

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan merupakan hal berharga pemberian dari Tuhan yang dimiliki oleh manusia untuk menunjang berbagai aktivitas sesuai dengan kebutuhan pribadi masing-masing. Kendati demikian, masih banyak orang yang kurang sadar pentingnya menjaga kesehatan dengan cara menerapkan pola hidup sehat sesuai anjuran para dokter. Ada banyak penyakit manusia yang hampir di setiap organ dari tubuh manusia ini memiliki jenisnya masing-masing sesuai dengan fungsi dari organ tubuh manusia tersebut. Dari banyaknya organ tubuh manusia, salah satu organ yang berperan penting dalam kelangsungan hidup manusia adalah jantung. Penyakit yang berkaitan dengan jantung salah satunya adalah penyakit kardiovaskular yang disebabkan oleh fungsi jantung dan pembuluh darah yang tidak bekerja semestinya. Beberapa contoh penyakit yang merepresentasikan dari gangguan tersebut adalah hipertensi, jantung koroner, stroke dan gagal jantung. Meninjau data yang tersedia dari Riset Kesehatan Dasar (RISKESDAS) pada tahun 2018 silam, di Indonesia terdapat kasus penyakit kardiovaskular hingga mencapai 2.784.064 penduduk menderita penyakit jantung dan mengalami peningkatan setiap tahunnya [1].

Penyakit kardiovaskular atau dalam Bahasa Inggris memiliki istilah *cardiovascular disease (CVD)* merupakan gangguan kesehatan manusia yang terjadi pada sistem organ jantung dan pembuluh darah. Lalu, ketika aliran darah dari jantung ke otak serta jantung tersumbat secara total dapat menyebabkan serangan jantung dan *stroke*. Tekanan darah sistolik tinggi tetap menjadi faktor risiko utama yang dapat dimodifikasi secara global untuk kematian kardiovaskular prematur yang disebabkan oleh 10,8 juta kematian penduduk akibat kardiovaskular dan 11,3 juta kematian secara keseluruhan pada tahun 2021, dan secara khusus dikaitkan dengan penyakit jantung iskemik dan kematian terkait stroke [2]. Adapun dua faktor penyebab dari penyakit kardiovaskular yang umum terjadi yaitu faktor risiko yang tidak bisa diubah contohnya penyakit keturunan dari keluarga, faktor usia, jenis kelamin, dan banyaknya kandungan gula dalam darah. Selanjutnya terdapat faktor risiko yang memungkinkan untuk diubah yaitu : *dislipidemia* yaitu kondisi kolesterol dalam tubuh yang abnormal, *diabetes melitus* atau peningkatan kadar glukosa, peningkatan tekanan darah, kondisi stres, diet tanpa berkonsultasi dengan ahli kesehatan serta kurangnya aktivitas fisik. Pencegahan penyakit kardiovaskular di Asia merupakan isu penting bagi kesehatan dunia, karena setengah dari populasi dunia tinggal di Asia [3]. Sebagian besar penyebab dari penyakit yang menyerang organ jantung ini adalah pola hidup yang kurang sehat serta

makanan yang kurang terkontrol. Selain itu, gizi yang tidak seimbang dalam suatu makanan juga dapat menjadi penyebab datangnya penyakit yang menyerang organ jantung, sebab jantung adalah bagian yang cukup krusial dalam tubuh yang seperti kita ketahui bersama bahwa fungsi jantung adalah memompa darah untuk dialirkan ke seluruh tubuh. Penyakit Kardiovaskular termasuk kedalam penyakit yang tidak menular, namun meskipun demikian penyakit kardiovaskular dapat menyebabkan kematian dikarenakan penyakit ini menyerang fungsi jantung. Mengingat berbahayanya penyakit kardiovaskular ini, perlu dilakukan klasifikasi dengan tujuan untuk memprediksi serta mendeteksi penyakit ini jika ditinjau dari faktor risiko yang menyebabkan terjadinya penyakit kardiovaskular.

Pada saat ini, *machine learning* dan *data mining* banyak digunakan untuk penelitian di berbagai bidang salah satunya adalah bidang kesehatan. Bidang kesehatan ini memiliki sekumpulan data terkait pasien, penyakit, pengobatan yang jika data tersebut dilakukan analisis, kumpulan data tersebut dapat membantu diagnosis, tindakan pencegahan, prediksi, maupun perumusan kebijakan dalam industri kesehatan [4]. *Machine Learning* merupakan ilmu pengembangan dari AI yang mempelajari data. *Machine Learning* ini memiliki fokus pada pengembangan sistem yang mampu melakukan pembelajaran tanpa adanya program secara manual oleh manusia. Untuk mendapatkan hasil *output* dari *testing* yang maksimal dan optimal, ketika proses *training*, *machine learning* membutuhkan data yang valid sebagai bahan belajar [5].

Algoritma *Neural Network* adalah suatu usaha dari manusia untuk memprogram logika komputer untuk meniru fungsi otak manusia. Otak manusia ini yang diyakini terdiri dari berjuta-juta unit pengolahan kecil yang disebut dengan *neuron*, dimana *neuron* ini saling terhubung satu dengan yang lain melalui koneksi *neuron* yang pada setiap *neuron* mengambil *input* dari satu set *neuron*. *Input* tersebut lalu diproses dan melewati *output* untuk satu set *neuron*. Hasil luaran tersebut kemudian dikumpulkan oleh *neuron* lain untuk diproses lebih lanjut [6].

Dalam prosesnya, klasifikasi menggunakan algoritma *Backpropagation Neural Network* ini membutuhkan *dataset* yang memiliki kelas beragam. Terdapat kelas *dataset* yang seimbang dan kelas *dataset* yang tidak seimbang. Pada *dataset* yang seimbang, proses klasifikasi dapat langsung dilakukan tanpa melalui tahapan tertentu, namun berbeda halnya ketika kelas *dataset* ini tidak seimbang perlu dilakukan teknik *sampling* untuk menangani *dataset* yang tidak seimbang supaya tidak menurunkan kinerja dari model prediksi. Dalam menangani ketidakseimbangan kelas *dataset*, dapat digunakan *over-sampling* dengan teknik SMOTE (*Synthetic Minority Oversampling Technique*) [7]. Teknik ini merupakan teknik yang cukup efisien dan dapat meningkatkan akurasi dari sebuah model yang dibuat.

Terdapat beberapa penelitian di bidang medis yang mana dalam penelitian tersebut menerapkan teknologi *deep learning* untuk melakukan klasifikasi penyakit kardiovaskular, salah satunya ialah penelitian pertama, yang mana dalam penelitian ini menganalisa akurasi terbaik dari beberapa *supervised learning* yaitu *logistic regression (LR)*, *k-nearest neighbors(k-NN)*, *random forest (RF)*, *stochastic gradient descent (SGD)*, dan *neural network (NN)* dengan hasil akurasi tertinggi adalah menggunakan algoritma *neural network* dengan hasil akurasi sebesar 89.60% [1]. Dalam penelitian selanjutnya ini disebutkan melakukan perhitungan dengan menggunakan *tools Rapidminer* dan menghasilkan akurasi terbaik sebesar 77.27% [6]. Dan penelitian lainnya yang memiliki kesimpulan bahwa sistem *call* adalah alat yang berguna untuk klasifikasi gangguan medis dan deteksi *Neural Network* [8].

Penelitian selanjutnya adalah penelitian yang melakukan kombinasi teknik *Synthetic Minority Oversampling Technique (SMOTE)* dan algoritma *Backpropagation Neural Network* yang bertujuan untuk mendeteksi apakah wanita mengalami kehamilan atau tidak jika wanita tersebut menggunakan kontrasepsi implan dengan menggunakan total 300 data yang terdiri dari 285 data mayoritas dan 15 data minoritas. Penelitian ini menghasilkan nilai akurasi sebesar 96% pada epoch ke-500 dan 1.000 untuk data yang menggunakan teknik SMOTE. Sedangkan data yang tidak menggunakan teknik SMOTE menghasilkan akurasi sebesar 95% pada epoch ke- 10 50, 100, 500 dan 1.000 [9].

Mengacu pada contoh penelitian tersebut, dapat dilihat bahwa *Neural Network* khususnya pada algoritma *Backpropagation Neural Network* menunjukkan hasil performa yang baik dalam klasifikasi penyakit kardiovaskular berdasarkan faktor risiko. Oleh karena itu, peneliti memutuskan untuk menggunakan teknologi *machine learning* dengan mengimplementasi algoritma *Backpropagation Neural Network* dalam pengklasifikasian kategori penderita kardiovaskular ke dalam dua kelas yaitu positif atau negatif. Alasan peneliti memilih model *Backpropagation Neural Network* adalah karena pada penelitian sebelumnya, model ini mampu menghasilkan performa akurasi yang tinggi [10]. Namun, pembeda penelitian ini dan penelitian sebelumnya yaitu menggunakan algoritma *Backpropagation Neural Network (BPNN)* untuk klasifikasi data. Selain itu digunakan *Synthetic Minority Oversampling Technique (SMOTE)* untuk menyeimbangkan data per kelas.

1.2 Perumusan Masalah

Pada bagian perumusan masalah, diuraikan mengenai permasalahan, metode usulan serta pertanyaan penelitian yang diuraikan sebagai berikut :

1.2.1 Permasalahan

Penyakit kardiovaskular merupakan penyakit dengan tingkat risiko kematian yang cukup tinggi terutama di negara dengan penghasilan yang sedang dan rendah. Adanya analisa faktor risiko terhadap penyakit kardiovaskular ini diperlukan supaya dapat dengan cepat dilakukan penanganan secara dini terhadap penyakit kardiovaskular sesuai dengan standar operasi prosedural penanganan penyakit kardiovaskular dari kemenkes RI. Saat ini, teknologi *artificial intelligence* seperti *machine learning* dapat digunakan untuk mendeteksi dan mengklasifikasikan penyakit kardiovaskular.

1.2.2 Metode Penelitian

Penelitian ini mengusulkan metode *Backpropagation Neural Network* untuk klasifikasi selain itu menggunakan teknik SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan data per kelas.

1.2.3 Pertanyaan Penelitian

Berdasarkan uraian latar belakang, dapat diambil rumusan masalah diantaranya sebagai berikut :

1. Berapa nilai akurasi, *recall*, presisi dan AUC dari klasifikasi menggunakan algoritma *Backpropagation Neural Network*?
2. Bagaimana pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) terhadap hasil klasifikasi?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Berdasarkan rumusan masalah yang ada, dapat ditentukan tujuan dari penelitian ini diantaranya adalah sebagai berikut :

1. Untuk mengetahui nilai akurasi yang dihasilkan menggunakan algoritma *Backpropagation Neural Network*.
2. Untuk mengetahui pengaruh SMOTE (*Synthetic Minority Oversampling Technique*) terhadap hasil klasifikasi.

1.3.2 Manfaat

Penelitian ini memiliki manfaat diantaranya sebagai berikut :

1. Menambah pengetahuan algoritma *Backpropagation Neural Network* serta implementasi pada klasifikasi penyakit kardiovaskular bagi penulis dan pembaca.
2. Sebagai alat bantu dokter atau tenaga kesehatan dalam mendiagnosa dan mengklasifikasikan penyakit kardiovaskular berdasarkan faktor risiko yang mempengaruhi.

1.4 Batasan Masalah

Penelitian ini memiliki batasan masalah dengan maksud agar penelitian yang dilakukan dapat terarah dan tidak menyimpang dari pokok permasalahan yang dibahas dalam penelitian ini. Berikut batasan masalah dari penelitian ini :

1. Data yang digunakan adalah *dataset* hasil penelitian [11] yang tersedia pada *repository* risiko kardiovaskular <https://github.com/laderast/cvdRiskData/tree/master/data-raw> . *Dataset* ini berisi beberapa faktor yang dapat mempengaruhi datangnya penyakit kardiovaskular.
2. Data yang digunakan berjumlah 425.195 data sampel dengan format .csv sebelum dilakukan SMOTE, dan berjumlah 750.650 data setelah dilakukan SMOTE
3. Data terdiri dari 14 fitur, 2 kelas dengan positif kardiovaskular dan negatif kardiovaskular
4. Distribusi kelas tidak merata sebelum dilakukan SMOTE dengan jumlah 49.870 kelas positif, dan 375.325 kelas negatif. Dengan perbandingan 10% kelas positif dan 90% kelas negatif. Setelah dilakukan SMOTE data mengalami perubahan sehingga baik kelas positif dan kelas negatif memiliki jumlah 375.325 data.
5. Fitur-fitur dari data yang digunakan adalah sebagai berikut :
 - a. *Htn* : Hipertensi [mmHg]
 - b. *Treat* : Perawatan hipertensi
 - c. *Smoking* : Perokok aktif
 - d. *T2d* : Penderita diabetes
 - e. *NumAge* : Umur pasien [tahun]
 - f. *Bmi* : Indeks massa tubuh [kg/m²]
 - g. *Tchol* : Total kolesterol [mg/dL]
 - h. *Sbp* : Tekanan darah pasien [mmHg]
 - i. *Male* : Laki-laki
 - j. *Female* : Perempuan
 - k. *AmInd* : Ras Amerika India
 - l. *Asian/PI* : Ras Asian / Pasifik India
 - m. *Black/AfAm* : Ras *Black* / Afrika Amerika
 - n. *White* : Ras Kulit Putih
6. *Hyperparameter* yang digunakan adalah :
 - a. *Learning rate* 0,01.
 - b. *Epoch* 1000
 - c. *Node input layer* 14, *node hidden layer* 3, *node output* 1.

7. Penelitian ini menggunakan 5 skenario uji coba dengan percobaan *hyperparameter* yang berbeda-beda.

1.5 Sistematika Penulisan Skripsi

Sistematika penulisan skripsi ini penulis susun dengan mengikuti aturan penulisan skripsi pada program studi Sistem Informasi Universitas Trunojoyo Madura. Berikut merupakan metodologi penulisan skripsi pada penelitian ini.

BAB I PENDAHULUAN

Bab ini menjelaskan terkait latar belakang, perumusan masalah, permasalahan penelitian, metode usulan, pertanyaan penelitian, tujuan dan manfaat penelitian serta batasan masalah yang dilakukan untuk klasifikasi penyakit kardiovaskular, serta sistematika penulisan skripsi.

BAB II KAJIAN PUSTAKA

Teori penunjang ini berisi landasan teori yang terkait dengan penyakit kardiovaskular, standar operasional prosedur kardiovaskular, *Artificial Intellegence*, *Neural Network*, *Backpropagation Neural Network*, *Supervised Learning*. Bab ini juga mencantumkan beberapa penelitian terdahulu yang relevan dengan penelitian ini.

BAB III METODE PENELITIAN

Bab ini membahas terkait perancangan sistem yang mencakup jenis penelitian, tahap penelitian, analisa kebutuhan sistem serta skenario pengujian klasifikasi penyakit kardiovaskular menggunakan algoritma *Backpropagation Neural Network* (BPNN) serta implementasi dari teknik SMOTE (*Synthetic Minority Oversampling Technque*) dalam *oversampling*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi spesifikasi perangkat, pengumpulan dan analisa data, proses pembuatan program dan hasil klasifikasi penyakit kardiovaskular menggunakan algoritma *Backpropagation Neural Network* (BPNN) serta SMOTE untuk menyeimbangkan kelas data. Hasil ini disajikan dalam bentuk grafik dan tabel disertai dengan analisa.

BAB V PENUTUP

Bab ini menjelaskan tentang kesimpulan berdasarkan penelitian yang dilakukan dan beberapa saran untuk penelitian selanjutnya yang dapat digunakan sebagai acuan dengan harapan dapat meningkatkan performa model

DAFTAR PUSTAKA

Pada daftar pustaka berisi kumpulan rujukan jurnal, prosiding dari penelitian terdahulu yang digunakan untuk menyusun skripsi ini.