

Penerapan *Clustering Algorithm* Untuk Mendukung Promosi Server Pulsa Reload

Firman Nurdiyansyah¹, Samsul Arifin², Fitri Marisa³
^{1,3} Teknik Informatika, Universitas Widyagama,
² Samsul Arifin, Teknik Informatika, STIMATA Malang
e-mail : ¹ firmansyah@gmail.com, ² samsularifin@stimata.ac.id,
³ fitrimarisa@gmail.com

ABSTRACT

Promotion is necessary for every business field, including server of reload pulsa. In the strategy of promotion, one of the necessary thing is how the data distribution of the customer and its characteristics. One method that can be applied to explore the distribution of customers and transaction characteristics is to group customers using clustering algorithm technique. The technique of clustering algorithm is suitable to be applied because it can bring up some groups that have previously unknown attributes. Thus it can be used as new knowledge for the owner to be used according to needs. This data training take attributes such as customer, transaction amount, and address. The number of centroid specified amounts to 3. The grouping results stop at the 4th iteration with the ratio value 0.0553 in the 3rd and 4th iterations. From the result of iteration then formed 3 groups namely: C1 {C, D, H}, C2 {A, B, E}, C3 {F, G}.

Keywords - clustering algoritm, reload pulse, K-Means, applied algorithm

ABSTRAK

Promosi sangat diperlukan bagi setiap bidang usaha, salah satunya server pulsa reload. Dalam strategi promo hal yang perlu diketahui salah satunya adalah data sebaran pelanggan dan karakteristiknya. Salah satu metode yang dapat diterapkan untuk menggali sebaran pelanggan dan karakteristik transaksinya adalah dengan mengelompokkan pelanggan menggunakan teknik clustering Algorithm. Teknik cluetering Algorithm cocok dipilih karena dapat memunculkan beberapa kelompok-kelompok yang memiliki keterhubungan yang sebelumnya tidak diketahui atributnya. Dengan demikian dapat dijadikan sebagai pengetahuan baru bagi pihak pemilik server pulsa reload untuk dapat dimanfaatkan sesuai kebutuhannya. Dalam data training ini mengambil atribut-atribut antara lain customer, jumlah transaksi, dan alamat. Jumlah centroid yang ditentukan berjumlah 3. Hasil pengelompokan berhenti pada iterasi ke-4 dengan nilai rasio 0.0553 di iterasi ke-3 dan ke-4. Dari hasil iterasi maka terbentuk 3 kelompok yaitu : C1 {C,D,H}, C2 {A,B,E}, C3 {F,G}.

Kata Kunci - clustering algoritm, server pulsa reload, K-Means.

1. PENDAHULUAN

Promosi sangat berperan penting bagi pengelola usaha. Promosi dilakukan dengan berbagai macam cara. Namun selama ini promosi khususnya yang dilakukan UKM adalah dengan cara konvensional menyebar brosur ke semua lapisan masyarakat tanpa melihat segmen. Big data adalah salah satu trend dari teknologi yang merupakan konskuensi dari transaksi data online. Server pulsa reload merupakan salah satu usaha transaksi online yang banjir data. Menurut [1] Server pulsa reload merupakan perusahaan bidang teknologi informasi yang menjual pulsa elektrik. cara kerja pulsa reload adalah dengan mendistribusikan pulsa elektrik dari operator seluler dari operator yang dijembatani distributor yang ditunjuk. Pulsa dimasukkan dalam chip-chip kemudian diolah dalam sms gateway an teknologi parsing yang didukung beberapa hardware dan software diantaranya modem, kabel data, dan aplikasi parsing.

Data adalah fakta yang dapat direkam dan memiliki arti secara implisit. Kumpulan data yang terhubung secara implisit itu disebut Database. [1] Menurut Ramakrishnan dan Gehrke (2003) dalam [2] mengatakan "basis data merupakan kelompok dari data, yang umumnya mendeskripsikan aktivitas sebuah organisasi / lebih dari beberapa organisasi yang berhubungan". DBMS (Database Management Systems) adalah kelompok perintah yang menjadikan user dapat menciptakan dan memelihara database. [1] Dapat dikatakan DBMS

adalah seperangkat GPSS (General Purpose Software System) yang memiliki fungsi untuk memberikan fasilitas proses-proses dalam database seperti sharing, definition, create, sharing, dan database manipulation.

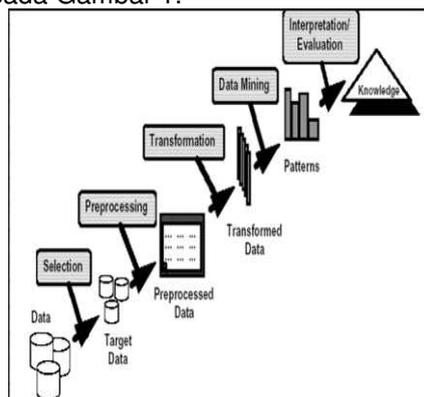
Entity Relationship (ER). Menurut [3] "Entitas adalah objek dalam dunia nyata yang dapat dibedakan dari objek lain. Contoh: mahasiswa, dan nilai. Relasi adalah hubungan antara dua atau lebih entitas. Misal: relasi menghubungkan mahasiswa dengan mtakuliah yang diambilnya." Atau dapat dikatakan "Entity relationship (ER) data model berdasarkan pada bagaimana persepsi terhadap kondisi nyata dengan kumpulan basis obyek atau hubungan antar obyek dan entitas [2].

Data mining adalah proses dalam rangka mencari hubungan antar item hingga terbentuk pola pengetahuan yang baru dari data yang berukuran besar dan dirproses dengan beberapa teknik yang melibatkan ilmu-ilmu antara lain statistik, dan matematika. [4]. Datamining mempunyai beberapa tahapan yaitu: Pembersihan data, seleksi data, transformasi dan integrasi data, proses datamining, evaluasi dan entepretasi. [5] Analisis cluster merupakan metode untuk mengelompokkan data dimana dikelompokkan berdasarkan karakter yang mirip dalam kelompok data. [6] Dalam pengembangannya server pulsa makin menjamur dan seiring kebijakan klasterisasi stok oleh distributor sehingga menyempitkan skup pangsa pasar server dan juga menambah persaingan yang sangat ketat antar server dalam suatu jangkauan wilayah klaster. Untuk itu diperlukan strategi promo yang efektif agar dapat mendukung peningkatan penjualan. Strategi promo selama ini adalah menggunakan metode menyebar brosur masyarakat tanpa melihat segmen dan sasaran. Hasil yang diperoleh tidak optimal dan sedikit mempengaruhi peningkatan transaksi. Oleh karena itu diperlukan strategi promo yang dapat mengoptimalkan promosi salah satunya adalah dengan menggali kelompok pelanggan yang berhubungan dengan jumlah transaksinya. Dalam pengelompokan ini teknik yang digunakan adalah clustering algorithm dengan menggunakan algoritma K-Means.

Dari pengujian penerapan algoritma menghasilkan pengelompokan dengan clustering algorithm yang terbentuk kelompok-kelompok dari atribut yang berhubungan yang sebelumnya syarat-syaratnya telah ditargetkan. Dari hasil proses selection dihasilkan data training dengan atribut anara lain customer, jumlah transaksi, dan alamat. Kemudian ditentukan jumlah centroid yaitu $K=3$. Iterasi dilakukan hingga 4 (empat) kali. Hasil pengelompokan berhenti pada iterasi ke-4 yaitu dengan nilai rasio 0.0553 di iterasi ke-4 yang sama dengan iletasi ke-3. Dari hasil pengelompokan maka terbentuk 3 kelompok customer yaitu : C1 {C,D,H}, C2 {A,B,E}, C3 {F,G}.

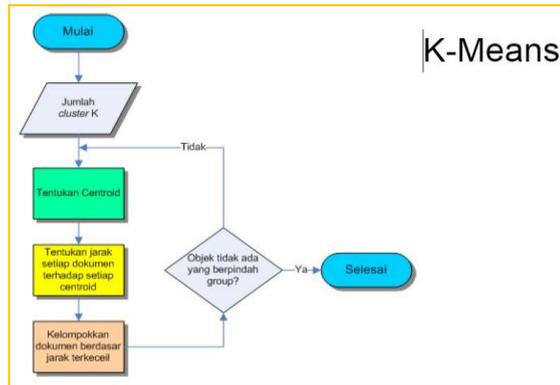
2. METODE PENELITIAN

Metode penelitian dalam penelitian ini menerapkan tahapan dari proses data mining yaitu meliputi: Data selection, preprocessing, transformation, data mining, dan interpretation/evaluation data. Tahapan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Data Mining

Dalam tahapan penerapan clustering algorithm telah digambarkan dalam flowchart pada Gambar 1.1 sebagai berikut:



Gambar 2. Flowchart Clustering Algorithm

3. HASIL DAN PEMBAHASAN

Berikut disajikan hasil dan pembahasan dimana tahapan ini mengikuti metode penelitian yang digunakan yaitu berdasarkan pendekatan tahapan proses data mining yaitu:

1. Data Selection dan preprocessing

Melakukan seleksi dari atribut terhadap query-query yang telah dimiliki oleh aplikasi server. Pemilihan ini berdasarkan tabel-tabel yang sudah terbentuk. Beberapa contoh tabel dapat dilihat pada gambar 2 dan gambar 3.

Name	Type	Null	Default
Kode	varchar(6)	No	
Tgl	date	Yes	
Nama	varchar(35)	Yes	
Alamat	varchar(55)	Yes	
Saldo	double	Yes	0
PJ	varchar(4)	No	

Gambar 3. Tabel Reseler

Name	Type	Null	Default
Faktur	int(10)	No	0
Tgl	date	Yes	
TglJam	datetime	Yes	
Reseller	varchar(6)	Yes	
Jumlah	double(12,2)	No	0.00
Lunas	double(12,2)	No	0.00
Operator	varchar(20)	Yes	
Waktu	varchar(20)	Yes	

Gambar 4. Tabel Deposit_reseler

Kemudian diambil data yang relevan dengan penggalan data sehingga terbentuk seperti Tabel 1

Tabel 1: Data yang relevan hasil seleksi

Pelanggan	Jumlah Transaksi / minggu	Alamat
A	75	Malang kota
B	100	Malang Kab.
C	125	malang kota
D	180	Malang kota
E	55	Malang kota
F	50	Luar Malang
G	35	Luar Jawa
H	200	Malang kab.

2. Transformation data

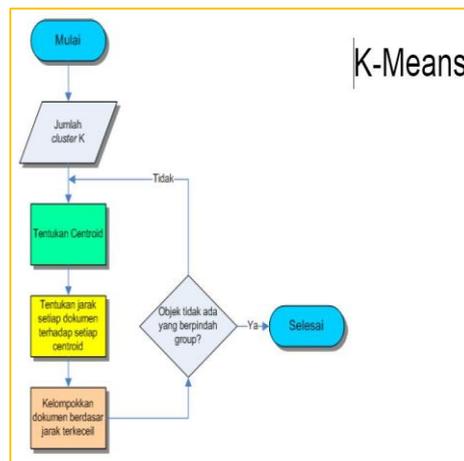
Tahap ini adalah menggabungkan data ke dalam bentuk yang sesuai sehingga dapat dilakukan pemrosesan berikutnya. Contoh atribut [kota] dikonversikan menjadi angka yang ditentukan skornya untuk dapat sejajar dengan atribut [jumlah transaksi]. Hasil dapat dilihat pada tabel 2.

Tabel 2. Data hasil transformasi

pelanggan	jumlah trx / minggu	domisili
A	75	100
B	100	75
C	125	100
D	180	100
E	55	100
F	50	25
G	35	50
H	200	75

3. Datamining dengan clustering K-Means.

Memberlakukan pengurutan proses mining dalam algoritma K-Means Sesuai urutan pada Gambar 4.



Gambar 4. Tahap Clustering K-Means

- Penentuan jumlah centroid, dinotasikan dalam K, dimana dalam ujicoba ini ditentukan K=3.
- Menentukan centroid (titik pusat) sesuai dengan jumlah cluster secara acak. Karena K=3 maka centroid yang ditunjuk secara acak jumlahnya 3 yaitu B, E, F yang dinotasikan sebagai berikut: M1 {100,75}, M2{55,100}, M3{50,25}
- Menghitung jarak yaitu, melakukan perhitungan jarak antara data dan centroid dengan menggunakan rumus ecludien yaitu sebagai berikut:

$$D(X_2, X_1) = || X_2 - X_1 ||_2 = \sqrt{\sum_{i=1}^p |X_{2j} - X_{1j}|^2}$$

D adalah jarak antara X1 dan X2.

Hasil iterasi ke-1 dapat dilihat pada Tabel 3.

Tabel 3. Hasil iterasi ke-1

pelanggan	C1	C2	C3	jarak	jarak^2	kelas
A	35.35533906	20	79.05694	20	400	C2
B	0	51.4782	70.71068	0	0	C1
C	35.35533906	70	106.066	35.35534	1250	C1
D	83.81527307	125	150.0833	83.81527	7025	C1
E	51.4781507	0	75.16648	0	0	C2
F	70.71067812	75.1665	0	0	0	C3
G	69.64194139	53.8516	29.15476	29.15476	850	C3
H	100	147.139	158.1139	100	10000	C1

- Menentukan BCV (Between Cluster Variation) dan WCV (Within Cluster Variation) bertujuan untuk mencari rasio dari keduanya.
 $BCV = d(m_1, m_2) + d(m_1, m_3) + d(m_2, m_3)$

$$WVC = \text{Jarak_centroid_terkecil_ke-1}^2 + \text{Jarak_centroid_terkecil_ke-N}^2$$

$$\text{Rasio} = \text{BCV} / \text{WCV}$$

$$\text{Rasio} = 197.35 / 19525$$

$$= 0.0101$$

Pada iterasi berikutnya untuk menentukan centroid adalah dengan menghitung rata-rata dari kelompok yang terbentuk pada iterasi pertama. Formula yang digunakan adalah sebagai berikut:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

Lihat tabel 4:

Tabel 4. tabel penentua centroid ke-2

cluster 1		
pelanggan	jml trx	kota
B	100	75
C	125	100
D	180	100
H	200	75
mean	151.25	87.5
cluster 2		
A	75	100
E	55	100
Mean	65	100
cluster 3		
F	50	25
G	35	50
mean	42.5	37.5

Didapat centroid baru :
 m1(151.25, 87.5) m2(37.5,65,100), m3(42.5)

hasil iterasi ke-2 dapat dilihat pada tabel 5:

Tabel 5. hasil iterasi ke-2

pelanggan	C1	C2	C3	jarak	jarak^2	kelas
A	77.2677973	10	70.44501	10	100	C2
B	52.75236962	43.0116	68.64765	43.01163	1850	C2
C	29.07425837	60	103.5012	29.07426	845.3125	C1
D	31.34984051	115	151.0381	31.34984	982.8125	C1
E	97.05829434	10	63.73774	10	100	C2
F	118.9866064	76.4853	14.57738	14.57738	212.5	C3
G	122.1487311	58.3095	14.57738	14.57738	212.5	C3
H	50.32705535	137.295	161.9027	50.32706	2532.813	C1

$$\text{Rasio BCV/WVC} = 273.27 / 6835.93$$

$$= 0.0399$$

Pada iterasi ke-3 menghasilkan rasio sebesar 0.0553.

Pada iterasi ke-4 menghasilkan rasio 0.0553 yaitu sama dengan rasio pada iterasi ke-3, maka proses iterasi dihentikan.

Tabel iterasi yang terbentuk dapat dilihat pada tabel 6

Tabel 6. hasil iterasi ke-4

pelanggan	C1	C2	C3	jarak	jarak^2	kelas
A	93.70461865	8.49837	70.44501	8.498366	72.22222	C2
B	70.33649282	28.6744	68.64765	28.67442	822.2222	C2
C	44.12734098	49.0465	103.5012	44.12734	1947.222	C1
D	14.33720878	103.669	151.0381	14.33721	205.5556	C1
E	113.6392929	23.214	63.73774	23.21398	538.8889	C2
F	135.8205515	71.8022	14.57738	14.57738	212.5	C3
G	139.6921218	58.9256	14.57738	14.57738	212.5	C3
H	35.78485092	124.454	161.9027	35.78485	1280.556	C1

Untuk itu dapat dideskripsikan cluster yang terbentuk adalah: C1 {C,D,H}, C2 {A,B,E}, C3 {F,G}.

4. Interpretasi pengetahuan

Tabel 7. *Data Training*

pelanggan	jumlah trx / m	domisili
A	75	100
B	100	75
C	125	100
D	180	100
E	55	100
F	50	25
G	35	50
H	200	75

Dalam tabel 7 diinterpretasikan hasil ujicoba penelitian ini dengan kesimpulan sebagai berikut: (C1) {C,D,H} adalah kelompok pelanggan yang memiliki transaksi banyak dan berdomisili di Malang Kota, C2 {A,B,E} adalah kelompok pelanggan dengan transaksi sedang yang berdomisili di Malang kota dan Malang Kabupaten, dan C3 {F,G} adalah kelompok pelanggan dengan jumlah transaksi sedikit dan berdomisili di luar malang dan luar jawa.

4. KESIMPULAN.

Dengan pendekatan clustering algorithm maka memungkinkan untuk memberikan pengetahuan baru bagi server pulsa reload untuk melakukan strategi promo. 3 Kelompok yang dihasilkan tersebut dapat dijadikan salah satu pegangan untuk memaksimalkan dan menyeimbangkan promosi di berapa daerah.

5. SARAN.

Penelitian ini masih terdapat banyak kekurangan, oleh karena itu perlu dikembangkan metode yang berbeda untuk menunjang strategi promo dan perlu berbandingan hasil dari metode-metode yang sudah diterapkan sehingga dapat dipilih yang lebih efektif.

UCAPAN TERIMAKASIH.

Ucapan terimakasih disampaikan kepada Kemenristek Dikti yang telah memberikan hibah dana Penelitian Dosen Pemula sehingga penelitian ini dapat terwujud. Terimakasih juga disampaikan kepada Universitas Widyagama melalui LPPM yang telah memberi kesempatan proposal untuk mengikuti kompetisi Hibah Ditlitabmas RISTEKDIKTI.

DAFTAR PUSTAKA

- [1] C. Setyo, Panduan Praktis Pemrograman Database Menggunakan MySQL dan Java, Bandung: Bandung, 2006, p. 10.
- [2] J. Simarmata and I. Paryudi, Basis Data, Yogyakarta: Andi Offset, 2006.
- [3] D. T. Octafian, "Desain Database Sistem Informasi Penjualan Barang," *Teknomatika*, vol. 1, no. 2, 2011.
- [4] A. S. Khazari, F. Marisa and I. D. Wijaya, "SISTEM REKOMENDASI PENENTUAN JUDUL SKRIPSI MENGGUNAKAN ALGORITMA DECISION TREE," *Jurnal Teknologi dan Manajemen Informatika*, vol. 3, no. 1, p. 124, 2017.
- [5] F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *Jurnal Teknologi Informasi*, vol. 4, no. 2, 2013.
- [6] R. Sitepu, I. Irmeilyana and B. Gultom, "Analisis Cluster Terhadap Tingkat Pencemaran Udara Pada Sektor Industri di Sumatra Selatan," *Jurnal Penelitian Sains*, vol. 14, no. 3, p. 12, 2011.