

Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means

Tonny Tendean¹, Windania Purba, M.Kom²

^{1,2}Sistem Informasi Universitas Prima Indonesia

¹tonnytt07@gmail.com, ²winda.nia04@gmail.com

Abstract-Food is material produced through agricultural products which has a great influence on human survival. Having agriculture spread throughout Indonesia, making Indonesia a country that always produces raw food. Raw food can be in the form of fruits, vegetables, rice, peanuts, and others. Clustering algorithm is applied to group the number of provinces according to the results of food production with K-Means. The data from this study were sourced from information on the number of provinces according to the results of food production produced by the National Statistics Agency. There are 34 provinces that will be used in this study. There are 5 variables used, namely corn production, peanut production, soybean production, rice production, and cassava production. The data collected will be processed by clustering in 3 clusters, namely high foodstuff production clusters, medium foodstuff production clusters, and low foodstuff production clusters. In its implementation using rapidminer software. The results obtained can be input for the government in analyzing the provinces according to food production.

Keywords- Data Mining, Foodstuffs, Clustering Algorithm, K-Means

Abstrak- Bahan pangan adalah bahan yang diproduksi melalui hasil pertanian yang memiliki pengaruh besar bagi kelangsungan hidup manusia. Memiliki pertanian yang tersebar di seluruh wilayah Indonesia, membuat Indonesia merupakan Negara yang selalu memproduksi bahan pangan mentah. Bahan pangan mentah tersebut dapat berupa buah-buahan, sayuran, padi, kacang tanah, dan lain-lain. Algoritma *clustering* diterapkan untuk mengelompokkan banyaknya provinsi menurut hasil produksi bahan pangan dengan K-Means. Data dari penelitian ini bersumber dari dokumen-dokumen keterangan banyaknya provinsi menurut hasil produksi bahan pangan yang dihasilkan oleh Badan Pusat Statistik Nasional. Terdapat 34 provinsi yang akan digunakan pada penelitian ini. Variable yang digunakan ada 5, yaitu produksi jagung, produksi kacang tanah, produksi kedelai, produksi padi, dan produksi ubi kayu. Data yang telah dikumpulkan akan diolah dengan melakukan clustering dalam 3 cluster, yaitu cluster produksi bahan pangan tinggi, cluster produksi bahan pangan sedang, dan cluster produksi bahan pangan rendah. Dalam implementasinya menggunakan *software rapidminer*. Hasil yang diperoleh dapat menjadi masukan bagi pemerintah dalam menganalisa provinsi menurut produksi bahan pangan.

Kata kunci- Data Mining, Bahan Pangan, Algoritma Clustering, K-Means

I. PENDAHULUAN

Manusia tidak dapat hidup tanpa energi, dalam setiap aktivitas mulai dari transportasi, pekerjaan, bahkan aktivitas yang paling mendasar dan dibutuhkan, makan (memasak), kita membutuhkan energi. Indonesia merupakan salah satu negara agraris yang sebagian besar penduduknya mayoritas bekerja di sektor pertanian. Sektor pertanian memiliki peranan yang sangat penting dalam perekonomian nasional, hal ini dapat dilihat dari banyaknya jumlah penduduk Indonesia yang hidup dan bekerja disektor tersebut. Sektor pertanian merupakan sektor yang mendapat perhatian yang cukup besar dari pemerintah karena memiliki peran dalam rangka pembangunan ekonomi jangka panjang maupun dalam rangka pemulihan ekonomi bangsa. Padi merupakan bahan pokok sehari-hari yang dibutuhkan setiap rumah tangga sebelum diolah menjadi nasi. Sedangkan bahan pangan lainnya seperti ubi kayu, jagung, padi, kedelai, dan kacang tanah merupakan bahan pelengkap bagi setiap rumah tangga.

Dengan mengetahui karakteristik masing-masing provinsi, pemerintah dapat mengambil kebijakan yang sesuai untuk masing-masing provinsi sesuai karakternya. Oleh karena itu, akan sangat baik jika ada analisis cluster

dari semua provinsi di Indonesia berdasarkan jenis produksi bahan pangan. Untuk setiap provinsi, bahan pangan utama dihitung dari persentase hasil produksi dari padi, kacang tanah, kedelai, jagung dan ubi kayu. Dalam tulisan ini, metode K-Means digunakan untuk mengelompokkan provinsi Indonesia berdasarkan produksi bahan pangan.

II. METODE PENELITIAN

A. PENGERTIAN DATA MINING

Data mining merupakan proses untuk menemukan karakteristik yang penting dalam data yang sedang digunakan dan menemukan model untuk menggambarkan kelas atau suatu konsep data. Data mining juga mengerjakan lebih dari satu teknik pembelajaran komputer untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Dalam artian data mining dapat melakukan suatu proses iteratif dan interaktif untuk mencari model atau pola baru dan juga dapat menemukan suatu hubungan pada nilai atribut dari beberapa data. Data mining adalah sekumpulan proses untuk mencari nilai tambah dari suatu kumpulan data yang berupa pengetahuan yang selama ini tidak diketahui secara manual[1][7].

B. CLUSTERING

Clustering dapat diartikan proses untuk mengorganisasikan sekelompok data ke dalam berbagai kelompok-kelompok sedemikian rupa sehingga objek-objek yang serupa akan menjadi satu *cluster* sedangkan objek-objek yang tidak serupa menjadi anggota *cluster* yang lain. Dalam setiap *cluster* berisi data yang semirip mungkin. Ukuran kemiripan biasanya dihitung dengan jarak. Jarak dalam satu *cluster* dibuat sedekat mungkin dan jarak antar *cluster* diusahakan untuk sejauh mungkin. Jadi dalam satu *cluster* harus sama dan dengan *cluster* yang lain harus berbeda. Defenisi ini mengasumsikan bahwa ada beberapa parameter penting yang mewakili kesamaan atau ketidaksamaan antar *cluster*[2].

C. ALGORITMA K-MEANS

K-Means merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. *K-Means* merupakan algoritma *clustering* dengan metode partisi (*partitioning method*) yang berbasis titik pusat (*centroid*) selain algoritma *k-Medoids* yang berbasis objek[3].

Berikut adalah langkah-langkah algoritma *K-Means*, yaitu[4]:

1. Tentukan jumlah *cluster* (*k*) pada data *set* sebagai nilai *centroid*.
2. Menghitung jarak antara data dan titik pusat *cluster* menggunakan rumus-rumus dari *Euclidian Distance*. Yang dapat dilihat pada teori jarak *Euclidian* yang dirumuskan sebagai berikut:

$$D_e = \sqrt{(X_i - S_i)^2 + (Y_i - t_i)^2}$$

Dimana :

- D_e adalah *Euclidean Distance*
 - i adalah banyaknya objek
 - (x, y) adalah koordinat objek
 - (s, t) adalah koordinat *centroid*
3. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat.
 4. Proses penentuan titik pusat *cluster* dan penempatan data dalam *cluster* diulangi terus-menerus sampai nilai *centroid* tidak berubah lagi.

D. RAPID MINER

Aplikasi yang digunakan dalam pengujian *clustering* dengan algoritma *K-Means* ini adalah salah satu aplikasi *Data Mining* yang disebut *Rapid miner*. *Rapid miner* merupakan perangkat lunak yang bersifat *Open Source* yang dapat melakukan analisis terhadap *Data Mining*, *Text Mining* dan analisis prediksi serta mampu menggunakan berbagai teknik deskriptif dan prediktif.

Dengan menggunakan *Rapid Miner*, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. *Rapid Miner* dikhususkan untuk penggunaan data mining.

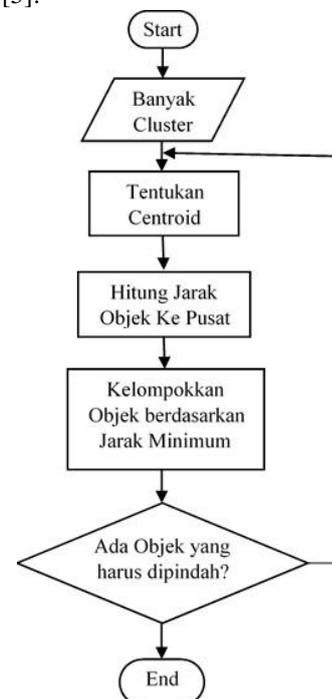
III. HASIL DAN PEMBAHASAN

A. TAHAP PENGUMPULAN DATA

Dalam pemanfaatan algoritma *clustering* dalam mengelompokkan banyaknya provinsi menurut jumlah produksi bahan pangan, diperlukan data terkait mengenai hal itu. Data tersebut diperoleh dari data yang dikumpulkan berdasarkan dokumen-dokumen keterangan dihasilkan oleh badan pusat statistik dan kementerian pertanian. Data yang digunakan pada penelitian ini adalah data hasil produksi beberapa bahan pangan menurut provinsi, yang terdiri dari 34 provinsi. Ada 4 variabel yang digunakan yaitu produksi jagung, produksi kacang tanah, produksi ubi kayu, produksi kedelai, dan produksi padi. Data diolah dengan melakukan *clustering* banyaknya provinsi menurut produksi bahan pangan sebanyak 3 *cluster* yaitu *cluster* tingkat produksi tinggi, *cluster* tingkat produksi sedang dan *cluster* tingkat produksi rendah.

B. TAHAP CLUSTERING

Pada tahap ini dilakukan pengolahan data menggunakan teori-teori dari data mining dengan menggunakan algoritma *clustering*. Berikut diperlihatkan urutan arsitektur sistem dari algoritma *K-Means Clustering* yang akan dilakukan dalam penelitian ini yang ditunjukkan pada gambar 1[5].



Gambar 1. Tahapan *Clustering*

C. TAHAP ANALISIS

Pada tahap analisis ini dilakukan analisis data banyaknya provinsi menurut produksi bahan pangan. Data yang diperoleh dari hasil pengumpulan data terdiri dari produksi jagung, produksi kacang tanah, produksi ubi kayu, produksi kedelai, dan produksi padi. Adapun kriteria yang digunakan yaitu sebanyak 5 kriteria penilaian seperti yang diperlihatkan pada tabel 1.

Tabel 1. Data Jumlah Produksi Bahan Pangan Menurut Provinsi

Provinsi	Jagung	Kacang Tanah	Kedelai	Padi	Ubi Kayu
Aceh	347735	3658	15835	2516221	15672
Sumatera Utara	1757126	4323	32758	5423154	803403
Sumatera Barat	1052408	4717	1117	2754079	201833
Riau	25723	1058	6488	391132	133738
Jambi	152158	1488	15400	855944	56605
Sumatera Selatan	935240	1099	14955	5076831	382043
Bengkulu	111816	2960	3477	699531	73491
Lampung	2581224	3333	72006	4556378	6683758
Kepulauan Bangka Belitung	3630	201	0	28310	59426
Kepulauan Riau	87	74	5	651	19722
DKI Jakarta	0	0	0	4183	0
Jawa Barat	1550966	39601	132099	12494919	1635031
Jawa Tengah	3688477	94291	130525	11401821	3267417
DI Yogyakarta	314179	107376	11093	878136	859393
Jawa Timur	6543359	150180	244442	13000475	2551840
Banten	174334	5348	18446	2470538	72616
Bali	51459	5993	3996	848698	92144
Nusa Tenggara Barat	2059222	33303	91724	2423285	58021
Nusa Tenggara Timur	859230	10682	16827	1213760	853468
Kalimantan Barat	166826	851	1260	1625355	147475
Kalimantan Tengah	158964	415	1413	783497	142852
Kalimantan Selatan	364489	6311	24647	2528593	88974
Kalimantan Timur	88105	1003	582	385544	86079
Kalimantan Utara	5977	194	1144	68793	44050
Sulawesi Utara	1531241	2442	50026	887758	41651
Sulawesi Tengah	380650	3366	27691	1154907	48405
Sulawesi Selatan	2341659	19152	35824	6196737	422601
Sulawesi Tenggara	192329	2606	8007	716156	209159
Gorontalo	1619649	341	3257	350256	2781
Sulawesi Barat	702339	391	15091	751531	22174
Maluku	40550	834	228	132852	85734
Maluku Utara	237778	2185	115	101054	122706
Papua Barat	4218	634	359	27736	22798
Papua	12476	1788	1761	288335	34173

Selanjutnya data tersebut akan dilakukan tahap *clustering* dengan menggunakan algoritma *K-Means* untuk mengelompokkan data menjadi tiga *cluster*.

D. CENTROID DATA

Dalam penelitian ini *centroid* awal ditentukan dengan cara *random* atau secara acak dan didapat *centroid* dari setiap *cluster* dapat ditunjukkan pada tabel 2. Penentuan *cluster* dibagi atas tiga, yaitu produksi bahan pangan rendah (C1), produksi bahan pangan sedang (C2), dan produksi bahan pangan tinggi (C3). Dimana titik *cluster* yang diambil adalah data ke 10,20, dan 27.

Tabel 2. Titik Pusat Awal *Cluster*

Data ke-I	Cluster	Jagung	Kacang Tanah	Kedelai	Padi	Ubi Kayu
10	C1	87	74	5	651	19722
20	C2	166826	851	1260	1625355	147475
27	C3	2341659	19152	35824	6196737	422601

E. CLUSTERING DATA

Setelah menentukan *centroid* data, maka langkah selanjutnya menghitung jarak data ke *centroidnya*, Selanjutnya adalah mengelompokkan data berdasarkan jarak minimumnya. Pengelompokkan tersebut adalah iterasi 1 dan dapat dilihat pada tabel 1 sebagai berikut.

Tabel 3. Hasil Perhitungan Jarak Pusat *Cluster*

Provinsi	C1	C2	C3
Aceh	2539534	918674,4	4205730
Sumatera Utara	5753779	4169362	1041801
Sumatera Barat	2953292	1435703	3682963
Riau	407646	1242350	6257255
Jambi	869627	775026,5	5783828
Sumatera Selatan	5174323	3543777	1798502
Bengkulu	709808,5	930408,1	5942614
Lampung	8475346	7559644	6477022
Kepulauan Bangka Belitung	48518,01	1607775	6606770
Kepulauan Riau	0	1638227	6636143
DKI Jakarta	20036,1	1636393	6634104
Jawa Barat	12694101	11058694	6463124
Jawa Tengah	12416293	10850843	6083947
DI Yogyakarta	1259094	1048008	5709411
Jawa Timur	14775105	13263330	8278858
Banten	2476665	848710,9	4324909
Bali	852712,6	787147,4	5827257
Nusa Tenggara Barat	3181222	2057934	3801967
Nusa Tenggara Timur	1704491	1071260	5216678
Kalimantan Barat	1638227	0	5069976
Kalimantan Tengah	808240,6	841907,5	5843554
Kalimantan Selatan	2555136	926773,3	4180440
Kalimantan Timur	400367,6	1243824	6242058
Kalimantan Utara	72602,94	1568265	6569016
Sulawesi Utara	1770418	1555396	5384017
Sulawesi Tengah	1216033	526841,4	5422725
Sulawesi Selatan	6636143	5069976	0
Selawesi Tenggara	764762,2	911672,5	5890925
Gorontalo	1656956	1938430	5905955
Sulawesi Barat	1028208	1032586	5700769
Maluku	153207,6	1499108	6494681
Maluku Utara	277827,2	1526153	6455631
Papua Barat	27578,31	1610706	6609209
Papua	288323,5	1350661	6362912

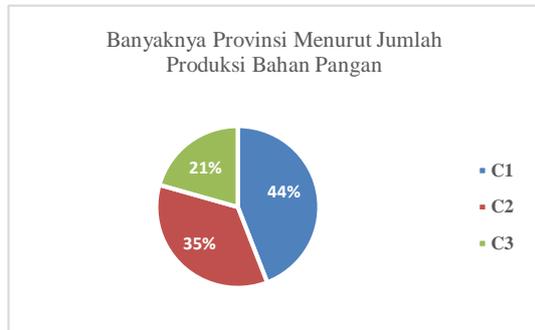
Setelah data didapat berdasarkan perhitungan antara pusat titik *cluster* terhadap data maka langkah selanjutnya mengelompokkan objek berdasarkan jarak minimumnya dimana setiap *cluster* yang jaraknya minimum dari *cluster* yang lain maka akan diberi tanda angka “1” sedangkan jarak *cluster* yang terbesar diberi angka “0” sehingga hasilnya seperti terlihat pada tabel 4.

Tabel 4. Jarak dengan *Centroid* Awal

Provinsi	C1	C2	C3
Aceh	0	1	0
Sumatera Utara	0	0	1
Sumatera Barat	0	1	0
Riau	1	0	0
Jambi	0	1	0
Sumatera Selatan	0	0	1
Bengkulu	1	0	0
Lampung	0	0	1
Kepulauan Bangka Belitung	1	0	0
Kepulauan Riau	1	0	0
DKI Jakarta	1	0	0
Jawa Barat	0	0	1
Jawa Tengah	0	0	1
DI Yogyakarta	0	1	0
Jawa Timur	0	0	1
Banten	0	1	0
Bali	0	1	0
Nusa Tenggara Barat	0	1	0
Nusa Tenggara Timur	0	1	0
Kalimantan Barat	0	1	0
Kalimantan Tengah	1	0	0
Kalimantan Selatan	0	1	0
Kalimantan Timur	1	0	0
Kalimantan Utara	1	0	0
Sulawesi Utara	0	1	0
Sulawesi Tengah	0	1	0
Sulawesi Selatan	0	0	1
Selawesi Tenggara	1	0	0

Gorontalo	1	0	0
Sulawesi Barat	1	0	0
Maluku	1	0	0
Maluku Utara	1	0	0
Papua Barat	1	0	0
Papua	1	0	0

Berdasarkan nilai minimum yang telah dihasilkan pada Tabel 4. tersebut di atas pada penentuan nilai *centroid* maka diperoleh hasil pengelompokkan seperti terlihat statistik pada gambar 2.



Gambar 2. Statistik Pengelompokkan Iterasi 1

Dari hasil *centroid* yang didapatkan pada iterasi 1, kemudian dilakukan lagi perhitungan yang sama sampai data tiap *cluster* tidak ada lagi yang berubah. Proses iterasi tersebut berhenti pada iterasi ke-9, dimana pada iterasi ke-8 memiliki data yang sama pada iterasi ke-9.

Tabel 5. Titik Pusat Iterasi 9

Cluster	Jagung	Kacang Tanah	Kedelai	Padi	Ubi Kayu
C1	394724,88 89	7415,5185 19	11851,4 4444	921734, 6296	133153,33 33
C2	1903812,2 5	6976,75	38885,7 5	5313275	2072951,2 5
C3	3927600,6 67	94690,666 67	169022	1229907 1,67	2484762,6 67

Proses yang sama dilakukan dengan mengelompokkan data berdasarkan jarak minimumnya. Pengelompokkan data pada iterasi ke-9 ditunjukkan pada tabel 6 berikut ini.

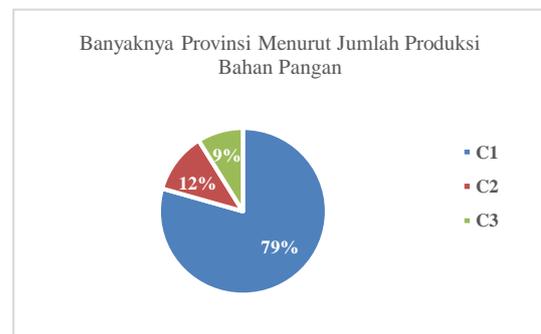
Tabel 6. Hasil Perhitungan Jarak Pusat Cluster Iterasi 9

Provinsi	C1	C2	C3
Aceh	1599508,3	3804974	10707370
Sumatera Utara	4750641,5	1282727	7405599,8
Sumatera Barat	1948043,2	3282817	10228474
Riau	646352,13	5613942	12750929
Jambi	262820,24	5196376	12293387
Sumatera Selatan	4197496,6	1963114	8097463,1
Bengkulu	364775,92	5338258	12448277
Lampung	7804195,5	4721486	8911304,9
Kepulauan Bangka Belitung	978156,94	5966355	13110640
Kepulauan Riau	1008560,9	6005452	13144978
DKI Jakarta	1007786,7	6009130	13145414
Jawa Barat	11728027	7204307	2532427,8
Jawa Tengah	11424809	6457414	1215018,1
DI Yogyakarta	738786,31	4866309	12089728
Jawa Timur	13770483	8995028	2710613,4
Banten	1565591,4	3882526	10795193
Bali	353427,95	5223837	12324579
Nusa Tenggara Barat	2244526,4	3526985	10340252
Nusa Tenggara Timur	905501,22	4402819	11618536
Kalimantan Barat	739851,28	4508524	11557314
Kalimantan Tengah	273760,61	5223994	12342310
Kalimantan Selatan	1607801,2	3749712	10673635

Kalimantan Timur	619596,85	5615024	12746164
Kalimantan Utara	941669,02	5935005	13074901
Sulawesi Utara	1141349,2	4883681	11914365
Sulawesi Tengah	249032,18	4869391	11947277
Sulawesi Selatan	5630337	1922506	6635484,8
Sulawesi Tenggara	298398,13	5247606	12382607
Gorontalo	1357994	5385093	12421652
Sulawesi Barat	368743,77	5143865	12241117
Maluku	866142,41	5853127	12996873
Maluku Utara	835717,22	5809294	12962362
Papua Barat	981878,58	5979133	13117819
Papua	746485,03	5743274	12869654

Tabel 7. Jarak dengan *Centroid* Iterasi 9

Provinsi	C1	C2	C3
Aceh	1	0	0
Sumatera Utara	0	1	0
Sumatera Barat	1	0	0
Riau	1	0	0
Jambi	1	0	0
Sumatera Selatan	0	1	0
Bengkulu	1	0	0
Lampung	0	1	0
Kepulauan Bangka Belitung	1	0	0
Kepulauan Riau	1	0	0
DKI Jakarta	1	0	0
Jawa Barat	0	0	1
Jawa Tengah	0	0	1
DI Yogyakarta	1	0	0
Jawa Timur	0	0	1
Banten	1	0	0
Bali	1	0	0
Nusa Tenggara Barat	1	0	0
Nusa Tenggara Timur	1	0	0
Kalimantan Barat	1	0	0
Kalimantan Tengah	1	0	0
Kalimantan Selatan	1	0	0
Kalimantan Timur	1	0	0
Kalimantan Utara	1	0	0
Sulawesi Utara	1	0	0
Sulawesi Tengah	1	0	0
Sulawesi Selatan	0	1	0
Sulawesi Tenggara	1	0	0
Gorontalo	1	0	0
Sulawesi Barat	1	0	0
Maluku	1	0	0
Maluku Utara	1	0	0
Papua Barat	1	0	0
Papua	1	0	0



Gambar 3. Statistik Pengelompokkan Iterasi 9

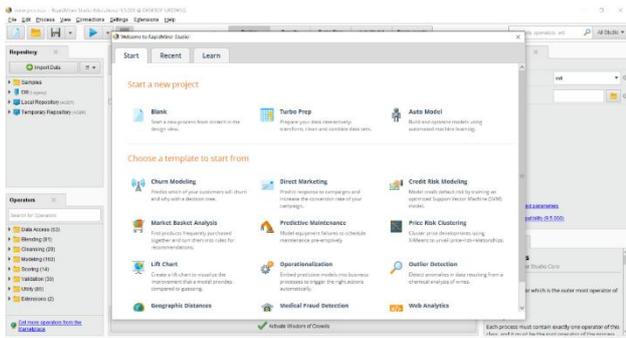
F. ANALISA DATA

Pada iterasi ke-9, pengelompokkan data memiliki hasil yang sama pada iterasi ke-8. Dari 34 provinsi didapatkan 27 provinsi pada cluster 1 (C1) yaitu produksi bahan pangan rendah, 4 provinsi pada cluster 2 (C2) yaitu produksi bahan pangan sedang, dan 3 provinsi pada cluster 3 (C3) yaitu produksi bahan pangan tinggi

G. IMPLEMENTASI SISTEM

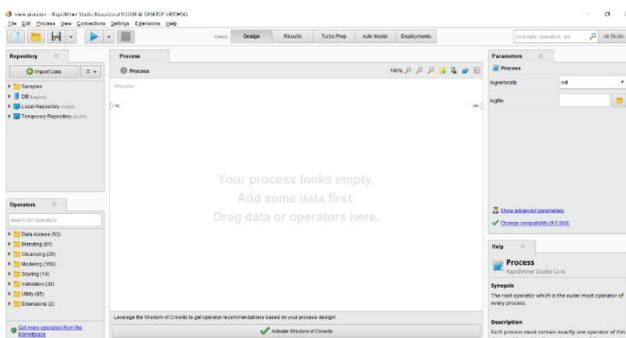
Tahap implementasi sistem adalah tahap penterjemahan perancangan berdasarkan hasil analisis ke dalam suatu bahasa pemrograman tertentu serta penerapan perangkat lunak yang dibangun pada lingkungan yang sesungguhnya. Penulis akan menjelaskan mengenai tahapan pengujian pengelompokan produksi bahan pangan menggunakan *K-Means Clustering*. Pengujian dilakukan dengan menggunakan software rapidminer 9.5.

Interface Rapid miner diawali dengan tampilan menu yang terdiri dari tampilan menu-menu yang terdiri dari *menu bar (File, Edit, Proses, View, Connections, Settings, Extensions, dan Help)* dan *menu shortcut (Start, Recent, dan Learn)* yang dapat dilihat pada gambar 4 berikut ini.



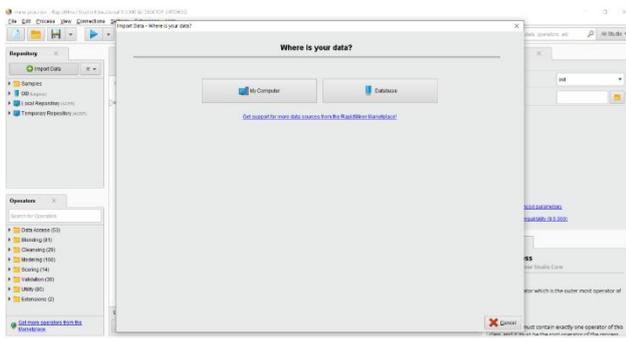
Gambar 4. Halaman Utama Rapid Miner 9.5

Setelah memilih menu *blank*, maka tampil halaman kosong seperti gambar 5 berikut ini.



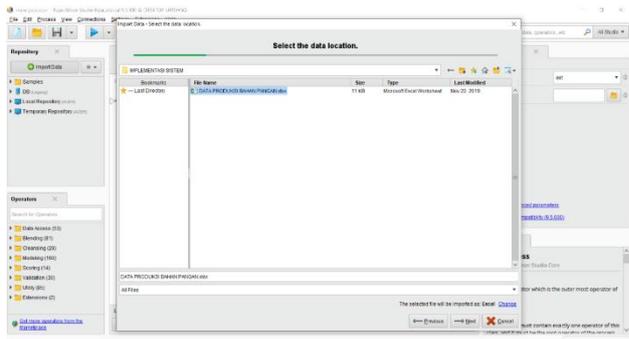
Gambar 5. Halaman New Process Rapid Miner 9.5

Lakukan *import data* dengan mengklik *button import data* yang terdapat pada menu *Repository*. Maka tampil halaman dengan pilihan dimana letak data yang kamu simpan seperti pada gambar 6 berikut ini.



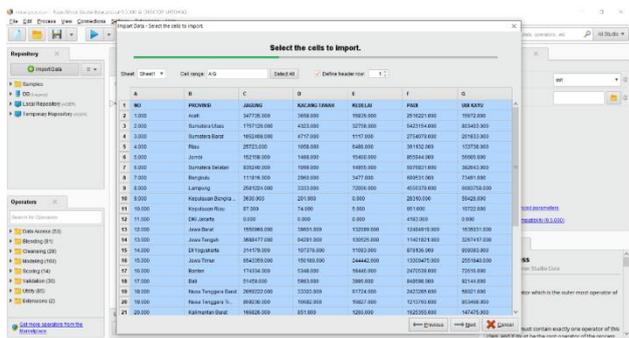
Gambar 6. Halaman Import Data Wizard Step 1 of 5

Langkah selanjutnya adalah mencari data dengan format *excel* pada media penyimpanan setelah itu klik tombol *next* seperti pada gambar 7 berikut ini.



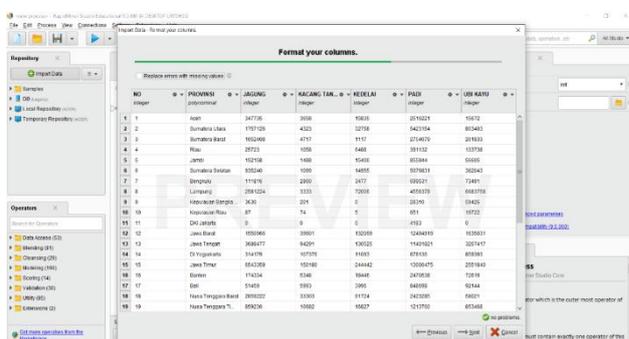
Gambar 7. Halaman Import Data Wizard Step 2 of 5

Setelah mengklik tombol *next*, maka akan tampil halaman data yang telah di import seperti pada gambar 8 berikut ini.



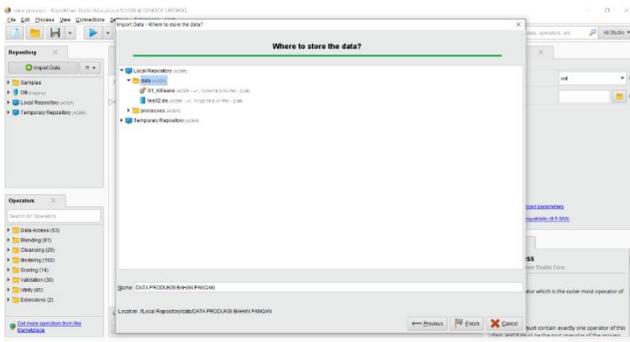
Gambar 8. Halaman Import Data Wizard Step 3 of 5

Selanjutnya klik tombol *next*, maka akan tampil halaman import data untuk memilih tipe data yang akan digunakan dalam proses clustering ini. Seperti pada gambar 9 berikut ini.



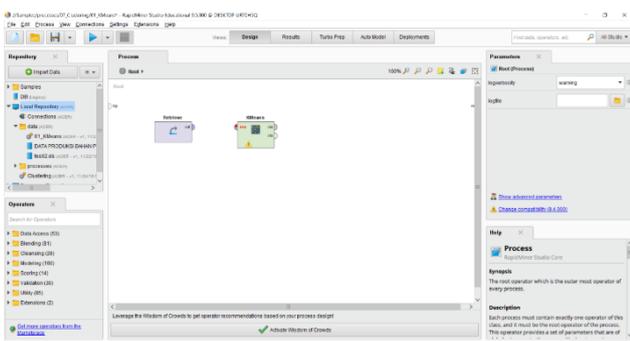
Gambar 9. Halaman Import Data Wizard Step 4 of 5

Setelah menentukan tipe data, maka tahap selanjutnya adalah klik tombol *next* dan akan tampil halaman penyimpanan data repository seperti pada gambar 10 berikut ini.



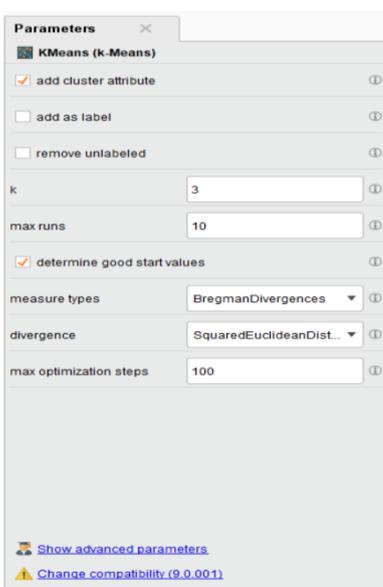
Gambar 10. Halaman *Import Data Wizard Step 5 of 5*

Kembali ke menu awal, lakukan *drag* dengan *mouse* pada *tab repository* untuk data yang sebelumnya sudah disimpan. Dalam hal ini nama data yang disimpan yaitu “data produksi bahan pangan”. Proses ini dapat dilihat pada gambar 11 berikut ini.



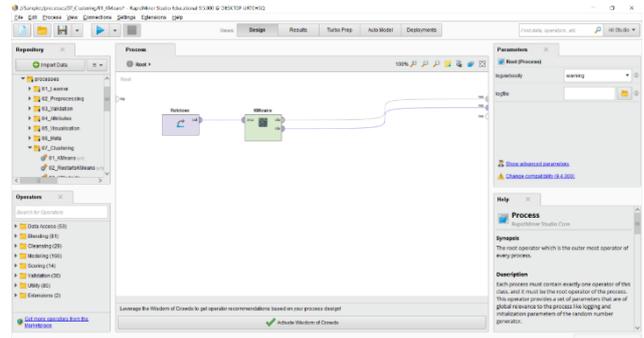
Gambar 11. Halaman Proses *K-Means*

Kemudian lakukan hal yang sama dengan algoritma *K-Means* dengan mencari pada *tab Repository-Samples-Processes-Clustering-KMeans*. Pada *clustering* lakukan pengisian jumlah *k cluster*, dalam pengujian ini *cluster* yang digunakan berjumlah 3 *cluster* sehingga ganti 2 dengan 3 pada *textbox* sebelah kanan. Seperti pada gambar 12 berikut ini.



Gambar 12. Halaman Penentuan Jumlah *Cluster*

Setelah melakukan penentuan jumlah *cluster*, maka langkah selanjutnya adalah menghubungkan data dengan *clustering k-means* untuk mengetahui *output* kearah *result*. Hal ini dapat dilihat pada gambar 13 berikut ini.



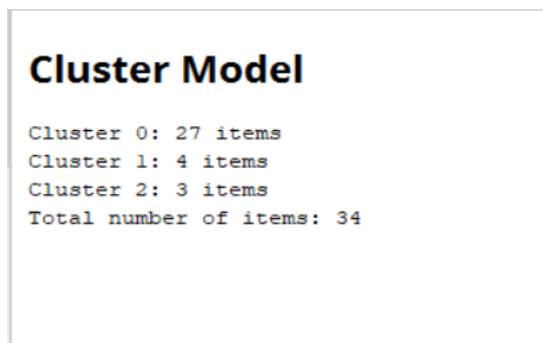
Gambar 13. Proses Pemindahan *K-Means*

Setelah proses dihubungkan klik tanda  (*Run*) pada bagian atas seperti tampak pada tampilan gambar 13. Tombol ini berfungsi untuk menjalankan program dan mendapatkan hasil dari pengolahan *data base*.

Pada hasil pengujian data, terdapat beberapa output yang dihasilkan oleh *software Rapid Miner 9.5* adalah sebagai berikut:

1. *Cluster Model (K-Means)*

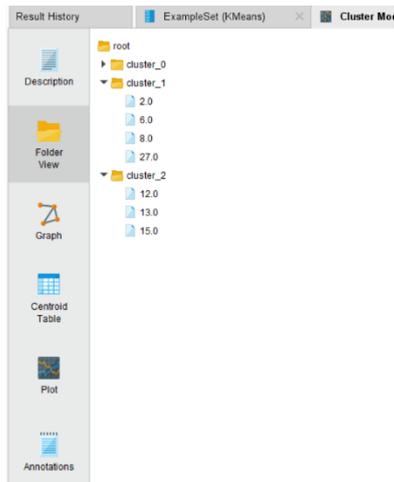
Pada *Cluster Model (K-Means)* dapat dilihat beberapa tampilan hasil *cluster*, yaitu *text view* yang merupakan tampilan hasil pengelompokkan berdasarkan *cluster* dan jumlah anggotanya. Tampilan *text view* dapat dilihat pada gambar 14 berikut ini.



Gambar 14. Tampilan *Cluster Model Text View*

2. *Folder View*

Folder view merupakan tampilan data bagian-bagian cluster secara keseluruhan, dimana masing-masing anggota cluster menampilkan nomor data itu berada. Tampilan *folder view* dapat dilihat pada gambar 15 berikut ini:



Gambar 15. Tampilan Folder View

3. Centroid Table

Centroid table merupakan titik pusat dalam menentukan banyaknya jumlah data yang ada pada setiap cluster. Tampilan centroid table dapat dilihat pada gambar 16 berikut ini:

Attribute	cluster_0	cluster_1	cluster_2
PROVINSI	18.963	10.750	13.333
JAGUNG	394724.889	1903812.250	3927600.667
KACANG TANAH	7415.519	6976.750	94690.667
KEDELAI	11851.444	38885.750	169022
PADI	921734.630	5313275	12299071.667
UBI KAYU	133153.333	2072951.250	2484782.667

Gambar 16. Tampilan Centroid Table

IV.PENUTUP

Berdasarkan hasil analisis terhadap data produksi bahan pangan menurut provinsi menggunakan metode K-Means, maka dapat disimpulkan hal-hal sebagai berikut:

1. Penerapan metode K-Means Clustering untuk mengelompokkan provinsi menurut jumlah produksi bahan pangan dimulai dari analisa data yang terdiri dari produksi jagung, produksi kacang tanah, produksi ubi kayu, produksi kedelai, dan produksi padi.
2. Diperoleh 3 provinsi tingkat tinggi dalam produksi bahan pangan yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur, 4 provinsi dengan tingkat sedang, dan 27 provinsi dengan tingkat produksi bahan pangan rendah.
3. Clustering yang dihasilkan dapat menjadi masukan kepada pemerintah dalam menganalisa provinsi mana saja yang memiliki jumlah produksi bahan pangan yang dominan terhadap jagung, kacang tanah, ubi kayu, kedelai, dan padi.
4. Berdasarkan hasil perbandingan cara manual dengan software menampilkan hasil akhir yang

sama yang dapat dilihat pada cluster model, folder view, dan centroid data.

V. REFERENSI

- [1] Astuti, F dan Hermawati. (2013). Data Mining. Yogyakarta: Andi.
- [2] Muflikhah, L., Ratnawati, D. E., dan Regasari, R. (2018). Data Mining. Malang: Universitas Brawijaya Press.
- [3] Irwansyah, E., dan Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi*. Yogyakarta: DeePublish.
- [4] Witten, I., dan Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. Amsterdam: Elsevier.
- [5] Sadewo, M. G., Windarto, A. P., dan Wanto, A. (2018). Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. Konferensi Nasional Teknologi Informasi dan Komputer. Vol 2 No 1.
- [6] Purba, W., Tamba, S., dan Saragih, J. (2018). *The effect of mining data k-means clustering toward students profile model drop out potential*. Journal of Physics: Conference Series.
- [7] Prasetyo, E. (2012). Data mining-Konsep dan Aplikasi Menggunakan MATLAB. Yogyakarta : Andi.