

DATA MINING CLUSTERING FOOD EXPENDITURE IN INDONESIA

Indri Tri Julianto^{*1}, Dede Kurniadi², Muhammad Rikza Nashrulloh³, Asri Mulyani⁴

^{1,2,3,4} Jurusan Ilmu Komputer, Institut Teknologi Garut, Indonesia

Email: ¹indritrijulianto@itg.ac.id, ²dede.kurniadi@itg.ac.id, ³rikza@itg.ac.id, ⁴asrimulyani@itg.ac.id

(Naskah masuk: 18 Mei 2022, Revisi: 8 Juni 2022, diterbitkan: 26 Desember 2022)

Abstract

The availability of food in a country is determined by a conducive climate. Prolonged droughts, floods, and natural disasters, especially for food crop production areas, will have an impact on the availability of natural disaster conditions faced by all countries including Indonesia is the Covid-19 pandemic, where this will affect food security in Indonesia. Data mining is the process of discovering the hidden meaning of a very large data set. The technique used in this study is Data Mining Clustering and the validity index used is Davies-Bouldin. This study aims to determine the Food Security Strategy in Indonesia through the Data Mining Clustering process based on food expenditure data and the Indonesian people's food expenditure per capita. The methodology used is Cross Industry Standard for Data Mining using the K-Means and K-Medoids Algorithm. The best cluster for the K-Means Algorithm is $K=7$ with a value of 0.341 and for the K-Medoids Algorithm, it is $K=7$ with a value of 0.362. This research produces the best algorithm, namely K-Means with a value of 0.341, which has a smaller value than K-Medoids with a value of 0.362. The results showed that the regional cluster with the highest average expenditure on food was cluster 5 covering the DKI Jakarta area, while the cluster with the lowest expenditure was cluster 6 covering Central Java, East Nusa Tenggara, Southeast Sulawesi, Gorontalo, and West Sulawesi. In cluster 6, it is necessary to implement a strategy to increase food security by increasing production capacity and food reserves in each region.

Keywords: clustering, data mining, food, k-means, k-medoids.

DATA MINING KLASTERISASI PENGELUARAN PANGAN DI INDONESIA

Abstrak

Ketersediaan pangan di suatu negara ditentukan oleh iklim yang kondusif. Kekeringan yang berkepanjangan, banjir, dan bencana alam khususnya untuk daerah produksi tanaman pangan akan berdampak pada jumlah ketersediaan. Kondisi bencana alam yang dihadapi oleh semua negara termasuk Indonesia yaitu pandemi Covid-19, dimana hal ini akan mempengaruhi ketahanan pangan di Indonesia. Data mining adalah proses menemukan makna tersembunyi dari kumpulan data yang sangat besar. Teknik yang digunakan dalam penelitian ini adalah Data Mining Klasterisasi dan indeks validitas yang digunakan adalah Davies-Bouldin. Penelitian ini bertujuan untuk mengetahui Strategi Ketahanan Pangan di Indonesia melalui proses Data Mining Clustering berdasarkan data pengeluaran pangan dan pengeluaran pangan per kapita masyarakat Indonesia. Metodologi yang digunakan adalah Cross Industry Standard for Data Mining dengan menggunakan Algoritma K-Means dan K-Medoids. Klaster terbaik untuk Algoritma K-Means adalah $K=7$ dengan nilai 0.341 dan untuk Algoritma K-Medoids adalah $K=7$ dengan nilai 0.362. Penelitian ini menghasilkan algoritma terbaik yaitu K-Means dengan nilai 0.341 yang memiliki nilai lebih kecil dari K-Medoids dengan nilai 0.362. Hasil penelitian menunjukkan bahwa klaster daerah dengan rata-rata pengeluaran pangan tertinggi adalah klaster 5 yang meliputi wilayah DKI Jakarta, sedangkan klaster dengan pengeluaran terendah adalah klaster 6 yang meliputi Jawa Tengah, Nusa Tenggara Timur, Sulawesi Tenggara, Gorontalo, dan Sulawesi Barat. Pada klaster 6 perlu diterapkan strategi peningkatan ketahanan pangan dengan meningkatkan kapasitas produksi dan cadangan pangan di masing-masing wilayah.

Kata kunci: data mining, klasterisasi, k-means, k-medoids, pangan.

1. PENDAHULUAN

Ketersediaan pangan di suatu negara itu ditentukan oleh iklim yang kondusif. Kemarau berkepanjangan, banjir serta bencana alam,

khususnya bagi wilayah-wilayah produksi tanaman pangan, akan berdampak terhadap ketersediaan pangan [1]. Kondisi bencana alam yang dihadapi seluruh negara termasuk Indonesia adalah pandemi covid-19, dimana hal ini akan mempengaruhi

terhadap ketahanan pangan di Indonesia. Persentase yang lebih tinggi dari pengeluaran konsumsi pangan masyarakat menunjukkan semakin rendah ketahanan pangan suatu daerah dan sebaliknya semakin rendah persentase pengeluaran konsumsi pangan maka ketahanan pangan suatu daerah semakin tinggi dan menunjukkan peningkatan kesejahteraan [2].

Penelitian ini bertujuan untuk mengelompokkan pengeluaran pangan masyarakat menurut provinsi di Indonesia. Data masukan yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), dimana data tersebut merupakan data pengeluaran per kapita perkotaan dan perdesaan menurut provinsi dan kelompok barang (rupiah). Pengolahan data ini memerlukan pengolahan *Data Mining*, karena sumber data yang tersedia cukup besar dan melalui teknik ini akan ditemukan pengetahuan baru dari data yang terkumpul. Luaran dari penelitian ini adalah pengelompokan wilayah dengan ketahanan pangan tinggi dan rendah. Kemudian hasil pengolahan Data Mining ini akan dijadikan acuan untuk menentukan strategi ketahanan penguatan pangan di daerah yang termasuk kategori rendah. *Data Mining* merupakan proses untuk menemukan keterkaitan baru, pola dan juga tren dengan istilah menambang sejumlah repository data dalam jumlah besar dengan menggunakan bidang ilmu teknik matematika serta statistika, kemudian data mining juga sering disebut *Knowledge Discovery in Database (KDD)* atau *Pattern Recognition* [3]–[5], dimana Data Mining ini dapat digolongkan menjadi empat kelompok, yang pertama adalah model prediksi (*prediction modelling*), kedua adalah analisis klasterisasi (*Cluster analysis*), analisis asosiasi (*association analysis*) serta deteksi anomali (*anomaly detection*) [6].

Data mining memiliki banyak algoritma yang dapat digunakan, akan tetapi terdapat 10 algoritma *Data Mining* yang menempati peringkat teratas, yaitu *C4.5*, *K-Means*, *Support Vector Machines*, *Apriori*, *EM*, *PageRank*, *Adaboost*, *k-Nearest Neighbors*, *Naive Bayes*, *Classification and Regression Trees* [7]. Data masukan yang digunakan pada proses *Data Mining* disajikan dalam bentuk gambar, seperti terlihat pada Gambar 1 dan Gambar 2.

Provinsi	Mekans						
	2013	2014	2015	2016	2017	2018	2019
Aceh	328.237	358.131	371.878	408.831	423.034	461.951	529.181
Sumatera Utara	338.342	344.467	369.383	398.932	424.566	472.220	522.706
Sumatera Barat	338.338	360.870	423.851	461.458	472.428	527.239	566.945
Riau	385.949	402.511	408.503	461.965	495.122	536.747	603.401
Jambi	324.297	339.737	377.133	403.639	428.427	464.381	538.260
Sulawesi Selatan	300.451	338.927	333.113	390.857	379.945	469.246	527.671
Bengkulu	294.975	330.123	346.161	384.146	413.263	447.384	523.129
Lampung	261.519	283.870	314.408	340.844	382.368	412.527	467.940
Kalimantan Tengah	362.656	418.488	462.111	511.949	578.811	599.361	721.396
Kalimantan Barat	451.248	470.371	508.569	574.814	602.071	661.525	759.544
DIY	467.668	519.028	605.269	623.386	625.486	662.207	797.828
Jawa Barat	287.960	340.757	371.861	389.219	423.880	457.619	562.767
Jawa Tengah	229.775	281.761	281.821	330.872	330.648	371.605	460.881
DI Yogyakarta	278.123	327.241	353.778	393.550	389.512	434.094	490.249
Bali	242.741	251.061	286.942	339.179	360.990	407.281	478.861
Banten	328.623	370.088	420.422	465.522	480.956	532.389	612.918
Bali	328.490	330.960	449.048	471.149	421.577	465.993	569.188
Nusa Tenggara Barat	246.890	281.278	324.656	368.485	395.094	453.122	467.628
Nusa Tenggara Timur	222.575	228.660	240.207	287.584	298.180	312.312	399.251
Kalimantan Barat	312.711	381.590	397.018	420.130	428.181	481.222	539.468

Gambar 1. Data Pengeluaran Pangan Masyarakat Indonesia

Kelompok Barang	Mekans						
	2013	2014	2015	2016	2017	2018	2019
Mekans	1370267	189944	70845	46095	34678	49274	49274
Profil dan	189.805	87.702	19.359	4.031	643	0	0
Umum-umum	10.019	287	33	0	0	0	0
Makan	132.828	2.738	213	152	215	0	0
Minum	36.435	2.332	237	118	207	0	0
Tarif dan jasa	54.517	4.845	767	1.030	217	0	0
Tarif-kecukupan	168.603	371	102	136	0	0	0
Kelembagaan	18.965	0	0	0	0	0	0
Konsumsi	53.863	330	216	0	61	0	0
Kelembagaan	64.849	14	0	0	0	0	0
Bahan makanan	46.538	0	42	0	0	0	0
Bumbu-bumbu	16.519	0	0	0	0	0	0
Kelembagaan	14.463	0	0	0	0	0	0
Makanan dan minuman jadi	234.863	16.846	9.283	9.305	13.204	32.707	32.707
Kelembagaan	113.633	76.520	40.204	30.214	17.879	16.568	16.568
Bahan Makanan	344.860	228.603	142.629	128.593	122.280	247.435	247.435
Pencukupan, bahan baku, penerangan, air	137.887	100.879	62.466	52.869	44.070	50.253	50.253
Kelembagaan	54.411	89.849	51.861	49.835	38.438	130.242	130.242
Bahan perikanan	52.184	3.941	3.464	2.068	4.068	342	342
Bahan kesehatan	22.630	4.355	4.074	5.332	18.729	38.961	38.961
Pakaian, jasa kesehatan, obat-obatan kesehatan	46.706	13.517	10.976	6.068	4.860	7.847	7.847
Barang yang bukan lain	30.725	8.175	5.701	8.311	8.704	49.860	49.860

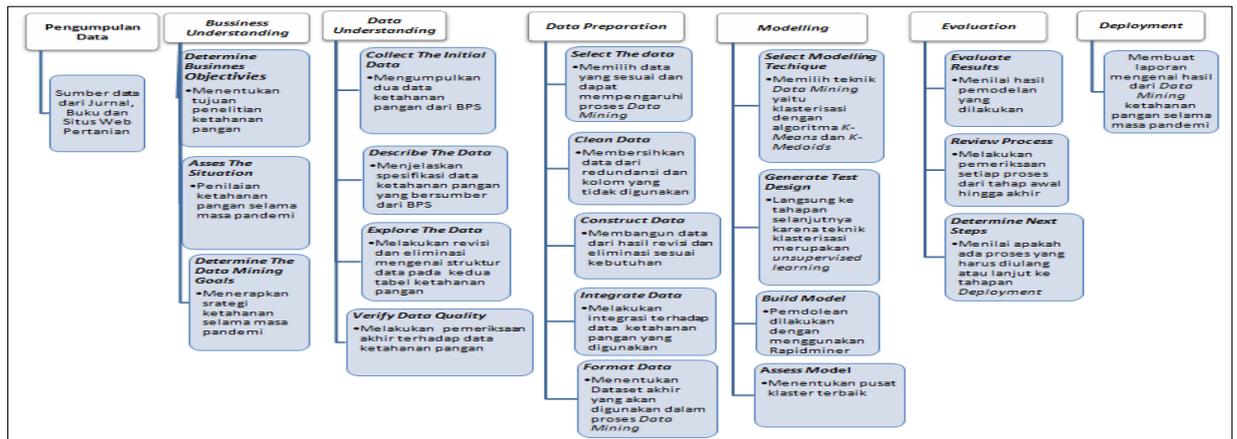
Gambar 2. Data Jenis Pengeluaran Pangan

Kedua data tersebut akan dijadikan masukan dalam proses *Data Mining* ketahanan pangan Indonesia. Data tersebut bersumber dari Badan Pusat Statistik Indonesia dari rentang tahun 2013-2019. Rentang waktu tersebut dipilih karena datanya relatif konsisten dibandingkan data-data sebelum 2013 dan data semenjak pandemi covid-19.

Terdapat beberapa penelitian yang telah membahas mengenai *Data Mining*, dimana penelitiannya menggunakan studi kasus Covid-19 serta algoritma yang digunakan merupakan algoritma klasterisasi. Secara keseluruhan ada yang menggunakan algoritma K-Medoid seperti penelitian [2], [8], [9], kemudian ada juga yang menggunakan algoritma K-Means seperti penelitian [10], serta ada yang menggunakan dua algoritma sekaligus yaitu algoritma *K-Means* dan *Hierarchical Clustering* seperti penelitian [11]. Berdasarkan penelitian-penelitian tersebut, diputuskan untuk menggunakan algoritma klasterisasi *K-Means* dan *K-Medoids* untuk mengetahui daerah pengeluaran pangan di Indonesia.

2. METODE PENELITIAN

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi yang digunakan dalam penelitian ini. Adapun tahapan-tahapannya disajikan dalam bentuk gambar sebagaimana tampak pada Gambar 3.



Gambar 3. Metodologi Penelitian Berdasarkan CRISP-DM

2. 1. Pengumpulan Data

Data yang digunakan merupakan hasil dari Studi Pustaka yang dilakukan. Adapun hasil yang didapatkan dari studi pustaka adalah dalam bentuk jurnal, buku, artikel dari suatu situs web serta data yang di rujuk dari Badan Pusat Statistik (BPS).

2. 2. Business Understanding

Pemahaman penelitian mengacu pada ketahanan pangan Indonesia selama menghadapi masa pandemi covid-19. Pada tahap ini diperlukan pemahaman tentang latar belakang, tujuan dan substansi pada penelitian ini. Adapun turunan dari tahapan ini adalah sebagai berikut:

2. 2. 1. Determine Business Objectives

Tujuan dari penelitian ini adalah melakukan klasterisasi dari data yang dirilis oleh BPS, sehingga nantinya akan menjadi rekomendasi bagi penerapan strategi yang cocok bagi klaster-klaster wilayah yang terbentuk;

2. 2. 2. Asses The Situation

Situasi selama masa pandemi ini sangat mempengaruhi ketahanan pangan serta daya beli masyarakat Indonesia, sehingga data yang digunakan adalah dari rentang 2013-2019 dikarenakan data yang sudah relatif konsisten serta dapat dijadikan acuan bagi strategi ketahanan pangan kedepannya;

2. 2. 3. Determine The Data Mining Goals

Tujuan dilakukannya proses Data Mining ini adalah untuk menggali pengetahuan baru dengan teknik *clustering* guna menunjang strategi ketahanan pangan di Indonesia.

2. 3. Data Understanding

Pemahaman data mengacu pada data BPS. Dilakukan pemahaman mengenai format data secara permukaan dan secara lebih mendalam mengenai

bentuk fisik data. Adapun tahapan yang dikerjakan dalam Data Understanding ini adalah sebagai berikut:

2. 3. 1. Collect The Initial Data

Data yang dihimpun merupakan data dari BPS mengenai rata-rata per kapita pengeluaran masyarakat untuk kebutuhan pangan per bulan berdasarkan daerah dalam rentang waktu 2013-2019 beserta data jenis makanannya;

2. 3. 2. Describe The Data

Data yang telah diunduh dari situs web BPS kemudian dibuka melalui Microsoft Excel, dimana untuk data tabel 1 didalamnya terdapat kolom untuk provinsi, kemudian kolom kebutuhan makanan berdasarkan tahun, kolom kebutuhan non makanan berdasarkan tahun dan kolom jumlah kebutuhan pangan dan non pangan berdasarkan tahun. Untuk tabel kedua terdapat kolom kelompok barang, kemudian rentang biaya yang dikeluarkan masyarakat dalam satuan rupiah dan kolom rata-rata pengeluaran per kapita;

2. 3. 3. Explore The Data

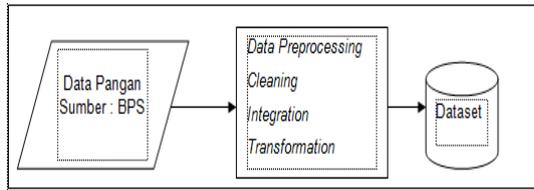
Berdasarkan temuan pada tahap *Describe The Data*, maka perlu dilakukan revisi atau eliminasi terhadap struktur dari data tersebut. Revisi yang dilakukan adalah pada kolom jumlah, seharusnya tidak perlu disertakan akumulasi dari kolom non makanan. Eliminasi dilakukan pada kolom kebutuhan non makanan yang tidak akan berpengaruh terhadap proses Data Mining yang akan dilakukan;

2. 3. 4. Verify Data Quality

Tidak ada penomoran di kedua tabel tersebut, misal untuk tabel pertama seharusnya ada kolom untuk nomor sehingga memudahkan dalam menemukan urutan dari suatu provinsi.

2. 4. Data Preparation

Adapun ilustrasi mengenai tahapan ini disajikan dalam bentuk gambar sebagaimana tampak pada Gambar 4.



Gambar 4. Ilustrasi *Data Preparation* [12]

Tugas-tugas yang dilakukan pada data preparation ini adalah sebagai berikut:

2. 4. 1. Select Data

Data yang telah diperoleh kemudian diseleksi, dimana hanya kolom dan baris yang berpengaruh terhadap proses *Data Mining* yang akan digunakan. Adapun atribut-atribut yang ada dalam data disajikan dalam bentuk tabel sebagaimana tampak pada Tabel 1.

Tabel 1. Atribut Yang Digunakan

No	Atribut	Tipe Atribut	Keterangan
1	Provinsi	Polynimial	Menampilkan nama daerah
2	Makanan	Polynominal	Menampilkan jenis makanan
3	Non-Makanan	Polynominal	Menampilkan jenis non-makanan
4	Tahun	Integer	Menampilkan pengeluaran berdasarkan tahun (2013-2019)
5	Jumlah	Integer	Menampilkan akumulasi perhitungan
6	Golongan pengeluaran per kapita sebulan (rupiah)	Integer	Menampilkan pengeluaran mulai kurang dari 10.000 s.d lebih dari 1.000.000
7	Rata-rata Per Kapita	Integer	Menampilkan rata-rata pengeluaran per kapita

Tidak semua atribut yang terdapat pada Tabel 1. akan digunakan, hanya atribut-atribut yang berpengaruh dalam proses Data Mining yang akan digunakan. Adapun atribut-atribut yang digunakan adalah sebagai berikut:

- a. Provinsi;
- b. Makanan;
- c. Tahun;
- d. Rata-rata per kapita.

Atribut-atribut tersebut dipilih berdasarkan pertimbangan pada tujuan awal penelitian ini yaitu klasterisasi ketahanan pangan Indonesia dimana kata kunci utama dalam pemilihan atribut ini adalah pangan atau makanan, sehingga golongan non makanan tidak dimasukkan dalam proses *Data Mining* ini.

2. 4. 2. Clean Data

Data akan dibersihkan dari data yang tidak layak seperti halnya data kosong, seperti halnya yang terdapat pada tabel rata-rata pengeluaran makanan per kapita sebagaimana tampak pada Tabel 2. serta pada tabel jenis makanan sebagaimana tampak pada Tabel 3. Data kosong tersebut akan

dihilangkan untuk mendapatkan data yang relevan [13].

Tabel 2. Pembersihan Data Kosong Pada Data Pengeluaran Pangan

Provinsi		2013	2014
Nusa Tenggara Barat		316 656	366 885
Nusa Tenggara Timur		240 207	267 584
Kalimantan Barat		367 018	420 130
Kalimantan Tengah		418 274	495 159
Kalimantan Selatan		440 803	456 699
Kalimantan Timur		477 325	516 036
Kalimantan Utara		-	-
Sulawesi Utara		379 814	395 996
Sulawesi Tengah		320 823	360 961
Sulawesi Selatan		302 903	330 220
Sulawesi Tenggara		284 683	291 923
Gorontalo		276 334	338 353
Sulawesi Barat		278 355	280 751

Pada awalnya data pada Tabel 2. terdapat nilai kosong yaitu pada Provinsi Kalimantan Utara. Sehingga untuk mendapatkan nilai yang relevan maka *records* Kalimantan Utara itu diambil nilai rata-rata.

Tabel 3. Pembersihan Data Kosong Pada Data Jenis Pangan

Nama Makanan	2013 Kurang dari 100.00
Padi-padian	29 722
Umbi-umbian	3 302
Ikan	5 300
Daging	-
Telur dan susu	263
Sayur-sayuran	6 430
Kacang-kacangan	1 303
Buah-buahan	1 931
Minyak dan lemak	2 505
Bahan minuman	1 527
Bumbu-bumbuan	1 466
Konsumsi lainnya	132
Makanan dan minuman jadi	2 839
Tembakau dan sirih	4 203

Pada Tabel 3. terdapat kekosongan pada bagian Daging, sehingga untuk mendapatkan nilai yang relevan, maka dilakukan pembersihan pada records Daging dengan cara mencari nilai rata-rata atau Means.

2. 4. 3. Construct Data

Setelah tahap clean data dilakukan, maka tahap ini menghasilkan struktur data baru untuk kedua tabel tersebut. Adapun contoh struktur tabel tersebut disajikan sebagaimana tampak pada Tabel 4 dan Tabel 5.

Tabel 4. Struktur Data Pengeluaran Untuk Makanan

No	Nama Provinsi	2013	2019
1	Aceh	371	559 300
		838		
2	Sumatera Utara	363	576 323
3	Sumatera Barat	419	609 258
		853		

No	Nama Provinsi	2013	2019
4	Riau	468	621 748
		503		
5	Jambi	377	560 542
		133		
...
34	Papua	379 876	665292

Tabel 5. Struktur Data Jenis Makanan

No	Jenis Makanan	Rata-Rata Pengeluaran Per Kapita		
		2013	2019
1	Padi-padian	62 449	559 300
2	Umbi-umbian	3 886	576 323
3	Ikan	26 150	609 258
4	Telur dan susu	14 168	560 542
...
14	Tembakau dan Sirih	41 323	70 654

2. 4. 4. Integrate Data

Tahapan ini menggabungkan kedua data pada tabel menjadi satu tabel yang akan diproses. Adapun hasil integrate data disajikan sebagaimana tampak pada Tabel 6.

Tabel 6. Hasil Integrate Data

No	Nama Provinsi	Rata-Rata Pengeluaran Per Kapita		
		2013	2019
1	Aceh	371 838	559 300
2	Sumatera Utara	363 363	576 323
3	Sumatera Barat	419 853	609 258
4	Riau	468	621 748
		503		
5	Jambi	377	560 542
		133		
.....
6	Papua	379 876	665292

2. 4. 5. Format Data

Format data dilakukan dengan menghasilkan tabel berupa rata-rata keuangan yang dikeluarkan masyarakat untuk makanan berdasarkan provinsi dimana strukturnya berupa kolom provinsi dan kolom rata-rata pengeluaran untuk makanan. Adapun tabel tersebut disajikan sebagaimana tampak pada Tabel 7.

Tabel 7. Hasil Format Data

No	Nama Provinsi	Rata-Rata Pengeluaran
1	Aceh	471 693
2	Sumatera Utara	471 784
3	Sumatera Barat	525 310
4	Riau	545 135
5	Jambi	477 357
.....
34	Papua	544 694

2. 5. Modelling

Tahapan ini secara langsung melibatkan teknik *Data Mining*. Pemilihan teknik *Data Mining*, algoritma dan menentukan parameter dengan nilai yang optimal [12]. Langkah-langkahnya pada modelling sebagai berikut:

2. 5. 1. Select Modelling Technique

Teknik *Data Mining* yang dipilih adalah klasterisasi dimana algoritma yang digunakan adalah K-Means dan K-Medoids;

2. 5. 2. Generate Test Design

Teknik *clustering* tidak memerlukan tahap pembelajaran karena *clustering* bersifat *unsupervised learning* dan melakukan pengelompokan secara alamiah berdasarkan kemiripan atributnya, dimana berbeda dengan teknik klasifikasi lainnya;

2. 5. 3. Build Model

Tahapan ini dilakukan dengan menggunakan aplikasi Rapidminer dimana nantinya data akan klaster dengan menggunakan 2 algoritma yakni *K-Means* dan *K-Medoids*;

2. 5. 4. Assess Model

Penilaian terhadap pusat klaster paling optimal dilakukan menggunakan 2 algoritma *K-Means* dan *K-Medoids*, dimana pusat klaster terbaiklah yang akan digunakan.

2. 6. Evaluation

Tahapan ini adalah fase interpretasi terhadap hasil *Data Mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap modelling sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding* [12]. Adapun kegiatan yang dilakukan dalam tahapan ini adalah sebagaimana berikut:

2. 6. 1. Evaluate Results

Tahapan ini menilai sejauh mana hasil pemodelan *Data Mining* memenuhi tujuan data mining yang ditentukan pada tahap *business understanding*;

2. 6. 2. Review Process

Tahapan ini melakukan pemeriksaan kembali tahapan dari awal untuk memastikan bahwa tidak ada faktor penting dalam proses tersebut yang terabaikan atau terlewat;

2. 6. 3. Determine Next Steps

Tahapan ini menentukan langkah apa yang diambil. Ada 2 pilihan: Kembali ke tahap awal (*business understanding*) atau melanjutkan ke tahap akhir (*deployment*).

2. 7. Deployment

Tahapan membuat laporan hasil kegiatan data mining. Laporan akhir mengenai pengetahuan yang didapat atau pengenalan pola pada data dalam proses data mining akan dijadikan acuan dalam pengambilan keputusan untuk meningkatkan ketahanan pangan di Indonesia.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini mengambil data yang dikeluarkan oleh Badan Pusat Statistik yang dirilis pada tahun 2020 mengenai ketahanan pangan di masing-masing wilayah di Indonesia. Data ini terdiri dari dua tabel yaitu tabel pertama mengenai jumlah rata-rata biaya yang dikeluarkan masyarakat untuk kebutuhan makanan dan tabel kedua adalah mengenai jenis dari makanannya. Data ini dihimpun dari rentang waktu 2013 sampai dengan 2019. Adapun hasil dari tahapan pengumpulan data disajikan dalam bentuk tabel sebagaimana tampak pada Tabel 8.

Tabel 8. Hasil Tahapan Pengumpulan Data

Nama Tahapan	Hasil tahapan	Keterangan
Pengumpulan Data	Data Ketahanan Pangan Indonesia bersumber dari BPS dan ditunjang dengan jurnal-jurnal ilmiah mengenai Pertanian, Data Mining serta Covid-19	Data tersebut terdiri dari dua table yaitu: 1. Tabel pertama Jumlah <i>Fields</i> = 22 Jumlah <i>Records</i> = 770; 2. Tabel kedua Jumlah <i>Fields</i> = 64 Jumlah <i>Records</i> = 1.600

3.2. Business Understanding

Hasil dari tahapan ini adalah menggambarkan tujuan dari penelitian ini, dimana tujuannya adalah melakukan klasterisasi mengenai ketahanan pangan di Indonesia selama masa pandemi covid-19. Data yang digunakan merupakan data dari BPS, nantinya dat tersebut akan di klaster menurut wilayah serta akhirnya akan bermuara pada penentuan strategi yang akan digunakan dalam rangka perkuatan ketahanan pangan Indonesia. Detail mengenai pembahasan dan turunan dari tahapan *Business Understanding* telah dibahas pada sub.bab sebelumnya.

3.3. Data Understanding

Hasil dari tahapan ini adalah pemahaman mengenai data yang telah diperoleh. Adapun hasil tersebut disajikan dalam bentuk tabel sebagaimana tampak pada Tabel 9.

Tabel 9. Hasil Tahapan *Data Understanding*

No	Nama Atribut	Tipe Atribut
1	Provinsi	<i>Polynominal</i>
2	Kebutuhan Makanan Berdasarkan Tahun	<i>Polynominal</i>
3	Kebutuhan Non Makanan Berdasarkan Tahun	<i>Polynominal</i>
4	Jumlah kebutuhan pangan dan non pangan berdasarkan tahun	<i>Polynominal</i>
5	Kelompok barang	<i>Polynominal</i>
6	Rentang biaya yang dikeluarkan masyarakat dalam satuan rupiah	<i>Polynominal</i>
7	Rata-rata pengeluaran per kapita	<i>Polynominal</i>

3.4. Data Preparation

Data hasil dari tahapan *Data Preparation* merupakan data akhir yang akan digunakan dalam proses *Data Mining*. Adapun hasil tahapan ini disajikan dalam bentuk tabel sebagaimana tampak pada Tabel 10.

Tabel 10. Hasil Tahapan *Data Preparation*

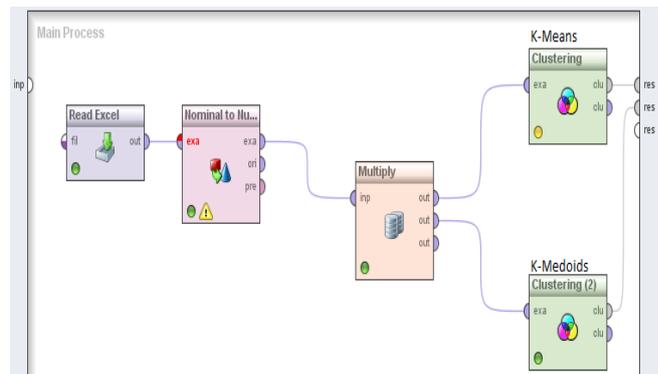
No	Nama Provinsi	Rata-Rata Pengeluaran
1	Aceh	471 693
2	Sumatera Utara	471 784
3	Sumatera Barat	525 310
4	Riau	545 135
5	Jambi	477 357
.....
34	Papua	544 694

3.5. Modelling

Tahapan ini langsung melibatkan aplikasi Rapidminer untuk membantu dalam proses *Data Mining* [14]. Tahap *Modelling* terdiri dari empat tahapan dimana dua tahapan telah dibahas sebelumnya yaitu poin satu dan juga dua. Dua poin lainnya yang akan dibahas adalah tahap *build model* dan *assess model*.

2.5.1. Build Model

Model yang dikonstruksi dalam aplikasi Rapidminer adalah pengolahan data ketahanan pangan yang diklasterisasi menggunakan dua algoritma *Data Mining* yaitu *K-Means* dan *K-Medoids*. Adapun modelnya disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 5.



Gambar 5. Model Data Mining Ketahanan Pangan

Proses konstruksi model pada Rapidminer, pada awal mulanya adalah data yang digunakan itu berekstensi .xlsx (file excel) kemudian masuk melalui operator *Nominal to Numerical* dimana fungsinya adalah merubah tipe data yang bukan numerik menjadi numerik. Selanjutnya adalah masuk ke operator *Multiply* yang berfungsi untuk memproses lebih dari 1 algoritma, dimana yang digunakan adalah *K-Means* dan *K-Medoids*. Kemudian adapaun parameter yang digunakan untuk algoritma *K-Means* disajikan pada Tabel 11

sedangkan untuk algoritma *K-Medoids* disajikan pada Tabel 12.

Tabel 11. Parameter Algoritma K-Means

Parameter	Nilai
K	2 sampai dengan 10
Max runs	10 (default)
Measure type	Numerical measurement
Numerical measure	EuclideanDistance
Max optimization step	100 (default)

Tabel 12. Parameter Algoritma K-Medoids

Parameter	Nilai
K	2 sampai dengan 10
Max runs	10 (default)
Max Optimization Step	100
Measure type	Numerical Measurement
Numerical measure	EuclideanDistance
Max optimization step	100 (default)

2. 5. 2. Assess Model

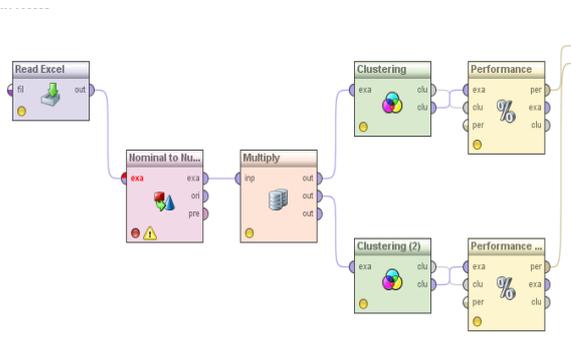
Pemodelan dilakukan dari K=2 sampai dengan K=10 agar dapat menilai pemodelan yang menghasilkan pusat kluster akhir yang paling optimal. Setelah itu maka tahapan selanjutnya adalah mencari K yang paling optimal dengan cara menggunakan indeks validitas Davies-Bouldin.

3. 6. Evaluation

Tahapan ini akan dilakukan evaluasi terhadap model yang telah dibuat pada sub.bab sebelumnya, dengan tujuan untuk mencari nilai K yang optimal, lalu setelah ditemukan, akan dilakukan analisis yang mencakup ciri-ciri dari setiap kluster yang telah terbentuk.

3. 6. 1. Evaluate Results

Indeks Davies-Bouldin digunakan sebagai indeks validitas untuk mengetahui nilai K yang terbaik dari hasil percobaan K=2 sampai dengan K=10. Adapun penggunaan indeks Davies-Bouldin ini disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 6.



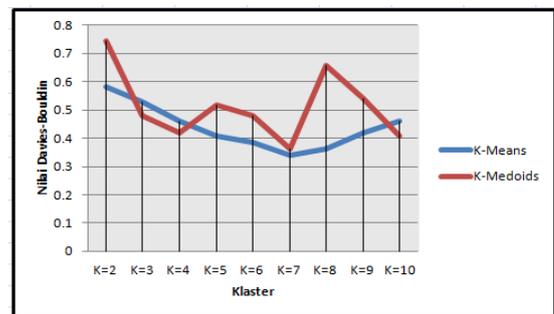
Gambar 6. Indeks Validitas Davies-Bouldin

Berdasarkan Gambar 6, maka hasil perhitungan indeks validitas Davies-Bouldin disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 13.

Tabel 13. Hasil Perhitungan Indeks Validitas Davies-Bouldin

No	Klaster	K-Means	K-Medoids
1	K=2	0,583	0,745
2	K=3	0,528	0,481
3	K=4	0,460	0,421
4	K=5	0,409	0,518
5	K=6	0,384	0,481
6	K=7	0,341	0,362
7	K=8	0,361	0,657
8	K=9	0,420	0,539
9	K=10	0,459	0,409

Hasil perhitungan yang tertera pada Tabel 13 ditranslasikan ke dalam bentuk grafik, sebagaimana tampak pada Gambar 7.



Gambar 7. Grafik Hasil Indeks Davies-Bouldin

Jumlah kluster terbaik dari indeks validitas Davies-Bouldin ditunjukkan dengan nilai indeks yang semakin kecil [15]. Berdasarkan Gambar 4.3, maka kluster terbaik untuk Algoritma *K-Means* adalah K=7 dengan nilai 0,341 dan untuk Algoritma *K-Medoids* adalah K=7 dengan nilai 0,362. Pada penelitian ini menghasilkan Algoritma terbaik yaitu *K-Means* dengan nilai 0,341, dimana memiliki nilai yang lebih kecil daripada *K-Medoids* dengan nilai 0,362. Maka model kluster untuk k=7 Algoritma *K-Means* disajikan dalam bentuk gambar sebagaimana tampak pada Gambar 8.

```

Cluster Model
Cluster 0: 7 items
Cluster 1: 8 items
Cluster 2: 9 items
Cluster 3: 1 items
Cluster 4: 2 items
Cluster 5: 1 items
Cluster 6: 5 items
Total number of items: 33
    
```

Gambar 8. Model Kluster K=7 K-Means

Setelah diketahui bahwa nilai K optimal adalah 7, maka langkah selanjutnya adalah menganalisis ciri-ciri dari setiap kluster yang terbentuk. Adapun

hasil analisis tersebut disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 14.

Tabel 14. Hasil Analisis Ciri-Ciri Tiap Klaster

Klaster	Nama Provinsi	Rata-Rata Pengeluaran
		PerKapita
Klaster 0	Aceh	Rp. 476.472
	Sumatera Utara	
	Jambi	
	Bengkulu	
	Jawa Barat	
Klaster 1	Kalimantan Barat	Rata-Rata Pengeluaran Per Kapita Rp. 540.129
	Sulawesi Utara	
	Sumatera Barat	
	Riau	
	Banten	
Klaster 2	Bali	Rata-Rata Pengeluaran Per Kapita Rp. 425.678
	Kalimantan Tengah	
	Kalimantan Selatan	
	Papua Barat	
	Papua	
Klaster 3	Sumatera Selatan	Rata-Rata Pengeluaran Per Kapita Rp. 425.678
	Lampung	
	DI Yogyakarta	
	Jawa Timur	
	Nusa Tenggara Barat	
Klaster 4	Sulawesi Tengah	Rata-Rata Pengeluaran Per Kapita Rp. 605.595
	Sulawesi Selatan	
Klaster 5	Maluku	Rata-Rata Pengeluaran Per Kapita Rp. 648.914
	Maluku Utara	
Klaster 6	Kalimantan Timur	Rata-Rata Pengeluaran Per Kapita Rp. 727.510
	Kepulauan Bangka Belitung	
Klaster 6	Kepulauan Riau	Rata-Rata Pengeluaran Per Kapita Rp. 366.455
	DKI Jakarta	
	Jawa Tengah	
	Nusa Tenggara Timur	
	Sulawesi Tenggara	
	Gorontalo	
	Sulawesi Barat	

Berdasarkan Tabel 14. maka diketahui pengeluaran rata-rata masyarakat per kapita untuk pangan tiap klaster dari yang tertinggi hingga yang terendah nilainya. Adapun ranking tiap klaster disajikan dalam bentuk tabel sebagaimana tampak pada Tabel 15.

Tabel 15. Ranking Klaster Berdasarkan Pengeluaran (Tertinggi-Terendah)

No	Nama Klaster	Rata-Rata Pengeluaran
1	Klaster 5	Rp. 727.510
2	Klaster 4	Rp. 648.914
3	Klaster 3	Rp. 605.595
4	Klaster 1	Rp. 540.129
5	Klaster 0	Rp. 476.472
6	Klaster 2	Rp. 425.678
7	Klaster 6	Rp. 366.455

Berdasarkan Tabel 15. dapat diketahui bahwa klaster daerah dengan rata-rata pengeluaran untuk makanan tertinggi adalah klaster 5 yang meliputi daerah DKI Jakarta, sedangkan klaster dengan pengeluaran terendah yaitu klaster 6 yang meliputi Jawa Tengah, Nusa Tenggara Timur, Sulawesi Tenggara, Gorontalo dan Sulawesi Barat. Pada klaster 6 perlu dilakukan strategi peningkatan ketahanan pangan melalui peningkatan kapasitas produksi serta peningkatan cadangan pangan di setiap daerah.

3. 6. 2. Review Process

Semua proses sudah dilakukan sesuai dengan metodologi yang digunakan yaitu CRISP-DM, sehingga dapat melanjutkan ke tahapan berikutnya.

3. 6. 3. Determine Next Step

Tahap ini akan menentukan apakah akan dilanjutkan ke tahap selanjutnya atau ada beberapa hal yang harus dikerjakan ulang. Semua proses telah dilakukan dan menghasilkan data dan pengetahuan baru terkait klasterisasi ketahanan pangan Indonesia di masa pandemi Covid-19, sehingga diputuskan penelitian ini dapat dilanjutkan ke tahap selanjutnya yaitu *Deployment*.

3. 6. 4. Deployment

Tahapan ini adalah pembuatan laporan atau dapat juga diartikan sebagai penerapan dari hasil *Data Mining*. Hasil yang diperoleh adalah “pengelompokan Data Pengeluaran Pangan menunjukkan bahwa klaster 6 merupakan klaster dengan tingkat pengeluaran pangan terendah dimana wilayahnya meliputi Jawa Tengah, NTT, Sulawesi Tenggara, Gorontalo, dan Sulawesi Barat. Strateginya adalah meningkatkan kapasitas produksi dan menambah cadangan pangan di masing-masing daerah”.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan:

1. Penentuan nilai K optimum untuk algoritma K -Means dan K -Medoids dilakukan dengan menggunakan Indeks Validitas Davies-Bouldin, dimana algoritma K -Means merupakan algoritma yang lebih baik dari algoritma K -Medoids, dengan hasil optimal untuk K -Means = 0,341, sedangkan untuk K -Medoids = 0,362;
2. Pengelompokan Pengeluaran Pangan Data menunjukkan bahwa klaster 6 merupakan klaster dengan tingkat pengeluaran pangan terendah dimana wilayahnya meliputi Jawa Tengah, NTT, Sulawesi Tenggara, Gorontalo, dan Sulawesi Barat. Strateginya adalah meningkatkan kapasitas produksi dan menambah cadangan pangan di masing-masing daerah.

DAFTAR PUSTAKA

- [1] R. Chairani, D. Agustanto, R. Amriza Wahyu, and P. Nainggolan, "Ketahanan Pangan Berkelanjutan," *J. Kependud. dan Pembang. Lingkung.*, vol. 1, no. 2, pp. 70–79, 2020, [Online]. Available: <http://jkpl.pjpj.unp.ac.id/index.php/JKPL/article/view/13>.
- [2] Z. R. S. Elsi *et al.*, "Utilization of Data Mining Techniques in National Food Security during the Covid-19 Pandemic in Indonesia," *J. Phys. Conf. Ser.*, pp. 1–7, 2020, doi: 10.1088/1742-6596/1594/1/012007.
- [3] Han and Kamber, *Data Mining Concepts and Technique*. San Francisco: Diane Cerra, 2006.
- [4] C. C. Aggarwal, *Data Mining: The Textbook*. New York: Springer, 2015.
- [5] N. Ye, *Data Mining: Theories, Algorithms, and Examples*. New York: Taylor & Francis Group, 2014.
- [6] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 2, pp. 173–178, 2018.
- [7] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. Minnesota: Taylor & Francis Group, LLC, 2009.
- [8] A. Alqurneh, A. Mustapha, and N. M. Sharef, "A Partitioning-based Approach for Clustering COVID-19 Drugs and Co-Medication for Safe Use," *Int. J. Integr. Eng.*, vol. 5, pp. 224–232, 2020.
- [9] S. Sindi, W. R. O. Ningse, I. A. Sihombing, P. P. P. A. N. . F. I. R.H.Zer, and D. Hartama, "Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020.
- [10] W. Yustanti, N. Rahmawati, and Y. Yamasari, "Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 di Propinsi Jawa Timur dengan Metode K-Means," *J. Inf. Eng. Educ. Technol.*, vol. 04, no. 01, pp. 1–9, 2020.
- [11] A. Doroshenko, "Analysis of the Distribution of COVID-19 in Italy Using Clustering Algorithms," *IEEE Third Int. Conf. Data Stream Min. Process.*, pp. 325–328, 2020.
- [12] I. Budiman, "Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan TRIDHARMA," Universitas Diponegoro Semarang, 2012.
- [13] F. N. Dhewayani, D. Amelia, D. N. ALifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.
- [14] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Comparison Of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends," *JUTIF*, vol. 3, no. 2, pp. 245–248, 2022.
- [15] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019.

