

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) adalah metode yang digunakan untuk dapat memperoleh pengetahuan yang berasal dari database yang ada. Hasil pengetahuan yang diperoleh dapat dimanfaatkan untuk basis pengetahuan (*knowledge base*) yang digunakan dalam keperluan mengambil keputusan (Gustientiedina, *et al.* 2019). *Data mining* adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara iteratif/ perulangan pada *database* yang besar, dengan tujuan memberikan informasi dan pengetahuan yang akurat dan berpotensi berguna untuk kerja pengetahuan yang berhubungan dengan pengambilan keputusan dan pemecahan masalah (M. Nanda Variestha Waworuntu dan Muhammad Faisal Amin, 2018).

Salah satu metode yang dapat digunakan dalam *Data Mining* adalah metode *Clustering* dengan algoritma *K-Means*. *K-Means* merupakan algoritma *clustering* yang berulang-ulang. Algoritma *K-Means* dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau *means* (Randi Rian Putra dan Cendra Wadisman, 2018).

Data Mining dapat diimplementasikan mengelompokkan data-data di berbagai kasus, seperti pengelompokkan Indikator HIV/AIDS di Indonesia berdasarkan dari 34 provinsi yang ada di Indonesia (Kiki Fatmawati dan Agus Perdana Windarto, 2018). Data Mining dengan algoritma *K-Means Clustering* juga pernah dimanfaatkan dalam menentukan pegawai yang layak mengikuti *Assessment Center* untuk program SDP (Iin Parlina, *et al.* 2018). Ada sebuah penelitian yang menggabungkan Metode *K-Means* dengan PSO dari hasil penelitian ini Hasil kinerja RMSE dari *K-Means* PSO terbukti meningkatkan akurasi untuk dataset, tapi *K-Means* PSO membutuhkan proses waktu lebih lama dari *K-Means* standar. Untuk metode pengukuran jarak *K-Means* PSO mendapatkan hasil nilai RMSE Euclidean 128, Manhattan 132, Canberra 146 dan *K-Means* standar mendapatkan hasil nilai RMSE euclidean 140, Manhattan 160, Canberra 160 (Elindra Ambar Pambudi, *et al.* 2018).

PT. Bougenville Anugrah merupakan sebuah perusahaan sebagai distributor dari PT. Pesona Amaranthine Cosmetiques yang bergerak dibidang Kosmetik Mazaya. Karena banyaknya data produk mazaya dan stok barang yang harus dianalisa, maka seorang manajemen harus bekerja keras dalam menganalisa dan menentukan barang yang akan dibeli berikutnya dilihat dari stok yang ada.

Berdasarkan permasalahan yang telah dijelaskan maka penelitian ini nantinya akan menghasilkan informasi mengenai pengelompokan penjualan barang produk mazaya dari yang “SANGAT LARIS”, “LARIS” dan “TIDAK LARIS”, selanjutnya akan diperhitungkan dari hasil penjualan yang ada menggunakan metode *K-Means Clustering* dengan aplikasi *RapidMiner Studio*, kemudian hasilnya dapat dijadikan pedoman atau dasar pengetahuan dalam merestok ulang barang khususnya untuk produk mazaya di PT. Bougenville Anugrah.

1.2 Perumusan Masalah

Berdasarkan uraian yang terdapat pada latar belakang diatas, maka permasalahan pokok yang akan dibahas yaitu :

1. Bagaimana menerapkan algoritma *K-Means Clustering* mampu mengelompokkan dari Sangat Laris, Laris dan Tidak Laris berdasarkan analisa stok produk mazaya?
2. Bagaimana menerapkan algoritma *K-Means Clustering* dengan menggunakan aplikasi *RapidMiner* mampu mengelompokkan dari Sangat Laris, Laris dan Tidak Laris berdasarkan analisa stok produk mazaya?

1.3 Batasan Masalah

Agar penulisan ini lebih terarah dan tujuan yang diharapkan dapat tercapai, maka penulis menetapkan batasan-batasan terhadap masalah yang akan diteliti dalam melakukan penelitian ini penulis memberikan batasan-batasan seperti berikut:

1. Algoritma yang digunakan dalam melakukan *Clustering* data produk mazaya adalah algoritma *K-Means*.
2. Data yang dijadikan rujukan dalam penelitian ini adalah data penjualan produk mazaya yang ada pada PT. Bougenville Anugrah.
3. Pengelompokan (*Clustering*) yang digunakan dalam menganalisis data stok terhadap produk mazaya terdapat 3 *cluster* yaitu Sangat Laris, Laris dan Tidak Laris.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menerapkan algoritma *K-Means* dalam analisa stok.
2. Menganalisa dan menguji hasil perhitungan manual menggunakan aplikasi RapidMiner untuk mengetahui minat pelanggan dari produk yang telah dikelompokkan Sangat Laris, Laris dan Tidak Laris dengan metode *K-Means*.

1.5 Manfaat Penelitian

Manfaat yang akan di capai dalam penelitian ini adalah :

1. Dengan menerapkan sistem *K-Means* dapat meningkatkan keuntungan penjualan PT. Bougenville Anugrah karena tidak ada lagi stok yang kurang sehingga dapat memenuhi semua kebutuhan pelanggan.
2. Dengan pengelompokkan ini dapat membantu pengelompokkan produk yang laris maupun yang tidak laris sesuai dengan minat pelanggan berdasarkan produk yang terjual.

1.6 Sistematika Penulisan

Adapun sistematika dari penulisan ini adalah sebagai berikut:

BAB I : Pendahuluan

Pada bab ini berisi Latar Belakang , Perumusan Masalah, Batasan Masalah , Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II : Landasan Teori

Pada bab ini membahas tentang konsep *Knowledge Discovery In Database* (KDD), Data Mining, *Clustering*, Metode K-Means, Algoritma K-Means, Analisis dan Stok.

BAB III : Metodologi Penelitian

Pada bab ini menjelaskan tentang kerangka kerja penelitian dan langkah-langkah dalam penelitian.

BAB IV : Analisa dan Pembahasan

Pada bab ini menjelaskan prinsip kerja penerapan teknik Data Mining yaitu Algoritma Metode K-Means dengan menggunakan aplikasi RapidMiner.

BAB V : Implementasi dan Hasil

Pada bab ini memuat tentang implementasi dan pengujian dengan menggunakan aplikasi RapidMiner.

BAB VI : Kesimpulan dan Saran

Pada bab ini berisi kesimpulan yang dapat ditarik setelah semua proses penelitian dilakukan, serta saran-saran yang dapat penulis berikan terkait dengan beberapa kekurangan yang ditemui selama proses penelitian berlangsung.