

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Sampai saat ini, kanker mulut rahim masih merupakan masalah kesehatan perempuan di Indonesia sehubungan dengan angka kejadian dan angka kematiannya yang tinggi[1]. Berdasarkan pernyataan Departemen Kesehatan Republik Indonesia pada tahun 2009 yang dikutip oleh Yustina Ananti, Kanker serviks merupakan penyakit yang disebabkan munculnya sel kanker ganas pada sel leher rahim. Keganasan sel tersebut disebabkan oleh virus yang bernama Human Papiloma Virus (HPV) tipe 16 dan 18 yang umum dijumpai di Indonesia[2]. Adapun faktor risiko terjadinya kanker serviks antara lain adalah aktivitas seksual pada usia muda, berhubungan seksual dengan multipartner, merokok, mempunyai anak banyak, sosial ekonomi rendah, pemakaian pil KB (dengan HPV negatif atau positif), penyakit menular seksual, dan gangguan imunitas. Menurut para ahli kanker, kanker serviks merupakan salah satu jenis kanker yang paling dapat dicegah dan paling dapat disembuhkan dari semua kasus kanker. Salah satu kegiatan deteksi dini kanker serviks yang paling umum di Indonesia adalah menggunakan metode *Pap Smear*. Sebagai contoh pada Rumah Sakit Umum Daerah (RSUD) Kediri dan Yayasan Kanker Indonesia (YKI) kabupaten Kediri merupakan penyedia layanan untuk tes *Pap Smear*. Tersedianya data histori rekam medis pasien tidak disertai dengan proses ekstraksi data menjadi sebuah pengetahuan atau informasi yang dapat berguna untuk keputusan klinis. Penggunaan teknik data mining sangat potensial untuk diterapkan dalam pelayanan kesehatan, khususnya untuk prediksi penyakit kanker serviks[3].

Beberapa penelitian terkait diantaranya adalah yang tertulis pada jurnal dengan judul “*Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means*”[5], Tool yang

digunakan adalah Weka. Hasil akurasi ditampilkan dengan bantuan *Confusion Matrix* dan kurva *ROC*, Nilai akurasi didapat sebesar 79,30% dengan *ROC* 79,17% pada *K-Means Euclidean Metric* sedangkan *K-Means Manhattan Metric* sebesar 67,83% dengan *ROC* 65,94%. Selanjutnya jurnal yang berjudul “Perbandingan Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah Menggunakan Algoritma *K-Means* Dan *K-Medoids*”. Didapatkan hasil pada klaster *K-Means* yaitu klaster puas sebanyak 276 siswa, klaster cukup puas sebanyak 216 siswa dan klaster kurang puas sebanyak 17 siswa. Kemudian pada algoritma *K-Medoids* yaitu klaster puas sebanyak 324 siswa, klaster cukup puas sebanyak 11 siswa dan klaster kurang puas sebanyak 174 siswa. Kajian-kajian tersebut hanya sebatas menghitung jumlah klaster yang terbentuk dari masing-masing algoritma, namun hasil perhitungan tersebut belum dilakukan validasi klaster dimana tujuan dari validasi klaster adalah untuk mengetahui seberapa baik proses kinerja pengklasteran.

Berdasarkan penelitian-penelitian sebelumnya, penulis akan mengambil judul “Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Klasterisasi Penyakit Kanker Serviks”, dimana alasan penulis menggunakan algoritma *K-Means* dan *K-Medoids* karena data yang akan digunakan tidak memiliki kelas, kemudian ditentukan nilai k sebagai input algoritma sebelum dilakukan pemodelan dan proses komputasi lebih cepat, selain itu juga untuk melihat perbandingan hasil klasterisasi apakah terdapat perbedaan hasil meskipun kedua algoritma tersebut memiliki kemiripan yaitu sama-sama memiliki sifat partisional dengan memecah data menjadi kelompok-kelompok tertentu. Hasil klaster yang terbentuk akan lebih optimal karena divalidasi dengan menggunakan metode *Davies Bouldin Index* dengan bantuan tool *RapidMiner Studio* dan *Excel*.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah yang akan dikaji dalam penelitian ini yaitu:

1. Bagaimana hasil pengklasteran penyakit kanker serviks?
2. Bagaimana perbandingan pengklasteran penyakit kanker serviks dengan menggunakan algoritma *K-Means* dan *K-Medoids* yang divalidasi dengan metode Davies Bouldin Index?

1.3 Batasan Masalah

Pada penulisan ini, penulis membatasi penelitian ini dengan menggunakan dua algoritma yaitu algoritma *K-Means* dan *K-Medoids*, dan menentukan nilai $k = 2$ untuk membentuk kluster positif dan negatif yang hasilnya akan divalidasi dengan menggunakan metode *Davies Bouldin Index*. Data yang digunakan adalah data *Cervical cancer (Risk Factors) Data Set* yang didapatkan dari *UCI Machine Learning Repository* dengan jumlah data 858 record dan 36 atribut, kemudian proses perhitungan dilakukan secara manual menggunakan excel dan diimplementasikan kedalam tool *RapidMiner Studio* Versi 9.

1.4 Tujuan Penelitian

Sesuai dengan rumusan masalah yang telah disusun, maka tujuan yang akan dicapai adalah untuk mengetahui hasil pengklasteran penyakit kanker serviks, dan mengetahui hasil perbandingan pengklasteran penyakit kanker serviks dengan menggunakan algoritma *K-Means* dan *K-Medoids*.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah mendapatkan gambaran bagaimana proses klasterisasi hingga tahap validasi dari data penyakit kanker serviks dengan menggunakan algoritma *K-Means* dan *K-Medoids* dengan menguji beberapa variable yang diperoleh dari dataset

publik dibantu dengan tool RapidMiner dan bagi institusi khususnya program studi Magister Teknik Informatika dapat digunakan sebagai referensi ilmiah dalam penelitian penerapan data mining serta memberikan masukan dalam ilmu pengetahuan untuk penelitian-penelitian selanjutnya terutama dalam pengembangan ilmu pengetahuan menggunakan algoritma *data mining*.

1.6 Sistematika Penulisan

Laporan penelitian tesis ini terdiri dari lima bab dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Dalam pendahuluan tercantum antara lain latar belakang, ruang lingkup, rumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat tentang teori-teori yang mendukung penelitian yang akan dilakukan oleh penulis/peneliti. Penelitian yang menggunakan analisis statistik, bab ini memuat kerangka pikir dan hipotesis (bila diperlukan).

BAB III METODOLOGI PENELITIAN

Dalam bab ini berisi objek penelitian, alat dan bahan, metode pengumpulan data, prosedur penelitian, pengukuran variabel dan metode analisis (metode-metode pendekatan penyelesaian permasalahan yang dipakai dan metode analisis data).

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini disajikan hasil, implementasi, analisis dan pembahasan penelitian.

Hasil dan implementasi dapat berupa gambar alat/program dan aplikasinya. Untuk penelitian lapangan hasil dapat berupa data (kualitatif maupun kuantitatif). Analisis dan pembahasan berupa hasil pengolahan data.

BAB V KESIMPULAN DAN SARAN

Dalam bab ini disajikan simpulan dan saran dari hasil pembahasan