

CLUSTERING UNTUK PENCARIAN TIPE RUMAH YANG COCOK BAGI CALON KONSUMEN BARU MENGGUNAKAN METODE *K-MEANS CLUSTERING* (STUDI KASUS PT. MAXIMA JAYA PERKASA)

Azyyati Nur Syarafina Wibowo

Program Studi Teknik Informatika, Universitas Bhayangkara Surabaya

Jl. Ahmad Yani No 114 Surabaya

Email : azyyatnur.072294@gmail.com

Abstrak

Bagi sebagian masyarakat Indonesia, rumah merupakan salah satu kebutuhan sekunder, sehingga dalam pemilihan perumahan yang tepat harus sesuai dengan keinginan konsumen. Dengan adanya perusahaan PT. Maxima Jaya Perkasa yang dirintis sejak tahun 2012 yang dimana data penjualan rumah perusahaan tersebut bertambah sangat pesat setiap tahunnya. Maka dapat dilakukan analisa data mining dengan menggunakan metode K-means Clustering. K-means Clustering adalah salah satu metode pengelompokan (clustering) data nonhierarchical (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda di kelompokkan ke dalam kelompok yang lain. Penelitian ini menggunakan data seperti penghasilan gaji, umur, status, harga rumah dan angsuran KPR.

Hasil penelitian ini dengan dilakukan dua kali pengujian menggunakan data latih sejumlah 12 data latih dan 100 data latih ditambah 1 sebagai data uji dan didapat nilai akurasi sebesar 83 % dan error sebesar 17%.

Kata Kunci : K-means Clustering, Rumah, Euclidean, Gaji, Harga Rumah, Angsuran KPR

1. PENDAHULUAN

Rumah merupakan salah satu kebutuhan sekunder, sehingga dalam pemilihan perumahan yang tepat agar sesuai dengan keinginan konsumen. Dalam pemilihan tipe rumah ada beberapa kriteria yang digunakan [2] seperti luas bangunan, harga rumah dan angsuran KPR.

Data konsumen pembeli rumah yang ada pada PT. Maxima Jaya Perkasa berlatar belakang dan berpenghasilan yang berbeda-beda oleh karena itu dalam penelitian ini dilakukanlah cara mengolah data-data yang telah didapatkan dari konsumen yang membeli rumah sebelumnya seperti penghasilan gaji,

umur dan status untuk mengetahui pola konsumen berdasarkan tipe rumah dari data-data tersebut.

Dengan memanfaatkan metode K-Means Clustering menggunakan 100 data konsumen yang ada di PT. Maima Jaya Perkasa sebagai data latih, diharapkan perusahaan mampu menentukan pengelompokan tipe rumah berdasarkan pola data konsumen sebelumnya dengan calon konsumen baru.

2. LANDASAN TEORI

K-means merupakan salah satu metode pengelompokan (*clustering*) data nonhierarchical (sekatan) yang berusaha mempartisi data yang ada ke

dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain[8].

Algoritma pengelompokan data secara umum metode K-Means :

- Tentukan jumlah kelompok
- Alokasikan data ke dalam kelompok secara acak
- Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing kelompok

Rumus :

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

Keterangan :

C_i = *centroid*

M = jumlah data

- Alokasikan masing-masing data ke sentroid terdekat.

Jarak data ke sentroid menggunakan rumus parameter *Euclidean*:

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

Keterangan :

D = jarak antara data x_2 dan x_1

|.| = nilai mutlak.

Data dialokasikan ulang secara tegas ke kelompok yang mempunyai centroid dengan jarak terdekat dari data tersebut. Pengalokasian data menggunakan rumus (MacQueen, 1967) :

$$a_{ij} = \begin{cases} 1 & d = \min \{(x_i C_1)\} \\ 0 & \text{lainnya} \end{cases}$$

Keterangan :

a_{ij} = nilai keanggotaan titik x_i ke pusat kelompok C_1

d = jarak terdekat dari data x_i ke K kelompok setelah dibandingkan

C_i = *centroid* (pusat kelompok) ke- i

- Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

Rumus perhitungan nilai fungsi objektif :

$$J = \sum_{i=1}^N \sum_{l=1}^K a_{il} D(x_i, C_l)^2$$

Keterangan :

N = jumlah data

K = kelompok

a_{il} = nilai keanggotaan titik data x_i ke pusat kelompok C_l

C_l = pusat kelompok ke- l

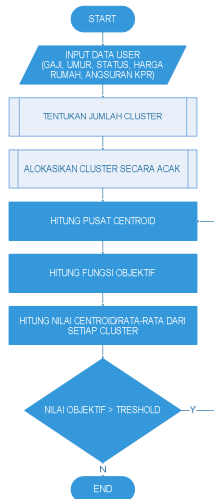
$D(x_i, C_l)$ = jarak titik x_i ke kelompok C_l

a mempunyai nilai 0 atau 1. Apabila suatu data merupakan anggota suatu kelompok, nilai $a_{il} = 1$. Jika tidak, nilai $a_{il} = 0$.

Iterasi dihentikan apabila nilai fungsi objektif=0.

3. PERANCANGAN SISTEM

Flowchart adalah penggambaran secara grafik dari langkah-langkah dan urutan-urutan prosedur dari suatu program. *Flowchart* biasanya mempermudah penyelesaian suatu masalah khususnya masalah yang perlu dipelajari lebih lanjut.



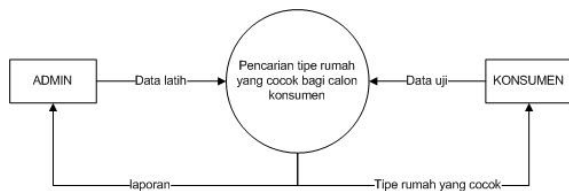
Gambar 1. Flowchart Sistem

Flowchart pada gambar 1 diatas menggambarkan alur proses pengelompokkan konsumen baru menggunakan metode k-means. Tahapan pertama yaitu input data konsumen baru yang didalamnya terdapat penghasilan/gaji, umur, status hidup, harga rumah, angsuran rumah. Tahapan selanjutnya akan diproses menggunakan metode k-means klustering.

3.1. Data Flow Diagram (DFD)

3.1.1. DFD Level 0

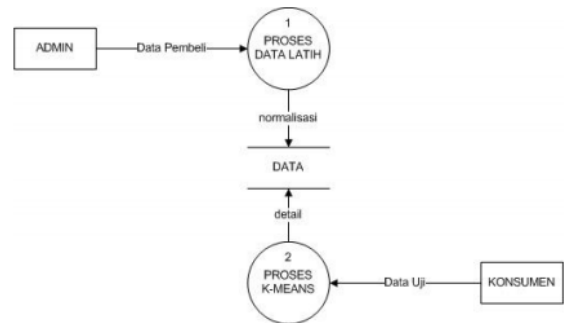
Admin sebagai pemberi masukan dan pengelolaan data yang ada didalam sistem atau posisi admin berada dibelakang sistem (*backend*). Konsumen rumah sebagai pengguna untuk memperoleh informasi yang ada pada sistem atau posisi calon pembeli rumah didepan sistem (*frontend*).



Gambar 2. DFD Level 0

3.1.2. DFD Level 1

Admin melakukan autentikasi dengan login, kemudian memasukkan data latih dan menentukan kelompok data secara acak ke dalam cluster kemudian disimpan di database. Dari sisi calon pembeli (konsumen) memasukkan data uji dan menyimpan ke database, kemudian dilanjutkan dengan proses normalisasi dan memperoleh nilai data uji yang dimasukkan kedalam sistem.

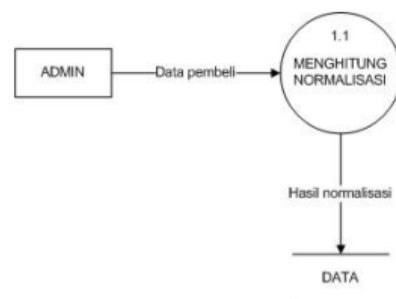


Gambar 3. DFD Level 1

3.1.3. DFD Level 2

a. Proses Data Latih

Langkah pertama yang dilakukan oleh admin adalah login, selanjutnya memasukkan data latih, dari data tersebut dinormalisasi kemudian disimpan ke *database*.

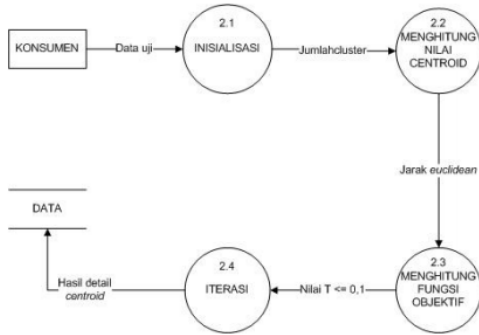


Gambar 4. DFD Level 2 Proses Data Latih

b. Proses K-Means Clustering

Inisialisasi semua data latih untuk mendapatkan cluster (kelompok) awal, kemudian setiap fitur dari masing-masing data dijumlah dan dibagi dengan

banyaknya jumlah data dalam satu kelompok. Kemudian setiap data dihitung terhadap nilai masing-masing centroid menggunakan jarak Euclidean untuk mendapatkan nilai fungsi objektif, nilai fungsi objektif inilah yang menentukan berhentinya iterasi perhitungan jika threshold (ambang batas) = 0.



Gambar 5. DFD Level 2 K-Means Clustering

4. PENGUJIAN

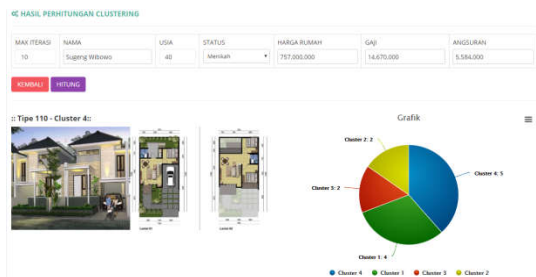
Pengujian pertama dan kedua menggunakan data uji yang sama.

MAX ITERASI	NAMA	USA	STATUS	HARGA RUMAH	GAJI	ANGSURAN
20	SUGENG WIB	40	Menikah	757.000.000	14.670.000	5.584.000

Gambar 6. Data Uji

a. Hasil Pengujian Pertama

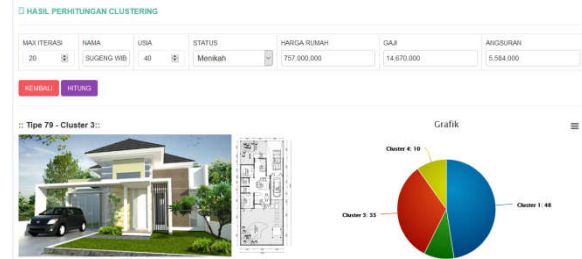
Hasil akhir cluster yang dapat direkomendasikan untuk calon konsumen baru berdasarkan 12 data latih dan 1 data uji yaitu cluster 4. Terlihat pada gambar 7.



Gambar 7. Hasil Pengujian Pertama

b. Hasil Pengujian Kedua

Hasil akhir cluster yang dapat direkomendasikan untuk calon konsumen baru berdasarkan 100 data latih dan 1 data uji yaitu cluster 3.



Gambar 8. Hasil Pengujian Kedua

5. ANALISA HASIL

Berdasarkan dari hasil pengujian yang sudah dilakukan, pada pengujian pertama ada 12 data latih yang dimana ada 2 data yang berpindah kluster dan ada 10 data yang tidak berpindah kluster saat dilakukan penambahan 88 data latih di pengujian kedua.

Tabel 1. Analisa Hasil Akhir

Data Ke-	Hasil Cluster		Hasil
	Pengujian 1	Pengujian 2	
1	4	4	TETAP
2	1	1	TETAP
3	4	3	PINDAH
4	4	4	TETAP
5	1	1	TETAP
6	1	1	TETAP
7	4	4	TETAP
8	3	3	TETAP
9	1	3	PINDAH
10	2	2	TETAP
11	2	2	TETAP
12	3	3	TETAP

Jadi semakin banyak data latih yang berkarakteristik dimasukkan maka akan semakin baik nilai validitas yang direkomendasikan untuk calon konsumen baru.

6. KESIMPULAN

1. Penentuan *clustering* pada tahap awal algoritma K-Means sangat berpengaruh pada hasil cluster, seperti pada hasil pengujian yang dilakukan dengan menggunakan 12 data latih dan 100 data latih dengan jumlah data pada tiap *cluster* yang berbeda menghasilkan hasil *cluster* yang berbeda juga.
2. Metode K-Means dapat melakukan pengelompokan calon konsumen baru berdasarkan variabel gaji, umur, harga rumah, status dan angsuran dengan jumlah data yang banyak tetapi belum efisien dalam mengelompokkan calon konsumen secara tepat.
3. Sistem pencarian tipe rumah yang cocok bagi calon konsumen baru menggunakan metode K-Means clustering memperoleh tingkat keberhasilan 83% dengan 2 kali pengujian.

7. DAFTAR PUSTAKA

- [1] Adrianto, Rizky dan Fahmi, Amiq (2016), *Penerapan metode clustering dengan algoritma K-means untuk rekomendasi pemilihan jalur peminatan sesuai kemampuan pada program studi teknik informatika SI Universitas Dian Nuswantoro*, **Journal of Information System** eISSN 2528-0236, Vol. 1, No. 2, hal 101-116.
- [2] Amborowati, Amardyah (2008), *Sistem Penunjang Keputusan Pemilihan Perumahan dengan metode AHP*

menggunakan Expert Choice, **Jurnal DASI**, Vol. 9, No. 1.

- [3] Amin, Choirul, dkk (2013), *110 Denah Rumah Pilihan*, Griya Kreasi, Cibubur, Jakarta Timur.
- [4] Frick, Heinz dan Widmer, Petra (2006) *Membangun, membentuk, menghuni*, Kanisius, Yogyakarta.
- [5] Khotimah, Tutik (2014), *Pengelompokan surat dalam Al-Qur'an menggunakan algoritma K-means*, **Jurnal SIMETRIS**, Vol. 5, No. 1, hal. 83-88.
- [6] Maulani, Desfa (2013) *"Implementasi algoritma K-means dalam menentukan berat badan ideal"*. Skripsi. Fakultas Ilmu Komputer dan Teknologi Informasi, Ilmu Komputer, Universitas Sumatera Utara Medan.
- [7] Ong, Johan Oscar (2013), *Implementasi Algoritma K-Means Clustering untuk menentukan strategi marketing*, **Jurnal Ilmiah Teknik Industri**, Vol. 12, No. 1, hal. 10-20.
- [8] Prasetyo, Eko (2012), *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Ed.1, Andi, Yogyakarta.
- [9] Ramadhani, Rima Dias (2014) *"Data mining menggunakan Algoritma K-Means Clustering untuk menentukan strategi promosi Universitas Dian Nuswantoro"*. Skripsi. Fakultas Ilmu Komputer, Jurusan Sistem Informasi, Universitas Dian Nuswantoro.
- [10] Rivani, Edmira (2010), *Aplikasi K-means Cluster untuk pengelompokan provinsi berdasarkan produksi padi, jagung, kedelai, dan kacang hijau tahun 2009*,

Jurnal Mat Stat, Vol. 10, No. 2, hal. 122-134.

- [11] Suprastiyono, Cahyo (2011) *“Pengelompokan Daerah Menurut Ketersediaan Pelayanan dan Kondisi Infrastruktur Puskesmas Di Jawa Timur Menggunakan Metode K-means”*. Skripsi. Fakultas Teknik, Teknik Informatika, Universitas Bhayangkara Surabaya.
- [12] Suprawoto, Totok (2016), *Klasifikasi data mahasiswa menggunakan metode K-means untuk menunjang pemilihan strategi pemasaran*, **Jurnal Informatika dan Komputer (JIKO)**, Vol. 1, No. 1, hal. 12-18.