

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Kanker adalah salah satu penyebab utama kematian di seluruh dunia. Kanker paru merupakan penyakit keganasan yang timbul (terutama) pada paru-paru itu sendiri [1]. Kanker paru-paru menempati urutan ketiga atau keempat diantara keganasan di rumah sakit di Indonesia. *Korelasi* antara merokok dengan risiko kanker paru dapat dilihat dari data RS Persahabatan tahun 2004 hingga 2006, dimana 83,4% pria dan 43,4% wanita merokok.

Kanker paru-paru paling sering terjadi akibat kebiasaan merokok. Meski demikian, orang yang tidak merokok juga dapat menderita kanker paru-paru, terutama jika sering terpapar polusi udara, tinggal di lingkungan yang tercemar zat berbahaya, atau memiliki keluarga yang menderita kanker paru-paru. Kanker paru-paru sering kali tidak menimbulkan gejala pada tahap awal. Gejala baru muncul ketika kanker sudah cukup besar atau telah menyebar ke jaringan dan organ sekitar. Beberapa gejala yang dapat dirasakan penderita kanker paru-paru adalah batuk kronis, sesak napas, batuk darah, dan nyeri dada.

Dalam penelitian ini, penulis menggunakan *dataset* kanker paru-paru dari *Kaggle.com*. Kumpulan data kanker paru-paru memiliki masalah dalam mengidentifikasi apakah orang Indonesia menderita kanker paru-paru. Catatan ini memiliki nomor kelas 2, “YES” dan “NO”. Penelitian lain tentang klasifikasi

menggunakan citra *dermoskopi* dengan metode *Naive Bayes* dan *Decision Tree* menunjukkan bahwa tingkat akurasi *Naive Bayes* memiliki akurasi yang lebih tinggi dibandingkan dengan metode *Decision Tree*. Banyak penelitian sebelumnya telah dilakukan oleh penelitian lain [2] menggunakan Algoritma *K-Nearest Neighbor*.

Banyak penelitian sebelumnya yang telah dilakukan oleh peneliti lain. Pada penelitian [3] Bahasa pemrograman *Python* digunakan dalam alat pemrograman *Google Colab* untuk meningkatkan kinerja *Gaussian Naive Bayes*, *Multinomial Naive Bayes*, dan *Bernoulli Naive Bayes*. Hasil penelitian ini menunjukkan bahwa rata-rata performansi metode *Bernoulli Naive Bayes* sangat baik, dengan rata-rata presisi 93,25%, rata-rata presesi 9,23%, dan rata-rata recall 9,69%. Ini menunjukkan akurasi diagnostik yang sangat tinggi dari *Bernoulli Naive Bayes*.

Berdasarkan permasalahan yang telah diuraikan maka judul yang diberikan oleh penulis pada karya ini adalah “**Penerapan Data Mining Dengan Algoritma *Naive Bayes* dan Algoritma *C4.5* Untuk Klasifikasi Kanker Paru-Paru**”.

1.2 RUMUSAN MASALAH

Adapun rumusan masalah yang muncul pada penelitian ini adalah :

1. Bagaimana penerapan *Algoritma Naive Bayes* dan *Algoritma C4.5* dalam klasifikasi kanker paru-paru?

2. Seberapa akurat *Algoritma Naive Bayes* dan *Algoritma C4.5* dalam mengklasifikasikan kanker paru-paru?

1.3 BATASAN MASALAH

Batasan masalah yang akan dibahas pada penelitian ini sebagai berikut :

1. Dataset yang digunakan dalam penelitian ini adalah kanker paru-paru (mengumpulkan hingga 309 data kanker paru-paru). Kumpulan data berasal dari situs web *kaggle.com* dan dapat diunduh dari tautan.:
<https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer>
2. Variabel berikut digunakan dalam penelitian ini adalah Jenis Kelamin, Umur, Merokok, Kulit Kuning, Kecemasan, Tekanan Lingkungan, Radang Usus, Kelelahan, Alergi, Asma, Mengonsumsi Alkohol, Batuk, Sesak Napas, Kesulitan Menelan, Nyeri Dada, Kanker Paru-Paru.
3. Alat bantu yang digunakan dalam penelitian adalah *tools WEKA*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan oleh penulis adalah

1. Penerapan *Algoritma Naive Bayes* dan *Algoritma C4.5* untuk klasifikasi kanker paru-paru
2. Untuk mengetahui seberapa akurat *Algoritma Naive Bayes* dan *Algoritma C4.5* untuk klasifikasi kanker paru-paru.

1.4.2 Manfaat Penelitian

Berikut adalah beberapa manfaat yang akan Anda dapatkan dari melakukan penelitian ini, yaitu :

1. Dapat memudahkan deteksi kanker paru-paru klasifikasi
2. Dengan menggunakan alat *WEKA* sebagai alat klasifikasi, penulis berharap dapat mencapai akurasi tinggi dalam klasifikasi kanker paru-paru.
3. Menambah wawasan bagi penulis saat mengelompokkan kumpulan data kanker paru-paru.

1.5 SISTEMATIKA PENULISAN

Sistematika dimaksudkan untuk memberikan gambaran yang jelas tentang penulisan akademik agar tidak terjadi pembahasan di luar topik dan judul kajian. Secara garis besar, sebuah karya tulis ilmiah terdiri dari lima bab, yang dapat dilihat dari sistematika karya tulis sebagai:

BAB I : PENDAHULUAN

Bab ini merupakan bab pendahuluan yang meliputi latar belakang masalah, rumusan masalah, definisi masalah, tujuan dan manfaat penelitian, serta struktur penulisan.

BAB II : LANDASAN TEORI

Bab ini memberikan landasan teoretis yang menjadi dasar untuk membahas laporan tertentu, dan memberikan definisi yang

mendasari penelitian yang dilakukan melalui tinjauan literatur sebagai dasar untuk melakukan analisis dan desain.

BAB III : METODOLOGI PENELITIAN

Bab ini merinci kerangka penelitian, metode atau pendekatan yang digunakan, dan alat yang digunakan untuk menjawab pertanyaan penelitian untuk mencapai tujuannya.

BAB IV : ANALISA

Pada bab ini, kami melakukan komputasi dan analisis menggunakan metode *algoritma naive bayes* pada *dataset* kanker paru. Hasil analisis disajikan dalam bentuk visualisasi dari *tools* yang digunakan yaitu *WEKA* dan *RapidMiner*.

BAB V : KESIMPULAN

Bab ini berisi kesimpulan dan saran yang diambil dari hasil analisis dan saran yang mencakup semua hasil yang dilakukan.