

BAB I PENDAHULUAN

1.1 Latar Belakang

Kanker Payudara adalah tumor ganas yang menyerang jaringan payudara yang berasal dari kelenjar, saluran kelenjar dan jaringan penunjang payudara. Kanker payudara terjadi karena adanya kerusakan gen yang mengatur pertumbuhan dan diferensiasi sehingga sel ini tumbuh dan berkembang biak tanpa dapat dikendalikan [1]. Kanker payudara menjadi pembunuh wanita terbanyak di dunia. Namun begitu, laki-laki juga bisa terkena penyakit ini, tetapi kemungkinan pada wanita 100 kali lipat dibandingkan pada laki-laki. Sebagian besar kanker payudara berasal dari sel-sel duktus (86%), kemudian lobular (12%), dan sisanya berasal dari jaringan lain [2]. salah satu cara pengobatan yang umum dilakukan adalah dengan pembedahan dan jika perlu dilanjutkan dengan kemoterapi maupun radiasi. Namun pengobatan tersebut tidak akan memberikan dampak yang signifikan jika kankernya sudah mencapai stadium akhir. Oleh karena itu apabila penyakit ini dapat dideteksi lebih awal, dampak buruk yang mungkin ditimbulkan oleh kanker payudara dapat dicegah. Hal ini menunjukkan bahwa kanker payudara merupakan penyakit yang sangat ganas dan mengharuskan penderitanya untuk melakukan pemeriksaan yang intensif [3].

Dengan adanya data kanker payudara dan pengklasifikasiannya, maka dibutuhkan teknologi informasi yang dapat mempermudah pengguna untuk memahami dan mempelajari data sehingga dapat menghasilkan informasi yang bermanfaat. Dengan berkembangnya teknologi, terdapat berbagai macam teknik yang dapat digunakan. Salah satunya ialah teknik klasifikasi. Teknik klasifikasi sendiri merupakan sebuah metode pembelajaran secara terbimbing (*Supervised Learning*). Teknik klasifikasi bertujuan untuk memprediksi kelas dari sebuah objek yang kelasnya belum diketahui sehingga dapat menghasilkan informasi yang baru. Klasifikasi memiliki prosedur-prosedur dalam menyelesaikan sebuah masalah. Prosedur dalam melakukan klasifikasi akan berpengaruh terhadap hasil yang diharapkan. Penelitian ini akan menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan data kanker payudara. Pemilihan *Naïve Bayes* berdasarkan beberapa penelitian yang telah dilakukan sebelumnya menggunakan algoritma

Naïve Bayes. Beberapa kelebihan dari *Naïve Bayes*, antara lain: memiliki performa yang lebih baik dari beberapa algoritma lainnya [4] [5], serta tidak membutuhkan data yang besar untuk memulai proses pelatihan [6], *Naïve Bayes* memiliki tingkat akurasi dan kecepatan yang tinggi ketika di aplikasikan pada data dalam jumlah besar [7] dan *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [8]. penelitian ini menggunakan *Naïve Bayes* sebagai algoritma klasifikasi dikarenakan performa yang dimilikinya.

Data dengan jumlah besar mengandung banyak fitur, baik fitur yang relevan, dan tidak relevan. Membiarkan fitur-fitur yang tidak relevan akan membingungkan proses klasifikasi data sehingga akan menurunkan kecepatan klasifikasi, meningkatkan biaya komputasional dan penggunaan memori, serta akan sangat berpengaruh terhadap hasil klasifikasi itu sendiri [7] [9]. Banyaknya fitur dapat menyebabkan *overfit* pada model sehingga menyebabkan penurunan performa yang dimiliki oleh model. Oleh karena itu, perlu dilakukan tahapan *preprocessing* untuk pemilihan fitur yang relevan.

Seleksi fitur merupakan teknik pemilihan fitur yang relevan berdasarkan kriteria tertentu sehingga dapat meningkatkan performa pelatihan seperti meningkatkan akurasi klasifikasi, menurunkan biaya komputasi dan interpretasi model yang lebih baik [10]. Seleksi fitur memiliki tiga model yang biasa digunakan yaitu *wrapper*, *filter*, dan *embedded*. Keunggulan dari seleksi fitur *filter* ialah memiliki waktu komputasi yang lebih rendah dari pada tipe lainnya, sederhana dan cepat, mengukur data dimensi tinggi dengan mudah, serta bersifat independen dari algoritma klasifikasi. Salah satu algoritma seleksi fitur tipe *filter* adalah *Chi Square*. Algoritma ini dapat menangani proses seleksi fitur dengan cepat dan lebih efektif dalam menghilangkan fitur dengan akurasi yang baik [11]. Dari kelebihan yang dipaparkan diatas, maka penelitian ini menggunakan seleksi fitur dengan tipe *filter*. Algoritma yang akan digunakan adalah *chi square*. Oleh sebab itu pada penelitian ini akan membahas mengenai pengaruh seleksi fitur terhadap model klasifikasi *Naïve Bayes* pada data kanker payudara Dengan penerapan seleksi fitur diharapkan dapat meningkatkan performa model klasifikasi *Naïve Bayes* pada data kanker payudara.

1.2 Perumusan Permasalahan

Ada 3 poin penting dalam perumusan masalah pada penelitian ini.

1.2.1 Permasalahan

Apakah pengaruh seleksi fitur dengan metode filter dalam meningkatkan akurasi klasifikasi Naïve bayes pada data kanker payudara?

1.2.2 Usulan Solusi

Metode naïve bayes untuk proses pengklasifikasian kanker payudara dan seleksi fitur chi square untuk menghilangkan fitur yang tidak relevan serta meningkatkan hasil akurasi.

1.2.3 Pertanyaan Penelitian

Pertanyaan peneliti pada peniliiian ini adalah :

Apa pengaruh penambahan seleksi fitur chi square terhadap nilai akurasi yang dihasilkan pada klasifikasi data kanker payudara menggunakan naïve bayes?

Tabel 1.1 Perumusan Masalah

Masalah	Solusi	Pertanyaan Penelitian	Tujuan Penelitian
Membiarkan fitur-fitur yang tidak relevan akan membingungkan proses klasifikasi kanker payudara sehingga dapat kecepatan klasifikasi dan berpengaruh terhadap hasil klasifikasi.	Metode naïve bayes untuk proses pengklasifikasi kanker payudara dan seleksi fitur chi square untuk menghilangkan fitur yang tidak relevan serta meningkatkan hasil akurasi.	Bagaimana pengaruh penambahan seleksi fitur chi square terhadap nilai akurasi yang dihasilkan pada klasifikasi data kanker payudara menggunakan naïve bayes?	tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui pengaruh penambahan seleksi fitur menggunakan chi square terhadap klasifikasi <i>Naïve Bayes</i> pada data kanker payudara.

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui pengaruh penambahan seleksi fitur menggunakan chi square terhadap klasifikasi *Naïve Bayes* pada data kanker payudara.

1.3.2 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah untuk mengetahui fitur-fitur yang relevan dari sebuah dataset kanker payudara setelah penerapan seleksi fitur sehingga dapat meningkatkan performa klasifikasi *Naïve Bayes*.

1.4 Batasan Masalah

1. Data berasal dari *UCI Machine Learning Repository*. Dataset yang digunakan merupakan dataset *Breast Cancer Wisconsin (original)* diperoleh dari University of Wisconsin Hospitals, Madison dari Dr. William H. Wolberg. Dataset ini terdiri dari 699 jumlah data dan 9 fitur.
2. Dataset yang akan digunakan merupakan dataset yang telah bersih dari *noise* dan *missing value*.
3. Metode seleksi fitur yang akan digunakan adalah tipe *filter* dengan algoritma *Chi square*.

1.5 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian ini, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini menjelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai kanker payudara dan fiturnya, seleksi fitur chi square , naïve bayes dan *k-fold cross validation*.

BAB III METODE USULAN

Pada bab ini menjelaskan tentang metode yang dijadikan solusi permasalahan yaitu *naïve bayes* menggunakan seleksi fitur *chi square* . pada bab

ini juga menjelaskan tahapan penelitian ,arsitektur sistem yang digunakan, rancangan evaluasi ,serta dataset yang akan digunakan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan hasil dari rancangan yang telah dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan selanjutnya.