

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Teknologi *machine learning* merupakan teknologi yang bertujuan untuk mengembangkan mesin agar dapat belajar dengan sendirinya tanpa arahan dari penggunanya. *Machine learning* dikembangkan berdasarkan beberapa ilmu seperti statistika, matematika dan *data mining* sehingga mesin dapat belajar menganalisa suatu data tanpa harus di program secara spesifik. *Machine learning* memiliki dua pendekatan yaitu *supervised learning* dan *unsupervised learning* yang keduanya dapat menghasilkan sebuah mesin cerdas yang akurat. *Supervised learning* adalah metode dalam *machine learning* yang digunakan untuk menggali pemahaman, pola, dan keterkaitan dari sejumlah data pelatihan yang telah diberi label. Data pelatihan berlabel ini merupakan kumpulan data yang memiliki nilai kebenaran yang diketahui, yang nantinya akan menjadi target variabel.

Jenis perawatan di rumah sakit dapat dibagi menjadi rawat inap dan rawat jalan. Sebagai bagian dari sistem otomasi rumah sakit, dibutuhkan sebuah model klasifikasi untuk memprediksi jenis perawatan pasien untuk mempermudah dan mempercepat pengambilan keputusan. Teknologi *machine learning* dapat menyajikan model kategorisasi perawatan pasien menggunakan beberapa algoritma machine learning. *Dataset* yang digunakan dikumpulkan dari Rumah Sakit swasta di Indonesia.

Pada dasarnya, semua algoritma *machine learning* dapat diaplikasikan di segala sektor, termasuk sektor kesehatan. Beberapa manfaat *machine learning* di sektor kesehatan adalah digunakan untuk mengklasifikasikan pasien ke dalam berbagai kelompok tertentu, memprediksi dan mengobati penyakit, diagnosis penyakit, menemukan dan mengembangkan obat baru, dan mengelola rekam medis. Prediksi berdasarkan data rekam kesehatan digital saat ini semakin dibutuhkan. Banyak bentuk prakiraan dan deteksi terkait kesehatan yang dapat diimplementasikan menggunakan metode *machine learning*. Model prediktif menggunakan EHR (*Electronic Health*

Record) masih jarang ditemukan karena kesulitan dalam mengumpulkan dan mengungkapkan data pasien.

Electronic Health Record menyertakan informasi penting tentang sebuah penyakit pasien dan pengobatannya, yang seringkali berisi data terstruktur dan tidak terstruktur. Salah satu contoh *Electronic Health Record* adalah data hasil tes laboratorium darah pasien. Kemajuan teknologi *machine learning* menjadi dorongan bagi peneliti untuk mendapatkan model prediksi dengan hasil yang lebih akurat dengan menggabungkan atau mengkombinasikan beberapa algoritma *machine learning*.

Terdapat banyak algoritma yang dapat digunakan untuk klasifikasi. *Ensemble* merupakan salah satu pendekatan pada *supervised learning* yang dapat menangani permasalahan pada *machine learning*. *Ensemble learning* adalah suatu pendekatan dalam *machine learning* di mana beberapa algoritma *machine learning* digabungkan untuk menghasilkan prediksi yang lebih akurat. Terdapat beberapa jenis *ensemble learning*, termasuk *bagging*, *boosting*, *stacking*, dan *voting*. Dalam penelitian ini, peneliti menggunakan dua jenis *ensemble learning*, yaitu *voting* dan *stacking*. *Voting* menggabungkan prediksi dari beberapa model untuk memilih hasil akhir berdasarkan mayoritas suara, sedangkan *stacking* menggabungkan prediksi dari beberapa model dengan menggunakan model *meta-learner* untuk menghasilkan prediksi yang lebih baik dimana algoritma yang akan dikombinasikan adalah *Logistic Regression*, *K-Nearest Neighbors*, *Support Vector Machine (SVM)*, *Decision Tree*, dan *Naive Bayes*.

Pada penelitian kali ini, peneliti ingin memprediksi jenis perawatan pada pasien di rumah sakit berdasarkan hasil tes laboratorium darah pasien. Peneliti ingin membuat sebuah model *ensemble learning* dan mencari tahu akurasi atau performa model apabila digunakan pada dataset *Electronic Health Record Predicting*.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang masalah, maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana cara membuat model klasifikasi yang akurat untuk memprediksi jenis perawatan pada pasien di Rumah Sakit?

2. Bagaimana cara mendapatkan parameter terbaik yang dihasilkan pada model *ensemble learning*?
3. Seberapa akurat algoritma *ensemble learning* untuk memprediksi jenis perawatan pada pasien di rumah sakit?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini berdasarkan rumusan masalah sebagai berikut:

1. Mendapatkan model klasifikasi yang akurat untuk memprediksi jenis perawatan pada pasien di Rumah Sakit berdasarkan hasil tes laboratorium darah pasien.
2. Mendapatkan parameter terbaik yang dihasilkan dari model *ensemble learning* untuk menghasilkan nilai akurasi yang tinggi.
3. Mendapatkan nilai akurasi dari penerapan *ensemble learning* dalam jenis perawatan pada pasien di rumah sakit berdasarkan data hasil tes laboratorium darah pasien.

1.4 Manfaat Penelitian

Adapun manfaat penelitian yang ingin dicapai antara lain :

- a. Bagi Peneliti
Mendapatkan pengetahuan dan pengalaman dalam pembuatan model *machine learning* dengan menggunakan algoritma *ensemble learning*.
- b. Bagi Pembaca
Mendapatkan pengetahuan terhadap model *machine learning* dengan menggunakan algoritma *ensemble learning* pada *dataset Electronic Health Record Predicting*.
- c. Bagi Institusi Kesehatan
Mendapatkan model klasifikasi yang akurat untuk memprediksi jenis perawatan pada pasien di Rumah Sakit menggunakan algoritma *ensemble learning*.
- d. Bagi Penelitian Selanjutnya

Dapat dijadikan referensi bagi peneliti yang melakukan penelitian dengan topik atau tema yang serupa.

1.5 Ruang Lingkup

Untuk menghindari kerancuan dalam proses penelitian, maka diberikan ruang lingkup penelitian sebagai berikut :

- a) Penelitian ini bertujuan mendapatkan model yang akurat untuk memprediksi jenis perawatan pada pasien di Rumah Sakit berdasarkan hasil tes laboratorium darah pasien.
- b) Jenis perawatan yang akan diprediksi meliputi rawat jalan atau rawat inap
- c) Penelitian ini ditujukan untuk mengukur tingkat akurasi algoritma *ensemble learning* untuk kasus klasifikasi.
- d) Data yang digunakan untuk membuat model *machine learning* adalah *dataset Electronic Health Record Predicting* yang diambil dari laman <https://www.kaggle.com/datasets/saurabhshahane/patient-treatment-classification?select=data-ori.csv> yang diakses pada tanggal 25 Oktober 2022.
- e) Penelitian ini menggunakan metode klasifikasi pada *supervised learning*.
- f) Menggunakan bahasa pemrograman python untuk memproses dan membangun model *machine learning*,

1.6 Luaran Yang Diharapkan

Luaran yang diharapkan dari penelitian ini adalah mendapatkan model yang memiliki nilai akurasi tinggi dari model *ensemble learning* serta dapat memprediksi jenis perawatan pada pasien di Rumah Sakit menggunakan data hasil tes laboratorium darah pasien.

1.7 Sistematika Penulisan

BAB I : PENDAHULUAN

Bab ini mengemukakan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, luaran yang diharapkan, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini menjelaskan tentang landasan teori yang mendukung dalam pelaksanaan penelitian serta penelitian terkait.

BAB III : METODOLOGI PENELITIAN

Bab ini menjelaskan tentang tahapan penelitian, model penelitian, waktu dan tempat penelitian, serta jadwal penelitian untuk dilaksanakan.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang data preprocessing, visualisasi data, data splitting, normalisasi data, hasil tuning parameter, hasil pengujian, perbandingan akurasi, dan model inference yang dilaksanakan

BAB V : PENUTUP

Bab ini menjelaskan tentang kesimpulan dan saran yang didapat setelah melakukan penelitian.

DAFTAR PUSTAKA

Daftar pustaka berisikan daftar bacaan yang menjadi sumber referensi yang digunakan di penelitian ini.

LAMPIRAN

Lampiran memuat informasi tambahan yang digunakan di penelitian ini.