

SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN MODEL QRF PADA PERUSAHAAN RINTISAN PENYEDIA TENAGA KERJA

Sari Anggar Kusuma Melati¹, Arief Wibowo²

^{1,2}Magister Ilmu Komputer Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta

¹sarianggarkm@gmail.com, ² arief.wibowo@budiluhur.ac.id

ABSTRAK

Sulitnya mendapatkan pekerjaan yang sesuai dengan minat dan spesialisasi seorang pekerja, serta sulitnya perusahaan mendapatkan seorang pekerja yang sesuai dengan kebutuhan perusahaan menyebabkan menjamurnya perusahaan konsultan atau penyedia tenaga kerja di Indonesia saat ini. Dengan semakin banyaknya pilihan perusahaan penyedia tenaga kerja, tentu daya saing bisnis industri sumber daya manusia semakin tinggi. Sehingga perlu dilakukan analisa untuk menentukan strategi bisnis yang tepat, seperti menentukan tujuan promosi perusahaan. Salah satunya adalah melakukan analisa terhadap segmentasi pelanggan yang sudah bekerja sama. Riset ini berhasil memodelkan segmentasi pelanggan berdasarkan teknik data mining klusterisasi dengan menggunakan data mining algoritma K-Means. Proses pemodelan QRF (*Quantity, Recency, Frequency*) yakni menganalisa perilaku pelanggan dari banyak nya jumlah permintaan di setiap transaksi yang dilakukan dalam rentang waktu tertentu, juga recency sebagai identifikasi rentang waktu transaksi terakhir, serta jumlah dari transaksi yang dilakukan dalam jangka waktu tertentu. Peneliti dilakukan dengan periode data selama satu tahun dengan menganalisis aktivitas pelanggan di perusahaan rintisan penyedia tenaga kerja selama tahun 2019, pada 86 pelanggan aktif. Berdasarkan hasil analisa yang diperoleh, segmentasi pelanggan pada dua kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index (DBI)* memperoleh nilai -0,482, sedangkan segmentasi pelanggan pada tiga kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index (DBI)* memperoleh nilai -0,469, dan segmentasi pelanggan pada empat kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index (DBI)* memperoleh nilai -0,526.

Kata Kunci— pelanggan, clustering, algoritma k-means, DBI, QRF

ABSTRACT

*The difficulty of getting a job that is in accordance with the interests and specialization of a worker, as well as the difficulty of the company getting a worker who suits the needs of the company causes the mushrooming of consulting firms or labor providers in Indonesia today. With the increasing number of companies providing labor, of course the competitiveness of the business industry in the human resources is increasingly high. So it needs to be analyzed to determine the right business strategy, such as determining the company's promotion goals. One of them is analyzing the segmentation of customers who have worked together. This research successfully modeled customer segmentation based on data mining clustering techniques using the K-Means data mining algorithm. The QRF (*Quantity, Recency, Frequency*) modeling process is analyzing the customer's behavior from the number of requests in each transaction carried out within a certain timeframe, as well as recency as the identification of the time span of the last transaction, as well as the number of transactions made within a certain time period. Researchers conducted a period of data for one year by analyzing customer activity in start-up providers of labor during 2019, on 86 active customers. Based on the analysis results obtained, customer segmentation in two clusters with QRF (*Quantity, Recency, Frequency*) modeling using *Davies Bouldin Index (DBI)* evaluation scored -0,482, while customer segmentation in three clusters using QRF (*Quantity, Recency, Frequency*) evaluation using *Davies Bouldin Index (DBI)* evaluation to obtain -0.469, and customer segmentation in four clusters with QRF (*Quantity, Recency, Frequency*) modeling using *Davies Bouldin Index (DBI)* evaluation to obtain -0,526.*

Keywords— pelanggan, clustering, algoritma k-means, DBI, QRF

1. PENDAHULUAN

Sulitnya mendapatkan pekerjaan yang sesuai dengan minat dan spesialisasi bagi seorang pekerja, serta sulitnya perusahaan mendapatkan seorang pekerja yang sesuai dengan kebutuhan perusahaan menyebabkan menjamurnya perusahaan konsultan atau penyedia tenaga kerja di Indonesia saat ini. Semakin banyaknya pilihan perusahaan penyedia tenaga kerja, tentu daya saing bisnis industri sumber daya manusia semakin tinggi. Sehingga perlu dilakukan analisis untuk menentukan strategi bisnis yang tepat.

Dalam menghadapi daya saing bisnis yang tinggi agar tetap bisa mempertahankan eksistensi perusahaan, maka pelanggan harus menjadi fokus utama bagi perusahaan. [1]

Perusahaan mengelompokkan kategori pelanggan berdasarkan catatan transaksi dalam kurun waktu tertentu. Proses klasterisasi dilakukan untuk mengetahui nilai tertentu pada pelanggan, sehingga dapat menentukan strategi pemasaran dengan analisis RFM (*Recency, Frequency, Monetary*) termodifikasi. Yaitu dengan memasukan atribut Quantity sebagai tambahan atau atribut pengganti untuk mendapatkan hasil klasterisasi yang optimum atau yang disebut dengan QRF (*Quantity, Recency, Frequency*). [2]

Untuk menerapkan kombinasi analisis QRF, maka digunakan juga metode *K-Means Clustering*. Salah satu metode data mining non hirarki yang membagi data menjadi satu kelompok atau lebih.

Disini peneliti melakukan riset selama satu tahun yakni meneliti aktifitas pelanggan aktif PT XYZ selama tahun 2019, sebanyak 86 data pelanggan aktif. Hal yang melatar belakangi penelitian ini karena klasterisasi data dengan menggunakan QRF akan mendapatkan hasil uji yang lebih optimum dibandingkan dengan menggunakan metode lain.

2. TINJAUAN PUSTAKA

a. Data Mining

Data mining digunakan untuk menguraikan suatu database dengan proses teknik statistik, matematis dan machine learning untuk mengidentifikasi informasi yang bermanfaat terkait dengan database dalam jumlah besar.

Jadi dapat dipahami bahwa data mining merupakan proses untuk menguraikan dan menemukan pola suatu data dengan menggunakan teknik statistik, matematik dan machine learning untuk mendapatkan informasi yang bermanfaat dari suatu data [3].

b. Algoritma *K-Means*

Metode data klaster non hirarki ini diawali dengan penentuan jumlah klaster yang diinginkan. Setelah jumlah klaster yang diinginkan dimasukkan, data akan diolah tanpa proses hirarki. *K-Means*

adalah salah satu algoritma yang membutuhkan suatu masukan sebanyak k dan membagi sekumpulan n *object* kedalam k *cluster*. Sehingga tingkat kemiripan sesama klaster sangat tinggi dan akan sangat rendah bila berbeda klaster [4].

Berikut algoritma dari metode *K-Means* :

- Masukan data yang akan diolah menjadi klaster
- Tentukan k sebagai jumlah klaster yang akan dibentuk
- Tentukan titik pusat *cluster* (*centroid*)

$$v = \sum_{i=1}^n x_i, i=1,2,3,\dots,n$$

Keterangan:

v sebagai centroid pada cluster, x_i adalah data ke- i dan n adalah banyaknya data yang menjadi anggota cluster.

d. Hitung jarak setiap data ke pusat klaster dengan menggunakan euclidian distance dengan rumus pada persamaan 2.5 sebagai berikut:

$$deuc(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Keterangan :

Jarak antara x_i dan y_i dinotasikan $d(x,y)$; x_i : data x ke- i ; y_i : data y ke- i dan n adalah banyaknya data

e. Hitung kembali pusat klaster dengan keanggotaan klaster yang baru, jika pusat klaster tidak berubah maka proses klaster telah selesai, jika belum maka ulangi langkah ke d sampai pusat klaster tidak berubah lagi. Berikut adalah rumus untuk menghitung kembali pusat cluster:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Keterangan :

μ_k = titik centroid dari cluster ke- K ; N_k = banyaknya data pada cluster ke- K ; x_q = data ke- q pada cluster ke- K

b. RFM Analysis

RFM (*Recency, Frequency, Monetary*) analisis merupakan suatu pendekatan untuk memahami perilaku pelanggan berdasarkan jarak waktu transaksi akhir (*Recency*), banyaknya jumlah transaksi (*Frequency*), banyaknya nilai pembelian (*Monetary*) [5].

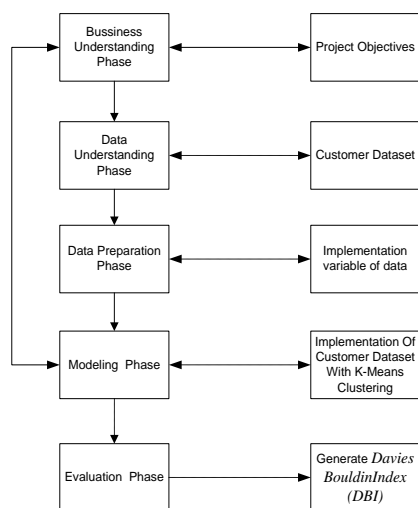
c. Tools RapidMiner

RapidMiner merupakan salah satu tools aplikasi yang digunakan untuk mengolah data mining dengan perhitungan dan analisis algoritma yang lengkap. RapidMiner yang digunakan pada penelitian kali ini adalah RapidMiner versi 5.3

3. METODOLOGI PENELITIAN

Tahapan penelitian berdasarkan dengan metodologi data mining adalah Cross Standard Process For Data Mining (CRISP-DM) yang terdiri dari 6 fase, yaitu : *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Dpeloymet*. [5]

Berikut gambaran alur proses penelitian berdasarkan CRISP-DM.



Gambar 1. Diagram alir penelitian

Tentukan objek penelitian sebagai tahap *business understanding* yang bertujuan untuk memahami proses bisnis dan permasalahan yang dihadapi pada suatu perusahaan. Kemudian kumpulkan data pelanggan yang akan diteliti untuk memenuhi kebutuhan penelitian sebagai tahap *data understanding*.

Apabila data sudah terkumpul, sebagai tahap *data preparation*, tentukan *variable* atau atribut data yang akan diteliti. Setelah atribut data sudah ditentukan, lalu tentukan algoritma yang akan digunakan sebagai tahap *modeling* pada data yang akan diteliti. Pada penelitian kali ini Algoritma yang digunakan adalah Algoritma *K-Means* yang digunakan untuk mengolah data.

Setelah data diolah menggunakan Algoritma *K-Means*, maka akan mendapatkan hasil akhir sebagai tahap *evaluasi* berupa nilai DBI (*Davies Bouldin Index*) sebagai standar evaluasi yang terbentuk.

4. HASIL PENELITIAN

Berdasarkan hasil dari seluruh proses tahapan penelitian. Pada bab ini akan dibahas mengenai proses pengolahan data pelanggan aktif berdasarkan CRISP-DM.

4.1. Pra-Pemrosesan Data.

Jenis data transaksi yang digunakan pada tahap ini adalah data pelanggan dengan banyak transaksi dalam satu kode pelanggan. Terdapat dua belas atribut dalam data, kemudian di seleksi menjadi. Seperti *kd_pelanggan, nm_pelanggan, industry, kota, nama_hrd, nmr_telepon, email, id_transaksi, tanggal_transaksi, jumlah_staff*.

Dari keseluruhan atribut dilakukan eksperimen pemodelan klasterisasi menggunakan atribut *Quantity, Recency, Frequency*.

4.2 Penyiapan Data

Dari 320 data pelanggan terdaftar selama tahun 2019, di seleksi ke dalam data pelanggan yang aktif melakukan transaksi. Sehingga menghasilkan 86 data pelanggan aktif. Hal ini dilakukan untuk mendapatkan data terpilih pelanggan yang loyal terhadap perusahaan.

Pada tahap selanjutnya di lakukan proses transformasi yaitu penghitungan antara banyaknya transaksi selama tahun 2019, serta jarak antara tanggal terakhir transaksi pada tahun 2019 dengan tanggal penelitian, dan banyaknya permintaan pada setiap transaksi selama tahun 2019.

Persiapan data dilakukan dengan pemodelan QRF (*Quantity, Recency, Frequency*).

Tabel 1. Transformasi Data Tahap 1

Kd Pelanggan	Nm Pelanggan	Industry	Kota	Nm_Hrd	Nmr_Telp	Email	Q	tgl_transaksi	tgl_riset	R	F
KD-00001	ati S	Perhotelan	Jakarta	Lu	536	lms	1	16/01/2019	16/07/2020	555	1
KD-00002	a	Makanan Minuman	Jakarta	nes	282	aga	1	10/01/2019	16/07/2020	555	1
KD-00003	nuo B	Perhotelan	Bali	nan	246	dgm	9	10/01/2019	16/07/2020	555	7
KD-00004	Mode	Perhotelan	Jakarta	li	817	lvi	2	16/01/2019	16/07/2020	548	2
KD-00005	eApu	Perhotelan	Jakarta	yan	138	yan	17	09/01/2019	16/07/2020	531	17
KD-00006	bil L	Logistik/Warehouse	Jakarta	yuk	388	yuk	1	24/01/2019	16/07/2020	520	1
KD-00007	gita	Makanan Minuman	Bali	gpi	188	min	1	29/01/2019	16/07/2020	515	1
KD-00008	hajo	Perhotelan	Jakarta	epi	212	ghe	2	24/01/2019	16/07/2020	510	2
KD-00009	ali N	Perhotelan	Bali	uS	738	lms	11	25/01/2019	16/07/2020	509	11
KD-00010	guri	Makanan Minuman	Jakarta	es	305	dgp	7	27/01/2019	16/07/2020	507	7

Pada table 1 transformasi tahap 1, data tersebut masih berupa dua belas atribut, berupa *kd_pelanggan, nm_pelanggan, industry, kota, nm_hrd, nmr_telepon, email, Q* sebagai *quantity, tgl_transaksi, tgl_riset, R* sebagai *recency, F* sebagai *frequency*.

Dari data tersebut sudah di implementasikan tahap transformasi berupa *recency* yakni hasil penghitungan tanggal transaksi dengan tanggal penelitian.

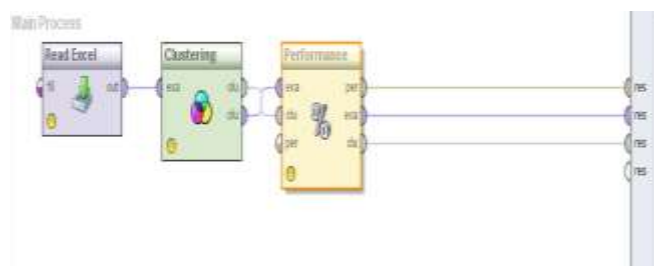
Tabel 2. Tranformasi Data Tahap 2

Kode Pelanggan	Q	R	F
KD-00001	1	559	1
KD-00002	1	555	1
KD-00003	9	555	7
KD-00004	2	519	2
KD-00005	17	531	17
KD-00006	1	520	1
KD-00007	1	515	1
KD-00008	2	510	2
KD-00009	11	509	11
KD-00010	7	507	7
KD-00011	15	502	9
KD-00012	17	500	8
KD-00013	18	499	2
KD-00014	5	499	5
KD-00015	5	495	5

Pada tabel 2 transformasi tahap 2, data tersebut menghasilkan empat atribut yang terdiri dari Kode Pelanggan, Q sebagai *Quantity* atau banyaknya permintaan di setiap transaksi, R sebagai *Recency* atau jarak waktu transaksi, dan F sebagai *Frequency* atau jumlah transaksi.

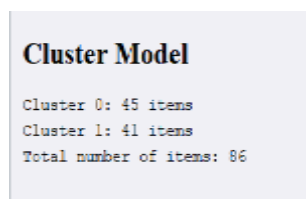
4.3. Pemodelan Klasterisasi

Pada tahap ini dilakukan pengolahan data ke dalam kelompok-kelompok atau klasterisasi dengan menggunakan Algoritma *K-Means* melalui perangkat lunak *RapidMiner* untuk mengetahui nilai *Davies Bouldin Index (DBI)* setiap klaster yang diteliti.

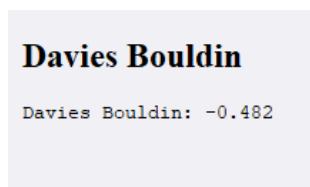


Gambar 2. Pengolahan data dengan Algoritma *K-Means* melalui perangkat lunak *RapidMiner* 5.3

4.3.1 Eksperimen Pemodelan Algoritma *K-Means* dan QRF dengan dua klaster



Gambar 3. Hasil pemodelan dua klaster



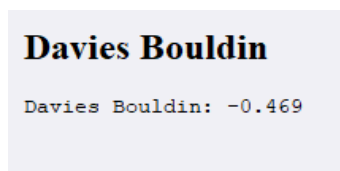
Gambar 4. Hasil DBI dengan dua klaster

Hasil diatas adalah hasil dari eksperimen pemodelan Algoritma *K-Means* dan QRF dengan dua klaster. Dengan nilai *Davies Bouldin Index* sebesar -0,482.

4.3.2 Eksperimen Pemodelan Algoritma *K-Means* dan QRF dengan tiga klaster



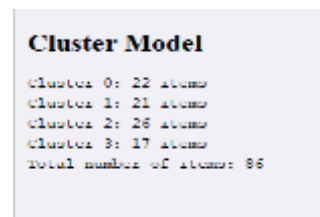
Gambar 5. Hasil pemodelan tiga klaster



Gambar 6. Hasil DBI dengan tiga klaster

Hasil diatas adalah hasil dari eksperimen pemodelan Algoritma *K-Means* dan QRF dengan tiga klaster. Dengan nilai *Davies Bouldin Index* sebesar -0,469.

4.3.3 Eksperimen Pemodelan Algoritma *K-Means* dan QRF dengan empat klaster



Gambar 5. Hasil pemodelan empat klaster



Gambar 6. Hasil DBI dengan empat kluster

Hasil diatas adalah hasil dari eksperimen pemodelan Algoritma *K-Means* dan QRF dengan empat kluster. Dengan nilai *Davies Bouldin Index* sebesar -0,526.

5. KESIMPULAN

Pada penelitian ini dapat ditarik kesimpulan bahwa berdasarkan hasil analisa yang diperoleh, bahwa segmentasi pelanggan pada dua kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index* (DBI) memperoleh nilai -0.482, sedangkan segmentasi pelanggan pada tiga kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index* (DBI) memperoleh nilai -0.469, dan segmentasi pelanggan pada empat kluster dengan pemodelan QRF (*Quantity, Recency, Frequency*) menggunakan evaluasi *Davies Bouldin Index* (DBI) memperoleh nilai -0.526.

Berdasarkan hasil eksperimen dapat kita simpulkan bahwa nilai optimum DBI terletak pada tiga kluster, yaitu senilai -0,469. Semakin kecil nilai *Davies Bouldin Index* (DBI) maka semakin optimum suatu data yang diuji.

DAFTAR PUSTAKA

- [1] D. P. Hidayatullah, R. I. Rokhmawati, and A. R. Perdanakusuma, "Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2018.
- [2] A. Wibowo and A. R. Handoko, "Segmentasi Pelanggan Ritel Produk Farmasi Obat Menggunakan Metode Data Mining Klusterisasi Dengan Analisis Recency Frequency Monetary (RFM) Termodifikasi," *J. Teknol. Inf. dan Ilmu Komput.*, 2020.
- [3] E. Turban, J. Aronson, and T. Llang, *Decision Support Systems and Intelligent Systems*. 2003.
- [4] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-MEANS Clustering Analysis pada penyakit menular manusia," *J. Chem. Inf. Model.*, 2013.
- [5] C. D. Rumiarti and I. Budi, "Segmentasi Pelanggan Pada Customer Relationship