

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker merupakan salah satu penyakit paling mematikan di dunia dan menjadi penyebab utama kematian di banyak negara. Berdasarkan data dari *World Health Organization* (WHO), diperkirakan terdapat 19,3 juta kasus kanker dan kisaran 10 juta kematian baru yang disebabkan oleh kanker secara keseluruhan di dunia pada tahun 2020. Angka ini diperkirakan akan terus meningkat seiring dengan pertumbuhan populasi global dan peningkatan usia harapan hidup. Jenis kanker yang sering ditemukan dan menjadi perhatian utama adalah kanker payudara, kanker paru-paru, kanker usus besar, dan kanker prostat. Deteksi dini dan diagnosis yang akurat sangat penting untuk meningkatkan peluang kesembuhan dan mengurangi angka kematian akibat kanker. Oleh karena itu, pengembangan metode yang efektif dan efisien untuk diagnosis kanker menjadi salah satu prioritas utama dalam penelitian medis.

Diagnosis kanker secara tradisional dilakukan melalui pemeriksaan fisik, pencitraan medis, dan biopsi. Prosedur ini biasanya melibatkan pemeriksaan histopatologi di mana jaringan yang dicurigai diambil dengan cara diperiksa menggunakan mikroskop oleh ahli patologi. Meskipun metode ini cukup akurat, mereka memiliki beberapa keterbatasan. Proses *biopsi* dapat menjadi *invasif* dan menyakitkan bagi pasien, analisis *histopatologi* digunakan dengan membutuhkan waktu yang cukup relatif lama, serta sangat bergantung pada keahlian dan pengalaman ahli patologi. Selain itu, interpretasi hasil bisa bervariasi antara satu ahli patologi dengan lainnya, yang dapat mengakibatkan ketidakpastian dalam diagnosis.

Dalam dekade terakhir, perkembangan teknologi komputasi dan kemajuan dalam bidang kecerdasan buatan (AI) dan *machine learning* (ML) telah membuka peluang baru untuk meningkatkan diagnosis kanker. Algoritma *machine learning* dapat digunakan untuk mengidentifikasi pola-pola kompleks dalam data medis yang mungkin tidak dapat dideteksi oleh manusia. Salah satu algoritma yang populer dan efektif dalam tugas klasifikasi adalah *Support Vector Machine* (SVM).

SVM dikenal karena kemampuannya untuk menangani data berdimensi tinggi dan menghasilkan model klasifikasi yang akurat. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas yang berbeda dengan margin maksimum.

Dataset Classification Cell Samples yang dikembangkan oleh *Asuncion dan Newman (2007)* merupakan *dataset* yang paling sering digunakan untuk pendekatan penelitian kanker. *Dataset* terdiri sebanyak 699 sampel sel kanker yang diambil melalui prosedur aspirasi jarum halus (*fine needle aspiration*) pada massa payudara. Setiap sampel dalam *dataset* ini dicirikan oleh sepuluh fitur yang relevan dengan sifat-sifat morfologis sel kanker, yaitu ketebalan benjolan (*clump thickness*), uniformitas ukuran sel (*uniformity of cell size*), uniformitas bentuk sel (*uniformity of cell shape*), adhesi marginal (*marginal adhesion*), ukuran sel epitel tunggal (*single epithelial cell size*), inti telanjang (*bare nuclei*), kromatin lembut (*bland chromatin*), nukleolus normal (*normal nucleoli*), dan mitosis (*mitoses*). Selain itu, setiap sampel diklasifikasikan sebagai jinak atau ganas, memberikan dasar yang kuat untuk pengembangan dan evaluasi model klasifikasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, penelitian ini terdapat beberapa rumusan masalah sebagai berikut:

1. Seberapa efektif algoritma SVM dalam mengklasifikasikan sel kanker menggunakan data dari penelitian *Asuncion dan Newman (2007)*?
2. Seberapa akurat prediksi yang dihasilkan oleh algoritma SVM pada *dataset* tersebut?
3. Faktor-faktor apa saja yang mempengaruhi kinerja algoritma SVM dalam klasifikasi sel kanker?

1.3 Tujuan

1. Mengaplikasikan algoritma SVM untuk klasifikasi sel kanker.
2. Menilai kinerja algoritma SVM dengan mengukur akurasi, presisi, *recall*, dan *F1-score*.

3. Mengidentifikasi faktor-faktor yang mempengaruhi kinerja SVM dalam klasifikasi sel kanker, seperti pemilihan kernel dan pengaturan *hyperparameter*.

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan, antara lain:

1. Jenis Data : Penelitian ini menggunakan data dari penelitian *Asuncion dan Newman (2007)*, sehingga hasilnya mungkin tidak dapat digeneralisasikan untuk semua jenis data kanker.
2. Metode Klasifikasi : Algoritma SVM digunakan untuk klasifikasi tanpa membandingkannya secara mendalam dengan algoritma lain seperti *Random Forest, Neural Networks, atau k-Nearest Neighbors*.
3. Parameter SVM : Penelitian ini tidak mengeksplorasi semua kemungkinan parameter dalam SVM, hanya beberapa parameter utama seperti jenis kernel dan parameter regulasi yang diuji.
4. Preprocessing Data : Data yang digunakan mungkin memerlukan preprocessing tertentu seperti normalisasi atau penanganan nilai yang hilang, yang dapat mempengaruhi hasil klasifikasi.

1.5 Manfaat

Manfaat dari Penelitian ini diharapkan memberikan manfaat diantaranya ?

1. Manfaat Teoretis : Menambah wawasan dan pengetahuan tentang penerapan algoritma SVM dalam bidang medis, khususnya dalam klasifikasi data kanker yang kompleks. Penelitian ini juga dapat memperkaya literatur terkait penggunaan *machine learning* dalam diagnosis medis.
2. Manfaat Praktis : Menyediakan metode yang efektif untuk klasifikasi sel kanker yang dapat membantu dalam proses diagnosis medis. Diharapkan hal ini dapat meningkatkan akurasi dan efisiensi dalam deteksi dini kanker, sehingga membantu dokter dalam merencanakan pengobatan yang tepat.

3. Manfaat Kebijakan : penelitian ini dapat dijadikan dasar bagi pengambilan kebijakan dalam mengembangkan sistem diagnosis kanker yang lebih canggih dan berbasis teknologi, guna meningkatkan layanan kesehatan.

