

BAB II

LANDASAN TEORI

A. Landasan Teori

1. Data Mining

Menurut Roiger J (2017: 529), Data Mining dapat didefinisikan sebagai berikut:

Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Data mining adalah proses menganalisa data dari yang berbeda dan menyimpulkannya menjadi informasi atau pengetahuan atau pola yang penting untuk meningkatkan keuntungan

Menurut Windarto (2018: 5), Data mining dapat didefinisikan sebagai berikut:

Data mining merupakan proses untuk menggali (mining) pengetahuan dan informasi baru dari data yang berjumlah banyak pada data *warehouse*, dengan menggunakan kecerdasan buatan (*Artificial Intelligence*), statistic dan matematika.

Dari definisi diatas dapat disimpulkan bahwa data mining adalah proses pengekstraksian informasi dari sekumpulan data dari perspektif yang berbeda dan menyimpulkan menjadi informasi-informasi penting yang dapat dipakai meningkatkan keuntungan

2. Clustering

a. Definisi Clustering

Menurut Sani (2017: 5), *Clustering* dapat didefinisikan sebagai berikut:

Clustering adalah proses pengelompokan satu set objek data (into multiple groups) atau cluster sehingga benda-benda dalam suatu kelompok memiliki kesamaan yang tinggi, tetapi sangat berbeda dengan objek di kelompok lain.

b. Tujuan Clustering

Tujuan *clustering* (pengelompokan) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan *clustering* untuk penggunaan. Windarto (2017: 50-51) Biasanya proses pengelompokan untuk tujuan pemahaman hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan seperti summarization (rata-rata, standar deviasi), pelabelan kelas untuk setiap kelompok sehingga dapat digunakan sebagai data training dalam klasifikasi *supervised*.

Contoh tujuan *clustering* untuk pemahaman diantaranya; dibidang Biologi (pengelompokan berdasarkan karakter tertentu secara hirarkis), pengelompokan gen yang memiliki fungsi sama. Dibidang *information retrieval (web search)*, bidang *klimatologi* (pengelompokan pola tekanan udara yang berpengaruh pada cuaca),

bidang bisnis (pengelompokan konsumen yang berpotensi untuk analisa dan strategi pemasaran).

3. Prediksi

Menurut Pambudi dkk (2020: 100-108), prediksi dapat didefinisikan sebagai berikut:

Prediksi (*forecasting*) adalah suatu kegiatan yang memperkirakan apa yang terjadi pada masa akan datang. Masalah pengambilan keputusan merupakan masalah yang dihadapi maka peramalan juga merupakan masalah yang harus dihadapi, karena peramalan berkaitan erat dengan pengambilan suatu keputusan.

Menurut Hay's dkk, (2017 : 29-32) prediksi dapat didefinisikan sebagai berikut:

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin / setepat mungkin yang akan terjadi.

Dari definisi diatas dapat diambil kesimpulan bahwa prediksi adalah suatu proses atau kegiatan yang memperkirakan secara sistematis

tentang suatu yang akan terjadi pada masa yang akan datang berdasarkan informasi masa lalu agar dapat memperkecil kesalahan.

4. Sistem

Menurut Ferdika & Kuswara (2017: 175-188), sistem dapat didefinisikan sebagai berikut:

sistem adalah sekumpulan objek-objek yang saling berelasi dan berinteraksi serta hubungan antara objek bisa dilihat sebagai satu kesatuan yang dirancang untuk mencapai satu tujuan. Dengan demikian, secara sederhana sistem dapat diartikan sebagai suatu kumpulan atau himpunan dari unsure variabel-variabel yang saling terorganisasi, saling berinteraksi, dan saling bergantung satu sama lain.

Sistem memiliki tiga komponen atau fungsi dasar yang berinteraksi, antara lain yaitu sebagai berikut:

1. *Input* (masukan) melibatkan penangkapan dan perakitan berbagai elemen yang memasuki sistem untuk diproses. *Input* yang dimaksud dalam hal ini berupa keseluruhan penginputan data yang berkaitan dengan transaksi dalam siklus pendapatan dan pengeluaran yang dilakukan oleh pihak yang berwenang.
2. Proses melibatkan tahap *transformasi* yang mengubah input menjadi *output*. Yang dimaksud tahap disini mencakup penghitungan dan kalkulasi dari data-data transaksi siklus pendapatan dan pengeluaran yang masuk ke sistem.

3. *Output* (keluaran) melibatkan perpindahan elemen yang telah diproduksi oleh proses. *Output* yang dimaksud adalah laporan keuangan dan laporan produk yang berhasil dijual yang dihasilkan dari sistem informasi akuntansi *revenue cycle*.

5. Metode K-Means

Menurut Metisan & Sari (2017: 111-115), metode *K-Means* dapat didefinisikan sebagai berikut:

K-Means merupakan *algoritma* yang menetapkan nilai-nilai *cluster* (*k*) secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* yang biasa disebut *centroid*. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus Euclidean hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid* hingga nilai *centroid* tidak berubah (stabil).

Berikut dasar *algoritma K-means* adalah:

- a. Tentukan nilai *k* sebagai jumlah *klaster* yang ingin dibentuk.
- b. Bangkitkan *k centroid* (titik pusat *klaster*) awal secara random/acak, kemudian untuk menghitung *centroid cluster k-i* berikutnya, digunakan rumus sebagai berikut:

$$v = \sqrt{\frac{\sum_{i=1}^n x_i}{n}} \quad \dots\dots\dots(2.1)$$

Dimana:

$i = 1, 2, 3, \dots, n$

$v = \text{centroid pada cluster}$

$x_i = \text{objek ke-}i$

n = banyaknya objek/jumlah objek yang menjadi anggota *cluster*.

- c. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \dots\dots\dots(2.2)$$

Keterangan :

d_{ij} = Jarak objek antara objek x dan y

n = Jumlah Atribut

X_i = Objek Data

Y_i = Data *Cluster*

- d. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya* (C).
- e. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

6. Evaluasi Kinerja

Menurut Mustofa dkk (2017: 151-162), evaluasi kinerja dapat didefinisikan sebagai berikut :

Evaluasi kinerja dilakukan untuk memberikan penilaian terhadap hasil kerja atau presentasi kerja yang diperoleh organisasi, tim atau individu. Evaluasi kinerja akan memberikan umpan balik terhadap tujuan dan sasaran kinerja, perencanaan dan proses pelaksanaan kinerja. Evaluasi kinerja dapat pula dilakukan dan proses penilaian, *review* dan pengukuran kinerja.

Menurut Sari dkk (2020: 53-68), evaluasi kinerja dapat didefinisikan sebagai berikut :

Evaluasi kinerja adalah memandang sebagai suatu proses mengevaluasi kinerja pekerja, membagi informasi dengan mereka, dan mencari cara memperbaiki kinerjanya. Evaluasi kinerja adalah suatu metode dan proses penilaian pelaksanaan tugas seseorang atau sekelompok orang atau unit kerja dalam suatu perusahaan atau organisasi sesuai dengan standar kinerja atau tujuan yang ditetapkan terlebih dahulu.

Sedangkan pengertian evaluasi kinerja *k-means* itu sendiri adalah suatu proses evaluasi kinerja yang mengukur tingkat keakurasian dari metode *k-means*.

7. Davies Bouldin Index (DBI)

Menurut Jumadi (2018: 8-10), *Davies Bouldin Index* dapat didefinisikan sebagai berikut :

Davies-Bouldin Index merupakan salah satu metode yang digunakan untuk mengukur validitas *cluster* pada suatu metode pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat cluster dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar titik pusat cluster terhadap *clusternya*. Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak *inter-cluster* antara *cluster* C_i dan C_j dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah *cluster*. Jika jarak *inter-*

cluster maksimal, berarti kesamaan karakteristik antar-masing-masing *cluster* sedikit sehingga perbedaan antar-cluster terlihat lebih jelas. Jika jarak *intra-cluster* minimal berarti masing-masing objek dalam *cluster* tersebut memiliki tingkat kesamaan karakteristik yang tinggi. Berikut adalah rumus untuk menghitung *Davies Bouldin Index*.

Rumus menghitung *sum of within cluster (SSW)* untuk menghitung matrik kohesi dalam sebuah *cluster* ke-i yaitu:

$$SSWi = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad \dots\dots\dots(2.3)$$

Keterangan :

m_i = jumlah data dalam *cluster* ke-i

c_i = *centroid cluster* ke-i

$d(x_j, c_i)$ = jarak *euclidean* setiap data ke *centroid*

Rumus menghitung *sum of square between cluster (SSB)* untuk mengetahui nilai separasi antara *cluster* yaitu :

$$SSBi,j = d(c_i, c_j) \quad \dots\dots\dots (2.4)$$

Keterangan :

$d(c_i, c_j)$ = jarak antar *cluster*

Setelah kohesi dan separasi diperoleh kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j.

$$R_{i,j} = \frac{SSWi + SSWj}{SSBi} \quad \dots\dots\dots (2.5)$$

Menghitung nilai *Davies Bouldin Index (DBI)*.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_i, j) \quad \dots\dots\dots (2.6)$$

Keterangan : k = jumlah *cluster* yang digunakan.

B. Kajian Pustaka

Dalam perancangan sebuah penelitian dibutuhkan kajian pustaka yang memuat penelitian-penelitian terdahulu untuk dijadikan sumber referensi sebagai berikut :

1. Oleh Siregar (2018: 83-91) fakultas teknik, universitas islam kuantan singing dengan judul *klasifikas* penjualan alat-alat bangunan menggunakan metode *k-means* studi kasus toko bangunan adi bangunan. Pada penelitian ini mengangkat masalah data penjualan dan pembelian ditoko adi bangunan tidak tersusun dengan baik sehingga data yang ada hanya sebagai arsip toko, oleh karena itu diperlukan solusi untuk memecahkan permasalahan tersebut untuk meningkatkan pemasaran pada toko adi bangunan. Dengan adanya permasalahan tersebut diperoleh solusi yaitu dengan membuat sistem pengambil keputusan dengan menggunakan data yang ada, teknologi data mining digunakan sebagai solusi serta dukungan infrastruktur dibidang teknologi dengan menerapkan algoritma *k-means* pengelompokan dipilih dalam masalah tersebut.
2. Oleh Indiriyani & Irfani (2019: 109-113) fakultas teknologi informasi, universitas bina sarana informatika dengan judul *clustering* data penjualan toko perlengkapan outdoor menggunakan metode *k-means*

method. Pada penelitian ini mengangkat masalah stok persediaan barang ditoko genta corp antara barang yang laku dan tidak laku belum terstruktur dengan baik karena pengelolaan data penjualan masih dilakukan secara manual yaitu menggunakan buku jurnal. Oleh karena itu diperlukan peramalan pendukung keputusan untuk mengatasi permasalahan tersebut perhitungan manual menggunakan metode *k-means* dan *software* menggunakan *rapid miner* dirasa sangat tepat untuk menyelesaikan permasalahan tersebut.

3. Oleh Normah, dkk (2021: 159-163) universitas bina sarana mandiri, universitas bina sarana informatika, universitas nusa mandiri dengan judul penerapan data mining metode k-means clustering untuk analisa penjualan pada toko fashion hijab banten. Pada penelitian ini mengangkat masalah toko helai bergerak dibidang penjualan fashion muslim data penjualan dan pembelian ditoko helai tidak tersusun dengan baik sehingga data yang ada hanya sebagai arsip toko dan tidak dimanfaatkan untuk perkembangan strategi pemasaran toko helai. Oleh kerena itu diperlukan data mining menggunakan metode *k-means* clustering untuk menentukan produk laris dan tidak laris, kemudian dipilih 3 cluster secara acak dan menerapkan metode *k-means* pada *rapid miner* dirasa tetap untuk menyelesaikan permasalahan ditoko helai.
4. Oleh Kesuma & Feryanto (2019: 67-71) Fakultas ilmu komputer, universitas prima indonesia dengan judul penerapan data mining untuk menentukan penjualan sparepart toyota dengan metode *k-means*

clustering. Pada penelitian ini mengangkat masalah CV.Terang jaya merupakan perusahaan yang bergerak dibidang otomotif yang melayani penjualan dan pembelian sperepat mobil. Namun demikian kurang dalam peninjauan produk apa saja yang dibutuhkan konsumen dan penyimpanan data kurang efektif, untuk mengatasi permasalahan tersebut analisis yang digunakan yaitu penerapan clustering dengan menggunakan algoritma *k-means* merupakan *algoritma* pengelompokan sejumlah data. Sehingga dengan adanya pengelompokan data perusahaan bisa mengetahui barang mana yang laris dan tidak laris terjual.

5. Oleh Annur (2019: 40-45) Universitas ichsan gorontalo dengan judul penerapan data mining menentukan strategi penjualan variasi mobil menggunakan metode *k-means clustering*. Pada penelitian ini mengangkat masalah toko luxor variasi mobil gorontalo merupakan toko terbesar yang bergerak di bidang variasi mobil dengan produk utama yaitu kaca film mobil,tape mobil, sarung jok mobil, dan aksesoris lainnya. Dimana perusahaan ini setiap harinya harus memenuhi kebutuhan konsumen, untuk dapat melakukan hal tersebut toko luxor variasi membutuhkan strategi penjualan untuk dapat menarik konsumen.
6. Dari beberapa jurnal diatas di dapatkan beberapa perbedaan dari judul yang akan diteliti yaitu jumlah data yang digunakan lebih banyak dari penelitian sebelumnya dan topik yang akan diteliti belum pernah digunakan pada penelitian sebelumnya.