BAB I

PENDAHULUAN

1.1. Latar Belakang

Kanker payudara merupakan salah satu penyakit kanker yang paling umum dan mematikan di seluruh dunia, terutama di kalangan wanita. Berdasarkan berbagai laporan epidemiologi terkini, insiden kanker payudara terus mengalami peningkatan secara global, sehingga menjadi tantangan serius bagi sistem kesehatan modern [1], [2]. Deteksi dini kanker payudara terbukti mampu meningkatkan peluang kesembuhan dan memperpanjang harapan hidup pasien secara signifikan, khususnya apabila diagnosis dilakukan pada tahap awal perkembangan sel kanker [14]. Oleh karena itu, pengembangan metode deteksi yang akurat dan efisien menjadi kebutuhan yang sangat penting.

Menurut World Health Organization (WHO), kanker payudara merupakan jenis kanker yang paling banyak didiagnosis pada perempuan di seluruh dunia. Pada tahun 2022 tercatat lebih dari 2,3 juta kasus baru kanker payudara dan sekitar 670.000 kematian akibat penyakit ini. Tingginya angka kejadian tersebut menjadikan kanker payudara sebagai salah satu prioritas utama dalam pengembangan teknologi diagnosis berbasis kecerdasan buatan untuk membantu tenaga medis melakukan deteksi dini secara lebih cepat dan akurat [1]. Algoritma Support Vector Machine (SVM) dipilih karena mampu menangani data berdimensi tinggi dan menghasilkan pemisahan kelas yang optimal dengan akurasi tinggi pada klasifikasi medis. Sementara itu, Random Forest dipilih karena mampu mengelola data kompleks, mengurangi overfitting, dan menghasilkan prediksi yang stabil melalui kombinasi banyak decision tree.

Secara konvensional, diagnosis kanker payudara dilakukan melalui pemeriksaan medis seperti mammografi dan biopsi jaringan. Meskipun metode tersebut masih menjadi standar klinis, pendekatan ini memiliki sejumlah keterbatasan, antara lain biaya yang relatif tinggi, sifatnya yang invasif, serta ketergantungan pada interpretasi tenaga medis yang berpotensi menimbulkan kesalahan diagnosis [1], [11]. Kondisi ini mendorong pemanfaatan teknologi kecerdasan buatan, khususnya machine learning, sebagai alternatif yang mampu meningkatkan akurasi dan efisiensi dalam proses deteksi kanker payudara berbasis data klinis.

Salah satu dataset yang paling banyak digunakan dalam penelitian prediksi kanker payudara adalah Breast Cancer Wisconsin (Diagnostic). Dataset ini terdiri dari 569 data pasien dengan 30 fitur numerik yang merepresentasikan karakteristik inti sel payudara, serta dua kelas label yaitu jinak (benign) dan ganas (malignant) [3], [5]. Struktur data yang jelas dan terstandarisasi menjadikan dataset ini sangat ideal untuk penerapan algoritma supervised learning dan telah digunakan secara luas dalam berbagai studi komparatif metode machine learning di bidang medis [14].

Seiring perkembangan teknologi informasi, machine learning telah banyak diterapkan dalam bidang kesehatan untuk membantu proses diagnosis penyakit. Algoritma machine learning mampu mempelajari pola dari data medis dan menghasilkan model prediksi yang dapat digunakan sebagai sistem pendukung keputusan. Berbagai penelitian menunjukkan bahwa penerapan machine learning pada diagnosis kanker payudara mampu meningkatkan

tingkat akurasi dan mengurangi potensi kesalahan diagnosis yang disebabkan oleh faktor subjektivitas manusia[11], [12], [15].

Support Vector Machine (SVM) merupakan salah satu algoritma pembelajaran mesin yang dikenal memiliki performa tinggi dalam menangani data berdimensi tinggi dan kompleks. Algoritma ini bekerja dengan membentuk hyperplane optimal yang memisahkan data ke dalam dua kelas secara maksimal[6][10]. Dalam konteks prediksi kanker payudara, SVM sering dilaporkan menghasilkan tingkat akurasi yang sangat tinggi dibandingkan metode lain seperti K-Nearest Neighbor (KNN), Logistic Regression, maupun Decision Tree [2], [6], [10]. Beberapa penelitian menunjukkan bahwa SVM mampu mencapai akurasi di atas 97% pada dataset Breast Cancer Wisconsin setelah dilakukan proses normalisasi dan pemilihan kernel yang tepat [11], [13].

Di sisi lain, Random Forest merupakan metode ensemble learning berbasis pohon keputusan yang mengombinasikan banyak decision tree untuk menghasilkan prediksi akhir melalui mekanisme voting mayoritas [7], [10]. Keunggulan utama Random Forest terletak pada kemampuannya menangani data kompleks, mengurangi risiko overfitting, serta memberikan performa yang stabil pada berbagai variasi data [5], [7], [12]. Studi terkini menunjukkan bahwa Random Forest mampu memberikan hasil klasifikasi yang kompetitif dan dalam beberapa kasus bahkan melampaui performa SVM, terutama ketika dikombinasikan dengan teknik feature selection yang tepat [12], [13], [15].

Meskipun kedua algoritma tersebut sama-sama menunjukkan performa yang sangat baik, hasil penelitian sebelumnya memperlihatkan adanya variasi kinerja antara SVM dan Random Forest ketika diterapkan pada dataset yang sama. Beberapa penelitian melaporkan bahwa SVM unggul dari sisi akurasi, sedangkan penelitian lain menemukan bahwa Random Forest lebih stabil dan robust terhadap variasi data serta noise [3], [7], [8]. Perbedaan hasil ini menunjukkan bahwa pemilihan algoritma terbaik sangat bergantung pada karakteristik data, teknik pra-pemrosesan, serta parameter model yang digunakan [13], [14], [15].

Penelitian ini bertujuan untuk membandingkan secara empiris kinerja algoritma Support Vector Machine dan Random Forest dalam memprediksi kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic). Proses pengolahan data dan eksperimen dilakukan menggunakan Google Colab. Hasil penelitian diharapkan dapat memberikan rekomendasi metode yang paling akurat dan efektif, serta menjadi referensi bagi pengembangan sistem pendukung keputusan klinis berbasis machine learning dalam deteksi dini kanker payudara.

Selain tingkat akurasi, pemilihan algoritma machine learning dalam bidang kesehatan juga perlu mempertimbangkan kemampuan model dalam mendeteksi kasus positif secara tepat. Dalam konteks kanker payudara, kesalahan klasifikasi berupa False Negative dapat menyebabkan keterlambatan diagnosis dan penanganan pasien, sehingga berpotensi meningkatkan risiko komplikasi maupun kematian [1], [18]. Oleh karena itu, evaluasi model tidak hanya dilakukan berdasarkan akurasi, tetapi juga mempertimbangkan sensitivitas, spesifisitas, dan Area Under Curve (AUC) agar diperoleh gambaran yang lebih komprehensif mengenai kualitas prediksi yang dihasilkan oleh masing-masing algoritma.

Berdasarkan berbagai penelitian sebelumnya, belum terdapat kesimpulan yang konsisten mengenai algoritma yang paling optimal untuk klasifikasi kanker payudara. Beberapa penelitian menunjukkan bahwa Support Vector Machine memiliki performa yang lebih baik pada dataset berdimensi tinggi, sedangkan penelitian lain menunjukkan bahwa Random Forest lebih unggul dalam menangani variasi data dan proses seleksi fitur [11], [12], [13], [15]. Oleh karena itu diperlukan penelitian komparatif untuk mengetahui metode yang memberikan performa terbaik pada dataset Breast Cancer Wisconsin (Diagnostic).

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa algoritma Support Vector Machine (SVM) dalam memprediksi kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic)?
2. Bagaimana performa algoritma Random Forest dalam memprediksi kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic)?
3. Bagaimana perbandingan performa SVM dan Random Forest berdasarkan metrik akurasi, presisi, recall, F1-Score, dan AUC?
4. Algoritma manakah yang memberikan performa terbaik dalam prediksi kanker payudara?

1.3. Batasan Masalah

Keterbatasan dalam penelitian ini agar tetap fokus dan terarah adalah sebagai berikut:

1. Penelitian ini hanya membahas klasifikasi kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic).
2. Dataset yang digunakan berasal dari UCI Machine Learning Repository dan diakses melalui platform kaggle.
3. Algoritma yang dibandingkan hanya Support Vector Machine (SVM) dan Random Forest (RF).
4. Evaluasi performa model menggunakan metrik akurasi, presisi, recall, F1-Score, dan AUC tanpa membahas kompleksitas komputasi.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis performa algoritma Support Vector Machine (SVM) dalam prediksi kanker payudara.
2. Menganalisis performa algoritma Support Vector Machine (SVM) dalam prediksi kanker payudara.
3. Membandingkan performa kedua algoritma berdasarkan akurasi, presisi, recall, F1-Score, dan AUC.
4. Menentukan algoritma dengan performa terbaik untuk prediksi kanker payudara.

1.5. Manfaat Penelitian

1.5.1 Bagi Peneliti

Penelitian ini memberikan pemahaman empiris mengenai kelebihan dan keterbatasan algoritma Support Vector Machine (SVM) dan Random Forest dalam memprediksi kanker payudara. Hasil penelitian ini juga dapat dijadikan sebagai referensi dan dasar bagi penelitian selanjutnya yang menggunakan algoritma machine learning lain atau pengembangan model prediksi pada bidang kesehatan.

1.5.2 Bagi Pasien dan Masyarakat

Penelitian ini memberikan wawasan mengenai pentingnya deteksi dini kanker payudara serta pemanfaatan teknologi machine learning dalam membantu proses diagnosis. Informasi yang dihasilkan diharapkan dapat meningkatkan kesadaran masyarakat terhadap upaya pencegahan dan penanganan kanker payudara secara lebih cepat dan akurat.

1.5.3 Bagi Tenaga Medis (Dokter dan Tenaga Kesehatan)

Tenaga medis dapat memanfaatkan hasil analisis prediktif dari penelitian ini sebagai sistem pendukung keputusan dalam membantu proses diagnosis kanker payudara. Model prediksi yang dihasilkan dapat membantu dokter dalam mengidentifikasi pasien dengan potensi kanker ganas sehingga keputusan medis dapat dilakukan secara lebih tepat dan efisien.

1.5.4 Bagi Institusi Kesehatan dan Pengembang Sistem

Hasil penelitian ini dapat digunakan oleh institusi kesehatan dan pengembang sistem sebagai dasar dalam pengembangan aplikasi diagnosis berbasis data dan kecerdasan buatan. Selain itu, penelitian ini dapat mendukung penerapan teknologi kesehatan digital (health informatics) guna meningkatkan kualitas pelayanan kesehatan dan efisiensi sistem diagnosis kanker payudara.