

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Di Indonesia dan di seluruh dunia kanker paru-paru merupakan salah satu penyakit dengan tingkat kematian tertinggi menduduki peringkat keenam dalam daftar penyebab kematian terbanyak. Kanker paru dalam arti luas merupakan semua penyakit keganasan di paru, mencakup keganasan yang berasal dari paru sendiri (primer) maupun keganasan dari luar paru (*metastasis*) [1]. Di Indonesia, kanker paru juga merupakan salah satu kanker dengan angka kejadian terbanyak, dan merupakan kanker dengan angka kematian paling tinggi [2]. Kanker paru-paru menempati peringkat enam dari sepuluh penyakit penyebab kematian terbanyak di Indonesia [3].

Kanker merupakan pertumbuhan dan penyebaran sel-sel abnormal yang memiliki karakteristik yang khas, Kanker yang sudah menyebar dan tidak dapat terkontrol lagi, biasanya akan menyebabkan kematian. Kanker paru paru lebih sering menyebabkan pria meninggal dibanding kanker lain, dimana yang sering menjadi penyebab kanker paru-paru adalah merokok [4]. Hal ini memerlukan upaya yang berkelanjutan dan cepat untuk menangani masalah ini. Penanganan yang diperlukan salah satunya adalah dengan melakukan pendeteksian dini pada pasien kanker paru-paru. Untuk mendapatkan hasil yang akurat dibutuhkan perhitungan data yang sesuai dengan menggunakan teknologi yang mendukung, untuk dapat melakukan pendeteksian resiko terkena kanker paru agar dapat segera diatasi sesuai dengan tingkat resikonya, dibutuhkan sebuah klasifikasi pada data

kanker paru. Salah satu metode untuk melakukan prediksi tersebut dengan menggunakan *data mining*.

Mengacu pada permasalahan tersebut maka diperlukan sebuah metode yang dapat digunakan untuk mengurangi tingkat kesalahan yaitu dengan menerapkan teknik *data mining*. *Data Mining* merupakan suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut [5]. *Data mining* merupakan suatu proses menggali nilai lebih yang ada pada suatu basis data dengan melihat pola-pola dari data sehingga menghasilkan informasi yang bermanfaat yang tidak dapat ditemukan secara manual [6]. Penerapan *data mining* dinilai tepat karena memudahkan dalam melakukan penggalian informasi terhadap data kanker paru-paru. Sehingga data yang telah diperoleh dapat diolah untuk mengetahui informasi-informasi yang tersembunyi dari data-data tersebut.

Berdasarkan permasalahan dan beberapa klasifikasi di atas penulis memilih menggunakan algoritma *Naïve Bayes* dikarenakan algoritma ini sering digunakan dalam klasifikasi data dan dinilai berpotensi baik dalam mengklasifikasi data. Metode *Naïve Bayes* merupakan salah satu pengklasifikasi statistik, dimana klasifikasi ini dapat memprediksi probabilitas keanggotaan kelas suatu data yang akan masuk ke dalam kelas tertentu, sesuai dengan perhitungan probabilitas [7]. *Naïve bayes* memiliki kinerja yang baik dan menghasilkan probabilitas rata-rata 71 persen dengan waktu proses yang lebih cepat dibanding algoritma pembelajaran mesin lain serta memiliki reputasi pada keakuratan prediksi [8]. Prediksi merupakan salah satu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan

informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil [9]. Oleh karena itu penulis tertarik dalam menggunakan algoritma *naïve bayes* ketika memprediksi kanker paru-paru karena dapat memberikan akurasi yang baik.

Sehubungan dengan permasalahan di atas, penulis tertarik untuk melakukan penelitian yang berjudul **“Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penyakit Kanker Paru-Paru”**.

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang yang telah diuraikan diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode *Naïve Bayes* dalam mengklasifikasi penyakit kanker paru-paru ?
2. Bagaimana mengukur tingkat akurasi metode *Naïve Bayes* apabila digunakan untuk mengklasifikasi penyakit kanker paru-paru ?

1.3 BATASAN MASALAH

Berdasarkan dari latar belakang masalah, maka adapun batasan dalam penelitian ini, yaitu:

1. Variable yang di gunakan dalam penelitian ini adalah *Gender, Age, Smoking, Yellow_Finger, Anxiety, Peer_Pressure, Chronic Disease, Fatigue, Allergy, Wheezing, Alcohol Consuming, Coughing, Shortness of Breath, Swallowing Difficulty, Chest Pain dan Lung Cancer* dengan menggunakan *RapidMiner*.

2. Penelitian ini hanya akan memfokuskan pada penerapan algoritma klasifikasi *Naive Bayes* sebagai model utama untuk deteksi kanker paru-paru.
3. Tools yang digunakan untuk penelitian ini adalah *RapidMiner*.

1.4 TUJUAN PENELITIAN

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu:

1. Menerapkan metode *Naive Bayes* dalam pengklasifikasian penyakit kanker paru-paru.
2. Mengukur akurasi yang tepat untuk melakukan klasifikasi pada data kanker paru-paru dengan menggunakan *Tools RapidMiner*.

1.5 MANFAAT PENELITIAN

Adapun manfaat dalam melakukan penelitian ini, yaitu:

1. Dapat menerapkan metode *Naive Bayes* dalam mengklasifikasi penyakit kanker paru-paru.
2. Mendapatkan tingkat akurasi tentang penyakit Kanker Paru-Paru menggunakan *Naive Bayes* dengan *Tools RapidMiner*.

1.6 SISTEMATIKA PENULISAN

Adapun sistematika penulisan yang diajukan dalam Skripsi ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini menerangkan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini terdiri dari penjelasan mengenai studi sebelumnya yang mendukung tugas akhir dan konsep atau teori-teori yang memiliki keterkaitan terhadap topik penelitian, yaitu kanker paru-paru, *Data Mining*, metode Naïve Bayes, dan memprediksi kanker paru-paru dengan mengolah data kanker paru-paru menggunakan aplikasi *RapidMiner*.

BAB III : METODOLOGI PENELITIAN

Pada bab ini berisikan tentang penjelasan mengenai kerangka kerja penelitian, dan metode-metode yang akan digunakan untuk melakukan penelitian pada tugas akhir ini, dan juga alat bahan yang akan digunakan.

BAB IV : ANALISIS

Pada bab ini akan menjelaskan dan memperlihatkan proses dari aktivitas perhitungan dan juga analisis yang akan

dilakukan pada tahap akhir untuk menampilkan hasil yang telah diperoleh.

BAB V : KESIMPULAN

Pada bab terakhir berisi kesimpulan yang berkaitan dengan tujuan penelitian dari hasil yang telah diperoleh, serta berisi saran yang berkaitan dengan hasil penelitian dan juga untuk peneliti selanjutnya.