

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kanker merupakan suatu pertumbuhan yang terjadi secara tidak normal dari sel-sel jaringan tubuh yang berubah menjadi ganas. Sel-sel tersebut dapat tumbuh lebih lanjut serta menyebar ke bagian tubuh lainnya yang dapat menyebabkan kematian. Sel tubuh yang mengalami perubahan mutasi dan mulai tumbuh dapat membelah lebih cepat dan tidak terkendali seperti sel normal. Sel kanker tidak mati setelah usianya melainkan dapat tumbuh terus dan bersifat *invasive* (menyerang) yang menyebabkan sel normal yang tumbuh terdesak atau malah mati (Kementerian Kesehatan Republik Indonesia, 2019).

Penyakit kanker adalah salah satu penyakit yang menyebabkan kematian utama di seluruh dunia. Pada tahun 2018, sekitar 9,6 juta kematian dikarenakan oleh kanker. Dimana kanker paru-paru dan payudara merupakan penyebab kematian terbesar tiap tahunnya (GLOBOCAN, 2018).

Berdasarkan Data *GLOBOCAN, International Agency for Research on Cancer (IARC)*, diberitahukan pada tahun 2018 terdapat kasus baru yang menyebabkan 18,1 juta korban kanker dan korban kematian kanker sekitar 9,6 juta di seluruh dunia. Kanker payudara dan kanker paru-paru merupakan jenis kanker yang sering muncul pada kasus baru sebesar 11,6%, dengan jumlah persentase kematiannya sekitar 6,6% untuk kanker payudara dan 18,4% untuk kanker paru-paru.

Untuk memprediksi dengan akurat menggunakan ciri-ciri yang ada pada setiap data kanker payudara dan kanker paru-paru, *machine learning* sangat cocok dalam penanganan masalah ini. Dikarenakan *Machine learning* merupakan bagian dari bidang kecerdasan buatan yang mempunyai fokus pada penerapan algoritma maupun metode tertentu untuk melakukan prediksi, pengenalan pola, dan klasifikasi. Pada umumnya tahapan pada

machine learning adalah koleksi data, pembentukan model, tahap pelatihan, serta tahap uji. Pada bidang kesehatan dan pengobatan, klasifikasi yang berbasis *machine learning* telah banyak dilakukan untuk membantu dokter dan ahli kesehatan dalam mengelompokkan jenis tumor, hingga menentukan tindakan pengobatan yang sebaiknya dilakukan (Handri, 2019).

Salah satu *mechine learning* adalah *Naïve Bayes*. *Naïve Bayes* merupakan yang menggunakan metode algoritma teori *bayes* terhadap data *mining* dalam penerapan klasifikasi. Teori keputusan *Bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang ditimbulkan dalam keputusan-keputusan tersebut. Ide dasar dari bayes adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek (Amalia, 2017). *Naïve Bayes* merupakan sekumpulan metode *supervised learning* yang dapat dikelompokkan berdasarkan distribusi dari menjadi bebrapa algoritma khusus antar lain *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes*, dan *Multinomial Naïve Bayes* (Saraswati, 2020).

Penelitian terkait yang sudah pernah dilakukan diantaranya oleh Hasabi (2018) dengan judul “Pengklasifikasian Breast Cancer Dengan Metode Naïve Bayes” yang bertujuan untuk mendeteksi resiko terkena kanker payudara dengan data *Wisconsin Breast Cancer*. Data yang digunakan sebanyak 683 data dengan jumlah 444 data jinak dan 239 data ganas, dengan atribut yang digunakan sebanyak 9 atribut dari total 10 atribut dan mendapatkan nilai *accuracy* sebesar 96%, penelitian lain yang dilakukan oleh Bunga, Djahi, Nabuasa (2018) dengan judul “Multinomial Naïve Bayes Untuk Klasifikasi Status Kredit Mitra Binaan Di Pt. Angkasa Pura I Program Kemitraan” yang bertujuan untuk mendeteksi kredit yang bermasalah dengan menggunakan data sebanyak 148 data mitra dengan 78 mitra yang bermasalah, dengan attribute yang digunakan sebanyak 5 atribut

dan mendapatkan nilai *accuracy* sebesar 86%, *precision* sebesar 73% dan *recall* sebesar 73%. Penelitian lain yang dilakukan Susanti (2016) dengan judul “Klasifikasi Dokumen Berkategori Menggunakan Algoritma Naive Bayes Berbasis Bernoulli” yang bertujuan untuk mengklasifikasi kategori dokumen dengan menggunakan algoritma *Naive Bayes* Berbasis *Bernoulli*. Klasifikasi ini ditekankan pada kategori dokumen diantaranya Ekonomi, Kesehatan, Hiburan dan Teknologi. Data yang digunakan sebanyak 60 dokumen. Pada proses pembuatan sistem ini hanya terdapat 1 user yang dimana user ini dapat menginputkan dokumen yang telah diketahui kategorinya agar sistem melakukan proses pembelajaran, setelah itu user dapat melakukan klasifikasi kepada dokumen yang belum diketahui kategorinya dan mendapatkan nilai *precision* sebesar 70%, *accuracy* 65% dan *recall* 70%

Berdasarkan latar belakang yang sudah dijelaskan tentang preferensi kasus kanker dan metode *naive bayes*, maka penulis akan membuat penelitian dengan menggunakan metode *naive bayes classifier* dalam pengklasifikasian kanker payudara dan kanker paru-paru, dengan judul **“Pengklasifikasian Kanker Payudara (*Breast Cancer*) dan Kanker Paru- Paru (*Lung Cancer*) dengan Metode *Gaussian Naive Bayes*, *Multinomial Naive Bayes*, dan *Bernoulli Naive Bayes*”**.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka penulis dapat merumuskan sebuah masalah dalam penelitian ini yaitu seberapa besar kinerja metode *Naive Bayes Gaussian*, *Multinomial*, dan *Bernoulli* dalam pengklasifikasian kanker payudara dan paru-paru.

1.3. Tujuan Penelitian

Untuk mengetahui perbandingan kinerja *accuracy*, *precision*, dan *recall*, dalam kinerja metode *Naive Bayes Gaussian*, *Naive Bayes*

Multinomial, dan *Naive Bayes Bernoulli* dalam pengklasifikasian kanker payudara dan paru-paru.

1.4. Manfaat

Diharapkan dengan dibuatnya tugas akhir ini dapat memberikan manfaat sebagai berikut ::

1. Diharapkan dapat diketahui *accuracy*, *precision*, dan *recall* untuk kinerja dari metode *Naive Bayes Gaussian*, *Naive Bayes Multinomial*, dan *Naive Bayes Bernoulli* dalam pengklasifikasian kanker payudara dan paru-paru.
2. Menambah metode yang bisa digunakan dalam penelitian lebih lanjut terhadap penyakit kanker payudara dan paru-paru, sehingga di masa depan pendeteksian dini kanker semakin meningkat.

1.5. Batasan Masalah

Yang menjadi batasan masalah dalam tugas akhir ini adalah :

1. Data yang digunakan adalah Database Lung Cancer(2017) (<https://data.world/sta427ceyin/survey-lung-cancer>) dan Database Breast Cancer(1995) (<https://data.world/uci/breast-cancer-wisconsin-diagnostic>)
2. Metode yang digunakan adalah metode *Naive Bayes Gaussian*, *Multinomial*, dan *Bernoulli*
3. Metode pengujian kinerja yang digunakan adalah *accuracy*, *precesion*, dan *recall*.
4. Tools pemrograman python yang digunakan adalah google colab.