

BAB I PENDAHULUAN

1.1 Latar Belakang

Kanker paru-paru merupakan salah satu penyakit yang paling mematikan di dunia, menurut WHO angka kematian pada pengidap kanker paru-paru mencapai 1.796.144 pada tahun 2020 maka deteksi dini terhadap penyakit ini sangat krusial dalam meningkatkan peluang kesembuhan pasien. Namun, dalam praktiknya, proses diagnosis masih menghadapi banyak tantangan, seperti keterbatasan alat, waktu diagnosis yang lama, serta subjektivitas dalam interpretasi hasil medis oleh tenaga profesional [1]. Seiring berkembangnya teknologi informasi dan komputasi, penerapan algoritma *machine learning* menjadi alternatif yang menjanjikan untuk membantu proses klasifikasi dan diagnosis penyakit secara otomatis, akurat, dan efisien [2]. Dua algoritma yang telah terbukti efektif dalam tugas klasifikasi penyakit berdasarkan berbagai studi sebelumnya adalah Support Vector Machine (SVM) [4] dan Extreme Gradient Boosting (XGBoost) [5], yang masing-masing menunjukkan performa tinggi pada berbagai kasus seperti klasifikasi kanker paru-paru, stroke, penyakit paru, hingga deteksi titik panas kebakaran hutan dan lahan.

Dalam konteks klasifikasi kanker paru-paru, SVM menghasilkan akurasi sebesar 62,3% pada pembagian data 80:20 ketika dikombinasikan dengan metode feature selection [1]. Pada kasus klasifikasi penyakit stroke, kernel polynomial dari SVM menunjukkan akurasi hingga 80% pada data tidak seimbang [2], dan akurasi 78,86% pada rasio data tertentu [3]. SVM juga terbukti fleksibel dan mampu memberikan hasil akurat dalam domain non-medis, seperti klasifikasi judul buku dalam perpustakaan digital, dengan akurasi mencapai 69,24% menggunakan kernel linear [4].

Sementara itu, algoritma XGBoost menunjukkan performa unggul dalam menangani data klasifikasi berskala besar dan kompleks. Pada klasifikasi titik panas kebakaran hutan, XGBoost yang dikombinasikan dengan metode feature

importance mampu mencapai akurasi hingga 89,52%, melampaui performa algoritma lain seperti Decision Tree, Logistic Regression, dan SVM [5]. Keunggulan utama XGBoost terletak pada kemampuannya dalam menangani outlier, missing value, serta efisiensi proses pelatihan dan penghindaran overfitting.

Meskipun berbagai studi telah menggunakan algoritma SVM maupun XGBoost, sebagian besar penelitian masih dilakukan secara terpisah dan belum melakukan perbandingan langsung antara kedua algoritma tersebut pada konteks dan dataset yang sama [6]. Selain itu, sebagian besar studi masih terbatas pada evaluasi akurasi tanpa mempertimbangkan metrik penting lainnya seperti precision, recall, dan f1-score [7], yang sebenarnya sangat krusial dalam konteks diagnosis penyakit yang sensitif terhadap kesalahan klasifikasi. Oleh karena itu, dibutuhkan penelitian yang secara langsung membandingkan performa algoritma SVM dan XGBoost dalam klasifikasi penyakit kanker paru-paru, menggunakan dataset yang relevan dan evaluasi metrik yang komprehensif, guna menghasilkan sistem klasifikasi yang lebih andal untuk mendukung proses diagnosis klinis.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Berapa akurasi masing-masing algoritma SVM dan XGBoost dalam mengklasifikasikan data medis pasien kanker paru-paru berdasarkan parameter polusi udara, penggunaan minuman beralkohol dan fatigue?
2. Bagaimana hasil perbandingan pengujian algoritma SVM dan XGBoost dalam mengklasifikasikan data medis pasien kanker paru-paru?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah yang ditetapkan untuk menjaga fokus dan kejelasan penelitian. Batasan-batasan tersebut adalah sebagai berikut:

1. Penelitian ini terbatas pada data medis yang diperoleh dari Website Kaggle. <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link> yang di upload di website Kaggle 3 tahun yang lalu
2. Parameter yang digunakan pada penelitian ini hanya berdasarkan polusi udara, penggunaan minuman beralkohol dan fatigue
3. Variabel utama yang diteliti adalah akurasi, presisi, recall, dan F1-score dari masing-masing algoritma dalam mengklasifikasikan penyakit kanker paru-paru. Variabel lain yang tidak terkait langsung dengan performa algoritma akan diabaikan.

Batasan-batasan ini diharapkan dapat menjadi fokus pada analisis komparatif yang mendalam terhadap dua algoritma yang dipilih, serta memberikan hasil yang akurat dan relevan dalam konteks klasifikasi penyakit kanker paru-paru.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah:

1. Menghasilkan analisis komparatif mengenai performa algoritma Support Vector Machine (SVM) dan XGBoost dalam mengklasifikasikan penyakit kanker paru-paru berdasarkan akurasi, presisi, recall, dan F-1 Score.
2. Mengetahui tingkat akurasi, presisi, recall, dan F1-score dari masing-masing algoritma (SVM dan XGBoost) dalam klasifikasi data medis pasien kanker paru-paru.

Tujuan ini menjadi bukti upaya peneliti memberikan rekomendasi berbasis bukti mengenai algoritma yang paling efektif dan akurat untuk digunakan dalam klasifikasi penyakit kanker paru-paru, sehingga dapat meningkatkan akurasi diagnosis dan efisiensi pengobatan.

1.5 Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat yang dapat dirasakan oleh berbagai pihak, baik secara teoritis maupun praktis, serta bagi objek penelitian dan peneliti selanjutnya:

Manfaat Teoritis :

1. Pengembangan Ilmu Pengetahuan:

Penelitian ini dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang machine learning dan data mining. Hasil penelitian ini diharapkan dapat menambah wawasan dan pengetahuan mengenai efektivitas algoritma SVM dan XGBoost dalam klasifikasi penyakit kanker paru-paru.

Penelitian ini juga dapat menjadi referensi bagi peneliti lain yang ingin mengembangkan metode klasifikasi dalam bidang kesehatan atau penyakit lainnya.

Manfaat Praktis :

1. Manfaat bagi Objek Penelitian (Institusi Kesehatan dan Pasien) :

Akurasi Diagnosa : Dengan aplikasi hasil penelitian ini, institusi kesehatan dapat meningkatkan akurasi diagnosa penyakit kanker paru-paru. Algoritma yang paling efektif dari hasil penelitian ini dapat digunakan untuk membantu dokter dalam mendiagnosa pasien dengan lebih tepat dan cepat.

Efisiensi Proses Diagnosa : Penggunaan algoritma yang teruji dalam penelitian ini dapat mempercepat proses analisis data medis, sehingga waktu yang dibutuhkan

untuk diagnosa bisa lebih singkat, yang pada akhirnya bisa menyelamatkan lebih banyak nyawa.

Pengurangan Biaya Medis : Dengan diagnosa yang lebih cepat dan akurat, biaya yang dikeluarkan oleh pasien dan institusi kesehatan dapat diminimalkan. Hal ini disebabkan karena pengurangan kebutuhan akan tes lanjutan atau prosedur diagnostik yang mahal.

2. Manfaat bagi Pengguna atau Organisasi (Rumah Sakit, Klinik, dan Peneliti Selanjutnya) :

Pengembangan Aplikasi Medis : Rumah sakit dan klinik dapat mengembangkan aplikasi medis berbasis machine learning yang memanfaatkan algoritma yang terbukti efektif dari penelitian ini. Hal ini dapat meningkatkan kualitas layanan kesehatan yang diberikan kepada pasien.

Pelatihan dan Pendidikan : Hasil penelitian ini dapat digunakan sebagai bahan pelatihan dan pendidikan bagi tenaga medis, khususnya dalam bidang radiologi dan onkologi, untuk lebih memahami penggunaan teknologi machine learning dalam praktek medis.

Penelitian Lanjutan : Peneliti selanjutnya dapat menggunakan hasil penelitian ini sebagai dasar untuk penelitian lebih lanjut, misalnya untuk mengembangkan metode atau algoritma baru yang lebih efisien atau untuk mengaplikasikan metode yang ada pada jenis penyakit lain.

Penelitian ini tidak hanya memberikan manfaat secara langsung kepada objek dan pengguna akhir, tetapi juga berkontribusi pada pengembangan ilmu pengetahuan dan praktik medis secara keseluruhan.

1.6 Sistematika Penulisan

Untuk memahami lebih jelas tugas skripsi ini, dilakukan dengan cara mengelompokkan materi menjadi beberapa sub bab dengan sistematika penulisan sebagai berikut ini:

BAB I PENDAHULUAN

Bab Pendahuluan memuat latar belakang, rumusan masalah, tujuan penelitian, batasan masalah dan manfaat penelitian yang dibagi menjadi manfaat teoritis dan manfaat praktis.

BAB II TINJAUAN PUSTAKA

Bab ini berisikan tinjauan pustaka yang digunakan peneliti sebagai referensi dan rujukan dari penelitian sebelumnya. Dalam bab ini juga memuat mengenai dasar-dasar teori yang digunakan dalam menjelaskan algoritma yang digunakan dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini berisikan tentang metode yang digunakan dalam penelitian, sumber data penelitian, pemrosesan data penelitian serta menjelaskan alur penelitian dan solusi yang diberikan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menampilkan ringkasan eksperimen mulai dari eksplorasi dan preprocessing data hingga pembangunan serta evaluasi model SVM dan XGBoost, disertai analisis metrik performa dan diskusi implikasi hasil.

BAB V PENUTUP

Bab ini merangkum temuan utama, menegaskan keunggulan XGBoost atas SVM dalam klasifikasi tingkat keparahan kanker paru-paru, serta mengemukakan rekomendasi pengembangan lanjutan dan penerapan praktis.