

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Kanker paru-paru menjadi kanker paling umum terjadi di Amerika Serikat di tahun 2020 dengan 2,20 juta kasus baru dengan menjadi nomor satu di Amerika Serikat dengan kematian terbanyak di antara semua kanker (*Cancer*, 2021.).

Kanker timbul karena berubahnya sel normal berubah sel tumor dalam proses yang panjang yang kebanyakan berkembang dari lesi pra-kanker menjadi tumor ganas. Berikut hasil interaksi antara faktor genetik seseorang dan tiga kategori agen eksternal :

1. karsinogen fisik, dengan contoh *ultraviolet* dan radiasi pengion;
2. karsinogen kimia, dengan contoh asbes, asap tembakau, *aflatoksin* (pencemar makanan), dan arsenik (pencemar air minum)
3. karsinogen biologis, dengan contoh infeksi dari virus, bakteri, atau parasit tertentu.

Insiden kanker meningkat terus-menerus seiring semakin menuanya usia, ini dikarenakan meningkatnya risiko kanker spesifik yang akumulasi risiko keseluruhan bersamanya dengan kecondongan mekanisme pembetulan sel yang tidak efektif seiring bertambahnya usia seseorang (*Cancer*, 2021.).

Sebagian besar pasien didiagnosis dengan kanker paru-paru secara tradisional diklasifikasikan sebagai 2 kategori utama: (1) non-sel kecil kanker paru-paru dan (2) kanker paru-paru sel kecil, yang lebih agresif dan memiliki tingkat kelangsungan hidup yang lebih buruk. Sekitar 85% dari kanker paru-paru . Risiko terkena kanker paru-paru sebagian besar didorong oleh usia dan status merokok, dengan merokok diperkirakan mencapai hampir 90% dari semua kanker paru-paru. Faktor risiko lain untuk paru-paru kanker termasuk paparan lingkungan,

terapi radisi, penyakit paru-paru, ras / etnis, dan riwayat keluarga (Jonas et al., 2021). Rokok merupakan penyebab utama dari penyakit kanker ini, menyumbang setidaknya 80-90% dari penyebab kanker paru-paru.

Untuk menganalisis data yang tersimpan di dalam *Database*, dengan menggunakan teknologi *Data Mining*. Potensi di bidang *Data Mining* kesehatan diakui secara luas. Banyak penelitian telah menggunakan teknik *Data Mining*, juga klasifikasi dan prediksi yang cocok untuk rekam medis elektronik. Penulis menggunakan Metode algoritma *Naïve Bayes* untuk mendiagnosis kanker paru-paru.

Proses yang dilakukan pada penelitian ini dengan menggunakan analisis prediksi *Naïve Bayes*, yaitu sebuah teknik mengklasifikasikan teorema bayes yang mampu mengolah data kuantitatif dan data diskrit untuk perhitungan estimasi peluang (Liu et al., 2020). Untuk melakukan prediksi, diperlukan data yang akurat dan valid, dan membutuhkan koresponden yang bisa dipercaya. Data yang saya akan uji disini memiliki beberapa klasifikasi seperti menguningnya jari, mempunyai penyakit kronis, umur, jenis kelamin, perokok atau bukan, dan lain sebagainya. Berdasarkan keluhan di jelaskan dalam bab 4

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang tersebut, terdapat beberapa masalah yang dapat dirumuskan dalam beberapa pertanyaan berikut:

1. Bagaimana memprediksi Penyakit kanker paru-paru menggunakan Algoritma *Naïve Bayes*.
2. Bagaimana mengevaluasi prediksi dengan menggunakan algoritma *Naïve Bayes*?

### **1.3. Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk memfasilitasi individu dalam memprediksi apakah ciri-ciri yang diderita seseorang dapat menjadi pertanda mengidap penyakit kanker paru-paru dan mampu menemukan akurasi dari prediksi penyakit kanker paru-paru.

### **1.4. Manfaat Penelitian**

1. Mengetahui penyebab utama dari kanker paru-paru.
2. Mengetahui ciri-ciri spesifik dari kanker paru-paru?
3. Membantu para dokter untuk mendiagnosis pasien yang memiliki gejala seperti kanker tersebut

### **1.5. Ruang Lingkup**

Pada penelitian ini, yang akan dianalisis adalah dataset yang sudah disediakan lalu akan penulis prediksi menggunakan analisis prediksi Naïve Bayes dan akan memprediksi manakah yang merupakan penderita dan mana yang bukan. hal-hal yang tidak berhubungan dengan lingkup analisis, tidak akan dijelaskan pada penelitian ini.

### **1.6. Sistematika Penulisan**

Berikut merupakan sistematika penulisan berupa gambaran secara terperinci mengenai bab-bab pada penulisan yang akan menjelaskan kesinambungan tiap bab satu sama lain yang akan dijelaskan :

#### **BAB I PENDAHULUAN**

Bab ini berisi Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Ruang Lingkup, Ruang Lingkupz', dan Sistematika Penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang teori-teori dasar yang digunakan untuk penelitian ini seperti Naïve Bayes, Python, kanker, kanker paru-paru.

### **BAB III METODE PENELITIAN**

Pada Bab ini berisi tentang kerangka berfikir, dan perkiraan jadwal penelitian, dan segala metode yang terdapat dalam penelitian ini.

### **DAFTAR PUSTAKA**