Analisis Metode *K-Medoids* Pada Penjualan Produk *Smartphone Vivo* Di Kota Pematangsiantar

Mita Yustika¹, Agus Perdana Windarto², Yuegilion Pranavarna Purba³

^{1,2,3} STIKOM Tunas Bangsa, Pematangsiantar – Indonesia

¹Mitayustika2017@gmail.com

Abstract

Globalization is a very promising era of openness and freedom, where a communication which is a means to obtain information is expected to be carried out easily and effectively. One of the communication media in question is a cellphone. The users of cellphones or cellphones used to be limited to the elite, but for now it has begun to penetrate various circles of society ranging from students, university students, civil servants and even ordinary people who have used it... *K-Medoids* Clustering is one of the one technique of one of the data mining functionality, the clustering algorithm is an algorithm for grouping a number of data into certain data groups (clusters). While data mining, often referred to as Knowledge Discovery In Database (KDD) is an activity that includes the collection, use of historical data to find regularities, patterns or relationships in large data sets aimed at finding out the *Vivo Smartphone* brand that is selling well in the market so that it can be done. early procurement. From this study, it was found that *Vivo smartphones* with the brand *S1 Pro 8+128*, *V17 Pro*, *V19 8+128GB*, *V19 8+256GB*, *Y11 2+32GB*, *Y12 3+64GB*, *Y17 4+128GB*, *Y19 6+128GB* are brands which sold a lot. With the results obtained can provide information to PT. *Vivo* is improving its service even better by increasing the stock of *Vivo smartphones* which are in great demand, especially in the city of Pematangsiantar.

Keywords: *K-Medoids*, Clustering, Data mining, Rapidminer, *Vivo Smartphone*.

Abstrak

Globalisasi merupakan era yang sangat menjanjikan keterbukaan dan kebebasan, dimana sebuah komunikasi yang merupakan sarana untuk mendapatkan informasi diharapkan dapat dilakukan dengan mudah dan efektif. Salah satu media komunikasi yang dimaksud adalah Handphone. Para pengguna handphone atau telepon genggam dulu hanya terbatas pada kalangan-kalangan elit, namun untuk saat ini sudah mulai merambah ke berbagai kalangan masyarakat mulai dari pelajar, mahasiswa, pegawai negeri bahkan masyarakat biasa sudah banyak yang menggunakannya. *K-Medoids* Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma clustering merupakan algoritma pengelompokkan sejumlah data menjadi kelompok-kelompok data tertentu (cluster). Sedangkan Data mining, sering juga disebut sebagai Knowledge Discovery In Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar yang bertujuan dalam mengetahui merk *Smartphone Vivo* yang laku dipasaran sehingga dapat dilakukan pengadaan barang secara dini. Dari penelitian ini didapatkan hasil bahwa *smartphone vivo* dengan merk *S1 Pro 8+128*, *V17 Pro*, *V19 8+128GB*, *V19 8+256GB*, *Y11 2+32GB*, *Y12 3+64GB*, *Y17 4+128GB*, *Y19 6+128GB* merupakan merk yang banyak terjual. Dengan hasil yang diperoleh dapat memberikan informasi kepada PT. *Vivo* dalam meningkatkan pelayanan yang lebih baik lagi dengan memperbanyak stok *smarphone Vivo* yang paling banyak diminati khususnya di kota Pematangsiantar.

Kata Kunci: *K-Medoids*, Clustering, Data mining, Rapidminer, *Smartphone Vivo*.

1. Pendahuluan

Istilah data mining dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses pengelompokan dapat dilakukan dengan mengimplementasikan *Data Mining* metode *clustering*, salah satunya yaitu metode K-Medoids. Tujuan dari *Data Mining* yaitu mencari trend atau pola yang diinginkan dalam database besar untuk membantu dalam pengambilan keputusan pada waktu yang akan datang [1]. Globalisasi merupakan era yang sangat menjanjikan keterbukaan dan kebebasan, dimana sebuah komunikasi yang merupakan sarana untuk mendapatkan informasi diharapkan dapat dilakukan dengan mudah dan efektif. Salah satu media komunikasi yang dimaksud adalah *Handphone*. Para pengguna *handphone* atau telepon genggam dulu hanya terbatas pada kalangan-kalangan elit, namun untuk saat ini sudah mulai merambah ke berbagai kalangan masyarakat mulai dari pelajar, mahasiswa, pegawai negeri bahkan masyarakat biasa sudah banyak yang menggunakannya.

Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma *clustering* merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (*cluster*). Adapun tahapan data mining salah satunya yaitu *Clustereng*. Untuk menghilangkan data noise (data yang tidak relevan/berhubungan langsung dengan tujuan akhir proses datamining, misal: data mining yang bertujuan untuk menganalisa hasil penjualan, maka data-data dalam kumpulan seperti "nama pegawai, umur", dan sebagainya dapat diignore) dan tidak konsisten. Penelitian sebelumnya antara lain menggunakan algoritma K-Medoids dan K-Means untuk pengelompokan wilayah sebaran cacat pada anak dapat dilakukan untuk proses pengelompokan sesuai dengan informasi yang dimiliki oleh data, sehingga dapat diketahui wilayah mana yang memiliki sebaran cacat yang tinggi dengan batasan usia dibawah 18 tahun [2][3]. Selanjutnya dengan tema *clustering* pada penanganan kasus diare di Indonesia. Untuk mengetahui daerah mana saja yang paling tinggi dan paling rendah dalam hal kasus diare yang ada di Indonesia [4].

2. Metodologi Penelitian

Metodologi penelitian menjelaskan langkah – langkah yang akan digunakan serta merupakan urutan-urutan yang dilakukan dalam melakukan sebuah penelitian dimana metodologi penelitian menjelaskan kerangka kerja penelitian dalam menyelesaikan permasalahan yang ada K-Medoids pada penelitian ini. Metodologi penelitian ini dilakukan dengan cara sistematis yang digunakan sebagai acuan dalam penelitian. Dalam metodologi penelitian ada kerangka kerja penelitian yang memilki beberapa tahapan yang berguna dalam menggambarkan proses penelitian dari awal hingga memperoleh hasil dari penelitian yang nantinya mudah dipahami. Metode K-Medoids merupakan teknik partisi klasik *clustering* yang mengelompokkan data set dari n objek ke dalam kelompok k dikenal apriori. Metode ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Objek yang terpilih untuk mewakili sebuah *cluster* disebut dengan medoid. *Cluster* di bangun dengan menghitung kedekatan yang dimiliki antara medoid dengan objek non-medoid. Langkah-langkah Metode K-Medoids:

- a) Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
- b) Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (1)$$

dimana $i=1, \dots, n$; $j=1, \dots, n$ dan p adalah banyak variable, serta V adalah matrik varian kovarian.

- c) Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru.

- d) Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
- e) Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
- f) Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing [5].

Rancangan atau model penelitian disajikan dalam rancangan *Flowchart* pada Gambar 1.

Gambar 1. Rancangan Penelitian

3. Hasil dan Pembahasan

Dalam penelitian ini, penulis mengambil sample data dengan menjumlahkan data penjualan vivo berdasarkan merk/bulan. Data hasil akhir yang nantinya akan diolah melakukan *clustering* dari berdasarkan merk yang dicantumkan dalam 3 *cluster* yaitu *cluster* tertinggi, *cluster* sedang dan *cluster* terendah. Berikut ini sample data yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Sample Data Yang Digunakan

Merk Hp	April	Mei	Juni	Juli	Agustus
S1 Pro 8+128	23	7	2	13	20
V17 Pro	4	3	6	46	64
V19 8+128GB	113	95	88	101	108
V19 8+256GB	9	6	9	39	72
Y11 2+32GB	71	118	98	111	128
Y12 3+32GB	907	712	751	1162	1085
Y12 3+64GB	5	1	25	31	31
Y154+64 GB	159	194	271	308	306
Y17 4+128GB	162	88	12	15	28
Y19 6+128GB	106	72	34	49	60
Y30 4+128GB	37	237	330	335	335
Y50 8+128GB	526	91	188	201	189
Y91C 2+32GB	3	401	522	532	525

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses *clustering* penjualan vivo menggunakan *algoritma* K-Medoids. Proses *clustering* dilakukan mulai dari penentuan data yang ingin di *cluster*. Dalam hal ini

variabel data yang ingin di *cluster*. Berikut adalah langkah-langkah penyelesaian yang dilakukan penulis dalam mengelompokkan Rumah Tangga Menguasai Komputer menggunakan algoritma K-Medoids:

- Menentukan Data yang diolah
Dalam penelitian, penulis mengambil *sample* data dengan menjumlahkan data penjualan *vivo* berdasarkan merk/bulan. Berikut tabel data yang akan digunakan.
- Menentukan Jumlah *Cluster*
Jumlah *Cluster* atau *k* yang digunakan sebanyak 3 yaitu *Cluster01*, *Cluster2* dan *Cluster3*.
- Menentukan Pusat Medoid
Menentukan Pusat Medoid dilakukan secara acak pada masing-masing *Cluster*. Pusat Medoid yang digunakan adalah :

Tabel 2. Pusat Medoid Iterasi 1

Y154+64 GB	159	194	271	308	306
V19 8+128GB	113	95	88	101	108
V19 8+256GB	9	6	9	39	72

- Menghitung Nilai *Euclidian*
Untuk menghitung jarak antara titik Medoid dengan titik tiap objek menggunakan Rumus *Euclidian Distance* :

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Maka perhitungan untuk jarak dari *Medoid* adalah sebagai berikut :

$$D_{x1,cl} = \sqrt{(23 - 159)^2 + (7 - 194)^2 + (2 - 271)^2 + (13 - 308)^2 + (120 - 306)^2} = 542,8139$$

$$D_{x2,cl} = \sqrt{(4 - 159)^2 + (3 - 194)^2 + (6 - 271)^2 + (46 - 308)^2 + (64 - 306)^2} = 507,876$$

Sampai $D_{x34,c3}$. Sehingga didapat tabel jarak dari *Medoid* dan mencari nilai minimal dari ketiga *Medoid*.

Tabel 3. Hasil Jarak *Medoid* Iterasi 1

Merk Hp	C1	C2	C3	Jarak Dekat	Cluster
S1 Pro 8+128	542,8139644	196,7943089	54,07402334	54,07402334	3
V17 Pro	507,8769536	178,9692711	10,67707825	10,67707825	3
V19 8+128GB	357,0140053	0	162,2837022	0	2
V19 8+256GB	503,5920968	173,5453831	0	0	3
Y11 2+32GB	337,5529588	53,78661544	165,8824885	53,78661544	2
Y12 3+32GB	1547,380044	1879,080626	1697,850406	1547,380044	1
Y12 3+64GB	523,2924612	187,8776197	44,38468204	44,38468204	3
Y154+64 GB	0	357,0140053	426,043425	0	1
Y17 4+128GB	491,3847779	148,3981132	179,0363092	148,3981132	2
Y19 6+128GB	448,841843	92,20629046	120,5985075	92,20629046	2
Y30 4+128GB	147,5940378	436,7940018	476,078775	147,5940378	1
Y50 8+128GB	421,0997507	444,0112611	566,0441679	421,0997507	1
Y91C 2+32GB	477,8315603	808,5431343	790,5264069	477,8315603	1

Dari Tabel 3. dapat dilihat hasil *cluster* yang diperoleh dari hasil iterasi Pengelompokkan merk *smartphone vivo* berdasarkan penjualan adalah *Cluster 1* yaitu Y12 3+32GB, Y154+64 GB, Y154+64 GB, Y30 4+128GB, *Cluster2* yaitu V19 8+128GB, Y11 2+32GB, Y17 4+128GB, Y19 6+128GB dan *Cluster3* yaitu S1 Pro 8+128, V17 Pro, V19 8+256GB, Y12 3+64GB. Dari tabel 4.3 dapat dilihat hasil jarak *Medoid* iterasi 1 akan digunakan untuk menentukan nilai *Cost*.

Nilai Jarak Dekat ini yang akan digunakan sebagai letak *cluster* atau kelompok berdasarkan nilai terkecil perbaris.

e) Menghitung Nilai *Cost*

Nilai *Cost* diperoleh dari total penjumlahan nilai jarak dekat *Medoid* yang diperoleh yaitu : $Cost = 2997,4321$

f) Lakukan Ulang Langkah 4 dan 5 dengan nilai Pusat *Medoid* baru.

Mengulangi proses menghitung nilai jarak dekat *Medoid* dengan nilai pusat *Medoid* baru secara acak yaitu :

Tabel 4. Pusat *Medoid* Iterasi 2

Y19 6+128GB	106	72	34	49	60
Y50 8+128GB	526	91	188	201	189
Y91C 2+32GB	3	401	522	532	525

Dari pengolahan menggunakan pusat *Medoid* baru diperoleh hasil sebagai Tabel 5. berikut :

Tabel 5. Hasil Jarak *Medoid* Iterasi 2

Merk Hp	C1	C2	C3	Jarak Dekat	Cluster
S1 Pro 8+128	122,6132	598,8038	824,9497	122,613213	1
V17 Pro	126,3883	594,1397	797,9323	126,3882906	1
V19 8+128GB	92,20629	444,0113	683,7763	92,20629046	1
V19 8+256GB	120,9711	588,6323	789,9025	120,9710709	1
Y11 2+32GB	126,115	477,153	649,3666	126,1150269	1
Y12 3+32GB	1963,335	1604,403	1131,64	1131,639519	3
Y12 3+64GB	128,4056	599,9783	806,5656	128,4056074	1
Y154+64 GB	448,8418	421,0998	421,8092	421,0997507	2
Y17 4+128GB	77,82031	473,2843	793,6259	77,82030583	1
Y19 6+128GB	0	490,1245	756,7932	0	1
Y30 4+128GB	526,3297	565,5732	317,5201	317,5200781	3
Y50 8+128GB	490,1245	0	770,5518	0	2
Y91C 2+32GB	898,0579	838,8337	0	0	3

Dari Tabel 5 dapat dilihat hasil jarak *Medoid* iterasi 2 untuk tahun 2018 yang akan digunakan untuk menentukan nilai *Cost*. Nilai Jarak Dekat ini yang akan digunakan sebagai letak *cluster* atau kelompok berdasarkan nilai terkecil perbaris. Dari tabel 4.4 dapat diperoleh nilai $Cost = 2664,7791$

g) Mencari nilai *S*

Nilai *S* diperoleh dengan cara mengurangi nilai *Cost* pada iterasi yang baru kepada iterasi awal. Jika nilai $S < 0$ maka pengolahan diteruskan dengan menggunakan nilai pusat *medoid* baru. Jika Nilai $S > 0$ atau nilai *Cost* iterasi baru lebih besar daripada nilai *Cost* iterasi lama maka proses dihentikan. Karena Nilai *Cost* Baru > Nilai *Cost* Lama, maka iterasi dihentikan dan iterasi terletak di iterasi 1

h) Mencari *Cluster* atau Pengelompokkan

Untuk menentukan *Cluster* diperoleh dengan menentukan nilai *Cluster* 1, 2 dan 3 terdekat dengan Nilai Jarak *Distance* sehingga diperoleh *Cluster* iterasi yang terletak pada iterasi 1 adalah sebagai berikut :

Tabel 6. *Cluster* Hasil Iterasi

Merk Hp	Cluster
S1 Pro 8+128	1
V17 Pro	1
V19 8+128GB	1
V19 8+256GB	1

Merk Hp	Cluster
Y11 2+32GB	1
Y12 3+32GB	3
Y12 3+64GB	1
Y154+64 GB	2
Y17 4+128GB	1
Y19 6+128GB	1
Y30 4+128GB	3
Y50 8+128GB	2
Y91C 2+32GB	3

Dari Tabel 6. dapat dilihat hasil *cluster* yang diperoleh dari hasil iterasi Pengelompokan merk *smartphone* vivo adalah *Cluster* 1 yaitu S1 Pro 8+128, V17 Pro, V19 8+128GB, V19 8+256GB, Y11 2+32GB, Y12 3+64GB, Y17 4+128GB, Y19 6+128GB . *Cluster* 2 yaitu Y50 8+128GB dan *Cluster* 3 yaitu Y12 3+32GB, Y30 4+128GB, Y91C 2+32GB .

4. Kesimpulan

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa: Diperoleh dari hasil iterasi Pengelompokan merk *smartphone* vivo berdasarkan penjualan April-Agustus 2020. Diperoleh *Cluster* 1 yaitu S1 Pro 8+128, V17 Pro, V19 8+128GB, V19 8+256GB, Y11 2+32GB, Y12 3+64GB, Y17 4+128GB, Y19 6+128GB. *Cluster* 2 yaitu Y50 8+128GB dan *Cluster* 3 yaitu Y12 3+32GB, Y30 4+128GB, Y91C 2+32GB. Dengan hasil yang sama antara perhitungan manual dan dengan Rapidminer maka algoritma K-Medoid dapat diterapkan dalam pengelompokan *smartphone* Vivo berdasarkan data penjualan dengan presentase keakuratan sebesar 100%.

Daftar Pustaka

- [1] Defiyanti, S., Jajuli, M., & W, N. rohmawati. (2017). Optimalisasi K - Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion. *TEKNOSI*, 03(01), 211–218.
- [2] Lambertus, S. (2018). *Optimasi Kinerja Firewall Menggunakan Teknik Data Mining*.
- [3] P a n g a n T. (2015). Analisis Pengelompokan Daerah Menggunakan Metode Non-Hierarchical Partitioning K-Medoids Dari Hasil Komoditas Pertanian Tanaman Pangan (Studi Kasus Kabupaten/Kota Se-Jawa Tengah Tahun 2009 – 2013). *None*, 4(4), 825–836.
- [4] F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723–732.
- [5] Pulungan, N., Suhada, S., & Suhendro, D. (2019). Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 329–334. <https://doi.org/10.30865/komik.v3i1.1609>
- [6] Riyanto, B. (2019). Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokkan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 562–568. <https://doi.org/10.30865/komik.v3i1.1659>
- [7] Ropianto, M. (2016). Pemahaman Penggunaan Unified Modelling Language. *Jt-Ibsi*, 1(1), 43–50.
- [8] Silitonga, D. A., Windarto, A. P., & Hartama, D. (2019). Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah

- Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019 ISBN*, 313–318.
- [9] Sindi, S., Ratnasari, W., Ningse, O., Sihombing, I. A., Zer, F. I. R. H., Hartama, D., & Kunci, K. (2020). *Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia*. 4(1), 166–173.
 - [10] Tanjung, Q. (2017). *Sistem Penunjang Keputusan Kelayakan Kredit Motor Menggunakan Metode Naive Bayes Pada NSC Finance Cikampek*.
 - [11] Urva, G., & Siregar, H. F. (2015). Pemodelan UML E-Marketing Minyak Goreng. *JTSI*, 1(9), 92–101.
 - [12] Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi. *Jurnal Terapan Sains & Teknologi*, 1(3), 54–69.
 - [13] Yudistira, P. (2019). *Jurnal selulosa*. 9(1).
 - [14] Yusuf, M., Danuri, D., & Jaroji, J. (2019). Aplikasi Penjualan Tiket Ro-Ro Bengkalis Berbasis Android. *INOVTEK Polbeng - Seri Informatika*, 4(2), 201. <https://doi.org/10.35314/isi.v4i2.1077>