

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Kemajuan teknologi saat ini tidak terlepas dari kehidupan masyarakat. Berbagai informasi bermunculan di berbagai belahan dunia yang kini dapat kita ketahui secara instan berkat kemajuan teknologi (*globalisasi*). Tentunya kemajuan teknologi tersebut telah menyebabkan perubahan yang luar biasa dalam kehidupan masyarakat dengan segala peradaban dan budayanya. Perubahan ini juga sangat mempengaruhi transformasi nilai-nilai yang ada di masyarakat, tidak terlepas dalam aspek Kesehatan salah satunya mengenai kanker payudara [1].

Kanker adalah penyakit tidak menular yang ditandai dengan pertumbuhan sel yang tidak normal atau terus menerus dan tidak terkendali, yang dapat merusak jaringan di sekitarnya dan menyebar ke tempat yang jauh dari asal. kanker dapat ditemukan baik pada pria maupun wanita. Perubahan tubuh yang disebabkan oleh kanker mudah dilihat dengan mata telanjang. Perubahan fisik yang terjadi ini dapat berupa pengobatan seperti operasi untuk mengangkat sel kanker pada wanita penderita kanker payudara atau lebih dikenal dengan *mastektomi*. Perubahan fisik lainnya yang disebabkan oleh kemoterapi terlihat dengan mata telanjang pada pria dan Wanita. Kanker payudara merupakan kanker nomor satu di Indonesia dan merupakan salah satu penyebab utama kematian akibat kanker. Pada tahun 2020 jumlah kasus baru kanker payudara mencapai 68.858 kasus (16,6%) dari total

396.914 kasus baru kanker di Indonesia. Sedangkan, mengenai jumlah kematiannya sudah mencapai lebih dari 22 ribu jiwa kasus [2]. Dengan kasus yang kian banyak, penerapan sebuah metode sangat diperlukan supaya meminimalisir kesalahan dalam menentukan keputusan, Sehingga, dengan metode yang tepat dalam proses implementasinya, dapat menggunakan sebuah metode yaitu metode *naïve bayes* dan *C4.5*.

Kinerja algoritma *naïve bayes* untuk klasifikasi kanker payudara memberikan nilai yang baik, dengan persentase rata-rata data yang terklasifikasi benar sebesar 96,9%, sedangkan persentase rata-rata data yang salah klasifikasi hanya 3,1% [3].

Akurasi yang diperoleh dari pemodelan algoritma *C4.5* dengan pembobotan atribut, yaitu sebesar 98,57%. Akurasi tersebut merupakan hasil dari kesesuaian antara prediksi klasifikasi dan hasil klasifikasi [4].

Sehingga dari beberapa penelitian mengenai kanker payudara yang sudah dicantumkan di atas dapat disimpulkan bahwa penggunaan kedua metode tersebut akan menjadi perbandingan mengingat hasil dari kedua metode tersebut yang tidak terlalu jauh. Sehingga hal tersebut yang melandasi penulis dalam membandingkan kedua metode tersebut untuk memperoleh hasil dengan persentase yang tinggi dalam penelitian ini. Dengan beberapa argumen diatas, maka penulis menentukan judul penelitian yang telah dipahami dan didiskusikan dengan beberapa pihak. Maka dari itu penulis mengangkat judul penelitian :

“ANALISIS DATA MINING UNTUK PREDIKSI KANKER PAYUDARA MENGGUNAKAN ALGORITMA KLASIFIKASI”

Dengan adanya penelitian ini di harapkan mampu untuk menentukan keputusan terhadap penderita kanker payudara dengan menggunakan metode *naïve bayes* dan *C4.5*.

1.2 RUMUSAN MASALAH

Masalah yang di bahas oleh penulis dalam penelitian ini meliputi :

1. Membuat sebuah model dengan akurasi yang tinggi dalam mengklasifikasi kanker payudara
2. Menerapkan suatu metode yang tepat serta terstruktur yaitu *naïve bayes* dan *C4.5* dalam menentukan metode yang lebih baik dalam penelitian mengenai penderita kanker payudara yang terindikasi ganas ataupun jinak
3. Memaksimalkan hasil analisis sebagai keputusan yang dapat digunakan sebagai landasan dan dapat dipertanggung jawabkan

1.3 BATASAN MASALAH

Batasan masalah yang digunakan dalam pembahasan penelitian ini adalah untuk tujuan membuat interpretasi atau elaborasi yang berorientasi rinci dan konsisten dengan tujuan penulis, oleh karena itu penulis menetapkan batasan masalah, yaitu :

1. Objek yang menjadi prioritas dalam penelitian ini adalah penderita kanker payudara
2. Metode yang digunakan dalam penelitian ini adalah *naïve bayes* dan *C4.5*

3. Keluaran yang dihasilkan yaitu mengenai apakah dengan indikasi-indikasi terkait pasien kanker payu dara dapat dikatakan sebagai kanker “ganas” atau “jinak” sehingga menjadi landasan dalam penelitian terkait penderita kanker payudara
4. Pengujian hasil analisis menggunakan aplikasi *WEKA*
5. Dataset dalam penelitian ini adalah dataset publik yang diperoleh melalui website *Kaggle.com* dengan jumlah data sebanyak 569 data dan dapat diakses pada link : <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset?resource=download>

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Tujuan yang akan disampaikan penulis dalam penelitian ini adalah :

1. Melakukan *data preparation* dan *exploratory data analysis*
2. Membuat model (komparasi kedua metode) untuk memperoleh akurasi yang tinggi dalam melakukan penelitian terkait penderita kanker payudara
3. Untuk mengetahui akurasi dari model yang di bangun dalam menentukan penderita kanker payudara dengan penggolongan “ganas” atau “jinak” dengan menggunakan *naïve bayes* dan *C4.5*

1.4.2. Manfaat Penelitian

Manfaat yang akan disampaikan penulis dalam penelitian ini adalah :

1. Dengan adanya pemanfaatan algoritma tersebut dapat membantu beberapa pihak dalam menentukan ataupun sebagai landasan dalam penerapan penderita kanker payudara yang terindikasi
2. Menentukan sebuah keputusan dengan tingkat akurasi yang tinggi dari model yang dibangun bagi penderita kanker payudara terkait penggolongan “ganas” ataupun “jinak”
3. Memperoleh hasil dari model yang dibangun dengan penerapan yang terstruktur dalam *data preparation* dan *exploratory data analysis* pada penelitian mengenai kanker payudara

1.5 SISTEMATIKA PENULISAN

Untuk mempermudah dalam memahami penulisan laporan penelitian ini, maka penulis sajikan sistematika penulisan penelitian ini sebagai berikut :

BAB I : PENDAHULUAN

Bab ini akan memaparkan latar belakang masalah, cara merumuskannya, batasan-batasannya, tujuan dan manfaat menulis, serta sistematisasi penulisan.

BAB II : LANDASAN TEORI

Dalam bab landasan teori ini akan dibahas teori-teori dan pendapat para ahli terkait dengan permasalahan yang dianalisis. Teori yang digunakan antara lain data mining, klasifikasi, pasien atau penderita, pernyataan terkait *Naive Bayes*, *C4.5*, *WEKA* dan *RapidMiner*.

BAB III : METODOLOGI PENELITIAN

Bab ini menjelaskan kerangka penelitian, metode pengumpulan data, metode klasifikasi, dan alat yang digunakan dalam penelitian ini.

BAB IV : HASIL ANALISIS DAN PEMBAHASAN

Dalam bab ini, perhitungan analitik menggunakan metode *Naive Bayes* dan C4.5 dilakukan pada data penderita yang akan digunakan dalam penelitian ini dan bab ini akan memaparkan hasil visualisasi dari analisis tools *WEKA* dan *RapidMiner* yang digunakan.

BAB V : PENUTUP

Dalam bab ini merupakan penutupan dari penelitian ini yang berisi kesimpulan dari pembahasan bab-bab sebelumnya dan juga saran-saran yang berguna bagi pihak-pihak yang berkaitan dan berhubungan terhadap penelitian ini