

BAB II

TINJAUAN PUSTAKA

A. Landasan Teori

Pada bab landasan teori memaparkan definisi, konsep serta proposisi tentang *variable* dari sebuah penelitian. Pada bab landasan teori ini memberikan pemahaman mengenai teori – teori yang relevan dengan penelitian yang sedang dilakukan. Landasan teori pada penelitian yang dilakukan adalah sebagai berikut :

1. Data mining

Menurut Dito Putro Utomo (2019: 847), *Data mining* dimaknai sebagai berikut :

Data mining merupakan sekumpulan proses yang berguna mengeksplorasi dan mencari nilai berupa informasi juga relasi – relasi kompleks yang selama ini tersimpan dari suatu basis data. Dengan melakukan penggalian pola informasi terhadap data yang berguna sebagai memanipulasi data menjadi sebuah informasi baru serta lebih bermanfaat yang didapatkan melalui cara mengekstraksi juga mengetahui pola - pola yang berharga atau menarik yang didapatkan dari data didalam basis data.

Data mining digunakan untuk pengelolaan data yang besar serta membantu proses penyimpanan data transaksi dan memproses data warehousing agar kiranya bisa mendapatkan informasi yang diperlukan bagi penggunaanya

Pada penelitian yang lain Dito Putro Utomo (2020: 439) juga memaparkan pengertian *data mining* yaitu sebagai berikut :

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, *statistic*, *database*, dan *visualisasi* untuk penanganan permasalahan pengambilan informasi dari *database* yang besar

Pengertian *data mining* yang lain yaitu menurut Dini Silvi

Purnia (2017:31) *data mining* yaitu sebagai berikut :

Data mining adalah 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 yang terikat dari berbagai *database* besar. Berdasarkan pengertian *data mining* yang telah dijelaskan di atas, maka *data mining* merupakan pengetahuan yang tersembunyi di dalam *database* yang di proses untuk menemukan pola dan teknik statistik matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan dari *database* tersebut

Dari penertian *data mining* diatas, dapat disimpulkan bahwa *data mining* adalah eksplorasi atau proses pencarian data dan analisis data dari suatu data yang besar untuk menemukan pola – pola yang berguna dan mempunyai arti.

2. *K-means Clustering*

Menurut Rony Setiawan (2016: 78) *Algoritma k-means* adalah sebagai berikut :

Salah satu *algoritma Clustering* pada *data mining* adalah *algoritma K-means Clustering* untuk dapat menghasilkan

kelompok yang memiliki kemiripan atribut yang sama. *Algoritma* pengelompokan *K-Means* yang akan menghasilkan kelompok catatan sebanyak k buah.

Langkah-langkah melakukan *Clustering* dengan metode *K-Means* adalah sebagai berikut:

- a. Pilih jumlah *cluster* k .
- b. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusatpusat *cluster* diberinilai awal dengan angka-angka random.
- c. 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:

$$Y^{ij} = \sum \sqrt{(X1 - X2)^2} \dots\dots\dots(2.1)$$

Y^{ij} = jarak antara titik pusat dan objek

$X1$ = titik pusat

$X2$ = Objek

- d. 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.
- e. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *Clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

Selain pengertian *k-means* diatas, Iin Parlina (2018:87) juga mamaparkan pengertian algoritma *K-means Clustering* yaitu sebagai berikut :

Algoritma K-means Clustering yang merupakan metode analisis kelompok yang mengarah pada partisian N objek pengamatan ke dalam K kelompok, di mana setiap objek

pengamatan sebuah kelompok data dengan *mean* (rata-rata) terdekat.

Algoritma K-means Clustering merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *K cluster* yang sudah ditetapkan di awal. *Algoritam K-means Clustering* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek.

Dari berbagai pengertian mengenai *algoritma K-means Clustering* diatas, dapat disimpulkan bahwa *algoritma K-means Clustering* merupakan suatu metode penganalisaan atau metode *data mining* yang melakukan proses pemodelan dan merupakan pengelompokan data dengan sistem partisi

B. Kajian Pustaka

1. Peneleitian Terkait

Dalam penelitian yang di lakukan oleh Benri Melpa Metisen dan Herlina Latipa Sari dalam artikelnya yang berjudul “Analisis *Clustering* Menggunakan Metode *K-means Clustering* Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila” menjelaskan bahwa swalayan Fadhilla masih memproses data penjualan secara manual. Swalayan Fadhilla juga tidak dapat mengelompokkan produk yang laris dan yang tidak laris terjual. Sehingga kesulitan yang dialami yaitu seringnya kekurangan stok produk yang laku karena penjualannya tinggi. Dan menumpuknya produk yang tidak laku di gudang karena penjualannya rendah. Dari

permasalahan tersebut dibutuhkan sistem informasi terkomputerisasi yang menunjang arus data dan informasi sesuai dengan kebutuhan dari proses- proses tersebut. Dalam pengembangannya, system tersebut memakai *algoritma K-means Clustering*.(Metisen, 2015 :110)

Elly Muningsih dan Sri Kiswati dalam artikelnya yang berjudul “Penerapan metode *K-Means* untuk *Clustering* Produk *Online Shop* dalam Penentuan Stok Barang” memaparkan bahwa pada online shop ragam jogja memiliki jumlah permintaan dari konsumen yang fluktuatif mengakibatkan stok yang harus disiapkan ragam jogja menjadi tidak stanil. Selain itu produk yang beragam dan banyak jenisnya menjadikan manajemen stok yang dilakukan menjadi tidak akurat. Manajemen stok yang tidak akurat akan mengakibatkan biaya simpan yang tinggi dan tidak ekonomis, selain itu juga mengakibatkan sering terjadi kekurangan atau kelebihan stok produk tertentu. Maka dari itu di butuhkan system untuk penentuan stok barang dengan menerapkan metode *K-means Clustering*. (Muningsih, 2015 : 10)

Yulia Darmi dan Agus Setiawan juga melakukan penelitian tentang K-Means dengan judul “Penerapan Metode *Clustering K-means Clustering* dalam Pengelompokan Penjualan Produk”. Yulia Darmi dan Agus Setiawan menyimpulkan bahwa pengklasteran dengan menggunakan metode *K-means Clustering* didapatkan untuk produk yang laku dengan yang tidak laku. Semakin banyak data penjualan barang yang diinput, maka semakn banyak *clustering*

centroid yang terbentuk di proses metode k-means. (Darmi, 2016 : 157)

2. Penelitian yang di lakukan

Penelitian yang dilakukan yaitu mengenai penerapan *algoritma k-means* untuk memprediksi stok bahan baku produksi. Perbedaan penelitian yang dilakukan dengan penelitian sebelumnya ialah objek penelitian. Objek penelitian yang di ambil ialah pada UD. Maharani Toys, sebuah usaha mikro kecil menengah (UMKM) yang memproduksi mainan edukasi anak usia dini. Permasalahan yang ada di UD. Maharani toys adalah penumpukan stok bahan baku produksi di karenakan kurang tepatnya dalam menentukan prioritas stok bahan baku sehingga di butuhkan sebuah system yang mampu membantu dalam perhitungan atau prediksi stok bahan baku produksi di UD. Maharani Toys. Perhitungan dalam memprediksi stok bahan baku produksi menggunakan *Algoritma K-means Clustering*.