

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker merupakan sebuah penyakit mematikan di dunia. Menurut (Batubara, 2021), 12% dari angka kematian di dunia disebabkan oleh penyakit kanker yang merupakan penyakit pembunuh nomor dua di dunia. Kanker bisa terjadi dikarenakan terjadinya pertumbuhan sel abnormal yang tidak terkendali di tubuh manusia. Tubuh manusia yang terbangun oleh triliunan sel yang memiliki jangka waktu hidup yang terbatas, sedangkan sel kanker yang terus bertumbuh tidak akan mati dengan sendirinya. Sel tersebut akan terus memperbanyak diri, dan inilah penyebab terjadinya kanker.

Kanker payudara merupakan salah satu varian dari penyakit kanker yang hanya bisa dialami oleh Wanita. Menurut (Andryan, Fajri, & Sulistyowati, 2022), kanker payudara sering ditemukan pada Wanita di Indonesia. Kanker payudara terjadi karena pertumbuhan *neoplasma* yang tidak biasa yang berasal dari *paranchyma*. Menurut data dari Departemen Kesehatan RI (2013) diambil dari (Batubara, 2021), bahwa penderita kanker tertinggi di Indonesia untuk Wanita adalah kanker payudara. Dimana 42,1 per 100.000 penduduk dengan rata-rata kematian 17 per 100.000 penduduk.

Dapat dilihat bahwa data diatas menunjukkan betapa tingginya kasus kanker payudara di Indonesia, maka karena itu pemerintah Indonesia telah menerapkan tata laksana deteksi dini untuk kanker payudara dengan melakukan metode pemeriksaan payudara sendiri atau yang dikenal sebagai SADARI (Batubara, 2021). Akan tetapi metode yang digunakan sampai saat ini adalah melakukan inspeksi sendiri dengan melakukan inspeksi dan perabaan untuk mencari tanda-tanda gejala terjadinya kanker payudara. Cara lain untuk

mendeteksi kanker payudara secara dini adalah melalui tumor payudara yang biasanya dibagi menjadi dua yaitu *benign* dan *malignant*.

Untuk mendapatkan hasil yang lebih akurat maka *machine learning* akan digunakan untuk mempelajari data set yang diberikan. *Machine learning* yang telah melewati tahap *training*/pelatihan akan mampu melakukan mendeteksi berdasarkan data yang diberikan apakah orang tersebut akan memiliki kanker payudara. Sebelumnya sudah terdapat penelitian yang menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)*. Akan tetapi algoritma diatas memiliki kelemahan dalam memiliki tingkat akurasi yang rendah, algoritma *Naïve Bayes* yang mengasumsikan bahwa semua prediktor atau fitur adalah independen, dimana sangat jarang terjadi dalam situasi kehidupan nyata. Ini akan membatasi penerapan algoritma dalam kasus penggunaan di dunia nyata. Sedangkan algoritma KNN, hanya mencari pola, perbedaan, dan kesamaan dari keseluruhan data (Chahboun & Maaroufi , 2021). Terdapat juga satu penelitian yang membandingkan dua algoritma *Support Vector Machine* dan *XGBoost*, dimana hasil penelitian tersebut membuktikan bahwa algoritma *XGBoost* jauh lebih unggul dibandingkan algoritma *Support Vector Machine* (Andryan, Fajri, & Sulistyowati, 2022).

Maka karena itu penulis mengusulkan untuk membandingkan algoritma *Random Forest* dengan algoritma *XGBoost* dalam melakukan pendeteksi kanker payudara untuk mengharapkan mendapatkan hasil yang lebih tinggi. Dibandingkan dengan algoritma *KNN*, *decision tree*, dan *Naïve Bayes*, Algoritma *Random Forest* memiliki tingkat *Variance* yang lebih rendah (Abdulkareem & Abdulazeez, 2021). Algoritma *Random Forest* yang hampir mirip dengan *decision trees*, akan tetapi *Random Forest* akan membuat subset secara acak dan membangun *decision trees* yang lebih kecil menggunakan subset tersebut.

Pada skripsi ini, penulis akan melakukan pendeteksi kanker payudara dengan implementasi algoritma *Random Forest* dalam *Machine Learning*. Untuk membantu para wanita dan dokter untuk mendapatkan hasil diagnosis kanker payudara yang lebih akurat. Oleh karena itu penulis akan mengangkat judul “Perbandingan Metode Algoritma *Random Forest* dengan *XGBoost* Dalam Mendeteksi Kanker Payudara”.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut maka terdapat rumusan masalah adalah sebagai berikut :

1. Bagaimana algoritma *Random Forest* dan algoritma *XGBoost* mengklasifikasi *dataset* kanker payudara?
2. Bagaimana *accuracy* dari hasil diagnosis *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*?
3. Bagaimana *precision* dari hasil diagnosis *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*?
4. Bagaimana *recall* dari hasil diagnosis *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*?

1.3 Tujuan Penelitian

Adapun tujuan yang diharapkan berdasarkan hasil penelitian ini adalah sebagai berikut:

1. Mengetahui bagaimana cara algoritma *Random Forest* dan algoritma *XGBoost* mengklasifikasi *dataset* kanker payudara.
2. Mengetahui *accuracy* dari hasil diagnosa *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*.

3. Mengetahui *precision* dari hasil diagnosa *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*.
4. Mengetahui *recall* dari hasil diagnosa *machine learning* dengan menggunakan algoritma *Random Forest* dan algoritma *XGBoost*.

1.4 Batasan Masalah

Supaya penelitian ini mampu mencapai hasil dan tujuan yang optimal, maka permasalahan yang dibatasi sebagai berikut:

1. Penelitian dilakukan dalam platform Google Colaboratory.
2. Penelitian hanya membahas mengenai pendeteksi kanker payudara melalui dua tipe tumor *benign* dan *malignant*
3. Penelitian ini berfokus pada implementasi algoritma *Random Forest* dan algoritma *XGBoost* pada *machine learning*.
4. Data set mengenai kanker payudara yang digunakan menggunakan data valid dari pasien di wilayah *Wisconsin*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah:

1. Penelitian ini diharapkan dapat menjadi sebuah panduan bagi para pengembang dalam mengembangkan sebuah sistem untuk mendeteksi kanker payudara yang lebih akurat.
2. Dapat menghasilkan sebuah hasil analisa yang efektif dan efisien dalam mendeteksi kanker payudara.

1.6 Sistematika Penulisan

Untuk mempermudah dalam penyusunan Tugas Akhir, Penulis menghasilkan penulisan sistematis dan penjelasan singkat mengenai laporan penelitian yang terdiri dari Bab I hingga Bab V. Sistematika penulisan skripsi ini antara lain adalah sebagai berikut ini:

BAB I PENDAHULUAN

Bab pertama menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan penelitian, Batasan masalah, manfaat penelitian dan sistematika penulisa.

BAB II TINJAUAN PUSTAKA

Bab kedua menjelaskan tentang landasan teori, teori-teori pendukung, dan hasil penelitian sebelumnya.

BAB III METODOLOGI PENELITIAN

Bab ketiga menjelaskan metode yang akan digunakan dalam penulisan seperti kerangka berpikir, metode pengumpulan data, dan cara melakukan analisis data.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab keempat menjelaskan tentang hasil ringkasan pada penelitian dan juga membahas hasil Analisa dari hasil penelitian data dan pembahasan mengenai data asli dan data prediksi dari data set kanker payudara.

BAB V

KESIMPULAN DAN SARAN

Bab kelima berisi kesimpulan, keterbatasan, dan saran dari hasil penelitian.

