

# **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  
2021**

# 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* (Widiastuti 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 (Karo, 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.

## BAB II

### KAJIAN LITERATUR

#### 2.1. Tinjauan Studi

Penyakit jantung menjadi tantangan tersendiri di dunia bagi industri pelayanan kesehatan saat ini (Prasetyo & Prasetyo, 2020). Dalam satu dekade terakhir, penyakit ini, merupakan penyakit yang paling utama menjadi penyebab kematian di seluruh dunia (Jothikumar & Siva Balan, 2016). Maka, diperlukan sistem yang dapat menangani pendeteksian penyakit jantung pada penderita secara akurat dan dengan biaya yang terjangkau (Wibisono & Fahrurrozi, 2019). Oleh karena itu, penelitian tentang penyakit jantung telah banyak dilakukan, penelitian tersebut dilakukan dengan menggunakan beberapa teknik dan algoritma untuk mendapatkan hasil prediksi yang semaksimal mungkin. Salah satu penelitian yang sudah dilakukan adalah penelitian yang dilakukan oleh Erwin Prasetyo dan Budi Prasetyo (Prasetyo & Prasetyo, 2020), penelitian tersebut dilakukan dengan menggunakan datasets *heart disease* dengan jumlah data sebanyak 303 data, dengan jumlah kolom yang digunakan adalah 13 kolom. Penelitian tersebut bertujuan untuk membandingkan performa algoritma C4.5 yang dikombinasikan dengan teknik *bagging*, dan algoritma C4.5 murni tanpa penambahan teknik apapun. Penelitian tersebut dilakukan dengan menggunakan bahasa pemrograman python dan dengan bantuan beberapa perpustakaan seperti *numpy*, *sklearn*, dan *pandas*. Untuk mengukur performa algoritma yang digunakan, penelitian tersebut menggunakan *confusion matrix*. Penelitian tersebut dilakukan dengan menggunakan algoritma C4.5 yang dikombinasikan dengan teknik *bagging*, untuk melakukan validasi hasil dari performa algoritma yang diuji dalam penelitian tersebut menggunakan *k-fold cross validation* dengan nilai  $k=10$ . Hasil dari penelitian yang dilakukan tersebut ditampilkan menggunakan tabel *confusion matrix*, dengan hasil akurasi dari algoritma C4.5 adalah 72,98% dan akurasi algoritma C4.5 yang dikombinasikan dengan teknik *bagging* adalah 81,84%. Dengan hasil tersebut, didapatkan kesimpulan bahwa teknik *bagging* yang dilakukan dengan algoritma C4.5 dapat meningkatkan akurasi model, dalam kasus tersebut meningkatkan akurasi sebanyak 8,86%.

Selanjutnya, penelitian yang dilakukan oleh Pandito Dewa Putra dan Dian Palupi Rini pada tahun 2019 (Putra & Rini, 2019). Penelitian tersebut dilakukan untuk mengetahui perbandingan beberapa algoritma klasifikasi seperti *naïve bayes*, *support vector machine*, C4.5, *logistic regression*, dan *back propagation* dalam melakukan klasifikasi untuk melakukan prediksi penyakit jantung. Datasets yang digunakan dalam penelitian tersebut adalah *Statelogs Heart Disease Datasets* yang berasal dari UCI *machine learning* yang dapat diunduh secara online. Datasets tersebut berjumlah 270 data dengan kolom atau *features* yang digunakan adalah 13 kolom yang berisi parameter seperti *age*, *sex*, *chest pain*, dan lain-lain. Untuk melakukan validasi, dalam penelitian tersebut menggunakan *cross validation* dan nilai yang akan dihitung untuk mengukur performa algoritma adalah akurasi, presisi, dan *recall*. Tahapan yang dilakukan di dalam penelitian tersebut meliputi input datasets, melakukan *preprocessing* terhadap data yang digunakan, pembuatan model klasifikasi yang dilakukan menggunakan algoritma yang ingin dibandingkan, proses validasi menggunakan *cross validation*, dan pengukuran



performa algoritma. Hasil dari penelitian tersebut adalah akurasi dari algoritma *naïve bayes* mendapatkan nilai tertinggi yaitu 84,07%. Kemudian algoritma yang memiliki nilai presisi tertinggi adalah algoritma *naïve bayes* dengan nilai presisi adalah 86,16%. Selanjutnya untuk pengukuran *recall*, algoritma yang memiliki *recall* paling tinggi adalah algoritma *support vector machine* dengan nilai *recall* mencapai 94,67%. Dari hasil penelitian yang dilakukan tersebut didapatkan kesimpulan bahwa algoritma *naïve bayes* tercatat memberikan hasil performa yang lebih baik dari algoritma lainnya baik dari segi akurasi dan presisi algoritma.

Selanjutnya adalah penelitian yang dilakukan oleh Ardea Bagas Wibisono dan Achmad Fahrurrozi pada tahun 2019 (Wibisono & Fahrurrozi, 2019). Penelitian tersebut dilakukan untuk mengimplemantasikan beberapa metode klasifikasi yang sudah ada seperti *Naive Bayes*, *K-Nearest Neighbor*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* untuk kasus pengenalan penyakit jantung koroner. Dengan hasil dari pengukuran yang didapat adalah nilai dari akurasi, *recall*, dan presisi. Datasets yang digunakan di dalam penelitian tersebut adalah *Cleveland Heart Disease* yang berisi data rekam jantung dengan jumlah data adalah 300 data, dengan 14 kolom. Parameter yang ada dalam datasets tersebut berjumlah 13, sesuai dengan kolom yang ada, 1 kolom sebagai data target dari hasil klasifikasi data tersebut. Parameter yang ada di dalam datasets yang digunakan dalam penelitian tersebut seperti *age*, *sex*, *chol*, *fbs*, *restecg*, *thalach*, dan lain-lain. Dalam penelitian tersebut dilakukan beberapa tahapan diantaranya adalah proses *import* dataset yang akan digunakan, kemudian melakukan pembagian data menjadi *data training* dan *data testing*, dengan perbandingan *data training* dan *data testing* adalah 80 banding 20. Kemudian setelah data dibagi menjadi dua, tahapan selanjutnya adalah pembuatan model klasifikasi dengan cara melakukan *training* pada *data training* menggunakan algoritma yang sudah ditentukan dalam penelitian tersebut. Selanjutnya, melakukan pengujian dan validasi model klasifikasi yang sudah dibuat. Proses pengujian dilakukan menggunakan *data testing*, dan untuk melakukan validasi terhadap performa model klasifikasi yang dibuat dilakukan dengan menggunakan *Cross Validation* dengan nilai  $K=5$  yang nantinya akan dilakukan 5 kali percobaan untuk setiap algoritma. Kemudian tahap terakhir adalah melakukan perhitungan untuk mengukur performa algoritma klasifikasi. Dari tahapan yang dilakukan dalam penelitian tersebut, dari hasil pengujian yang dilakukan terhadap beberapa algoritma yang diuji dalam penelitian tersebut mendapatkan hasil bahwa akurasi tertinggi didapatkan pada algoritma *Random Forest* dengan nilai akurasi mencapai 85,67%. Algoritma *Naïve Bayes* dan algoritma *Decision Tree* mendapat hasil akurasi yang sama, yaitu sebesar 80,33%. Algoritma *K-Nearest Neighbor* mendapat akurasi sebesar 69,67%. Dari penelitian tersebut dapat disimpulkan bahwa algoritma *Random Forest* memiliki performa yang paling baik dalam melakukan klasifikasi dengan menggunakan data pasien jantung koroner.

Kemudian, penelitian serupa juga dilakukan oleh Riski Annisa pada tahun 2019 (Annisa, 2019). Penelitian tersebut dilakukan untuk mencari algoritma terbaik dengan cara membandingkan algoritma klasifikasi meliputi *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor*, *Random Forest*, dan *Decision Stump* dengan menggunakan uji

parametrik dengan *t-test*. Datasets yang digunakan untuk melakukan penelitian perbandingan algoritma tersebut adalah datasets laki-laki penderita penyakit jantung yang terdiri dari 8 atribut, datasets tersebut tersedia secara online pada UCI *machine learning*. Dalam penelitian tersebut dilakukan validasi dengan menggunakan *Cross Validation*, dan menggunakan nilai  $K=10$  fold cross validation yang berarti *data training* tersebut akan dipecah menjadi 10 bagian, yang nantinya masing-masing bagian akan menjadi *data testing* secara bergantian sebanyak 10 kali percobaan. Untuk menghitung kinerja algoritma yang sedang dibandingkan, dalam penelitian tersebut pengujian dilakukan dengan menggunakan uji t (*t-test*). Tahapan yang dilakukan di dalam pengujian penelitian tersebut adalah pembagian datasets menjadi dua bagian, yaitu *data training* dan *data testing*. Kemudian diterapkan evaluasi menggunakan AUC atau *Area under Curve*. Sedangkan untuk hasil dari akurasi algoritma yang didapatkan dapat dilihat menggunakan kurva *Receiver Operating Characteristic* (ROC) dan dalam bentuk *confusion matrix*. Kemudian, di dalam penelitian tersebut juga dilakukan uji t atau *t-test* yaitu untuk membandingkan hubungan antara dua variabel, yaitu variabel respon dan variabel *predictor* yang ada di dalam datasets yang digunakan. Hasil dari penelitian tersebut nilai akurasi tertinggi didapatkan dengan menggunakan algoritma *Random Forest* dengan akurasi sebesar 80,38%. Berdasarkan pengukuran menggunakan AUC, algoritma *Random Forest* juga tergolong ke dalam *good classification*. Sedangkan untuk algoritma *K-Nearest Neighbor*, C4.5, dan algoritma *Decision Stump* tergolong ke dalam *fair classification*. Sedangkan untuk hasil *t-test* diketahui algoritma C4.5 dan algoritma *Naïve Bayes* tidak menunjukkan perbedaan yang signifikan, algoritma *Naïve Bayes* tidak menunjukkan perbedaan yang signifikan dengan algoritma *Random Forest* dan *Decision Stump*. Algoritma C4.5 memiliki perbedaan yang signifikan dengan algoritma *Random Forest* dan *Decision Stump*. Algoritma *K-Nearest Neighbor* berbeda secara signifikan dengan algoritma lain yang dibandingkan di dalam penelitian tersebut yang berarti algoritma K-NN kurang baik diimplementasikan dalam kasus dan datasets yang digunakan dalam penelitian tersebut.

Selanjutnya, penelitian yang dilakukan oleh Mei Lestari pada tahun 2014 (Lestari, 2014). Penelitian tersebut dilakukan menggunakan algoritma *K-Nearest Neighbor* (K-NN) yang diterapkan untuk mendeteksi penyakit jantung. Dalam penelitian tersebut datasets yang digunakan adalah data yang bersumber dari UCI yang berjumlah 110 *record* data dengan jumlah atribut yang terdapat dalam datasets adalah sebanyak 13 atribut yang mewakili setiap parameter data tersebut. Seperti *age*, *sex*, *resting blood pressure*, *old peak*, dan lain-lain. Tahap-tahapan yang dilakukan di dalam penelitian tersebut adalah melakukan *import* datasets yang akan digunakan yaitu berjumlah 110 data. Selanjutnya, membagi datasets menjadi dua bagian yaitu *data training*, dan *data testing*, *data training* yang digunakan di dalam penelitian tersebut berjumlah 100 data, dan *data testing* yang digunakan berjumlah 10 data. Kemudian dilanjutkan dengan proses membuat model klasifikasi menggunakan algoritma K-NN dan menentukan nilai k di dalam algoritma K-NN tersebut dengan nilai  $k=9$ . Kemudian, proses training pada *data training* dan proses pengujian menggunakan *confusion matrix* dan kurva ROC untuk mengukur performa algoritma K-NN dalam melakukan deteksi penyakit jantung.

Hasil dari penelitian tersebut adalah algoritma K-NN yang digunakan menggunakan nilai  $k=9$ , maka proses yang terjadi di dalam algoritma tersebut akan mengecek 9 buah tetangga terdekat untuk masing-masing data, yang nantinya akan digunakan untuk menentukan klasifikasi. Berdasarkan hasil pengujian yang dilakukan menggunakan *data testing* yang berjumlah 10 data, akurasi yang didapatkan dari algoritma K-NN melalui penelitian tersebut sebesar 70%. Sedangkan metode pengukuran performa lainnya yang dilakukan di dalam penelitian tersebut adalah menggunakan kurva ROC dan AUC. Dari pengujian yang dilakukan, nilai AUC yang didapatkan dari algoritma tersebut adalah 0.875 yang berarti dapat dikatakan metode algoritma KNN yang dilakukan pada penelitian tersebut tergolong algoritma yang baik untuk melakukan klasifikasi deteksi penyakit jantung.

Penelitian selanjutnya yang masis membahas tentang penyakit jantung adalah penelitian yang dilakukan oleh Abdul Rohman, Vincent Suhartono dan Catur Supriyadi pada tahun 2017 (Rohman et al., 2017). Penelitian tersebut membahas tentang prediksi penyakit jantung yang dilakukan dengan menggunakan algoritma *Decision Tree* atau C4.5 namun dikombinasikan dengan metode *Adaboost*. Tujuan penelitian tersebut melakukan kombinasi dari algoritma dan metode tersebut adalah untuk mengoptimalkan atribut-atribut yang dimiliki oleh datasets yang digunakan, dan diharapkan dengan menerapkan metode *Adaboost* dalam algoritma C4.5 dapat meningkatkan performa algoritma tersebut dalam melakukan klasifikasi. Datasets yang digunakan di dalam penelitian tersebut adalah data hasil gabungan dari datasets Cleveland yang berjumlah 303 data, datasets Statlog yang berjumlah 270 data, dan datasets Hungaria yang berjumlah 294 data. Sehingga jumlah keseluruhan data yang digunakan di dalam penelitian tersebut berjumlah 867 data, dengan rincian data adalah 364 data masuk ke dalam klasifikasi sakit, dan 503 data masuk ke dalam klasifikasi sehat. Semua datasets yang digunakan di dalam penelitian tersebut bersumber dari UCI *machine learning*. Tahapan yang dilakukan di dalam penelitian tersebut sama seperti tahapan penelitian yang dilakukan pada penelitian-penelitian sebelumnya. Mulai dari import data, *preprocessing* data, pembuatan model, dan pengujian dan pengukuran performa algoritma yang digunakan. Pada tahap *preprocessing*, untuk mendapatkan data yang berkualitas dilakukan seleksi data terlebih dahulu. Proses *preprocessing* data mendapatkan hasil akhir yaitu data yang sudah berkualitas dengan jumlah data sebanyak 567 data, dengan data yang tergolong dalam klasifikasi sakit adalah 257 data, dan sebanyak 310 data tergolong ke dalam klasifikasi sehat. Dalam pembuatan model algoritma yang dilakukan pada penelitian tersebut, dilakukan validasi dengan menggunakan *K-Fold Cross Validation* dengan nilai  $k=10$ . Dan dalam pengujian model, pengujian yang dilakukan dalam penelitian tersebut menggunakan kurva *Area Under Curve* (AUC). Hasil penelitian tersebut algoritma C4.5 mendapatkan nilai akurasi sebesar 86,59% dengan nilai AUC adalah 0,957. Sedangkan hasil perhitungan dari algoritma C4.5 yang dikombinasikan dengan metode *Adaboost* mendapat akurasi sebesar 92,24% dan nilai AUC sebesar 0.982. Dengan demikian, dapat disimpulkan bahwa metode *Adaboost* yang dikombinasikan dengan algoritma C4.5 dapat

meningkatkan performa algoritma yang signifikan dibandingkan algoritma C4.5 murni tanpa penambahan metode apapun.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Nur Aeni Widiastuti, Stefanus Santosa, dan Catur Supriyanto pada tahun 2014 (Widiastuti et al., 2014). Penelitian tersebut membahas tentang klasifikasi penyakit jantung yang dilakukan dengan algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*. PSO atau *Particle Swarm Optimization* dipilih dalam penelitian tersebut karena dirasa mudah diterapkan dan terdapat beberapa parameter untuk menyesuaikan. Pada penelitian tersebut dataset yang digunakan berjumlah 300 data, dengan pembagian menjadi *data training* sebanyak 75% dan sebanyak 25% untuk data testing. Tahapan yang dilakukan dalam penelitian tersebut secara umum dibagi menjadi dua, yang pertama adalah tahapan penelitian pembuatan model klasifikasi menggunakan algoritma *Naïve Bayes* tanpa dikombinasikan dengan metode lainnya. Dan yang kedua adalah tahapan penelitian menggunakan algoritma *Naïve Bayes* yang dikombinasikan dengan metode *Particle Swarm Optimization*. Kedua tahapan tersebut nantinya akan dibandingkan hasil performa algoritmanya, dari kedua perbandingan tersebut nantinya dapat diketahui apakah penggunaan PSO yang dikombinasikan dengan algoritma *Naïve Bayes* akan meningkatkan performa model yang dibuat, dan apakah hasil pengukuran model yang dibuat dengan algoritma *Naïve Bayes* yang dikombinasikan dengan PSO akan lebih tinggi performanya. Hasil pengujian dari penelitian tersebut, pengukuran performa model yang dibuat dengan menggunakan algoritma *Naïve bayes* murni tanpa dikombinasikan dengan apapun memiliki akurasi sebesar 82,14% dan nilai AUC sebesar 0,686. Sedangkan hasil pengukuran performa model klasifikasi yang dibuat dengan menggunakan algoritma *Naïve Bayes* yang dikombinasikan dengan PSO mendapat akurasi sebesar 92,86% dengan nilai AUC adalah 0,839. Berdasarkan hasil dari penelitian tersebut, dapat disimpulkan bahwa penambahan metode *Particle Swarm Optimization* dapat meningkatkan performa model klasifikasi yang dibuat, dan dalam kasus tersebut algoritma yang dikombinasikan adalah algoritma *Naïve Bayes*.

Selanjutnya, terdapat penelitian yang dilakukan oleh Dito Putro Utomo dan Mesran pada tahun 2020 (Utomo & Mesran, 2020). Penelitian tersebut dilakukan untuk membandingkan performa dari algoritma C5.0 dan algoritma *Naïve Bayes*, kedua algoritma tersebut dikombinasikan dengan metode *Principal Component Analysis* (PCA) untuk mereduksi atau mengurangi jumlah atribut sehingga tersisa hanya atribut-atribut yang memiliki bobot paling tinggi dan dirasa paling berpengaruh dalam proses klasifikasi nantinya. Tahapan yang dilakukan di dalam penelitian tersebut adalah mencari dataset terlebih dahulu. Dataset yang digunakan di dalam penelitian tersebut adalah dataset yang bersumber dari *UCI machine learning*, datasets tersebut memiliki atribut sebanyak 57 dan memiliki 2 kelas target yaitu *CAD* dan *Normal*. Penelitian ini menggunakan PCA untuk mengurangi jumlah atribut yang dirasa terlalu banyak dan bisa mengurangi efektifitas model dalam melakukan proses *training* dan klasifikasi, hasil dari PCA akan menyisakan 10 atribut yang memiliki bobot paling tinggi. Kemudian, tahapan selanjutnya adalah melakukan *preprocessing* data untuk menghasilkan data yang berkualitas dan siap digunakan untuk tahapan selanjutnya. Selanjutnya, dilakukan proses

pembuatan model dengan menggunakan algoritma C5.0 dan algoritma *Naïve Bayes*. Kemudian data yang digunakan akan dilakukan proses PCA untuk mereduksi jumlah atribut sehingga dapat mengurangi atribut. Tahap terakhir adalah melakukan klasifikasi ulang pada data yang sudah direduksi atributnya menggunakan kedua algoritma yang sama dan melakukan pengujian serta pengukuran algoritma, kemudian membandingkan hasil dari masing-masing pengujian dan menentukan kesimpulan berdasarkan hasil dari penelitian yang dilakukan tersebut. Hasil dari penelitian tersebut adalah pengujian algoritma C5.0 tidak merubah akurasi yaitu sebesar 95,38% baik menggunakan data yang sudah direduksi ataupun data yang belum direduksi menggunakan PCA. Sedangkan pada algoritma *Naïve Bayes*, pengujian menggunakan data yang belum direduksi menghasilkan akurasi sebesar 99,01% dan pengujian menggunakan data yang sudah direduksi mendapatkan akurasi dengan nilai sebesar 98,53%. Dengan demikian dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki performa yang lebih baik, bahkan setelah data direduksi dan hanya menyisakan 10 atribut algoritma tersebut masih dapat melakukan klasifikasi dan mendapatkan nilai akurasi yang tinggi. Hal ini juga didukung bahwa algoritma *Naïve Bayes* tidak memerlukan *rule* seperti algoritma C5.0. Oleh karena itu, algoritma *Naïve Bayes* dapat melakukan klasifikasi lebih baik.

Kemudian penelitian yang dilakukan oleh Dwi Normawati dan Sri Winiarti pada tahun 2017 (Normawati & Winarti, 2017). Penelitian tersebut membahas tentang diagnosis pada data penyakit jantung menggunakan teknik seleksi fitur bernama *Variable Precision Rough Set* (VPRS) yang merupakan pengembangan dari metode *Rough Set*. Dalam penelitian tersebut, juga dilakukan penggabungan metode yaitu metode VPRS dan metode seleksi fitur berbasis medis atau *Motivated Feature Selection* (MFS) agar menghindari atribut yang dianggap penting oleh medis terseleksi apabila hanya menggunakan metode VPRS. Penggabungan dua metode tersebut diharapkan dapat meningkatkan performa model klasifikasi yang dibuat untuk mendiagnosis penyakit jantung. Data yang digunakan dalam penelitian tersebut adalah datasets *Cleveland Heart Disease* yang berjumlah 303 data yang bersumber dari UCI *machine learning*. Datasets tersebut memiliki 7 data yang rusak, oleh karena itu 7 data tersebut dihapus supaya tidak mempengaruhi hasil klasifikasi yang dilakukan oleh model nantinya. Metodologi yang dilakukan dalam penelitian tersebut adalah pengumpulan datasets, *preprocessing* data yang dilakukan dengan cara membersihkan data yang dirasa rusak atau *missing value*, dan merubah kelas data menjadi *binary class* dengan asumsi label 0 yang berarti data tersebut masuk ke dalam klasifikasi sehat, dan label 1 yang berarti data tersebut masuk ke dalam klasifikasi sakit. Kemudian, tahapan selanjutnya adalah diskritisasi data, yaitu merubah data numerik menjadi diskrit. Selanjutnya dilakukan proses seleksi fitur, dalam penelitian ini seleksi fitur yang dilakukan dengan menggunakan dua cara, yaitu teknik VPRS dan teknik MFS. Kemudian tahapan selanjutnya adalah pembuatan *rule* yang berisi aturan-aturan yang akan dijadikan untuk proses klasifikasi menggunakan data tersebut. Langkah terakhir adalah pengujian dan evaluasi model klasifikasi untuk mengetahui performa metode tersebut. Evaluasi performa yang dilakukan dalam penelitian tersebut dilakukan dengan menggunakan perhitungan *confusion matrix*. Hasil dari evaluasi program tersebut mendapatkan nilai

akurasi, presisi, dan *recall*. Metode VPRS yang digunakan untuk klasifikasi mendapatkan akurasi sebesar 84,84% metode MFS yang digunakan untuk klasifikasi menghasilkan model yang memiliki nilai akurasi sebesar 86,86%. Sedangkan kombinasi dari kedua metode VPRS dan metode MFS saat digunakan untuk proses klasifikasi menghasilkan model yang memiliki akurasi sebesar 84,84%. Dari hasil penelitian tersebut dapat disimpulkan bahwa, menggunakan metode seleksi fitur VPRS dinilai lebih baik karena menghasilkan model yang memiliki performa yang lebih baik daripada proses klasifikasi yang dilakukan tanpa menggunakan seleksi fitur. Sedangkan menggunakan metode VPRS yang dikombinasikan dengan metode MFS akan menghasilkan *rule* yang lebih sedikit, namun masih tetap mendapatkan nilai akurasi yang baik yaitu sebesar 84,84%.

Selanjutnya ada penelitian yang dilakukan oleh Vincentius Adbi Gunawan, Leonardus Sandy Ade Putra, dan Ignitia Imelda Fitriana pada tahun 2020 (Gunawan et al., 2020). Penelitian tersebut dilakukan untuk membuat sistem yang dapat melakukan diagnosis penyakit jantung dengan menggunakan citra mata, khususnya pada bagian iris. Menurut penelitian tersebut, apabila seseorang terkena penyakit jantung maka akan terjadi penyempitan pembuluh darah dan akan sangat berkaitan dengan aliran darah yang terjadi di bagian mata. Oleh karena itu penelitian ini bermaksud untuk melakukan proses pendeteksian penyakit jantung yang dilakukan dengan input berupa citra foto bola mata. Metode yang digunakan untuk mendeteksi organ tubuh menggunakan iris mata adalah Iridologi. Setiap organ yang ada di tubuh kita dapat dicerminkan melalui iris, iris mata kanan akan mencerminkan kondisi organ tubuh yang berada di kanan, begitu juga dengan iris mata kiri yang akan mencerminkan kondisi organ tubuh di bagian kiri. Data yang digunakan dalam penelitian tersebut adalah data citra foto mata yang berjumlah 70 foto, terdiri dari 35 foto dengan klasifikasi normal atau sehat, dan 35 foto lainnya yang tergolong ke dalam klasifikasi mata abnormal. Tahapan yang dilakukan di dalam penelitian tersebut adalah pengolahan citra digital dan pembagian data menjadi *data training* sebanyak 30 foto dan *data testing* sebanyak 40 foto. Kemudian dilakukan proses ekstraksi citra dengan menggunakan GLCM, dan kemudian proses klasifikasi dengan menggunakan algoritma SVM, kemudian pengujian dan pengukuran performa algoritma untuk mengetahui kinerja sistem yang dibuat. Hasil pengujian model klasifikasi yang dibuat dengan menggunakan algoritma SVM mendapatkan akurasi sebesar 87,15%. Dari penelitian yang sudah dilakukan tersebut, didapatkan kesimpulan bahwa hubungan penyakit jantung dengan iris mata adalah, saat seseorang menderita atau memiliki penyakit jantung maka akan terjadi masalah pada syaraf yang ada di matanya. Sedangkan untuk orang normal yang tidak memiliki penyakit jantung, maka tidak akan ada masalah pada syaraf yang ada di iris matanya.

Selanjutnya masih ada penelitian yang membahas tentang penyakit jantung namun menggunakan *deep learning*, yaitu penelitian yang dilakukan oleh Majzoob K. Omer pada tahun 2018 (Omer et al., 2018). Penelitian tersebut dilakukan untuk membuat sistem pendeteksian penyakit jantung, karena dirasa penyakit jantung menjadi penyakit yang perlu diwaspadai namun untuk melakukan pengecekan sebelumnya harus melakukan konsultasi ke dokter. Proses tersebut tentunya memerlukan waktu dan biaya

serta proses yang tidak efektif, ditambah tidak semua rumah sakit memiliki dokter jantung yang sangat berkompeten untuk dapat melakukan pengecekan penyakit jantung. Oleh karena itu penelitian tersebut dilakukan untuk membuat sistem yang dapat digunakan secara universal untuk mendeteksi penyakit jantung menggunakan metode *deep neural network*. Datasets yang digunakan di dalam penelitian tersebut adalah data pasien yang terkait dengan IHD yang bersumber dari sebuah rumah sakit di Jakarta. Datasets yang digunakan berjumlah 305 data yang memiliki atribut sebanyak 10 atribut. Data tersebut dibagi menjadi dua bagian, yaitu *data training* yang berjumlah sebanyak 250 data, dan *data testing* yang berjumlah 55 data. Metode yang digunakan untuk menyelesaikan permasalahan deteksi penyakit jantung dalam penelitian ini menggunakan *deep neural network* dengan konfigurasi memiliki 152 neuron masukan, 52 neuron keluaran, dan memiliki 4 *hidden layer*. Hasil dari pengujian yang dilakukan sebanyak 5 rangkaian percobaan pada 55 data yang dijadikan *data testing* adalah, dari 5 kali percobaan pengujian yang dilakukan menggunakan metode *neural network* konvensional mendapatkan rata-rata akurasi sebesar 99,055% sedangkan pengujian yang dilakukan dengan menggunakan metode *deep neural network* mendapatkan rata-rata akurasi sebesar 99,787%. Berdasarkan penelitian yang dilakukan tersebut didapatkan kesimpulan bahwa melakukan klasifikasi menggunakan teknik *deep neural network* dapat meningkatkan performa yang signifikan yaitu meningkat sebesar 0,7322% apabila dibandingkan dengan klasifikasi yang dilakukan dengan menggunakan teknik *neural network* konvensional.

Penelitian selanjutnya adalah penelitian yang dilakukan oleh Wiharto, Esti Suryani, dan Vicka Cahyawati pada tahun 2019 (Wiharto et al., 2019). Prediksi penyakit jantung yang dilakukan untuk sepuluh tahun kedepan dirasa masih bisa dilakukan, oleh karena itu data-data sumber daya harus digunakan secara maksimal khususnya adalah data pasien jantung koroner yang dapat digunakan untuk membuat sistem yang dapat memprediksi apakah seseorang tersebut menderita penyakit jantung atau tidak. Penelitian tersebut dilakukan dengan menggunakan metode jaringan syaraf tiruan *multi layer perceptron* (MLP-ANN) dan *Duo Output Ensemble Artificial Neural Network* (DOANNE). Data yang digunakan dalam penelitian tersebut adalah data yang bersumber dari RSUD Dr.Moewardi, Surakarta. Datasets tersebut memiliki atribut sebanyak 12 atribut, dengan jumlah data adalah 72 data, 36 pasien terdiagnosis penyakit jantung koroner positif, dan sisa data lainnya tergolong dalam klasifikasi negative atau sehat. Sebanyak 24 data digunakan sebagai *data testing* yang nantinya akan dilakukan pengujian. Tahapan yang ada dalam penelitian tersebut adalah pengumpulan data, kemudian melakukan *preprocessing* untuk membersihkan data yang *missing value* atau rusak, kemudian melakukan proses training menggunakan teknik jaringan syaraf tiruan. Dari algoritma *neural network* tersebut akan mengklasifikasikan data sesuai dengan klasifikasi yang diprediksi oleh model yang dibuat. Hasil dari penelitian tersebut menjelaskan bahwa melakukan klasifikasi menggunakan teknik DOANNE-LM dapat meningkatkan performa model klasifikasi dan juga dapat mencegah *overfitting* pada model. Dibuktikan dengan klasifikasi yang dilakukan dengan teknik DOANNE-LM dapat mendapat akurasi sebesar 86,875%. Model klasifikasi yang dibangun dengan

menggunakan teknik DOANNE-LM mampu menekan *overfitting* sebesar 49,09% dibandingkan dengan klasifikasi yang dilakukan dengan menggunakan JST-LM.

Selanjutnya terdapat penelitian yang dilakukan oleh Pareza Alam Jusia pada tahun 2018 (Pareza Alam Jusia, 2018). Penelitian tersebut dilakukan untuk *improve classification accuracy* atau *ensemble methods technique* dengan cara mengkombinasikan algoritma *Decision Tree* dan teknik *Particle Swarm Optimization* (PSO) yang ditambahkan dengan metode *Adaboost* untuk melakukan prediksi penyakit jantung pada seseorang menggunakan data yang sudah ada. Datasets yang digunakan di dalam penelitian tersebut adalah data sekunder yang bersumber dari UCI *machine learning*, data tersebut memiliki jumlah 270 data dengan rincian adalah 120 data masuk dalam klasifikasi negative dan sebanyak 150 data tergolong ke dalam klasifikasi dengan label positif terkena penyakit jantung. Dalam penelitian tersebut terdapat beberapa tahapan yang dilakukan supaya penelitian tersebut dapat berjalan sesuai dengan rencana dan tidak menimbulkan kesalahan pada saat penelitian dilakukan. Tahapan tersebut antara lain proses *import* datasets, proses *preprocessing* data, pembuatan model menggunakan algoritma C4.5 dan algoritma C4.5 yang sudah dikombinasikan dengan PSO dan teknik *Adaboost*. Kemudian dilakukan pengujian dan pengukuran performa algoritma yang sedang diuji. Dalam penelitian tersebut pengujian dan pengukuran performa algoritma dilakukan menggunakan *Confusion Matrix* dan AUC untuk mengukur akurasi, presisi, dan recall dari model klasifikasi yang sedang dibuat. Hasil dari penelitian tersebut adalah pembuatan model klasifikasi yang dibuat dengan menggunakan algoritma C4.5 murni mendapatkan nilai akurasi sebesar 79,26%. Model klasifikasi yang dibuat dengan menggunakan algoritma C4.5 yang dikombinasikan dengan seleksi fitur PSO mendapatkan nilai akurasi paling tinggi, yaitu sebesar 82,59%. Sedangkan model klasifikasi yang dibuat dengan menggunakan algoritma C4.5 yang dikombinasikan dengan teknik *Adaboost* mendapatkan nilai akurasi yang sama dengan algoritma C4.5 murni, yaitu sebesar 79,26%.

Selanjutnya terdapat penelitian yang dilakukan oleh Tri Retnasari dan Eva Rahmawati yang dilakukan pada tahun 2017 (Retnasari & Rahmawati, 2017). Dalam penelitian tersebut dijelaskan bahwa pentingnya menciptakan sistem yang mampu mendiagnosis seseorang apakah terkena penyakit jantung atau tidak dengan menggunakan sistem computer, supaya dapat dilakukan dengan mudah dan efisien. Beberapa penelitian sudah dilakukan dan menggunakan beberapa algoritma untuk melakukan prediksi penyakit jantung tersebut. Dalam penelitian tersebut penulis melakukan penelitian dengan menggunakan algoritma *Naïve Bayes* dan algoritma C4.5 yang kemudian akan dibandingkan untuk mendapatkan algoritma yang lebih baik dalam menangani klasifikasi penyakit jantung. Tahapan yang dilakukan dalam penelitian tersebut adalah melakukan pengumpulan data, data yang digunakan di dalam penelitian tersebut adalah datasets sekunder yang bersumber dari UCI *machine learning*. Setelah melakukan pengumpulan data, tahapan selanjutnya adalah melakukan *preprocessing* data, *preprocessing* dilakukan untuk memilih data yang baik, membersihkan data, ataupun mentransformasikan data ke dalam bentuk yang diinginkan sebelum dilakukan pemodelan. Kemudian, dilakukan proses pembuatan model klasifikasi dengan



menggunakan algoritma *Naïve Bayes* dan algoritma C4,5. Kemudian dilakukan proses pengujian dan pengukuran untuk mengevaluasi model klasifikasi yang dibuat dengan kedua algoritma tersebut, dan menentukan algoritma terbaik yang memiliki performa yang lebih baik dalam menangani klasifikasi penyakit jantung. Untuk mengukur performa algoritma yang digunakan, di dalam penelitian tersebut digunakan *confusion matrix* dan kurva AUC. Berdasarkan penelitian yang sudah dilakukan algoritma *Naïve Bayes* mendapatkan akurasi sebesar 86,67% dan AUC 0,090. Sedangkan algoritma C4,5 mendapatkan akurasi sebesar 83,70% dan AUC sebesar 0,834. Dari penelitian tersebut dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki performa yang lebih baik daripada algoritma C4,5 walaupun hanya memiliki selisih yang tidak terlalu signifikan.

Kemudian, terdapat penelitian tentang penyakit jantung yang dilakukan oleh Syafitri Hidayatul Annur Aini, Yuita Arum Sari, dan Achmad Arwan pada tahun 2018 (Aini et al., 2018). Penelitian tersebut dilakukan untuk mengurangi dimensi data dengan menggunakan *information gain*, yang nantinya hanya menyisakan atribut-atribut yang penting atau memiliki bobot paling tinggi terhadap hasil klasifikasi yang dihasilkan. Kemudian setelah melakukan *information gain* untuk mengurangi atribut, dilakukan klasifikasi menggunakan algoritma *Naïve Bayes* dan algoritma *K-Nearest Neighbor*. Datasets yang digunakan di dalam penelitian tersebut adalah data sekunder yang bersumber dari *UCI machine learning* dengan data yang berjumlah 270 data dan memiliki atribut sebanyak 13, dengan memiliki dua kelas target yaitu terkena penyakit jantung (TPJ) dan tidak terkena penyakit jantung (TTPJ). Tahapan yang dilakukan di dalam penelitian tersebut adalah, pertama data dikonversikan dahulu dari yang semula data bersifat numerik, dikonversikan menjadi data yang bersifat kategoris. Kemudian data yang sudah dikonversikan dilakukan proses pengurangan dimensi atribut dengan menggunakan *information gain*. Kemudian menggunakan data yang belum dikonversi, dilakukan proses klasifikasi menggunakan algoritma KNN pada data numerik, dan dilanjutkan dengan perhitungan data yang bersifat kategoris dengan menggunakan algoritma *Naïve Bayes*. Pengujian yang dilakukan di dalam penelitian tersebut terbagi menjadi dua macam, yaitu pengujian yang dilakukan dengan data latih dengan kelas seimbang, dan pengujian yang dilakukan dengan menggunakan data latih dengan kelas tidak seimbang. Untuk jumlah atribut yang digunakan pada masing-masing pengujian adalah menggunakan 6 atribut dan 4 atribut, sedangkan nilai K yang digunakan berkelipatan 10 mulai dari K=5 sampai dengan K=95. Berdasarkan dari hasil penelitian yang didapatkan dapat disimpulkan bahwa saat pengujian dilakukan dengan menggunakan data latih kelas seimbang, mendapatkan akurasi tertinggi yaitu 92,31% dengan atribut sejumlah enam dan nilai K=25. Sedangkan pengujian yang dilakukan pada data latih tidak seimbang, nilai akurasi tertinggi yaitu 92,31% dengan menggunakan empat atribut dan nilai K=35.

Selanjutnya terdapat penelitian yang dilakukan dengan menggunakan metode XgBoost. Penelitian tersebut dilakukan oleh Ichwanul Muslim Karo pada tahun 2020 (Karo, 2020). Penelitian tersebut bertujuan untuk melakukan klasifikasi titik api penyebab kebakaran hutan pada data yang bersumber dari *Global Forest Watch* (GFW) dengan menggunakan algoritma XgBoost. Data yang digunakan di dalam penelitian

tersebut berjumlah 300 dengan jumlah atribut adalah 12 atribut, dengan label kelas yang menjadi target klasifikasi adalah empat kelas label. Tahapan yang dilakukan di dalam penelitian tersebut antara lain adalah tahap *preprocessing* yang dilakukan untuk memilih data yang baik, melakukan normalisasi data, dan melakukan *feature important* untuk mengurangi atribut dan hanya menyisakan atribut yang penting. Kemudian pembuatan model klasifikasi dilakukan dengan menggunakan algoritma XgBoost dengan beberapa parameter seperti *max depth*=5, *seed*=7, *test size*=0,35, *feature*=1-9, dan *learning rate*=0,05. Setelah pembuatan model dilakukan dengan menggunakan algoritma XgBoost kemudian dilakukan validasi model untuk mengukur performa model klasifikasi, pengukuran model klasifikasi dilakukan dengan menggunakan *confusion matrix* untuk mengetahui nilai akurasi, presisi, dan *recall*. Berdasarkan penelitian yang dilakukan, hasil yang didapatkan adalah dengan menggunakan *feature important*, dapat mengurangi jumlah atribut yang tadinya 12 atribut menjadi 9 atribut yang dirasa penting. Namun saat dilakukan validasi menggunakan algoritma XgBoost, ternyata akan lebih optimal dan memiliki performa yang baik apabila dibuat klasifikasi dilakukan dengan menggunakan 6 atau 7 atribut yang paling berpengaruh terhadap data. Hasil dari pengukuran performa, model algoritma XgBoost yang dilakukan untuk klasifikasi pada penelitian tersebut mendapatkan nilai SE sebesar 91,32% nilai SP sebesar 93,16%, dan nilai MCC sebesar 92,75%.

Selanjutnya terdapat penelitian yang dilakukan oleh Sherla Yualinda, Dr. Dedy Rahman Wijawa, S.T., M.T., dan Elis Hernawati, S.T., M.Kom. yang dilakukan pada tahun 2020 (Yualinda et al., 2020). Penelitian tersebut dilakukan untuk membuat sistem prediksi kemiskinan menggunakan *machine learning* dengan metode algoritma *Naïve Bayes* dan algoritma XgBoost. Tahapan yang dilakukan untuk membuat model prediksi adalah melakukan import data, kemudian melakukan *preprocessing* data seperti normalisasi data, kemudian dilakukan proses dengan menggunakan teknik *Similarity Based Feature Selection* untuk mengurangi dimensi atribut dan menyisakan atribut yang penting saja. Kemudian dilakukan proses pemodelan klasifikasi dengan menggunakan algoritma *Naïve Bayes* dan algoritma XgBoost. Setelah itu, dilakukan proses evaluasi dengan menghitung nilai RMSE dan R. Hasil dari penelitian tersebut dapat diciptakan sistem prediksi kemiskinan yang ada di Indonesia dengan hasil tampilan akan menghasilkan grafik prediksi kemiskinan yang ada di Indonesia.

Selanjutnya terdapat penelitian yang dilakukan pada tahun 2020 oleh Muhamad Syukron, Rukun Santoso, dan Tatik Widiyari (Syukron et al., 2020). Penelitian tersebut dilakukan untuk membuat model klasifikasi yang digunakan untuk mengklasifikasikan tingkat penyakit hepatitis C. Data yang digunakan di dalam penelitian tersebut memiliki jenis *Imbalance Data*, yang berarti data yang ada memiliki kelas target klasifikasi yang tidak seimbang jumlahnya. Apabila dipaksakan membuat model klasifikasi dengan menggunakan data yang memiliki kelas tidak seimbang tersebut, model akan cenderung memprediksi dengan hasil kelas klasifikasi yang lebih banyak jumlahnya. Oleh karena itu, peneliti menggunakan teknik SMOTE untuk menangani ketidak seimbangan kelas data tersebut. Dalam penelitian ini peneliti juga membandingkan performa model yang dihasilkan dengan menggunakan algoritma *Random Forest*, dan algoritma XgBoost

untuk mendapatkan kesimpulan algoritma yang memiliki performa terbaik. Tahapan yang digunakan di dalam penelitian tersebut antara lain ada *preprocessing* data, data yang akan digunakan tentunya belum baik atau memiliki beberapa data yang harus dibuang. Dalam kasus ini, data yang digunakan ternyata memiliki *outlier*, oleh karena itu perlu dihapus supaya model yang dibuat nantinya dapat lebih optimal saat melakukan proses *training* data. Kemudian dibuat model klasifikasi dengan menggunakan algoritma *Random Forest* dan *XgBoost* dan menggunakan *Random Search* untuk menentukan parameter terbaik yang akan digunakan model. Setelah itu dilakukan pengujian sekaligus pengukuran algoritma menggunakan *confusion matrix* untuk mengetahui nilai akurasi, presisi, dan *recall* dari algoritma tersebut. Hasil dari penelitian tersebut dapat disimpulkan bahwa algoritma *Random Forest* yang dikombinasikan dengan teknik SMOTE ternyata memiliki akurasi paling tinggi yaitu 80,97% dibandingkan algoritma *XgBoost* dan SMOTE yang mendapatkan akurasi sebesar 78,63%.

Kemudian, terdapat penelitian yang dilakukan oleh Ngakan Nyoman Pandika Pinata, I Made Sukarsa, Ni Kadek Dwi Rusjyanthi pada tahun 2020 (Pinata et al., 2020). Penelitian tersebut membahas tentang prediksi kecelakaan lalu lintas yang ada di Bali. Sistem klasifikasi tersebut dibuat dengan menggunakan algoritma *XgBoost* dan dengan menggunakan bahasa pemrograman python. Data yang digunakan di dalam penelitian tersebut adalah data jumlah kecelakaan lalu lintas yang ada di Bali dari tahun 1996 sampai dengan tahun 2019 dengan jumlah data adalah 24 data sesuai dengan tahun yang ada pada data tersebut. Data yang digunakan dalam penelitian tersebut dibagi menjadi dua bagian yaitu *data training* yang digunakan untuk proses pelatihan pembuatan model klasifikasi berjumlah 20 data, dan *data testing* yang digunakan untuk pengujian model klasifikasi yang dibuat berjumlah 4 data. Pengukuran yang digunakan di dalam penelitian tersebut adalah menggunakan RMSE untuk mengukur kesalahan dari model klasifikasi yang dibuat. Berdasarkan penelitian yang sudah dilakukan, model klasifikasi yang dibuat dengan menggunakan algoritma *XgBoost* dapat menghasilkan performa yang baik dibuktikan dengan nilai RMSE yang cukup rendah. Nilai *error* pada kategori jumlah kejadian adalah 21,69. Nilai RMSE pada kategori jumlah orang meninggal dunia adalah 4,92. Nilai RMSE pada kategori luka berat adalah 4,11. Dan pada kategori luka ringan mendapatkan nilai RMSE sebesar 77,24.

Selanjutnya terdapat penelitian yang dilakukan oleh Ahmedbahaldin Ibrahim Ahmed Osman, Ali Najah Ahmed pada tahun 2020 di Malaysia (Ibrahim Ahmed Osman et al., 2020). Penelitian tersebut dilakukan untuk membuat sistem yang dapat memprediksi ketinggian air di daerah Selangor negara Malaysia. Data yang digunakan di dalam penelitian tersebut adalah data dari tanggal 20 Oktober 2017 hingga 24 Juli 2018, data tersebut memiliki atribut antara lain adalah curah hujan, suhu, evaporasi, tinggi air. Data yang digunakan tersebut terlebih dahulu dibagi menjadi dua bagian, yaitu *data training* dan *data testing* dengan perbandingan 70% untuk *data training* dan 30% untuk *data testing*. Metode yang digunakan dalam penelitian ini menggunakan beberapa algoritma seperti *XgBoost*, JST, dan SVR dimana ketiga algoritma tersebut akan dibandingkan hasil performanya untuk mengetahui algoritma mana yang memiliki performa terbaik dalam menangani kasus prediksi pada penelitian tersebut. Proses

pengukuran performa algoritma pada penelitian tersebut menggunakan *R Square*. Hasil dari penelitian tersebut membuktikan bahwa algoritma XgBoost dinilai memiliki performa yang lebih baik daripada kedua algoritma JST dan SVR. Hal tersebut dibuktikan dalam pengukuran kinerja, algoritma XgBoost mendapatkan nilai MAE sebesar 0,086 algoritma JST yang mendapatkan nilai MAE sebesar 0,254 dan algoritma SVR dengan nilai sebesar 0,111.

No.	Penulis	Judul	Tahun	Metode	Hasil
1.	Erwin Prasetyo & Budi Prasetyo	Peningkatan Akurasi Klasifikasi Algoritma C4,5 Menggunakan Teknik Bagging Pada Diagnosis Penyakit Jantung	2020	Algoritma C4,5 dan <i>Bagging</i>	
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					
10.					
11.					
12.					
13.					
14.					
15.					
16.					
17.					
18.					
19.					
20.					

## 2.2.Tinjauan Pustaka

## 2.3.Kerangka Pemikiran

## DAFTAR PUSTAKA

- Aini, S. H. A., Sari, Y. A., & Arwan, A. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2546–2554.
- 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>
- Gunawan, V. A., Fitriani, I. I., & Putra, L. S. A. (2020). Sistem Diagnosis Otomatis Identifikasi Penyakit Jantung Coroner Menggunakan Ekstraksi Ciri GLCM dan Klasifikasi SVM. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 15(1), 13. <https://doi.org/10.30872/jim.v15i1.2495>
- Ibrahim Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2020). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, xxxx. <https://doi.org/10.1016/j.asej.2020.11.011>
- 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.
- Jothikumar, R., & Siva Balan, R. (2016). C4.5 classification algorithm with back-track pruning for accurate prediction of heart disease. *ISSN: 0970-938X (Print)*. <https://www.biomedres.info/biomedical-research/c45-classification-algorithm-with-backtrack-pruning-for-accurate-prediction-of-heart-disease.html>
- Karo, I. M. K. (2020). Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan. 1(1), 10–16.
- Lestari, M. (2014). Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung. *Faktor Exacta*, 7(September 2010), 366–371.
- 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>
- Normawati, D., & Winarti, S. (2017). Seleksi Fitur Menggunakan Penambangan Data Berbasis Variable Precision Rough Set (VPRS) untuk Diagnosis Penyakit Jantung Koroner. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 3(2), 100. <https://doi.org/10.26555/jiteki.v3i2.8072>
- 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>
- Omer, M. K., Sheta, O. E., Adrees, M. S., Stiawan, D., Riyadi, M. A., & Budiarto, R. (2018). Deep neural network for heart disease medical prescription expert system. *Indonesian Journal of Electrical Engineering and Informatics*, 6(2), 217–224.  
<https://doi.org/10.11591/ijeei.v6i2.456>
- Pareza Alam Jusia. (2018). Analisis komparasi pemodelan algoritma decision tree menggunakan metode particle swarm optimization dan metode adaboost untuk prediksi awal penyakit jantung. *Seminar Nasional Sistem Informasi 2018*, 1048–1056.
- Pinata, N. N. P., Sukarsa, I. M., & Rusjyanthi, N. K. D. (2020). *Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python*. 8(3), 188–196.
- 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>
- Putra, P. D., & Rini, D. P. (2019). Prediksi Penyakit Jantung dengan Algoritma Klasifikasi. *Prosiding Annual Research Seminar 2019*, 5(1), 978–979.
- 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.
- Rohman, A., Suhartono, V., & Supriyanto, C. (2017). Penerapan Agoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung. *Jurnal Teknologi Informasi*, 13, 13–19.
- Syukron, M., Santoso, R., & Widiharhi, T. (2020). *Perbandingan Metode Smote Random Forest dan Smote XgBoost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data*. 9, 227–236.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- 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>
- Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Optimization Untuk Deteksi Penyakit Jantung. *Pseudocode*, 1, 11–14.
- Wiharto, W., Suryani, E., & Cahyawati, V. (2019). The methods of duo output neural network ensemble for prediction of coronary heart disease. *Indonesian Journal of Electrical Engineering and Informatics*, 7(1), 50–57.  
<https://doi.org/10.11591/ijeei.v7i1.458>
- Yualinda, S., Wijaya, D. R., & Hernawati, E. (2020). Aplikasi Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naive Bayes, XgBoost dan Similarity Based Feature Selection. *Jurnal Borneo Cendekia*, 3(2), 40–46.
- 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 dn Rasio Kolesterol (HDL) pada Penderita Penyakit Jantung Koroner di Poliklinik Jantung RSUD Dr Moewardi Surakarta. *Jurnal Kesehatan*, 2(1), 11–18.