



CLUSTERING ALGORITHM UNTUK PENGELOMPOKAN PELANGGAN DALAM BIDANG USAHA SERVER RELOAD

Firman Nurdiyansyah¹, Samsul Arifin², Fitri Marisa³

^{1,3}Teknik Informatika, Universitas Widyagama

²Teknik Informatika, STIMATA Malang

Email: ¹firmansyah@gmail.com, ²samsularifin@stimata.ac.id, ³fitrimarisa@gmail.com

Abstrak

Server pulsa menjadi teknologi perdagangan modern masyarakat di era ini dan pelakunya kebanyakan pengusaha kecil. Dalam pelaksanaannya terdapat beberapa permasalahan. Salah satu permasalahan server pulsa adalah strategi promo. Strategi promo dalam usaha server pulsa umumnya menggunakan metode klasik dengan cara membagikan brosur ke siapa saja tanpa mempertimbangkan sasaran. Hasil yang diperoleh juga tidak efektif. Dalam makalah ini akan dibahas salah satu ide tentang bagaimana memanfaatkan data yang ada untuk strategi promo. Setiap hari telah banyak data transaksi yang dihasilkan dalam server pulsa. Data-data ini sebagian besar menjadi data-data pasif yang tidak memiliki manfaat bagi pemilik usaha, sedangkan keberadaan data sangat banyak dan menumpuk dalam area backup data. Berdasarkan dua hal yang telah disampaikan maka perlu dilakukan studi pendahuluan tentang pentingnya pengelompokan segmen pelanggan menggunakan teknik clustering algorithm. Teknik cocok dipilih karena mengingat karakteristik clustering yang dapat memunculkan kelompok-kelompok yang sebelumnya tidak diketahui atributnya sehingga dapat menjadi pengetahuan bagi pihak pemilik data dalam memanfaatkannya. Sebagai ilustrasi rancangan, maka atribut-atribut yang terlibat antara lain : pelanggan, jumlah transaksi, dan domisili untuk membantu kelompok-kelompok cluster.

Kata kunci : *clustering algorithm, server pulsa, big data*

Abstract

Server pulsa became a modern trading technology community in this era and the perpetrators are mostly small entrepreneurs. In the implementation there are several problems. One of the problems of the pulse server is the promo strategy. The promo strategy in the pulse server business generally uses the classical method by distributing flyers to anyone without considering the target. The results obtained are also not effective. In this paper we will discuss one of the ideas on how to use the existing data for the promo strategy. Every day has a lot of transaction data generated in the pulse server. These data mostly become passive data that has no benefits for business owners, while the existence of data is very much and accumulate in the area of data backup. Based on two things that have been submitted, it is necessary to do a preliminary study about the importance of grouping customer segment using clustering algorithm technique. The suitable technique is chosen because it is considering the clustering characteristics that can bring up previously unknown groups of attributes so that it can become knowledge for the data owner in using it. As an illustration of the design, the attributes involved are: customer, transaction number, and domicile to establish cluster groups.

Keywords: *clustering algorithm, server pulsa,, big data*



I. PENDAHULUAN

Big data menjadi salah satu trend teknologi di masa sekarang dan ini salah satunya diakibatkan maraknya usaha bidang online. Salah satu usaha online adalah server reload. Server reload adalah perusahaan jasa di bidang IT yang melayani penjualan pulsa secara elektronik. Adapun cara kejanya adalah dengan membagi pulsa elektrik dari operator seluler yang ber dari dealer asal dari dealer resmi dengan media chip-chip yang berisi pulsa, setelah itu diolah dalam sms gateway dan integrasi beberapa perangkat antara lain modem sebagai tempat chip, mekanisme parsing. (Marisa, 2015).

Dalam perjalanannya terdapat beberapa permasalahan dalam server pulsa. Namun masalah yang paling krusial adalah tentang promosi. Strategi promo yang digunakan selama ini dalam server reload secara umum menggunakan metode tradisional dengan menyebar brosur ke end user dan konter-konter tanpa mempertimbangkan sasaran. Hasil yang diperoleh tidak efektif dan tidak terlalu memberi pengaruh signifikan terhadap peningkatan pelanggan maupun transaksi.

Berangkat dari permasalahan tersebut maka perlu ada gagasan untuk meningkatkan strategi promo terhadap server reload. Salah satu ide yang dapat diimplementasikan adalah pengelompokan pelanggan. Salah satu teknik pengelompokan yang dapat digunakan adalah clustering. Clustering memungkinkan untuk terbetuknya kelompok-kelompok dari data yang sebelumnya sudah ditargetkan syarat-syaratnya. Kondisi eksisting bahwa Setiap hari banyak data yang dihasilkan dari transaksi. Sementara perlakuan terhadap data selama ini hanya disimpan dalam backup tanpa digali kebermanfaatannya. Oleh karena itu

maka perlu dilakukan studi pendahuluan tentang pentingnya pengelompokan segmen pelanggan menggunakan teknik clustering algorithm. Teknik cocok dipilih karena mengingat karakteristik clustering yang dapat memunculkan kelompok-kelompok yang belum diketahui atributnya sehingga dapat menjadi pengetahuan bagi pihak pemilik data dalam memanfaatkannya. Sebagai ilustrasi rancangan, maka atribut-atribut yang terlibat antara lain : pelanggan, jumlah transaksi, dan domisili untuk membantu kelompok-kelompok cluster.

II. TINJAUAN PUSTAKA

A. Pengertian Database

Data merupakan fakta yang bisa direkam dan mempunyai arti secara implisit. Kumpulan data yang terhubung secara implisit itu dinamakan Database. [1] Menurut Ramakrishnan dan Gehrke (2003) dalam [2] menyatakan “basis data adalah kelompok data, dimana biasanya menjelaskan aktivitas organisasi ataupun lebih dari satu organisasi yang berhubungan”. DBMS (Database Management Systems) yaitu kelompok program yang mengkondisikan user untuk menciptakan dan memelihara kumpulan data (database). [1] Dengan kata lain DBMS merupakan seperangkat GPSS (General Purpose Software System) yang mempunyai fungsi untuk memberi wadah proses-proses dalam kumpulan data seperti definisi, pembuatan, berbagi, ataupun manipulasi database.

Dalam bidang database, dikenal istilah Entity Relationship (ER). Menurut [3] “Entitas merupakan sebuah objek dalam dunia nyata yang dapat dibedakan dari objek lain. Misal: pelanggan, dan barang. Relasi merupakan hubungan antara dua atau lebih entitas. Contoh: relasi menghubungkan pelanggan dengan barang yang dibelinya.”. dengan kata lain

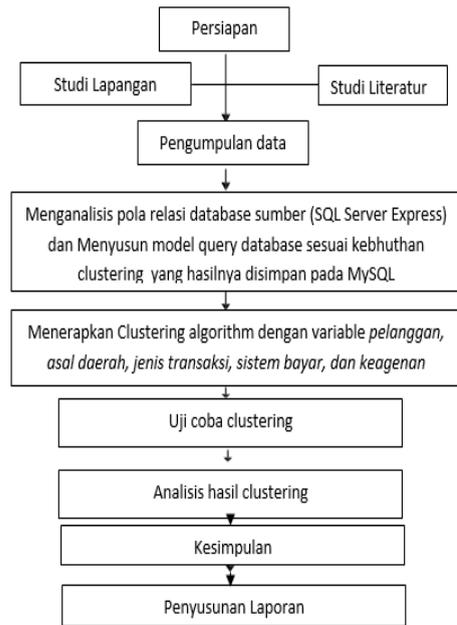
“Entity relationship (ER) data model berdasarkan pada bagaimana mempersepsikan kondisi riil dengan kumpulan basis obyek atau entitas dan hubungan antar objek [2].

B. Data Mining

Data mining merupakan pola untuk menemukan hubungan antar item sehingga membentuk pola pengetahuan baru dari data yang besar dan diselesaikan dengan teknik tertentu yang melibatkan ilmu statistik, dan matematika. [4]. Tahapan data mining terdiri dari 6 tahap yaitu: Data cleaning, data selection, data transformation, mining process, evaluation. [5] Analisis cluster adalah sebuah cara untuk mengelompokkan data dengan karakter yang mirip ke dalam kelompok. [6]

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah menggunakan pada pendekatan Waterfall dengan konsep dasar SDLC (System Development Life Cycle). Terdapat beberapa tahapan antara lain: Planning, Analysis, Implementation. Kegiatan ini diawali dengan mengkaji data yang diperoleh dari transaksi pulsa secara online. Kemudian dipilih atribut-atribut yang cocok untuk pengolana data pelanggan. Selanjutnya dilakukan penghitungan cluster dengan algoritma K-meas. Atribut dipilih sesuai dengan kebutuhan data yang nanti akan dihasilkan yaitu” nama pelanggan, asal daerah, kejis transaksi sistem bayar dan keagenan. Secara bagan dapat diliha paa Gambar 1.



Gambar 1: Metode Penelitian

IV. HASIL DAN PEMBAHASAN

Langkah-langkah clustering adalah sebagai berikut:

1. Seleksi data.

Dari hasil query dihasilkan beberapa tabel dan dari tabel tersebut dapat dipilih mana-mana yang menjadi atribut, dimana tabel-tabel bisa 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 2: 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	

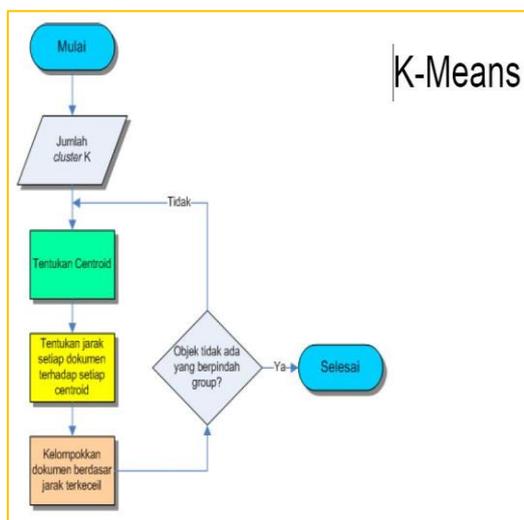
Gambar 3: Tabel Deposit_reseller

2. Integrasi data

Dalam integrasi data ini dilakukan penyamaan format terhadap atribut yang dipilih. Contoh atribut [kota] dikonversikan menjadi angka untuk dapat sejajar dengan atribut [jumlah transaksi].

3. Datamining dengan pendekatan clustering.

Dalam tahap ini mulai diperlakukan tahap demi tahap proses clustering dengan menggunakan algoritma K-Means. Urutan tahap dapat dilihat pada gambar 4.



Gambar 4: Tahap Clusteri dalam K-Means

a. Penentuan jumlah cluster adalah dinotasikan dalam K dan dapat ditentukan secara bebas. Contoh K=3

b. Menentukan centroid yaitu menentukan titik pusat sesuai dengan jumlah cluster secara acak. Jika K= 3 maka centroid yang ditunjuk secara acak berjumlah 3.

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

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

c. Menghitung jarak yaitu, melakukan perhitungan jarak antara data dan centroid dengan menggunakan rumus eucliden 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.

d. Menentukan BCV (Between Cluster Variation) dan WCV (Within Cluster Variation) yang tujuannya adalah untuk mencari rasio antar keduanya.

$$BCV = d(m_1, m_2) + d(m_1, m_3) + d(m_2, m_3)$$

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

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

e. Melakukan iterasi hingga rasio hasil iterasi tabel terakhir sama dengan 1 tabel sebelumnya, dengan kata lain centroid sudah tidak ada perubahan sehingga pola luster sudah ditemukan.

$$\text{Contoh : } C1 = \{A, C\}, C2 = \{E, G, H\}, C3 = \{F, B\}$$

4. Interpretasi pengetahuan

Dalam interpretasi hasil ini disimpulkan dengan melihat data training dan dapat dibaca hasil pengelompokannya dan



dihubungkan dengan konten dari data-data tersebut.

Contoh: Dari hasil cluster telah terbentuk tiga kelompok dengan penjabaran kelompok pertama (C1) adalah kelompok mahasiswa dengan nilai IPK stabil dan jumlah absensi yang minim, C2 adalah kelompok mahasiswa yang IPK sedang dengan absensi yang minim, dan C3 adalah kelompok mahasiswa dengan IPK kecil dengan jumlah absen yang cenderung banyak.

V. KESIMPULAN DAN SARAN

Dengan penerapan datamining pendekatan cluster maka dimungkinkan dapat menggali data untuk dihasilkan pengelompokan sesuai dengan jenis dan karakteristik data yang dimiliki untuk digunakan sebagai rekomendasi terhadap kebijakan-kebijakan suatu organisasi.

Penelitian ini masih berupa studi kajian sebagai langkah awal dalam penerapan datamining khususnya teknik cluster, untuk itu perlu pengembangan yang sifatkan melanjutkan dari ide dan rancangan yang sudah disampaikan dalam artikel ini.

REFERENSI

- C Setyo, Panduan Praktis Pemrograman Database Menggunakan MySQL dan Java, Bandung: Bandung, 2006, p. 10.
- J. Simarmata and I. Paryudi, Basis Data, Yogyakarta: Andi Offset, 2006.
- D. T. Octafian, "Desain Database Sistem Informasi Penjualan Barang," *Teknomatika*, vol. 1, no. 2, 2011.
- 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.
- F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *Jurnal Teknologi Informasi*, vol. 4, no. 2, 2013
- 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.