

Revised: August 8, 2023

Accepted: August 23, 2023

Page : 77-83

Abstrak : Sosial ekonomi merupakan suatu kedudukan yang diatur secara sosial dan menempatkan seseorang pada posisi tertentu dalam masyarakat. Kondisi sosial ekonomi masyarakat dapat dilihat dari melalui aspek pendidikan, pekerjaan dan penghasilan. Kondisi sosial ekonomi masyarakat dapat dikatakan baik apabila telah terpenuhi kebutuhan sehari-harinya baik itu kebutuhan sandang, pangan maupun papan, sebaliknya apabila salah satu dari kebutuhan itu tidak terpenuhi maka dapat dikatakan bahwa kondisi sosial ekonominya buruk. Penulisan proposal ini menggunakan metode *clustering* yang merupakan salah satu teknik data Mining untuk pengelompokan Data Dasar Keluarga Dengan menggunakan metode *clustering* algoritma *k-means*. Dengan menerapkan 20 sampel data alternatif data siswa dan memberikan jumlah *cluster* sebanyak 3, dan memanfaatkan 3 kriteria utama sebagai penelitian di proposal ini, menghasilkan jumlah *cluster* 1 sebanyak 11 data, *cluster* 2 sebanyak 7 data. Dan *cluster* 3 sebanyak 3 data. Sistem pengelompokan Data Dasar Keluarga ini di rancang dengan Bahasa pemrograman aplikasi *MATLAB* dan memanfaatkan *Gui* sebagai *interfacenya*.

Kata kunci: Algoritma_K-Means, Clustering, Data_Mining, Data Dasar Keluarga, Matlab

Abstract : Socio-economic is a position that is regulated socially and places a person in a certain position in society. Socio-economic conditions of the community can be seen through the aspects of education, employment and income. The socio-economic condition of the community can be said to be good if their daily needs have been met, be it the need for clothing, food or shelter, conversely if one of these needs is not fulfilled then it can be said that the socio-economic condition is bad. Writing this proposal uses the clustering method which is one of the data mining techniques for grouping basic family data using the k-means

algorithm clustering method. By applying 20 alternative data samples of student data and providing a total of 3 clusters, and utilizing 3 main criteria as research in this proposal, the resulting number of cluster 1 is 11 data, cluster 2 is 7 data. And cluster 3 as much as 3 data. This Family Basic Data grouping system is designed with the MATLAB application programming language and utilizes the GUI as its interface.

Keywords: *K-Means_Algorithm, Clustering, Data_Mining, Basic Family Data, Matlab*



Journal of Mathematics and Technology (MATECH) This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

1 Pendahuluan (or Introduction)

Sosial ekonomi merupakan suatu kedudukan yang diatur secara sosial dan menempatkan seseorang pada posisi tertentu dalam masyarakat. Kondisi sosial ekonomi masyarakat dapat dilihat dari melalui aspek pendidikan kepala keluarga, pekerjaan kepala keluarga dan penghasilan kepala keluarga. Kondisi sosial ekonomi masyarakat dapat dikatakan baik apabila telah terpenuhi kebutuhan sehari-harinya baik itu kebutuhan sandang, pangan maupun papan, sebaliknya apabila salah satu dari kebutuhan itu tidak terpenuhi maka dapat dikatakan bahwa kondisi sosial ekonominya buruk [1].

Desa Tanjung Merahe terdiri dari 8 wilayah dusun. Pendataan dan klasifikasi status sosial ekonomi keluarga pada Desa Tanjung Merahe masih dilakukan secara manual sehingga prosesnya kurang efektif dan efisien karena harus menunggu data dari masing-masing kepala dusun yang ada di Desa Tanjung Merahe, setelah data dikumpulkan oleh semua kepala dusun maka perangkat desa akan merekap semua datanya untuk dijadikan laporan.

Proses klasifikasi dan validasi data status sosial ekonomi keluarga Desa Tanjung Merahe dilakukan hanya pada saat diperlukan saja, misalnya apabila ada permintaan data dari pemerintah pusat untuk pemberian bantuan, pihak Pemerintah Desa Tanjung Merahe baru melakukan pendataan melalui kepala dusun dan hanya mengklasifikasikan pada keluarga yang tergolong kelas menengah ke bawah. Sehingga pihak Pemerintah Desa tidak dapat mengetahui secara pasti dan akurat tentang status sosial keluarga mana yang tergolong kelas atas, menengah atau bawah.

Dengan pemanfaatan data mining menggunakan data masukan berupa data dasar keluarga (DDK) pada Desa Tanjung Merahe, data tersebut akan diolah dengan metode clustering menggunakan algoritma k-means dengan menggunakan variabel pendidikan kepala keluarga, variabel pekerjaan kepala keluarga dan variabel penghasilan kepala keluarga pada data dasar keluarga (DDK) Desa Tanjung Merahe. Berdasarkan hasil pengelompokan tersebut, pengolahan data yang dilakukan dapat menghasilkan informasi yang berguna terhadap pihak Desa Tanjung Merahe dalam pengambilan keputusan.

2 Tinjauan Literatur (or Literature Review)

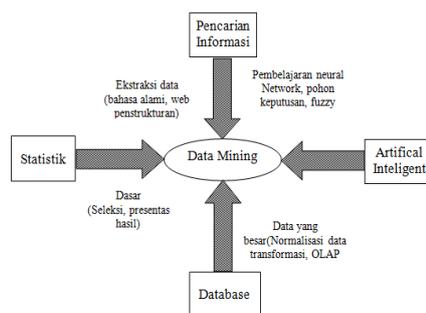
Dalam penelitian ini, penulis sedikit banyak terinspirasi dan mererensi dari penelitian-penelitian sebelumnya yang berkaitan dengan latar belakang masalah pada penelitian ini yaitu:

Menurut (Suhartini et al.2021) menyatakan bahwa *Data Mining* dapat menghasilkan informasi yang bermanfaat, dimana dapat dibuktikan bahwa hasil dari sistem yang dibuat dengan metode *clustering* yang dapat mengelompokkan data penduduk di wilayah Sukumulia Timur yang memang

dikatakan tergolong penduduk miskin untuk memudahkan dalam proses pembagian bantuan agar merata kepada penduduk di wilayah Sukamulia Timur. Dengan menggunakan Algoritma *K-Means* dalam mengelompokan data dapat dilakukan dengan sangat efektif. *Software* yang digunakan yaitu *rapid miner* dan *ms. Excel* [2]. Penelitian sebelumnya membahas tentang pengelompokan data penduduk miskin berdasarkan pekerjaan, penghasilan dan jumlah anak yang,sekolah dan jumlah anggota keluarganya sedangkan penelitian penulis tentang bagaimana mengelompokan data keluarga berdasarkan status sosial ekonominya berdasarkan pendidikan, pekerjaan dan penghasilanya dengan menggunakan *software matlab* dan *ms. Excel*.

Menurut (Noviana et al. 2019) Menyatakan bahwa *Data Mining* dapat membantu untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dalam pengolahan data penerima bantuan beras sejahtera (rastra) dan dapat membantu dalam menentukan penduduk yang menjadi prioritas utama untuk mendapatkan bantuan di tengah banyaknya data penduduk di kecamatan muara papalik. *Software* yang digunakan yaitu *weka* [3]. Penelitian sebelumnya membahas tentang pengelompokan data prioritas penerima bantuan beras sejahtera (rastra) berdasarkan jumlah tanggungan, kondisi rumah, jumlah penghasilan dan status kesejahteraannya sedangkan penelitian penulis tentang bagaimana mengelompokan data keluarga berdasarkan status sosial ekonominya dengan menggunakan *software matlab* dan *ms. excel*.

Menurut Kusriani dan emha Taufiq Lutfhi (Nisa Hanum Harani et al.2020) menyatakan bahwa *Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam *database*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (*machine learning*) untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan berpengetahuan yang terkait dari berbagai *database* besar [4].



Gambar 1 Bidang Ilmu Data Mining

Sumber Kusriani dan emha Taufiq Lutfhi (Nisa Hanum Harani et al.2020)

Menurut (Cahyo Prianto et al. 2020 hal:14) *clustering* adalah “suatu metode mempartisi data-set menjadi beberapa sub-set atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki *set property* yang *dishare* bersama, dengan tingkat similaritas yang tinggi dalam suatu kelompok dan tingkat similaritas antar kelompok yang rendah”. Sehingga *cluster* tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sampel yang ada salah satunya dengan menggunakan rumus jarak *ecluidean*. Tujuan utama dari metode *clustering* adalah pengelompokkan sejumlah data/obyek kedalam *cluster (group)* sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin [5].

Menurut (Nurhayati.2022 hal:7) menyatakan bahwa “ *K-Means* merupakan salah satu metode pengelompokan data non hierarki (sekatan) yang berusaha mempartisi data ke dalam *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama akan dikelompokan ke dalam satu *cluster* yang sama dan data yang memiliki karakteristik yang berbeda dikelompokan ke dalam kelompok yang lain [6].



Menurut (Indah Werdiningsih.2022 hal:55) menyatakan bahwa *Matlab* merupakan singkatan dari Matrix Laboratory. *Matlab* merupakan sebuah platfor pemogramana dan komputasi numeric yang digunakan untuk menganalisis data, mengembangkan algoritma dan membuat model. Leh karena itu *matlab* banyak digunakan beberapa engineer dan ilmuwan di berbagai bidang riset yang memerlukan komputasi numeric yang kompleks [7].

Menurut (I Made Andik Setiawan.2022 hal:1) *Graphical User Interface (GUI)* *Matlab* adalah faslitas yang disediakan oleh *matlab* untuk programing berbasis object [8] .

Menurut Hughes dan Kapoor (Martha Rianty., S.E., MSi.2021 hal:12) usaha atau bisnis adalah suatu kegiatan individu yang terorganisasi untuk menghasilkan dan menjual barang dan jasa guna mendapatkan keuntungan dalam memenuhi kebutuhan masyarakat [9] .

Menurut (hamdani, Se., Msi.2020 hal:1) Usaha Mikro Kecil dan Menengah adalah bentuk kegiatan ekonomi rakyat yang berskala kecil dan memenuhi kriteria kekayaan bersih atau hasil penjualan tahunan serta kepemilikan sebagaimana diatur dalam undang-undang [10] .

Menurut (S. Purwaningsih.2020 hal:1) Masyarakat merupakan sekumpulan manusia yang hidup secara berdampingan dengan segala kebudayaan dan kepribadiannya [11].

3 Metode Penelitian (or Research Method)

Metodologi penelitian ini dilakukan untuk mencari sesuatu secara sistematis dengan menggunakan metode ilmiah serta sumber yang berlaku. Dengan adanya proses ini, dapat memberikan hasil penelitian yang baik dan tepat. Atas dasar metodologi penelitian yang digunakan, maka dapat dibuat suatu alur kegiatan sebahai berikut :

1. Persiapan

Tahap ini adalah kegiatan awal dari penelitian, yaitu dengan penentuan dari latar belakang masalah kemudian dilakukan batasan masalah dan selanjutnya dilakukan penentuan tujuan serta manfaat yang dilakukan dalam penyusunan proses kerja sistem. Setelah itu, penulis menentukan kebutuhan data penelitian diantaranya menentukan metode yang akan digunakan, kriteria dari data dan hasil kelayakan. Setelah itu penulis membuat surat izin penelitian yang ditujukan ke Kantor Kepala Desa Tanjung Merahe untuk mengumpulkan data yang dibutuhkan dalam penelitian ini.

2. Kajian Teori

Dalam tahap ini penulis mengumpulkan berbagai teori baik dari buku yang dipinjam dari perpustakaan, jurnal maupun internet untuk mendukung penelitian yang akan dilakukan. Teori yang dikumpulkan antara lain mengenai, Data Mining, Metode *Clustering*, algoritma *K-Means*, Keluarga, Status Sosial Ekonomi, *Matlab*, *Gui*, *Flowchart* dan teori yang mendukung lainnya.

3. Pengumpulan Data

Tahap ini merupakan pengumpulan data yang diperlukan dalam pembuatan skripsi, yaitu Data Dasar Keluarga yang diperoleh Dari Kantor Kepala Desa Tanjung Merahe.

4. Analisa Data

Tahapan ini berupa merupakan tahapan mengelola dan menganalisa data yang telah diperoleh yaitu Data Dasar Keluarga, kemudian data tersebut dilakukan transformasi untuk dapat dilakukan analisa yang kemudian diproses dengan menggunakan metode *Clustering* untuk mendapatkan sebuah informasi yang baru atau dikelompokan sesuai dengan variabel yang telah ditentukan.

5. Pengujian dan implementasi

Tahap ini merupakan tahapan yang melakukan pengujian validasi dan implementasi data yang telah dianalisa sebelumnya serta penyusunan program. Pada tahap ini juga dilakukan pengujian dengan memproses data yang telah diproses dengan menggunakan *software matlab* sebagai implementasi proses pengelompokan data dasar keluarga.

6. Tahap Akhir

Tahap ini merupakan tahap pengambilan kesimpulan dan saran yang dapat dilakukan dalam penyusunan penelitian ini. Dengan adanya kesimpulan maka akan diketahui hasil dari keseluruhan penelitian dan diharapkan dengan saran akan ada perbaikan dan manfaat bagi yang lain.

4 Hasil dan Pembahasan (or Results and Analysis)

Tampilan Halaman *Home*

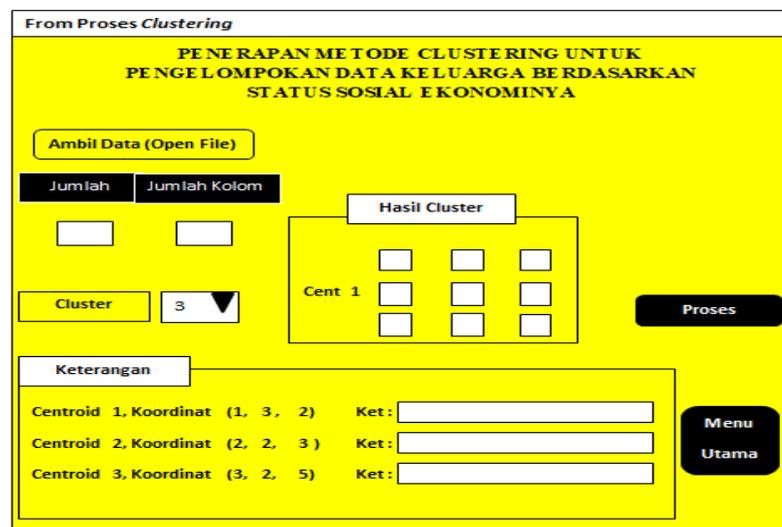
Halaman ini tampil ketika pertama kali membuka aplikasi *Matlab* untuk data mining yang akan dirancang, pada halaman ini akan terlihat semua *link-link* yang digunakan dalam aplikasi ini. Di sini hanya menampilkan gambar, tulisan, logo, judul aplikasi dan *button close*. Dengan desain form sebagai berikut :



Gambar 2 Tampilan *Halaman Home*

Tampilan Proses Clustreing

Pada halaman ini akan terlihat keseluruhan proses *Data Mining* sampai pada pemunculan grafik dan keterangan *Centroid* sebagai hasil dari perhitungan dengan metode *clustering* menggunakan algoritma *K-means* :



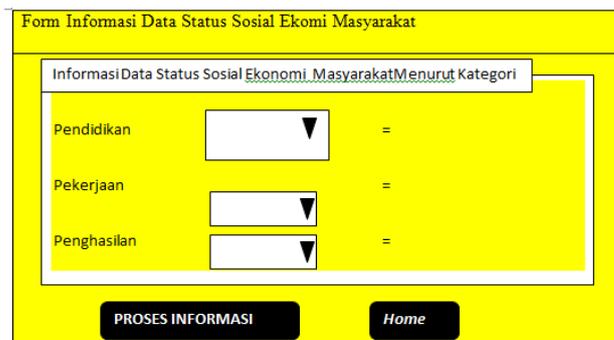
Gambar 3 Tampilan *Proses Clustering*

Keterangan Gambar :

1. *Button* Ambil Data berfungsi untuk mengambil data/import file dari *microsoft excel* yang akan dilakukan proses *clustering*.
2. Jumlah baris dan jumlah kolom untuk menampilkan jumlah baris dan kolom sesuai dengan data yang ada pada *microsoft excel* yang diimport untuk selanjutnya diproses *clustering*.
3. Jumlah *cluster* untuk menentukan jumlah pusat *cluster* yang akan diproses *clustering* dengan memilih angka *cluster* di menu *pop up* yang disediakan.
4. *Button* proses *cluster* digunakan untuk memproses perhitungan *clustering* dengan algoritma *k-means* dalam proses *data mining* tersebut, kemudian akan muncul hasil *clustering* pada masing-masing *centroid* dan juga muncul grafik hasil *clustering* berikut keterangan dari *centroid* tersebut.
5. *Button* Menu utama untuk kembali ke menu utama.

Halaman Informasi Data Status Sosial Ekonomi Masyarakat

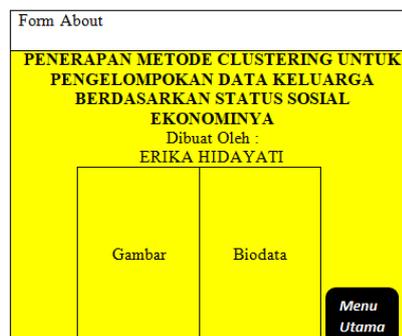
Pada halaman Informasi Data Status Sosial Ekonomi Masyarakat akan muncul data status social ekonomi masyarakat menurut kategori yang tersimpan di Microsoft Excel yang telah terkoneksi dengan Matlab, di mana data tersebut mengalami proses seleksi berdasarkan Jenjang Pendidikan, Pekerjaan, dan Penghasilan dari data status sosial ekonomi masyarakat pada halaman menu ini. Di mana desain formnya dalah sebagai berikut :



Gambar 4 Tampilan Menu Informasi Data Sosial Ekonomi Masyarakat

Halaman About

Pada halaman *about* akan muncul data dan gambar dari penulis selaku pembuat program *data mining*. Seperti tampak pada desain form berikut :



Gambar 4 Tampilan Menu About**5 Kesimpulan (or Conclusion)**

Setelah melakukan penelitian ini, dapat disimpulkan bahwa :

1. Pengelompokan Data Status Sosial Ekonomi Masyarakat Menggunakan Metode *Clustering* Pada Kantor Desa Tanjung Merahe merupakan sistem yang efektif untuk mengetahui informasi tentang Pendidikan kepala keluarga, Pekerjaan kepala keluarga dan Penghasilan kepala keluarga yang bisa dijadikan dasar pengambilan keputusan dalam meningkatkan pelayanan kepada masyarakat.
2. Sistem yang dibangun mampu diterapkan untuk memproses data Status sosial ekonomi masyarakat dan memudahkan pihak Pemerintah Desa Tanjung Merahe dalam mengambil keputusan apabila ada permintaan data dari pemerintah pusat maupun apabila ada pemberian bantuan sosial agar tepat sasaran.
3. Dengan telah dilakukannya pengelompokan ini agar kedepannya sistem ini dapat dikembangkan lebih baik lagi dan sistem ini dapat dijadikan bahan dan pengembangan penelitian lebih lanjut sehingga dapat memberikan manfaat lebih banyak lagi.

Referensi (Reference)

- [1] D. Dermawan, "ANALISIS DAMPAK COVID-19 PADA KONDISI SOSIAL EKONOMI GURU DI SD SWASTA GEMBALA BAIK II PONTIANAK," p. 2023, 2023.
- [2] S. Suhartini and R. Yuliani, "Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021, doi: 10.29408/jit.v4i1.2986.
- [3] T. Noviana, J. Jasmir, and Y. Novianto, "Penerapan Data Mining Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra Dengan Clustering K-Means," *Progr. Stud. Tek. Inform. Stikom Din. Bangsa*, pp. 159–174, 2019.
- [4] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *J. Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.2683.
- [5] C. Prianto and S. Bunyamin, *Pembuatan aplikasi clustering gangguan jaringan menggunakan metode K-Means clustering*. in Knowledge. Bandung.: Kreatif, 2020. [Online]. Available: <https://books.google.co.id/books?id=y8TgDwAAQBAJ>
- [6] *PEMODELAN K- MEANS ALGORITMA DAN BIG DATA ANALYSIS (PEMETAAN DATA MUSTAHIQ)*. Pascal Books, 2022. [Online]. Available: https://books.google.co.id/books?id=%5C_bJmEAAAQBAJ
- [7] I. Werdiningsih, D. C. R. Novitasari, and D. Z. Haq, *Pengelolaan Data Mining dengan Pemrograman Matlab*. Airlangga University Press, 2022. [Online]. Available: <https://books.google.co.id/books?id=CgOdEAAAQBAJ>
- [8] M. A. Setiawan, *Graphical User Interface (GUI) Matlab dan Interface dengan Arduino*. Bangka Belitung: Politeknik Manufaktur Negeri Bangka Belitung, 2022. [Online]. Available: <https://books.google.co.id/books?id=FeR3EAAAQBAJ>
- [9] M. Rianty and P. F. Rahayu, "Pengaruh E-Commerce Terhadap Pendapatan UMKM Yang Bermitra Gojek Dalam Masa Pandemi Covid-19," *Akunt. dan Manaj.*, vol. 16, no. 2, pp. 153–167, 2021, doi: 10.30630/jam.v16i2.159.
- [10] S. E. M. S. Hamdani, *MENGENAL USAHA MIKRO KECIL DAN MENENGAH (UMKM) LEBIH DEKAT*. uwais inspirasi indonesia. [Online]. Available: <https://books.google.co.id/books?id=aGwQEAAAQBAJ>
- [11] S. Purwaningsih, *Pranata Sosial dalam Kehidupan Masyarakat*. Alprin, 2020. [Online]. Available: <https://books.google.co.id/books?id=uAYBEAAAQBAJ>