

PENERAPAN K-MEANS CLUSTERING GUNA IDENTIFIKASI PENDUDUK MISKIN EKSTREM YANG BELUM TERAKSES PROGRAM PERLINDUNGAN SOSIAL

Muhammad Ikhsan, Dwi Safiroh Utsalina* dan Tubagus Mohammad Akhriza

S1 Sistem Informasi, STMIK PPKIA Pradnya Paramita, Malang, Jawa Timur

ikhsan24510010@stimata.ac.id, utsalina@stimata.ac.id, akhriza@stimata.ac.id

* Corresponding Author

Abstrak

Kemiskinan masih menjadi hambatan utama dalam mewujudkan kesejahteraan sosial. Berdasarkan data BPS Kabupaten Malang tahun 2023, jumlah penduduk miskin mencapai 251,36 ribu jiwa. Upaya penghapusan kemiskinan ekstrem menjadi prioritas strategis daerah dengan memanfaatkan data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) yang diperbarui setiap tahun oleh Kemenko PMK. Namun, belum adanya model segmentasi sasaran sering menimbulkan masalah seperti duplikasi penerima, *inclusion error*, *exclusion error*, hingga ketidakakuratan data. Penelitian ini menerapkan metode K-Means Clustering dalam kerangka kerja CRISP-DM untuk melakukan segmentasi penduduk miskin dengan atribut jenis pekerjaan, kepemilikan rumah, kepemilikan PKH, dan jumlah anggota keluarga berisiko stunting. Dari 7.857 data, jumlah klaster optimal ditentukan sebanyak $k=5$ menggunakan uji Silhouette Coefficient dan WCSS. Hasilnya terbentuk lima segmen prioritas, yaitu Sangat Prioritas (2.332 KK), Prioritas Tinggi (1.353 KK), Prioritas Sedang (1.545 KK), Kurang Prioritas (284 KK), dan Bukan Prioritas (2.343 KK). Segmentasi ini dapat menjadi dasar perumusan kebijakan penyelaruan program perlindungan sosial yang lebih tepat sasaran dan efektif.

Kata Kunci : K-Means, Silhouette Coefficient, WCSS, Segmentasi Data Kemiskinan,

PENDAHULUAN

Pembangunan pada dasarnya bertujuan meningkatkan kesejahteraan masyarakat, yang mencakup pemenuhan kebutuhan dasar, pendidikan, kesehatan, serta kehidupan yang aman dan layak [1] [2]. Namun, kemiskinan masih menjadi hambatan utama. Di Kabupaten Malang, tahun 2023 tercatat 9,45% penduduk miskin, dengan 0,30% di antaranya tergolong miskin ekstrem [3] [4]. Kecamatan Bantur sebagai wilayah rawan bencana memiliki tingkat kerentanan sosial-ekonomi yang tinggi, sehingga program perlindungan sosial berperan penting untuk memperkuat ketahanan masyarakat. [5] Akan tetapi, pelaksanaannya menghadapi kendala berupa ketidadaan model segmentasi data, yang memicu duplikasi penerima, *inclusion error*, *exclusion error*, serta ketidaktepatan sasaran bantuan [6].

Untuk mengatasi persoalan tersebut, K-Means Clustering dipandang relevan dalam mengelompokkan data penduduk miskin berdasarkan karakteristik sosial-ekonomi. Algoritma ini dikenal sederhana, efisien, serta mampu mengolah data dalam

jumlah besar dengan baik [7]. Optimasi jumlah klaster dilakukan menggunakan Silhouette Coefficient dan Within Cluster Sum of Squares (WCSS) agar segmentasi yang dihasilkan lebih valid dan dapat dijadikan dasar penetapan prioritas sasaran penerima perlindungan sosial.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan menerapkan K-Means Clustering untuk mengidentifikasi penduduk miskin ekstrem yang belum terakses program perlindungan sosial di Kecamatan Bantur. Penelitian difokuskan pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) dengan atribut jenis pekerjaan, kepemilikan rumah, jumlah anggota keluarga berisiko stunting, dan kepemilikan Program Keluarga Harapan (PKH). Analisis dilakukan menggunakan KNIME Analytics Platform guna menghasilkan segmentasi prioritas yang dapat mendukung kebijakan penurunan kemiskinan ekstrem.

Penerapan algoritma K-Means dalam data mining telah banyak dilakukan, salah satunya oleh Rivallinata yang menggunakan K-Means untuk mengklaster

data kemiskinan provinsi di Indonesia berdasarkan empat indikator utama. [7] Penelitian tersebut menghasilkan tiga klaster (tinggi, sedang, rendah) dengan evaluasi menggunakan Silhouette Coefficient. Penelitian lain oleh Febriansyah dan Muntari menerapkan K-Means untuk klasterisasi penerima Program Keluarga Harapan (PKH) di Kota Pagar Alam dengan 471 record, menghasilkan tiga klaster, serta dievaluasi menggunakan Davies Bouldin Index (DBI) [8].

Kedua penelitian tersebut menunjukkan efektivitas K-Means dalam segmentasi data sosial, namun pengelompokan data kemiskinan di tingkat lokal masih belum optimal. Oleh karena itu, penelitian ini menggunakan K-Means dengan kerangka kerja CRISP-DM melalui tahapan *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* guna menghasilkan segmentasi penduduk miskin ekstrem yang lebih terarah

METODE PENELITIAN

Kemiskinan ekstrem adalah kondisi ketidakmampuan memenuhi kebutuhan dasar, seperti pangan, air bersih, sanitasi, kesehatan, pendidikan, perumahan, dan layanan sosial. Batasnya ditetapkan sebesar USD 1,9 PPP per hari atau Rp 10.739/orang/hari setara Rp 322.170/orang/bulan; sehingga keluarga dengan empat anggota dikategorikan miskin ekstrem bila pengeluaran kurang atau lebih dari Rp 1.288.680/bulan [9]. Strategi percepatan penghapusannya diatur dalam KMK PMK No. 32/2022 melalui tiga kebijakan utama: pengurangan beban pengeluaran, peningkatan pendapatan, serta penurunan kantong kemiskinan [10]. Adapun penghitungan jumlah dan persentase penduduk miskin ekstrem dilakukan melalui Susenas oleh BPS.

Data mining merupakan salah satu tahap dalam *Knowledge Discovery in Database (KDD)* yang berfungsi untuk mengklasifikasikan, memprediksi, memperkirakan, serta mengekstraksi informasi bermanfaat dari data berukuran besar [11]. Secara konsep, data mining adalah proses penggalian dan identifikasi

pengetahuan dari basis data besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin [12]. Metode ini memungkinkan ditemukannya pola serta informasi berharga dari sekumpulan data sehingga dapat dimanfaatkan dalam pengambilan keputusan [13].

Clustering adalah proses pengelompokan objek data ke dalam beberapa cluster berdasarkan kemiripan karakteristik, di mana objek dalam satu cluster memiliki kesamaan tinggi, sedangkan antar cluster berbeda secara signifikan [14]. Teknik ini merupakan metode utama dalam *data mining* untuk menemukan pola dan pengetahuan baru yang berguna dalam pengambilan keputusan. Tahapan umum dalam clustering meliputi:

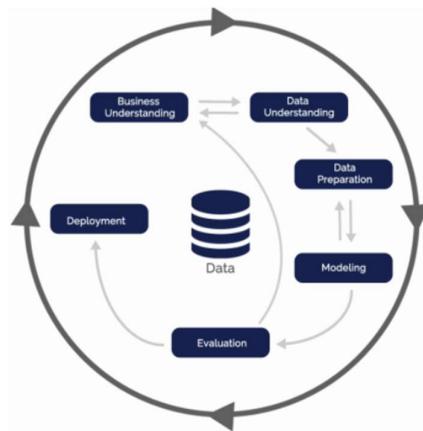
1. Pemilihan Data: memilih data yang relevan, berkualitas, dan memiliki jumlah yang memadai untuk klasterisasi.
2. Persiapan Data (Preprocessing): meliputi seleksi atribut, transformasi data (normalisasi, standardisasi, diskritisasi), serta pembersihan dari missing values dan outlier.
3. Pemilihan Algoritma: menentukan algoritma klasterisasi yang sesuai dengan jenis data, bentuk cluster, jumlah cluster, sensitivitas terhadap noise/outlier, serta skalabilitas (misalnya K-Means, DBSCAN, atau Hierarchical Clustering).
4. Penentuan Jumlah Cluster: menentukan jumlah cluster optimal dengan metode seperti Silhouette Coefficient atau Elbow Method melalui perhitungan Sum of Square Error (SSE).
5. Penerapan Algoritma: menjalankan algoritma klasterisasi pada data yang telah dipersiapkan untuk membentuk cluster.
6. Evaluasi Hasil: menilai kualitas hasil klasterisasi menggunakan metrik internal (Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index) maupun eksternal (Purity, Rand Index, F-Measure).
7. Interpretasi dan Visualisasi: memahami karakteristik utama tiap cluster dan memvisualisasikannya untuk

mendapatkan wawasan yang lebih bermakna

CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah salah satu *framework* data mining yang awalnya dikembangkan oleh lima perusahaan: Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation, dan OHRA, kemudian disempurnakan oleh ratusan organisasi di Eropa sehingga menjadi metodologi standar non-proprietary bagi data mining [15]. Proses CRISP-DM terdiri dari enam tahapan utama sebagai berikut:

1. Business Understanding: memahami tujuan bisnis, merumuskan permasalahan data mining, menetapkan batasan, serta menyusun strategi awal untuk mencapai tujuan.
2. Data Understanding: mengumpulkan dan memeriksa data, membuat ringkasan, mengidentifikasi masalah (seperti *missing values* dan *outlier*), serta menyiapkan dasar analisis awal.
3. Data Preparation: memperbaiki masalah data, melakukan transformasi, pembersihan, serta membagi data menjadi *training set* dan *testing set* agar siap digunakan pada tahap modeling.
4. Modeling: membangun model prediktif atau deskriptif menggunakan teknik statistik maupun *machine learning* (misalnya klasifikasi, clustering, atau scoring) dengan algoritma data mining yang sesuai.
5. Evaluation: mengevaluasi hasil pemodelan untuk memastikan kesesuaian model dengan tujuan bisnis yang ditetapkan pada tahap awal.
6. Deployment: menerapkan model ke dalam lingkungan nyata, baik untuk pengambilan keputusan maupun integrasi ke sistem operasional.



Gambar 1. Siklus CRISP-DM

Metode K-Means

K-Means merupakan salah satu metode *clustering* non-hierarki yang mempartisi data ke dalam sejumlah *cluster* berdasarkan kemiripan karakteristik objek [16] [17]. Tujuan utama *clustering* adalah meminimalkan variasi dalam *cluster* dan memaksimalkan variasi antar *cluster*. Dibandingkan dengan metode *clustering* hierarki, K-Means lebih sederhana, cepat, dan efisien dalam membagi kumpulan data ke dalam jumlah *cluster* tertentu (k). Algoritma dasar K-Means meliputi langkah-langkah:

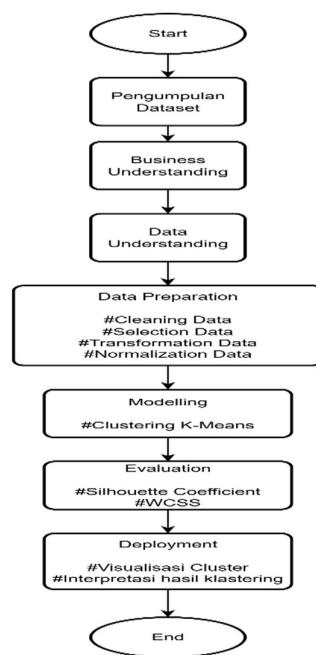
1. Menentukan jumlah cluster (k).
2. Menentukan pusat cluster awal (*centroid*) secara acak.
3. Menghitung jarak setiap data ke *centroid*, lalu mengelompokkan data ke *cluster* terdekat.
4. Memperbarui posisi *centroid* berdasarkan rata-rata data dalam *cluster*.
5. Mengulangi langkah 3 dan 4 hingga posisi *centroid* stabil atau tidak ada perubahan signifikan dalam keanggotaan *cluster*.

KNIME Analytics Platform

KNIME Analytics Platform adalah perangkat lunak *open-source* untuk analisis data, pelaporan, dan integrasi yang dikembangkan oleh University of Konstanz sebagai kerangka kerja *data mining* modular dengan antarmuka intuitif [18]. KNIME banyak dimanfaatkan di berbagai bidang, seperti analisis uji klinis dalam industri

farmasi, segmentasi pelanggan pada pemasaran digital, deteksi penipuan di sektor keuangan, analisis sentimen media sosial, serta pengelompokan data menggunakan modul *clustering* untuk menemukan pola dan pengetahuan baru [19].

Pada penelitian ini dilakukan segmentasi penduduk miskin di Kabupaten Malang menggunakan algoritma K-Means. Segmentasi ini memudahkan pemerintah dalam menentukan sasaran prioritas program bantuan sosial berdasarkan data dasar penduduk miskin ekstrem. Jumlah klaster optimal ditentukan dengan metode Silhouette Coefficient dan WCSS, kemudian digunakan sebagai dasar segmentasi. Hasil segmentasi diinterpretasikan sesuai atribut data, dengan tahapan penelitian ditunjukkan pada gambar berikut.



Gambar 2. Tahapan Penelitian

Pengumpulan Dataset

Penelitian dilaksanakan di Bappeda Kabupaten Malang pada bidang Pemerintahan dan Pembangunan Manusia (PPM). Data yang diperoleh berupa P3KE Kecamatan Bantur, mencakup atribut tingkat kesejahteraan, status kependudukan, tempat tinggal, sosial ekonomi, pendidikan, kesehatan, dan kepemilikan bantuan sosial.

Business Understanding

Target penghapusan kemiskinan ekstrem di Kabupaten Malang sebesar 0% pada tahun 2024 menuntut intervensi yang sistematis dari tingkat desa hingga daerah. Penyaluran bantuan sosial yang masih berbasis data mentah berpotensi tidak tepat sasaran, sehingga diperlukan pengelompokan data P3KE menggunakan algoritma K-Means untuk menghasilkan segmentasi masyarakat yang dapat menjadi dasar kebijakan.

Data Understanding

Data yang digunakan berupa file Excel (.xls) berisi 7.857 record kepala keluarga dengan 27 atribut pendukung. Masing-masing atribut menunjukkan indikator kesejahteraan, namun hanya atribut yang relevan yang dipertahankan untuk proses clustering agar hasil analisis lebih terfokus.

Data Preparation

Tahap persiapan data meliputi pembersihan, seleksi, transformasi, dan normalisasi. Data terbukti bersih dari nilai kosong atau salah ketik. Seleksi atribut dilakukan dengan mengacu pada P3KE, sehingga dipilih 4 (empat) atribut utama yaitu jenis pekerjaan, kepemilikan rumah, kepemilikan PKH, dan jumlah anggota keluarga berisiko stunting. Atribut non-numerik kemudian ditransformasikan ke bentuk numerik menggunakan kodeifikasi, dan semua atribut dinormalisasi dengan metode min-max normalization di KNIME agar berada pada skala yang seragam.

Modelling

Pemodelan dilakukan menggunakan algoritma K-Means pada KNIME Analytics Platform dengan node utama Excel Reader, Normalizer, K-Means, Denormalizer, dan Parallel Coordinates Plot Chart. Data dibaca melalui Excel Reader, dinormalisasi menggunakan metode min-max, lalu diklasterisasi dengan K-Means menggunakan jumlah awal 4 (empat) klaster. Hasil klasterisasi selanjutnya dikelompokkan kembali melalui

denormalisasi untuk mengembalikan nilai ke bentuk aslinya sebelum evaluasi dilakukan.

Evaluation

Evaluasi dilakukan dengan menghitung nilai Silhouette Coefficient dan Within-Cluster Sum of Squares (WCSS) menggunakan KNIME. Visualisasi grafik digunakan untuk menentukan jumlah klaster optimal, yaitu titik dengan nilai *silhouette* tertinggi atau titik siku pada grafik WCSS yang menunjukkan penurunan terbesar.

Deployment

Tahap akhir berupa visualisasi dan interpretasi hasil klaster. Visualisasi menggunakan Parallel Coordinates Plot agar dapat menggambarkan data multidimensi secara jelas. Selanjutnya, karakteristik setiap klaster diinterpretasikan untuk menunjukkan perbedaan antar kelompok. Informasi ini dapat menjadi dasar bagi pemangku kebijakan dalam menentukan sasaran program bantuan sosial secara lebih tepat.

HASIL DAN PEMBAHASAN

Untuk mencapai tujuan penelitian, digunakan metodologi berbasis kerangka kerja data mining yang meliputi tahapan business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Setiap tahapan dijelaskan secara sistematis guna memastikan proses analisis berjalan terarah dan hasil yang diperoleh dapat mendukung pengambilan keputusan secara tepat.

Business Understanding

Tahap business understanding bertujuan untuk merumuskan tujuan implementasi data mining melalui pengajuan data kepada penanggung jawab P3KE serta wawancara teknis mengenai pengelolaan dan pemanfaatan data. Hasil analisis menunjukkan belum tersedia model segmentasi penduduk miskin ekstrem sebagai dasar penentuan sasaran bantuan sosial, sehingga berpotensi menimbulkan ketidaktepatan sasaran dan duplikasi penerima bantuan.

Data Understanding

Tahap data understanding dilakukan untuk mengenali komposisi data mentah yang akan dianalisis. Data yang digunakan berupa 7.857 record keluarga miskin ekstrem dengan 27 atribut dalam format spreadsheet (.xls), yang selanjutnya menjadi dasar pemrosesan lebih lanjut.

Tabel 1. Sampel Data Mentah

Atribut	Dataset P3KE		
No	14978	14979	...
ID Keluarga P3KE	1324278 74	1327488 69	...
Kecamatan	BANTU R	BANTU R	...
Desa/Kelurahan	WONO KERTO	WONOK ERTO	...
Desil Kesejahteraan	1	1	...
Alamat	DUSUN WONO GIRI DESA WONO KERTO	DUSUN GAMPI NGAN DESA WONOK ERTO	...
Kepala Keluarga	ABDUL GUFRON	MUCHLIS	...
Padan Dukcapil	Ya	Ya	...
Jenis Kelamin	Laki-laki	Laki-laki	...
Tanggal Lahir	05/11/19 77 0:00:00	05/06/19 70 0:00:00	...
Pendidikan	Tamat SD/seder ajat	Tamat SD/seder ajat	...
Pekerjaan	Wiraswa sta	Wiraswa sta	...
Kepemilikan Rumah	Milik Sendiri	Milik Sendiri	...
Memiliki Simpanan Uang/Perhiasan/T ernak/Lainnya	Tidak	Ya	...
Jenis Atap	Genteng	Genteng	...
Jenis Dinding	Tembok	Tembok	...
Jenis Lantai	Semen	Semen	...
Sumber Penerangan	Listrik Pribadi s/d 900 Watt	Listrik Pribadi s/d 900 Watt	...
Bahan Bakar Memasak	Listrik/G as	Listrik/G as	...

Atribut	Dataset P3KE		
Sumber Air Minum	Sumur Terlindung	Ledeng/PAAM	...
Memiliki fasilitas Buang Air Besar	Ya, dengan Septic Tank	Tidak, Jamban Umum/Bersama	...
Penerima BPNT	Ya	Ya	...
Penerima BPUM	Tidak	Tidak	...
Penerima BST	Ya	Ya	...
Kepemilikan PKH	Ya	Ya	...
Penerima SEMBAKO	Ya	Ya	...
Jumlah Anggota Keluarga Beresiko Stunting	0	1	...

Data yang diperoleh memiliki memiliki 27 atribut pembentuk. Dilihat dari nilai dari masing-masing atribut, tipe data atribut sangat beragam mulai dari numerik, string, hingga kategorikal. Untuk memudahkan dalam pemrosesan klasterisasi data mentah diolah terlebih dahulu pada tahapan Data Preparation

Tabel 2. Penjelasan Atribut Data

No	Atribut	Tipe data	Keterangan
1	No	Number	Urutan data
2	ID Keluarga P3KE	String	Nomor identitas keluarga dalam Data P3KE
Atribut Demografi			
3	Kecamatan	Number	Nama kecamatan tempat tinggal
4	Desa/Kelurahan	String	Nama desa/kelurahan tempat tinggal
5	Desil Kesejahteraan	Number	Tingkat desil kemiskinan keluarga, disajikan dalam skala 1-3. semakin kecil angka semakin miskin status keluarga
6	Alamat	String	Nama alamat tempat tinggal keluarga
7	Kepala Keluarga	String	Nama kepala keluarga
8	Padan Dukcapil	Kategorikal	Status verifikasi

No	Atribut	Tipe data	Keterangan
			identitas kependudukan dalam database kependudukan, disajikan dalam pilihan Ya/Tidak. Bila opsi bernilai Tidak maka kepala keluarga belum terverifikasi dalam database kependudukan
9	Jenis Kelamin	Kategorikal	Jenis kelamin kepala keluarga
10	Tanggal Lahir	Date	Tanggal lahir kepala keluarga
Atribut Sosial Ekonomi			
11	Pendidikan	Kategorikal	Pendidikan terakhir kepala keluarga, menunjukkan jenjang pendidikan terakhir
12	Pekerjaan	Kategorikal	Jenis pekerjaan kepala keluarga
13	Kepemilikan Rumah	Kategorikal	Kepemilikan rumah oleh keluarga,
14	Memiliki Simpanan Uang/Perhiasan/Ternak/Lainnya	Kategorikal	Kepemilikan simpanan uang/perhiasan/ternak lainnya oleh kepala keluarga
15	Jenis Atap	Kategorikal	Jenis Atap rumah
16	Jenis Dinding	Kategorikal	Jenis Dinding rumah
17	Jenis Lantai	Kategorikal	Jenis Lantai rumah
18	Sumber Penerangan	Kategorikal	Sumber Penerangan rumah
19	Bahan Bakar Memasak	Kategorikal	Bahan Bakar Memasak di rumah
20	Sumber Air Minum	Kategorikal	Sumber Air Minum di rumah
21	Memiliki fasilitas Buang Air Besar	Kategorikal	Jenis fasilitas Buang Air Besar yang dimiliki di rumah
22	Penerima BPNT	Kategorikal	Status kepemilikan bantuan sosial Bantuan Pangan Non

No	Atribut	Tipe data	Keterangan
			Tunai (BPNT), disajikan dalam pilihan Ya/Tidak
23	Penerima BPUM	Kategorikal	Status kepemilikan bantuan sosial Bantuan Penerima Iuran Manfaat (BPUM), disajikan dalam pilihan Ya/Tidak
24	Penerima BST	Kategorikal	Status kepemilikan bantuan sosial Bantuan Sosial Tunai Kemensos (BST), disajikan dalam pilihan Ya/Tidak
25	Kepemilikan PKH	Kategorikal	Status kepemilikan bantuan sosial Program Keluarga Harapan (PKH) Kemensos, disajikan dalam pilihan Ya/Tidak
26	Penerima SEMBAKO	Kategorikal	Status kepemilikan bantuan sosial Sembako Kemensos, disajikan dalam pilihan Ya/Tidak
27	Jumlah Anggota Keluarga Beresiko Stunting	Number	Jumlah anggota keluarga beresiko stunting, disajikan dalam bentuk angka

Data Preparation

Tahap data preparation dilakukan untuk mengolah data mentah menjadi data siap analisis melalui proses *cleaning*, *selection*, *transformation*, dan *normalization* menggunakan node pada KNIME Analytics Platform. Proses *cleaning* memastikan tidak terdapat data kosong maupun kesalahan input, dan pada penelitian ini seluruh data dinyatakan valid tanpa nilai hilang.

Proses *selection* data dilakukan dengan memilih atribut yang relevan untuk kebutuhan analisis sesuai tujuan pada tahap business understanding, yaitu mengidentifikasi segmen masyarakat prioritas penerima bantuan sosial. Dari 27 atribut tersedia, dipilih empat atribut utama yang dianggap paling representatif, yakni jenis pekerjaan, kepemilikan rumah, kepemilikan PKH, serta jumlah anggota keluarga berisiko stunting.

Tahap *transformation* data bertujuan menyesuaikan dataset agar dapat diproses oleh algoritma K-Means yang hanya menerima input numerik. Dari atribut terpilih, beberapa variabel seperti jenis pekerjaan, kepemilikan rumah, kepemilikan PKH, dan fasilitas sanitasi merupakan data non-numerik sehingga perlu dikonversi ke dalam bentuk angka. Konversi dilakukan secara manual di Microsoft Excel menggunakan fungsi *find and replace*, dengan pemberian kode numerik berdasarkan frekuensi kemunculan nilai atribut.

Tabel 3. Transformasi nilai atribut

Jenis Pekerjaan	Frekuensi	Transformasi
Petani	3963	1
Wiraswasta	2527	2
Pekerja Lepas	625	3
Pedagang	277	4
Pegawai Swasta	237	5
Tidak/belum bekerja	183	6
Pensiunan	20	7
Nelayan	17	8
Lainnya	8	9

Proses *normalization* dilakukan untuk menyeragamkan skala nilai pada setiap atribut sehingga lebih mudah dibandingkan dan diolah. Normalisasi dilakukan menggunakan node Normalizer pada KNIME Analytics Platform, dengan data awal dibaca terlebih dahulu melalui node Excel Reader yang memproses file spreadsheet (.xls). Hasil proses normalisasi membuat nilai dari atribut pekerjaan, kepemilikan rumah, dan penerima PKH

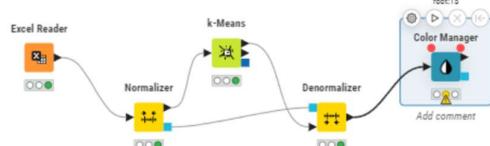
menjadi nilai dengan rentang 0 sampai dengan 1

RowID	Kecamatan	Desa/Kelurahan	Pekerjaan	Kependidikan Rtu.	Penerima PKH	Resiko Stunting
1	Rosd	BANTUR	WONOKERTO	0.125	0	0.5
2	Ros1	BANTUR	WONOKERTO	0.125	0	1
3	Ros2	BANTUR	WONOKERTO	0	0	0.5
4	Ros3	BANTUR	WONOKERTO	0	0	1
5	Ros4	BANTUR	WONOKERTO	0	0	0.5
6	Ros5	BANTUR	WONOKERTO	0	0	0.5
7	Ros6	BANTUR	WONOKERTO	0.375	0	1
8	Ros7	BANTUR	WONOKERTO	0.5	0	0.5
9	Ros8	BANTUR	WONOKERTO	0	0	1
10	Ros9	BANTUR	WONOKERTO	0	0	0
11	Ros.	BANTUR	WONOKERTO	0	0	0
12	Ros.	BANTUR	WONOKERTO	0	0	1
13	Ros.	BANTUR	WONOKERTO	0	0	0
14	Ros.	BANTUR	WONOKERTO	0	0	1
15	Ros.	BANTUR	WONOKERTO	0	0	0.5
16	Ros.	BANTUR	WONOKERTO	0	0	0
17	Ros.	BANTUR	WONOKERTO	0	0	1
18	Ros.	BANTUR	WONOKERTO	0.125	0	0
19	Ros.	BANTUR	WONOKERTO	0	0	1
20	Ros.	BANTUR	WONOKERTO	0	0	1

Gambar 3. Tampilan Data Hasil Normalisasi

Modelling

Tahap *modelling* dilakukan dengan menerapkan algoritma K-Means pada KNIME Analytics Platform. Proses dimulai dengan menghubungkan node Normalizer dan K-Means, kemudian dilakukan konfigurasi jumlah klaster awal sebanyak empat, penentuan titik centroid secara acak (*Random Initialization*), serta jumlah maksimum iterasi sebanyak sepuluh kali. Setelah dijalankan, node K-Means menghasilkan kolom baru berisi label klaster serta informasi centroid masing-masing kelompok. Selanjutnya, data hasil klasterisasi diproses dengan node Denormalizer untuk mengembalikan nilai ke bentuk aslinya, serta node Color Manager untuk memberikan visualisasi warna pada tiap klaster sehingga lebih mudah diinterpretasikan.

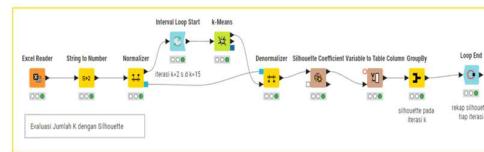


Gambar 4. Clustering menggunakan node KNIME Analytics Platform

Evaluation

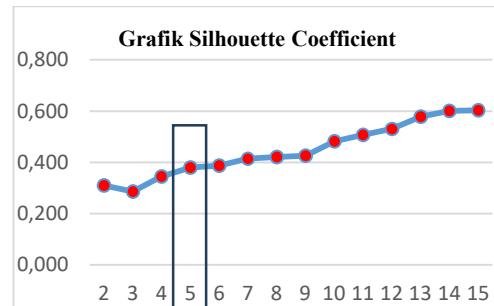
Pada tahap Evaluation dilakukan pengujian jumlah kluster yang optimal menggunakan *Silhouette Coefficient* dan WCSS. Rentang kluster yang diuji mulai dari 2 sampai dengan 15. Pengujian dilakukan menggunakan KNIME Analytics

Platform dengan cara membuat *workflow* baru dengan sumber dari data olah.



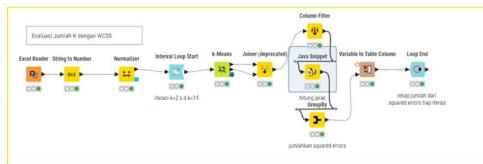
Gambar 5. Evaluasi Silhouette Coefficient menggunakan KNIME Analytics Platform

Untuk memudahkan analisis, hasil perhitungan divisualisasikan dalam bentuk *Line Chart* yang menunjukkan tren peningkatan nilai Silhouette Coefficient. Grafik memperlihatkan kenaikan signifikan pada iterasi ke-3 dengan jumlah klaster (k) = 5, sedangkan peningkatan berikutnya pada k = 10 dan k = 13 dianggap kurang relevan karena menghasilkan terlalu banyak klaster dan menyulitkan interpretasi. Oleh karena itu, berdasarkan evaluasi Silhouette Coefficient, jumlah klaster optimal ditetapkan pada k = 5.



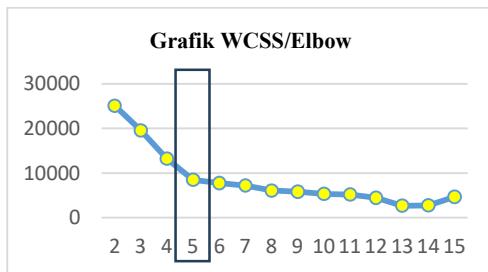
Gambar 6. Grafik Hasil Evaluasi menggunakan Silhouette Coefficient

Berdasarkan evaluasi dengan metode Silhouette Coefficient, hasil yang diperoleh kurang konsisten karena nilai silhouette terus mengalami kenaikan. Untuk itu, diperlukan evaluasi tambahan menggunakan metode WCSS dengan menghitung jarak antar titik centroid dan memvisualisasikannya dalam bentuk *Line Chart*. Penentuan jumlah klaster optimal dilakukan dengan pendekatan *Elbow Method* melalui identifikasi titik siku pada grafik yang dihasilkan di KNIME Analytics Platform.



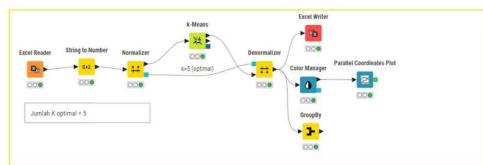
Gambar 7 Evaluasi WCSS menggunakan KNIME Analytics Platform

Hasil evaluasi disajikan dalam bentuk grafik Line Chart. Grafik memperlihatkan penurunan tajam pada $k=2$ hingga $k=5$, kemudian melambat hingga $k=15$. Titik siku terlihat pada $k=5$, sehingga jumlah klaster optimal ditetapkan sebanyak 5 klaster.



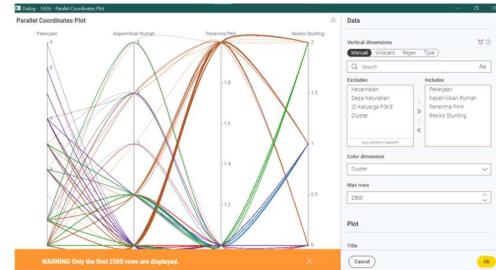
Gambar 8. Grafik Hasil Evaluasi menggunakan WCSS

Berdasarkan hasil evaluasi menggunakan Silhouette Coefficient dan Metode WCSS ditemukan jumlah k yang optimal adalah $k=5$. Hasil evaluasi diterapkan kembali kedalam K-Means untuk membuat klasterisasi yang optimal



Gambar 9. Clustering dengan menggunakan jumlah k yang optimal

Visualisasi dilakukan menggunakan node Parallel Coordinates Plot, di mana atribut direpresentasikan sebagai sumbu koordinat dan garis menunjukkan hubungan antar atribut. Grafik ini dipilih karena sesuai untuk dataset multivariat dengan variabel yang beragam.

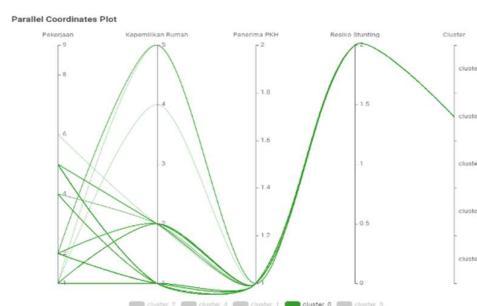


Gambar 10. Visualisasi data menggunakan Grafik Parallel Coordinates Plot

Deployment

Deployment adalah tahapan terakhir dari serangkaian proses data mining untuk menemukan pengetahuan dari dataset yang telah diolah. Dalam penelitian ini, proses penyajian informasi menggunakan visualisasi grafik Parallel Coordinates Plot dan tabel perbandingan antar atribut. Interpretasi Grafik Parallel Coordinates Plot berdasarkan warna garis yang menghubungkan masing-masing atribut data.

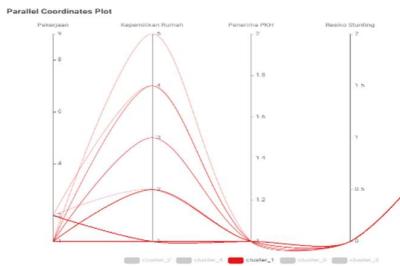
Cluster 0 (Hijau) didominasi kepala keluarga dengan pekerjaan rendah hingga menengah, sebagian tidak bekerja, mayoritas memiliki rumah sendiri atau menumpang, serta tidak menerima PKH. Kelompok ini merepresentasikan masyarakat berpenghasilan rendah-menengah dengan risiko stunting yang sangat tinggi..



Gambar 11. Visualisasi Cluster 0 dengan Parallel Coordinates Plot

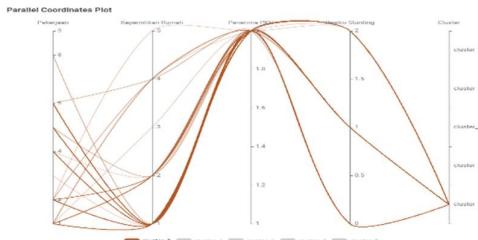
Cluster 1 (Merah) didominasi oleh pekerjaan sebagai Petani, Wiraswasta, dan Pekerja Lepas, dengan mayoritas memiliki rumah sendiri atau menumpang serta tidak menerima PKH. Klaster ini merupakan

kelompok paling heterogen dalam pekerjaan dengan risiko stunting yang sangat rendah



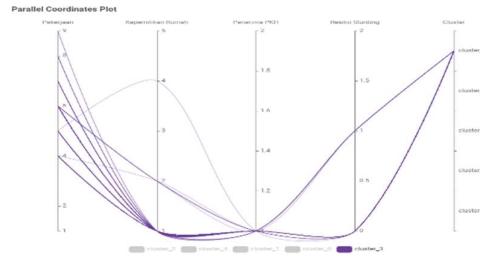
Gambar 12. Visualisasi Cluster 1 dengan Parallel Coordinates Plot

Cluster 2 (Coklat) memiliki pekerjaan yang beragam dari tingkat rendah hingga atas, dengan proporsi besar kepala keluarga yang tidak bekerja, mayoritas memiliki rumah sendiri, serta banyak yang menerima PKH. Klaster ini menunjukkan variasi risiko stunting dari rendah hingga tinggi, dengan sebagian anggota keluarga berisiko stunting tinggi.



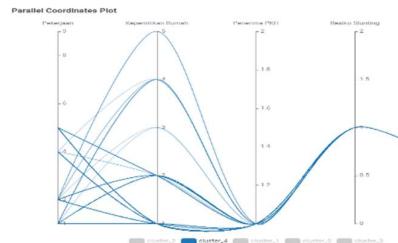
Gambar 13. Visualisasi Cluster 2 dengan Parallel Coordinates Plot

Cluster 3 (Ungu) didominasi oleh pekerjaan sebagai Pedagang, Pegawai Swasta, Pensiunan, dan Nelayan, dengan sebagian kepala keluarga tidak bekerja, mayoritas sudah memiliki rumah sendiri, serta bukan penerima PKH. Klaster ini merepresentasikan kelompok berpenghasilan rendah dengan risiko stunting yang bervariasi, terutama rendah hingga sedang.



Gambar 14. Visualisasi Cluster 3 dengan Parallel Coordinates Plot

Cluster 4 (Biru) didominasi oleh kepala keluarga yang bekerja sebagai Petani, Wiraswasta, Pedagang, dan Pegawai Swasta tanpa ada yang tidak bekerja, mayoritas memiliki rumah sendiri atau menumpang, serta tidak menerima PKH. Klaster ini merepresentasikan kelompok berpenghasilan rendah dengan risiko stunting moderat.



Gambar 15. Visualisasi Cluster 4 dengan Parallel Coordinates Plot

Pada tabel perbandingan atribut, setiap atribut disajikan dalam kolom beserta nilai masing-masing, kemudian ditampilkan akumulasi jumlah kepala keluarga per atribut sesuai klaster yang terbentuk.

Tabel 4. Hasil Clustering menggunakan K-Means

Klaster	Kepemilikan PKH	Jenis Pekerjaan	Kepemilikan Rumah	Jumlah Anggota Keluarga Berisiko Stunting
cluster_0	Tidak memiliki PKH: 1353	Lainnya: 1 Nelayan: 1 Pedagang: 57 Pegawai Swasta: 62 Pekerja Lepas: 90 Pensiunan: 1 Petani: 398 Tidak/belum bekerja: 13 Wiraswasta: 730	Bebas Sewa: 2 Kontrak/Sewa: 8 Lainnya: 9 Menumpang: 130 Milik Sendiri: 1204	> 1 anggota :1353
cluster_1	Tidak memiliki	Pekerja Lepas: 110 Petani: 1808	Bebas Sewa: 5	-

Klaster	Kepemilikan PKH	Jenis Pekerjaan	Kepemilikan Rumah	Jumlah Anggota Keluarga Berisiko Stunting
	PKH: 2343	Wiraswasta: 425	Kontrak/Sewa : 6 Lainnya: 8 Menumpang: 55 Milik Sendiri: 2269	
cluster_2	Tidak memiliki PKH: -	Lainnya: 1 Nelayan: 2 Pedagang: 30 Pegawai Swasta: 30 Pekerja Lepas: 160 Pensiunan: 1 Petani: 849 Tidak/belum bekerja: 46 Wiraswasta: 426	Bebas Sewa: 8 Kontrak/Sewa : 2 Lainnya: 2 Menumpang: 69 Milik Sendiri: 1464	> 1 anggota: 246 1 anggota: 698
cluster_3	Tidak memiliki PKH: 284	Lainnya: 6 Nelayan: 14 Pedagang: 75 Pegawai Swasta: 47 Pensiunan: 18 Tidak/belum bekerja: 124	Bebas Sewa: 1 Lainnya: 3 Menumpang: 18 Milik Sendiri: 262	1 anggota: 33
cluster_4	Tidak memiliki PKH: 2332	Pedagang: 115 Pegawai Swasta: 98 Pekerja Lepas: 265 Petani: 908 Wiraswasta: 946	Bebas Sewa: 11 Kontrak/Sewa : 5 Lainnya: 11 Menumpang: 247 Milik Sendiri: 2058	1 anggota: 2332

Analisis klaster terhadap lima kelompok (Cluster 0 sampai 4) menunjukkan mayoritas kepala keluarga tidak menerima PKH, dengan pekerjaan dominan sebagai petani, wiraswasta, dan pegawai swasta, serta kepemilikan rumah mayoritas "Milik Sendiri". Pola ini mengindikasikan kepemilikan aset yang relatif stabil, namun masih bergantung pada sektor informal dan pertanian yang rentan terhadap fluktuasi. Cluster 1 dan 4 menonjol dengan jumlah terbesar kepala keluarga tanpa PKH, sementara Cluster 0 memiliki kasus stunting tertinggi (1353 anggota keluarga). Sebaliknya, Cluster 1 tidak mencatat risiko stunting, yang dapat menunjukkan keberhasilan intervensi kesehatan atau keterbatasan data. Temuan ini menekankan perlunya kajian lebih lanjut terkait faktor penyebab stunting di Cluster 0 dan 2, serta potensi replikasi praktik baik dari Cluster 1 untuk menekan risiko stunting di klaster lain.

Hasil analisis lima klaster menunjukkan variasi tingkat prioritas intervensi:

1. Cluster 4 ditetapkan sebagai *sangat prioritas* karena memiliki jumlah

terbesar kepala keluarga tanpa PKH (2332 KK) sekaligus kasus stunting yang juga sangat tinggi (2332 anggota keluarga), mencerminkan masalah kemiskinan dan kesehatan yang sistemik.

2. Cluster 0 masuk kategori *prioritas tinggi* dengan kasus stunting signifikan (1353 anggota keluarga) meskipun angka non-PKH lebih rendah dari Cluster 4.
3. Cluster 2 dikategorikan *prioritas sedang* karena masih terdapat masalah gizi cukup besar (246 KK dan 698 KK berisiko stunting) dengan kerentanan bervariasi akibat jenis pekerjaan yang beragam.
4. Cluster 3 berada pada *prioritas rendah* karena kasus stunting relatif kecil (33 KK), meskipun terdapat kelompok nelayan yang rentan secara ekonomi sehingga perlu dukungan diversifikasi mata pencarian.
5. Cluster 1 tidak menjadi prioritas karena meski memiliki banyak KK non-PKH (2343 KK), tidak ditemukan kasus stunting, sehingga intervensi lebih diarahkan pada penguatan ketahanan ekonomi petani.

Dari hasil pengujian dan pembahasan segmentasi penduduk miskin ekstrem yang belum terakses program perlindungan sosial melalui *K-Means clustering* menunjukkan bahwa segmentasi penduduk miskin ekstrem menggunakan K-Means Clustering pada 7857 kepala keluarga (KK) menghasilkan 5 (lima) segmen yaitu:

1. Sangat Prioritas (Cluster 4) beranggotakan 2332 KK yaitu 29,7% dari keseluruhan jumlah penduduk miskin ekstrem
2. Prioritas Tinggi (Cluster 0) beranggotakan 1353 KK yaitu 17,2% dari keseluruhan jumlah penduduk miskin ekstrem.
3. Prioritas Sedang (Cluster 2) beranggotakan 1545 KK yaitu 19,7% dari keseluruhan jumlah penduduk miskin ekstrem
4. Kurang Prioritas (Cluster 3) beranggotakan 284 KK yaitu 3,6% dari

- keseluruhan jumlah penduduk miskin ekstrem
5. Bukan Prioritas (Cluster 1) beranggotakan 2343 KK yaitu 29,8% dari keseluruhan jumlah penduduk miskin ekstrem
- Adapun hasil pengujian $k=2$ sampai dengan $k=15$ menggunakan metode *Silhouette Coefficient* dan WCSS/Elbow menghasilkan jumlah klaster optimal pada $k=5$ yang diterjemahkan ke dalam 5 segmen prioritas penduduk miskin ekstrem.

UCAPAN TERIMA KASIH

Akan ditambahkan setelah lolos proses *review*.

PENUTUP

Penelitian ini mengidentifikasi segmentasi 7.857 KK miskin ekstrem yang belum terakses program perlindungan sosial menggunakan metode K-Means Clustering. Hasil analisis menunjukkan bahwa jumlah cluster optimal adalah lima segmen ($k = 5$), yaitu Sangat Prioritas (2.332 KK), Prioritas Tinggi (1.353 KK), Prioritas Sedang (1.545 KK), Kurang Prioritas (284 KK), dan Bukan Prioritas (2.343 KK). Temuan ini dapat dimanfaatkan sebagai dasar perumusan kebijakan dalam penentuan prioritas penerima program perlindungan sosial agar lebih tepat sasaran.

Namun, penelitian ini masih memiliki keterbatasan, khususnya pada aspek pemilihan atribut. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan *feature selection* dan *feature engineering* guna mengidentifikasi variabel yang paling berpengaruh serta meningkatkan kualitas segmentasi yang dihasilkan

DAFTAR PUSTAKA

- [1] I. R. Adi, Ilmu Kesejahteraan Sosial: Pengantar dan Kajian, Jakarta: Rajawali Pers, 2015.
- [2] A. Fahrudin, Pengantar Kesejahteraan Sosial, Bandung: Refika Aditama, 2014.
- [3] Badan Pusat Statistik Kabupaten Malang, "Kecamatan Bantur Dalam Angka 2023," BPS Kabupaten Malang, Malang, 2023.
- [4] Badan Pusat Statistik, Profil Kemiskinan di Indonesia, Jakarta: BPS, 2021.
- [5] Peraturan Daerah Nomor 3 Tahun 2021 tentang Rencana Pembangunan Jangka Menengah Daerah Kabupaten Malang Tahun 2021–2025., Malang, 2021.
- [6] Badan Perencanaan Pembangunan Nasional, "Strategi Nasional Percepatan Penghapusan Kemiskinan Ekstrem Tahun 2021," Bappenas, Jakarta, 2021.
- [7] A. V. Rivallinata, "Segmentasi Penduduk Miskin Di Indonesia Menggunakan Algoritma K-Means," STMIK PPKIA Pradnya Paramita, Malang, 2023.
- [8] S. M. Febriansyah Febriansyah, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin Pada Kota Pagar Alam," *JISKA (Jurnal Informatika Sunan Kalijaga)*, pp. 66-77, 2023.
- [9] Badan Pusat Statistik, "Pengeluaran untuk Konsumsi Penduduk Indonesia," Badan Pusat Statistik, Malang, 2021.
- [10] Keputusan Menteri Koordinator Pembangunan Manusia dan Kebudayaan Nomor 32 tahun 2022 Tentang Pedoman Umum Pelaksanaan Program Percepatan Penghapusan Kemiskinan Ekstrem, 2022.
- [11] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, p. 2, 2016.
- [12] R. P. Utomo and M. , Data Mining: Konsep dan Implementasi, Medan: Universitas Potensi Utama, 2020.

- [13] U. R. H. R. d. M. A. A. Damuri, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *Jurnal Riset Komputer*, p. 6, 2021.
- [14] S. S. N. A. E. d. D. M. M. Fatkhudin, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa," *Jurnal Minfo Polgan*, pp. 12(2), 297-302, 2022.
- [15] P. Chapman, CRISP-DM 1.0 Step-by-Step Data Mining Guide, The CRISP-DM Consortium, 2000.
- [16] Y. Agusta, "K Means – Penerapan, Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, pp. 47-60, 2007.
- [17] M. Mahanani, D. S. Utsalina and T. M. Akhriza, "Rekomendasi Solusi Atas Kasus Perceraian Di Kabupaten Lamongan Berbasis K-Means Clustering," *Prosiding Seminar SeNTIK*, vol. 8, no. 1, pp. 489-496, 2024.
- [18] N. K. B. A. d. S. J. Abdulkadir, "KNIME-Based Clustering Technique on Twitter Trends Detection," *International Symposium on Innovative Engineering*, p. 4, 2018.
- [19] E. Nensy, "Segmentasi Tingkat Kemiskinan Pada Data Penduduk Miskin Di Indonesia Selama Pandemi Menggunakan Metode Algoritma K-Medoids," *STMIK PPKIA Pradnya Paramita*, 2022.