

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Data mining disebut juga dengan *Knowledge Discovery in Database (KDD)*. Istilah *KDD* atau disebut penemuan pengetahuan data karena tujuan utama Data Mining adalah untuk mengolah data sehingga menghasilkan informasi baru yang bermanfaat (Hidayat dkk, 2021).

Proses penggalian informasi dari sebuah *data set* atau kumpulan data yang sangat besar disebut dengan Data Mining (Hasanah dkk, 2019). Data Mining digunakan dalam aplikasi yang lebih luas dan diterapkan di berbagai bidang kehidupan seperti, *sains*, penelitian, bisnis, industri dan bidang lainnya (Joseph, 2019).

Clustering adalah pengelompokan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam klaster atau kelompok (Dewi dkk, 2021). Teknik *Clustering* di antaranya adalah algoritma *K-Means Clustering*. *K-Means Clustering* menjadi salah satu algoritma yang paling banyak digunakan untuk mengelompokkan data (Mirantika dkk, 2021). Pengelompokan data dengan algoritma *K-Means* bisa diterapkan pada data penjualan dengan dua standar utama yaitu penjualan tinggi dan penjualan rendah (Ahmed dkk, 2019).

Penerapan Algoritma *K-Means* pernah digunakan untuk pengelompokan data penjualan toko retail perlengkapan *Outdoor* yang menghasilkan tiga *cluster*. Pengelompokan data tersebut digunakan untuk meningkatkan manajemen stok dan strategi penjualan (Indriyani & Irfiani, 2019).

Penelitian lainnya menerapkan algoritma *K-Means* pada *data set* ulasan obat yang tidak berlabel untuk mengelompokkan obat dengan penggunaan dan manfaat yang serupa. Hasil dari analisis kluster ini dapat digunakan untuk menyarankan pemilihan resep obat dalam kondisi tertentu (Bhargava, 2019).

Pada penelitian ini akan dilakukan pengelompokan data gejala penyakit dengan cara menghitung jarak terdekat antara data dengan titik tengah (*centroid*) suatu *cluster*. Sehingga pengelompokan tersebut akan menghasilkan beberapa *cluster* gejala penyakit ISPA yaitu ISPA sedang, ISPA ringan dan ISPA berat. Penerapan Algoritma *K-Means Clustering* diharapkan dapat membantu pihak Klinik Rahmatan Lil Alamin dalam meningkatkan pengelompokan gejala penyakit ISPA.

Berdasarkan uraian di atas dapat dilakukan penelitian yang berupa tesis dengan judul “ DATA MINING MENGGUNAKAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* PENYAKIT INFEKSI SALURAN PERNAFASAN AKUT “.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang ada sebelumnya, maka penulis dapat merumuskan beberapa detail permasalahan pada penelitian sebagai berikut:

1. Bagaimana menerapkan Data Mining menggunakan metode *K-Means Clustering* untuk mengklusterisasi penyakit ISPA dan mendapatkan hasil analisa dengan cepat dan akurat?

2. Bagaimana hasil akurasi pengklasteran penyakit ISPA dengan menggunakan metode *K-Means Clustering*?

1.3 Batasan Masalah

Agar penulisan tidak keluar dari permasalahan yang ada dan hasil penelitian dapat diperoleh dengan baik, maka penulis menetapkan batasan ruang lingkup pembahasan sebagai berikut:

1. *Clustering* dibentuk berdasarkan data historis penyakit ISPA dari bulan januari 2021 sampai agustus 2021.
2. Atribut pembentuk *clustering* yang digunakan adalah umur, jenis kelamin dan gejala-gejala yang menjadi penyebab ISPA.
3. Data yang digunakan hanya kasus dari pasien dewasa dan anak / bukan balita yang terkena ISPA.
4. Algoritma yang digunakan adalah Algoritma *K-Means Clustering*.

1.4 Tujuan Penelitian

Dalam penelitian ini dan pelaksanaannya ada beberapa tujuan yang akan dicapai, di antaranya:

1. Memahami metode penerapan Data Mining untuk mengklasterisasi penyakit ISPA.
2. Menganalisa *clustering* penyakit ISPA dengan metode *K-Means Clustering*.
3. Melakukan pengujian dengan menggunakan software Data Mining.
4. Mengimplementasikan hasil *clustering* penyakit ISPA dengan metode *K-Means Clustering*.
5. Menguji hasil penelitian dalam mengklasterisasi penyakit ISPA menggunakan metode *K-Means Clustering*.

BAB III**METODOLOGI PENELITIAN**

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

BAB IV**ANALISA DAN PERANCANGAN SISTEM**

Bagian ini menjelaskan tentang proses simulasi *K-Means Clustering* dan hasil yang didapatkan.

BAB V**IMPLEMENTASI DAN HASIL**

Bagian ini membahas tentang implementasi sistem yang dirancang dengan menggunakan metode *K-Means Clustering*, berupa langkah-langkah penggunaan sistem, tampilan hasilnya dan evaluasi sistem.

BAB VI**KESIMPULAN DAN SARAN**

Bagian ini berisikan kesimpulan dari penyusunan tesis serta saran-saran untuk pengembangan selanjutnya.