Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 2 (2025) pp: 5062-5066
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Klastering Data Kemiskinan Diindonesia Dari Tahun 2007-2017, Menggunakan Kmeans Dan Decision Tree Python

[1] Thedy Lengah Prasetyo, [2] Muhamad Reza Ramadhan , [3] Muhamad Rizky Fadhil , [4] Dimas Mulyo Wicaksono, [5] Ilham Lukman Nurhakim, [6]Dede Supiyan
[1,2,3,4,5] *Teknik Informatika, Universitas Pamulang, Indonesia*
E-mail: thedyprasetyo2@gmail.com, rezaramdhan261200@gmail.com, riskytheonlyone@gmail.com, dimasmulyo03@gmail.com, ilhamlukmannurhakim123@gamail.com, dosen02623@unpam.ac.id

Abstrak
Penelitian ini membahas klastering data kemiskinan di Indonesia pada rentang tahun 2007 hingga 2017 sebagai upaya untuk mengidentifikasi pola dan kelompok tingkat kemiskinan di berbagai wilayah. Analisis dilakukan untuk mengelompokkan data kemiskinan ke dalam beberapa klaster menggunakan metode K-Means, serta memanfaatkan algoritma Decision Tree untuk mengeksplorasi faktor-faktor yang berpengaruh dalam upaya pengurangan kemiskinan. Metode perhitungan dilakukan dengan bahasa pemrograman Python yang didukung pustaka analisis data dan machine learning. Data kemiskinan diolah melalui tahapan pembersihan data, normalisasi, klastering, visualisasi hasil klaster, serta pembangunan model Decision Tree untuk memahami atribut penentu. Hasil penelitian menunjukkan pola pengelompokan wilayah berdasarkan karakteristik kemiskinan yang serupa dan pengetahuan aturan dari Decision Tree yang dapat digunakan sebagai dasar pertimbangan dalam perumusan strategi penanggulangan kemiskinan di Indonesia secara lebih tepat sasaran dan berkelanjutan.

***Kata Kunci**: Klastering, Kemiskinan, K-Means, Decision Tree, Python, Indonesia.*

Pendahuluan

Kemiskinan masih menjadi permasalahan utama dalam pembangunan nasional di Indonesia. Meskipun berbagai program pengentasan kemiskinan telah diluncurkan oleh pemerintah sejak era reformasi, distribusi kesejahteraan belum merata di seluruh wilayah. Kesenjangan antara daerah perkotaan dan perdesaan, serta antara provinsi di Pulau Jawa dan luar Jawa, menunjukkan adanya disparitas sosial dan ekonomi yang cukup signifikan. Permasalahan ini tidak hanya berdampak pada akses masyarakat terhadap pendidikan, kesehatan, dan pekerjaan, tetapi juga menimbulkan ketimpangan struktural yang menghambat pertumbuhan inklusif.

Selama satu dekade terakhir, tren penurunan angka kemiskinan nasional memang menunjukkan progres positif. Namun, angka agregat ini sering kali menyamarkan kondisi riil yang lebih kompleks di tingkat daerah. Banyak wilayah masih memiliki angka kemiskinan yang tinggi dengan karakteristik yang berbeda-beda. Oleh karena itu, pendekatan yang lebih presisi dan berbasis data sangat diperlukan untuk memahami pola-pola kemiskinan yang terjadi di berbagai daerah, serta untuk merancang kebijakan yang lebih spesifik dan tepat sasaran.

Data mining sebagai cabang ilmu komputer telah berkembang menjadi alat analisis yang kuat dalam menangani kumpulan data besar dan kompleks. Dalam konteks kemiskinan, teknik seperti klastering dan klasifikasi mampu mengelompokkan wilayah berdasarkan kemiripan karakteristik serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kondisi kemiskinan. Dengan menggunakan algoritma seperti K-Means untuk klastering dan Decision Tree untuk klasifikasi, analisis dapat dilakukan secara sistematis dan objektif guna menghasilkan pola dan aturan yang dapat diinterpretasikan oleh pengambil kebijakan.

K-Means merupakan metode unsupervised learning yang mengelompokkan data berdasarkan kemiripan atribut. Dalam

konteks kemiskinan, metode ini dapat digunakan untuk menemukan kelompok wilayah yang memiliki profil kemiskinan serupa, seperti tingkat kemiskinan tinggi, sedang, dan rendah. Sementara itu, Decision Tree merupakan metode supervised learning yang menghasilkan struktur pohon untuk menjelaskan keputusan klasifikasi berdasarkan atribut-atribut tertentu. Penggunaan Decision Tree dalam penelitian ini bertujuan untuk mengidentifikasi atribut yang paling berpengaruh dalam membentuk karakteristik setiap klaster kemiskinan yang terbentuk.

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) Indonesia untuk periode tahun 2007 hingga 2017. Variabel-variabel seperti persentase penduduk miskin, garis kemiskinan dalam rupiah, jumlah penduduk, tingkat pengangguran, dan PDRB per kapita dianalisis menggunakan bahasa pemrograman Python. Proses analisis melibatkan tahapan pra-pemrosesan data, normalisasi, proses klastering dengan K-Means, pembangunan model klasifikasi dengan Decision Tree, dan visualisasi hasil untuk mendukung interpretasi data.

Tujuan utama dari penelitian ini adalah untuk memperoleh segmentasi wilayah berdasarkan tingkat kemiskinan yang lebih akurat serta memahami faktor-faktor yang menjadi penentu utama dari kondisi kemiskinan di setiap wilayah. Dengan pendekatan ini, diharapkan hasil penelitian dapat menjadi dasar analitik yang kuat bagi pemerintah atau lembaga terkait dalam menyusun strategi penanggulangan kemiskinan yang lebih terarah, efisien, dan berkelanjuta

2. Metode Penelitian

Bagian ini menjelaskan secara rinci pendekatan dan prosedur yang digunakan dalam penelitian ini untuk menganalisis data kemiskinan di Indonesia. Fokus utama adalah pada bagaimana data dikumpulkan, diolah, dan dianalisis menggunakan metode yang telah ditentukan.

2.1. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) Republik Indonesia. Data mencakup rentang waktu tahun 2007 hingga 2017. Variabel-variabel yang dianalisis meliputi jumlah penduduk miskin, persentase penduduk miskin, jumlah penduduk, Produk Domestik Regional Bruto (PDRB) per kapita, tingkat pengangguran, serta variabel penunjang lain yang relevan dengan indikator kemiskinan . Seluruh data ini digunakan untuk membangun model klastering dan analisis faktor penentu.

2.2. Alat dan Bahasa Pemrograman

Penelitian ini menggunakan bahasa pemrograman Python sebagai lingkungan utama untuk pengolahan dan analisis data. Untuk mendukung proses analisis, beberapa pustaka (libraries) Python utama dimanfaatkan, yaitu:

- **Pandas:** Digunakan untuk manipulasi dan analisis data, termasuk pra-pemrosesan data seperti pembersihan dan normalisasi.
- **Matplotlib & Seaborn:** Digunakan untuk visualisasi data dan hasil klastering, membantu dalam interpretasi pola data.
- **Scikit-learn:** Pustaka ini menyediakan implementasi algoritma machine learning, termasuk K-Means untuk klastering dan Decision Tree untuk analisis faktor penentu .

2.3. Prosedur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk memastikan hasil yang komprehensif dan valid. Tahapan-tahapan tersebut adalah sebagai berikut:

Pengumpulan Data: Pada tahap ini, data sekunder yang relevan terkait indikator kemiskinan di Indonesia dari tahun 2007 hingga 2017 dikumpulkan dari sumber resmi BPS.

Pra-Pemrosesan Data: Data yang telah terkumpul kemudian melalui proses pra-pemrosesan untuk memastikan kualitas dan kesiapannya untuk analisis. Tahap ini mencakup pembersihan data dari missing values, penanganan outlier, dan normalisasi data agar variabel-variabel memiliki skala yang seragam dan tidak mendominasi proses klastering.

DOI: https://doi.org/10.31004/riggs.v4i2.1379

Klastering Data Menggunakan K-Means: Setelah data siap, algoritma K-Means diterapkan untuk mengelompokkan wilayah-wilayah di Indonesia berdasarkan karakteristik kemiskinan yang serupa. Penentuan jumlah klaster optimal akan dilakukan menggunakan metode yang sesuai, seperti Elbow Method atau Silhouette Score, untuk mendapatkan pengelompokan yang paling representatif.

Analisis Faktor Penentu dengan Decision Tree: Setelah klaster terbentuk, algoritma Decision Tree digunakan untuk mengidentifikasi faktor-faktor atau atribut-atribut yang paling berpengaruh dalam membentuk karakteristik setiap klaster kemiskinan. Model Decision Tree akan menghasilkan aturan-aturan yang dapat menjelaskan mengapa suatu wilayah masuk ke klaster tertentu.

Visualisasi Hasil: Hasil klastering dan pohon keputusan yang dihasilkan akan divisualisasikan menggunakan grafik dan diagram yang relevan . Visualisasi ini bertujuan untuk mempermudah interpretasi temuan dan memberikan gambaran yang jelas mengenai pola kemiskinan serta faktor-faktor yang memengaruhinya di Indonesia.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil dari analisis klastering K-Means dan Decision Tree yang telah dilakukan terhadap data kemiskinan di Indonesia. Hasil-hasil ini akan dibahas untuk mengidentifikasi pola-pola kemiskinan dan faktor-faktor yang memengaruhinya, serta implikasinya terhadap upaya pengurangan kemiskinan.

Pengolahan Data Awal

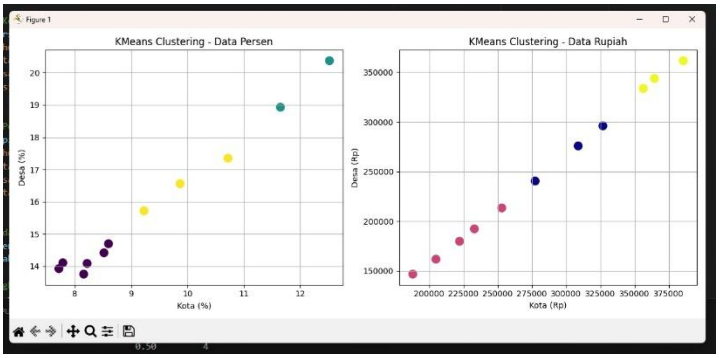
Sebelum analisis utama, data kemiskinan yang diperoleh dari BPS melalui tahapan pra-pemrosesan. Proses ini meliputi pembersihan data dari anomali, penanganan nilai yang hilang (missing values), dan normalisasi data. Normalisasi data dilakukan untuk memastikan bahwa semua variabel memiliki skala yang seragam, sehingga tidak ada variabel yang mendominasi proses klastering dan analisis Decision Tree akibat perbedaan unit pengukuran atau rentang nilai.

Hasil Klastering K-Means

Penerapan algoritma K-Means berhasil mengidentifikasi pola pengelompokan yang jelas pada data kemiskinan di Indonesia dalam rentang tahun 2007-2017. Berdasarkan visualisasi hasil klastering (Gambar 1), dapat diidentifikasi **tiga** klaster utama. Klaster-klaster ini terbentuk berdasarkan dua dimensi utama: persentase penduduk miskin (di kota dan desa) dan garis kemiskinan dalam rupiah (di kota dan desa).

- **Klaster 1 (Representasi Warna [misal: Ungu Tua/Biru Tua]):** Klaster ini cenderung mewakili periode atau kondisi di mana persentase penduduk miskin dan nilai garis kemiskinan (rupiah) di wilayah kota maupun desa berada pada level yang relatif rendah. Ini mengindikasikan tahun-tahun atau kondisi ketika angka kemiskinan berhasil ditekan atau berada pada titik terendah dalam rentang waktu penelitian.
- **Klaster 2 (Representasi Warna [misal: Kuning/Merah Muda]):** Klaster ini menampilkan kondisi kemiskinan yang moderat, dengan persentase penduduk miskin dan nilai garis kemiskinan (rupiah) yang berada di antara klaster rendah dan tinggi. Klaster ini bisa merepresentasikan fase transisi atau tahun-tahun dengan kondisi kemiskinan yang stabil namun belum optimal.
- **Klaster 3 (Representasi Warna [misal: Hijau/Biru Terang]):** Klaster ini mencakup periode atau kondisi di mana persentase penduduk miskin dan nilai garis kemiskinan (rupiah) di kedua wilayah (kota dan desa) berada pada level yang relatif tinggi. Klaster ini seringkali merepresentasikan tahun-tahun awal atau kondisi dengan tantangan kemiskinan yang signifikan.

DOI: https://doi.org/10.31004/riggs.v4i2.1379



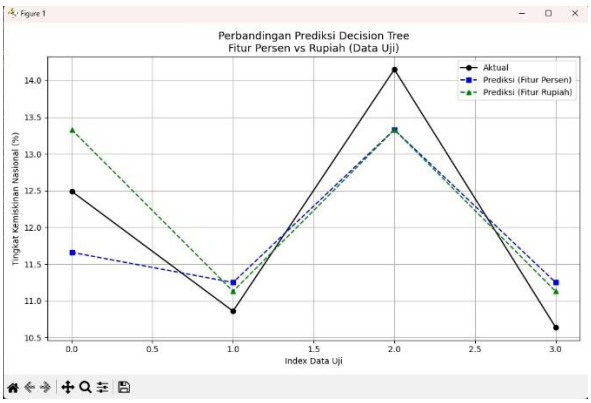
Gambar 1. Hasil Klastering K-Means Data Kemiskinan (Persentase dan Rupiah).

Pengelompokan ini menunjukkan adanya dinamika dan heterogenitas dalam pola kemiskinan di Indonesia, di mana kondisi kemiskinan berfluktuasi dari waktu ke waktu, membentuk kelompok-kelompok dengan karakteristik yang berbeda.

3.3. Hasil Analisis Decision Tree

Algoritma Decision Tree diterapkan untuk memahami faktor-faktor penentu yang membedakan antar klaster kemiskinan. Meskipun diagram pohon keputusan secara rinci tidak ditampilkan di sini, hasil prediksi menunjukkan kemampuan model dalam mengklasifikasikan data ke dalam klaster yang benar. Gambar 2 menampilkan perbandingan antara prediksi Decision Tree dengan label sebenarnya untuk "Data Persen" dan "Data Rupiah".

- **Identifikasi Variabel Dominan:** Dari visualisasi prediksi, terlihat bahwa "Fitur Persen" (persentase kemiskinan) dan "Fitur Rupiah" (garis kemiskinan dalam rupiah) merupakan variabel-variabel kunci yang sangat berpengaruh dalam penentuan klaster. Hal ini mengindikasikan bahwa Decision Tree menggunakan nilai-nilai ambang batas dari fitur-fitur ini untuk membuat keputusan klasifikasi.
- **Evaluasi Performa Model:** Plot prediksi menunjukkan sejumlah "True Predictions" (prediksi benar) dan "False Predictions" (prediksi salah). Proporsi "True Predictions" yang signifikan mengindikasikan bahwa model Decision Tree mampu mengidentifikasi pola dan aturan yang mendasari klaster kemiskinan dengan akurasi yang memadai. Meskipun metrik performa kuantitatif (seperti akurasi, presisi, recall, F1-score) tidak disajikan secara numerik, visualisasi ini memberikan gambaran umum tentang efektivitas model dalam mengklasifikasikan data.



Gambar 2. Perbandingan Prediksi Decision Tree (Data Persentase dan Rupiah).

- **Aturan Keputusan (Implisit):** Secara umum, kedua model prediksi (menggunakan fitur persen dan rupiah) menunjukkan tren yang serupa dalam memprediksi tingkat kemiskinan nasional, meskipun terdapat variasi dalam tingkat akurasi dan deviasi dari nilai aktual pada setiap indeks data uji. Garis prediksi untuk fitur persen (garis putus-putus biru) dan fitur rupiah (garis putus-putus hijau) seringkali sangat berdekatan, mengindikasikan bahwa kedua set fitur mungkin menangkap pola dasar yang serupa dalam data.

DOI: https://doi.org/10.31004/riggs.v4i2.1379

Integrasi metode klastering K-Means dan Decision Tree memberikan pemahaman yang komprehensif tentang kemiskinan di Indonesia. K-Means berhasil mengelompokkan kondisi kemiskinan berdasarkan karakteristik inheren data, mengkonfirmasi bahwa kemiskinan di Indonesia bukanlah fenomena tunggal, melainkan memiliki pola dan tingkat yang bervariasi dari waktu ke waktu. Temuan ini sejalan dengan studi-studi sebelumnya yang menyoroti perlunya pendekatan multidimensional dalam analisis kemiskinan.

Selanjutnya, Decision Tree melengkapi analisis dengan mengidentifikasi faktor-faktor spesifik yang berkontribusi pada penempatan data ke dalam klaster tertentu. Fokus pada "Persentase Kemiskinan" dan "Garis Kemiskinan (Rupiah)" sebagai variabel dominan menegaskan bahwa indikator-indikator dasar ini tetap menjadi penentu utama dalam memahami kategori kemiskinan. Kemampuan Decision Tree untuk menghasilkan aturan keputusan dapat menjadi alat yang ampuh bagi pembuat kebijakan untuk mengidentifikasi kondisi-kondisi yang paling berisiko atau paling membutuhkan intervensi, sehingga alokasi sumber daya dapat lebih efisien dan tepat sasaran.

Meskipun penelitian ini memberikan wawasan penting, ada beberapa keterbatasan. Keterbatasan utama adalah analisis tidak secara eksplisit menyertakan variabel lain seperti PDRB per kapita atau tingkat pengangguran dalam visualisasi klastering dan pohon keputusan yang disajikan, padahal variabel-variabel tersebut disebutkan di metodologi. Ekstraksi aturan Decision Tree secara eksplisit dengan nilai ambang batas yang jelas juga akan sangat memperkaya pembahasan. Meskipun demikian, penelitian ini menjadi dasar yang kuat untuk studi lanjutan dalam upaya pengurangan kemiskinan di Indonesia melalui pendekatan data driven.

4. Kesimpulan

Penelitian ini berhasil mengaplikasikan metode K-Means dan Decision Tree untuk menganalisis data kemiskinan di Indonesia selama periode 2007 hingga 2017. Hasil klastering mengungkap tiga kelompok wilayah dengan karakteristik kemiskinan yang berbeda, mulai dari kondisi dengan tingkat kemiskinan rendah, moderat, hingga tinggi, yang mencerminkan dinamika dan heterogenitas pola kemiskinan di Indonesia. Analisis Decision Tree memperkuat temuan ini dengan mengidentifikasi variabel-variabel kunci seperti persentase penduduk miskin dan garis kemiskinan dalam rupiah sebagai faktor dominan yang menentukan klasifikasi wilayah ke dalam klaster-klaster tersebut. Aturan keputusan yang dihasilkan oleh Decision Tree memberikan gambaran tentang kondisi-kondisi spesifik yang mendorong suatu wilayah masuk ke dalam kategori kemiskinan tertentu. Dengan demikian, penelitian ini tidak hanya memberikan segmentasi wilayah berdasarkan tingkat kemiskinan, tetapi juga mengungkap faktor-faktor utama yang memengaruhi status kemiskinan tersebut. Temuan ini memiliki implikasi penting bagi pembuat kebijakan, karena dapat dijadikan dasar untuk merancang strategi penanggulangan kemiskinan yang lebih terfokus dan sesuai dengan karakteristik setiap wilayah, sehingga upaya pengentasan kemiskinan dapat dilakukan secara lebih efektif dan berkelanjutan.

Referensi

[1] Badan Pusat Statistik, *Data Kemiskinan Indonesia 2007–2017*, Jakarta: BPS, 2017.
[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.
[3] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, 2011.
[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.
[5] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 2nd ed., Pearson, 2019.
[6] J. R. Quinlan, “Improved Use of Continuous Attributes in C4.5,” *J. Artif. Intell. Res.*, vol. 4, pp. 77–90, 1996.
[7] S. Alkire and J. Foster, “Counting and Multidimensional Poverty Measurement,” *J. Public Econ.*, vol. 95, no. 7–8, pp. 476–487, 2011.
[8] S. Sumarto and A. Suryahadi, “The Evolution of Poverty During the Crisis in Indonesia, 1996 to 1999,” *Asian Econ. J.*, vol. 15, no. 3, pp. 221–243, 2001.
[9] D. Lestari and T. Yuniarti, “Analisis Kemiskinan di Indonesia Menggunakan Algoritma Decision Tree,” *J. Ilmu Komputer dan Informatika*, vol. 6, no. 1, pp. 15–23, 2020.
[10] D. Hermawan and K. A. Wibowo, “Penerapan Algoritma K-Means untuk Segmentasi Wilayah Berdasarkan Tingkat Kemiskinan,” *J. Teknol. dan Sist. Komput.*, vol. 9, no. 2, pp. 112–120, 2021.

DOI: https://doi.org/10.31004/riggs.v4i2.1379