

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) adalah proses yang melibatkan sekumpulan data, dimana tahapan pada KDD digunakan untuk pencarian data di dalam sebuah dataset. (Sudarsono and Lestari, 2021) Proses *Knowledge Discovery in Database* diawali dengan menetapkan tujuan dan diakhiri dengan evaluasi (Permadi dan Wijaya, 2023). Tahapan penelitian dilakukan dengan mengacu pada proses *Knowledge Discovery* sebagai proses dari penggunaan metode *Data Mining* untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data. Terdapat lima tahap dalam proses *Knowledge Discovery in Database* yaitu *selection, preprocessing, transformation, Data Mining, dan interpretation/evaluation* (Mirantika, et al., 2023).

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih, (Gunawan and Purwayoga, 2022) dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan (Romadhona, et al., 2022). Salah satu metode yang terjadi di dalam *Data Mining* adalah *Clustering* (Hidayat, 2022)

Clustering merupakan salah satu metode *Clustering analysis* non-hirarki (Praseptian M and Fadlil, 2022) yang berusaha untuk mempartisi objek yang ada kedalam satu atau lebih *Cluster*. (Pangestu, et al., 2023) *Clustering* dapat pengelompokan item data ke dalam kelompok-kelompok yang lebih kecil sehingga setiap kelompok memiliki kemiripan yang signifikan (Nurrohman and Sukmasetya, 2023)

Metode *K-Means Clustering* merupakan metode unsupervised learning untuk mendefinisikan pusat *Cluster* dan pengelompokan data berdasarkan data yang sama (Indah, *et al.*, 2023) Algoritma *k-means clustering* merupakan teknik perulangan, numerik, tanpa pengawasan dan non-deterministik (Raheem and Alabbas, 2022). Algoritma *K-Means Clustering* saat ini banyak digunakan pada perusahaan-perusahaan berskala besar yang memiliki banyak sekali data, salah satunya pada perusahaan teknologi transportasi karna sangat mudah untuk diimplementasikan (Rahmatika, Sedyono and Widodo, 2020). Tidak jarang suatu pabrik transportasi mengeluarkan banyak sekali model kendaraan, setidaknya pada tahun 2022 perusahaan mobil Toyota telah menjual 300 ribu lebih unit mobil dari semua type.

Dari banyaknya total penjualan mobil ini tentu saja pengelompokan data penjualan akan sulit kalau hanya dilakukan dengan perhitungan secara manual dikarenakan penjualan produk Toyota tersebar di seluruh provinsi di Indonesia. Bagi konsumen dengan banyaknya jenis kendaraan yang ada dipasaran konsumen perlu riset dalam menentukan jenis kendaraan yang sesuai dengan kebutuhan, seperti riset perbandingan harga, jenis mobil, kapasitas penumpang, dan lain-lain.

Penelitian terdahulu menggunakan metode *K-Means Clustering* untuk menentukan penjualan *sparepart* Toyota dengan metode *K-Means Clustering*, data penjualan yang digunakan sebanyak 50 data. Output yang diharapkan adalah menghasilkan 3 *Clustering* yaitu (C1) data penjualan paling laris, (C2) data penjualan yang laris, dan (C3) data penjualan kurang laris (Kesuma dan Tamba, 2020).

Penelitian selanjutnya menggunakan metode *K-Means Clustering* untuk prediksi penjualan mobil dengan data riset berdasarkan pada website resmi GaiKindo sebanyak 900 data. *Cluster* yang terbentuk setelah dilakukan proses *K-Means Clustering* terbagi menjadi tiga *Clustering* yaitu *Cluster 0* jumlah anggota 235 dengan presentase 26% dikategorikan Laris, *Cluster 1* jumlah anggota 604 dengan presentase 67% dikategorikan Kurang Laris, dan *Cluster 2* jumlah anggota 61 dengan presentase 7% dikategorikan Paling Laris, dari proses *Clustering* di atas dapat diperoleh validasi DBI (Davies Bouldin Index) dengan nilai 0,341 (Butsianto dan Mayangwulan, 2020).

Berdasarkan penelitian yang ada sebelumnya, penulis melakukan penelitian dengan judul Penerapan Algoritma *K-Means Clustering* Dalam Pengelompokan Data Penjualan Pada Pabrik Mobil Toyota Indonesia.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan di atas, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana mengelompokkan data penjualan ke dalam sebuah *Clustering* dengan metode Algoritma *K-Means Clustering*?
2. Bagaimana metode mencari mobil yang sesuai berdasarkan minat pasar menggunakan algoritma *K-Means Clustering*?
3. Bagaimana cara menerapkan konsep *Data Mining* menggunakan algoritma *K-Means (Clustering)* untuk menganalisa serta menentukan produk potensial?

1.3 Batasan Masalah

Berdasarkan pada perumusan masalah yang telah dijelaskan di atas, agar penelitian dapat memberikan pemahaman yang terarah maka diberi batasan-batasan sebagai berikut:

1. Data penelitian yang digunakan adalah data penjualan mobil Toyota.
2. Pengujian dilakukan pada aplikasi *RapidMiner* dengan menggunakan tool-tool yang ada pada menu aplikasi.

1.4 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai penulis yaitu:

1. Memahami dan mengetahui konsep *Data Mining* untuk mengelompokkan data penjualan menggunakan Algoritma *K-Means Clustering*.
2. Analisa data tingkat penjualan produk mobil menggunakan Algoritma *K-Means Clustering*.
3. Merancang model yang dihasilkan dari Algoritma *K-Means Clustering* berdasarkan parameter yang digunakan.
4. Implementasi *Data Mining* menggunakan Algoritma *K-Means Clustering* untuk menentukan produk potensial.
5. Menguji data penjualan dengan Algoritma *K-Means Clustering* menggunakan *Tools RapidMiner*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Dengan metode *Data Mining* Algoritma *K-Means Clustering* peneliti dapat mengelompokkan data penjualan ke dalam sebuah *Cluster*.
2. Dengan menggunakan metode *K-Means Clustering* dapat merekondisikan dan menentukan jenis mobil yang sesuai.
3. Dengan konsep *Data Mining* menggunakan algoritma *K-Means Clustering* dapat menganalisa serta menentukan produk potensial.
4. Dengan adanya hasil clusterisasi data dapat digunakan perusahaan sebagai rujukan dalam pengambilan keputusan, kebijakan, dan strategi promosi.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian ini terdiri dari 5 bab yang saling berkaitan sesuai dengan rumusan masalah yang ada sebagai berikut:

BAB I PENDAHULUAN

Pendahuluan berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Landasan teori berisi pembahasan tentang konsep dan penerapan *Data Mining, Clustering*, dan Algoritma *K-Means Clustering*.

BAB III METODOLOGI PENELITIAN

Metodologi Penelitian membahas tentang metode yang digunakan dalam analisa dan penggunaan Algoritma *K-Means Clustering*.

BAB IV ANALISA DAN PERANCANGAN

Pada bab ini akan dibahas tentang analisa dan pengolahan data yang meliputi: analisa data, hasil *Clustering, Centroid* dan *Iterasi*.

BAB V IMPLEMENTASI DAN HASIL

Pada implementasi dan hasil akan dibahas bagaimana implementasi Algoritma *K-Means Clustering* digunakan, dan pengujian data dengan menggunakan *Software RapidMiner*.

BAB VI KESIMPULAN DAN SARAN

Bab ini menyimpulkan hasil dari penelitian yang telah dilakukan serta memberikan saran-saran yang relevan untuk penelitian berikutnya