

BAB I

PENDAHULUAN

1.1 Latar Belakang

Knowledge Discovery in Database (KDD) adalah proses *nontrivial* dalam mengidentifikasi pola yang valid, baru, berpotensi, berguna, dan bermakna dari volume data yang besar. KDD memiliki istilah lain yang biasa disebut dengan Data Mining, di mana Data Mining digunakan dalam mengekstraksi pengetahuan yang berguna dari himpunan data (Yan dkk, 2020). Salah satu metode yang terdapat pada Data Mining adalah *Clustering*. *Clustering* digunakan untuk menemukan struktur *cluster* dalam kumpulan data yang di karakterisasi dengan kesamaan terbesar dalam *cluster* yang sama dan perbedaan terbesar antara *cluster* yang berbeda (Sinaga & Yang, 2020). Pada *clustering* terdapat beberapa metode salah satu di antaranya yaitu metode K-Medoids. Metode ini muncul untuk mengatasi kekurangan dari metode K-Means yaitu sensitif terhadap *outlier* data (Rahman dkk, 2020). Metode K-Medoids memiliki tujuan untuk meminimalkan jarak antara titik yang ditentukan sebagai pusat *cluster* dan titik data *cluster* (Hooshyar dkk, 2020). Metode K-Medoids bekerja dengan cara memisahkan dataset menjadi beberapa kelompok. (Muruganandham, 2018).

Beberapa penelitian mengenai metode K-Medoids telah dilakukan sebelumnya, di antaranya yaitu penelitian Sundari dkk (2019), penelitian ini mengelompokkan data imunisasi campak balita (Sundari dkk, 2019). Penelitian Irwansyah dkk (2020), penelitian ini mengangkat topik tentang pengelompokan pasien penyakit cardiovascular (Irwansyah dkk, 2020). Penelitian Lee dkk (2020), penelitian ini mengangkat topik tentang pengelompokan laktasi kurva sapi perah (Lee dkk, 2020). Penelitian Samudi dkk (2020), penelitian ini mengangkat topik tentang

pengelompokan aplikasi pembelajaran pada masa pandemi Covid-19 (Samudi dkk, 2020). Penelitian Rifa dkk (2020), topik yang di bahas pada penelitian ini adalah tentang pengelolaan resiko gempa di Indonesia (Rifa dkk, 2020).

Penyakit merupakan suatu kondisi di mana pikiran dan tubuh mengalami semacam gangguan dan ketidaknyamanan bagi yang mengalaminya (Andini & Arifin, 2020). Semakin hari, jumlah pasien di Puskesmas Kuok semakin meningkat dengan beragam jenis penyakit yang berbeda-beda. Peningkatan jumlah pasien dapat dilihat pada laporan yang disebut dengan data rekam medis (Andini & Arifin, 2020).

Data rekam medis adalah suatu informasi berupa riwayat penyakit yang diderita pasien setiap melakukan pengobatan di Rumah Sakit, Puskesmas dan Klinik (Ordila dkk, 2020). Data rekam medis ini harus selalu diperbaharui sesuai dengan jumlah pasien yang semakin hari semakin meningkat. Pembaharuan data rekam medis pasien mengakibatkan terjadinya penumpukan data. Penumpukan data rekam medis di Puskesmas Kuok hanya sebatas laporan yang berisikan informasi jumlah pasien dan penyakit yang diderita. Berdasarkan data tersebut, pihak Puskesmas memerlukan jenis informasi baru yaitu berupa informasi mengenai penyakit yang paling rentan dan banyak dialami oleh pasien. Informasi ini dapat membantu dalam mengambil tindakan untuk mengantisipasi bagaimana pencegahan dan pengobatan yang akan dilakukan. Informasi ini juga dapat menjadi bahan pertimbangan dalam perencanaan pembangunan fasilitas, penambahan obat dan penambahan *staff* (Andini & Arifin, 2020). Informasi-informasi yang terdapat dalam data rekam medis pasien Puskesmas Kuok akan diolah dengan metode K-Medoids *Clustering* dengan menggunakan 3 *cluster*: golongan sebab sakit, jumlah pasien laki-laki dan jumlah pasien perempuan sehingga menghasilkan informasi mengenai penyakit yang paling rentan dan banyak dialami oleh pasien.

Berdasarkan masalah di atas, maka perlu dilakukan penelitian dengan judul “Data Mining dalam Pengelompokan Penyakit Pasien dengan Metode K-Medoids (Studi Kasus di Puskesmas Kuok Kabupaten Kampar)”.

1.2 Perumusan Masalah

Sesuai dengan latar belakang di atas maka didapatkan rumusan masalah yaitu:

1. Bagaimana mengelompokkan penyakit pasien dengan menggunakan metode K-Medoids?
2. Bagaimana hasil dari proses setelah melakukan pengelompokan penyakit pasien menggunakan metode K-Medoids?

1.3 Batasan Masalah

Pada penelitian ini terdapat batasan masalah sebagai berikut.

1. Data yang diolah merupakan data rekam medis pasien Puskesmas Kuok pada tahun 2020.
2. Metode yang digunakan yaitu metode K-Medoids dengan menggunakan 3 atribut *cluster* berdasarkan golongan sebab sakit, jumlah pasien laki-laki, dan jumlah pasien perempuan.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Memahami metode K-Medoids dalam mengelompokkan data penyakit pasien di Puskesmas Kuok.
2. Menganalisa tahapan K-Medoids dalam melakukan pengelompokan data penyakit.
3. Merancang alur dan proses dalam mengelompokkan data penyakit pasien di Puskesmas Kuok.
4. Menerapkan aplikasi *Rapid Miner* untuk mengelompokkan data penyakit pasien di Puskesmas Kuok.
5. Menguji hasil *cluster* metode K-Medoids dengan menggunakan *Silhouette Coefficient*.

1.5 Manfaat Penelitian

Manfaat pada penelitian ini adalah sebagai berikut.

1. Peneliti memiliki wawasan dan pengetahuan baru tentang topik dan metode penelitian yang di teliti.
2. Hasil penelitian ini dapat menjadi acuan bagi pihak Puskesmas dan organisasi kesehatan lainnya, serta masyarakat untuk dapat selalu waspada dan melakukan tindakan pencegahan agar terhindar dari penyakit yang mendominasi.

1.6 Sistematika Penulisan

Bab I : Pendahuluan

Bab ini berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

Bab II : Landasan Teori

Pada bab ini dilakukan kajian literatur yang membahas mengenai topik penelitian didasari oleh penelitian yang dilakukan oleh peneliti-peneliti sebelumnya.

Bab III : Metodologi Penelitian

Bab ini membahas tentang tahapan penelitian yang dilakukan dan tahapan proses pada metode K-Medoids.

Bab IV : Analisa dan Perancangan

Bab ini membahas tentang analisa perhitungan K-Medoids dan perancangan penelitian berdasarkan metodologi penelitian.

Bab V : Implementasi dan Pengujian

Bab ini membahas tentang hasil *clustering* K-Medoids yang didapat dan mengimplementasikannya pada aplikasi *Rapid Miner*. Hasil *clustering* yang didapat kemudian diuji untuk menentukan hasil *cluster* yang didapat termasuk baik atau tidak dengan menggunakan *Silhouette Coeficient*.

Bab VI : Penutup

Bab ini membahas tentang kesimpulan dari hasil penelitian yang dilakukan dalam pengelompokan penyakit pasien dengan metode K-Medoids dan memberikan saran bagi peneliti berikutnya.