

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Kanker payudara masih menjadi hal yang menakutkan bagi wanita di Indonesia. Oleh karena itu, pentingnya kesadaran untuk mendeteksi sedari dini. Menurut Ayustawati [1] Kanker payudara merupakan sebuah tumor ganas yang dihasilkan dari pertumbuhan sel-sel abnormal di payudara. Kanker payudara umumnya terjadi pada wanita, namun tidak menutup kemungkinan bahwa pria juga bisa terkena kanker payudara meskipun dengan tingkat presentase yang sangat kecil. Jika kanker ini terdeteksi sejak dini dan belum menyebar ke luar payudara, pengobatan masih bisa dilakukan. Menurut Ridwan Sholihin [2] Faktor-faktor yang berkontribusi terhadap kanker payudara antara lain usia, genetik, obat-obatan, menstruasi dini, tidak menikah, tidak memiliki anak, melahirkan di usia lanjut, tidak pernah menyusui, dan stres. Kanker payudara menempati urutan kelima sebagai penyebab kematian setelah kanker paru, kanker rahim, kanker hati, dan kanker usus [2].

Risiko terkena kanker payudara memang tidak bisa dihindari. Namun jika kanker payudara terdiagnosis, pencegahan dan deteksi dini dapat dilakukan. Salah satu metode yang dapat digunakan dalam permasalahan ini adalah *data mining* dengan teknik klasifikasi. Dengan teknik klasifikasi *data mining*, peneliti

mendapatkan hasil berdasarkan data jenis tumornya. Jenis tumor kanker payudara terbagi menjadi dua, yaitu *Benign* (jinak) dan *Malignant* (ganas).

Beberapa penelitian data mining dengan berbagai metode klasifikasi untuk mengetahui kinerja algoritma klasifikasi dalam mendeteksi penyakit kanker payudara dilakukan perbandingan algoritma diantaranya: Agung Mulyo Widodo, Nizirwan Anwar, Bambang Irawan, Lista Meria, dan Andika Wisnujati dengan perbandingan algoritma KNN, *Bagging*, dan *Random Forest* dalam mengklasifikasikan penyakit kanker payudara yang berulang dan tidak berulang. KNN memiliki akurasi sebesar 74.37% dengan akurasi yang lebih baik dari algoritma lainnya [3]. Selanjutnya penelitian yang dilakukan oleh Annisa Nugraheni, Rima Dias Ramadhani, Amalia Beladinna Arifa, dan Agi Prasetiadi dalam mendiagnosa pasien penyakit kanker payudara atau orang yang sehat dengan menggunakan perbandingan algoritma *Naïve Bayes* dan KNN. Hasilnya KNN mendapatkan akurasi, presisi sehat, recall sakit dan AUC yang lebih tinggi dibandingkan *Naïve Bayes*. Dapat dikatakan bahwa algoritma KNN lebih baik dibandingkan dengan *Naïve Bayes* [4]. Berikutnya penelitian yang dilakukan oleh Lusa Indah Prahartiwi dan Wulan Dari menggunakan algoritma *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine* mengenai diagnosa yang mengalami kanker payudara. Dari ketiga algoritma tersebut, diperoleh tingkat akurasi tertinggi sebesar 74.29% dari algoritma *Support Vector Machine*. Sehingga algoritma *Support Vector Machine* (SVM) memiliki performa terbaik dibandingkan dengan algoritma *Naïve Bayes* dan *Decision Tree* [5].

Data mining memiliki beberapa algoritma untuk mengolah data khususnya dalam teknik klasifikasi yaitu *Naïve Bayes* dan C4.5. Dalam penelitian Achmad Ridwan, *Naïve Bayes* terbukti memiliki akurasi yang baik dalam mengklasifikasi penyakit diabetes mellitus dengan hasil akurasi sebesar 90.20% [6]. Selain itu, Sanni Ucha putri, Eka Irawan, dan Fitri Rizky melakukan penelitian tentang prediksi penyakit diabetes menggunakan algoritma C4.5 yang mendapatkan hasil memuaskan yaitu sebesar 90.00% [7]. Berdasarkan penelitian tersebut, algoritma yang digunakan berbeda namun menghasilkan akurasi yang sama-sama baik. Disisi lain kinerja algoritma klasifikasi tidak hanya dilihat dari nilai akurasi. Oleh karena itu, perlu dilakukan komparasi kinerja algoritma klasifikasi dengan memperhatikan nilai akurasi, presisi, *recall*, dan ROC.

Pada penelitian ini, peneliti melakukan komparasi algoritma *Naïve Bayes* dan C4.5. Perbandingan algoritma dilakukan untuk mengetahui kinerja algoritma mana yang lebih baik dan mendapatkan hasil yang akurat dalam mengklasifikasi penyakit kanker payudara. Berdasarkan pemaparan latar belakang diatas, maka penulis melakukan penelitian yang berjudul **“KOMPARASI ALGORITMA NAÏVE BAYES DAN DECISION TREE C4.5 UNTUK KLASIFIKASI PENYAKIT KANKER PAYUDARA”**

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dari penelitian ini adalah “Bagaimana kinerja dari algoritma *Naïve Bayes* dan C4.5 dalam mengklasifikasi penyakit kanker payudara?”

1.3 BATASAN MASALAH

Agar tidak meluasnya permasalahan yang ada dan lebih terarah maka dilakukan pembatasan masalah sebagai berikut:

1. Menggunakan *dataset Breast Cancer Wisconsin (Original)* dari website *UCI Machine Learning Repository*
2. Atribut yang digunakan yaitu: *Sample Code Number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses,* dan *Class*
3. Hasil klasifikasi jenis kanker payudara dalam penelitian ini terbagi menjadi dua kategori, yaitu *Benign* (tumor jinak) dan *Malignant* (tumor ganas).
4. Pengukuran kinerja algoritma *Naïve Bayes* dan *C4.5* menggunakan nilai akurasi, presisi, *recall*, dan *ROC*
5. Alat bantu yang digunakan dalam menganalisis adalah *WEKA*

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dari penelitian ini adalah:

1. Mengaplikasikan algoritma *Naïve Bayes* dan *C4.5* untuk klasifikasi penyakit kanker payudara
2. Menganalisa kinerja algoritma *Naïve Bayes* dan *C4.5* dalam klasifikasi penyakit kanker payudara

1.4.2 Manfaat Penelitian

Adapun manfaat yang di harapkan oleh peneliti dalam penelitian ini adalah sebagai berikut:

1. Mendapatkan algoritma dengan kinerja terbaik dalam mengklasifikasi penyakit kanker payudara
2. Dapat menambah wawasan dan informasi bagi peneliti serta hasil dari penelitian ini bisa menjadi pertimbangan atau masukan sebagai bahan acuan bagi penelitian selanjutnya

1.5 SISTEMATIKA PENULISAN

Untuk memberikan gambaran umum mengenai keseluruhan penulisan ilmiah, dapat dilihat melalui sistematika penulisan yang meliputi:

BAB I PENDAHULUAN

Pada bab ini menguraikan tentang latar belakang masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab landasan teori ini membahas tentang teori-teori dan pendapat para ahli yang berhubungan dengan pembahasan yang dianalisis. Teori-teori yang digunakan antara lain mengenai *Data Mining*, *Naïve Bayes*, *Decision Tree*, *C4.5*, *WEKA*, dan kanker payudara.

BAB III METODOLOGI PENELITIAN

Dalam bab ini menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, proses yang dilakukan, prosedur penelitian, dan alat bantu (*tools*) yang digunakan untuk mendukung penelitian.

BAB IV ANALISIS

Bab ini berisi tentang perhitungan, hasil, visualisasi, serta analisis permasalahan penelitian yang berupa pengujian mengenai algoritma yang digunakan untuk klasifikasi penyakit kanker payudara.

BAB V PENUTUP

Bab ini disajikan kesimpulan dari hasil penelitian dan saran-saran yang ditujukan kepada semua pihak yang bersangkutan.

