

## BAB 6

### KESIMPULAN DAN SARAN

Bab ini merangkum seluruh perjalanan penelitian tugas akhir, memberikan kesimpulan terhadap temuan-temuan fundamental, serta menawarkan saran-saran konstruktif untuk pengembangan penelitian ini di masa mendatang. Secara keseluruhan, bab ini menutup perjalanan penelitian dengan memberikan gambaran tentang kontribusi penelitian ini terhadap pemahaman penyakit jantung serta memberikan pijakan untuk penelitian lanjutan dalam bidang ini.

#### 6.1 Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan pada tugas akhir ini, berikut ini adalah beberapa kesimpulan yang dapat diambil:

- Berdasarkan hasil analisis dalam penelitian ini, dapat disimpulkan bahwa faktor umur, jenis kelamin, lokasi nyeri pada dada, nyeri dada pada saat bergerak, nyeri pada dada hilang setelah berhenti bergerak, tipe nyeri dada, tekanan darah sistolik istirahat, riwayat darah tinggi, kadar serum kolesterol, gula darah, hasil EKG, asupan oksigen, adanya angina, gradien EKG pada saat istirahat, dan gradien EKG pada titik tertinggi menunjukkan adanya hubungan dengan ketepatan prediksi penyakit jantung. Eksplorasi data dan analisis model memberikan gambaran bahwa setiap faktor tersebut memiliki peran yang berguna dalam menentukan risiko atau kecenderungan seseorang untuk mengalami penyakit jantung. Dengan mengetahui dan memahami keterkaitan antara faktor tersebut dengan penyakit jantung, dapat membantu dalam pengembangan strategi pencegahan dan diagnosis dini secara lebih efektif.
- Tujuan dari penelitian ini adalah untuk menerapkan tahapan *data mining* pada studi kasus prediksi penyakit jantung. Dalam konteks *data mining*, penelitian ini melibatkan pemahaman terkait penyakit jantung, pengumpulan dan eksplorasi data, penerapan metode statistika, serta penggunaan algoritma klasifikasi untuk memprediksi keberadaan penyakit jantung. Kemudian, tahapan eksplorasi dan penyiapan data digunakan untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik. Proses penyiapan data melibatkan seleksi fitur, transformasi data, serta penerapan teknik seperti *one-hot encoding* dan *train-test split*. Peranan *domain expert* sangat berguna dalam penentuan (seleksi) fitur yang akan dipakai, yang sangat berpengaruh terhadap hasilnya. Evaluasi dan analisis model, yang merupakan bagian dari proses *data mining*, mengacu pada pengujian dan pengevaluasian berbagai model klasifikasi, termasuk eksperimen dengan *hyperparameter* dan metrik evaluasi. Langkah-langkah *data mining* ini bertujuan untuk menghasilkan model yang terbaik dan siap untuk dilibatkan dalam proses prediksi.
- Dengan menggunakan algoritma klasifikasi, dapat dibuat model sehingga dapat dilakukan prediksi. Pada tahap evaluasi dan analisis model, dilakukan evaluasi dengan menggunakan empat algoritma klasifikasi yaitu *Naive Bayes*, *Support Vector Machine*, *Decision Tree* dan *Random Forest*. Evaluasi tersebut dilakukan dengan *hyperparameter default* dan didapatkanlah algoritma terbaik pada kasus ini yaitu *Random Forest*. Setelah itu dilanjutkan dengan eksplorasi *hyperparameter Random Forest* untuk mendapatkan hasil *false negative* paling akurat untuk menghindari adanya kesalahan klasifikasi yang fatal. *False negative* sangat

bermanfaat karena merupakan jumlah dari pasien yang didiagnosa memiliki penyakit jantung dan diklasifikasi sebagai tidak memiliki penyakit jantung. *Cross-validation* juga digunakan untuk memvalidasi keandalan model. Pada akhirnya model dengan fitur pilihan dokter memiliki tingkat *recall* tertinggi dan *false negative* terendah. Hasil pada *confusion matrix* didapatkan bahwa *true negative* bernilai 96, *true positive* bernilai 128, *false negative* bernilai 4 dan *false positive* bernilai 66. Sedangkan pada metrik evaluasi didapatkan bahwa akurasi bernilai 0.761, presisi bernilai 0.659, dan *recall* bernilai 0.969. Tahap ini mencakup evaluasi model khusus untuk fitur "merokok", mengidentifikasi dampak variabel ini terhadap hasil prediksi.

- Proses perancangan *website* aplikasi prediksi penyakit jantung dimulai dengan identifikasi fitur-fitur utama yang akan dimiliki oleh website. Fitur-fitur tersebut mencakup kemampuan pengguna untuk memasukkan data untuk diklasifikasikan (dimasukan kedalam model klasifikasi), lalu melihat hasil prediksi yang dihasilkan oleh model.

## 6.2 Saran

Berdasarkan penelitian yang sudah dilakukan, terdapat beberapa saran pengembangan untuk penelitian lanjutan yaitu diantaranya:

- Penelitian dapat dikembangkan dengan menggunakan data terbaru dengan jumlah data yang lebih banyak.
- Sebaiknya data penelitian ini menggunakan data dengan demografi yang sama dengan demografi dimana aplikasi ini dirilis.
- Penelitian dapat dilakukan dengan menggunakan *deep learning* atau model *ensemble* lain seperti *boosted random forest* untuk melakukan klasifikasi penyakit jantung.
- Dalam mempermudah penggunaannya melakukan prediksi, fitur yang digunakan seharusnya lebih sederhana (tidak perlu hasil lab) dan tidak terlalu banyak.

## DAFTAR REFERENSI

- [1] Janosi, A., Steinbrunn, W., Pfisterer, M., dan Detrano, R. Heart disease data set. <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
- [2] Zhang, K., Song, W., Li, D., dan Jin, X. (2017) Apigenin in the regulation of cholesterol metabolism and protection of blood vessels. *Experimental and Therapeutic Medicine*, **13(5)**, 1719–1724.
- [3] Fentem, P. H. (1994) Benefit of exercise in health and disease. *ABC of Sports Medicine*, **308**, 1291–1295.
- [4] Jeffrey B. Lakier, M. (1992) Smoking and cardiovascular disease. *The American Journal of Medicine*, **93**, S8–S12. The Effects of Cigarette Smoking: A Global Perspective.
- [5] Sher, Y., Lolak, S., dan Maldonado, J. R. (2010) The impact of depression in heart disease. *Springer Science+Business Media*, **12**, 255–264.
- [6] Louridi, N., Douzi, S., dan Ouahidi, B. E. (2021) Machine learning-based identification of patients with a cardiovascular defect. *Journal of Big Data*, **8**, 133.
- [7] Dangare, C. S. dan Apte, S. S. (2012) A data mining approach for prediction of heart disease using neural networks. *International Journal Of Computer Engineering & Technology (IJCET)*, **3**, 30–40.
- [8] J. Mackay, G. M. (2004) *The Atlas of Heart Disease and Stroke*. World Health Organization, 6-7 Old Steine, Brighton BN1 1EJ, UK.
- [9] Ni, H., Morotti, S., dan Grandi, E. (2018) A heart for diversity: Simulating variability in cardiac arrhythmia research. *Frontiers in Physiology*, **9**, 1–19.
- [10] Allison L. Cirino MS, C. dan Carolyn Y. Ho, M. (2013) Genetic testing for inherited heart disease. *Circulation*, **128**, e4–e8.
- [11] Newbold, P., Carlson, W. L., dan Thorne, B. M. (2013) Using numerical measures to describe data. Bagian dari Yagan, S. (ed.), *Statistics for Business and Economics*. Pearson Education Limited, Harlow.
- [12] Mendenhall, W., Beaver, R. J., dan Beaver, B. M. (2018) Large-sample estimation. *Introduction to Probability and Statistics*. Cengage Learning, Boston.
- [13] Ananda, R. dan Fadhli, M. (2018) *Statistik Pendidikan Teori Dan Praktik Dalam Pendidikan*. CV. Widya Puspita, Medan, Indonesia.
- [14] William Mendenhall, I., Beaver, R. J., dan Beaver, B. M. (2018) *Introduction to Probability and Statistics, Metric Version*, 15th edition. Cengage, United States of America.
- [15] Han, J., Kamber, M., dan Pei, J. (2012) *Data Mining Concepts and Techniques*, 3rd edition. Morgan Kaufmann, Amerika.

- 
- [16] Baker, R. S. (2010) Data mining for education. Technical report. Carnegie Mellon University, Amerika.
- [17] Lenzerini, M. (2002) Data integration: A theoretical perspective. *Data Integration: A Theoretical Perspective*, New York, NY, USA, April PODS '02 233–246. Association for Computing Machinery.
- [18] Beniwal, S. dan Arora, J. (2012) Classification and feature selection techniques in data mining. *International Journal of Engineering Research and Technology*, **1**.
- [19] Brodlie, K., Osorio, R. S. A., dan Lopes, A. (2012) A review of uncertainty in data visualization. *White Rose Research Online*, **12**, 81–110.
- [20] Chen, H., Hu, S., Hua, R., dan Zhao, X. (2021) Improved naive bayes classification algorithm for traffic risk management. *EURASIP*, **30**, 1–12.
- [21] Gold, C. (2018) Model selection for support vector machine classification. *Computation and Neural Systems*, **1**, 1–38.
- [22] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning*. Springer New York Heidelberg Dordrecht London, New York, USA.
- [23] Mohan, S., Thirumalai, C., dan Srivastava, G. (2019) Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, **7**, 81542–81554.
- [24] Morgan, T. O., Anderson, A. I., dan MacInnis, R. J. (2001) Ace inhibitors, beta-blockers, calcium blockers, and diuretics for the control of systolic hypertension. *American Journal of Hypertension*, **14**, 241–247.