

## KLASIFIKASI DATA MINING PADA RUMAH TANGGA MENURUT PROVINSI DAN STATUS KEPEMILIKAN RUMAH KONTRAK/SEWA MENGUNAKAN *K-MEANS* CLUSTERING METHOD

Erene Gernaria Sihombing

*Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri)  
Jl. Damai Warung Jati No.8, Jakarta Selatan, Indonesia*

*erene.egs@nusamandiri.ac.id*

Page | 74

**Abstrak** — Perumahan adalah kelompok rumah yang berfungsi sebagai lingkungan tempat tinggal atau hunian yang dilengkapi dengan prasarana lingkungan yaitu kelengkapan dasar fisik lingkungan, misalnya penyediaan air minum, pembuangan sampah, tersedianya listrik, telepon, jalan, yang memungkinkan lingkungan pemukiman berfungsi sebagaimana mestinya. Penelitian ini membahas tentang pengelompokan pertumbuhan rumah tangga menurut provinsi dan status kepemilikan rumah kontrak/sewa yang dilihat dari persentase. Metode yang digunakan adalah *Datamining Clustering K-Means*. *Clustering* adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengklompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh. Pengumpulan data dari rumah tangga terpilih dilakukan melalui wawancara tatap muka antara pencacah dengan responden. Untuk pertanyaan-pertanyaan dalam kuesioner Susenas yang ditujukan kepada individu diusahakan agar individu yang bersangkutan yang menjadi responden. Sedangkan pertanyaan-pertanyaan dalam kuesioner yang ditujukan kepada rumah tangga dikumpulkan melalui wawancara dengan kepala rumah tangga, suami/isteri kepala rumah tangga atau anggota rumah tangga lain yang mengetahui tentang karakteristik yang ditanyakan. Proses Survey Sosial Ekonomi Nasional (Susenas) dilakukan dalam beberapa tahap mulai dari susenas 2004-2009. Semua kegiatan ini terekam di situs Badan Pusat Statistik (BPS) dengan alamat url <https://www.bps.go.id/>. Data yang digunakan dalam penelitian ini adalah data Persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dari tahun 1999-2016 yang terdiri dari 34 provinsi. Variabale yang digunakan adalah jumlah rata-rata persentasi Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa. Data akan diolah dengan melakukan clustering yang dibagi dalam 3 cluster yaitu cluster status kepemilikan rimah kontrak dengan tingkatan tinggi cluster status kepemilikan rimah kontrak dengan tingkatan sedang dan cluster status kepemilikan rimah kontrak dengan tingkatan rendah. Metode clustering yang digunakan dalam penelitian ini adalah metode K-Means. Centroid data untuk cluster tingkatan tinggi 28,7986, Centroid data untuk cluster tingkatan sedang 13,0389 serta Centroid data untuk cluster tingkatan rendah 5,2515. Proses iterasi berlangsung 4 kali sehingga diperoleh penilaian persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dengan 2 provinsi cluster tingkatan tinggi yakni Kepulauan Riau dan DKI Jakarta, 12 provinsi cluster tingkatan sedang yakni Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Jawa Tengah, Banten, Bali, Kalimantan Selatan, Kalimantan Timur, Kalimantan Tengah, Papua Barat dan Papua dan 20 provinsi cluster tingkatan rendah yakni Aceh, Jambi, Sumatera Selatan, Lampung, Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, NTB, NTT, Kalimantan Barat, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo dan Sulawesi Barat . Data ini dapat menjadi masukan bagi pemerintah untuk memberikan kebijakan kepada provinsi tentang status kepemilikan rumah kontrak/sewa menjadi milik sendiri dengan mempertimbangkan segala aspek yang ada.

**Kata kunci** — Data Mining, K-Means, Klastering, Perumahan, Kontrakan/Sewa

### I. PENDAHULUAN

Perumahan adalah kelompok rumah yang berfungsi sebagai lingkungan tempat tinggal atau hunian yang dilengkapi dengan prasarana lingkungan yaitu kelengkapan dasar fisik lingkungan, misalnya penyediaan air minum, pembuangan sampah, tersedianya listrik, telepon, jalan, yang memungkinkan lingkungan pemukiman berfungsi

sebagaimana mestinya. menurut uu no. 4 tahun 1992 tentang perumahan dan permukiman, rumah adalah bangunan yang berfungsi sebagai tempat tinggal atau hunian dan sarana pembinaan keluarga. Menurut John F.C Turner, 1972, dalam bukunya *freedom to build* mengatakan, “Rumah adalah bagian yang utuh dari permukiman, dan bukan hasil fisik sekali jadi semata, melainkan merupakan suatu proses yang terus berkembang dan terkait dengan mobilitas sosial

ekonomi penghuninya dalam suatu kurun waktu. yang terpenting dan rumah adalah dampak terhadap penghuni, bukan wujud atau standar fisiknya. selanjutnya dikatakan bahwa interaksi antara rumah dan penghuni adalah apa yang diberikan rumah kepada penghuni serta apa yang dilakukan penghuni terhadap rumah". Perumahan memiliki status kepemilikan atau Status Penguasaan Tempat Tinggal yakni (1) Milik sendiri, jika tempat tinggal tersebut pada waktu pencacahan betul-betul sudah milik krt atau salah satu seorang art. Rumah yang dibeli secara angsuran melalui kredit bank atau rumah dengan statussewa beli dianggap sebagai rumah milik sendiri, (2) Kontrak, jika tempat tinggal tersebut disewa oleh krt/art dalam jangka waktu tertentu berdasarkan perjanjian kontrak antara pemilik dan pemakai, misalnya 1 atau 2 tahun. Cara pembayarannya biasanya sekaligus di muka atau dapat diangsur menurut persetujuan kedua belah pihak, (3) Sewa, jika tempat tinggal tersebut disewa oleh krt atau salah seorang art dengan pembayaran sewanya secara teratur dan terus menerus tanpa batasan waktu tertentu, (4) Bebas sewa milik orang lain, jika tempat tinggal tersebut diperoleh dari pihak lain (bukan famili/orang tua) dan ditempati/didiami oleh rt tanpa mengeluarkan suatu pembayaran apapun, (5) Rumah milik orang tua/sanak/saudara, jika tempat tinggal tersebut bukan milik sendiri melainkan milik orang tua/sanak/saudara, dan tidak mengeluarkan suatu pembayaran apapun untuk mendiami tempat tinggal tersebut, (6) Rumah dinas, jika tempat tinggal tersebut diperoleh dari pihak lain (bukan famili/orang tua) dan ditempati/didiami oleh RT tanpa mengeluarkan suatu pembayaran apapun dan (7) Lainnya, jika tempat tinggal tersebut tidak dapat digolongkan ke dalam salah satu kategori di atas, misalnya tempat tinggal milik bersama, rumah adat. Undang-undang Republik Indonesia nomor 1 tahun 2011 tentang perumahan dan kawasan permukiman dalam point d menyebutkan bahwa "Pertumbuhan dan pembangunan wilayah yang kurang memperhatikan keseimbangan bagi kepentingan masyarakat berpenghasilan rendah mengakibatkan kesulitan masyarakat untuk memperoleh rumah yang layak dan terjangkau". Bangsa Indonesia adalah bangsa majemuk. Data sensus menghitung 236.728.379 warga negara Indonesia serta 73.217 warga negara asing yang berada di Indonesia dan 839.730 tidak diketahui keberadaannya dengan jumlah provinsi sebanyak 34. Dengan data jumlah penduduk tersebut, pemerintah harus dapat melakukan pembangunan secara merata guna tercapainya tujuan pembangunan nasional. Penelitian ini membahas tentang pengelompokan pertumbuhan rumah tangga menurut provinsi dan status kepemilikan rumah kontrak/sewa yang dilihat dari persentase. Metode yang digunakan adalah Datamining *Clustering K-Means*. *Clustering* adalah metode yang digunakan dalam data mining yang cara

kerjanya mencari dan menglompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh. Ciri khas dari teknik data mining ini adalah mempunyai sifat tanpa arahan (*unsupervised*), yang dimaksud adalah teknik ini diterapkan tanpa perlunya data *training* dan tanpa ada *teacher* serta tidak memerlukan target *output*[1]. Metode *clustering* yang mempunyai sifat efisien dan cepat yang dapat digunakan salah satunya adalah metode k-means, metode ini bertujuan untuk membuat *cluster* objek berdasarkan atribut menjadi *k* partisi.cara kerja metode ini adalah mula – mula ditentukan *cluster* yang akan dibentuk, pada elemen pertama dalam tiap *cluster* dapat dipilih untuk dijadikan sebagai titik tengah (*centroid*), selanjutnya akan dilakukan pengulangan langkah – langkah hingga tidak ada objek yang dapat dipindahkan lagi[2].

Berdasarkan permasalahan yang telah dipaparkan di atas maka penelitian ini akan menerapkan metode k-means untuk menghasilkan informasi mengenai pengelompokan pertumbuhan rumah tangga menurut provinsi dan status kepemilikan rumah kontrak/sewa. Dari penelitian ini akan didapatkan hasil daerah propinsi mana yang memiliki status kepemilikan rumah kontrak/sewa yang paling tinggi. Ini dapat menjadi masukan bagi pemerintah untuk memberikan kebijakan kepada provinsi tentang status kepemilikan rumah kontrak/sewa menjadi milik sendiri dengan mempertimbangkan segala aspek yang ada.

## II. LANDASAN TEORI

### A. Data Mining

Data mining yang juga dikenal dengan istilah *pattern recognition* merupakan suatu metode yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang diolah. Data yang diolah dengan teknik data mining ini kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan di masa depan[3].

Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Dalam data mining juga terdapat metode – metode yang dapat digunakan seperti klasifikasi, *clustering*, regresi, seleksi variabel, dan market basket analisis[3].

Data mining juga bisa diartikan sebagai rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data – data tersebut dapat disimpan dalam database, data warehouse atau penyimpanan informasi. Ada beberapa ilmu yang mendukung teknik data mining

diantaranya adalah data analisis, *signal processing*, *neural network* dan pengenalan pola[4].

### B. Clustering

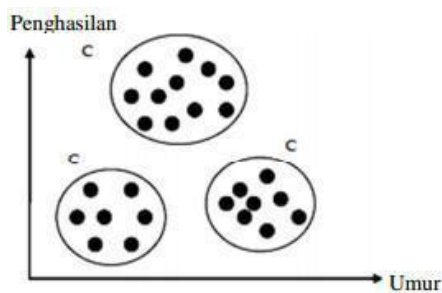
*Clustering* atau pengklasteran adalah suatu teknik data mining yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam mengelompokkan data atau lebih tepatnya mempartisi dari dataset ke dalam subset. Pada teknik *clustering* targetnya adalah untuk kasus pendistribusian (objek, orang, peristiwa dan lainnya) ke dalam suatu kelompok, hingga derajat tingkat keterhubungan antar anggota *cluster* yang sama adalah kuat dan lemah antara anggota *cluster* yang berbeda[5].

Teknik *cluster* mempunyai dua metode dalam pengelompokkannya yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *hierarchical clustering* merupakan suatu metode pengelompokkan data yang cara kerjanya dengan mengelompokkan dua data atau lebih yang mempunyai kesamaan atau kemiripan, kemudian proses dilanjutkan ke objek lain yang memiliki kedekatan dua, proses ini terus berlangsung hingga *cluster* membentuk semacam *tree* dimana ada hirarki atau tingkatan yang jelas antar objek dari yang paling mirip hingga yang paling tidak mirip. Namun secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*[4].

Pada proses analisis *cluster* metode yang digunakan untuk membagi data menjadi subset data berdasarkan kesamaan atau kemiripan yang telah ditentukan sebelumnya. Jadi analisis *cluster* secara umum dapat dikatakan bahwa [5]:

- a. Data yang terdapat dalam satu *cluster* memiliki tingkat kesamaan yang tinggi, dan
- b. Dan yang terdapat dalam suatu *cluster* yang berbeda memiliki tingkat kesamaan yang rendah

Sebagai contoh dapat dilihat pada gambar 1 dibawah ini :



Gambar 1 Grafik Clustering

Pada gambar 1 dapat dilihat kita misalkan data tersebut merupakan data konsumen sederhana yang terdapat dua atribut didalamnya, yaitu umur dan

penghasilan. Pada data yang berdasarkan dua atribut tersebut kemudian dibagi menjadi tiga *cluster* yaitu *cluster* C1 yang terdiri dari konsumen usia muda dan berpenghasilan rendah, *cluster* C2 terdiri dari konsumen usia muda dan tua berpenghasilan tinggi, dan *cluster* C3 terdiri dari konsumen usia tua dan berpenghasilan relatif rendah.

### C. K-Means

K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – berda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses[5].

Sehingga dalam penggunaan algoritma k-means terdapat aturan sebagai berikut [5] :

- a. Berapa jumlah *cluster* yang perlu dimasukkan
- b. Hanya memiliki atribut bertipe numeric

Pada dasarnya algoritma k-means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat *cluster* awal, pada penentuan pusat *cluster* ini dipilih secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing – masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat *cluster*. Selanjutnya posisi pusat *cluster* akan dihitung kembali samapi semua komponen data digolongkan ke dalam tiap – tiap *cluster* dan terakhir akan terbentuk *cluster* baru[5].

Algoritma K-Means pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat *cluster* dan proses pencarian anggota dari tiap-tiap *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan dikluster,  $X_{ij}$  ( $i=1, \dots, n; j=1, \dots, m$ ) dengan  $n$  adalah jumlah data yang akan dikluster dan  $m$  adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara bebas (sembarang),  $C_{kj}$  ( $k=1, \dots, k; j=1, \dots, m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean. Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak data tersebut ke pusat *cluster*- $k$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lain[5].

Proses dasar algoritma k-means dapat dilihat di bawah ini :

1. Tentukan jumlah kluster yang ingin dibentuk dan tetapkan pusat cluster  $k$ .
2. Menggunakan jarak *euclidean* kemudian hitung setiap data ke pusat cluster.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots(2.1)$$

3. Kelompokkan data ke dalam cluster dengan jarak yang paling pendek dengan persamaan

$$\text{Min} \sum_k^k = d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots(2.2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \dots\dots\dots(2.3)$$

Dengan :

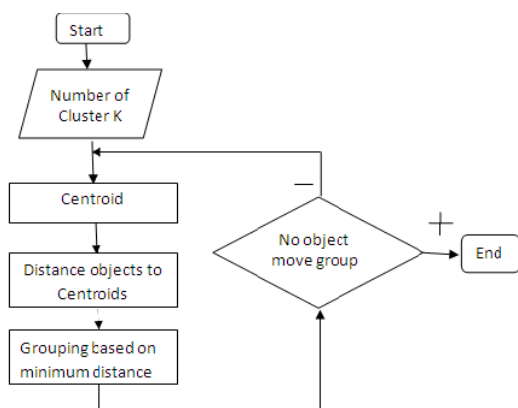
$x_{ij} \in$  Kluster ke -  $k$

$p$  = banyaknya anggota kluster ke -  $k$

5. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

#### D. Kerangka Pemikiran

Kerangka pemikiran merupakan garis besar dari langkah – langkah penelitian yang sedang dilakukan, kerangka pemikiran dijadikan acuan untuk melakukan tahap – tahap yang sedang dilakukan dalam penelitian.



Gambar 2. Kerangka Kerja K-Means

### III. METODE PENELITIAN

#### A. Tahap Pengumpulan Data

Pengumpulan data dari rumah tangga terpilih dilakukan melalui wawancara tatap muka antara pencacah dengan responden. Untuk pertanyaan-pertanyaan dalam kuesioner Susenas yang ditujukan kepada individu diusahakan agar individu yang bersangkutan yang menjadi responden. Sedangkan pertanyaan-pertanyaan dalam kuesioner yang

ditujukan kepada rumah tangga dikumpulkan melalui wawancara dengan kepala rumah tangga, suami/isteri kepala rumah tangga atau anggota rumah tangga lain yang mengetahui tentang karakteristik yang ditanyakan. Proses Survey Sosial Ekonomi Nasional (Susenas) dilakukan dalam beberapa tahap mulai dari sesunas 2004-2009. Semua kegiatan ini terekam di situs Badan Pusat Statistik (BPS) dengan alamat url <https://www.bps.go.id/>. Data yang digunakan dalam penelitian ini adalah data Persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dari tahun 1999-2016 yang terdiri dari 34 provinsi. Variabel yang digunakan adalah jumlah rata-rata persentasi Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa. Data akan diolah dengan melakukan clustering yang dibagi dalam 3 cluster yaitu cluster status kepemilikan rumah kontrak dengan tingkatan tinggi cluster status kepemilikan rumah kontrak dengan tingkatan sedang dan cluster status kepemilikan rumah kontrak dengan tingkatan rendah. Metode clustering yang digunakan dalam penelitian ini adalah metode K-Means

#### B. Tahap Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu untuk dapat diclustering. Dalam tahap sebelumnya, data Persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dari tahun 1999-2016 yang terdiri dari 34 provinsi adalah dengan mengambil nilai rata-rata setiap provinsi sehingga pada tahapan ini sudah diperoleh perhitungan nilai yang akan diproses pada tahap clustering.

#### C. Tahap Analisis

Pada tahapan ini dilakukan analisis data Persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dengan aplikasi tool. Rapid Miner. RapidMiner adalah sebuah lingkungan *machine learning data mining*, *text mining* dan *predictive analytics* [6]. Data yang diperoleh diolah dengan menggunakan perhitungan bobot dari tiap indeks. Pada tahapan sebelumnya, telah ditentukan akan dicluster ke dalam 3 cluster yakni cluster tingkatan tinggi, cluster tingkatan sedang dan cluster tingkatan rendah. Pada tahapan inilah akan dianalisis hasilnya.

TABEL 1  
PERSENTASE RUMAH TANGGA MENURUT PROVINSI DAN STATUS KEPEMILIKAN RUMAH KONTRAK, 1999-2016

Provinsi	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2011*	2012	2012*	2013	2013*	2014	2015	2015*	2016	2016*	Total
Aceh	6,71	-	-	29,35	5,85	6,51	7,06	6,91	6,96	6,82	6,57	7,34	6,37	6,12	5,91	5,84	6,01	5,92	5,76	5,52	6,26	7,55	6,27	7,55
Sumatera Utara	11,52	11,88	12,16	11,30	11,00	12,30	11,69	12,54	12,62	13,65	13,55	13,98	12,50	12,38	11,67	11,83	11,59	11,28	11,48	11,76	12,75	12,14	12,14	12,14
Sumatera Barat	12,53	11,26	10,08	10,32	10,27	9,00	10,41	10,55	11,46	11,44	11,37	10,85	9,41	9,83	10,52	10,45	9,67	9,53	9,64	9,58	10,55	10,42	10,42	10,42
Riau	15,41	15,42	16,15	15,96	15,58	15,36	14,09	14,64	15,25	19,22	19,17	18,52	15,48	14,99	14,66	14,00	13,86	13,34	14,72	14,37	15,39	15,41	15,41	15,41
Jambi	9,26	8,40	9,34	8,41	7,53	7,57	9,55	8,54	8,82	9,24	9,54	8,66	7,14	6,83	7,50	7,05	6,84	6,55	5,96	5,91	6,29	7,00	7,00	7,00
Sumatera Selatan	9,63	8,50	9,38	7,92	7,41	7,26	7,57	7,51	8,76	9,05	8,08	9,08	7,91	7,30	6,38	6,01	6,67	6,42	6,31	6,36	6,18	7,60	7,60	7,60
Bengkulu	14,90	10,94	9,04	8,69	9,22	8,32	8,51	9,51	9,94	10,75	10,24	9,39	7,92	7,54	8,40	7,87	6,52	6,83	7,86	5,48	6,86	8,82	8,82	8,82
Lampung	4,30	4,95	3,21	3,68	3,81	3,32	4,54	4,17	4,20	4,41	5,10	4,44	4,53	4,43	4,47	4,19	3,07	3,00	3,06	2,87	3,21	3,95	3,95	3,95
Kepulauan Bangka Belitung	-	-	6,34	3,94	4,00	5,03	6,86	6,34	6,33	7,05	6,85	6,73	9,24	8,73	5,92	5,44	6,07	5,69	6,52	4,75	5,47	6,27	6,27	6,27
Kepulauan Riau	-	-	-	-	-	-	20,07	19,92	26,14	17,95	17,41	24,30	27,74	26,27	28,21	24,48	28,65	26,51	24,18	25,89	24,44	24,01	24,01	24,01
DKI Jakarta	26,91	30,15	34,92	30,61	31,45	31,95	34,35	34,04	35,73	34,48	36,12	40,61	35,02	33,34	31,33	34,31	33,30	33,71	34,13	35,76	35,59	35,59	35,59	35,59
Jawa Barat	7,25	7,50	6,87	5,97	5,65	6,01	6,28	6,46	6,97	9,32	8,68	11,45	8,60	8,27	8,21	7,81	8,24	8,03	8,44	8,54	8,71	7,77	7,77	7,77
Jawa Tengah	2,86	2,86	2,86	2,73	2,76	2,62	2,77	2,32	2,47	2,96	2,84	3,40	2,74	2,63	2,74	2,63	2,56	2,46	2,96	2,29	2,36	2,71	2,71	2,71
DI Yogyakarta	17,45	14,72	18,48	15,00	17,16	17,54	17,92	19,68	16,09	16,32	13,65	16,94	13,98	14,98	14,01	13,87	14,21	13,90	15,48	13,58	14,07	15,67	15,67	15,67
Jawa Timur	6,17	6,30	5,86	5,56	5,24	5,09	5,24	5,00	5,16	5,43	4,99	5,69	5,07	4,97	4,47	4,31	4,85	4,69	4,91	4,55	4,32	5,14	5,14	5,14
Banten	-	-	14,45	10,74	10,96	10,20	10,58	10,53	13,90	15,28	13,08	17,68	14,88	14,90	13,94	13,31	15,02	14,57	13,52	13,17	13,13	13,47	13,47	13,47
Bali	10,30	12,03	10,71	11,83	9,93	11,24	10,59	13,43	12,81	13,82	14,12	17,74	17,11	17,52	16,51	15,69	17,77	17,22	19,66	17,10	19,69	14,61	14,61	14,61
Nusa Tenggara Barat	2,87	3,45	3,36	2,53	2,70	2,75	3,06	3,08	2,78	3,85	4,04	3,32	2,82	2,66	2,69	2,38	2,25	2,15	2,76	2,90	2,78	2,91	2,91	2,91
Nusa Tenggara Timur	3,72	2,86	3,20	3,48	3,20	3,50	3,66	3,48	3,57	4,57	4,80	5,71	3,91	4,37	4,44	4,29	3,94	3,92	4,29	4,04	4,81	3,99	3,99	3,99
Kalimantan Barat	3,26	4,72	4,44	5,05	3,90	4,05	3,58	3,72	4,43	3,83	4,12	5,44	4,62	4,12	3,17	3,11	2,62	2,55	2,93	2,58	3,94	3,94	3,94	3,94
Kalimantan Tengah	8,02	8,20	7,03	6,23	6,97	7,16	7,61	7,55	8,72	9,41	8,48	9,99	7,79	7,57	7,32	7,09	7,93	7,62	8,17	7,67	7,69	7,82	7,82	7,82
Kalimantan Selatan	9,83	10,32	10,12	10,16	8,81	9,75	8,74	9,22	10,51	9,88	11,15	11,52	9,91	9,60	10,41	10,05	9,74	9,42	10,95	9,77	9,85	10,03	10,03	10,03
Kalimantan Timur	15,18	19,87	18,25	18,28	18,05	16,48	16,95	17,60	19,06	20,09	19,22	22,62	18,73	17,94	18,04	17,19	16,70	18,02	16,32	15,32	17,86	18,08	18,08	18,08
Kalimantan Utara	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13,02	15,95	15,95	15,95
Sulawesi Utara	5,22	4,79	6,59	6,09	5,91	7,10	4,84	6,06	6,47	7,03	6,95	7,15	7,07	7,20	6,33	6,43	5,26	5,27	5,47	4,78	4,85	6,04	6,04	6,04
Sulawesi Tengah	5,88	4,94	4,75	5,25	4,89	4,99	5,84	5,84	5,56	5,96	6,10	6,89	4,97	4,89	4,38	4,17	4,76	4,41	4,73	4,52	5,22	5,19	5,19	5,19
Sulawesi Selatan	5,53	5,56	6,68	6,18	6,03	5,88	6,52	6,71	8,06	6,93	6,83	7,16	5,84	6,03	5,92	5,81	5,47	5,34	5,24	4,68	5,55	6,08	6,08	6,08
Sulawesi Tenggara	4,06	3,26	3,26	3,61	3,91	4,18	5,30	5,06	5,01	5,77	4,67	6,35	5,01	5,18	5,37	5,07	5,15	4,93	4,67	6,00	6,09	4,96	4,96	4,96
Gorontalo	-	-	1,30	1,39	0,91	1,79	2,05	1,71	1,95	1,83	1,58	3,26	2,31	2,51	1,48	1,53	1,18	1,20	2,29	1,56	2,79	1,88	1,88	1,88
Sulawesi Barat	-	-	-	-	-	-	1,95	2,04	2,90	2,67	4,09	3,38	3,53	2,21	2,26	2,07	2,14	2,49	2,57	2,26	2,36	2,56	2,56	2,56
Maluku	4,99	-	5,53	12,04	2,78	3,83	5,00	4,66	5,08	5,65	5,99	9,52	6,11	6,13	5,62	5,32	5,13	5,14	5,76	5,65	6,93	5,84	5,84	5,84
Maluku Utara	-	-	1,51	12,30	2,53	2,77	3,15	3,47	5,78	6,04	5,80	6,61	4,87	4,54	5,29	5,11	4,75	4,58	4,25	4,86	5,59	4,91	4,91	4,91
Papua Barat	-	-	-	-	-	-	-	10,95	12,44	10,80	8,83	14,50	13,31	13,23	14,30	13,13	13,35	12,10	12,38	13,39	13,38	12,42	12,42	12,42
Papua	6,29	6,63	7,76	30,86	8,80	8,32	10,76	9,56	10,16	8,69	10,03	8,91	8,69	8,52	8,15	7,71	8,03	7,64	8,15	7,79	6,96	9,45	9,45	9,45
Total	7,81	8,06	8,36	7,68	7,30	7,42	7,87	7,79	8,36	9,08	8,88	10,32	8,68	8,49	8,24	7,88	8,37	8,06	8,57	8,08	8,51	8,51	8,51	8,51

Sumber : Badan Pusat Statistik , url : <https://www.bps.go.id>

Data tersebut kemudian diakumulasikan berdasarkan 1 atribut yakni jumlah rata-rata persentasi Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa seperti yang ditunjukkan pada tabel 2 berikut:

TABEL 2  
DATA JUMLAH RATA-RATA PERSENTASI RUMAH TANGGA MENURUT PROVINSI DAN STATUS KEPEMILIKAN RUMAH KONTRAK/SEWA

No	Provinsi	Persen
1	Aceh	7,5
2	Sumatera Utara	12,1
3	Sumatera Barat	10,4
4	Riau	15,4
5	Jambi	7,8
6	Sumatera Selatan	7,6
7	Bengkulu	8,8
8	Lampung	4,0
9	Kepulauan Bangka Belitung	6,3
10	Kepulauan Riau	24,0
11	DKI Jakarta	33,6
12	Jawa Barat	7,8
13	Jawa Tengah	2,7
14	DI Yogyakarta	15,7
15	Jawa Timur	5,1
16	Banten	13,5
17	Bali	14,6
18	Nusa Tenggara Barat	2,9
19	Nusa Tenggara Timur	4,0
20	Kalimantan Barat	3,8
21	Kalimantan Tengah	7,8
22	Kalimantan Selatan	10,0
23	Kalimantan Timur	18,1
24	Kalimantan Utara	16,0
25	Sulawesi Utara	6,0
26	Sulawesi Tengah	5,2
27	Sulawesi Selatan	6,1
28	Sulawesi Tenggara	4,9
29	Gorontalo	1,9
30	Sulawesi Barat	2,6
31	Maluku	5,8
32	Maluku Utara	4,9
33	Papua Barat	12,4
34	Papua	9,4

Setelah diakumulasikan maka akan didapatkan nilai dari seluruh *persentasi*. Kemudian data tersebut akan masuk ke tahapan *clustering* dengan



menerapkan algoritma *K-Means* menggunakan *rapid minner* untuk mengcluster data menjadi tiga cluster. Data yang telah diakumulasikan akan dimasukkan ke dalam *tool* Rapid Minner. Sehingga dapat *diclustering* menjadi tiga. Setelah dimasukkan ke dalam Rapid Minner. Hasil akumulasi data dapat dilihat pada tabel 2.

Page | 79

#### D. Centroid Data

Dalam penerapan algoritma *K-means* dihasilkan nilai titik tengah atau centroid dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 3, Penentuan cluster dibagi atas tiga bagian yakni cluster tingkat tinggi (C1), cluster tingkat sedang (C2) dan cluster tingkat rendah (C3). maka nilai titik tengah atau centroid juga terdapat 3 titik. Penentuan titik cluster ini dilakukan dengan mengambil nilai terbesar (*maksimum*) untuk *cluster* tingkat tinggi (C1), nilai rata-rata (*average*) untuk *cluster* tingkat sedang (C2) dan nilai terkecil (*minimum*) untuk *cluster* tingkat rendah (C3). Nilai titik tersebut dapat diketahui pada Tabel 3 berikut:

TABEL 3  
CENTROID DATA AWAL

Nilai	C1	C2	C3
Jlh Pesentasi	%	%	%
	33,6	9,4	1,9

#### E. Clustering Data

Dengan menggunakan centroid tersebut maka dapat *dicluster* data yang telah didapat menjadi 3 cluster. Proses *cluster* dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data persentasi Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa didapatkan pengelompokan pada *iterasi* 1 untuk 3 cluster tersebut. *Cluster* tingkat tinggi (C1) terdiri dari 2 provinsi yakni Kepulauan Riau dan DKI Jakarta, *cluster* tingkat sedang (C2) terdiri dari 21 provinsi dan *cluster* tingkat rendah (C3) terdiri dari 11 provinsi. Pengelompokan data pada *iterasi* 1 dan *Clustering* data dapat digambarkan pada tabel dan gambar berikut:

TABEL 4  
PERHITUNGAN JARAK PUSAT CLUSTER ITERASI 1

No	Provinsi	Persen	C1	C2	C3	Jarak Terpendek
1	Aceh	7,5	26	2	6	2
2	Sumatera Utara	12,1	21	3	10	3
3	Sumatera Barat	10,4	23	1	9	1
4	Riau	15,4	18	6	14	6
5	Jambi	7,8	26	2	6	2
6	Sumatera Selatan	7,6	26	2	6	2
7	Bengkulu	8,8	25	1	7	1
8	Lampung	4,0	30	5	2	2
9	Kepulauan Bangka Belitung	6,3	27	3	4	3
10	Kepulauan Riau	24,0	10	15	22	10
11	DKI Jakarta	33,6	0	24	32	0
12	Jawa Barat	7,8	26	2	6	2
13	Jawa Tengah	2,7	31	7	1	1
14	DI Yogyakarta	15,7	18	6	14	6
15	Jawa Timur	5,1	28	4	3	3
16	Banten	13,5	20	4	12	4
17	Bali	14,6	19	5	13	5
18	Nusa Tenggara Barat	2,9	31	6	1	1
19	Nusa Tenggara Timur	4,0	30	5	2	2
20	Kalimantan Barat	3,8	30	6	2	2
21	Kalimantan Tengah	7,8	26	2	6	2
22	Kalimantan Selatan	10,0	24	1	8	1
23	Kalimantan Timur	18,1	16	9	16	9
24	Kalimantan Utara	16,0	18	7	14	7
25	Sulawesi Utara	6,0	28	3	4	3
26	Sulawesi Tengah	5,2	28	4	3	3
27	Sulawesi Selatan	6,1	28	3	4	3
28	Sulawesi Tenggara	4,9	29	5	3	3
29	Gorontalo	1,9	32	8	0	0
30	Sulawesi Barat	2,6	31	7	1	1
31	Maluku	6,2	27	3	4	3
32	Maluku Utara	4,9	29	4	3	3
33	Papua Barat	12,4	21	3	11	3
34	Papua	9,4	24	0	8	0

TABEL 5  
PENGELOMPAKAN DATA ITERASI 1

No	Provinsi	C1	C2	C3
1	Aceh		1	
2	Sumatera Utara		1	
3	Sumatera Barat		1	
4	Riau		1	
5	Jambi		1	
6	Sumatera Selatan		1	
7	Bengkulu		1	
8	Lampung			1
9	Kepulauan Bangka Belitung		1	
10	Kepulauan Riau	1		
11	DKI Jakarta	1		
12	Jawa Barat		1	
13	Jawa Tengah			1
14	DI Yogyakarta		1	
15	Jawa Timur			1
16	Banten		1	
17	Bali		1	
18	Nusa Tenggara Barat			1
19	Nusa Tenggara Timur			1
20	Kalimantan Barat			1
21	Kalimantan Tengah		1	
22	Kalimantan Selatan		1	
23	Kalimantan Timur		1	
24	Kalimantan Utara		1	
25	Sulawesi Utara		1	
26	Sulawesi Tengah			1
27	Sulawesi Selatan		1	
28	Sulawesi Tenggara			1
29	Gorontalo			1
30	Sulawesi Barat			1
31	Maluku		1	
32	Maluku Utara			1
33	Papua Barat		1	
34	Papua		1	

Proses *K-Means* akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data *iterasi* sebelumnya. Dengan kata lain, proses akan terus melakukan *iterasi* sampai data pada *iterasi* terakhir sama dengan *iterasi* sebelumnya. Pada *iterasi* 1 diperoleh cluster data data persentasi Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa, yang dapat dilihat pada gambar 2. Proses ini berlangsung bekal kali sampai diperoleh hasil *iterasi* terakhir sama dengan *iterasi* sebelumnya. Untuk kasus ini proses berakhir di *iterasi* 4 dimana jumlah pengelompokan data sama dengan *iterasi* 3. Proses pencarian nilai centroid data dapat digambarkan pada tabel dan gambar berikut:

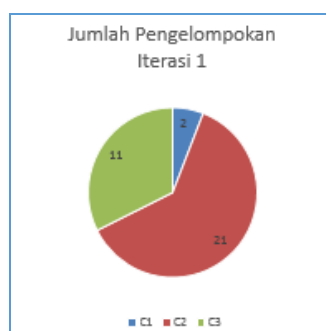
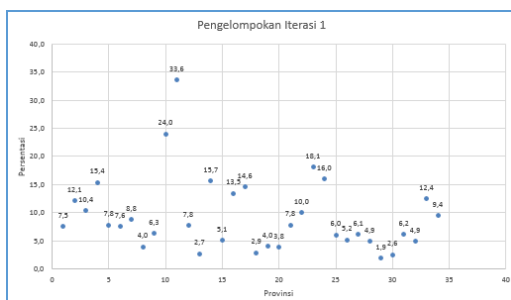
TABEL 6  
CENTROID DATA ITERASI 4

Nilai	C1	C2	C3
Jlh Pesentasi	%	%	%
	28,79865	13,03889	5,251482

Setelah mendapatkan nilai titik tengah atau *centroid*, proses sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terpendek, pengelompokan data pada *iterasi* 4 dan *Clustering* data dapat digambarkan pada tabel dan gambar berikut:

TABEL 7  
PERHITUNGAN JARAK PUSAT CLUSTER ITERASI 4

No	Provinsi	Persen	C1	C2	C3	Jarak Terpendek
1	Aceh	7,5	21	5	2	2
2	Sumatera Utara	12,1	17	1	7	1
3	Sumatera Barat	10,4	18	3	5	3
4	Riau	15,4	13	2	10	2
5	Jambi	7,8	21	5	3	3
6	Sumatera Selatan	7,6	21	5	2	2
7	Bengkulu	8,8	20	4	4	4
8	Lampung	4,0	25	9	1	1
9	Kepulauan Bangka Belitung	6,3	23	7	1	1
10	Kepulauan Riau	24,0	5	11	19	5
11	DKI Jakarta	33,6	5	21	28	5
12	Jawa Barat	7,8	21	5	3	3
13	Jawa Tengah	2,7	26	10	3	3
14	DI Yogyakarta	15,7	13	3	10	3
15	Jawa Timur	5,1	24	8	0	0
16	Banten	13,5	15	0	8	0
17	Bali	14,6	14	2	9	2
18	Nusa Tenggara Barat	2,9	26	10	2	2
19	Nusa Tenggara Timur	4,0	25	9	1	1
20	Kalimantan Barat	3,8	25	9	1	1
21	Kalimantan Tengah	7,8	21	5	3	3
22	Kalimantan Selatan	10,0	19	3	5	3
23	Kalimantan Timur	18,1	11	5	13	5
24	Kalimantan Utara	16,0	13	3	11	3
25	Sulawesi Utara	6,0	23	7	1	1
26	Sulawesi Tengah	5,2	24	8	0	0
27	Sulawesi Selatan	6,1	23	7	1	1
28	Sulawesi Tenggara	4,9	24	8	0	0
29	Gorontalo	1,9	27	11	3	3
30	Sulawesi Barat	2,6	26	10	3	3
31	Maluku	6,2	23	7	1	1
32	Maluku Utara	4,9	24	8	0	0



Gambar 3. Clustering Data

33	Papua Barat	12,4	16	1	7	1
34	Papua	9,4	19	4	4	4

TABEL 8  
PENGELOMPAKAN DATA ITERASI 4

No	Provinsi	C1	C2	C3
1	Aceh			1
2	Sumatera Utara		1	
3	Sumatera Barat		1	
4	Riau		1	
5	Jambi			1
6	Sumatera Selatan			1
7	Bengkulu		1	
8	Lampung			1
9	Kepulauan Bangka Belitung			1
10	Kepulauan Riau	1		
11	DKI Jakarta	1		
12	Jawa Barat			1
13	Jawa Tengah			1
14	DI Yogyakarta		1	
15	Jawa Timur			1
16	Banten		1	
17	Bali		1	
18	Nusa Tenggara Barat			1
19	Nusa Tenggara Timur			1
20	Kalimantan Barat			1
21	Kalimantan Tengah			1
22	Kalimantan Selatan		1	
23	Kalimantan Timur		1	
24	Kalimantan Utara		1	
25	Sulawesi Utara			1
26	Sulawesi Tengah			1
27	Sulawesi Selatan			1
28	Sulawesi Tenggara			1
29	Gorontalo			1
30	Sulawesi Barat			1
31	Maluku			1
32	Maluku Utara			1
33	Papua Barat		1	
34	Papua		1	

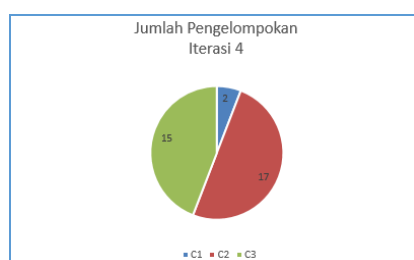
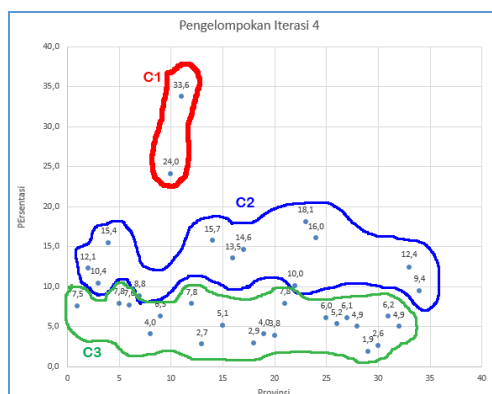
Pada iterasi 4, pengelompokan data yang dilakukan terhadap 3 cluster dengan iterasi yang didapatkan sama dengan hasil yang pada iterasi 3. Centroid data untuk cluster tingkatan tinggi 28,7986, Centroid data untuk cluster tingkatan sedang 13,0389 serta Centroid data untuk cluster tingkatan rendah 5,2515. Sehingga diperoleh penilaian persentase Rumah Tangga menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa dengan 2 provinsi cluster tingkatan tinggi yakni Kepulauan Riuan dan DKI Jakarta, 12 provinsi cluster tingkatan sedang yakni Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Jawa Tengah, Banten, Bali, Kalimantan Selatan, Kalimantan Timur, Kalimantan Tengah, Papua Barat dan Papua dan 20 provinsi cluster tingkatan rendah yakni Aceh, Jambi, Sumatera Selatan, Lampung, Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, NTB, NTT, Kalimantan Barat, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo dan Sulawesi Barat.

#### IV. KESIMPULAN DAN SARAN

Untuk melakukan penilaian terhadap hasil ekspor buah-buahan berdasarkan negara tujuan dapat menerapkan metode clustering K-Means. Data diolah untuk memperoleh nilai dari produksi ekspor buah-buahan berdasarkan negara tujuan. Data tersebut diolah menggunakan *Rapidminer* untuk ditentukan nilai *centroid* dalam 3 cluster yaitu cluster tingkat ekspor tinggi, cluster tingkat ekspor sedang dan cluster tingkat ekspor rendah. Centroid data untuk cluster tingkat ekspor tinggi 904.276,5, Centroid data untuk cluster tingkat ekspor sedang 265.501 dan Centroid data untuk cluster tingkat ekspor rendah 34.280,1. Sehingga diperoleh penilaian berdasarkan indeks ekspor buah-buahan dengan 2 negara cluster tingkat ekspor tinggi yakni India dan Pakistan, 3 negara cluster tingkat ekspor sedang yakni Singapura, Bangladesh dan Negara lainnya dan 6 negara cluster tingkat ekspor rendah yakni Hongkong, Tiongkok, Malaysia, Nepal, Vietnam dan Iran. Hasil yang dari penelitian dapat digunakan untuk mengetahui jumlah ekspor buah-buahan menurut negara tujuan.

#### REFERENSI

- [1] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10 - 20, 2013.
- [2] Y. Ardilla, H. Tjandrasa dan I. Arieshanti, "Deteksi Penyakit Epilepsidengan Menggunakan Entropi Permutasi, K-Means Clustering, dan Multilayer Perceptron," *Jurnal Teknik POMITS*, vol. 3, no. 1, pp. A70 - A74, 2014.
- [3] N. Atthina dan L. Iswari, "Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat kesehatan Daerah dengan Metode K-Means," *Seminar Nasional*



Gambar 4. Clustering Data Iterasi 4



*Aplikasi Teknologi Informasi (SNATI)*, Vol. %1 dari %2ISSN 1907 - 5022, pp. B52 - B59, 2014.

- [4] Anindya Khrisna Wardhani, "Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan", *Jurnal Transformatika*, Volume 14, Nomor 1, Juli 2016, Pp 30-37
- [5] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana dan F. Khusnu, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," Universitas Brawijaya Malang, Malang, 2012.
- [6] Soni, N., & Ganatra, A., 2012, *Categorization of several Clustering algorithms from different perspective: a review*, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(8), 63