



## KLASTERISASI DATA MEMBER INDOGROSIR JAMBI MENGUNAKAN METODE K-MEANS

**Mohd. Yuliandani<sup>1</sup>, Pariyadi<sup>2</sup>, Teuku Djauhari<sup>3</sup>**

Program Studi Teknik Informatika, STMIK Nurdin Hamzah, Jambi

E-mail: [1mohd.yuliandani@gmail.com](mailto:mohd.yuliandani@gmail.com), [2pariyadi.twn@gmail.com](mailto:pariyadi.twn@gmail.com), [3teknikom@gmail.com](mailto:teknikom@gmail.com)

**Abstract** - The high level of the visit of Jambi Indogrosir members is a reflection of the high interest in shopping from members. For Indogrosir members are the top priority for increasing shopping interest. Indogrosir members are given many interesting promotions. Because the number of existing members of Indogrosir Jambi has difficulty in being able to provide the right promotions for the members. To be able to provide the types and types of promotions that are right on target, an application is built to classify member data based on the number of member visits. The method used to build a member data clustering application is K-Means data mining. Data mining is the process of analyzing data to find a pattern from a data set. Data mining can analyze large amounts of data into information that has meaning for decision support. Data to be input by the user is in the form of member data, the number of visits, member shopping data, member shopping data based on the number of products purchased by the unit. The making of this application uses the PHP programming language and MySQL database. The purpose of making this application is expected to help the Indogrosir Jambi in designing promotional strategies to the maximum.

**Keywords:** Clustering, Data Mining, Indogrosir, Member, PHP, Promotion.

### I. PENDAHULUAN

#### 1.1. Latar Belakang

Banyaknya data yang dimiliki sebuah instansi/perusahaan bisa menyebabkan kesulitan dalam mengelola data untuk kepentingan instansi/perusahaan. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin dikelola. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik *clustering*. Penggunaan teknik DM *clustering* berbeda-teknik *Data Mining* (DM) yang lainnya, seperti *association rule mining* dan *classification* yang memerlukan tahapan *training* dan evaluasi. Teknik ini menggunakan metode *unsupervised learning* yang berarti DM tidak perlu melakukan *training* terlebih dahulu tapi bisa langsung menggunakannya untuk pengelompokan. Teknik ini masih jarang digunakan dibanding dengan teknik-teknik DM yang lain.

Indogrosir merupakan jaringan pusat perkulakan dengan format distribusi ke pedagang retail/eceran tradisional dan modern. Kehadiran sebuah pusat perkulakan menjadi suatu solusi karena para pedagang eceran akan dengan mudah memperoleh barang dagangannya pada satu tempat saja sehingga menghemat waktu dan biaya, yang pada akhirnya akan meningkatkan keuntungan dan daya saing atas barang yang mereka jual. Dan Indogrosir memberikan kemudahan untuk pelanggan

yang ingin menjadi *member* dari Indogrosir dengan cara mengajukan diri untuk menjadi *member* tetap Indogrosir. Dengan adanya kartu *member* ini banyak keuntungan yang didapat oleh pelanggan yang menjadi *member* dari Indogrosir Jambi.

Banyaknya *member* Indogrosir yang ada membuat sulit untuk mengelola dan mengklasifikasinya, sehingga membuat Indogrosir Jambi harus lebih ekstra lagi dalam memvalidkan data-data *member* tersebut. Kegiatan klasifikasi yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan. Peneliti tertarik untuk melakukan penelitian menggunakan *K-Means* sebagai metode pengelompokan dengan menghitung jarak di mana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan pada kelompok yang sama sehingga data dapat terklusterisasi.

Berdasarkan latar belakang tersebut maka peneliti tertarik untuk melakukan penelitian dalam bentuk skripsi yang berjudul : “**KLASTERISASI DATA MEMBER INDOGROSIR JAMBI MENGGUNAKAN METODE K-MEANS**”.

#### 1.2. Rumusan Masalah

Adapun perumusan masalah dalam penelitian ini adalah : “Bagaimana cara mengetahui kebiasaan belanja dari setiap *member* Indogrosir Jambi ?

### 1.3. Batasan Masalah

Adapun Batasan Masalah Dalam Penelitian Ini Adalah :

1. Penelitian dilakukan di Indogrosir Jambi.
2. Penelitian dilakukan dengan menggunakan metode data mining dengan algoritma K-means *clustering*.
3. *Member* pelanggan akan terbagi dalam 3 klaster yaitu, member terbaik, baik dan cukup baik. Pengklasteran member berdasarkan dari jumlah kunjungan, dan produk yang dibeli. Aplikasi yang digunakan adalah berbasis Web.
4. Hasil penelitian berupa data belanja dari setiap *member* Indogrosir Jambi.

### 1.4. Tujuan Penelitian

Tujuan penelitian ini adalah agar Perusahaan dapat mengklasterisasikan *member* dari Indogrosir Jambi.

### 1.5. Manfaat Penelitian

Manfaat penelitian ini adalah :

1. Data *member* yang berjumlah banyak akan cepat diproses dan dipisah berdasarkan kemiripan data dengan metode *data mining* menggunakan metode K-Means.
2. Hasil penelitian ini dapat memberikan gambaran dari pola belanja setiap *member*.
3. Bagi Indogrosir Jambi, dapat digunakan sebagai pertimbangan untuk memaksimalkan promosi dari hasil statistik yang ada.

## II. TINJAUAN PUSTAKA

### 2.1. Pengertian Data Mining

Data mining menurut Witten, Ian H. Frank adalah proses ekstraksi suatu data (sebelumnya tidak diketahui, bersifat implisit, dan dianggap tidak berguna) menjadi informasi atau pengetahuan atau pola dari data yang jumlahnya besar (Suntoro, 2019:2).

Data mining merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Dalam data mining juga terdapat metode yang dapat digunakan seperti klasifikasi, *clustering*, regresi, seleksi variabel, dan market basket analisis (Sihombing, 2017:75).

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machining

learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Hartono & Nursikuwagus, 2016:702)

### 2.2. Pengertian *Clustering*

Menurut (He & Tan, 2012), *clustering* adalah data yang tidak mempunyai label/kelas sehingga sering disebut dengan teknik *unsupervised learning* (Suntoro, 2019:85).

Menurut (Santoso, 2013), pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target *output*. Dalam data mining ada dua jenis metode *clustering* yang digunakan dalam pengelompokkan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*.

Menurut Xiang, Zhu, Ma, Meng, & An *clustering* adalah pengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan. (Suntoro, 2019:85).

Maka dapat disimpulkan *cluster* adalah menemukan kumpulan objek hingga objek-objek dalam suatu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan). Dengan objek-objek dalam kelompok lain. Tujuan dari analisa *cluster* adalah meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antara *cluster*.

### 2.3. Pengertian K-Means

K-means *clustering* merupakan salah satu metode data *clusternon-hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berbeda dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Widodo, 2013).

K-Means merupakan salah satu metode *clustering* yang *non hirarki* yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data kedalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain (Prasetyo, 2012).

K-means merupakan algoritma *clustering* yang berulang-ulang. Algoritma K-means dimulai



dengan pemilihan secara acak  $K$ ,  $K$  disini merupakan banyaknya *cluster* yang ingin di bentuk. Kemudian tetapkan nilai-nilai  $K$  secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau bisa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid (Widodo, 2013).

Diproleh keanggotaan *cluster* tidak berubah dibandingkan dengan pengelompokan data sebelumnya, sehingga algoritma dapat dihentikan. Algoritma K-means pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada langkah ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya K-means menguji masing-masing komponen didalam populasi data dan menndai komponen tersebut ke salah satu pusat *cluster* yang telah didenifikasikan tergantung jarak minimum antar komponen dengan tiap puast *cluster*. Posisi pusat *cluster* akan dihitung kembali kedalam tiap-tiap *cluster* dan terakhir akan terbentuk posisi *cluster* baru.

Algoritma k-means adalah algoritma yang mempartisi data ke dalam *cluster – cluster* sehingga data yang memiliki kemiripan berada pada satu *cluster* yang sama dan data yang memiliki ketidaksamaan berada pada *cluster* yang lain. Sarwono mengemukakan secara lebih detail, algoritma K-Means adalah sebagai berikut (Jarogi, 2016:59):

- 1) Menentukan  $k$  sebagai jumlah *cluster* yang ingin di bentuk.
- 2) Membangkitkan nilai random untuk pusat *cluster* awal (centroid) sebanyak  $K$ .
- 3) Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$D = \sqrt{(p_k - q_k)^2 + (p_q - q_q)^2}$$

Dengan ketentuan :  $n$  adalah jumlah dimensi (atribut),  $p_k$  dan  $q_k$  adalah atribut ke  $-k$  untuk obyek data  $p$  dan  $q$ .

- 4) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
- 5) Memperbaharui nilai centroid. Nilai centroid baru di peroleh dari rata-rata *cluster* yang bersangkutan.
- 6) Melakukan perulangan dari langkah 3 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

### III. HASIL DAN PEMBAHASAN

#### 3.1. Kebutuhan Sistem

Merupakan pemilihan kebutuhan sistem yang harus diwujudkan dalam perangkat lunak yang meliputi fungsi-fungsi yang dibutuhkan, proses masukan dan keluaran data dan antar muka yang diharapkan. Adapun kebutuhan analisis yang dibutuhkan sebagai berikut :

##### 3.1.1. Kebutuhan Input

Untuk kebutuhan masukan dibutuhkan data-data yang saling mendukung antara lain:

1. Data *member*. Data ini meliputi nama, kode *member*, jumlah barang belanjamember yang diisi oleh admin.
2. Data *cluster*. Data ini meliputi nilai yang telah ditetapkan bertujuan untuk mengelompokkan data dengan karakteristik yang sama dan data dengan karakteristik yang berbeda.
3. Data Centroid. Data ini meliputi pusat dari penyebaran data yang sudah ada.

##### 3.1.2. Kebutuhan Proses

Kebutuhan proses antara lain :

1. Pengolahan data masukan yaitu data *member*, *cluster* dan data centroid.
2. Pengolahan data mining, melakukan perhitungan dengan metode K-means.

##### 3.1.3. Kebutuhan Output

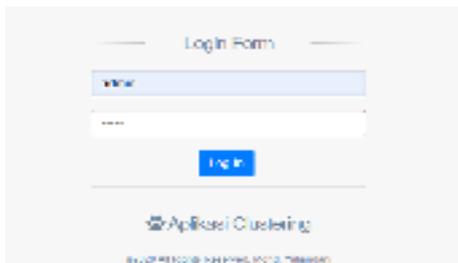
Data keluaran (*output*) yang dihasilkan pada tahap ini adalah berupa hasil analisis sistem berdasarkan data masukan (*input*), antara lain :

1. Data *member*. Data ini meliputi nama, jumlah banyaknya barang belanja sehingga menjadikan kebutuhan laporan.
2. Data hasil perhitungan. Data ini meliputi hasil pengelompokan *member* dengan metode K-Means sehingga menjadikan kebutuhan laporan.

#### 3.2. Kebutuhan Antarmuka

Kebutuhan antarmuka yang diinginkan oleh pemakai dan berhubungan dengan pemilik kost ini diinginkan sebaik mungkin sehingga bersifat ramah pengguna, artinya pengguna dapat menggunakan perangkat lunak yang dibuat nyaman mungkin dan





Gambar 5. Implementasi Login

## 2. Implementasi Halaman Utama

Tampilan halaman menu utama/beranda merupakan *form* pertama kali ditampilkan. *Form* menu utama ini dirancang secara sederhana agar pengguna mudah dalam menggunakan sistem ini.



Gambar 6. Implementasi Halaman Utama

## 3. Implementasi Halaman Member

Pada halaman ini menampilkan data *member* serta dapat menambah *member* dan *update* data *member*.



Gambar 7. Implementasi Halaman Member

## 4. Implementasi Halaman Perhitungan

Pada halaman ini menampilkan hasil perhitungan.



Gambar 8. Implementasi Halaman Perhitungan

## 5. Implementasi Halaman Laporan Member

Pada halaman ini menampilkan laporan *member* yang masuk kedalam *clustering*



Gambar 9. Implementasi Halaman user Buka Kost

## IV. PENUTUP

Berdasarkan Pengujian dan penjelasan cara klasterisasi data *member* indogrosir jambi menggunakan metode k-means pada bab-bab sebelumnya, maka pada bab ini peneliti menarik suatu kesimpulan:

### 4.1. Kesimpulan

1. Setelah dilakukan pengelompokan data *member* menggunakan K-Means *clustering* terbentuk tiga *cluster*.
2. Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma K-Means bisa digunakan untuk mengelompokkan *member* berdasarkan jumlah kunjungan dan beberapa atribut belanja dari setiap *member*.
3. Hasil penelitian dapat dijadikan strategi yang tepat sasaran untuk setiap *cluster* yang terbentuk adalah dengan memberikan banyaknya promosi yang sesuai dengan pada setiap *cluster*.
4. Metode K-Means *clustering* telah berhasil diimplementasikan untuk mengelompokkan *member* dengan kategori *member* terbaik, baik, dan cukup baik. Pembagian *member* untuk setiap *cluster* ditentukan berdasarkan jumlah kunjungan, jumlah produk yang dibeli dan jumlah produk dalam satuan untuk setiap *cluster*. Sistem pada penelitian ini dibuat dengan menggunakan bahasa pemrograman PHP.

### 4.2. Saran

Agar kinerja dari program klasterisasi data *member* indogrosir jambi menggunakan metode k-means dirancang lebih optimal, maka peneliti berharap:

1. Pengelompokan terhadap data *member* Indogrosir sebaiknya dilakukan secara rutin setiap tahun untuk menghindari penurunan minat belanja dari setiap *member*.
2. Penelitian ini dapat dijadikan sebagai salah satu referensi bagi pihak Indogrosir jambi untuk mengambil keputusan dalam melakukan strategi promosi.
3. Setelah mengetahui tingkat kunjungan *member*, diharapkan Indogrosir jambi lebih memperhatikan produk tersedia.

- 4 Pengembangan lebih lanjut dari sistem ini adalah dengan melengkapi sistem pengentrian data member dengan lebih mudah melalui Microsoft Excel dan menambahkan fitur-fitur baru.

#### DAFTAR REFERENSI

- Aprilla, Dennis., Donny Aji Baskoro., Lia Ambarwati., dan I Wayan Simri Wicaksana. 2012. *“Belajar Data Mining dengan RapidMiner”*. Jakarta: Gramdia Pustaka Utama.
- Ayu, F., & Permatasari, N. 2018. *“Perancangan Sistem Informasi Pengolahan Data Praktek Kerja Lapangan (PKL) Pada Devisi PT. Pengadaian”*. Jurnal Intatech Volume 2, No. 2 Oktober 2018.
- Enterprise, J. 2018. *“Html, Php, Dan Mysql Untuk Pemula”*. ISBN: 978-602048622-2, Yogyakarta.
- Ginjar, Abdurrahman. 2016. *“Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-means”*. Journal of Sistem & Teknologi Informasi Indonesia.
- Jogiyanto. 2014. *“Analisis & Desain Sistem Informasi”*. Yogyakarta : Penerbit CV. Andi Offset.
- Joko, Suntoro, KN. 2019. *“Data Mining”*. Jakarta : PT. Elex Media Komputindo Indonesia.
- Kurniawan, B., & Herryanto, D. 2018. *“System Informasi Kepegawaian Berbasis Web Dengan Menggunakan Framework Codeigniter Pada Dinas Pendidikan Kabupaten Ogan Komering Ulu”*. JIK, Vol. 9, No. 2 Desember 2018.
- Nugroho. Bunafit. 2019. *“Aplikasi Pemrograman Web Dinamis dengan PHP dan MySQL”*. Cetakan ke-1. Yogyakarta : PENERBIT GAVA MEDIA.
- Nurul, Chafid., & Ismail Ari Wibowo. 2018. *“Implementasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Demam Berdarah Di Kota Tangerang Selatan Menggunakan Algoritma K-means”*. Journal of Satya Informatika.

Purwanto, Agus., & Darmadi Agus. 2018. *“Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma”*. Vol. 2 Maret 2018.

Sidik, Betha. 2012. *“Pemrograman Web dengan PHP”*. Edisi Revisi. Bandung : Informatika Bandung.

Siregar, S.R.S., Sundari, P. 2016. *“Rancangan Sistem Informasi Pengelolaan Data Kependudukan Desa (Studi Kasus Di Kantor Desa Sangiang Kecamatan Sepatan Timur)”*. Jurnal Sisfotek Global ISSN : 2088-1762 Vol. 6 No.1/Maret 2016.

Yuliansyah, H. 2014. *“Perancangan Replikasi Basis Data Mysql Dengan Mekanisme Pengamanan Menggunakan Ssl Encryption”*. Jurnal Informatika Vol. 8, No. 1.

#### IDENTITAS PENULIS

Nama : Mohd. Yuliandani  
 NIM : 1602013  
 TTL : Bajubang, 10Juli 1998  
 Alamat Rumah : Perumahan Pesona Jambi Mayang Mangurai Kecamatan Kota Baru  
 Telp. : 089630068817  
 Email : [mohd.yuliandani@gmail.com](mailto:mohd.yuliandani@gmail.com)

Nama : Pariyadi, M.Kom  
 NIDN/NIK : 1013029002  
 NIDN/NIK : 1013029001 / 16.104  
 TTL : Jambi/13 Februari 1990  
 Jabatan Fungsional: Asisten Ahli  
 Alamat Rumah : Jl. Kolonel Abunjani, Sipin, Jambi  
 Telp. : 085266369055  
 Email : [pariyadi.twn@gmail.com](mailto:pariyadi.twn@gmail.com)

Nama : Teuku Djauhari, S.Kom, M.SI  
 NIDN : 1012116702  
 TTL : Kupang, 12 November 1967  
 Golongan/Pangkat : III B  
 Jabatan Fungsional: Asisten Ahli  
 Alamat Rumah : Jl. Pemancar TVRI, No. 26, RT. 36, Kec. Telanaipura, Kota Jambi.  
 Telp. : 0811749440  
 Email : [teknikom@gmail.com](mailto:teknikom@gmail.com)