

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit serangan jantung adalah penyakit yang mematikan. Penyakit ini terjadi secara mendadak dan parahnya mengakibatkan penderitanya kehilangan nyawa. Menurut WHO kematian akibat jantung mencapai 17,8 juta kematian atau satu dari tiga kematian di dunia setiap tahun [1]. Dan menurut *Institute for Health Metrics and Evaluation*, di Indonesia angka kematian dari jantung dari tahun 1990 hingga tahun 2019 terus mengalami kenaikan setiap tahunnya [1]. Dengan angka ini menunjukkan bahwa masyarakat harus lebih *concern* terhadap kondisi kesehatan masing - masing dalam upaya pencegahan terjadinya penyakit jantung.

Dalam data penyakit serangan jantung terdapat beberapa faktor yang mempengaruhinya. Faktor tersebut dapat dianalisis menggunakan *machine learning*. *Machine learning* adalah sistem yang dapat belajar dan membuat keputusan sendiri dengan sekali pemrograman oleh komputer berdasarkan data[2]. Secara umum, tahapan pengembangan *machine learning* yaitu eksplorasi data, preprocessing data, pembuatan model, dan evaluasi. Salah satu teknik analisa dalam *machine learning* adalah dengan klasifikasi. Klasifikasi adalah teknik pengelompokan data ke dalam suatu kelas tertentu. Menurut cara belajarnya, sistem ini dibagi menjadi 2 yaitu, *supervised learning* atau teknik belajar yang menggunakan label dan kelas nya telah ditentukan dalam sebuah dataset, lalu yang kedua adalah *unsupervised learning* atau teknik belajar dengan mengelompokkan kelas berdasarkan kemiripan karakteristik data dan kelasnya belum ditentukan[2].

Dalam metode klasifikasi, model yang digunakan ada berbagai macam seperti, *random forest*, *support vector machine*, *logistic regression*, *gradient boosted tree*. Masing - masing dari model tersebut memiliki kelebihan dan kekurangan. Dalam berbagai penelitian, model yang paling banyak digunakan

dalam analisis klinis adalah *logistic regression*[3]. Model *logistic regression* mampu melakukan analisis faktor yang signifikan dalam memprediksi risiko sehingga dapat membantu dalam pengembangan strategi pencegahan dan deteksi dini penyakit jantung[4].

Random forest adalah model yang diperkenalkan pada tahun 2001 dan merupakan model hasil pengembangan dari metode *Classification* dan *Regression Tree (CART)*[5]. Dalam jurnal model ini, dengan 12 pengujian menunjukkan hasil yang baik dengan nilai akurasi rata - rata 80% [5]. Dari kesimpulan jurnal tersebut maka model *random forest* dinilai mampu dalam melakukan analisa terhadap penyakit jantung. Selain itu, *random forest* juga memiliki kelebihan yaitu dapat meningkatkan akurasi walaupun terdapat data yang hilang, dapat mengatasi adanya *outliers*, dan efisien dalam penyimpanan data[6]. Model ini dapat bekerja secara efektif pada big data dengan adanya seleksi fitur[6].

Support Vector Machine (SVM) adalah model dalam metode klasifikasi pada *machine learning* yang dikenalkan pada tahun 1992 oleh Vapnik[7]. Model ini merupakan model klasifikasi yang efisien dalam masalah nonlinier. Kelebihan dari model ini yaitu, mampu menentukan jarak sehingga komputasi menjadi cepat. Dalam sebuah jurnal, model SVM juga dinilai mampu dengan nilai akurasi, *recall*, dan presisi diatas 90% [8].

Gradient Boosted Trees (GBT) merupakan model klasifikasi yang menangani data numerik, nominal, maupun missing data[9]. Cara kerja dari model ini adalah dengan membangun *decision tree* dan juga meningkatkan struktur pohon yang lemah dan memperbaiki kesalahan pada pohon. Kelebihan dari model ini adalah mampu mencegah terjadinya *overfitting*[9]. Dalam sebuah penelitian analisis sentimen terhadap layanan Indihome di twitter, GBT mendapatkan nilai akurasi 99.31%[9].

Untuk meningkatkan performa dari setiap model dapat dilakukan metode optimalisasi. Salah satu metode optimalisasi yang dapat digunakan dalam

metode klasifikasi *machine learning* adalah *hyperparameter tuning*. *Hyperparameter tuning* adalah sebuah metode optimalisasi yang bekerja dengan penyesuaian parameter agar performa model dapat meningkat. Dalam jurnal, *hyperparameter tuning* digunakan dalam model *XBoost* terhadap penyakit *jantung* dapat meningkatkan nilai *AUC* sebesar 0.1% dengan random search sedangkan saat menggunakan grid search, performa dapat naik sebesar 0.2%.

Berdasarkan masalah dan pembahasan diatas, maka tujuan perbandingan *machine learning* dengan 4 model *machine learning* yaitu *random forest*, *SVM*, *logistic regression* dan *GBT* menggunakan metode optimalisasi *hyperparameter tuning* terhadap penyakit serangan jantung adalah untuk mengidentifikasi model *machine learning* mana yang paling unggul, menentukan faktor apa saja yang berpengaruh signifikan terhadap penyakit serangan jantung serta bagaimana *hyperparameter tuning* mempengaruhi performa dari setiap model *machine learning*.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan sebuah masalah yang dapat disajikan dalam list berikut.

1. Faktor - faktor apa saja yang signifikan dalam mempengaruhi analisis serangan data serangan jantung?
2. Bagaimana performa dari masing - masing model dalam melakukan Analisa terhadap penyakit *jantung*?
3. Bagaimana *Hyperparameter tuning* dapat meningkatkan hasil analisa di tiap model?

1.3 Batasan Masalah

Dari rumusan masalah diatas, agar penelitian lebih fokus dalam mencapai tujuan, maka perlu dilakukan pembatasan masalah. Berikut ini beberapa pembatasan masalah dalam penelitian ini :

1. Penelitian ini berfokus untuk melakukan perbandingan beberapa model machine learning dalam analisis data penyakit serangan jantung.
2. Penelitian ini menggunakan dataset publik yang bersumber dari platform online Kaggle, dengan limitasi pada jumlah fitur dan sampel yang tersedia dalam dataset.
3. Penelitian ini hanya akan menyajikan hasil analisis dalam bentuk kajian literatur atau artikel ilmiah tanpa pengembangan aplikasi atau implementasi lebih lanjut.
4. Model evaluasi dari penelitian ini berfokus pada *confusion matrix* dengan beberapa variabel perbandingan hasil yaitu akurasi, presisi, *FI score* dan *recall*.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisa perbandingan model *machine learning random forest*, *support vector machine*, *logistic regression*, dan *gradient boosted trees* dengan optimisasi *hyperparameter tuning* terhadap penyakit serangan jantung. Hasil dari perbandingan tersebut dapat menentukan model mana yang cocok dan sesuai untuk melakukan analisis terhadap penyakit serangan jantung.

1.5 Manfaat Penelitian

Secara praktis, dalam dunia teknologi kesehatan, penelitian ini mampu memberikan kontribusi yang signifikan dalam pengembangan sistem deteksi dini penyakit serangan jantung. Dengan pengembangan model yang lebih akurat, sistem ini dapat membantu pemantauan kondisi kesehatan secara individu, sehingga mampu mengidentifikasi gejala maupun resiko jauh sebelum gejala klinis muncul.

Penelitian ini juga diharapkan mampu memberikan pertimbangan terhadap penelitian selanjutnya terkait pemilihan model *machine learning* yang tepat, akurat dan efisien. Dengan mempertimbangkan faktor – faktor penyebab penyakit jantung, penelitian ini diharapkan dapat membuka peluang untuk pengembangan algoritma maupun model *machine learning* yang lebih baik. Selanjutnya penelitian ini dapat

memberikan wawasan baru dalam kolaborasi antara teknologi dan kedokteran khususnya dalam peningkatan diagnosis penyakit jantung.

Secara umum, penelitian ini dapat meningkatkan kesadaran masyarakat terhadap pentingnya menjaga kesehatan. Dengan pendekatan teknologi, dapat meningkatkan ketertarikan masyarakat sehingga mampu pro aktif dalam menjaga gaya hidup, yang pada akhirnya dapat menekan angka kematian akibat jantung.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN.

BAB I ini berisi Latar belakang yang membahas alasan dan urgensi penelitian ini dalam konteks deteksi dini penyakit jantung, rumusan masalah atau pertanyaan utama yang ingin dijawab oleh peneliti, batasan masalah atau lingkup penelitian, dan manfaat penelitian terkait perbandingan model machine learning.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian literatur dan teori dasar yang mendasari penelitian, termasuk studi terkait yang relevan dan model machine learning yang akan dibandingkan.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi: **Gambaran umum objek penelitian**, Analisis masalah, Solusi yang ditawarkan, Rancangan eksperimen dan alat analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil dari perbandingan performa model machine learning, analisis performa, serta interpretasi dari hasil yang didapat.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian mengenai hasil perbandingan dan penentuan model yang cocok dalam analisis penyakit serangan jantung serta saran-

saran untuk penelitian selanjutnya.

