

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Menurut *Turban et al* mengatakan bahwa “*Data mining* merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai *database* besar (*Kusrini, 2009*).

Data mining juga diartikan sebagai suatu proses otomatis data-data yang sangat besar dan bertujuan untuk mendapatkan hubungan atau pola yang memberikan manfaat. Menurut *Kusumo et al* menjelaskan bahwa data mining juga merupakan suatu proses pendukung pengambil keputusan dimana mencari pola informasi dalam data. Pencarian ini dapat dilakukan oleh pengguna. Pencarian ini disebut *discovery*. *Discovery* merupakan proses pencarian dalam basis data dalam menemukan pola yang tersembunyi tanpa ide yang didapatkan sebelumnya atau hipotesa tentang pola yang ada. Dengan kata lain aplikasi mengambil inisiatif untuk menemukan pola dalam data tanpa pengguna berpikir mengenai pertanyaan yang relevan terlebih dahulu (*Sari, 2013*).

Biasanya pengambilan keputusan ini berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil kesimpulan (*Susanto et al., 2015*). Berdasarkan penjelasan mengenai pengertian *data mining*, dapat disimpulkan bahwa *data mining* merupakan suatu proses pengambilan suatu keputusan yang bertujuan untuk membentuk suatu hubungan yang memberikan manfaat didalam proses *database*.

2.2 Pengelompokan *Data Mining*

Menurut Larose data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu (Kusrini *et al.*, 2009) adalah sebagai berikut:

1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, peninjauan estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contohnya adalah pada prediksi harga beras dalam tiga bulan yang akan datang.

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Contohnya adalah pada meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler.

2.3 Strategi Pemasaran

Menurut *Kotler* menjelaskan bahwa pemasaran merupakan suatu proses sosial dan manajerial yang didalamnya individu dan kelompok mendapatkan apa yang mereka butuhkan dan inginkan dengan menciptakan menawarkan, dan mempertukarkan produk yang bernilai kepada pihak lain (*Hartono et al.*, 2012).

Menurut (*Jensen, n.d.*) menjelaskan bahwa strategi pemasaran adalah “*marketing is the process of satisfying the organization’s stakeholders and creating value for these stake-holders*”. Penjelasan di atas menjelaskan bahwa pemasaran merupakan suatu proses yang dilakukan untuk memuaskan keinginan stakeholder dalam suatu organisasi dan menciptakan nilai bagi pemegang saham.

2.4 Aktifitas Pemasaran

Pada umumnya aktifitas pemasaran (*marketing activity*) yang dilakukan sebagai berikut (*Jensen, n.d.*) adalah sebagai berikut:

1. Meneliti pasar dan pelanggan.
2. Melakukan perencanaan dan mengelola diperlukan untuk mencapai perusahaan tujuan.
3. Penilaian yang tepat untuk harga produk atau layanan.
4. Mengembangkan produk baru dan produk-produk yang sudah ada.

2.5 Clustering

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 pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, 2007).

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Santoso, 2010).

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, atau lain sebagainya). Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki. Adapun metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010).

2.6 Algoritma *K-Means Clustering*

K-Means clustering merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster* atau kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* atau kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster* yang lain sehingga data yang berada dalam satu *cluster* memiliki tingkat variasi yang kecil (Agusta, 2007). Menurut Santosa (2007), langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut:

1. Pilih jumlah *cluster* k .
2. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.

3. Alokasikan semua data objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i,j) = (x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2 \dots \text{dst.}$$

Keterangan:

$D(i,j)$ = Jarak antara data dengan titik pusat awal *cluster* tinggi

i = Data Penjualan

j = Titik pusat awal (*centroid*) *cluster* tinggi

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

2.7 Keaslian Penelitian

Adapun dibawah ini adalah beberapa *literature* yang penulis gunakan dalam penelitian ini yaitu jurnal yang berkaitan dengan *Algoritma K-Means clustering* serta *Data Mining*, dapat dilihat pada tabel 2.1 adalah sebagai berikut:

Tabel 2.1 Tinjauan *Literature*

No	Nama Penulis	Judul Penelitian	Metode yang Digunakan	Hasil
1.	Andika Budi Pratama.	Verifikasi Citra Sidik Jari Poin <i>Minutiae</i> dalam Visum Et Repertum (VER) menggunakan <i>K-Means Clustering</i>	<i>K-Means Clustering</i>	Algoritma <i>K-Means clustering</i> dapat diimplementasikan dengan model perangkat lunak untuk verifikasi citra sidik jari poin.
2.	Yulia Darmi, Agus Setiawan.	Penerapan Metode Clustering K-Means dalam Pengelompokan Penjualan Produk.	<i>K-Means Clustering</i>	Algoritma <i>K-Means Cluster Analysis</i> pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran.
3.	Benri Melpa Metisen, Herlina Latipa Sari.	Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila.	<i>K-Means Clustering</i>	<i>Penerapan clustering dengan menggunakan Algoritma K-means. Dari data yang diolah dengan sampel data yang diambil di Swalayan Fadhilla Bengkulu, maka menghasilkan dua jenis kelompok data. Yaitu data penjualan rendah dan data penjualan tinggi.</i>

4.	Mhd Gading Sadewo, Agus Perdana Windarto, Dedy Hartama.	Penerapan <i>Data Mining</i> Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan <i>K-Means Clustering</i>	K-Means Clustering	Data akan diolah dengan melakukan <i>clustering</i> dalam 3 <i>cluster</i> yaitu <i>cluster</i> tingkat populasi tinggi, <i>cluster</i> tingkat populasi sedang dan rendah. <i>Centroid</i> data untuk <i>cluster</i> tingkat populasi tinggi 4711403141, <i>Centroid</i> data untuk <i>cluster</i> tingkat populasi sedang 304240647, dan <i>Centroid</i> data untuk <i>cluster</i> tingkat populasi rendah 554200.
5.	Anggraeni Triningsih, Heru Supriyono.	Aplikasi <i>Data Mining</i> Berbasis <i>Web</i> Menggunakan Metode <i>K-Means Clustering</i> Untuk Pengelompokan Penjualan Terlaris Produk Kacamata	<i>K-Means Clustering</i>	Hasil pengujian <i>blackbox</i> dapat disimpulkan bahwa sistem dapat berjalan dengan baik. Hasil pengujian validasi menghasilkan rata-rata validasi sebesar 98,48% sehingga dapat disimpulkan bahwa sistem sudah sesuai dengan kebutuhan dan dapat diterima oleh pemilik <i>optic</i> .

Berdasarkan penelitian sebelumnya dan tahap pemetaan masalah dimana menurut Andika Budi Pratama (2008) menjelaskan *Algoritma K-Means clustering* dapat diimplementasikan dengan model perangkat lunak untuk verifikasi citra sidik jari *point minutiae*. Menurut Yulia Darmi, Agus Setiawan (2016) menjelaskan Algoritma K-Means *Cluster Analysis* pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran serta menurut Anggraeni Triningsih, Heru Supriyono (2016) menjelaskan *Algoritma K-means clustering* bisa digunakan untuk pengelompokan penjualan terlaris produk kacamata. Maka penelitian ini diusulkan untuk menganalisa data penjualan pada PT. *Fast Food* Indonesia dengan menggunakan *Algoritma K-means Clustering*, karena *Algoritma K-means Clustering* sudah banyak digunakan pada penelitian sebelumnya selain itu juga *Algoritma K-means Clustering* mudah dipahami.