

BAB I

PENDAHULUAN

A. Latar Belakang

Agenda *Sustainable Development Goals* (SDGs) yang diadopsi oleh Perserikatan Bangsa-Bangsa menempatkan kesehatan sebagai salah satu pilar utama pembangunan berkelanjutan, khususnya pada tujuan ketiga (*Good Health and Well-being*).¹ Tantangan kesehatan global saat ini tidak hanya terbatas pada penyakit menular, tetapi juga pada meningkatnya beban Penyakit Tidak Menular (PTM) yang mengancam kualitas hidup dan produktivitas masyarakat. Pemanfaatan inovasi digital menjadi strategi kunci untuk memperkuat sistem kesehatan, mempercepat deteksi dini, serta meningkatkan efektivitas intervensi promotif dan preventif. Integrasi inovasi berbasis data ini diharapkan mampu mendorong pencapaian target SDGs dengan cara yang lebih adaptif, terukur, sekaligus memperkuat respons global terhadap tantangan epidemiologi di masa depan.²

Transformasi digital kesehatan yang sejalan dengan agenda *Sustainable Development Goals* (SDGs) membuka peluang besar untuk membangun sistem kesehatan yang lebih tanggap dan berorientasi pada pencegahan.³ Indonesia melalui Kementerian Kesehatan tengah mendorong agenda transformasi digital kesehatan sebagai upaya memperkuat tata kelola pemanfaatan *big data* kesehatan secara nasional. Kebijakan ini mencakup digitalisasi rekam medis, integrasi sistem informasi kesehatan, serta penerapan teknologi analitik untuk mendukung pengambilan keputusan berbasis bukti (*evidence-based*).² Perubahan ini menjadi semakin relevan ketika dihadapkan dengan tantangan penyakit yang terus berkembang dan memerlukan strategi pengendalian yang komprehensif.

Salah satu tantangan terbesar yang dihadapi hampir seluruh negara adalah meningkatnya beban Penyakit Tidak Menular (PTM), yang menuntut perhatian khusus dari sisi pencegahan maupun pengendalian. Penyakit Tidak Menular (PTM) menyumbang lebih dari 74% kematian global dan merupakan penyebab utama disabilitas dengan kecenderungan meningkat pada kelompok usia yang

produktif, terutama di negara berpenghasilan rendah dan menengah, termasuk Indonesia.⁴ Saat ini terjadi pergeseran pola penyakit dari penyakit tidak menular ke penyakit degeneratif dalam satu dekade terakhir yang membutuhkan pengelolaan jangka panjang dan biaya besar.⁵

Data dari *Global Burden of Disease Study* tahun 2020, mencatat bahwa penyakit jantung iskemik merupakan penyebab kematian tertinggi di Indonesia, dengan peningkatan angka kematian lebih dari 25%.⁶ Kondisi ini menandakan semakin dominannya kelompok PTM khususnya penyakit jantung yang menjadi penyebab utama tingginya angka kematian dalam struktur morbiditas nasional.⁷ Selain itu beban yang ditimbulkan oleh PTM tidak hanya berdampak pada sistem kesehatan, tetapi juga menurunkan produktivitas masyarakat serta meningkatkan ketergantungan terhadap layanan kesehatan jangka panjang.⁸

Berdasarkan laporan Badan Kebijakan Pembangunan Kesehatan (BKPK) penyakit jantung menyumbang sebagian besar pembiayaan dalam sistem Jaminan Kesehatan Nasional (JKN). Pada tahun 2022, total biaya klaim untuk PTM mencapai lebih dari Rp17,92 triliun, menjadikannya sebagai penyakit katastropik dengan beban tertinggi.⁹ Data dari *Institute for Health Metrics and Evaluation* (IHME) yang menjelaskan bahwa penyakit jantung merupakan penyebab utama kematian mencapai angka 53,2%.¹⁰ Hal ini tidak hanya mengindikasikan prevalensi yang tinggi, tetapi juga menandakan bahwa penyakit jantung cenderung terdiagnosis pada stadium lanjut sehingga membutuhkan intervensi dan terapi kompleks serta mahal.¹¹

Penyakit jantung termasuk kedalam golongan penyakit katastropik yaitu penyakit yang memerlukan biaya pengobatan tinggi, berpotensi mengancam jiwa, dan membutuhkan perawatan jangka panjang.¹² Faktor risiko penyakit jantung bersifat multifaktorial dan melibatkan interaksi antara karakteristik individu, gaya hidup, kondisi klinis, dan faktor sosial ekonomi.¹³ Secara umum, faktor risiko Penyakit Tidak Menular (PTM) khususnya penyakit jantung diklasifikasikan menjadi dua, yaitu faktor yang tidak dapat dimodifikasi (usia, jenis kelamin, dan genetik), serta faktor yang dapat dimodifikasi (merokok, pola

makan, aktivitas fisik, konsumsi buah dan sayur, tekanan darah, indeks masa tubuh, dan kadar glukosa).^{14,15}

Sebagian besar penelitian epidemiologi kesehatan di Indonesia masih menggunakan data lama seperti Riset Kesehatan Dasar (Riskesdas) 2018 atau *Indonesia Family Life Survey* (IFLS) 2014, yang memiliki keterbatasan dalam cakupan variabel serta tidak representatif hingga tingkat kabupaten atau kota.^{16–18} Selain itu, pendekatan analisis yang digunakan mayoritas masih berbasis regresi linier atau logistik, dengan keterbatasan dalam mengungkap pola-pola tersembunyi dan interaksi kompleks antar faktor risiko.¹⁹ Padahal, karakteristik penyakit jantung sangat dipengaruhi oleh interaksi antar variabel yang tidak selalu linier, terutama ketika mempertimbangkan determinan sosial dan perilaku kesehatan masyarakat yang beragam antar wilayah di Indonesia.²⁰ Oleh karena itu, dibutuhkan pendekatan yang lebih representatif dan fleksibel untuk memahami profil risiko penyakit jantung secara menyeluruh.²¹

Keterbatasan data tersebut direspons oleh pemerintah melalui Kementerian Kesehatan RI dengan peluncuran Survei Kesehatan Indonesia (SKI) 2023 sebagai integrasi dari Riskesdas dan Survei Status Gizi Indonesia (SSGI). Survei Kesehatan Indonesia (SKI) 2023 berhasil mengumpulkan sebanyak lebih dari 586.000 rumah tangga di seluruh Indonesia dan menghasilkan *dataset* mikro-level yang mencakup variabel klinis, demografis, sosial-ekonomi, dan perilaku kesehatan.²² Keunggulan SKI 2023 terletak pada kemampuannya untuk menggambarkan kondisi kesehatan masyarakat hingga tingkat kabupaten atau kota, serta menyediakan basis data yang lebih mutakhir untuk analisis kesehatan masyarakat dan evaluasi kebijakan.²³

Selain memanfaatkan data SKI, pendekatan analisis juga perlu diperbarui. Salah satu pendekatan yang relevan untuk menangani data besar dan kompleks adalah *Machine Learning* (ML).²⁴ ML merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan membentuk model prediksi klasifikasi berdasarkan pola yang ditemukan.²⁵ Dibandingkan dengan analisis statistik konvensional, ML mampu menangani interaksi

variabel yang kompleks, mengenali pola *non-linier*, serta mengurangi risiko *overfitting* melalui validasi silang.

Machine Learning dalam konteks kesehatan digunakan secara luas di berbagai negara untuk memprediksi risiko penyakit, klasifikasi klinis, hingga pengembangan sistem peringatan dini (*early warning system*).²⁶ Urgensi tersebut menekankan peranan *Machine Learning* (ML) dalam membangun pendekatan analisis berbasis teknologi relevan untuk menjawab tantangan kompleks dalam analisis risiko Penyakit Tidak Menular (PTM).²⁷ *Machine Learning* (ML) merupakan bagian kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa dilakukan program secara eksplisit, dan efektif mengenali pola tersembunyi, menangani data berskala besar, dan membangun model prediksi yang adaptif terhadap beragam variabel.²⁸

Sejumlah studi terbaru mendemonstrasikan keunggulan ML dalam konteks kesehatan masyarakat. Sejumlah penelitian menggunakan algoritma ML seperti *gradient boosting* dan *Random Forest* untuk memprediksi hipertensi berbasis data *check-up* rutin, dan menghasilkan akurasi tinggi dengan *Area Under the Curve* (AUC) lebih dari 0.90.²⁹ Fitur-fitur penting dalam model tersebut antara lain usia, tekanan darah, indeks masa tubuh, dan kadar glukosa puasa, variabel-variabel yang juga tersedia dalam data Survei Kesehatan Indonesia (SKI) 2023 dan sangat relevan untuk konteks pemodelan faktor risiko penyakit dalam skala nasional. Hal ini membuka peluang penerapan algoritma serupa pada data SKI 2023 untuk menghasilkan klasifikasi risiko penyakit jantung yang lebih akurat dan berbasis bukti.

Studi lain mengenai *Machine Learning* (ML) kerap menggunakan pendekatan serupa pada data survey nasional di Asia Selatan dengan algoritma *XGBoost* dan *Random Forest* untuk memprediksi hipertensi menunjukkan bahwa model ML mampu mengenali kombinasi faktor risiko yang kompleks dan mengklasifikasikan populasi berisiko tinggi dengan sensitivitas lebih dari 99%.³⁰ Penelitian dalam konteks penyakit kronik lainnya membandingkan model klasifikasi stroke dan infark miokard pada pasien diabetes tipe 2 dengan metode konvensional *logistic regression*.³¹ Hasilnya, *Machine Learning*

menunjukkan kinerja prediktif yang lebih baik, terutama dalam mengidentifikasi pasien yang tidak terdeteksi oleh model statistik pada umumnya.³²

Beberapa algoritma *Machine Learning* seperti *logistic regression*, *decision tree*, *Random Forest*, dan *gradient boosting* telah terbukti efektif dalam membangun model prediktif untuk berbagai penyakit kronis.³³ Terbukti dengan adanya penelitian yang relevan mengenai model *Machine Learning* untuk manajemen hipertensi dan stroke telah membantu meningkatkan efektivitas program skrining dan pengendalian PTM berbasis risiko.³⁴ Potensi tersebut juga sangat relevan untuk pengembangan penelitian di Indonesia, terutama dalam mendukung transformasi digital sistem kesehatan sebagaimana diamanatkan dalam Rencana Pembangunan Jangka Panjang Nasional (RPJMN) 2020-2024 dan program Satu Sehat dan Kementrian Kesehatan.

Selain berperan dalam analisis faktor risiko, *Machine Learning* (ML) berpotensi besar diintegrasikan dengan sistem Rekam Medis Elektronik (RME) untuk memperkuat transformasi digital kesehatan. Integrasi ini memungkinkan analisis *real-time* terhadap data klinis pasien guna mengidentifikasi pola kesehatan, mendeteksi risiko dini, dan memberikan rekomendasi intervensi yang tepat. Integrasi *Machine Learning* (ML) dalam RME juga mengoptimalkan manajemen layanan kesehatan melalui terapi personal, pemantauan pengobatan, serta prediksi beban penyakit populasi, sehingga bermanfaat bagi praktik klinis maupun kebijakan kesehatan publik.

Pemanfaatan *Machine Learning* dalam Rekam Medis Elektronik (RME) membantu tenaga kesehatan mengakses riwayat medis pasien secara komprehensif, meningkatkan akurasi diagnosis, serta mempercepat pengambilan keputusan berbasis bukti. *Machine Learning* menjadi pendekatan yang sangat ideal untuk membangun model prediksi penyakit dapat diadopsi dalam sistem informasi kesehatan digital nasional.³⁵ Hingga saat ini, belum ada penelitian berbasis SKI 2023 yang memanfaatkan pendekatan *Machine Learning* untuk mengklasifikasikan penyakit jantung.

Penelitian ini dirancang untuk menganalisis faktor risiko utama yang berkontribusi terhadap kejadian penyakit jantung di Indonesia dengan mengintegrasikan pendekatan analitik berbasis *Machine Learning* (ML) dalam membangun model klasifikasi risiko. Pendekatan ini memungkinkan identifikasi variabel yang paling signifikan dalam memprediksi risiko, baik bersifat individual maupun faktor komunal. Implementasi model ini juga membuka peluang integrasi ke dalam *platform digital* berbasis risiko individual guna mendukung pelayanan kesehatan primer secara lebih efisien dan responsif.³⁶ Kontribusi penelitian ini tidak hanya bersifat praktis dalam konteks pelayanan kesehatan, tetapi juga teoritis dalam memperkaya kajian epidemiologi prediktif dan pemanfaatan teknologi *big data*.^{37,38}

Penggunaan *dataset* Survei Kesehatan Indonesia (SKI) 2023 yang komprehensif, penerapan algoritma *Random Forest* untuk mengidentifikasi faktor risiko dominan, serta orientasi pada pengembangan sistem skrining berbasis risiko yang terintegrasi dengan layanan kesehatan primer dalam Rekam Medis Elektronik (RME) menjadi tiga pilar kebaruan utama dalam penelitian ini. Integrasi data nasional terbaru dengan menggunakan *Machine Learning* (ML), dan dukungan terhadap transformasi digital kesehatan nasional ini belum banyak ditemukan pada studi epidemiologi di Indonesia, sehingga berpotensi memberi kontribusi signifikan bagi kebijakan preventif dan deteksi dini penyakit Penyakit Tidak Menular (PTM) di Indonesia melalui pendekatan berbasis bukti yang presisi.³⁹

Integrasi data mikro SKI 2023 yang representatif secara nasional dengan pemodelan *Machine Learning* (ML) memungkinkan identifikasi faktor risiko penyakit jantung secara lebih akurat dan komprehensif dibandingkan dengan pendekatan konvensional. Hasil klasifikasi mampu memberikan *framework* yang dapat diimplementasikan pada sistem skrining kesehatan berskala nasional. Penelitian ini berorientasi pada transformasi digital kesehatan, sehingga model yang dihasilkan berpotensi menjadi instrumen strategis bagi tata kelola sistem kesehatan yang adaptif, prediktif, dan meningkatkan kualitas hidup untuk mendukung pembangunan kesehatan nasional yang berkelanjutan.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini fokus pada upaya untuk mengidentifikasi faktor risiko dominan yang berkontribusi terhadap kejadian penyakit jantung menggunakan data mikro Survei Kesehatan Indonesia (SKI) 2023. Permasalahan utama yang ingin dijawab adalah faktor risiko apa saja yang paling berpengaruh terhadap kejadian penyakit jantung, seberapa besar kontribusi masing-masing variabel prediktor melalui analisis *feature importance*, serta bagaimana algoritma *Machine Learning* (ML), khususnya *Random Forest*, dapat digunakan untuk melakukan klasifikasi risiko penyakit jantung berbasis data mikro SKI 2023.

C. Tujuan Penelitian

1. Tujuan Umum

Mengidentifikasi faktor risiko utama dan membangun klasifikasi risiko penyakit jantung berbasis *Machine Learning* (ML) dengan menggunakan data mikro Survei Kesehatan Indonesia (SKI) 2023.

2. Tujuan Khusus

- a. Mengidentifikasi faktor-faktor risiko dominan terhadap penyakit jantung dengan metode *feature importance* menggunakan *Information Gain* (IG) dan *Gain Ratio* (GR).
- b. Membangun model klasifikasi risiko penyakit jantung berbasis *Machine Learning* menggunakan algoritma *Random Forest* menggunakan metrik *sensitivity* dan *specificity*.

D. Manfaat Penelitian

1. Manfaat Akademis

- a. Memberikan kontribusi ilmiah dalam pengembangan kajian rekam medis dan informasi kesehatan, khususnya pada penerapan *Machine Learning* (ML) dalam klasifikasi risiko penyakit jantung.
- b. Memperkaya literatur mengenai analisis faktor risiko Penyakit Tidak Menular (PTM) berbasis data mikro survei nasional yang representatif.

- c. Menjadi referensi bagi penelitian selanjutnya yang ingin mengembangkan pendekatan epidemiologi prediktif dengan memanfaatkan *big data* kesehatan.
- d. Mendukung perkembangan metodologi penelitian di bidang kesehatan dengan memadukan pendekatan *Knowledge Discovery in Database* (KDD) dan analisis *feature importance*.

2. Manfaat Praktis

- a. Menjadi dasar bagi tenaga kesehatan dan pembuat kebijakan dalam mengembangkan sistem skrining penyakit jantung berbasis teknologi yang lebih presisi dan efisien.
- b. Memberikan solusi berbasis data untuk memprioritaskan intervensi terhadap kelompok risiko tinggi secara cepat dan terukur.
- c. Mendukung agenda transformasi digital Kementerian Kesehatan Republik Indonesia melalui pendekatan promotif dan preventif yang berbasis kecerdasan buatan.
- d. Mengintegrasikan hasil model pada tingkat nasional sekaligus mengadaptasikan pada kebijakan daerah, sehingga intervensi dapat disesuaikan dengan karakteristik populasi lokal dan memperkuat relevansi program pencegahan di berbagai wilayah.

E. Keaslian Penelitian

Tabel 1. 1 Keaslian Penelitian

Peneliti	Judul Penelitian	Metode Penelitian	Variabel Penelitian	Letak Perbedaan
Mellya Putri & Erlin (2024)	Prediksi Penyakit Stroke Menggunakan <i>Machine Learning</i> dengan Algoritma <i>Random Forest</i>	Eksperimen kuantitatif dengan jenis algoritma <i>Random Forest</i>	Usia, jenis kelamin, hipertensi, riwayat penyakit jantung, status merokok, BMI	Berbeda jenis target penyakit, fokus pada penyakit stroke, dan data klinis institusional.

Peneliti	Judul Penelitian	Metode Penelitian	Variabel Penelitian	Letak Perbedaan
Changchun Hei et al. (2024)	<i>Machine Learning Analysis of Gene Expression Profiles of Pyroptosis-related Genes in Ischemic Stroke</i>	Bioinformatika, LASSO, SVM-RFE, <i>Random Forest</i>	Data ekspresi gen (PRDEGs) pasien stroke iskemik	Berbeda jenis target penyakit, fokus pada stroke dengan data genomik dan identifikasi biomarker genetik.
Mohammed Ahmed & Idress Husien (2024)	<i>Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review</i>	Studi literatur dan eksperimen ensemble learning (RF, GB, CART, SVM, KNN)	Data kesehatan jantung (usia, tekanan darah, kolesterol, ECG, dll.)	Bersifat tinjauan literatur dengan pendekatan <i>hybrid/ensemble</i> , bukan implementasi pada <i>dataset</i> survei populasi.
Aida Khalisatifa et al. (2024)	Klasifikasi Risiko Penyakit Serangan Jantung dengan Menggunakan Algoritma C4.5	<i>Data mining, Decision Tree</i> C4.5, CRISP-DM	Usia, jenis kelamin, <i>chest pain</i> , tekanan darah, kolesterol, dll.	Menggunakan algoritma C4.5 pada <i>dataset</i> UCI berskala kecil dengan <i>RapidMiner</i> .
Kwandy Chandra & Juan S. Prasetyo (2024)	Prediksi Penyakit Jantung Koroner Menggunakan K-NN dan Regresi Logistik	CRISP-DM dengan algoritma K-NN dan <i>Logistic Regression</i>	Kondisi kesehatan, aktivitas fisik, merokok, alkohol, dll.	Menggunakan algoritma K-NN dan Regresi Logistik dengan <i>dataset</i> berbeda.