

## ABSTRAK

Penyakit jantung koroner merupakan penyakit jantung yang disebabkan oleh adanya penyempitan atau penyumbatan pada pembuluh darah koroner. Di Indonesia, penyakit jantung koroner masih menjadi penyebab kematian tertinggi dan Kemenkes memperkirakan jumlah kematiannya akan terus meningkat. Penyakit jantung koroner memiliki gejala utama yaitu adanya nyeri dada atau *angina*, yang dapat dipastikan dengan berkonsultasi langsung pada dokter spesialis jantung dan pembuluh darah. Akan tetapi, jumlah dokter spesialis jantung dan pembuluh darah di Indonesia masih sangat terbatas. Terbatasnya jumlah dokter spesialis jantung dan pembuluh darah dapat menyebabkan terjadinya keterlambatan penanganan pasien. Permasalahan tersebut dapat diatasi dengan melakukan pendekatan *machine learning* menggunakan klasifikasi. Klasifikasi penyakit jantung koroner dapat dilakukan menggunakan salah satu algoritma klasifikasi terbaik yaitu algoritma *K-Nearest Neighbor*. Namun, kelemahan dari algoritma tersebut yang sensitif terhadap adanya fitur data yang tidak relevan dapat menyebabkan akurasi menjadi rendah.

Masalah pada algoritma *K-Nearest Neighbor* tersebut dapat diatasi dengan melakukan seleksi fitur. Seleksi fitur adalah proses memilih atau menyeleksi fitur data yang paling penting untuk digunakan yang dilakukan pada tahap *pre-processing* data. Pada penelitian ini, seleksi fitur dilakukan menggunakan *Information gain* yang merupakan salah satu metode paling efisien dalam menyeleksi fitur data dengan hasil yang baik, sehingga dipercaya dapat mengatasi kelemahan yang dimiliki algoritma *K-Nearest Neighbor*.

Dalam penelitian ini, pengujian dilakukan berdasarkan *confusion matrix* untuk menghitung nilai akurasi yang dihasilkan dari klasifikasi penyakit jantung koroner menggunakan algoritma *K-Nearest Neighbor* yang dilakukan dengan seleksi fitur menggunakan *information gain* maupun tanpa seleksi fitur, dengan split dataset 90:10 dan 80:20. Berdasarkan pengujian yang telah dilakukan, didapatkan bahwa algoritma *k-nearest neighbor* yang diterapkan untuk klasifikasi pada data penyakit jantung koroner dengan seleksi fitur menggunakan *information gain* mampu meningkatkan akurasi yang dihasilkan, dengan hasil terbaik yaitu dari klasifikasi dengan seleksi fitur dengan *split* dataset 90:10 dan nilai  $K=5$  sebesar 90% menggunakan 4 fitur hasil seleksi, yaitu *thal*, *cp*, *ca*, dan *oldpeak*. Pada klasifikasi dengan nilai  $k$  serta *split* dataset yang sama, terjadi peningkatan akurasi sebesar 10% setelah diterapkan seleksi fitur. Pada pengujian dengan split dataset 80:20 menghasilkan akurasi tertinggi sebesar 86,67% dari klasifikasi yang dilakukan menggunakan nilai  $K=15$  dan 3 fitur hasil seleksi, yaitu *thal*, *cp*, dan *ca*.

**Kata kunci:** Penyakit Jantung Koroner, Klasifikasi, Seleksi Fitur, *K-Nearest Neighbor*, *Information Gain*