

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jantung merupakan organ berongga dan memiliki empat ruang atau bilik yang terletak di antara kedua paru-paru di bagian tengah rongga toraks (Sloane, 2004). Jantung mempunyai fungsi yang penting di dalam tubuh manusia yakni sebagai pemompa yang menekan darah agar dapat mengalir ke semua bagian tubuh melalui pembuluh darah arteri ataupun vena (Susilawati, Rachman, Nurulniza, & Utomo, 2014). Penyakit yang disebabkan oleh bertumpuknya plak di dalam arteri koroner yang mensuplai oksigen ke otot jantung, sehingga mengakibatkan kerusakan yang berat pada jantung disebut penyakit jantung koroner (PJK) (Rajkumar & Reena, 2010).

Menurut laporan yang dipublikasikan oleh (WHO, 2017) pada 2015, sekitar 17,7 juta orang meninggal disebabkan penyakit kardiovaskuler. Penyakit kardiovaskuler termasuk PJK di Indonesia diestimasikan menyebabkan lebih dari 470.000 kematian setiap tahun. Menurut survei *Sample Registration System* (SRS), angka kematian PJK adalah sebesar 12,9% dari seluruh kematian (Kementerian Kesehatan Republik Indonesia, 2017). Prevalensi PJK menurut diagnosis dokter yang dilakukan oleh Riset Kesehatan Dasar (Riskesdas) pada tahun 2013 menunjukkan bahwa PJK berada pada posisi ketujuh tertinggi Penyakit Tidak Menular (PTM) di Indonesia (Badan Penelitian dan Pengembangan Kesehatan, 2013).

Diagnosa PJK awal biasanya memakai pemeriksaan fisik dan riwayat medis, kemudian diteruskan dengan uji lanjutan. Dari uji lanjutan ini, *coronary angiography* adalah standar emas dalam mendiagnosis keterjangkitan PJK (Phibbs, 2007). Uji *coronary angiography* lebih dipilih oleh pakar jantung dalam mendiagnosis keberadaan PJK terhadap pasien dengan akurasi yang tinggi meskipun *invasive*, beresiko, dan juga mahal. Bila melihat kekurangan uji

lanjutan ini, maka perlu dikembangkan sebuah metode yang mampu mendiagnosis keberadaan PJK sebelum dilakukan *coronary angiography* (Setiawan, Venkatachalam, & Hani, 2009).

(Sultana, Haider, & Uddin, 2016) melakukan penelitian terhadap beberapa teknik *data mining*, yakni *KStar*, *Decision Tree*, *Sequential Minimal Optimization* (SMO), *Bayes Net*, dan MLP. Hasil penelitian menunjukkan bahwa SMO memiliki akurasi tertinggi sebesar 89% dibandingkan 4 metode yang diujikan lainnya dalam menangani prediksi penyakit jantung.

SMO terbukti memiliki kemampuan klasifikasi yang paling baik dalam menangani prediksi PJK. Namun di dalam *data mining*, apabila data memiliki terlalu banyak *noise*, *redundant* dan tidak relevan, maka proses *knowledge discovery* akan mengalami kesulitan. Teknik seleksi fitur adalah salah satu teknik untuk menyeleksi fitur data dengan cara mengidentifikasi dan menghilangkan informasi yang *redundant* dan tidak relevan. Dengan pemilihan teknik seleksi fitur yang tepat, dimensi data akan berkurang, sehingga ruang hipotesis akan semakin mengecil dan pembelajaran algoritma akan berjalan lebih cepat (Purbasari & Nugroho, 2013), bahkan pada banyak kasus akurasi pada teknik klasifikasi *data mining* dapat meningkat.

(Purbasari & Nugroho, 2013) meneliti pengaruh 6 macam seleksi fitur, yakni *Information Gain* (IG), *ReliefF*, *Principal Component Analysis* (PCA), *Correlation-based Feature Selection* (CFS), *Consistency-based Subset Evaluation* (CNS), dan *Wrapper Subset Evaluation* terhadap 2 algoritma klasifikasi, yakni *Decision Tree C4.5* dan *Naïve Bayes* untuk prediksi 25 macam dataset dari koleksi (UCI Machine Learning Repository, 1998). Kesimpulan yang dihasilkan adalah teknik *Wrapper* adalah seleksi fitur terbaik dengan total menang 17 kali dalam berhasil menaikkan akurasi dari kedua *classifier*, apabila kelambanan bukanlah masalah. Kalau iya, maka *ReliefF* dengan 13 kali menang dan CNS dengan 9 kali menang setelah *Wrapper* adalah pemilik rata-rata performa yang bagus.

Pada diagnosis medis, menurut (Prasetyo, 2012) reduksi data adalah masalah yang penting. Namun, data medis seringkali mengandung sejumlah fitur yang tidak relevan dan *redundant*, yang dalam sejumlah kasus secara dramatis mempengaruhi kualitas dari diagnosis suatu penyakit. Maka seleksi fitur bisa digunakan untuk menyeleksi fitur yang relevan pada data medis. Namun, untuk menghindari hilangnya faktor-faktor medis yang dianggap penting, perlu dilakukan proses seleksi fitur yang didasarkan pada pakar medis atau yang disebut *medical knowledge motivated feature selection* (MFS).

(Nahar, Imam, Tickle, & Chen, 2013) menguji pengaruh teknik seleksi fitur MFS dan *Computer-based Feature Selection* (CFS) yang digabungkan dengan beberapa algoritma klasifikasi seperti *Naïve Bayes*, SMO, IBK, *AdaBoostM1*, J48, dan PART untuk diagnosis PJK. Semua algoritma klasifikasi ini pada dataset Sick-4 berhasil meraih akurasi sebesar 97.05% baik menggunakan MFS maupun MFS+CFS, kecuali *Naïve Bayes* yang hanya dapat mencapai akurasi sebesar 97.05% menggunakan MFS saja, karna saat *Naïve Bayes* menggunakan MFS+CFS akurasi yang dihasilkan hanyalah sebesar 91.17%. Sehingga menghasilkan kesimpulan baik MFS maupun MFS+CFS merupakan teknik yang menjanjikan untuk digunakan untuk diagnosis PJK.

Deteksi dini terhadap PJK diperlukan sekali untuk menyelamatkan banyak nyawa. Menyadari kegunaan dari *data mining* untuk membantu diagnosis terhadap PJK menjadi sangat penting. Beberapa penelitian sebelumnya telah menyediakan detail perbandingan dari beberapa *classifier* untuk deteksi PJK di mana SMO telah menunjukkan potensi sebagai algoritma klasifikasi pada area ini, terutama ketika mempertimbangkan total akurasi sebagai ukuran performa (Nahar, Imam, Tickle, & Chen, 2013). (Purbasari & Nugroho, 2013) telah menunjukkan bahwa teknik seleksi fitur *ReliefF* dan CNS memiliki kehandalan dalam mereduksi fitur yang tidak relevan dan *redundant*, begitu pula dengan MFS yang telah ditunjukkan oleh (Nahar, Imam, Tickle, & Chen, 2013). Mengetahui ini, maka penulis akan mencoba melakukan penelitian untuk prediksi PJK

menggunakan algoritma klasifikasi SVM berbasis teknik seleksi fitur MFS yang digabungkan dengan *ReliefF* dan CNS dengan harapan dapat memperbaiki akurasi dari penelitian-penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan dari uraian latar belakang sebelumnya maka dapat dirumuskan suatu permasalahan, antara lain:

1. Bagaimana algoritma klasifikasi *Support Vector Machine* dapat memprediksi penyakit jantung koroner?
2. Bagaimana teknik seleksi fitur MFS, *ReliefF*, dan CNS dapat mereduksi dimensi data penyakit jantung koroner?
3. Bagaimana pengaruh teknik seleksi fitur MFS, *ReliefF*, dan CNS terhadap akurasi algoritma klasifikasi *Support Vector Machine* dalam memprediksi penyakit jantung koroner?

1.3 Tujuan Penelitian

Dari rumusan masalah yang terjadi maka penelitian ini bertujuan untuk:

1. Mengetahui bagaimana algoritma klasifikasi *Support Vector Machine* dapat memprediksi penyakit jantung koroner.
2. Mengetahui bagaimana teknik seleksi fitur MFS, *ReliefF*, dan CNS dapat mereduksi dimensi data penyakit jantung koroner.
3. Mengetahui bagaimana pengaruh teknik seleksi fitur MFS, *ReliefF*, dan CNS terhadap akurasi algoritma klasifikasi *Support Vector Machine* dalam memprediksi penyakit jantung koroner.

1.4 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat bermanfaat untuk membantu peneliti selanjutnya untuk merancang-bangun sistem prediksi penyakit jantung koroner yang dapat menghasilkan hasil diagnosa yang tidak bias terhadap diagnosis yang dilakukan oleh pakar. Selain itu, juga diharapkan dapat menjadi saran bagi tenaga medis untuk mendiagnosis penyakit jantung koroner sebelum dilakukan

metode *gold standard*, sehingga prosedur diagnosis penyakit jantung koroner yang *invasive*, beresiko, serta tidak ekonomis terhadap pasien dapat dihindari.

1.5 Batasan Masalah

Agar dalam penulisan tugas akhir ini tidak menyimpang dari permasalahan dan lebih terarah, maka perlu adanya batasan masalah pada ruang lingkup bahasan antara lain adalah sebagai berikut:

1. Menggunakan dan membahas teknik seleksi fitur MFS, *ReliefF*, dan CNS beserta pengaruhnya terhadap algoritma klasifikasi SVM dalam memprediksi PJK.
2. Aplikasi yang akan digunakan untuk memproses data adalah *Waikato Environment for Knowledge Analysis* (Weka).
3. Data set PJK yang akan digunakan dalam penelitian ini diambil dari (UCI Machine Learning Repository, 1988).

1.6 Sistematika Pembahasan

Untuk memahami lebih jelas isi dari tugas akhir ini, maka materi-materi yang tertera pada penelitian ini dikelompokkan menjadi beberapa sub bab dengan sistematika pembahasan sebagaimana berikut:

- Bab I : Merupakan pendahuluan yang berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika pembahasan.
- Bab II : Yaitu tinjauan pustaka yang menjelaskan tentang pengertian konsep penyakit jantung koroner, data mining, teknik seleksi fitur, *ReliefF*, *Consistency-based Feature Selection* (CNS), dan algoritma klasifikasi *Support Vector Machine* (SVM) yang diambil dari kutipan buku atau jurnal yang berkaitan dengan penyusunan laporan skripsi serta beberapa *literature review* yang berhubungan dengan penelitian.

- Bab III : Metodologi penelitian yang berisi tentang metode analisis kebutuhan alat yang digunakan, teknik pengumpulan data, teknik pemahaman data, teknik persiapan data, pemodelan algoritma pembelajaran, dan teknik evaluasi yang akan diusulkan untuk mengetahui seberapa berhasilnya metode yang telah diusulkan oleh peneliti.
- Bab IV : Merupakan hasil dan pembahasan yang meliputi perancangan penelitian, spesifikasi alat yang dibutuhkan, implementasi dari rancangan yang telah dibuat dan penjelasan proses prediksi penyakit jantung koroner menggunakan metode yang telah diusulkan oleh peneliti.
- Bab V : Adalah penutup dari tugas akhir ini yang berisi tentang kesimpulan dan saran-saran.