

K-Modes Analysis with Validation of the DBI in Grouping Provinces in Indonesia based on Indicators of Poor Households

Syifa Azahra, Zilrahmi*, Dodi Vionanda, dan Fadhilah Fitri

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: zilrahmi@fmipa.unp.ac.id

Submitted : 07 Mei 2024

Revised : 24 Mei 2024

Accepted : 29 Mei 2024

ABSTRACT

Poverty is the most pressing social problem in Indonesia. Efforts to alleviate poverty are to group provinces in Indonesia based on indicators of poor households using the K-modes algorithm. The data used is data from the 2017 Indonesian Demographic and Health Survey (IDHS) on the Household List. The analysis includes data noise detection, data clustering using K-Modes algorithm, and cluster validation with Davies Bouildin Index (DBI). Based on the clustering that has been done, two clusters are obtained, where cluster 1 consists of 26 provinces and cluster 2 consists of 8 provinces. cluster 1 is a cluster that fulfills 9 indicators of poor households and cluster 2 only a few indicators of poor households. So that the government can prioritize these 8 provinces to overcome poverty in Indonesia. For the DBI value obtained is 1.89 which means that 2 clusters are already well used in the algorithm.

Keywords: Clustering, Davies Bouildin Index (DBI), K-modes, Poverty.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Pertumbuhan penduduk di Indonesia cukup tinggi setiap tahunnya dan persebaran penduduk serta lahan untuk tinggal yang tidak merata yang menimbulkan banyak permasalahan sosial, salah satunya adalah kemiskinan. Kemiskinan menurut Rencana Pembangunan Jangka Menengah Nasional (2004-2009) tidak hanya dipahami sebagai ketidakmampuan ekonomi, tetapi juga tidak terpenuhinya hak-hak dasar seperti, kebutuhan pangan, kesehatan, pekerjaan, perumahan, air bersih, pertahanan dan lingkungan hidup. Menurut Kusumaningtyas (2006) kemiskinan masyarakat diartikan sebagai ketidakmampuan untuk memenuhi hak dasar, kekurangan *asset* rumah tangga, kondisi kekurangan dalam memenuhi kehidupan sehari-hari, pekerjaan yang tidak menentu, pendidikan dan keahlian yang sangat rendah.

Menurut Badan Pusat Statistik (BPS), tingkat penurunan kemiskinan makin berkurang dari tahun ke tahun, dimana dari tahun 1999 sampai tahun 2004 persentase kemiskinan turun sekitar 6,77% dalam 5 tahun, namun dari tahun 2014 sampai tahun 2019 penurunan kemiskinan hanya berkurang 1,74% saja yang artinya ada hambatan dari pengentas kemiskinan di Indonesia. Menurut data BPS (2013-2023) Provinsi Papua, Papua Barat, Nusa Tenggara Timur merupakan tiga provinsi yang memiliki tingkat kemiskinan tertinggi. Kemiskinan suatu provinsi dapat diukur menggunakan indikator rumah tangga miskin. Sehingga perlu dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator rumah tangga miskin, untuk melihat provinsi mana saja yang perlu diutamakan dan indikator rumah tangga miskin apa saja yang ada pada provinsi tersebut. Melalui pengelompokan, pemerintah dapat menyesuaikan strategi pembangunan atau membuat program penanggulangan kemiskinan yang sesuai dengan kebutuhan dan karakteristik masing-masing kelompok provinsi, agar pemerintah tidak perlu mengeluarkan biaya yang besar, waktu yang lama dan juga tenaga dalam mengatasi kemiskinan di Indonesia.

Data yang bisa digunakan untuk melakukan pengelompokan provinsi di Indonesia berdasarkan indikator rumah tangga miskin adalah data Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017. SDKI merupakan survei sosial kependudukan yang dilaksanakan sejak tahun 1987 dan dilakukan setiap 5 tahun sekali. Salah satu kuisioner yang digunakan dalam pelaksanaan SDKI yaitu kuisioner rumah tangga yang digunakan untuk mencatat seluruh ke adaan di dalam rumah tangga dan anggota rumah tangga tersebut. Mengolah data SDKI menggunakan proses data mining digunakan untuk mendapatkan informasi yang berguna dari data mentah yang perlu diubah agar dapat membantu dalam

pengambilan keputusan. Salah satu tujuan data mining yaitu deskriptif, digunakan untuk menemukan pola-pola yang menjelaskan karakteristik dalam suatu data. Teknik data mining yang termasuk data mining deskriptif adalah *clustering*. Yang membedakan penelitian ini dengan penelitian sebelumnya yaitu pada penelitian Sukma Direja tahun 2021, melakukan penelitian tentang pengaruh karakteristik individu rumah tangga terhadap kemiskinan menggunakan regresi berganda sedangkan pada penelitian ini menggunakan algoritma K-Modes untuk mengelompokkan provinsi yang termasuk ke dalam rumah tangga miskin.

Algoritma *clustering* yang digunakan untuk mengelompokkan data bertipe kategorik adalah *K-Modes*. Yulianton (2021) menulis dalam jurnalnya beberapa kelebihan metode *K-Modes* yaitu, bisa diterapkan pada data kategorik, interpretasi yang lebih mudah, dan waktu dalam proses pembentukan *cluster* yang relatif lebih singkat. Validasi *cluster* yang digunakan pada data bertipe kategorik adalah validasi *Davies Bouildin Index* (DBI). Validasi DBI memiliki beberapa kelebihan yaitu bisa digunakan pada data bertipe kategorik, perhitungannya yang lebih sederhana dan tidak rumit, serta melakukan perhitungan jarak antar *cluster* yang memberikan gambaran lebih lengkap tentang pola dan karakteristik data. Maka dari itu tujuan dari penelitian ini yaitu, pertama untuk mengetahui berapa *cluster* yang terbentuk dan provinsi mana saja yang berada pada *cluster* tersebut, kedua untuk mengetahui karakteristik masing-masing *cluster*, dan terakhir untuk mengetahui seberapa baik hasil *cluster* yang digunakan pada algoritma *K-modes* menggunakan validasi DBI pada data SDKI tahun 2017.

II. METODE PENELITIAN

A. Sumber Data

Jenis penelitian yang dilakukan adalah penelitian terapan dan jenis data yang digunakan adalah data sekunder. Data diperoleh dari kuisioner Survei Demografi dan Kesehatan Indonesia 2017 tentang Daftar Rumah Tangga yang terdiri dari 47.083 amatan dengan 12 variabel. Berikut adalah variabel yang digunakan dalam penelitian ini.

Tabel 1. Variabel Penelitian

No	Variabel	Keterangan	Skala Pengukuran
1.	X ₁	Memiliki Mobil	Nominal
2.	X ₂	Sumber Air Minum	Nominal
3.	X ₃	Jenis Toilet	Nominal
4.	X ₄	Bahan Lantai	Nominal
5.	X ₅	Bahan Atap	Nominal
6.	X ₆	Bahan Dinding	Nominal
7.	X ₇	Memiliki Listrik	Nominal
8.	X ₈	Bahan Bakar Memasak	Nominal
9.	X ₉	Menggunakan Rekening Bank	Nominal
10.	X ₁₀	Luas Lantai	Nominal
11.	X ₁₁	Jumlah Anggota Rumah Tangga	Nominal
12.	X ₁₂	Memiliki Motor	Nominal

B. Teknik Analisis Data

Analisis pada penelitian dilakukan menggunakan software *Python*. Setelah variabel dalam analisis dipilih, selanjutnya dilakukan analisis data dengan tahapan sebagai berikut:

1. Melakukan proses *cleaning* pada data atau variabel yang sudah dipilih. *Cleaning* data yang dilakukan yaitu mengidentifikasi data *noise* dengan melihat pada deskriptif data. Data *noise* merupakan kesalahan struktural seperti kesalahan ketik, penggunaan huruf besar dan kecil di dalam data. Penanganan yang dilakukan terhadap data *noise* yaitu dengan menghapus data *noise* pada dataset dikarenakan *noise* tidak bermakna dan kurang dari 5%.
2. Menentukan jumlah *cluster* yang dibentuk pada data menggunakan metode *elbow*. Caranya yaitu dengan melihat perbandingan hasil *Sum of Square Error* (SSE) pada masing-masing jumlah *cluster* dengan persamaan SSE sebagai berikut (Irwanto, 2012).

$$SSE = \sum_{i=1}^k \sum_{x_i \in C_k} |x_i - C_k|^2 \quad (1)$$

Dimana k merupakan jumlah *cluster*, $x_i \in C_k$ merupakan nilai keanggotaan titik data x_i ke pusat kelompok C_k , C_k merupakan pusat *cluster* ke- k dan x_i merupakan data pada objek ke- i . Untuk C_k atau penentuan *clusternya* dilakukan terlebih dahulu menggunakan metode *K-modes* untuk melihat *cluster* sementara.

3. Memilih objek yang paling sering muncul sebagai *centroid* untuk setiap *cluster*.
4. Menghitung tingkat kemiripan masing-masing objek ke *centroid* menggunakan teknik *simple matching* pada persamaan berikut.

$$d(X, Y) = \sum_{j=1}^r \delta(x_j, y_j) \quad (2)$$

Dimana $d(X, Y)$ merupakan jarak data X ke Y , x_j merupakan nilai variabel ke- j dari X , y_j merupakan nilai variabel ke- j dari Y , dan r merupakan jumlah variabel.

Untuk menghitung nilai pencocokan menggunakan persamaan berikut ini.

$$\delta(x_j, y_j) = \begin{cases} 0; & \text{if } x_j = y_j \\ 1; & \text{if } x_j \neq y_j \end{cases} \quad (3)$$

5. Dari hasil perhitungan kemiripan objek (*simple matching*), cari modus dari masing-masing variabel dan bentuk *centroid* baru.
6. Setelah mendapatkan *centroid* baru, selanjutnya membandingkan *centroid* baru dengan *centroid* sebelumnya. Cara membandingkannya adalah dengan melihat kesamaan *centroid* baru dengan *centroid* sebelumnya, jika ada *cluster* yang berubah kembali ke langkah 4 untuk melanjutkan iterasi baru menggunakan *centroid* baru dan jika tidak ada maka lanjutkan ke tahap berikutnya.
7. Memilih modus dari *cluster* pada masing-masing provinsi sehingga hasil *clustering* dapat diperoleh.
8. Menghitung nilai validasi *Davies Bouildin Index* (DBI), untuk melihat seberapa baik hasil dari *cluster* yang digunakan pada algoritma *K-Modes*. Terdapat beberapa tahapan untuk menghitung nilai DBI, pertama yaitu menghitung jarak antar cluster menggunakan persamaan berikut .

$$d(Y_a, Y_b) = \sum_{j=1}^r \delta(y_{aj}, y_{bj}) \quad (4)$$

Dimana $d(Y_a, Y_b)$ merupakan jarak antara *centroid cluster* ke- a dengan *centroid cluster* ke- b , y_{aj} merupakan nilai variabel ke- j pada *centroid cluster* ke- a , y_{bj} merupakan nilai variabel ke- j pada *centroid cluster* ke- b , dan r merupakan jumlah variabel. Menghitung nilai pencocokan menggunakan persamaan berikut ini.

$$\delta(y_{aj}, y_{bj}) = \begin{cases} 0; & \text{if } y_{aj} = y_{bj} \\ 1; & \text{if } y_{aj} \neq y_{bj} \end{cases} \quad (5)$$

Kedua yaitu menghitung jarak masing-masing anggota *cluster* terhadap *centroid* masing-masing *cluster* menggunakan persamaan berikut:

$$d(X_i, Y_a) = \sum_{j=1}^r \delta(x_{ij}, y_{aj}) \quad (6)$$

Dimana $d(X_i, Y_a)$ merupakan jarak antara data *cluster* ke- i dengan *centroid cluster* ke- a , x_{ij} merupakan nilai variabel ke- j pada data ke- i , y_{aj} merupakan nilai variabel ke- j pada *centroid* ke- a , dan r merupakan jumlah variabel. Mengitung nilai pencocokan menggunakan persamaan berikut ini.

$$\delta(x_{ij}, y_{aj}) = \begin{cases} 0; & \text{if } x_{ij} = y_{aj} \\ 1; & \text{if } x_{ij} \neq y_{aj} \end{cases} \quad (7)$$

Ketiga yaitu menghitung dispersi *cluster* menggunakan persamaan berikut:

$$S_a = \left[\frac{1}{n_a} \sum_{X_i \in C_a, i=1}^{n_a} d^2(X_i, Y_a) \right]^{\frac{1}{2}} \quad (8)$$

Dimana S_a merupakan ukuran dispersi cluster, n_a merupakan banyaknya anggota *cluster* ke- a , C_a merupakan *cluster* ke- a , dan X_i merupakan *centroid* ke- i pada cluster ke- a . Tahapan keempat yaitu menghitung rasio menggunakan persamaan berikut:

$$R_{a,b} = \frac{S_a + S_b + \dots + S_n}{d(Y_a, Y_b) + \dots + d(Y_a, Y_b)_{na,nb}} \quad (9)$$

Dimana R_{ab} merupakan Rasio, S_a merupakan dispersi *cluster* ke- a , S_b merupakan dispersi *cluster* ke- b , dan $d(Y_a, Y_b)$ merupakan jarak *centroid cluster* ke- a dengan *centroid cluster* ke- b . Tahapan terakhir yaitu menghitung nilai DBI menggunakan persamaan sebagai berikut:

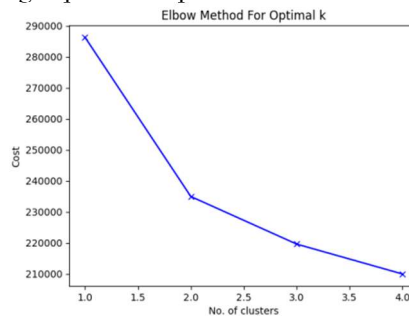
$$DBI = \frac{1}{k} \sum_{a,b=1}^k \max_{a \neq b} (R_{a,b}) \quad (10)$$

Dimana DBI merupakan validasi *Davies Bouldin Index*, $R_{a,b}$ merupakan Rasio antara *cluster a* dan *b*, \max merupakan rasio antara *cluster* yang terbesar, dan k merupakan jumlah *cluster*.

III. HASIL DAN PEMBAHASAN

Data *noise* merupakan kesalahan struktural dalam data, seperti kesalahan ketik hingga penggunaan huruf besar atau kecil yang tidak konsisten. Setelah mendeteksi data *noise* pada analisis ini, terdapat beberapa variabel yang memiliki data *noise* yaitu variabel mobil, listrik, rekening bank, dan motor, yang terdapat kategori angka 9. Dimana kategori tersebut tidak ada di dalam survei yang dilaksanakan. Maka dari itu dilakukan penanganan dengan menghapus data *noise* yang ada di dalam variabel dikarenakan kurang dari 5%.

Menentukan *cluster* yang optimum pada analisis ini dengan menggunakan metode *elbow*. Metode *elbow* disajikan dalam bentuk grafik dengan sumbu x sebagai jumlah *cluster* dan sumbu y sebagai nilai dari SSE yang didapatkan pada masing-masing jumlah *cluster*. Jumlah *cluster* optimum dilihat dari titik-titik pada grafik yang mengalami penurunan drastis yang membentuk patahan menyerupai siku (*elbow*), sehingga pada analisis jumlah *cluster* optimum yang didapatkan yaitu berjumlah 2 *cluster* yang dapat dilihat pada Gambar 1.



Gambar 1. Menentukan jumlah *cluster* optimum

Analisis *clustering* dapat dilakukan setelah menentukan jumlah *cluster* optimum dengan metode *elbow*, selanjutnya dilakukan analisis *clustering* menggunakan metode *K-modes* dengan hasil *cluster* yang diperoleh seperti pada Tabel 2.

Tabel 2. Hasil *clustering* metode *K-Modes*

Cluster	Jumlah anggota cluster	Provinsi
0	26 provinsi	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, Jakarta, Jawa Barat, Jawa Tengah, Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kalimantan Barat, Sulawesi Utara, Sulawesi Selatan, Gorontalo, Nusa Tenggara Barat..
1	8 provinsi	Papua, Papua Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat.

Berdasarkan hasil *clustering* yang telah diperoleh pada Tabel 2, diketahui karakteristik *cluster* 0 dan *cluster* 1 yang dijelaskan pada tabel 3.

Tabel 3. Karakteristik masing-masing *cluster*

<i>Cluster</i>	Karakteristik <i>Cluster</i>
0	Tidak memiliki mobil, sumber air minum dari air yang dibeli, jenis toilet dengan tangki septik, bahan lantai dari porselen, bahan atap genteng, bahan dinding dari batu bata dipleser, memiliki listrik, bahan bakar memasak menggunakan LPG, menggunakan rekening bank, luas lantai lebih dan sama dari 8m ² , jumlah anggota rumah tangga lebih dan sama dari 5, dan memiliki motor.
1	Tidak memiliki mobil, sumber air minum dari sumur, jenis toilet sungai, bahan lantai dari semen, bahan atap seng, bahan dinding dari serat alami, memiliki listrik, bahan bakar memasak menggunakan serat alami, tidak menggunakan rekening bank, luas lantai kurang dari 8m ² , jumlah anggota rumah tangga kurang dari 5, dan memiliki motor.

Menurut Isdijoso (2016), provinsi yang tergolong ke dalam rumah tangga miskin adalah provinsi yang memenuhi 9 indikator rumah tangga miskin. Berdasarkan Tabel 2 dan Tabel 3 terlihat bahwa *cluster* 1 merupakan *cluster* yang memenuhi 9 indikator rumah tangga miskin sehingga Provinsi Papua, Papua Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara, dan Sulawesi Barat merupakan provinsi yang tergolong miskin. Menurut BPS (2016), Provinsi Papua, Papua Barat, Maluku Utara, Nusa Tenggara Timur, Maluku, Sulawesi Tengah, merupakan provinsi wilayah timur yang mendominasi persentase penduduk miskin tertinggi. Sehingga sesuai dengan analisis yang telah dilakukan, dimana wilayah timur tengah juga mendominasi provinsi dengan tingkat kemiskinan tertinggi di Indonesia. Terdapat beberapa faktor yang menyebabkan kemiskinan di wilayah timur, pertama yaitu beberapa wilayah di Indonesia bagian timur sulit diakses karena kondisi geografis yang sulit, seperti pergunungan, pulau-pulau terpencil, dan kurangnya infrastruktur memadai. Kedua yaitu investasi yang cenderung lebih tinggi di wilayah yang lebih maju secara ekonomi seperti Jawa dan Sumatera dapat menyebabkan wilayah timur ketinggalan dalam pembangunan infrastruktur, pendidikan, dan layanan kesehatan. Ketiga yaitu konflik internal, ketidakstabilan politik, dan masalah sosial yang dapat mempengaruhi pertumbuhan dan kesejahteraan di wilayah timur. Maka dari itu pemerintah perlu memprioritaskan 8 provinsi tersebut untuk menangani masalah kemiskinan di Indonesia berdasarkan indikator rumah tangga miskin, agar target penurunan kemiskinan di Indonesia tercapai setiap tahunnya.

Adapun metode validasi *cluster* yang digunakan adalah *Davies Bouildin Index* (DBI), dimana validasi ini digunakan untuk melihat seberapa baik hasil *cluster* dalam algoritma *clustering*. Nilai DBI dikatakan baik apabila nilainya kecil dari 2, dan nilai DBI dikatakan sangat baik apabila nilai tersebut berkisaran antara 0 sampai dengan satu. Setelah dilakukan perhitungan DBI menggunakan *software python* didapatkan hasilnya sebesar 1,94063 dimana nilai ini kecil dari 2 dan mendekati angka 0, yang artinya $k = 2$ merupakan jumlah *cluster* yang baik digunakan pada *cluster* ini. Saran untuk penelitian selanjutnya adalah peneliti bisa menggunakan penangan data *noise* lain seperti mengganti *noise* dengan modus dan untuk validasinya menggunakan *silhoutte*.

IV. KESIMPULAN

Dari hasil pengelompokan provinsi di Indonesia menggunakan analisis *K-Modes* pada data SDKI tahun 2017, didapatkan jumlah *cluster* optimumnya adalah 2, dimana *cluster* 1 merupakan *cluster* yang memenuhi 9 indikator rumah tangga miskin di Indonesia. Provinsi yang ada dalam *cluster* 1 adalah Provinsi Papua, Papua Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara, dan Sulawesi Barat. Berdasarkan data BPS tahun 2017-2023, Provinsi Papua, Papua Barat, Nusa Tenggara Timur masuk ke dalam 3 provinsi yang memiliki tingkat kemiskinan tertinggi, Sehingga pemerintah perlu memprioritaskan provinsi yang berada pada *cluster* 1 untuk menangani masalah kemiskinan di Indonesia, agar pemerintah bisa menyusun strategi atau program penanggulangan kemiskinan sesuai dengan karakteristik *cluster* 1. Hasil validasi DBI adalah sebesar 1,94063 dimana nilai ini mendekati angka 0, yang artinya $k = 2$ baik digunakan untuk algoritma *K-modes*.

DAFTAR PUSTAKA

- Az-zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila, R., & Madjida, W. O. Z. (2021). Penerapan Algoritma *K-Modes Clustering* dengan Validasi *Davies Bouildin Index* pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Matematika dan Statistika serta Aplikasinya*. 9(1), 24-34.
- Badruttaman, A., Sudarno., & Maruddani, D. A. I. (2020). Penerapan Analisis Klaster *K-Modes* Dengan Validasi *Davies Bouildin Index* Dalam Menentukan Karakteristik Kanal Youtube Di Indonesia. *Jurnal Gaussian*. 9(3), 263-266.
- BKKBN, BPS, Kemenkes RI. (2018). *Survei Demografi dan Kesehatan Indonesia*, Jakarta.
- BPS. (2023). *Profil Kemiskinan di Indonesia September 2022*. Jakarta : BPS Indonesia.
- BPS Provinsi Jawa Timur. (2024). *Persentase Penduduk Miskin Menurut Provinsi (Persen)*. Surabaya: BPS Provinsi Jawa Timur.
- Davies, D. L., & Bouildin, D. W. (1979) *A Cluster Separation Measure*. *IEEE Transaaction on Pattern Analysis and Machine Inteligence*. *PAM-1*(2), 224-227.
- Fayyadh, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. AAAI and The MIT Pres. 17(3), 37-53.
- Han, J., & M. Kamber. (2006). *Data Mining Concept and Techniques Second Edition*, San Francisco: Morgan Kaufmann.
- Hidayah, I. N., Suhery, C., Hidayati, R. (2022). Implementasi Metode *K-Modes* Dalam Pengelompokan Penerima Bantuan Langsung Tunai. *Jurnal Riset Komputer*. 9(6), 2102-2108.
- Huang, J. Z. (2009). *Clustering Categorical Data with K-Modes*, Hongkong: Universitas Hongkong.
- Huang, Z. (1997). *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining*, Canberra: Cooperative Research Centre for Advanced Computational Systems.
- Irwanto, Purwananto, Y., & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klasterisasi *K-Means* untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*. 1(1), 197-202.
- Isdijoso, W., Suryahadi, A., & Akhmadi. (2016). *Penetapan Kriteria Variabel Pendataan Penduduk Miskin yang Komprehensif dalam Rangka Perlindungan Penduduk Miskin di Kabupaten/Kota*, Jakarta: The SMERU Research Institute.
- Kusumaningtyas, A. D. (2006). *Kemiskinan Masyarakat di Sekitar Kawasan Industri Jababeka (Studi Kasus Desa Pasir Gombang, Kecamatan Cikarang Utara, Kabupaten Bekasi, Provinsi Jawa Barat)*. (Skripsi Sarjana, Institut Pertanian Bogor)
- Yulianton, H., Sutanto, F. A., & Mulyani, S. (2021). Pengelompokan Mahasiswa Berbasis *Categorical Variables* Menggunakan Metode *K-Modes Clustering*. *Proceeding SENDIU*. 424-249.