

Penerapan Algoritma K-means Untuk Segmentasi Konsumen Menggunakan R

by Karina Auliasari

Submission date: 21-Jun-2020 06:22AM (UTC-0700)

Submission ID: 1347376871

File name: 10_JTMI_V5_N2_2019_PAPER_ABDIMASY-24-31.pdf (875.5K)

Word count: 3227

Character count: 19537

PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI KONSUMEN MENGUNAKAN R

Karina Auliasari¹, Mariza Kertaningtyas²

^{1,2} Fakultas Teknologi Industri, Institut Teknologi Nasional Malang
Jl. Raya Karanglo Km. 2, Malang

Email: karina.auliasari86@gmail.com , mariza_kertaningtyas@lecturer.itn.ac.id

13

ABSTRAK — Means merupakan salah satu teknik analisis data yang digunakan untuk mengelompokkan konsumen berdasarkan beberapa variabel, dimana teknik ini menjadi sangat handal ketika konsumen berinteraksi dengan produk yang ditawarkan. Teknik ini juga membantu perusahaan dalam memodelkan hubungan antara kecenderungan produk yang dibeli konsumen yang mungkin tidak diperhatikan oleh pihak perusahaan sebelumnya. Dengan demikian, penelitian ini berfokus pada penargetan konsumen yang menguntungkan dengan analisis cluster dan upaya untuk mengklasifikasikan konsumen dari perusahaan berdasarkan karakteristiknya menggunakan algoritma K-Means menggunakan bahasa pemrograman R. Dari hasil *cluster* dataset konsumen PT. Super Sukses Niaga pada tahun 2018 menunjukkan bahwa konsumen yang paling berpotensi ada pada *cluster* 2, di mana *cluster* 2 memiliki nilai rata-rata Total_Belanja_Setahun paling tinggi yaitu 4198.6656. Pada *cluster* 2 ini konsumen didominasi oleh jenis usaha PT karena nilai rata-rata Jenis.1 bernilai 1.667 yang mendekati nilai 2 (di mana nilai 2 merupakan label untuk jenis usaha PT). Hal serupa juga dihasilkan dari nilai rata-rata Lokasi.1 pada *cluster* 2 yang bernilai 1.667 yang mendekati nilai 2 (di mana nilai 2 merupakan label untuk lokasi pulau Kalimantan). Oleh karena itu, dapat disimpulkan konsumen yang memiliki nilai tinggi dalam melakukan transaksi adalah PT. BIS INDUSTRI, PT. REA KALTIM dan CV. MEDCO MINING pada *cluster* 2.

Kata Kunci: Segmentasi, Konsumen, K-Means, R.

ABSTRACT — Means is often used to clustering consumers and the company's market share. This technique is used to group consumers based on several variables, where this technique becomes very reliable when consumers interact with the products offered. This technique also helps companies measure the relationship between the tendency of products purchased by consumers that may not have been noticed by the company before. This study focuses on targeting profitable consumers with cluster analysis based on their characteristics using the K-Means algorithm and the R programming language. The results shows that the most potential consumers are in cluster 2, where cluster 2 has the highest value of Total_Belanja_Setahun, which is 4198.6656. In cluster 2, consumers are dominated by PT type of business because the average value of Jenis.1 is 1,667 which is close to 2 (where 2 is a label for PT). The same thing is also produced from the average value of Lokasi.1 in cluster 2 which is 1,667 which is close to 2 (where 2 is a label for Kalimantan). Therefore, it can be concluded that consumers who have high value in conducting transactions are PT. BIS INDUSTRI, PT. REA KALTIM and CV. MEDCO MINING on cluster 2.

Index terms: Segmentation, Customer, K-Means, R.

1. PENDAHULUAN

Dalam rangka mengidentifikasi konsumen yang berharga, beberapa perusahaan menerapkan *customer relationship management* (CRM). CRM sendiri merupakan serangkaian strategi bisnis yang digunakan perusahaan yang berfokus pada konsumen untuk menarik lebih banyak loyalitas konsumen dan membangun hubungan jangka panjang dan menguntungkan [1]. Penerapan CRM juga membantu perusahaan untuk meningkatkan nilai konsumen dan mencapai laba tinggi [2]. Segmentasi konsumen merupakan salah satu teknik dasar dalam langkah-langkah CRM [3] dan juga merupakan pengelompokan konsumen berdasarkan karakteristik yang serupa [4]. Dengan teknik segmentasi konsumen menjadi portofolio, perusahaan memperhatikan konsumen dalam berbagai aspek dan mengidentifikasi konsumen yang paling menguntungkan. Ini juga membantu perusahaan untuk membangun loyalitas dengan konsumen dengan mengembangkan dan menawarkan produk dan layanan yang konsumen inginkan dan membuat konsumen lama puas serta mampu menarik konsumen yang baru [5]. Segmentasi konsumen mengelompokkan data konsumen menjadi kelompok-kelompok konsumen tertentu dimana masing-masing kelompok memiliki karakteristik yang serupa yang terkait dengan pemasaran produk, seperti umur, jenis kelamin, kecenderungan memilih produk tertentu ataupun seberapa banyak uang yang dikeluarkan konsumen dalam berbelanja. Segmentasi konsumen menjadi cara yang handal dalam mengidentifikasi pangsa pasar suatu perusahaan. Dengan menggunakan data konsumen perusahaan kemudian dapat merencanakan strategi seperti meningkatkan layanan, memberikan promosi yang menarik atau mengembangkan produk baru. Mampu mengelompokkan konsumen memungkinkan perusahaan untuk lebih

memahami konsumen lebih jauh lagi mengetahui siapa saja yang menjadi target konsumen dengan cara yang lebih efisien. Pada umumnya perusahaan memberi label secara manual pada konsumen sebagai pembeda berdasarkan kebiasaan belanja yang dapat diamati. Hal ini memungkinkan terjadi kesalahan atau bias bila dilakukan oleh manusia.

Oleh karena itu peran komputer dalam mengelompokkan konsumen sangat dibutuhkan. Komputer akan mengelompokkan berdasarkan beberapa variabel yang disebut sebagai klasifikasi tanpa pengawasan (*unsupervised learning*) dan menjadi sangat handal ketika konsumen berinteraksi dengan produk yang ditawarkan dalam berbagai cara. Teknik ini membantu dalam memodelkan hubungan antara tindakan dan kepribadian konsumen yang mungkin tidak diperhatikan sebelumnya. Beberapa penelitian dilakukan menggunakan teknik clustering untuk mengidentifikasi konsumen, diantaranya penelitian yang dilakukan oleh Dursun dan Caber di tahun 2016, dimana dalam penelitian ini menggunakan algoritma *self organizing maps* (SOM) untuk menentukan jumlah cluster pada dataset konsumen hotel yang berada di Antalya Turki. Dari hasil penelitian Dursun dan Caber dihasilkan cluster-cluster konsumen berdasarkan data profil konsumen yang menghasilkan variable-variabel berdasarkan analisis RFM (*Recency of the last purchases, Frequency of the purchases, dan Monetary value of the purchases*) [6]. Dalam beberapa penelitian, kombinasi RFM (*Recency of the last purchases, Frequency of the purchases, dan Monetary value of the purchases*) dan berbagai metode pengelompokan telah digunakan untuk mengelompokkan pelanggan. Cheng et al. (2009) mengusulkan model pengelompokan dua tahap untuk mengklasifikasikan pasien dari rumah sakit. Data pasien dianalisis dengan RFM kemudian mereka menggunakan K-Means untuk mengklasifikasikan departemen rumah sakit dan akhirnya mengidentifikasi departemen yang menghasilkan laba tertinggi [7]. Berbagai macam teknik telah digunakan untuk segmentasi konsumen, karena sejumlah besar data yang

diperoleh setiap hari dari konsumen, teknik data mining seperti algoritma clustering telah banyak digunakan untuk segmentasi konsumen. *Data Science* sendiri merupakan keilmuan yang semakin dibutuhkan saat ini. Berangkat dari kebutuhan akan keilmuan *data science* maka ada beberapa bahasa pemrograman *open source* yang ikut berkembang karena banyak digunakan untuk mengaplikasikan *data science*. Salah satu dari beberapa bahasa pemrograman tersebut adalah R, dimana berdasarkan peringkat IEEE bahasa pemrograman R menempati peringkat kelima terpopuler setelah C++ untuk analisis data [8]. Ini menunjukkan bahwa Bahasa pemrograman R unggul jika dibandingkan dengan pesaingnya seperti Python dan C [8]. Bahasa pemrograman R juga didukung oleh banyak paket *library* analisis data yang tersedia pada *repository* CRAN (*Comprehensive R Archive Network*) yang dibuat dan dikembangkan oleh komunitas yang sangat aktif [9]. Dengan demikian, penelitian ini berfokus pada penargetan konsumen yang menguntungkan dengan analisis cluster dan upaya untuk mengklasifikasikan konsumen dari perusahaan berdasarkan karakteristiknya menggunakan algoritma K-Means menggunakan bahasa pemrograman R. Segmentasi konsumen dapat membantu manajer untuk menghasilkan strategi CRM yang tepat untuk setiap kelompok konsumen dan dapat ditekan pada CRM.

Algoritma *K-Means* diperkenalkan oleh J.B. MacQueen pada tahun 1976. Metode ini mempartisi data ke dalam *cluster* (kelompok) sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok yang lain [10]. Berikut adalah langkah-langkah dari algoritma *K-Means* [10]:

Step 1 : Menentukan banyak *K-cluster* yang ingin dibentuk.

Step 2 : Membangkitkan nilai random untuk pusat *cluster* awal (*centroid*) sebanyak k.

Step 3 : Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidian* (*Euclidian Distance*) hingga ditemukan

jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidian Distance* (persamaan 1):

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

Step 4 : Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

Step 5 : Mengupdate nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus yang ditunjukkan pada persamaan 2:

$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j=S_j} x_j \quad (2)$$

dimana:

$\mu_j(t+1)$ = centroid baru pada iterasi ke $(t+1)$,

N_{S_j} = banyak data pada cluster S_j

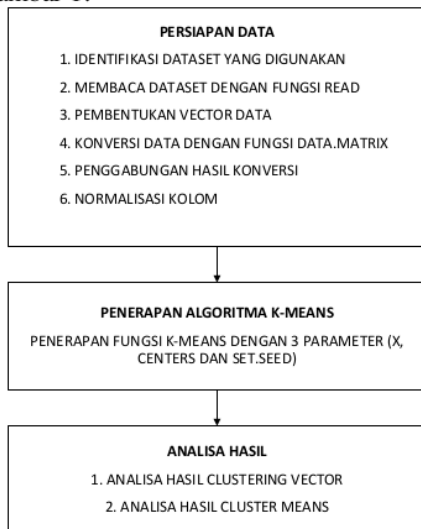
Step 6 : Melakukan n perulangan dari langkah 2 hingga 5 hingga anggota tiap *cluster* tidak ada yang berubah.

Step 7 : Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk *Radial Basis Function* yang ada di *hidden layer*.

2. METODOLOGI PENELITIAN

Beberapa tahapan yang dilakukan untuk melakukan segmentasi data konsumen sehingga perusahaan mampu mengenal konsumennya dengan lebih baik. Tahapan pertama adalah persiapan data tahap ini penting dilakukan karena dataset yang digunakan harus memiliki keseragaman tipe data berupa tipe data angka. Dalam tahapan pertama terdapat detail langkah-langkah yang harus dilakukan yaitu mengenal dataset yang digunakan, membaca dataset dengan fungsi *read*, membuat *vector*, mengubah data ke dalam bentuk *matrix*, menggabungkan data yang sebelumnya telah diubah, menormalisasi data, dan membuat data master. Tahap selanjutnya setelah persiapan data adalah tahap eksplorasi data, dimana dalam tahap ini diterapkan fungsi *kmeans* untuk selanjutnya dilakukan analisa hasil pengelompokkan. Dengan menganalisa

hasil pengelompokkan data, nantinya dapat diketahui seperti apa kedekatan tiap titik data dari tiap kelompok yang terbentuk sehingga bisa menjadi inisiasi penentuan jumlah *cluster* yang optimal. Gambaran blok diagram tahapan penelitian dari tiap penjabarannya tadi ditunjukkan pada Gambar 1.



Gambar 1. Tahapan segmentasi data Dataset yang digunakan

Dataset yang digunakan dalam penelitian ini adalah data konsumen yang melakukan transaksi pembelian pada PT. Super Sukses Niaga pada tahun 2018 dengan lima atribut data yaitu Customer_ID, Nama_Pelanggan, Jenis, Lokasi, dan Total_Belanja_Setahun. Detail penjelasan dari keenam kolom tersebut ditunjukkan pada Tabel 1.

Tabel 1. Atribut data konsumen dan deskripsi data

Atribut tabel			
No	Nama Atribut	Tipe Data	Deskripsi
1	Customer_ID	Character & Integer	Kode konsumen
2	Nama_Pelanggan	Character	Nama konsumen
3	Jenis	Character	Jenis konsumen (PT/CV)

4	Lokasi	Character	Lokasi konsumen (Kepulauan)
5	Total_Belanja_Setahun	Integer	Total nominal belanja yang dikeluarkan konsumen dalam setahun (dalam satuan juta rupiah)

Dari Tabel 1 terlihat jika tidak semua atribut data bertipe data angka, ada dua atribut yang bertipe data *character* yaitu Jenis dan Lokasi sisanya atribut Total_Belanja_Setahun bertipe data angka. Untuk atribut data Customer_ID dan Nama_Pelanggan dianggap unik karena kedua atribut dijadikan identitas data yang harus memiliki nilai unik yang mewakili tiap individu sehingga berbeda dari tiap baris data.

Persiapan data

Proses awal pada persiapan data adalah membaca dataset konsumen, pada penelitian ini bahasa pemrograman R yang digunakan akan secara otomatis membaca dataset dengan fungsi *read*. Dataset konsumen yang digunakan disini merupakan file teks dengan format .txt (data_konsumen_ssn.txt). Untuk membaca data_konsumen_ssn.txt digunakan fungsi *read.csv*, disini file csv merupakan format *comma separated value*, fungsi ini akan membaca isi dari file teks menjadi *data.frame* pada R, hasil dari fungsi *read.csv* ditunjukkan pada Gambar 2, dimana ada beberapa baris data transaksi yang dilakukan konsumen PT. Super Sukses Niaga.

```

> konsumen <- read.csv("D:/R/data_konsumen_ssn.txt", sep="\t")
> konsumen
  Customer_ID      Nama_Pelanggan
1  CUST-001      MEDCO MINING
2  CUST-002      PT.BIS INDUSTRI
3  CUST-003      PT.REA KALTIM PLANTATIONS
4  CUST-004      PT.HEXINDO ADI PERKASA
5  CUST-005      DIESELINDO NUSA UTAMA
6  CUST-006      PT.DUTA PUTERA SUMATERA
7  CUST-007      SUPER SUKSES GROUP
8  CUST-008      PT.UNITED EQUIPMENT INDONESIA
9  CUST-009      PT.ORICA MINING SERVICE
10 CUST-010      PT.SUKSES INTI SOLUSINDO

Jenis  Lokasi  Total_Belanja_Setahun
CV Kalimantan  4812938365
PT Jawa  4387246588
PT Kalimantan  3395811782
PT Jawa  1079100000
CV Jawa  1053461750
PT Sumatera  445511000
PT Jawa  285462463
PT Jawa  215682500
PT Kalimantan  172393650
PT Jawa  151120200

```

Gambar 2. Tampilan Fungsi *read.csv*

Sesuai memanggil fungsi *read.csv* selanjutnya dilakukan pembentukan *vector* untuk atribut data Jenis, dan Lokasi sehingga dapat digunakan berulang kali dalam script R. *Vector* merupakan jenis data di bahasa pemrograman R dengan struktur data yang menyimpan deretan nilai (lebih dari satu nilai) dengan tipe data yang sama. Oleh karena itu, apabila suatu data bertipe data *character* maka seluruh data harus bertipe data *character* begitu pula sebaliknya jika tipe datanya angka maka keseluruhan data harus bertipe data angka. *Vector* didefinisikan dengan nama function

```

> konsumen <- read.csv("D:/R/data_konsumen_ssn.txt", sep="\t")
> konsumen_matrix <- data.matrix(konsumen[c("Jenis", "Lokasi")])
> konsumen <- data.frame(konsumen, konsumen_matrix)
> Lokasi <- unique(konsumen[c("Lokasi", "Lokasi.1")])
> Jenis <- unique(konsumen[c("Jenis", "Jenis.1")])

```

Gambar 3. Tampilan Langkah Pembentukan *Vector*

Selanjutnya perlu dilakukan normalisasi data pada kolom Total_Belanja_Setahun dimana pada berisi data dengan nilai jutaan. Kolom Total_Belanja_Setahun digunakan untuk proses pengelompokkan oleh karena itu nilainya dinormalisasi agar mempermudah dan menyederhanakan perhitungan tanpa mengurangi tingkat akurasi, proses normalisasi data ini sangat penting dilakukan mengingat kemungkinan penambahan data konsumen di masa depan. Normalisasi data Total_Belanja_Setahun dilakukan dengan cara membagi data sehingga nilai jutaan menjadi puluhan.

yang hanya memiliki satu huruf saja yaitu *c*. Setelah pembentukan *vector*, isi data dari dua atribut yaitu Jenis dan Lokasi yang bertipe data *character* harus diubah (dikonversi) menjadi tipe data numerik agar bisa digunakan dalam tahap penerapan algoritma K-Means. Untuk mengubah data pada *vector* menjadi tipe data numerik digunakan fungsi *data.matrix*. Fungsi *data.matrix* akan mengubah data pada kolom Jenis dan Lokasi yang diwakili oleh `konsumen[c("Jenis", "Lokasi")]` menjadi numerik. Setelah proses konversi ke angka, perlu dilakukan penggabungan data ke variabel asal, hal ini bertujuan untuk mempermudah proses segmentasi menggunakan metode K-Means. Untuk menggabungkan data hasil konversi (*data.matrix*) ke dalam data konsumen, digunakan kembali fungsi *data.frame* hal ini dikarenakan kolom Jenis dan Lokasi sudah ada maka ditambahkan akhiran *.1* pada kolom **Jenis** dan **Lokasi** yang terdapat pada **konsumen_matrix**. Langkah pembentukan *vector* hingga menggabungkan hasil konversi ditunjukkan pada Gambar 3.

Penerapan Algoritma K-Means

Data transaksi konsumen dibagi berdasarkan 3 kolom yaitu Jenis, Lokasi dan Total_Belanja_Setahun, ketiga kolom tersebut disimpan ke dalam sebuah *vector field_yang_digunakan*, selanjutnya dilakukan tahapan segmentasi terhadap data konsumen. Untuk melakukan segmentasi data konsumen digunakan fungsi *kmeans* menggunakan 3 parameter yaitu *x* sebagai data yang digunakan (keseluruhan data harus berupa numerik), *centers* yang merupakan jumlah cluster yang diinginkan dan *nstart* yaitu jumlah kombinasi acak

yang dihasilkan oleh library R. Pada fungsi *kmeans* disertai dengan pemanggilan fungsi *set.seed*, fungsi ini untuk menyeragamkan daftar nilai acak yang sama dari algoritma K-Means sehingga didapatkan output yang sama. Urutan langkah baris program untuk mengelompokkan data konsumen PT. Super Sukses Niaga menggunakan algoritma K-Means diperlihatkan pada Gambar 4. Pada Gambar 4 ketika baris program R berhasil dieksekusi didapatkan empat hasil yang

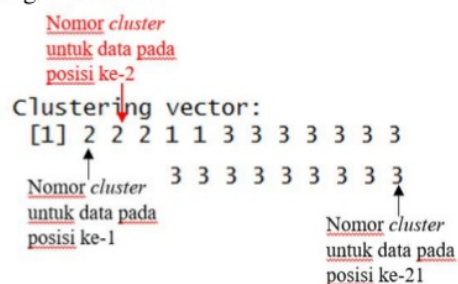
```
> konsumen$Total_Belanja_Setahun <- konsumen$Total_Belanja_Setahun/1000000
> kolom_yang_digunakan = c("Jenis.1", "Lokasi.1", "Total_Belanja_Setahun")
> set.seed(100)
> segmentasi <- kmeans(x=konsumen[field_yang_digunakan], centers=3, nstart=25)
> segmentasi
K-means clustering with 3 clusters of sizes 2, 3, 16
Cluster means:
  Jenis.1 Lokasi.1 Total_Belanja_Setahun
1 1.500000 1.000000          1066.2809
2 1.666667 1.666667          4198.6656
3 1.875000 1.500000           113.8933
Clustering vector:
[1] 2 2 2 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
Within cluster sum of squares by cluster:
[1] 329.1599 1057469.4053 221657.3371
(between_SS / total_SS = 97.1 %)
```

Gambar 4. Tampilan fungsi K-Means

3. HASIL DAN PEMBAHASAN

Jika diamati hasil dari *cluster* yang ditunjukkan pada Gambar 4 memperlihatkan hasil *clustering vector*, dimana merupakan rangkaian *vector* yang berisi angka *cluster*. Pada Gambar 5 *vector* berisi angka 1 sampai dengan 3 sesuai inputan jumlah *cluster* yang ingin dibentuk. *Vector* pada Gambar 5 dimulai dari angka 2, yang artinya data pertama dari dataset konsumen dialokasikan pada *cluster* nomor 2. Dari Gambar 5 juga terlihat isi *vector* keempat bernilai 1, ini menunjukkan bahwa data keempat dari keseluruhan dataset dialokasikan pada *cluster* nomor 1, dan seterusnya hingga posisi data terakhir (ke-21) memiliki *cluster* nomor 3.

masing-masing telah dilabeli dengan angka 1 sampai dengan 4. Label nomor 1 menunjukkan ada tiga kelompok/*cluster* yang masing-masing *cluster* memiliki jumlah titik data yang berbeda, label nomor 2 menunjukkan nilai rata-rata (*centroid*) dari tiap *cluster*, label nomor 3 merupakan pembagian *cluster* dari tiap elemen data berdasarkan posisi data dan terakhir label nomor 4 sebagai jumlah jarak kuadrat dari tiap titik ke *centroid*.



Gambar 5. Hasil Clustering Vector

Pada hasil *clustering* algoritma K-Means seperti yang ditunjukkan Gambar 3, bagian awal menunjukkan **K-Means clustering with 3 clusters of sizes 2, 3, 16**, ini artinya menggunakan algoritma K-Means dataset konsumen telah dibagi menjadi 3 *cluster*. Tiga *cluster* yang telah dibagi tadi masing-masing *cluster* ke-1 memiliki 2 data, *cluster* ke-2 memiliki 3 data dan *cluster* ke-3 memiliki 16 data, dimana jumlah total data 21 data yang merupakan jumlah total dataset konsumen. Jika ingin melihat *vector* mana saja yang masuk ke dalam masing-masing

cluster maka digunakan fungsi *which* seperti yang ditunjukkan pada Gambar 6 hingga Gambar 8. Pada Gambar 6 memperlihatkan bahwa cluster 1 memiliki 2 data (data ke-4 dan ke-5), cluster 2 pada Gambar 7 memiliki 3 data (data ke-1, ke-2 dan ke-3) dan cluster 3 pada Gambar 8 memiliki 16 data (data ke-6 sampai dengan data ke-21).

```
> segmentasi <- kmeans(x=konsumen[field_yang_digunakan],
centers=3, nstart=25)
> konsumen$cluster <- segmentasi$cluster
> konsumen[which(konsumen$cluster == 1),]
Customer_ID      Nama_Pelanggan Jenis Lokasi
4 CUST-004 PT.HEXINDO ADI PERKASA PT Jawa
5 CUST-005 DIESELINDO NUSA UTAMA CV Jawa
Total_Belanja_Setahun Jenis.1 Lokasi.1 cluster
1079.100          2          1          1
1053.462          1          1          1
```

```
> segmentasi <- kmeans(x=konsumen[field_yang_digunakan],
centers=3, nstart=25)
> konsumen$cluster <- segmentasi$cluster
> konsumen[which(konsumen$cluster == 3),]
Customer_ID      Nama_Pelanggan Jenis Lokasi
6 CUST-006 PT.DUTA PUTERA SUMATERA PT Sumatera
7 CUST-007 SUPER SUKSES GROUP PT Jawa
8 CUST-008 PT.UNITED EQUIPMENT INDONESIA PT Jawa
9 CUST-009 PT.ORICA MINING SERVICE PT Kalimantan
10 CUST-010 PT.SUKSES INTI SOLUSINDO PT Jawa
11 CUST-011 PT.PANATAMA RAJASA PT Jawa
12 CUST-012 PT.HIRUP HURIP TRANSINDO PT Kalimantan
13 CUST-013 PT.ROAD TECHNOLOGY INDONESIA PT Jawa
14 CUST-014 PT.BERLIAN CRANCERCO IND PT Jawa
15 CUST-015 PT.CHAKRA JAWARA PT Sumatera
16 CUST-016 PT.ADITI ALIANSI MOTOR PT Jawa
17 CUST-017 PT.INTERNATIONAL TRANSNUSA INDONESIA PT Sumatera
18 CUST-018 CV.INDABO JAYA MANDIRI CV Jawa
19 CUST-019 PT.YUNMA PERMATA MANDIRI PT Jawa
20 CUST-020 PT.MATA JUARA GEMILANG PT Jawa
21 CUST-021 UD.KEVINDO PRIMA CV Jawa
Total_Belanja_Setahun Jenis.1 Lokasi.1 cluster
445.51100          2          3          3
285.46246          2          1          3
215.68250          2          1          3
172.39365          2          2          3
151.12020          2          1          3
166.78549          2          1          3
108.01560          2          2          3
59.05548           2          1          3
35.52778           2          1          3
53.24067           2          3          3
29.90551           2          1          3
24.00640           2          3          3
19.33470           1          1          3
15.16020           2          1          3
29.49770           2          1          3
11.59400           1          1          3
```

Gambar 8. Tampilan vector cluster 3

Hasil cluster mean yang ditunjukkan pada label nomor 2 pada Gambar 3, merupakan nilai rata-rata centroid dari seluruh titik data tiap cluster. Untuk kolom **Jenis.1** menunjukkan nilai rata-rata dari data jenis usaha yang telah dirubah menjadi numerik dengan angka 1 mewakili CV dan angka 2 mewakili PT. Pada cluster 1 untuk kolom Jenis.1 (jenis usaha) terlihat bahwa nilai cluster mean bernilai 1.50 ini artinya bahwa data bersifat campuran (antara CV dan PT jumlahnya seimbang), sedangkan untuk cluster 2 dan 3 data

lebih banyak cenderung berjenis usaha PT karena nilainya 1.67 dan 1.87 yang mendekati nilai 2 yang berjenis usaha PT. Pada kolom **Lokasi.1** menunjukkan nilai rata-rata dari data lokasi konsumen yang juga telah diubah menjadi numerik dengan angka 1 mewakili pulau Jawa, angka 2 mewakili pulau Kalimantan dan angka 3 mewakili pulau Sumatera. Pada ketiga cluster terlihat bahwa lokasi konsumen cenderung ke pulau Jawa. Kolom terakhir yang dianalisa yaitu kolom **Total_Belanja_Setahun** memiliki nilai

Gambar 6. Tampilan vector cluster 1

```
> segmentasi <- kmeans(x=konsumen[field_yang_digunakan],
centers=3, nstart=25)
> konsumen$cluster <- segmentasi$cluster
> konsumen[which(konsumen$cluster == 2),]
Customer_ID      Nama_Pelanggan Jenis Lokasi
1 CUST-001 MEDCO MINING CV Kalimantan
2 CUST-002 PT.BIS INDUSTRI PT Jawa
3 CUST-003 PT.REA KALTIM PLANTATIONS PT Kalimantan
Total_Belanja_Setahun Jenis.1 Lokasi.1 cluster
4812.938          1          2          2
4387.247          2          1          2
3395.812          2          2          2
```

Gambar 7. Tampilan vector cluster 2

yang cukup signifikan pembagiannya untuk tiap *cluster*, dimana *cluster* ke-2 memiliki total belanja lebih tinggi dibandingkan *cluster* ke-1 dan ke-3. Data konsumen pada *cluster* ke-2 inilah yang bisa menjadi fokus untuk mempertahankan layanan dan pemasaran kepada konsumen pada *cluster* ke-2 dan perusahaan bisa meningkatkan pemasaran dan evaluasi pelayanan pada konsumen pada *cluster* 1 dan 3.

4. KESIMPULAN

Dari hasil *cluster* dataset konsumen PT. Super Sukses Niaga pada tahun 2018 menunjukkan bahwa konsumen yang paling berpotensi ada pada *cluster* 2, di mana *cluster* 2 memiliki nilai rata-rata Total_Belanja_Setahun paling tinggi yaitu 4198.6656. Pada *cluster* 2 ini konsumen didominasi oleh jenis usaha PT karena nilai rata-rata Jenis.1 bernilai 1.667 yang mendekati nilai 2 (di mana nilai 2 merupakan label untuk jenis usaha PT). Hal serupa juga dihasilkan dari nilai rata-rata Lokasi.1 pada *cluster* 2 yang bernilai 1.667 yang mendekati nilai 2 (di mana nilai 2 merupakan label untuk lokasi pulau Kalimantan). Oleh karena itu, dapat disimpulkan konsumen yang memiliki nilai tinggi dalam melakukan transaksi adalah PT. BIS INDUSTRI, PT. REA KALTIM dan CV. MEDCO MINING pada *cluster* 2.

Untuk penelitian segmentasi konsumen selanjutnya perlu dikembangkan pengujian *cluster* dengan beberapa nilai K untuk kemudian dilakukan pengujian hasil *cluster* terbaik berdasarkan variasi nilai K.

5. REFERENSI

- [1] Wu, S-I. and Lu, C-L. 2012. "The relationship between CRM, RM, and business performance: a study of the hotel industry in Taiwan". International Journal of Hospitality Management, Vol. 31, No. 1, pp.276–285.
- [2] Thakur, R. and Workman, L. 2016. "Customer portfolio management (CPM) for improved customer relationship management (CRM): are your customers platinum, gold, silver, or bronze?". Journal of Business Research, Vol. 69, No. 10, pp.4095–4102.
- [3] Smith, B. 2007. "Marketing master class: excellence in medical marketing: origins, definition and precursors". Journal of Medical Marketing, Vol. 7, No. 1, pp.25–32.
- [4] Chen, Y-S., Cheng, C-H., Lai, C-J., Hsu, C-Y. and Syu, H-J. 2012. "Identifying patients in target customer segments using a two stage clustering-classification approach: a hospital-based assessment". Computers in Biology and Medicine, Vol. 42, No. 2, pp.213–221.
- [5] Wei, J-T., Lin, S-Y., Weng, C-C. and Wu, H-H. 2012. "A case study of applying LRFM model in market segmentation of a children's dental clinic". Expert Systems with Applications, Vol. 39, No. 1, pp.5529–5533.
- [6] Dursun dan M. Caber, "Using Data Mining Techniques for Profiling Profitable Hotel Customers: an Application of RFM Analysis,". Tourism Management Perspectives, vol. 18, p. 53–160, 2016.
- [7] Cheng, C-H. and Chen, Y-S. (2009) "Classifying the segmentation of customer value via RFM model and RS theory". Expert Systems with Applications, Vol. 36, No. 2, pp.4176–4188.
- [8] B Stephen Cass. 2016. The 2016 top programming languages. [Online] <http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages>.
- [9] Tim Smith, Kevin Ushey. 2016. "Arrgh: a newcomer's (angry) guide to R". [Online] <http://arrgh.tim-smith.us/>.
- [10] Prasetyo E. 2014. "Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB". Penerbit ANDI OFFSET, Yogyakarta.

Penerapan Algoritma K-means Untuk Segmentasi Konsumen Menggunakan R

ORIGINALITY REPORT

12%

SIMILARITY INDEX

9%

INTERNET SOURCES

9%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

- 1** Serhat Peker, Altan Kocyigit, P. Erhan Eren. "LRFMP model for customer segmentation in the grocery retail industry: a case study", *Marketing Intelligence & Planning*, 2017
Publication 3%
- 2** Submitted to Thames Valley University
Student Paper 1%
- 3** *Kybernetes*, Volume 42, Issue 3 (2013-05-27)
Publication 1%
- 4** Hung Nguyen, Norma Harrison. "Leveraging customer knowledge to enhance process innovation", *Business Process Management Journal*, 2019
Publication 1%
- 5** Sabbir Hossain Shihab, Shyla Afroge, Sadia Zaman Mishu. "RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study", 2019 International 1%

Conference on Electrical, Computer and Communication Engineering (ECCE), 2019

Publication

6	Submitted to 87988 Student Paper	1%
7	Fintri Indriyani, Eni Irfiani. "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means", JUITA : Jurnal Informatika, 2019 Publication	1%
8	Julian Thomé, Lwin Khin Shar, Domenico Bianculli, Lionel C. Briand. "JoanAudit: a tool for auditing common injection vulnerabilities", Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering - ESEC/FSE 2017, 2017 Publication	1%
9	Drew Schmidt, Wei-Chen Chen, Michael A. Matheson, George Ostrouchov. "Programming with BIG Data in R: Scaling Analytics from One to Thousands of Nodes", Big Data Research, 2017 Publication	1%
10	jurnal.ugm.ac.id Internet Source	<1%
11	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1%

12	komputasi.wordpress.com Internet Source	<1%
13	etheses.uin-malang.ac.id Internet Source	<1%
14	jimfeb.ub.ac.id Internet Source	<1%
15	Submitted to Universitas Negeri Jakarta Student Paper	<1%
16	a2fahmi.com Internet Source	<1%
17	Submitted to iGroup Student Paper	<1%
18	Submitted to Universitas Diponegoro Student Paper	<1%
19	Submitted to Udayana University Student Paper	<1%
20	www.tandfonline.com Internet Source	<1%
21	aldragon.net Internet Source	<1%

Exclude bibliography On