

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cardiovascular Disease (CVD) merupakan kelompok gangguan medis yang menyerang sistem jantung dan pembuluh darah. CVD tidak merujuk pada satu jenis penyakit saja, melainkan mencakup berbagai kondisi seperti penyakit jantung koroner (*coronary heart disease*), stroke, hipertensi, penyakit arteri perifer, dan gagal jantung. Penyakit ini berkembang secara progresif akibat interaksi berbagai faktor risiko yang bersifat biologis, demografis, klinis, dan gaya hidup, sehingga menjadikannya sebagai permasalahan kesehatan global yang kompleks dan multidimensional [1]. Faktor-faktor tersebut meliputi tekanan darah tinggi, kadar kolesterol, gula darah puasa, indeks massa tubuh (BMI), lingkar perut, riwayat keluarga, kebiasaan merokok, pola makan tinggi lemak dan rendah nutrisi, kurangnya aktivitas fisik, serta konsumsi alkohol berlebihan [2]. Interaksi berbagai faktor risiko tersebut tidak bersifat tunggal, melainkan saling memengaruhi secara simultan dan membentuk pola hubungan *nonlinier* yang kompleks. Kegagalan dalam melakukan penilaian risiko sejak dini berpotensi meningkatkan risiko kejadian kardiovaskular secara signifikan. Selain berdampak pada aspek kesehatan, CVD juga memberikan implikasi signifikan terhadap sektor ekonomi dan bisnis, khususnya industri asuransi. Penyakit ini menjadi salah satu penyebab utama beban finansial bagi individu maupun sistem kesehatan akibat tingginya biaya perawatan dan pengobatan, sehingga ketidakakuratan dalam penilaian risiko dapat memengaruhi penentuan profil nasabah dan berpotensi menimbulkan kerugian [3]. Berdasarkan pemodelan yang dipublikasikan pada tahun 2025, beban penyakit kardiovaskular diproyeksikan terus meningkat hingga tahun 2050. Kondisi ini menunjukkan bahwa hingga tahun 2026, CVD masih menjadi isu kesehatan yang sangat relevan untuk diteliti, khususnya dalam upaya mengembangkan metode klasifikasi risiko yang lebih akurat [4].

Kompleksitas hubungan antar faktor risiko tersebut menuntut pendekatan analisis yang mampu mengolah data dalam jumlah besar serta menangkap pola hubungan *nonlinier* secara optimal. Dalam konteks ini, perkembangan teknologi kecerdasan buatan menawarkan solusi berbasis komputasional untuk mendukung analisis risiko penyakit kardiovaskular. Berdasarkan beberapa penelitian sebelumnya mengenai deteksi CVD, berbagai algoritma telah digunakan, seperti K-Nearest Neighbors (KNN) dengan akurasi

70%, Support Vector Machine (SVM) 72%, Decision Tree 63%, Naïve Bayes 71%, Random Forest 70%, Logistic Regression 73%, *Artificial Neural Network* (ANN) 73%, Voting Classifier 73%, dan *Gradient Boosting* 73%, dengan nilai akurasi yang berkisar antara 63% hingga 73% [5]. Penelitian lainnya juga menunjukkan hasil prediksi CVD menggunakan beberapa algoritma *machine learning* dengan performa yang berbeda, di mana metode KNN memperoleh akurasi sebesar 80,1%, Random Forest 83,5%, eXtreme *Gradient Boosting* (XGBoost) 83,0%, dan *Gradient Boosted Decision Trees* (GBDT) 84,5%, sementara metode *Deep Neural Network* (DNN) menunjukkan kinerja terbaik dengan akurasi sebesar 85,0% [6]. Selain memiliki performa akurasi yang tinggi, metode DNN juga dinilai mampu mempelajari pola hubungan nonlinier yang kompleks pada data medis melalui beberapa *hidden layer* yang saling terhubung. Kemampuan tersebut membuat DNN lebih efektif dalam mengekstraksi representasi fitur secara mendalam dibandingkan metode *machine learning* tradisional. Penelitian sebelumnya menunjukkan bahwa penerapan DNN pada prediksi penyakit kardiovaskular mampu meningkatkan performa klasifikasi serta memberikan hasil yang lebih optimal pada data medis dengan kompleksitas tinggi dan mampu mencapai akurasi sebesar 90,08% [7]. Berdasarkan penelitian-penelitian terdahulu tersebut, metode DNN menunjukkan performa yang baik dalam klasifikasi penyakit kardiovaskular. Oleh karena itu, metode DNN dipilih sebagai dasar pada penelitian ini dan dikombinasikan dengan teknik SMOTE untuk meningkatkan performa klasifikasi.

Namun demikian, selain pemilihan metode klasifikasi yang tepat, karakteristik *dataset* yang digunakan juga menjadi faktor penting yang dapat memengaruhi kinerja model dalam proses klasifikasi. *dataset* yang digunakan dalam penelitian ini memiliki kondisi yang tidak seimbang (*imbalanced*), yang terlihat dari distribusi pada kolom *Risk* yaitu *Low* sebanyak 220, *Intermediary* sebanyak 581, dan *High* sebanyak 728. Kondisi ketidakseimbangan tersebut berpotensi memengaruhi kemampuan model dalam mempelajari karakteristik setiap kelas [8]. Oleh karena itu, penerapan teknik *Synthetic Minority Oversampling Technique* (SMOTE) dapat menjadi salah satu pendekatan yang efektif untuk mengatasi permasalahan tersebut. Beberapa penelitian sebelumnya telah mengombinasikan metode *Convolutional Neural Network* (CNN) dengan teknik SMOTE dalam memprediksi penyakit kardiovaskular dan menunjukkan hasil akurasi yang cukup tinggi, yaitu mencapai 91% [9]. Penelitian lain juga melaporkan bahwa penggabungan teknik SMOTE dengan DNN pada pemetaan tipe hutan mampu menghasilkan akurasi hingga 94% [10]. Selain itu, penerapan metode *Artificial Neural Network* (ANN) yang

dikombinasikan dengan SMOTE juga terbukti mampu meningkatkan kinerja model dalam menangani data yang tidak seimbang, seperti pada tugas klasifikasi biner dengan menghasilkan akurasi sebesar 99,8% [11]. Berdasarkan penelitian-penelitian terdahulu tersebut, SMOTE terbukti efektif dalam mengatasi ketidakseimbangan data dan meningkatkan performa klasifikasi. Oleh karena itu, penelitian ini menerapkan kombinasi DNN dan SMOTE untuk mengklasifikasikan tingkat risiko penyakit kardiovaskular.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengklasifikasikan tingkat risiko penyakit kardiovaskular dengan menggunakan metode DNN yang dikombinasikan dengan teknik SMOTE. Penerapan tersebut dilakukan untuk mengatasi permasalahan ketidakseimbangan kelas pada *dataset* sehingga model dapat mempelajari karakteristik setiap kelas secara lebih optimal. Melalui pendekatan tersebut, model diharapkan mampu menangkap pola hubungan *nonlinier* yang kompleks antar faktor risiko serta meningkatkan performa klasifikasi risiko penyakit kardiovaskular dibandingkan dengan metode *machine learning* lainnya. Oleh karena itu, penelitian ini diangkat dengan Judul “**Klasifikasi Risiko Penyakit Kardiovaskular Berdasarkan Faktor Demografis, Klinis, dan Gaya Hidup Menggunakan Deep Neural Network**”.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah bagaimana penggabungan metode DNN dan SMOTE dapat mengatasi kompleksitas hubungan *nonlinier* dan multidimensional antara faktor risiko serta ketidakseimbangan distribusi data untuk menghasilkan klasifikasi tingkat risiko penyakit kardiovaskular yang lebih optimal dan meminimalisir bias klasifikasi pada kelas minoritas.

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Mengklasifikasikan tingkat risiko penyakit kardiovaskular berdasarkan faktor demografis, klinis, dan gaya hidup menggunakan metode DNN yang mampu mempelajari pola hubungan nonlinier antar faktor risiko.
2. Menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan distribusi kelas pada dataset sehingga dapat meningkatkan performa model klasifikasi DNN.

1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini antara lain:

1. Penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya terkait penggunaan teknik SMOTE dalam mengatasi permasalahan ketidakseimbangan kelas pada *dataset*.
2. Penelitian ini diharapkan dapat memberikan gambaran mengenai pemanfaatan teknologi kecerdasan buatan dalam membantu proses klasifikasi risiko penyakit kardiovaskular secara lebih efektif dan akurat.

1.5 Ruang Lingkup

Ruang lingkup dari penelitian ini meliputi:

1. *Dataset* yang digunakan merupakan *dataset* penyakit kardiovaskular yang diperoleh dari Mendeley data [8], dengan jumlah 1.529 data yang mencakup variabel demografis, klinis, dan gaya hidup, serta variabel target berupa CVD *Risk Level*.
2. Penelitian berfokus pada permasalahan klasifikasi multikelas yang terdiri atas kategori *Low*, *Intermediary*, dan *High*.
3. Penelitian ini menggunakan pembagian *dataset* menjadi data latih, data validasi, dan data uji dengan rasio 80:10:10. Data latih digunakan untuk proses pelatihan model DNN, data validasi digunakan untuk memantau kinerja model selama proses pelatihan, sedangkan data uji digunakan untuk mengevaluasi performa akhir model.
4. Penelitian mempertimbangkan kondisi ketidakseimbangan distribusi kelas pada *dataset* yang ditangani menggunakan teknik SMOTE.
5. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.