

Prediksi Penyakit Jantung Dengan Menggunakan Algoritma XgBoost dan Randomized Search Optimizer

TUGAS AKHIR

Sebagai syarat untuk memperoleh gelar sarjana S-1 di Program Studi Informatika, Jurusan
Teknik Informatika, Fakultas Teknik Industri, Universitas Pembangunan Nasional
“Veteran” Yogyakarta



Disusun oleh:

Reo Sahobby

123170067

**PROGRAM STUDI INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
YOGYAKARTA
2020**

BAB I

PENDAHULUAN

1.1.Latar Belakang

Dunia kesehatan akhir-akhir ini sedang ramai dibicarakan karena kemunculan virus baru di tahun 2019 yang bernama *corona virus* atau sering disebut dengan istilah covid 19. Virus tersebut menyerang pada bagian pernafasan atau paru-paru manusia. Namun, selain virus yang sedang *booming* tersebut, kita juga harus memperdulikan tentang kesehatan jantung yang tidak kalah pentingnya. Jantung merupakan organ dalam manusia yang fungsinya sangatlah penting yaitu untuk mengedarkan darah yang berisi oksigen dan nutrisi ke seluruh tubuh manusia dan untuk mengangkut sisa hasil metabolisme tubuh, sehingga tubuh dapat bekerja dengan optimal. Akan sangat fatal apabila di dalam organ jantung terdapat gangguan, seperti penyumbatan pembuluh darah dan lain-lain. Sehingga menyebabkan jantung tidak dapat bekerja dan dapat menyebabkan kematian.

Berdasarkan data dari WHO terdapat sebanyak 7,3 juta penduduk di seluruh dunia meninggal karena penyakit jantung. Penyakit jantung adalah penyakit yang menyerang pada organ jantung yang berkaitan dengan pembuluh darah, contohnya adalah pembuluh darah di organ jantung yang tersumbat. Penyakit ini menyerang pada pembuluh darah arteri karena terjadi proses *arterosklerosis* pada dinding arteri yang menyebabkan penyempitan (Marleni & Alhabib, 2017). Penyakit jantung juga bisa disebut dengan istilah *sudden death* (Widyastuti et al., 2014). Karena penyakit jantung tersebut sering kali tidak menimbulkan gejala, namun tiba-tiba pembuluh darah di jantung yang tersumbat tidak dapat memompa darah dan menyalurkannya ke seluruh tubuh, sehingga dapat menyebabkan kematian.

Penyebab penyakit jantung dapat berupa beberapa faktor kebiasaan hidup, berbagai penelitian sudah banyak dilakukan untuk menemukan penyebab dari penyakit jantung tersebut. Diantaranya adalah kebiasaan merokok, pola makan yang tidak sehat, jarang berolahraga, dan lain-lain (Marleni & Alhabib, 2017). Dan tentunya kebiasaan dan faktor penyebab penyakit jantung tersebut sangat bergantung sesuai dengan umur, jenis kelamin, kondisi geografis atau tempat tinggal, tingkat kolesterol, obesitas dan kecenderungan *stress* seseorang tersebut (Anwar, 2004). Sedangkan menurut (Zulaekah et al., 2009), faktor-faktor penyebab terjadinya penyakit jantung juga dipengaruhi oleh kondisi tubuh dan nutrisi seseorang, seperti asupan lemak yang tinggi dan kurangnya tubuh melakukan aktivitas fisik seperti olahraga, sehingga jantung tidak terbiasa dengan aktivitas tubuh yang berat. Kadar kolesterol darah yang tinggi dipengaruhi oleh kebiasaan mengonsumsi makanan yang berlemak, semakin banyak mengonsumsi makanan yang berlemak, peluang untuk menaikkan kadar kolesterol di dalam darah akan semakin tinggi dan akan menurunkan kadar *high density lipoprotein*. Kandungan HDL yang rendah di dalam darah akan mempengaruhi rasio total kolesterol darah dan HDL, semakin tinggi angka rasio total kolesterol dan HDL, maka akan semakin tinggi pula risiko terjadinya penyakit jantung (Zulaekah et al., 2009).

Gejala penyakit jantung yang sering ditemui adalah penderita terkadang merasa sesak napas, kondisi fisik penderita yang mudah lelah, penderita mungkin saja

mengalami gangguan seksual, dan penderita sering merasakan nyeri dada (Nuraeni, 2016). Selain itu seseorang yang menderita penyakit jantung juga memiliki gejala non fisik seperti sering merasa cemas, ketakutan berlebihan, dan sering merasakan depresi. Namun, selain penderita penyakit jantung yang dapat merasakan gejala dari penyakit yang dialaminya, terdapat juga penderita penyakit jantung yang tidak merasakan gejala apa-apa. Selama 50 tahun terakhir, semakin banyak penderita penyakit jantung koroner yang penderitanya tidak merasakan gejala apa-apa, baik gejala fisik maupun gejala non fisik (Zahrawardani et al., 2013). Kondisi tersebut terbilang lebih berbahaya daripada kondisi penderita penyakit jantung yang dapat merasakan gejala, karena penderita tidak akan menyadari kondisi tubuh mereka dan tidak melakukan pencegahan penyakit jantung yang sedang dialami. Karena menurut jurnal (Indrawati, 2014), terdapat hubungan antara kesadaran dan pengetahuan tentang penyakit jantung dan kondisi diri sendiri untuk dapat melakukan upaya pencegahan penyakit jantung.

Di Indonesia sendiri, penyakit jantung sering kali tidak dihiraukan oleh masyarakat, masyarakat masih belum terlalu sadar untuk mengubah pola hidup mereka menjadi pola hidup yang lebih sehat. Padahal jika masyarakat tahu memiliki pengetahuan yang cukup tentang penyakit jantung koroner dan faktor risikonya, maka akan mudah untuk melakukan pencegahan penyakit tersebut (Zahrawardani et al., 2013). Angka kematian karena penyakit jantung di Indonesia meningkat, dari yang tadinya sebesar 41,7% pada tahun 1995, menjadi sebesar 59,5% pada tahun 2007 (Depkes RI, 2009). Di rumah sakit Siti Khadijah di Palembang, tercatat pada tahun 2015 jumlah pasien di poli jantung mengalami peningkatan pasien mencapai 354 pasien, dan pada tahun 2016 terdapat 274 pasien penderita penyakit jantung di rumah sakit tersebut (Marleni & Alhabib, 2017). Di Jawa Tengah, dari laporan puskesmas di daerah tersebut terdapat 26,38 kasus penyakit jantung dari 1.000 penduduk di daerah tersebut (Zahrawardani et al., 2013). Oleh karena itu, di Indonesia penyakit jantung juga harus tetap diwaspadai.

Proses pendeteksian apakah seseorang tersebut terkena penyakit jantung atau tidak dapat dilakukan dengan cara melakukan konsultasi kepada dokter spesialis jantung yang nantinya akan dilakukan pemeriksaan laboratorium dan dikonsultasikan oleh dokter spesialis jantung (Wibisono & Fahrurrozi, 2019). Namun cara tersebut tidaklah efektif, selain memakan waktu yang lama karena proses pemeriksaan, menunggu hasil pemeriksaan, dan konsultasi tentunya memakan waktu yang lama, juga karena memakan biaya yang cukup tinggi. Oleh karena itu perlu dilakukan pendeteksian penyakit jantung secara digital supaya dapat meningkatkan efektifitas kerja. Banyak penelitian yang sudah menciptakan pendeteksian penyakit jantung secara digital, yaitu dengan menggunakan data-data hasil rekam jantung yang ada, yang nantinya dipelajari pola-pola datanya dan akan menghasilkan prediksi, berdasarkan data tersebut apakah seseorang ini berpotensi menderita penyakit jantung atau tidak. Teknik yang digunakan dalam melakukan prediksi tersebut dinamakan teknik klasifikasi. Klasifikasi adalah jenis analisis data yang digunakan untuk memprediksi label kelas dari data tersebut (Annisa, 2019).

Dalam klasifikasi terdapat beberapa teknik atau algoritma yang dapat dilakukan untuk mengerjakan klasifikasi, diantaranya adalah dapat menggunakan algoritma KNN, algoritma *Naïve Bayes*, algoritma *Support Vector Machine*, algoritma *Decision Tree*,

algoritma *Random Forest*, dan lain-lain. Dalam kasus prediksi penyakit jantung ini, penelitian-penelitian sebelumnya telah banyak dilakukan dengan menggunakan berbagai algoritma klasifikasi yang ada. Diantaranya adalah penelitian yang dilakukan oleh Retnasari dan Rahmawati, yang melakukan penelitian dengan menggunakan algoritma *Naïve Bayes* dan algoritma C4.5. Penelitian tersebut dilakukan dengan menggunakan 270 data yang bersumber dari *UCI Machine Learning Repository* dengan jumlah *features* yaitu 13, penelitian tersebut dilakukan dengan menggunakan *rapid mider* dan *confusion matrix* untuk menghitung akurasi masing-masing algoritma. Hasil dari penelitian yang dilakukan tersebut menunjukkan bahwa algoritma *Naïve Bayes* lebih baik dengan mendapatkan nilai akurasi sebesar 86,67% dan algoritma C4.5 mendapat akurasi sebesar 83,70% (Retnasari & Rahmawati, 2017). Penelitian selanjutnya yang dilakukan oleh Ardea dan Achmad, penelitian tersebut dilakukan untuk mencari algoritma terbaik dengan cara membandingkan masing-masing hasil dari algoritma tersebut. Algoritma yang dibandingkan di dalam penelitian tersebut adalah algoritma *Naïve Bayes*, algoritma *Random Forest*, algoritma *Decision Tree*, dan algoritma *K-Nearest Neighbor*. Hasil dari penelitian tersebut untuk masing-masing algoritma dihitung dengan menggunakan *confusion matrix* dan didapat hasil akurasi untuk masing-masing algoritma sebagai berikut. Algoritma *Random Forest* memiliki nilai akurasi tertinggi dengan 85,67%, kemudian algoritma *Naïve Bayes* dan algoritma *Decision Tree* memiliki nilai akurasi yang sama dengan nilai akurasi 80,33%, dan algoritma *K-Nearest Neighbor* memiliki nilai akurasi paling rendah yaitu 69,67%. Dengan hasil tersebut, algoritma yang terbaik adalah algoritma *Random Forest* (Wibisono & Fahrurrozi, 2019). Selanjutnya penelitian yang dilakukan oleh Erwin Prasetyo dan Budi Prasetyo, penelitian tersebut dilakukan dengan menerapkan teknik *bagging* pada algoritma C4.5 untuk melihat apakah teknik *bagging* dapat meningkatkan akurasi dari model klasifikasi yang dibuat. Data yang digunakan dalam penelitian tersebut adalah data *Heart Disease* yang diambil dari *UCI Machine Learning* sejumlah 300 data. Hasil dari penelitian tersebut membuktikan bahwa penerapan teknik *bagging* pada algoritma C4.5 dapat meningkatkan akurasi model yang dibuat dengan kenaikan yaitu 8,86% dengan hasil akurasi algoritma C4.5 sebesar 72,98% dan akurasi algoritma C4.5 yang dikombinasikan dengan teknik *bagging* adalah 81,84% (Prasetyo & Prasetyo, 2020).

Dari berbagai macam algoritma yang sudah digunakan dalam penelitian-penelitian sebelumnya, tentunya masing-masing algoritma memiliki kelebihan dan kelemahan. Sebenarnya beberapa metode yang sudah digunakan dalam penelitian sebelumnya sudah menghasilkan nilai akurasi model yang baik, namun seringkali apabila model memiliki nilai akurasi yang terlalu tinggi, maka model akan terlalu fokus mempelajari data *training* sehingga nilai akurasi model sangatlah tinggi, namun pada saat dilakukan prediksi menggunakan data real yang belum pernah ditemui oleh model, hasil prediksi seringkali tidak tepat, kondisi tersebut dinamakan *overfitting*. Untuk mencegah terjadinya *overfitting* pada model yang sudah dibuat, perlu diterapkan teknik *regularization* untuk mengurangi *overfitting* pada model. Pada penelitian ini algoritma yang dipilih adalah menggunakan algoritma XgBoost. Algoritma XgBoost adalah algoritma *gradien boosting* yang dibuat dengan *tree base* yang dapat membuat *boosted*

tree secara efisien dan dapat dikerjakan secara paralel (Muslim et al., 2020). Algoritma tersebut sudah memiliki operasi *regularization*, sehingga algoritma tersebut dapat mencegah terjadinya *overfitting*. Tujuan dari penelitian ini adalah untuk mengetahui apakah pendeteksian penyakit jantung dapat dilakukan dengan menggunakan algoritma XgBoost, dan bagaimana hasil akurasi dari model yang dibuat dengan algoritma tersebut.

1.2.Rumusan Masalah

Sesuai dengan uraian latar belakang yang sudah dijelaskan di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

- a. Penerapan algoritma klasifikasi XgBoost pada kasus prediksi penyakit jantung.
- b. performa model yang dibuat dengan algoritma XgBoost dalam menyelesaikan permasalahan prediksi penyakit jantung.

1.3.Batasan Masalah

Batasan masalah yang ada di dalam penelitian ini adalah sebagai berikut:

- a. Data yang digunakan dalam penelitian ini adalah data *Heart Disease* yang diambil dari *UCI Machine Learning*.
- b. Metode algoritma klasifikasi yang digunakan dalam penelitian ini adalah menggunakan algoritma XgBoost.

1.4.Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

- a. Menerapkan algoritma klasifikasi XgBoost dalam kasus prediksi penyakit jantung.
- b. Mengetahui performa model yang dibuat dengan algoritma XgBoost dalam menyelesaikan permasalahan prediksi penyakit jantung.

1.5.Manfaat Penelitian

Manfaat yang ingin dicapai dari penelitian yang ingin dilakukan ini adalah sebagai berikut:

- a. Manfaat bagi peneliti, peneliti dapat menerapkan ilmu yang didapat selama perkuliahan, dapat mengimplementasikan algoritma XgBoost untuk menyelesaikan permasalahan prediksi penyakit jantung.
- b. Manfaat penelitian ini bagi industri kesehatan adalah, untuk kedepannya diharapkan mampu membantu dalam proses pendeteksian penyakit jantung supaya lebih efektif.

1.6.Tahapan Penelitian

Pada penelitian yang akan dilakukan ini, terdapat beberapa tahapan yang akan dilakukan yaitu sebagai berikut:

- a. Study Literatur

Tahap pertama yang dilakukan dalam penelitian ini adalah melakukan *study literature*. *Study literature* dilakukan untuk mencari referensi, penelitian sebelumnya, data yang akan digunakan, dan lain-lain. Study literature dapat dicari dari jurnal-jurnal yang membahas penelitian serupa.

- b. Pengumpulan Data

Tahap selanjutnya adalah melakukan pengumpulan data, data yang akan digunakan dalam penelitian ini adalah data sekunder, yaitu data *Heart Disease* yang bersumber dari *UCI Machine Learning Repository*.

c. Analisis Sistem

Selanjutnya adalah melakukan analisis kebutuhan perangkat lunak yang ada dibuat di dalam penelitian ini.

d. Pembuatan Model *Machine Learning*

Tahap selanjutnya adalah pembuatan model *machine learning*. Pada tahap ini dilakukan pembuatan model prediksi yang menggunakan algoritma dan teknik yang sudah dipilih.

e. Pengujian dan Evaluasi Model

Setelah model prediksi *machine learning* sudah dibuat, tahap selanjutnya adalah memastikan model yang dibuat memiliki performa yang baik dalam menangani data. Apabila model dirasa belum maksimal, dapat dilakukan pembuatan model ulang dengan *hyper parameter* yang berbeda dan dilakukan pengujian lagi, diharapkan mendapat peningkatan performa.

f. Implementasi Perangkat Lunak

Selanjutnya, setelah model yang dibuat dirasa memiliki performa yang bagus, model tersebut diimplementasikan dalam bentuk perangkat lunak yang bisa digunakan oleh pengguna. Dalam pembuatan perangkat lunak ini, menggunakan metodologi *waterfall*.

g. Pengujian dan Evaluasi Perangkat Lunak

Setelah perangkat lunak selesai dibuat, dilakukan pengujian perangkat lunak untuk memastikan perangkat lunak yang dibuat berjalan normal tanpa ada kendala.

h. Kesimpulan dan Saran

Setelah semua tahap dilakukan, didapatkan kesimpulan dari penelitian yang sudah dilakukan tentang bagaimana performa algoritma XgBoost dalam menangani kasus permasalahan yang dipilih.

1.7. Sistematika Penulisan

Penelitian ini disusun berdasarkan sistematika penulisan yang terdiri dari 5 bab yang terdiri dari:

BAB 1 PENDAHULUAN

Pada BAB I ini, membahas latar belakang penelitian ini dilakukan, rumusan masalah yang ada di dalam penelitian ini, batasan masalah, tujuan, dan manfaat penelitian ini dilakukan, serta sistematika penulisan laporan mengenai penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Dalam BAB II ini, berisi landasan teori mengenai obyek penelitian dan metode yang akan dilakukan di dalam penelitian ini, kemudian juga membahas penelitian-penelitian serupa yang sudah dilakukan sehingga menjadi referensi penulis dalam mengadakan melakukan penelitian ini.

BAB III METODE PENELITIAN

Pada BAB III ini berisi penjelasan tentang metode yang akan digunakan oleh penulis di dalam melakukan penelitian ini. Metode-metode yang dipilih nantinya akan digunakan untuk menyelesaikan permasalahan pada kasus yang sedang diteliti, yaitu prediksi penyakit jantung.

BAB IV HASIL DAN PEMBAHASAN

Pada BAB IV ini, berisi pemaparan dan penjelasan hasil dari tahapan demi tahapan penelitian yang sudah dilakukan oleh penulis dengan menggunakan metode yang sudah dijelaskan pada bab sebelumnya. Penjelasan hasil penelitian akan berisi evaluasi performa model yang sudah dibuat dengan menggunakan algoritma yang dipilih.

BAB V PENUTUP

Bab ini akan berisi kesimpulan hasil dari penelitian yang sudah dilakukan oleh penulis. Kemudian penulis juga menambahkan kekurangan dari penelitian yang sudah dilakukan ditambahkan dengan saran yang bisa dilakukan pada penelitian yang akan datang, dapat berupa saran perbaikan data ataupun saran mengenai perbaikan metode supaya penelitian yang akan datang dapat menghasilkan hasil yang lebih maksimal.

DAFTAR PUSTAKA

- Annisa, R. (2019). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *Jurnal Teknik Informatika Kaputama (JTIK)*, 3(1), 22–28. <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- Anwar, T. B. (2004). Faktor Risiko Penyakit Jantung Koroner. *E-USU Repository*, 01(Medan), 1–15. <http://repository.usu.ac.id/bitstream/handle/123456789/3472/gizi-bahri4.pdf?sequence=1>
- Indrawati, L. (2014). Hubungan Antara Pengetahuan, Sikap, Persepsi, Motivasi, Dukungan Keluarga dan Sumber Informasi Pasien Penyakit Jantung Koroner dengan Tindakan Pencegahan Sekunder Faktor Risiko (Studi Kasus di RSPAD Gatot Soebroto Jakarta). *Jurnal Ilmiah Widya*, 2(3), 30–36.
- Marleni, L., & Alhabib, A. (2017). Faktor Risiko Penyakit Jantung Koroner di RSI SITI Khadijah Palembang. *Jurnal Kesehatan*, 8(3), 478. <https://doi.org/10.26630/jk.v8i3.663>
- Muslim, I., Karo, K., Informatika, F., & Telkom, U. (2020). *Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan*. 1(1), 10–16.
- Nuraeni, A. (2016). Faktor yang Memengaruhi Kualitas Hidup Pasien dengan Penyakit Jantung Koroner. *Jurnal Keperawatan Padjadjaran*, v4(n2), 107–116. <https://doi.org/10.24198/jkp.v4n2.1>
- Prasetyo, E., & Prasetyo, B. (2020). *PENINGKATAN AKURASI KLASIFIKASI ALGORITMA C4.5 MENGGUNAKAN TEKNIK BAGGING PADA DIAGNOSIS PENYAKIT JANTUNG*. 7(5), 1035–1040. <https://doi.org/10.25126/jtiik.202072379>
- Retnasari, T., & Rahmawati, E. (2017). Diagnosa Prediksi Penyakit Jantung Dengan Model Algoritma Naïve Bayes Dan Algoritma C4.5. *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, 7–12.
- Wibisono, A. B., & Fahrurrozi, A. (2019). Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170. <https://doi.org/10.35760/tr.2019.v24i3.2393>
- Widyastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particle Swarm Optimization Untuk Deteksi Penyakit Jantung. *Nature Methods*, 7(1), 34. <https://doi.org/10.1038/nmeth.f.284>
- Zahrawardani, D., Herlambang, K. S., & Anggraheny, H. D. (2013). Analisis Faktor Risiko Kejadian Penyakit Jantung Koroner di RSUP Dr Kariadi Semarang. *Jurnal Kedokteran Muhammadiyah*, 1(3), 13. <http://jurnal.unimus.ac.id/index.php/kedokteran/article/view/1341>
- Zulaekah, S., Rahmawati, A. C., & Rahmawaty, S. (2009). Aktivitas Fisik dan Rasio Kolesterol (HDL) pada Penderita Penyakit Jantung Koroner di Poliklinik Jantung RSUD Dr Moewardi Surakarta. *Jurnal Kesehatan*, 2(1), 11–18.