

BAB I

PENDAHULUAN

1.1 Latar Belakang

Melakukan esktraksi data dalam perkembangan saat ini terdapat beberapa metodologi yang dapat digunakan salah satunya yaitu *Knowledge Discovery in Database* (KDD). KDD merupakan suatu proses analisa terstruktur yang bertujuan untuk mendapatkan informasi terbaru dan benar, serta pola dari suatu data yang kompleks dan berguna (Fakhri dkk, 2021). KDD memiliki arti proses dari pemahaman dalam bidang aplikasi, pengolahan data target dari data mentah dari dalam database, serta *preprocessing* dan pembersihan data (Virgo dkk, 2021). Salah satu tahapan inti dari KDD adalah *data mining* (Fakhri dkk, 2021).

Data mining berperan penting dalam ekstraksi informasi tersembunyi atau pola dan hubungan dalam jumlah data yang besar. Pada *data mining* proses ekstraksi dan pemilihan fitur menjadi komponen yang sangat penting. Selain itu, pada saat jumlah fitur deskriptif meningkat terjadi peningkatan pada waktu komputasi. Salah satu yang termasuk metode dalam *Data Mining* adalah *clustering* (Mohammadi dkk, 2021).

Clustering atau yang biasa disebut pengelompokkan adalah sebagai alat yang digunakan dalam bidang ilmu data (Sinaga & Yang, 2020). Secara umum, *cluster* adalah sekumpulan objek yang lebih mirip antara satu dengan yang lain dari pada anggota *cluster* lainnya (Singh & Singh, 2020). *Clustering* termasuk kategori *unsupervised learning* yang memiliki fungsi mempartisi data tanpa penamaan label ke dalam satu kelompok data (Elda dkk, 2021). Algoritma K-Means sebagai proses kuantisasi data yang sering digunakan dalam klasifikasi data (Paulino dkk, 2020).

Pada K-Means ini menggunakan metode *data clustering non hierarchical* yang mana Jarak antara data ke klasternya harus diminimalisir (Virgo dkk, 2021).

Dari pembahasan penelitian yang dilakukan untuk menarik kesimpulan melalui *data mining* wilayah kelayakan tanam jagung dengan menggunakan algoritma K-Means *clustering*. Terdapat dua kelompok wilayah dengan jumlah panen jagung. Kelompok terbanyak adalah Penengahan dengan 79448.30257 dengan 2 nilai pusat. Untuk wilayah dengan jumlah panen jagung terendah adalah Candipuro dengan sentroid 1424.036868 dengan 2 nilai untuk ditingkatkan produksinya (Aldino dkk, 2021).

Berdasarkan hasil penelitian dari analisis *cluster* kepuasan pengguna terhadap layanan Shopee dengan nilai *cluster* yang digunakan 2, 3, 4 dan 5. Data yang digunakan terdiri dari 8 variabel data seperti data pribadi, *reliabilitas*, *information quality*, data *web design and layout*, adanya *interaction quality*, efisien, kontak, dan terakhir *user satisfaction*. Mengevaluasi *cluster* menggunakan *Davies Bouldin Index (DBI)* yang sudah dibentuk mendapatkan DBI paling kecil merupakan *cluster* yang optimal. Hasil yang didapatkan kepuasan responden terdapat dalam variabel *reliabilitas*, kualitas interaksi, efisien dan kontak cepat (Patimah dkk, 2021).

Pada penelitian ini algoritma K-Means *clustering* berfungsi sangat baik untuk menentukan persediaan barang. Keuntungan menggunakan algoritma ini menjadikan manajemen mampu mengatur stok persediaan barang. Terdapat tiga *cluster* dimana *cluster* 1 terdapat dua barang, *cluster* 2 terdapat 9 barang dan *cluster* 3 sisa dari 25 barang lainnya (Nurdiyansyah & Akbar, 2021).

Dari penerapan algoritma K-Means *clustering* dapat menghasilkan pemilihan dan pembagian kelas unggulan. Data kemampuan siswa yang digunakan diambil dari 119 data siswa yang dibagi menjadi 3 *cluster*. *Cluster* tertinggi terdapat pada kelas A dengan 40 siswa, nilai rata-rata sedang terdapat pada kelas B dengan 37 siswa dan nilai rata-rata rendah terdapat pada kelas C dengan 42 siswa. Dengan ini dapat memberikan kemudahan solusi dalam menentukan kelas unggulan (Novi & Mubarak, 2021).

Dari uraian diatas maka penulis melakukan penelitian dalam bentuk tesis dengan judul “Data Mining Menggunakan Algoritma K-Means *Clustering* dalam Analisis Tingkat Potongan Harga terhadap Harga Jual Sepeda Motor Honda”.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan sebelumnya agar penelitian sesuai dengan tujuan yang ingin di capai, maka didapatkan perumusan masalah sebagai berikut:

1. Bagaimana menerapkan Algoritma K-Means *Clustering* dapat menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda ?
2. Bagaimana merancang dan menerapkan *data mining* dengan algoritma K-Means *Clustering* dapat menentukan tingkat potongan harga terhadap harga jual sepeda motor Honda ?
3. Bagaimana menerapkan Algoritma K-Means *Clustering* dalam menentukan tingkat potongan harga terhadap harga jual sepeda motor Honda ke dalam bahasa pemrograman *PHP*?

1.3 Batasan Masalah

Pembahasan yang dilakukan dapat lebih terarah dan tidak menyimpang, maka diperlukan batasan terhadap masalah yang diteliti. Batasan masalah dalam penelitian ini meliputi:

1. *Data mining* yang diterapkan menggunakan algoritma K-Means *Clustering*.
2. Data yang digunakan adalah data penjualan sepeda motor Honda pada PT. Menara Agung.
3. Sistem dibuat terbatas pada analisis tingkat potongan harga jual sepeda motor Honda.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang diteliti, maka tujuan yang diharapkan dalam penelitian penerapan metode K-Means *clustering* dalam analisis tingkat potongan harga jual sepeda motor Honda adalah sebagai berikut:

1. Memahami dan menerapkan Algoritma K-Means *Clustering* untuk menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda.
2. Menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda berdasarkan data penjualan.
3. Menguji dan mengimplementasikan bahasa pemrograman PHP dengan Algoritma K-Means *Clustering* untuk menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda.

1.5 Manfaat Penelitian

Berdasarkan analisis masalah yang telah di teliti ini diharapkan dapat memberikan manfaat kedepannya, adapun manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Memberikan beberapa pengetahuan untuk menjadi pedoman kepada pihak *marketing* sepeda motor Honda dalam menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda.
2. Diharapkan penerapan ini dapat memberikan manfaat dalam menentukan tingkat potongan harga terhadap harga jual sepeda motor Honda.

1.6 Sistematika Penulisan

Pembahasan agar lebih terinci, maka dalam penulisan tesis ini penulis membagi atas beberapa bab yang saling berhubungan sesuai dengan ruang lingkup masalah. Secara umum gambaran ini masing-masing bab sebagai berikut:

Bab I : Pendahuluan

Pada bab pendahuluan ini berisikan Latar Belakang, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penelitian.

Bab II : Landasan Teori

Pada bab ini membahas dasar-dasar teori pada Algoritma K-Means *Clustering* dan bagaimana penerapan Algoritma K-Means *Clustering* dengan masalah yang dibahas pada penelitian ini serta penelitian sebelumnya yang mendukung mengenai Algoritma K-Means *Clustering*.

Bab III : Metodologi Penelitian

Bab ini menjelaskan tentang jenis penelitian yang dilakukan, pendekatan dan sumber data yang digunakan, lokasi penelitian, metode dan alat pengumpulan data serta teknik pengolahan dan analisis data.

Bab IV : Analisa dan Perancangan

Bab ini menjelaskan tentang kondisi pada objek penelitian, kondisi yang dijelaskan meliputi penggambaran yang ada pada objek penelitian dan penguraian proses yang terjadi pada objek penelitian hingga analisa dan desain.

Bab V : Implementasi dan Hasil

Bab ini menjelaskan bagaimana implementasi dan pengujian terhadap penerapan *data mining* yang telah dibuat menggunakan algoritma K-Means *Clustering*.

Bab VI : Kesimpulan dan Saran

Bab ini memuat kesimpulan dari penelitian tentang K-Means *Clustering* untuk menganalisis tingkat potongan harga terhadap harga jual sepeda motor Honda agar dapat berguna untuk penentuan potongan harga sepeda motor Honda, serta memberikan saran kepada peneliti selanjutnya.