

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kanker payudara merupakan tumor ganas yang menyerang jaringan sel-sel payudara. Kanker payudara merupakan masalah paling besar bagi wanita di seluruh dunia dan menyebabkan kematian utama bagi penderita kanker payudara. Penyakit kanker payudara di negara berkembang menunjukkan bahwa penyakit kanker dengan persentase kasus tertinggi, kurang lebih 43% kasus dan persentase kematian yaitu 12,9% (*American Cancer Society Inc*, 2012) dalam (Mulyasari, et al, 2017). Menurut WHO sekitar 8-9% wanita menderita penyakit kanker payudara. Kasus kanker payudara terus meningkat lebih dari 250,000 kasus baru, di Eropa dilakukan penelitian kanker payudara oleh *American Cancer Society*(ACS) hampir 178.000 wanita yang telah diagnosis kanker payudara dan jumlah tersebut ditambah 2 juta wanita yang memiliki riwayat penyakit ini (*American Cancer Society Inc*, 2012) dalam (Mulyasari, et al, 2017). Penyakit kanker payudara masih menjadi masalah utama dalam dunia kesehatan, terbukti dari berbagai kasus komplikasi fisik fungsional dan dapat juga menyebabkan gangguan kualitas hidup. Penurunan kualitas hidup wanita penderita kanker payudara dapat dilihat dari sisi kesehatan fisik, status psikologi, hubungan sosial, kemandirian dan spiritual. Kualitas hidup merupakan persepsi individu dalam kemampuan, keterbatasan psikologi dalam konteks budaya dan sistem nilai untuk mengetahui peran dan fungsi (Rabin, E. G et al, 2008). Bahaya dari penyakit ini bukan hanya berdampak pada fisik pada penderita melainkan pada sisi psikologi juga. Keterlambatan mengetahui akan merambatnya penyakit ini adalah salah satu faktor yang paling sering dialami. Ketidaktahuan informasi penanganan sejak dini terhadap tanda-tanda kemunculan penyakit ini menyebabkan penderita harus mengalami operasi, bahkan meninggal dunia (Laksono, 2018). Deteksi dini pada penyakit kanker sangatlah dibutuhkan agar penderita dapat segera ditolong. Deteksi dini menggunakan Teknik Data Mining pada penyakit kanker sudah banyak dikembangkan (Farahdiba & Nugroho, 2016)

Salah satu teknik *data mining* yang dipakai dalam penelitian terhadap dunia kesehatan adalah klasifikasi. Meliputi *Decicion tree*, *K-NN*, *Bayesian* dan lain-lain. Berikut ini beberapa penelitian terhadap kanker payudara menggunakan metode klasifikasi. Menurut Uma Ojha dalam penelitiannya *A study on prediction of breast cancer recurrence using data mining techniques* menggunakan metode *decision tree C5.0* dan *SVM*, hasil yang didapat bahwa dalam pengukuran performansi *C5.0* memiliki tingkat akurasi 81%. Dalam penelitian Shubham Sharma berjudul *Breast Cancer Detection Using Machine Learning* terhadap data *breast cancer wisconsin* menggunakan metode *Random Forest*, *kNN (k-Nearest-Neighbor)* dan *Naïve Bayes* dimana pada hasil yang diperoleh adalah metode *K-NN* memiliki tingkat akurasi paling tinggi yaitu 95,90%. *Gaussian Naive Bayes* adalah salah satu metode klasifikasi yang juga digunakan dalam penelitian dunia medis diantaranya pada penelitian Ahmad Zaki Hakimi berjudul *Klasifikasi Sel Darah Putih Menggunakan Gaussian Naïve Bayes* menyebutkan dalam kesimpulannya yaitu didapatkan bahwa fitur dari inti sel darah putih sangat penting dalam mengklasifikasi sel darah putih serta penerapan algoritma *Gaussian Naïve Bayes* menghasilkan *F1 score* sebesar 91,26%. Untuk mencapai akurasi tersebut, *preprocessing* data berupa standarisasi serta seleksi fitur dengan metode *sequential feature selection* perlu dilakukan. *Gaussian naive bayes* digunakan ketika tipe data tidak dapat didistribusikan menggunakan pendekatan distribusi kelas (kategorikal). Performansi klasifikasi *Naive Bayes* akan lebih baik ketika atribut numerik didiskritisasi daripada diasumsikan dengan pendekatan distribusi (kategorikal) (Dougherty, et al, 1995). Identifikasi penyakit kanker sejak dini dan pengembangan menggunakan Teknik data minning menjadi landasan alasan peneliti melakukan penelitian dengan judul *Klasifikasi Data Penderita Breast Cancer Menggunakan Metode Gaussian Naive Bayes*.

1.2. Rumusan Masalah

Dari uraian di atas penulis merumuskan masalah yang akan menjadi pokok penelitian yaitu :

1. Berapakah akurasi tertinggi yang dihasilkan metode *Gaussian Naive Bayes* dengan skenario uji *k-fold cross validation*?

2. Berapakah presisi tertinggi yang dihasilkan metode *Gaussian Naive Bayes* dengan skenario uji *k-fold cross validation*?

1.3. Tujuan

Sesuai rumusan masalah yang telah dibangun, maka tujuan dari penelitian ini adalah:

1. Mengetahui tingkat akurasi dari metode *Gaussian Naive Bayes* terhadap klasifikasi data penderita kanker payudara.
2. Mengetahui tingkat presisi dari metode *Gaussian Naive Bayes* terhadap klasifikasi data penderita kanker payudara.

1.4. Manfaat

Dengan hasil dari penelitian ini diharapkan akan adanya manfaat terhadap peneliti, peneliti lain dan dunia kesehatan diantaranya

1. Manfaat terhadap penulis adalah pengembangan *data mining* dan implementasi metode *Gaussian Naive Bayes* dalam dunia kesehatan khususnya pada pendeteksian kanker payudara.
2. Manfaat terhadap peneliti lain adalah sebagai bahan literatur review dalam penelitian serupa juga sebagai dasar untuk pengembangan penelitian selanjutnya.
3. Manfaat terhadap dunia medis adalah sebagai bahan pertimbangan dalam penanganan pasien penderita breast cancer.

1.5. Batasan Masalah

Penelitian ini berfokus pada usulan pada rumusan masalah yang dibangun. Agar pembahasan dapat mengerucut dan berfokus pada pokok permasalahan berikut batasan masalah pada penelitian ini.

1. Data yang digunakan dalam penelitian ini adalah data pasien penderita kanker yang dihimpun University Hospital Centre of Coimbra dan diserahkan pada *UCI machine learning* dengan tahun data 2018.

2. *Output* yang dihasilkan yaitu *healty control* dengan kode 1 dan *patient* dengan kode 2.
3. Atribut yang digunakan diantaranya *age*, *BMI*, *glucose*, *insulin*, *HOMA*, *leptin*, *adiponectin*, *resistin* dan *MCP* (merujuk pada data breast cancer UCI machine learning pada penelitian *Hyperresistinemia and metabolic dysregulation: a risky crosstalk in obese breast cancer* dan *Using Resistin, glucose, age and BMI to predict the presence of breast cancer*).
4. Skenario uji yang digunakan adalah *k-fold cros validation* dengan nilai $k = 2, 4, 5, 8$ dan 10 .
5. Uji performansi yang dilakukan adalah pengukuran akurasi dan presisi.

