

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung atau penyakit *kardiovaskuler* dan pembuluh darah merupakan salah satu masalah kesehatan yang penting baik di negara maju maupun negara berkembang. Penyakit ini merupakan penyebab utama kematian di dunia setiap tahunnya. Sekitar 17,3 juta kematian disebabkan oleh penyakit *kardiovaskuler* sejak tahun 2008. Lebih dari tiga juta dari kematian ini terjadi sebelum usia enam puluh tahun. Di negara berpenghasilan tinggi, tingkat kematian dini akibat penyakit jantung berkisar antara 4 persen dan 42 persen di negara berpenghasilan rendah. Pada tahun 2030, diperkirakan jumlah kematian akibat penyakit *kardiovaskuler*, terutama penyakit jantung, akan mencapai 23,3 juta. Di Indonesia penyakit *kardiovaskuler* terus meningkat sehingga menimbulkan beban sosial ekonomi bagi keluarga pasien, masyarakat dan negara. Terjadinya penyakit jantung di Indonesia sebesar 0,5% pada tahun 2013 berdasarkan diagnosis medis. Pada saat yang sama, itu adalah 1,5%, berdasarkan diagnosis dokter terhadap gejalanya. Terjadinya gagal jantung di Indonesia pada tahun 2013 berdasarkan diagnosis dokter adalah 0,13%. (Kemenkes, 2022)

Konsumsi tembakau, aktivitas fisik yang kurang, diet yang tidak sehat, dan penggunaan alkohol berbahaya adalah penyebab utama penyakit kardiovaskular. Faktor resiko penyakit jantung terdiri dari yang dapat dimodifikasi dan yang tidak dapat dimodifikasi. Faktor yang tidak dapat dimodifikasi termasuk riwayat keluarga, umur, jenis kelamin, dan obesitas. Faktor yang dapat dimodifikasi termasuk *diabetes melitus*, hipertensi, *dislipidemia*, kurang aktivitas fisik, diet yang tidak sehat, dan *stress*. (WHO, 2016).

Untuk memprediksi penyakit jantung diperlukan beberapa tes. Kurangnya keahlian staf medis dapat menghasilkan prediksi yang salah (Pouriyeh et al., 2017). Penyakit ini menjadi sangat sulit untuk didiagnosa secara dini. Penanganan bedah penyakit jantung semakin sulit, terutama di negara berkembang, di mana diagnosis dan perawatan yang tepat bagi pasien dengan penyakit jantung tidak tersedia. Sebuah evaluasi

yang didasarkan pada prediksi yang akurat tentang resiko penyakit jantung akan sangat membantu pasien untuk menghindari serangan jantung yang parah dan meningkatkan tingkat keselamatan pasien. Salah satu metode yang efektif untuk menemukan dan memprediksi penyakit jantung adalah dengan menggunakan algoritma *Machine Learning*. *Machine learning* mampu mengatasi kerumitan dalam mendiagnosis penyakit jantung dengan model prediksi seperti *Logistic Regresion*, *KNN (K-Nearest Neighbors)*, *CNN (Convolutional Neutral Network)*, *Random Forest Clasifier*. (Ghwanmeh et al., 2013).

Random Forest termasuk bagian dari algoritma *Machine learning* digunakan untuk mengklasifikasikan set data yang sangat besar dengan menggabungkan pohon dalam pohon keputusan dengan metode *training dataset*. Algoritma ini menggunakan pohon keputusan atau *Decision Tree* untuk memulai proses seleksi, di mana pohon akan dibagi secara rekursif berdasarkan data dari kelas yang sama. Dalam hal ini, penggunaan lebih banyak pohon akan meningkatkan akurasi. (Goldstein, Poley dan Briggs, 2011).

Metode *Random Forest* telah digunakan dalam beberapa penelitian. Salah satunya adalah yang dilakukan oleh Widya Apriliah, Ilham Kurniawan, Muhamad Baydhowi dan Tri Haryati (2020). Dalam penelitiannya merancang model yang dapat menghitung kemungkinan diabetes pada pasien dengan menggunakan teknik *data mining* yang mengkategorikan data untuk diprediksi dan dianalisis dengan menggunakan *Suport Vector Machine*, *Naive Bayes*, dan *Random Forest*. Hasil dari penelitian di dapat akurasi *Random Forest* yang lebih tinggi dengan tingkat akurasi 97,88% dibandingkan dengan algoritma lainnya.

Penelitian lainnya dilakukan oleh Abdul Hafiz Muttaqien (2020). Dalam penelitiannya bertujuan untuk memperhatikan kesehatan kucing dengan memerhatikan penyakit kulit pada kucing. Oleh karena itu penelitian ini mengklasifikasi penyakit kucing khususnya pada kulit dengan metode *Random Forest* dengan klasifikasinya menggunakan tiga kelas yaitu *scabies*, *dermatophytosis*, dan sehat. Penelitian ini memiliki tingkat akurasi sebesar 96% dan penggunaan *tree* dalam jumlah besar yaitu 100.

Berdasarkan latar belakang dan penelitian sebelumnya *implementasi Machine Learning* untuk prediksi kemungkinan penyakit jantung menggunakan *Random Forest* dapat memiliki akurasi yang tinggi dalam bidang kesehatan.

1.2 Perumusan Masalah

Berdasarkan uraian dari latar belakang di atas, maka yang menjadi permasalahan dalam penelitian ini, yaitu :

1. Bagaimana mengimplementasikan *Machine Learning* dengan algoritma *Random Forest* untuk menganalisa tingkat akurasi model ?
2. Bagaimana mengimplementasikan model *Random Forest* ke *Streamlit* untuk memprediksi kemungkinan seseorang mengalami penyakit jantung ?

1.3 Tujuan Penelitian

Dengan mempertimbangkan masalah yang dihadapi, maka ditentukan tujuan yang ingin dicapai, yaitu:

1. Mengimplementasikan algoritma *Random Forest* untuk menganalisa tingkat akurasi model
2. Mengimplementasikan model *Random Forest* menggunakan *Streamlit* untuk memprediksi kemungkinan seseorang mengalami penyakit jantung.

1.4 Batasan Masalah

Untuk menghindari meluasnya permasalahan, maka diperlukan ruang lingkup, yaitu:

1. Dataset yang digunakan untuk membuat model *Machine Learning* terdiri dari data riwayat penyakit jantung dalam format *.csv* yang diambil dari *UCI Machine Learning Repository*.
2. Implementasi prediksi menggunakan dengan algoritma *Random Forest*
3. Sistem ini dirancang dengan menggunakan *python* dan *Jupyter Notebook*
4. Implementasi model *Random Forest* menggunakan *Streamlit*

1.5 State of The Art

Penyusunan tugas akhir ini mengambil beberapa referensi penelitian sebelumnya termasuk jurnal-jurnal yang berhubungan dengan penelitian ini, diantaranya :

Judul Jurnal	Pembahasan
<p>Judul Klasifikasi Penyakit Jantung Menggunakan <i>Decision Tree</i> dan <i>Random Forest</i></p> <p>Peneliti Sabrina Adnin Kamila, RR Sri Sulistijowati dan Irwan Susanto</p> <p>Lokasi Universitas Nusantara PGRI Kediri</p> <p>Tahun 2023</p>	<p><u>Hasil Penelitian</u> Berdasarkan hasil analisis, klasifikasi penyakit jantung menggunakan <i>Random Forest</i> menghasilkan nilai akurasi 81,82%, nilai <i>precision</i> 87,04%, nilai <i>recall</i> 79,82%, dan nilai <i>f1 score</i> 83,13%. Ini menunjukkan bahwa klasifikasi <i>Random Forest</i> menghasilkan hasil yang lebih baik daripada klasifikasi <i>Decision Tree</i>.</p> <p><u>Alasan Menjadi Tinjauan Penelitian</u> Pada penelitian memiliki kasus yang sama dan pada hasil perbandingan <i>Random Forest</i> menghasilkan akurasi yang lebih baik sebesar 81% dan dari penelitian akan digunakan sebagai perbandingan apakah dengan kasus yang sama dan metode yang sama akan memiliki angka akurasi yang sama atau bisa lebih besar.</p>
<p>Judul Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi <i>Random Forest</i></p> <p>Peneliti Widya Apriliah, Ilham Kurniawan,</p>	<p><u>Hasil Penelitian</u> Dalam penelitian ini, model yang dapat digunakan untuk memprakirakan kemungkinan terkena diabetes pada pasien dengan menggunakan teknik data <i>mining</i> yang mengkategorikan data untuk diprediksi dan dianalisis dengan menggunakan <i>Support Vector Machine</i>, <i>Naive Bayes</i>, dan <i>Random Forest</i>. Dataset</p>

<p>Muhamad Baydhowi dan Tri Haryati</p> <p>Lokasi Universitas Bina Insani, Bekasi</p> <p>Tahun 2020</p>	<p>yang digunakan dari rumah sakit di Bangladesh yang bersumber dari <i>UCI repository</i>. Hasil dari penelitian di dapat akurasi <i>Random Forest</i> yang lebih tinggi dengan tingkat akurasi 97,88% dibandingkan dengan algoritma lainnya.</p> <p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Pada penelitian ini menggunakan metode yang sama yaitu <i>Random Forest</i> dan sama - sama di bidang kesehatan, hanya saja beda pada kasusnya.</p>
<p>Judul Implementasi Algoritma <i>Random Forest</i> dalam Menentukan Kualitas Susu Sapi</p> <p>Peneliti I Putu Ryan Paramaditya dan Cokorda Pramatha</p> <p>Lokasi Universitas Udayana, Bali</p> <p>Tahun 2022</p>	<p><u>Hasil Penelitian</u></p> <p>Hasil analisis menunjukkan bahwa algoritma <i>Random Forest</i> memiliki tingkat akurasi sebesar 98% dalam klasifikasi. nilai <i>precision</i>, <i>recall</i>, dan <i>f1 score</i> untuk setiap variabel masing-masing memiliki nilai sebesar 98%. Selanjutnya, variabel respons dengan tingkat akurasi tertinggi, dengan <i>AUC</i> sebesar 1,00, berada dalam kategori "<i>medium</i>", meskipun perbedaan hasil akurasi dalam kategori lain tidak begitu besar. Oleh karena itu, algoritma <i>Random Forest</i> sangat cocok untuk digunakan dalam proses klasifikasi data kualitas susu sapi.</p>

	<p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Penelitian ini menggunakan metode yang sama dengan penelitian yang akan dibahas pada tugas akhir ini, hanya saja berbeda di bidang implementasinya.</p>
<p>Judul</p> <p>Komparasi Tingkat Akurasi <i>Random Forest</i> dan <i>KNN</i> untuk Mendiagnosis Penyakit Kanker Payudara</p> <p>Peneliti</p> <p>Vincent Angkasa dan Jefri Junifer Pangaribuan</p> <p>Lokasi</p> <p>Universitas Pelita Harapan, Tangerang Selatan</p> <p>Tahun</p> <p>2022</p>	<p><u>Hasil Penelitian</u></p> <p>Berdasarkan hasil penelitian dan analisis yang dilakukan, dapat disimpulkan bahwa algoritma <i>KNN</i> dan <i>Random Forest</i> yang dilatih dengan <i>library Scikit-learn</i> dapat digunakan untuk membuat model machine learning dengan tingkat akurasi yang sangat tinggi pada kasus prediksi klasifikasi, seperti pada <i>Breast Cancer Wisconsin (Diagnostic)</i> dataset. Meskipun algoritma <i>Random Forest</i> memiliki <i>starting point</i> yang lebih tinggi pada graf <i>ROC curve</i>, namun tidak menjamin bahwa kalibrasi <i>Random Forest</i> berikutnya akan menghasilkan <i>True Positive Rate</i> dan <i>False Positive Rate</i> yang lebih baik daripada <i>KNN</i>. Selain itu, meskipun algoritma <i>KNN</i> memiliki proses yang sederhana, namun mampu menghasilkan nilai <i>ROC</i> yang hampir sama dengan <i>Random Forest</i> yang memiliki proses yang lebih kompleks. Implementasi code untuk membangun model <i>machine learning</i> dengan <i>library Scikit-learn</i> juga terbukti mudah dipahami dan dilakukan dengan <i>tuning</i> yang minimum.</p>

	<p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Algoritma perbandingan untuk <i>Random Forest</i> akurasi lebih tinggi dari pada algoritma <i>KNN</i></p>
<p>Judul Klasifikasi penyakit kucing menggunakan metode <i>Random Forest</i></p> <p>Peneliti Abdul Hafiz Muttaqien</p> <p>Lokasi Palembang</p> <p>Tahun 2020</p>	<p><u>Hasil Penelitian</u></p> <p>Penelitian ini bertujuan untuk memperhatikan kesehatan kucing dengan memerhatikan penyakit kulit pada kucing. Oleh karena itu penelitian ini mengklasifikasi penyakit kucing khususnya pada kulit dengan metode <i>Random Forest</i> dengan klasifikasinya menggunakan 3 kelas yaitu <i>scabies</i>, <i>dermatophytosis</i>, dan sehat. Penelitian ini memiliki tingkat akurasi sebesar 96% dan penggunaan tree dalam jumlah besar yaitu 100.</p> <p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Pada penelitian ini sama - sama menggunakan <i>Random Forest</i> dan mendapatkan tingkat akurasi yang sangat baik yaitu 96%, dan akan dijadikan komparasi apakah akan memiliki akurasi yang dengan kasus yang berbeda</p>

1.6 Sistematika Penulisan

Adapun sistematika penulisan dalam menyusun tugas akhir antara lain sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, *state of the art* dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini merupakan penjelasan tentang teori yang berkaitan dengan penelitian.

BAB III METODE PENELITIAN

Bab ini membahas mengenai analisis yang dilakukan dengan metode *Random Forest*.

BAB IV PEMBAHASAN

Bab ini berisi tentang implementasi dan analisis hasil dari metode *Random Forest*.

BAB V KESIMPULAN DAN SARAN

Dalam bab ini membahas tentang kesimpulan dan saran yang sudah dilakukan dalam penelitian memprediksi serangan jantung dengan metode *Random Forest*.