

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) merupakan kegiatan yang terdiri dari mengumpulkan data, menggunakan data untuk menemukan pola atau keterkaitan, informasi serta pengetahuan yang bermanfaat dari sekumpulan data yang besar (Wibowo *et al*, 2021). Data Mining merupakan kombinasi dari beberapa disiplin ilmu komputer yang mendefinisikan sebagai proses untuk menemukan pola-pola baru dari sekumpulan data yang besar dengan menggunakan beberapa teknik seperti statistika, *artificial intelligent*, *machine learning* dan *database system* (Dewi *et al*, 2022). Terdapat lima teknik dalam Data Mining yaitu estimasi, prediksi, klasifikasi, klusterisasi, dan asosiasi. Kombinasi dua teknik dalam Data Mining diusulkan dalam penelitian ini yaitu klusterisasi dan klasifikasi. Kombinasi dua teknik ini dinamakan dengan *Hybrid Data Mining*. *Hybrid Data Mining* merupakan proses yang membantu pengambilan keputusan dengan melakukan kombinasi algoritma dan berbagai fitur seleksi (Mandala *et al*, 2022).

Clustering atau klusterisasi adalah proses mengelompokkan satu objek fisik atau abstrak yang memiliki kesamaan (Rejito *et al*, 2021). Proses klusterisasi menggunakan algoritma K-Means. Algoritma K-Means merupakan teknik Data Mining yang membagi entitas menjadi K grup berdasarkan atribut atau fitur, di mana K adalah bilangan positif (Alghanam *et al*, 2022). Klasifikasi adalah salah satu metode Data Mining yang digunakan untuk membagi data atau mengelompokkan data serta menemukan pola untuk memisahkan kelas data (Putri dan Mandala, 2022). Proses klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN). Algoritma K-Nearest Neighbor atau KNN merupakan metode untuk melakukan klasifikasi objek berdasarkan jarak data pembelajaran baru berdasarkan K tetangga yang paling dekat (Dewi *et al*, 2022).

Penelitian sebelumnya yang berkaitan dengan algoritma *hybrid* K-Means dan KNN dilakukan oleh (Mandala *et al*, 2022) yang membahas tentang klasifikasi pinjaman nasabah, penelitian ini menghasilkan bahwa menggunakan metode *hybrid* antara K-Means dan KNN memiliki pengaruh yang signifikan dalam menentukan mana yang berhak menerima pinjaman dan jumlah pinjaman yang diberikan kepada nasabah. Selanjutnya penelitian (Putri dan Mandala, 2022) melakukan klusterisasi produk untuk klasifikasi penjualan di minimarket, klusterisasi digunakan untuk mengelompokkan produk yang terjual, sementara klasifikasi digunakan untuk mengklasifikasikan penjualan.

Penelitian ini mengusulkan *Hybrid Data Mining* dengan teknik klusterisasi dan klasifikasi untuk memprediksi penjualan pada *marketplace* Shopee. Shopee dianggap sebagai pasar terbesar di Indonesia dan juga merupakan *platform* yang menyediakan bagi usaha kecil dan individu untuk membuat toko *online* mereka sendiri. Salah satu badan usaha yang menggunakan Shopee sebagai *platform* untuk menjual produk adalah Soraya Bedsheet. Soraya Bedsheet merupakan perusahaan yang bergerak dalam bidang industri yang memproduksi produk-produk penghasil dan penyedia kamar seperti *sprey*, *bed cover*, dan lain-lain.

Pada *marketplace* Shopee terdapat beberapa persoalan yang dialami pada toko Soraya Bedsheet. Persoalan yang pertama yaitu tidak dapat memperkirakan stok produk, kedua ketika produk sudah *soldout* Soraya Bedsheet tidak dapat mengupdate stok, dan ketiga nomor pembeli tidak bisa dilihat. Untuk menangani hal tersebut, maka dilakukan prediksi penjualan pada *marketplace* Shopee di toko Soraya Bedsheet. Prediksi yang dilakukan bertujuan untuk mendapatkan informasi dan pengetahuan dengan menggunakan *Hybrid Data Mining* dengan teknik klusterisasi dan klasifikasi.

Tahapan yang dilakukan diawali dengan melakukan klusterisasi produk yang dijual pada *marketplace* Shopee di Soraya Bedsheet. Proses klusterisasi menggunakan algoritma K-Means untuk membagi data latih menjadi beberapa kluster. Selanjutnya tahapan kedua melakukan klasifikasi penjualan menggunakan algoritma K-Nearest Neighbor (KNN). Hasil dari klasifikasi menggunakan algoritma KNN berupa data penjualan yang dapat digunakan untuk melakukan prediksi penjualan pada *marketplace* Shopee di Soraya Bedsheet. Berdasarkan latar belakang yang telah dijelaskan, maka penulis akan menerapkan algoritma *Hybrid Data Mining* K-Means dan K-Nearest Neighbor untuk memprediksi penjualan produk pada *marketplace* Shopee di toko Soraya Bedsheet.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah maka perumusan masalah dalam penelitian ini adalah:

1. Bagaimana menerapkan algoritma *hybrid* K-Means dan K-Nearest Neighbor untuk memprediksi penjualan?
2. Bagaimana mengukur tingkat akurasi algoritma *hybrid* K-Means dan K-Nearest Neighbor dalam memprediksi penjualan?

1.3 Batasan Masalah

Dalam penelitian ini terdapat batasan yang diharapkan dapat tetap fokus dan konsisten terhadap pada permasalahan. Adapun ruang lingkup masalah pada penelitian ini yaitu:

1. Penelitian ini menggunakan algoritma *Hybrid* Data Mining klusterisasi dan klasifikasi untuk memprediksi penjualan pada *marketplace* Shopee di Soraya Bedsheet
2. Proses klusterisasi dalam penelitian ini menggunakan pendekatan algoritma K-Means yang dilakukan untuk mengklusterisasikan produk penjualan pada *marketplace* Shopee di Soraya Bedsheet
3. Proses klasifikasi dalam penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) yang dilakukan setelah proses klusterisasi. Proses Klasifikasi ini dilakukan untuk memberikan prediksi penjualan pada *marketplace* Shopee di Soraya Bedsheet
4. Data yang digunakan dalam penelitian adalah data penjualan pada *marketplace* Shopee Soraya Bedsheet sejak September 2022 – Maret 2023
5. Aplikasi yang digunakan dalam penelitian ini adalah *RapidMiner*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Memahami proses penjualan produk pada *marketplace* Shopee di Soraya Bedsheet

2. Menganalisa dengan menggunakan algoritma *hybrid* K-Means dan K-Nearest Neighbor untuk melakukan prediksi penjualan pada *marketplace* Shopee di Soraya Bedsheet
3. Merancang model yang dihasilkan algoritma *hybrid* K-Means dan K-Nearest Neighbor berdasarkan parameter yang digunakan
4. Menerapkan aplikasi *RapidMiner* untuk proses pengolahan data dalam melakukan prediksi penjualan pada *marketplace* Shopee di Soraya Bedsheet
5. Menguji hasil algoritma *hybrid* K-Means dan K-Nearest Neighbor untuk mendapatkan akurasi terbaik.

1.5 Manfaat Penelitian

Manfaat penelitian dalam penelitian ini adalah:

1. Menghasilkan pengetahuan baru dari hasil penerapan algoritma *hybrid* K-Means dan K-Nearest Neighbor untuk memprediksi tingkat penjualan pada *marketplace* Shopee di Soraya Bedsheet
2. Memberikan informasi prediksi kepada Soraya Bedsheet tentang produk yang termasuk ke dalam penjualan rendah, dan penjualan tinggi.

1.6 Sistematika Penulisan

Sistematika penulisan laporan Tesis ini disusun agar pembuatan laporan dapat terstruktur dan lebih mudah dalam memahami penelitian yang dilakukan. Adapun sistematika penulisan tersebut yaitu:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini akan berisi penjelasan tentang teori-teori dan penerapan algoritma yang berhubungan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini akan membahas tentang metodologi penelitian dan metodologi dalam menganalisa perhitungan yang digunakan dalam penyusunan tesis.

BAB IV ANALISA DAN PERANCANGAN

Bab ini akan membahas tentang langkah-langkah dalam perhitungan menggunakan algoritma K-Means dan K-Nearest Neighbor.

BAB V IMPLEMENTASI DAN HASIL

Bab ini berisikan penjelasan tentang implementasi langkah-langkah penggunaan tools yang digunakan untuk melakukan perhitungan algoritma K-Means dan K-Nearest Neighbor menggunakan *RapidMiner*.

BAB VI KESIMPULAN DAN SARAN

Berisi kesimpulan dan saran yang berkaitan dengan analisa perhitungan berdasarkan yang telah diuraikan pada bab-bab sebelumnya.