

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan salah satu penyebab kematian baik di negara maju maupun di negara yang sedang berkembang. Beban kanker meningkat di negara yang sedang berkembang akibat meningkatnya gaya hidup yang berkaitan dengan kanker, misalnya merokok, aktivitas fisik yang kurang, dan konsumsi diet (Yuliana, 2018). Berdasarkan data dari (GLOBOCAN, 2012) dirilis oleh Badan Kesehatan Dunia (WHO) menyebutkan bahwa kanker paru-paru menempati peringkat pertama dalam jumlah kasus baru sebesar 2,094 juta kasus diseluruh dunia. Jumlah kasus baru tertinggi berikutnya adalah kanker payudara, kanker kolorektal, kanker prostat, dan kanker lambung.

Kanker payudara merupakan kanker yang berasal dari sel-sel yang terdapat di payudara, bisa dari sel-sel saluran air susu atau sel-sel kelenjar penghasil air susu atau jaringan lain. Kanker ini terjadi hampir seluruhnya padawanita, tetapi dapat juga terjadi pada pria. Di Amerika Serikat, diperkirakan 231.840 wanita didiagnosis kanker payudara pada tahun 2015. Pada tahun 2012, kanker payudara menempati peringkat kedua setelah kanker paru sebagai penyebab kematian utama karena kanker di Amerika Serikat. Di Indonesia, diperkirakan 51.136 wanita didiagnosis kanker payudara pada tahun 2015. Penyakit ini juga menjadi penyebab kematian utama karena kanker di Indonesia pada tahun 2012. Oleh sebab itu sangat penting untuk dilakukan deteksi dini

terhadap penyakit kanker payudara agar dapat segera dilakukan pengobatan (Yuliana, 2018).

Pada praktiknya, di dunia kedokteran seringkali data rekam medis seperti penyakit kanker payudara disimpan untuk berbagai tujuan. Namun kenyataannya, proses penyimpanan dan pengolahan data rekam medis di beberapa rumah sakit masih belum memanfaatkan media komputer sehingga data-data rekam medis yang tersimpan sering hilang ataupun rusak serta memakan waktu lama ketika akan melakukan proses pencarian. Selain itu, data rekam medis yang tercatat dan terkumpul biasanya diolah dan dimanfaatkan menjadi sebuah pengetahuan dimana pengetahuan tersebut dapat digunakan untuk melakukan prediksi. Namun untuk mengelola kumpulan data rekam medis menjadi pengetahuan tentunya tidak dapat dilakukan secara konvensional dikarenakan proses pengelolaan secara konvensional memiliki tingkat keakuratan yang kurang baik sehingga hasil kesimpulan prediksi yang diambil pun tentunya akan berbeda dan tidak begitu akurat.

Oleh karena permasalahan tersebut, maka perlu dibangun sebuah sistem informasi dengan penerapan *Data Mining* dalam dunia kesehatan khususnya pengelolaan data rekam medis. *Data Mining* merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. *Data Mining* itu sendiri disalurkan dalam bentuk algoritma-algoritma kompleks yang memiliki tingkat keakuratan yang cukup baik. Terdapat beberapa penelitian yang mencoba membangun sistem

pengelolaan data rekam medis yaitu penelitian analisa data rekam medis menggunakan teknik *Data Mining Association* dengan algoritma *Clustering* (Kurniawan E. , Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering, 2017). Selanjutnya penelitian untuk membangun sebuah sistem klusterisasi data rekam medis dengan menggunakan algoritma *K-Means Clustering* (Ali, Klasifikasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo, 2019). Kedua penelitian tersebut sudah cukup baik, namun masih memiliki kekurangan pada tingkat akurasi. Pada penelitian ini, algoritma *Data Mining* yang digunakan adalah *Decision Tree C4.5*. Algoritma C4.5 sering digunakan untuk memprediksi kejadian yang akan datang, seperti penelitian yang dilakukan oleh (Irfansyah, 2016) yang menggunakan algoritma C4.5 dalam penentuan promosi jabatan dan penelitian yang dilakukan oleh (Yulianti, 2019) yang menggunakan algoritma C4.5 dalam prediksi keberhasilan *software*. Kedua penelitian tersebut memberikan hasil prediksi yang baik. Selain keakuratan hasil prediksinya, kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan dan efisien dalam menangani atribut bertipe diskret dan numerik (Kamagi & Hansun, 2014). Selain itu, terdapat juga penerapan algoritma C4.5 dalam melakukan prediksi penyakit yaitu penelitian yang dilakukan oleh (Handayani, Nurlalah, Raharjo, & Ramdani, 2019) membahas mengenai prediksi penyakit Liver menggunakan algoritma *Decision Tree C4.5* dan *Neural Network*. Hasil penelitian

menunjukkan bahwa algoritma *Decision Tree C4.5* lebih akurat dengan akurasi sebesar 75,56% dan algoritma *Neural Network* dengan akurasi sebesar 74,17%.

Berdasarkan uraian tersebut, maka penulis tertarik untuk melakukan penelitian dengan mengambil judul “**IMPLEMENTASI ALGORITMA DATA MINING DECISION TREE C4.5 DALAM PREDIKSI PENYAKIT KANKER PAYUDARA**”.

1.2 Rumusan Masalah

Berikut ini akan dipaparkan perumusan masalah dalam penelitian ini yaitu:

1. Bagaimana membangun sebuah sistem informasi pencatatan data rekam medis penyakit kanker payudara agar dapat memudahkan dalam proses pencarian data serta menjaga keamanan data agar tidak hilang ataupun rusak.
2. Bagaimana mengimplementasikan algoritma *Decision Tree C4.5* dalam melakukan prediksi data rekam medis penyakit kanker payudara secara akurat sehingga dapat digunakan acuan dalam deteksi dini penyakit kanker payudara.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun sebuah sistem informasi pencatatan data rekam medis penyakit kanker payudara.

2. Implementasi algoritma *Decision Tree* C4.5 untuk prediksi penyakit kanker payudara.
3. Analisis keakuratan hasil prediksi dengan algoritma *Decision Tree* C4.5.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan dataset penyakit kanker payudara yang berasal dari *UCI Machine Learning Repository*.
2. *Input* data yang dimasukkan adalah nama pasien, nama dokter, dan variabel penyakit kanker payudara yang terdiri dari umur, BMI, *Glucose*, *Insulin*, HOMA, *Leptin*, *Adiponectin*, *Resistin*, dan MCP1.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini yaitu:

1. Sistem yang dibangun dapat dijadikan sebagai media penyimpanan dan pengolah data rekam medis penyakit kanker payudara secara cepat, mudah, dan terkomputerisasi.
2. Hasil prediksi yang didapatkan dapat dijadikan acuan dalam melakukan deteksi dini penderita penyakit kanker payudara.
3. Informasi yang dihasilkan dapat digunakan oleh pihak dinas kesehatan setempat sebagai pertimbangan dalam mengambil keputusan lebih lanjut untuk menangani penyakit kanker payudara.

1.6 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam merancang sistem pada penelitian ini terdiri dari:

1. Pengumpulan Data

a. Studi Literatur

Penulis mengumpulkan informasi melalui buku, maupun bahan referensi lainnya yang berhubungan dengan sistem informasi serta bahan referensi yang berhubungan dengan kegiatan pencatatan data rekam medis pasien.

b. Studi Lapangan

Penulis secara langsung mengamati proses operasional yang dilakukan di klinik-klinik di kota Medan. Dalam mengumpulkan data dilakukan tanya jawab dengan beberapa manajer klinik.

2. Analisis Sistem

Pada tahap ini dilakukan analisis sistem dengan menguraikan cara kerja algoritma *Decision Tree* C4.5 dalam melakukan prediksi penyakit kanker payudara dan melakukan pemodelan analisis sistem dengan menggunakan *Use Case Diagram*.

3. Desain Sistem

Pada tahap ini dilakukan perancangan sistem dengan menggunakan *software* Balsamiq Mockup 3.

4. Pembangunan Sistem

Adapun beberapa bahas pemrograman yang digunakan dalam membangun sistem yaitu:

- a. *Framework* Laravel untuk *Back End Programming*.
- b. *Framework* Bootstrap untuk *Front End Programming*.

1.7 Sistematika Penulisan

Untuk mempermudah memahami permasalahan yang dibahas, maka penelitian ini disusun dengan sistematika sebagai berikut:

1. Bab I berisi tentang latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.
2. Bab II berisi tentang tinjauan pustaka yang menjabarkan teori-teori yang menjadi pendukung dalam penelitian ini serta peneliti terdahulu dan menjelaskan mengenai profil perusahaan.
3. Bab III berisi tentang analisis dan perancangan dimana membahas mengenai analisis proses cara kerja algoritma, analisis sistem berjalan, pemodelan kebutuhan sistem dan perancangan sistem.
4. Bab IV berisi tentang hasil dan pembahasan dimana membahas mengenai hasil penelitian yang telah dilakukan serta pembahasan terkait kelebihan dan kekurangan sistem yang dibangun.

5. Bab V berisi tentang kesimpulan dan saran dimana membahas mengenai kesimpulan yang didapatkan dari penelitian yang telah dilakukan serta saran yang mungkin dapat dilanjutkan bagi peneliti berikutnya.

