

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Manusia memiliki organ yang selalu bekerja untuk bertahan hidup. Jantung adalah salah satu alat vital yang paling utama dalam kehidupan manusia. Jantung tidak akan pernah berhenti bekerja walaupun manusia sedang istirahat, sehingga jika manusia mempunyai kemampuan untuk mengontrol keadaan dan denyut jantung mereka, maka manusia sendiri akan kewalahan dalam mengatur jantung mereka sendiri dan mengakibatkan datangnya penyakit jantung bahkan bisa terjadi gagal jantung [1].

Kematian yang terjadi diseluruh dunia saat ini didominasi oleh penyakit jantung. Menurut WHO, kematian global yang disebabkan penyakit jantung memiliki persentase 31%. Beberapa faktor resiko penyakit jantung yaitu diabetes, kolesterol, tekanan darah tinggi dan lainnya merupakan faktor yang sulit untuk diidentifikasi [2].

Machine learning sendiri bukanlah sebuah teknologi yang memiliki tingkat keakuratan tinggi untuk melakukan analisa data dan mendapat kesimpulan berdasarkan analisa data tersebut. Namun keakuratan yang dihasilkan cukup efektif sehingga machine learning telah terbukti membantu di bidang kesehatan [3]. Perkembangan teknologi komputer sangat berkembang khususnya pada sistem pengenalan atau pengidentifikasian yang banyak digunakan sebagai objek riset dan pengembangan. Sistem ini menjadi salah satu riset yang populer dalam hal ini untuk mengidentifikasi suatu objek [4].

Pada penelitian [5] juga terdapat meneliti tentang penyakit jantung yang menggunakan metode berbeda namun dengan dataset yang sama yaitu penyakit jantung, pada penelitian [5] mengusulkan judul hybrid machine learning model untuk memprediksi penyakit jantung dengan metode logistic regression dan random forest, peneliti tersebut bertujuan untuk memprediksi penyakit jantung menggunakan algoritma (ML) sebagai langkah preventif dini pada sistem informasi berbasis desktop. Dengan Machine Learning models yaitu hybrid model dapat meningkatkan nilai akurasi dari sebuah metode ML yang ditambahkan dengan metode ML lainnya. Nilai akurasi yang didapatkan dari hybrid model machine learning dengan menggunakan metode Random Forest dan Logistic Regression sebesar 84,48% yang meningkat sebesar 1,32%.

Selanjutnya penelitian [3] melakukan deteksi penyakit jantung menggunakan machine learning dengan algoritma logistic regression bertujuan untuk mengimplementasikan penggunaan algoritma yaitu logistic regresi, dimana algoritma tersebut memakai fungsi logistik untuk menghasilkan binary atau nol dan satu sebagai penentuan klasifikasi. Setelah eksperimen dilakukan dengan algoritma logistik regresi memberikan hasil yang memiliki keunggulan yang berbeda-beda terhadap metode lainnya berdasarkan model analisa confusion matrix. Pada data training, metode logistik regresi mempunyai nilai sensitivity yang paling tinggi yaitu 88.54% dibanding metode lainnya. Pada data testing, metode logistik regresi mempunyai nilai kekhususan yang paling tinggi yaitu 87.50% dibanding metode lainnya.

Pada penelitian [6] melakukan algoritma neural network untuk prediksi penyakit jantung bertujuan untuk memprediksi penyakit jantung menggunakan jaringan syaraf tiruan dan peningkatan kinerja algoritma jaringan syaraf tiruan diimplementasikan pada data pasien penyakit jantung. Dari hasil pengujian dengan metode pengukuran menggunakan berbasis neural network diketahui bahwa algoritma neural network menghasilkan nilai akurasi sebesar 91.45%, presisi sebesar 92.79%, recall sebesar 94.27% dan nilai AUC yang diperoleh sebesar 0.937. dengan melihat keakuratannya, jaringan saraf tiruan berbasis algoritma masuk dalam kategori kelompok sangat baik, karena nilai AUC antara 0,90 –1,00.

Pada penelitian [7] melakukan Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote yang bertujuan untuk melakukan pendekatan imputasi *mean/mode* diusulkan untuk menangani *missing value replacement*, Min-Max Normalization untuk menangani *smoothing noisy data*, K-Fold Cross Validation untuk menangani data *validation*, dan pendekatan ensemble menggunakan metode *Weighted Vote* (WV) yang dapat menyatukan kinerja tiap-tiap metode machine learning untuk mengambil keputusan klasifikasi sekaligus untuk mereduksi unbalanced class. Hasil penelitian ini menunjukkan bahwa metode yang diusulkan tersebut memberikan akurasi sebesar 85,21%, sehingga mampu meningkatkan kinerja akurasi metode-metode machine learning, selisih 7,14% dengan *Artificial Neural Network*, 2,77% dengan *Support Vector Machine*, 0,34% dengan C4.5, 2,94% dengan Naïve Bayes, dan 3,95% dengan k-Nearest Neighbor.

Secara keseluruhan, kesimpulan dari keempat penelitian ini menegaskan pentingnya pemilihan dan kombinasi metode machine learning yang tepat untuk meningkatkan akurasi prediksi penyakit jantung. hybrid models dan ensemble methods, ensemble methods berguna untuk memberikan solusi prediksi terbaik yang menunjukkan potensi besar dalam meningkatkan performa prediksi, sementara algoritma individual seperti logistic regression dan neural network juga menunjukkan hasil yang sangat baik dengan keunggulan spesifik masing-masing. Kombinasi metode dan pendekatan preprocessing yang tepat dapat secara signifikan meningkatkan keakuratan dan reliabilitas model prediksi penyakit jantung. Diperlukannya perbandingan metode yang tepat untuk mendeteksi penyakit jantung menggunakan machine learning dengan performa yang terbaik dari penelitian sebelumnya, contohnya metode k-nearest neighbor (K-NN) dan support vector machine (SVM). Penelitian yang saat ini sedang dilakukan menggunakan metode k-nearest neighbor (K-NN) dan support vector machine (SVM) untuk melakukan perbandingan kedua metode tersebut, berikut penjelasannya dari metode k-nearest neighbor (K-NN) dan support vector machine (SVM).

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang diuji. Algoritma K-NN sangatlah sederhana, bekerja berdasarkan jarak terpendek [4]. KNN akan mengelempokkan hasil perhitungan dengan data latih yang mempunyai kerabat terbanyak dalam nilai jangkauan yang ditentukan. Jarak antara data latih dan data uji dihitung menggunakan persamaan

Euclidean [8]. Algoritma Nearest Neighbor (K-NN) merupakan algoritma klasifikasi berdasarkan kedekatan jarak suatu data dengan data yang lain. Pada algoritma K-NN, data berdimensi q , jarak dari data tersebut ke data yang lain dapat dihitung. Nilai jarak inilah yang digunakan sebagai nilai kedekatan/kemiripan antara data uji dengan data latih. Nilai K pada K-NN berarti K -data terdekat dari data uji [9]. Selanjutnya penjelasan dari metode support vector machine (SVM) yang akan digunakan pada penelitian yang sedang dilakukan.

Support Vector Machine (SVM) adalah teknik seleksi algoritmik yang mencapai ambang akurasi klasifikasi tinggi dengan membandingkan seperangkat parameter standar dengan diskerta kandidat, juga dikenal sebagai kualitas yang dirasakannya. Keuntungan dari Support Vector Machines (SVM) adalah kemampuannya untuk menangani masalah yang kompleks tanpa terpengaruh oleh berbagai faktor. SVM adalah kemampuan untuk melakukan pelatihan secara efektif, and harus berbagaian dari pembelajaran teknik [10]. Berikut adalah sebuah penelitian perbandingan yang menggunakan metode k-nearest neighbor (K-NN) dan support vector machine (SVM) dengan dataset yang berbeda untuk digunakan yaitu *Heart Failure Clinical Records*.

Tujuan penelitian ini [1] adalah untuk mengomparasi akurasi dari metode support vector mahince (SVM), random forest (RF) dan k-nearest neighbors (KNN). Dalam mengomparasi ketiga metode tersebut, peneliti juga mengomparasi pembagian data latih dan data uji dengan rincian pembagiannya adalah 90% : 10%, 80% : 20%, 70% : 30%, 60% : 40% sehingga menghasilkan confusion matrix yang berguna untuk melihat precision, recall, F1 score, accuracy yang dihasilkan. Hasil

terbaik menurut akurasi dan waktu yang dibutuhkan untuk running program, diperoleh metode SVM dan RF dengan pembagian data training 90% dan testing 10%, dimana akurasi dari kedua metode tersebut adalah 97% dan running program yang dibutuhkan untuk metode SVM adalah 2.82 detik sedangkan untuk metode RF adalah 7.29 detik. Parameter yang digunakan oleh metode SVM adalah $C = 1$, $\gamma = 0.01$, $kernel = linear$. Sedangkan parameter yang digunakan oleh metode RF adalah $n_estimators = 30$ dan $random_state = 0$. Kelemahan dari penelitian ini yaitu semakin besar data yang diuji maka akurasi pada setiap metode akan semakin menurun sehingga peneliti tersebut tidak membagi dataset dengan rincian data uji lebih besar dari 40% karena data uji tidak digunakan untuk melatih dataset yang diberikan. Jika data uji yang digunakan lebih besar daripada data latih, maka tidak akan diketahui outcome dari dataset tersebut dan menyebabkan akurasi semakin menurun atau *out of sample testing*.

Berdasarkan pada penelitian sebelumnya dapat dilihat bahwa peneliti sebelumnya menggunakan metode yang berbeda-beda, dapat dilihat bahwa terdapat kurangnya hasil yang memuaskan pada penelitian tersebut dan adanya kekurangan dari masing masing peneliti sebelumnya, dari penjelasan penelitian sebelumnya peneliti ingin meningkatkan performa dan hasil akurasi dari penelitian sebelumnya yang menggunakan metode knn dan svm untuk melakukan perbandingan visualisasi pada dataset gagal jantung, maka dari itu peneliti tertarik untuk melakukan perbandingan visualisasi antar dua metode yaitu knn dan svm dengan menggunakan dataset gagal jantung.

Berdasarkan dari penjelasan yang telah dipaparkan tersebut maka peneliti akan melakukan pengujian terhadap algoritma K-Nearest Neighbour (K-NN) dan Support Vector Machine (SVM) untuk melakukan perbandingan visualisasi pada dataset gagal jantung. Dari uraian tersebut, maka tugas penelitian akhir ini akan meliputi perbandingan visualisasi data dalam prediksi penyakit gagal jantung menggunakan metode knn dan svm yang dimana data sekunder dari sumber asli.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang diatas, dapat dirumuskan sebagai berikut :

1. Bagaimana menggunakan perbandingan prediksi algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) untuk memprediksi penyakit jantung.
2. Manakah diantara metode K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) yang terbaik untuk melakukan prediksi penyakit jantung.

1.3 BATASAN MASALAH

Untuk menghindari terjadinya pembahasan diluar tema dan judul penelitian, maka masalah yang akan dijadikan acuan agar tidak mencakup bahan yang terlalu luas. Adapun batas masalahnya sebagai berikut :

1. Penelitian ini memfokuskan pada penggunaan metode K-nearest neighbor (KNN) dan Support Vector Machine (SVM) dalam melakukan prediksi terhadap penyakit jantung.
2. Model identifikasi penyakit jantung akan dikembangkan menggunakan teknik machine learning dengan implementasi algoritma K-nearest neighbor (KNN) dan Support Vector Machine (SVM) menggunakan tools google collaboration.

1.4 TUJUAN DAN MANFAAT PENELITIAN

Adapun tujuan dan manfaat pada penelitian ini adalah sebagai berikut :

1.4.1 Tujuan Penelitian

Adapun tujuan pada penelitian ini yaitu :

1. Menerapkan visualisasi berbentuk gambar dan grafik terhadap penyakit jantung dengan menggunakan algoritma K-nearest neighbor dan Support Vector Machine (SVM) yang merupakan tujuan utama penelitian ini.
2. Mengukur akurasi menggunakan algoritma K-nearest neighbor dan Support Vector Machine (SVM) untuk prediksi penyakit jantung.

1.4.2 Manfaat Penelitian

Adapun manfaat pada penelitian ini yaitu :

1. Dapat membuktikan bahwa algoritma K-nearest neighbor (KNN) dan Support Vector Machine (SVM) dapat melakukan prediksi penyakit jantung.
2. Untuk memberikan wawasan kepada peneliti selanjutnya agar dapat menggunakan algoritma K-nearest neighbor (KNN) dan Support Vector Machine (SVM) dengan mudah.

1.5 SISTEMATIKA PENULISAN

Sistematika penulisan penelitian ini dimaksudkan untuk mempermudah pemahaman terhadap hasil penelitian. Adapun garis besar, penulisan laporan penelitian ini dibagi menjadi lima bab, yaitu :

BAB I : PENDAHULUAN

Pada bab ini memberikan penjelasan tentang latar belakang, definisi masalah, batasan masalah, tujuan penelitian, dan sistematika penulis.

BAB II : TINJAUAN PUSTAKA

Ide-ide penting yang mendasari penelitian serta hipotesis terkait lainnya akan diuraikan dalam bab ini.

BAB III : METODOLOGI PENELITIAN

Metodologi penelitian, yang merinci proses melakukan penelitian, akan disajikan pada bab ketiga.

BAB IV : ANALISA DAN PERANCANGAN

Kajian tentang prosedur pengumpulan data untuk perbandingan visualisasi pada algoritma KNN dan SVM dibahas dalam bab ini.

BAB V : PENUTUP

Kesimpulan bab ini merangkum temuan penelitian secara keseluruhan dan menawarkan rekomendasi untuk meningkatkan standar penelitian lebih lanjut.