

IMPLEMENTASI *DATA MINING CLUSTERING* UNTUK MENGETAHUI POTENSI PRODUKTIFITAS KACANG TANAH DI INDONESIA

Herliyani Hasanah¹⁾, Nurmalitasari²⁾, Nugroho Arif Sudibyo³⁾

^{1,2,3)} Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta

e-mail: herliyani_hasanah@udb.ac.id¹⁾, nurmalitasari@udb.ac.id²⁾, nugroho.arif.sudibyo@gmail.com³⁾

ABSTRAK

Kondisi produktifitas kacang tanah yang mengalami fluktuasi, pemerintah perlu memberlakukan kebijakan tertentu di bidang pertanian. Dengan kebijakan tersebut diharapkan dapat mewujudkan ketahanan pangan tingkat nasional. Salah satu informasi yang dapat digunakan untuk mendukung kebijakan tersebut adalah pengelompokan provinsi di Indonesia berdasarkan hasil panen kacang tanah di Indonesia. Pengelompokan ini dilakukan karena beragamnya potensi hasil produksi kacang tanah di masing-masing provinsi, sehingga perlu dilakukan pengelompokan untuk mengetahui provinsi mana yang memiliki hasil produksi kacang tanah tertinggi sehingga dapat membantu untuk mengoptimalkan program-program pemerintah dibidang pertanian tanaman pangan. Pengelompokan tersebut menggunakan analisis cluster k-means. Dalam penelitian ini simulasi pemetaan menggunakan metode *K-Means Clustering* berbasis *mobile*. Proses pemetaan dilakukan berdasarkan 2 (dua) variabel yaitu luas panen (Ha) dan produksi (ton). Hasil pemetaan akan dikelompokkan menjadi dua (2) cluster, yaitu hasil panen tinggi dan rendah. Dari perhitungan manual, aplikasi *mobile* dan simulasi menggunakan Rapid Miner diperoleh hasil yang sama yaitu cluster 1 meliputi Provinsi no 12 (Jawa Barat), 13 (Jawa Tengah) dan 15 (Jawa Timur). Sedangkan sisanya ada 20 data masuk ke *cluster* 0.

Kata Kunci: *cluster k-means*, kacang tanah, *mobile*, pertanian.

ABSTRACT

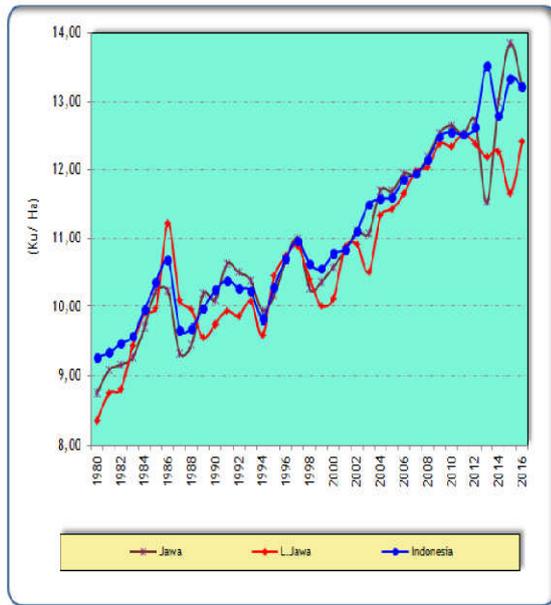
As peanut productivity conditions increase fluctuations, the government needs to impose certain policies in agriculture. With this policy, it is hoped that national food security will be realized. One of the information that can be used to support the policy is the grouping of provinces in Indonesia based on peanut yields in Indonesia. This grouping is done because of the diverse potential of peanut production in each province, so it is necessary to do the grouping to find out which provinces have peanut production so that it can help to improve government programs in the field of food crop agriculture. The grouping uses k-means cluster analysis. In this study, the simulation uses the mobile-based K-Means Clustering method. The mapping process is based on 2 (two) variables, namely harvest area (Ha) and production (tons). The mapping results will be grouped into two (2) clusters, namely high and low yields. From manual calculations, mobile applications and simulations using Rapid Miner the same results are obtained, namely cluster 1 distributing provinces no 12 (West Java), 13 (Central Java) and 15 (East Java). While there are 20 data entered into cluster 0.

Keywords: *cluster k-means*, peanuts, *mobile*, agriculture

I. PENDAHULUAN

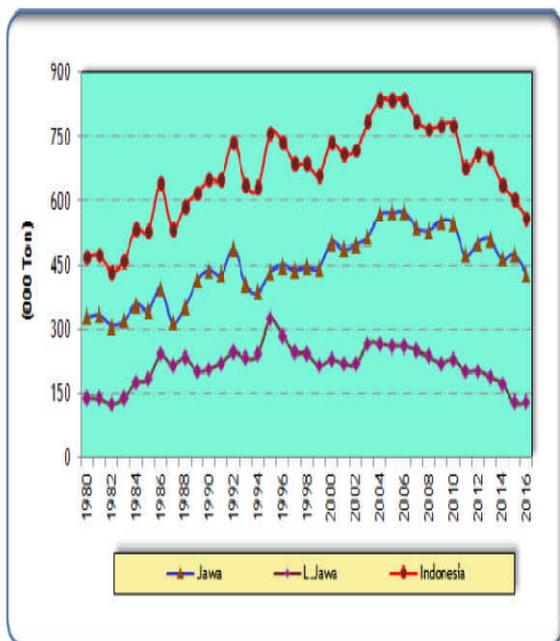
KACANG tanah merupakan salah satu sumber gizi untuk masyarakat dikarenakan mengandung sumber protein nabati yang tinggi. Produksi nasional kacang tanah selama tiga dekade terakhir menunjukkan pertumbuhan yang cukup positif. Akan tetapi produksi tersebut belum dapat memenuhi permintaan yang semakin meningkat, sehingga jumlah impor kacang tanah meningkat tajam. Negara Indonesia menjadi negara importir nomor kedua dunia yang mengimpor kacang tanah dengan rata-rata sebesar 137,17 ribu ton berdasarkan data FAO pada tahun 2009–2013. Perkembangan produktivitas kacang tanah nasional pada periode 1980–2016 mengalami peningkatan

(Gambar 1). Pertumbuhan produktivitas kacang tanah secara nasional lima tahun terakhir yaitu periode 2012–2016 naik 1,17% per tahun. Trend perkembangan produktivitas kacang tanah per wilayah sama, yaitu rata - rata 12 kuintal per hektar. Rata-rata produktivitas kacang tanah di Pulau Jawa mengalami peningkatan dibandingkan produktivitas di luar Pulau Jawa [1].



Gambar 1. Perkembangan Produktivitas Kacang Tanah Indonesia Tahun 1980–2016 [1]

Perkembangan produksi kacang tanah di Indonesia pada periode 2012–2016 fluktuasi dengan rata-rata pertumbuhan minus 3,58% per tahun (Gambar 2). Pada Gambar 2 dapat disimpulkan bahwa produksi kacang tanah baik di Jawa maupun di luar Jawa cenderung mengalami fluktuasi.



Gambar 2. Perkembangan Produksi Kacang Tanah Indonesia Tahun 1980–2016 [1]

Dengan kondisi produktivitas kacang tanah yang mengalami fluktuasi, maka pemerintah perlu memberlakukan kebijakan tertentu dibidang pertanian. Dengan kebijakan tersebut diharapkan dapat mewujudkan ketahanan pangan tingkat nasional. Salah satu informasi yang dapat digunakan untuk mendukung kebijakan tersebut adalah pengelompokan provinsi di Indonesia

berdasarkan hasil panen kacang tanah di Indonesia. Pengelompokan ini dilakukan karena beragamnya potensi hasil produksi kacang tanah di masing-masing provinsi, sehingga perlu dilakukan pengelompokan untuk mengetahui provinsi mana yang memiliki hasil produksi kacang tanah tertinggi sehingga dapat membantu untuk mengoptimalkan program-program pemerintah dibidang pertanian tanaman pangan.

Dalam beberapa tahun terakhir data pertanian dan lingkungan telah meningkat dalam tingkat eksponensial dengan penggunaan alat dan sistem pengumpulan data otomatis yang luas. Data hasil dari aplikasi pertanian presisi telah menjadi salah satu kontributor terakhir dalam peningkatan ini. Sejumlah besar data yang dikumpulkan oleh prakiraan cuaca, penginderaan jauh dan sistem informasi geografis telah digunakan sejak lama. Selain penggunaan progresif dan intensif dari jaringan sensor dan komputer di area budidaya, lumpung dan rumah unggas telah memainkan peran penting dalam peningkatan data pertanian. Karena pertumbuhan yang luar biasa ini, teknik data mining akan sangat membantu untuk menemukan informasi yang berguna atau berarti dalam data besar pertanian. Namun, data mining adalah bidang yang relatif baru di bidang pertanian, pangan, lingkungan dan bidang terkait lainnya.

Pengelompokan tersebut menggunakan analisis *cluster k-means*. *K-means* dilaporkan cepat, kuat, dan mudah diimplementasikan. Seperti yang dilaporkan dalam banyak studi itu memberikan hasil yang relatif baik jika kelompok dalam kumpulan data berbeda atau terpisah dengan baik. Itu juga diperiksa bahwa *K-Means* relatif efisien dalam kompleksitas waktu komputasi [2].

II. STUDI PUSTAKA

A. Penelitian Terkait

Guido (2018) telah melakukan penelitian yang mengimplementasikan pengelompokan palawija di Kalimantan menggunakan metode *K-Means Clustering*. Dataset tanaman palawija yang diproses adalah jagung, ubi jalar, ketela pohon, kacang tanah dan kedelai. Hasil dari pengelompokan ini jumlah kluster yang direkomendasikan untuk tiap dataset rata-rata pada $k=3$ dimana kluster ke-0 merupakan kelompok dengan rata-rata produktivitas terendah, kluster ke-1 sedang dan kluster ke-2 tertinggi [3].

Etik (2015) melakukan penelitian analisis pengelompokan daerah menggunakan metode *Non-Hierarchical Partitioning K-Medoids* dari hasil komoditas tanaman pertanian tanaman pangan se-Jawa Tengah. Hasil pengelompokan Kabupaten/Kota di Jawa Tengah dari tahun 2009-2013 berdasarkan hasil produksi padi menunjukkan bahwa secara umum Kabupaten/Kota di

Jawa Tengah dapat di kelompokkan menjadi 7 cluster. Hasil pengelompokan Kabupaten/Kota di Jawa Tengah dari tahun 2009-2013 berdasarkan hasil produksi palawija menghasilkan 2 cluster, dengan jumlah anggota cluster ke-1 selalu lebih banyak dari pada anggota cluster ke-2 [4].

Pada penelitian ini terdapat perbedaan dengan penelitian sebelumnya berupa cakupan yang akan diteliti tingkat nasional dan aplikasi yang dibangun berbasis mobile, selain itu dalam penelitian ini akan bertujuan untuk dijadikan sebagai bahan serta informasi bagi penyusunan kebijakan dan program pengembangan komoditas tanaman pangan khususnya kacang tanah di masa yang akan datang.

B. K-Means Clustering

K-Means adalah sebuah algoritma clustering partisi, metode ini adalah untuk mengklasifikasikan objek tanggal diberikan ke k cluster yang berbeda melalui iteratif, konvergen ke minimum lokal. Demikian hasil yang dihasilkan cluster kompak dan independen. Algoritma terdiri dari dua fase terpisah. Fase pertama memilih pusat k secara acak, di mana nilai k diperbaiki di muka. Fase selanjutnya adalah membawa setiap objek data ke pusat terdekat. Jarak Euclidean umumnya dianggap menentukan jarak antara setiap objek data dan pusat cluster. Kapan semua objek data termasuk dalam beberapa kelompok, langkah pertama adalah selesai dan pengelompokan awal dilakukan. Menghitung ulang rata-rata dari cluster yang terbentuk awal. Proses berulang ini terus berulang hingga fungsi kriteria menjadi minimum [5]. Misalkan objek target adalah x, xi menunjukkan rata-rata cluster Ci, fungsi kriteria didefinisikan sebagai berikut:

$$E = \sum_{i=1}^k \sum_{x \in C_i} |x - x_i|^2 \dots \dots \dots (1)$$

E adalah jumlah kesalahan kuadrat dari semua objek dalam database. Jarak fungsi kriteria adalah jarak Euclidean, yaitu digunakan untuk menentukan jarak terdekat antara setiap data objek dan pusat cluster. Jarak Euclidean antara satu vektor x = (x1, x2, ... xn) dan vektor lain y = (y1, y2, ... yn). Jarak Euclidean d (xi, yi) dapat diperoleh sebagai berikut:

$$d(x_i, y_i) = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2} \dots \dots \dots (2)$$

Proses dari algoritma k-means meliputi:

Masukan:

Jumlah dari cluster, k, dan sebuah database D={d1, d2,...dn} yang berisi data objek.

Luaran:

Satu set dari k cluster

Tahapan:

- 1) Memilih secara random k object data dari dataset D sebagai pusat cluster.

- 2) Mengulangi tahap 1;
- 3) Menghitung jarak antara setiap objek data di (1 <= i <= n) dan semua pusat cluster k cj (1 <= j <= k) dan menetapkan data objek di ke cluster terdekat.
- 4) Untuk setiap cluster j (1 <= j <= k), menghitung ulang cluster pusat.
- 5) Sampai tidak ada perubahan di pusat cluster [5].

III. METODE PENELITIAN

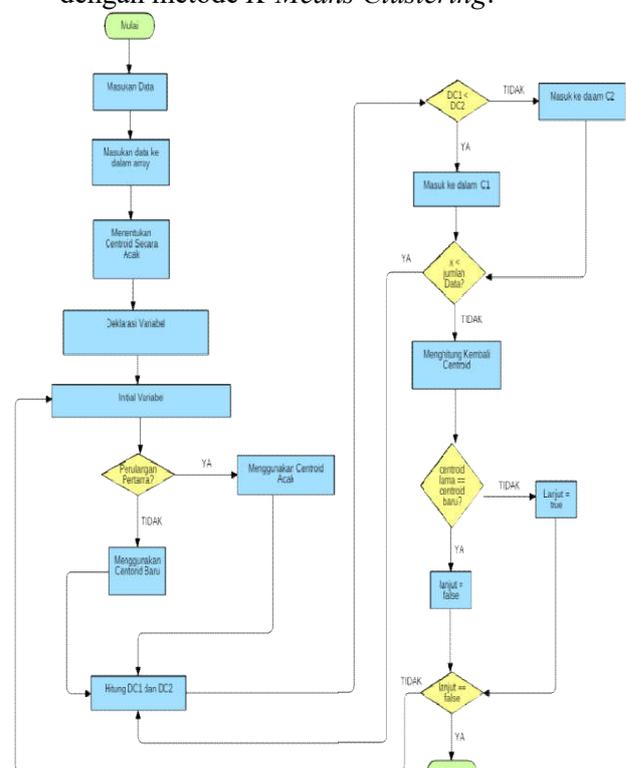
Metode pengembangan sistem yang digunakan dalam penelitian ini adalah metode RUP (Rational Unified Process). Tahapan-tahapan yang terdapat dalam metode RUP adalah :

a. Inception

Dalam penelitian ini simulasi pemetaan menggunakan metode K-Means Clustering berbasis mobile. Proses pemetaan dilakukan berdasarkan 2 (dua) variabel yaitu luas panen (Ha) dan produksi (ton). Hasil pemetaan akan dikelompokkan menjadi dua (2) cluster, yaitu hasil panen tinggi dan rendah. Pada penelitian ini menggunakan data publik dengan alamat url: <http://data.go.id/dataset/tanaman-kacang-tanah-per-provinsi>.

b. Elaboration

Tahap ini lebih difokuskan pada perancangan flowchart aplikasi dan perhitungan manual. Berikut ini adalah flowchart aplikasi pemetaan dengan metode K-Means Clustering:



Gambar 3. Perancangan Flowchart Aplikasi dengan K-Means Clustering

c. Construction

Tahap ini lebih pada implementasi sistem yang fokus pada implementasi perangkat lunak pada kode program dan menghasilkan produk perangkat lunak berbasis *mobile*.

d. *Transition*

Tahap ini lebih pada *deployment* atau instalasi agar dapat dimengerti oleh user dan aktifitas yang dilakukan.

Iterasi ke-2

Tabel 3. Penentuan pusat *cluster* iterasi ke-2

C	Luas Panen	Produksi
1	73323,16667	75357,6667
0	6410,647059	6584,94118

IV. HASIL DAN PEMBAHASAN

A. Pengolahan Data

Berikut ini adalah proses perhitungan manual pemetaan tanaman kacang tanah di Indonesia menggunakan *K-Means Clustering* dalam dua (2) *cluster*, yaitu hasil panen tinggi (C1) dan rendah (C0):

Tabel 1. Penentuan pusat *cluster* inialisasi awal

C	Luas Panen	Produksi
1	8770	8594
0	3656	3799

Tabel 2. Penentuan *cluster* inialisasi awal

No	Nama Provinsi	Luas Panen	Produksi	C1	C0
1	Nanggroe Aceh Darussalam	16756	19056	√	
2	Sumatera Utara	24607	24801	√	
3	Sumatera Barat	7229	6894	√	
4	Riau	3656	3799		√
5	Jambi	3131	3405		√
6	Sumatera Selatan	18199	18854	√	
7	Bengkulu	8770	8594	√	
8	Lampung	10820	12123	√	
9	Kepulauan Bangka Belitung	0	0		√
10	Kepulauan Riau	0	0		√
11	DKI Jakarta	27	28		√
12	Jawa Barat	89306	97780	√	
13	Jawa Tengah	113920	11769	√	
14	Prov. D I Yogyakarta	36106	30762	√	
15	Jawa Timur	156080	15895	√	
16	Banten	0	0		√
17	Bali	11870	13936	√	
18	Nusa Tenggara Barat	19920	22153	√	
19	Nusa Tenggara Timur	7424	6284	√	
20	Kalimantan Barat	2388	1865		√
21	Kalimantan Tengah	994	1105		√
22	Kalimantan Selatan	14311	12671	√	
23	Kalimantan Timur	3406	3330		√

Tabel 4. Penentuan *cluster* iterasi ke-2

No	Nama Provinsi	d_C1	d_C2	C1	C0
1	Nanggroe Aceh Darussalam	79810,53	16203,50		√
2	Sumatera Utara	70208,55	25747,46		√
3	Sumatera Barat	95161,50	874,7679		√
4	Riau	99870,70	3917,850		√
5	Jambi	100519,2	4568,162		√
6	Sumatera Selatan	78938,82	17014,55		√
7	Bengkulu	92868,17	3098,848		√
8	Lampung	88911,57	7079,017		√
9	Kepulauan Bangka Belitung	105143,0	9190,094		√
10	Kepulauan Riau	105143,0	9190,094		√
11	DKI Jakarta	105104,1	9151,198		√
12	Jawa Barat	27535,64	123240,3		√
13	Jawa Tengah	58654,05	154606,0		√
14	Prov. D I Yogyakarta	58085,20	38292,87		√
15	Jawa Timur	117633,7	213584,7		√
16	Banten	105143,0	9190,094		√
17	Bali	86885,63	9156,560		√
18	Nusa Tenggara Barat	75383,25	20612,30		√
19	Nusa Tenggara Timur	95466,59	1057,094		√
20	Kalimantan Barat	102141,9	6201,575		√
21	Kalimantan Tengah	103657,9	7705,181		√
22	Kalimantan Selatan	86093,28	9972,747		√
23	Kalimantan Timur	100381,2	4429,734		√

Iterasi ke-3

Tabel 5. Penentuan pusat *cluster* iterasi ke-3

C	Luas Panen	Produksi
1	119768,6667	124810
0	9480,7	9483

Tabel 6. Penentuan *cluster* iterasi ke-3

No	Nama Provinsi	d_C1	d_C2	C1	C0
1	Nanggroe Aceh Darussalam	147633,0519	12023,82298		√
2	Sumatera Utara	138049,0597	21527,79774		√
3	Sumatera Barat	163001,1032	3431,191322		√
4	Riau	167707,5236	8138,487949		√
5	Jambi	168355,3364	8789,810811		√
6	Sumatera Selatan	146775,5808	12799,39045		√
7	Bengkulu	160707,3821	1138,163209		√
8	Lampung	156742,3744	2960,291285		√
9	Kepulauan Bangka Belitung	172979,9688	13409,36097		√
10	Kepulauan Riau	172979,9688	13409,36097		√
11	DKI Jakarta	172941,0716	13370,47003		√
12	Jawa Barat	40725,85126	119031,2511	√	
13	Jawa Tengah	9212,644885	150388,68	√	
14	Prov. D I Yogyakarta	125874,8112	34083,75626		√
15	Jawa Timur	49845,7504	209366,0201	√	
16	Banten	172979,9688	13409,36097		√
17	Bali	154709,942	5053,510017		√
18	Nusa Tenggara Barat	143206,8989	16416,6953		√
19	Nusa Tenggara Timur	163308,7163	3803,106085		√
20	Kalimantan Barat	169981,4517	10408,66549		√
21	Kalimantan Tengah	171494,4561	11925,39144		√
22	Kalimantan Selatan	153936,5934	5787,498777		√
23	Kalimantan Timur	168219,0851	8646,466856		√

Tabel 7. Penentuan pusat *cluster* iterasi ke-4

C	Luas Panen	Produksi
1	119768,6667	124810
0	9480,7	9483

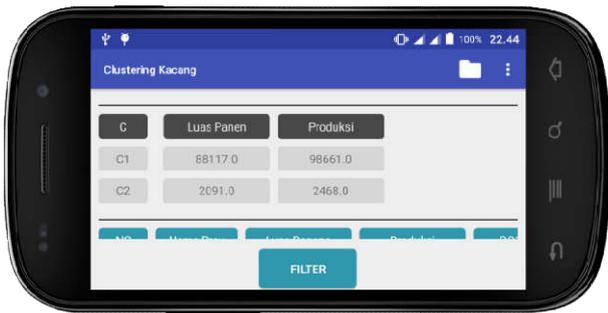
Tabel 8. Penentuan Cluster iterasi ke - 4

No	Nama Provinsi	d_C1	d_C0	C1	C0
1	Nanggroe Aceh Darussalam	147633,0519	12023,82298		√
2	Sumatera Utara	138049,0597	21527,79774		√
3	Sumatera Barat	163001,1032	3431,191322		√
4	Riau	167707,5236	8138,487949		√
5	Jambi	168355,3364	8789,810811		√
6	Sumatera Selatan	146775,5808	12799,39045		√
7	Bengkulu	160707,3821	1138,163209		√
8	Lampung	156742,3744	2960,291285		√
9	Kepulauan Bangka Belitung	172979,9688	13409,36097		√
10	Kepulauan Riau	172979,9688	13409,36097		√
11	DKI Jakarta	172941,0716	13370,47003		√
12	Jawa Barat	40725,85126	119031,2511	√	
13	Jawa Tengah	9212,644885	150388,68	√	
14	Prov. D I Yogyakarta	125874,8112	34083,75626		√
15	Jawa Timur	49845,7504	209366,0201	√	
16	Banten	172979,9688	13409,36097		√
17	Bali	154709,942	5053,510017		√
18	Nusa Tenggara Barat	143206,8989	16416,6953		√
19	Nusa Tenggara Timur	163308,7163	3803,106085		√
20	Kalimantan Barat	169981,4517	10408,66549		√
21	Kalimantan Tengah	171494,4561	11925,39144		√
22	Kalimantan Selatan	153936,5934	5787,498777		√
23	Kalimantan Timur	168219,0851	8646,466856		√

Dari perhitungan manual diperoleh hasil *cluster* 1 yaitu provinsi no 12 (Jawa Barat), 13 (Jawa Tengah) dan 15 (Jawa Timur). Sedangkan sisanya masuk ke *cluster* 0.

B. Implementasi Program

1) Tampilan Pemilihan Pusat *Cluster*



Gambar 4. Tampilan Pemilihan Pusat *Cluster* Secara Acak

Hal pertama yang dilakukan oleh sistem adalah pemilihan pusat *cluster*, dimana pada aplikasi ini telah dilakukan pemilihan secara acak pusat *cluster* yang akan digunakan. Oleh karena pemilihan pusat *cluster* dilakukan secara acak, jadi memungkinkan jumlah perulangan berbeda-beda pada setiap pembukaan sebuah file.

2) Tampilan Menu Filter

Menu filter digunakan untuk mengelompokkan data yang telah diproses. Berikut tampilan setelah ditekan tombol filter :



Gambar 5. Tampilan Hasil Kelompok *Cluster* 1

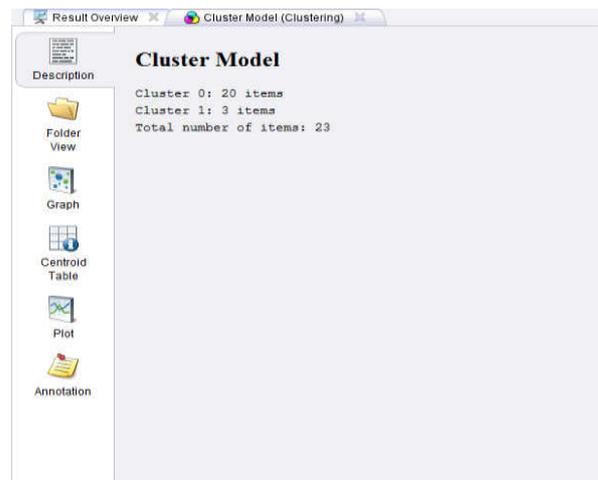


Gambar 6. Tampilan Hasil Kelompok *Cluster* 2

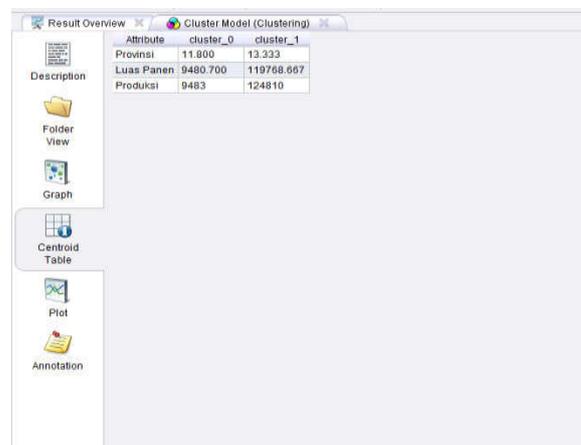
Data yang telah diproses dibagi menjadi dua buah *cluster*, dan hasil yang diperoleh sudah sesuai dengan hasil perhitungan manual yang telah dilakukan.

C. Pengujian Sistem

Pengujian sistem pada penelitian ini dengan membandingkan perhitungan manual, hasil aplikasi dengan hasil simulasi menggunakan Rapid Miner. Gambar 7 menunjukkan *cluster* model, yaitu *cluster* 1 sebanyak 3 items dan *cluster* 0 sebanyak 20 items.

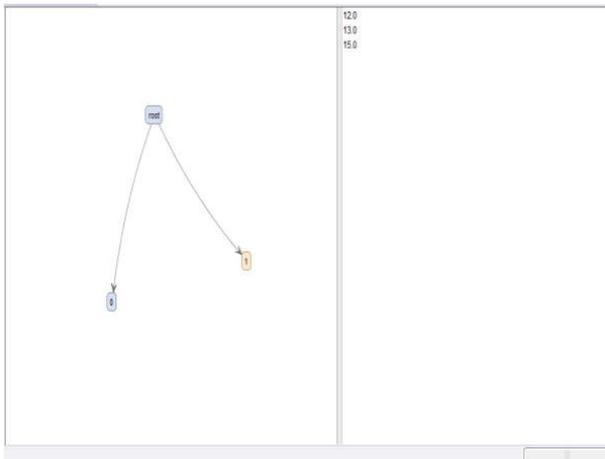


Gambar 7. *Cluster* Model Hasil Simulasi Rapid Miner



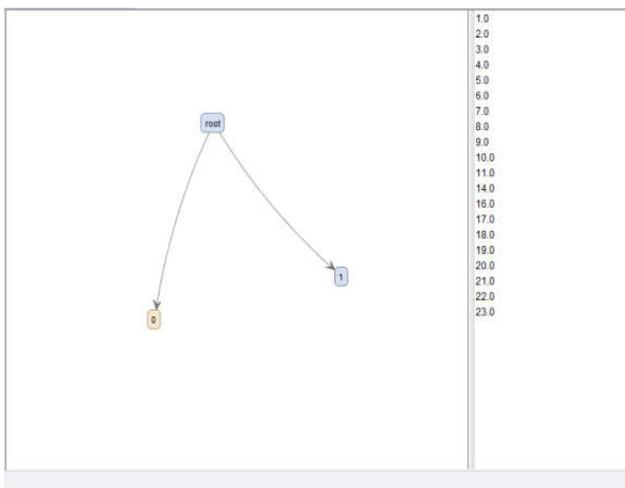
Gambar 8. Pusat *Cluster* Iterasi Terakhir Hasil Simulasi Rapid Miner

Pada gambar 8 ditunjukkan pusat *cluster* 0 dan *cluster* 1. Untuk pusat *cluster* 0 luas panen 9480,7 dan produksi 9483. Sedangkan untuk pusat *cluster* 1 yaitu luas panen 119768,667 dan produksi sebesar 124810. Hasil tersebut menunjukkan besar yang sama dengan perhitungan manual.



Gambar 9. Cluster 1 Hasil Simulasi Rapid Miner

Dari simulasi diperoleh hasil *cluster* 1 yaitu provinsi no 12 (Jawa Barat), 13 (Jawa Tengah) dan 15 (Jawa Timur) terlihat pada gambar 9. Sedangkan sisanya masuk ke *cluster* 0 yaitu provinsi no 1–11, 14, 16–23 yang ditunjukkan pada gambar 10.



Gambar 10. Cluster 1 Hasil Simulasi Rapid Miner

V. KESIMPULAN

Dalam penelitian ini simulasi pemetaan menggunakan metode *K-Means Clustering* berbasis *mobile*. Proses pemetaan dilakukan berdasarkan 2 (dua) variabel yaitu luas panen (Ha) dan produksi (ton). Hasil pemetaan akan dikelompokkan menjadi dua (2) *cluster*, yaitu hasil panen tinggi dan rendah. Dari perhitungan manual, aplikasi *mobile* dan simulasi menggunakan Rapid Miner diperoleh hasil yang sama yaitu *cluster* 1 meliputi provinsi no 12 (Jawa Barat), 13 (Jawa Tengah) dan 15 (Jawa Timur). Sedangkan sisanya ada 20 data masuk ke *cluster* 0.

DAFTAR PUSTAKA

- [1] Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian, “Keragaan Kacang Tanah Nasional,” dalam *Outlook Komoditas Pertanian Sub Sektor Tanaman Pangan Kacang Tanah*, Jakarta : Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian, 2016, hal.7-10.
- [2] Berkhin P, “*A Survey of Clustering Data Mining Techniques*,” In: Kogan J., Nicholas C., Teboulle M. (eds) *Grouping Multidimensional Data*. Springer, Berlin, Heidelberg, 2006.
- [3] Guido, “Peta Potensi Palawija di Kalimantan Menggunakan Metode *K-Means Clustering*,” Program Studi Teknik Informatika Jurusan Teknik Informatika. Fakultas Sains dan Teknologi Universitas Sanata Dharma Yogyakarta, 2018.
- [4] Etik Setyowati, Agus Rusgiyono, Moch. Abdul Mukid, “Analisis Pengelompokan Daerah Menggunakan Metode *Non-Hierarchical Partitioning K-Medoids* Dari Hasil Komoditas Paertanian Tanaman Pangan,” *Jurnal Gaussian*, Vol 4, No 4, Tahun 2015 Halaman 825-836.
- [5] Shi Na, Guan Yong and Liu Xumin, "*Research on k-means Clustering Algorithm*", *Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010.