

BAB II TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban dkk. 2005). Terdapat beberapa istilah lain yang memiliki makna sama dengan *data mining*, yaitu *Knowledge discovery in databases* (KDD), ekstraksi pengetahuan (*knowledge extraction*), Analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan data *archaeology* dan data *dredging* (Larose, 2005).

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu:

1. Fungsi Prediksi (*prediction*). Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.

2. Fungsi Deskripsi (*description*). Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.

3. Fungsi Klasifikasi (*classification*). Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.

4. Fungsi Asosiasi (*association*). Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

2.2 Clustering

Clustering merupakan proses untuk mengelompokkan sekumpulan objek data yang memiliki kemiripan tinggi ke dalam satu kelompok atau *cluster* dan setiap kelompok yang terbentuk tidak memiliki kemiripan dengan kelompok lainnya . Kemiripan dinilai berdasarkan nilai atribut yang mendeskripsikan objek data. *Clustering* atau juga disebut sebagai segmentasi ini adalah salah satu metode penggalian data yang *unsupervised*, karena tidak ada atribut yang digunakan sebagai panduan atau tidak adanya label pada data dalam proses pembelajaran. Teknik dalam clustering dilakukan untuk menemukan pengetahuan dari kumpulan data . Analisis cluster dalam konteks penggalian data yaitu proses menempatkan pelanggan atau prospek kedalam kelompok yang memiliki ciri-ciri yang sama .

Algoritma *clustering* membangun sebuah model dengan melakukan serangkaian pengulangan dan berhenti ketika model tersebut telah terpusat dan batasan dari segmentasi telah stabil. Hasil dari *clustering* yang bagus tergantung dengan ukuran kesamaan dan metode yang digunakan. Pendekatan dalam *cluster* berdasarkan saran dari Fraley dan Raftery, membagi metode pengelompokan menjadi dua kelompok utama yaitu metode hirarkis dan metode partisi .

Pada sistem ini proses metode clustering yang digunakan yaitu metode partisi karena memiliki tujuan untuk membuat kelompok yang setiap pelanggannya hanya berada pada satu kelompok tertentu.

2.3 Algoritma K-Means

Metode ini termasuk teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke k daerah bagian yang terpisah. Pada k-means, setiap objek harus masuk dalam kluster tertentu, tetapi dalam satu tahapan berikutnya objek akan berpindah ke kluster lain.

1. Menentukan nilai k sebagai jumlah *cluster* yang dibentuk.
2. Menentukan nilai awal *centroid* atau titik pusat cluster. Pada tahapan ini ditentukan nilai *centroid* secara acak, namun untuk tahapan berikutnya menggunakan rumus.

$$\overline{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Dimana :

V_{ij} = *centroid cluster* ke-i untuk variabel ke-j

N_i = banyaknya data pada *cluster* ke-i

i,k = indeks dari *cluster*

j = indeks dari variabel

X_{kj} = nilai data ke-k yang ada di dalam *cluster* untuk variabel ke-j

3. Menghitung jarak antara titik *centroid* dengan titik setiap objek, dapat dilakukan dengan menggunakan Euclidean Distance dengan rumus.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_t - t_t)^2}$$

Dimana:

D_e = *Euclidean Distance*

- i = banyaknya data
 (x,y) = koordinat data
 (s,t) = koordinat centroid.

4. Mengelompokan data hingga terbentuk *cluster* dengan titik centroid dari setiap *cluster* merupakan titik centroid yang terdekat. Penentuan anggota cluster adalah dengan memperhitungkan jarak minimum objek.
5. Memperbarui nilai *centroid* setiap cluster.
6. Mengulangi langkah ke-2 hingga akhir sampai nilai dari titik centroid tidak lagi berubah.

2.4 Segmentasi Pelanggan

Segmentasi merupakan salah satu cara untuk memiliki komunikasi yang lebih dengan pelanggan. Tujuan dari segmentasi adalah untuk menyesuaikan produk, jasa, dan pesan pemasaran untuk setiap segmen. Proses segmentasi menempatkan pelanggan sesuai dengan karakteristik kelompok pelanggan yang serupa. Segmentasi pelanggan merupakan langkah persiapan untuk mengklasifikasikan setiap pelanggan sesuai dengan kelompok pelanggan yang sudah ditetapkan. Karakteristik pelanggan dapat direpresentasikan oleh beberapa kategori variabel yang terkait dengan pengelompokan, seperti berikut ini :

- *Demographics*: Umur, jenis kelamin, besarnya keluarga, besarnya kediaman, siklus kehidupan keluarga, pemasukan, pekerjaan atau profesi, pendidikan, kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.
- *Psychographics*: kepribadian, gaya hidup, nilai-nilai, sikap.
- *Behaviour*: manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian
- *Geographic*: negara, provinsi, kota, kode pos, iklim

Skema segmentasi yang berbeda dapat dikembangkan menurut tujuan bisnis yang spesifik dari organisasi. Segmentasi umumnya digunakan melalui riset data pasar untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, dan opini tentang perusahaan dan kompetisi. Segmentasi pelanggan berdasarkan pada riset pasar dan demografi seringkali membutuhkan pemahaman karakteristik semua pelanggan agar lebih efektif mengetahui.

2.5 Customer Relationship Management (CRM)

CRM merupakan sebuah pendekatan baru dalam mengelola hubungan korporasi dan pelanggan pada level bisnis sehingga dapat memaksimalkan komunikasi, pemasaran melalui pengelolaan berbagai kontak yang berbeda dengan pelanggan. Pendekatan ini memungkinkan untuk mempertahankan pelanggan dan memberikan nilai tambah terus menerus pada pelanggan, selain juga memperoleh keuntungan yang berkelanjutan.

Hal yang perlu dipahami adalah bahwa dari luar, pelanggan yang berinteraksi dengan perusahaan hanya memahami bisnis yang dilakukan oleh perusahaan sebagai satu entitas, tidak lebih; meskipun pelanggan juga berinteraksi dengan sejumlah pekerja yang berbeda peran dan departemennya. Semua itu tetap dianggap sebagai satu kesatuan. Dengan CRM, dukungan pada proses bisnis, informasi tentang pelanggan dan interaksinya dapat dimasukkan, disimpan, diakses oleh semua staf pada berbagai unit kerja dengan tujuan untuk meningkatkan layanan yang diberikan pada pelanggan, dan menggunakan informasi kontak pelanggan untuk target pemasaran.

2.6 Penelitian Terdahulu

2.6.1 Penelitian Dewi Astria dan Suprayogi (2017)

Sistem yang diajukan Dewi Astria dan Suprayogi menerapkan Algoritma Fuzzy C-Means Untuk Clustering Pelanggan Pada CV. Mataram Jaya Bawen menjadi 3 kelompok. Data yang digunakan merupakan data transaksi pelanggan periode September – Desember 2015. Hasil akhir jumlah pelanggan golden 28, silver 14 dan bronze 33, dengan keakuratan cluster adalah 59 %.

2.6.2 Penelitian Fina Nasari dan Charles Jhony Manto Slanturi (2016)

Sistem yang diajukan Fina Nasari dan Charles Jhony Manto Slanturi menerapkan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat. Data yang digunakan yaitu jumlah desa, jumlah kecamatan dan jumlah penderita diare pada Kabupaten Langkat. Hasil yang diperoleh pada cluster pertama yaitu kecamatan Bartang Serangan, Brandan Barat dan Pematang Jaya merupakan daerah penyebaran diare untuk jumlah penderita tingkat menengah atau bukan merupakan pusat penyebaran diare. Cluster kedua yaitu kecamatan Hinai, Sei Bingai dan Serapit merupakan daerah-daerah pusat penyebaran diare.

2.6.3 Penelitian Rima Dias Ramadhani (2014)

Sistem yang diajukan Rima Dias Ramadhani menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. Hasilnya terdapat tiga cluster yaitu, cluster satu dengan jumlah 804 mahasiswa dengan rata-rata IPK 3.16, cluster dua dengan jumlah 2792 mahasiswa dengan rata-rata IPK 3.15 dan cluster tiga dengan jumlah 223 mahasiswa dengan rata-rata IPK 3.2. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan cluster yang terbentuk adalah dengan mengirimkan tim admisi UDINUS yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan potensi akademik mahasiswa dengan penyesuaian menggunakan promotion mix.

2.6.4 Penelitian Mike Indra Istiana (2013)

Sistem yang diajukan Mike Indra Istiana menggunakan Segmentasi Pelanggan Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran pada LAROIBA Seluler. Algoritma K-Means menghasilkan tabel visual cluster dengan tools Rapidminer 5.2 yang merepresentasikan jumlah pelanggan di setiap kluster dengan menggunakan atribut RFM (Recency, Frequency dan Monetary). Kelompok pelanggan yang terbentuk adalah empat kelas dengan kluster pertama 4 pelanggan, kluster kedua 76 pelanggan, kluster ketiga 20 pelanggan dan kluster keempat dengan jumlah pelanggan 173. dari hasil tersebut dilakukan strategi potongan harga 10% yang digunakan untuk meningkatkan penerimaan perusahaan dari pelanggan dapat diterapkan pada kelas pertama dan kedua dengan cara mengirimkan SMS (Short Message Service) atau pemberitahuan informasi ke nomor pelanggan dalam database pelanggan LAROIBA Seluler dengan id pelanggan tersebut.

[Halaman ini sengaja dikosongkan]