

# TUGAS AKHIR

## APLIKASI UNTUK PREDIKSI PENYAKIT JANTUNG



Febrianda Putra

NPM: 6181901007

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2024

**FINAL PROJECT**

**HEART DISEASE PREDICTION APPLICATION**



**Febrianda Putra**

**NPM: 6181901007**

**DEPARTMENT OF INFORMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2024**

# LEMBAR PENGESAHAN

## APLIKASI UNTUK PREDIKSI PENYAKIT JANTUNG

Febrianda Putra

NPM: 6181901007

Bandung, 16 Januari 2024

Menyetujui,

Pembimbing

Digitally signed  
by Vania Natali

Vania Natali, M.T.

Ketua Tim Penguji

Digitally signed  
by Luciana  
Abednego

Luciana Abednego, M.T.

Anggota Tim Penguji

Digitally signed  
by Natalia

Natalia, M.Si.

Mengetahui,

Ketua Program Studi

Digitally signed  
by Lionov

Lionov, Ph.D.

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa tugas akhir dengan judul:

### **APLIKASI UNTUK PREDIKSI PENYAKIT JANTUNG**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
Tanggal 16 Januari 2024



Febrianda Putra  
NPM: 6181901007

## ABSTRAK

Kesehatan merupakan aspek vital yang perlu diperhatikan oleh setiap individu. Berbagai faktor, seperti pola makan buruk, gaya hidup tidak sehat, dan kondisi mental yang tidak stabil dapat berkontribusi terhadap memburuknya kesehatan manusia dan berujung pada penyakit jantung. Penyakit jantung dapat dicegah berdasarkan gejala yang timbul dari pasien dan didiagnosis oleh dokter, tetapi diagnosis dokter membutuhkan pasiennya untuk mengeluarkan biaya, dan biaya tersebut terkadang tidaklah sedikit. Dalam upaya mengurangi biaya yang perlu dikeluarkan pada saat konsultasi dokter, analisis dapat dilakukan menggunakan perangkat lunak berbasis web yang memanfaatkan data medis untuk memprediksi kemungkinan seseorang terkena penyakit jantung. Metode *data mining* dapat digunakan untuk membantu mengidentifikasi pola-pola kompleks dalam data medis, prediksi penyakit jantung dan memberikan informasi yang berguna untuk ditampilkan pada halaman aplikasi prediksi penyakit jantung.

Dilakukan eksplorasi dari data medis tersebut untuk dilakukan pemilihan fitur yang berguna dalam menentukan apakah seseorang memiliki penyakit jantung atau tidak. Pemilihan fitur tersebut dilakukan dengan berbagai teknik seperti statistika, *domain knowledge* dan *expertise*. Setelah dilakukan pemilihan fitur, dilakukan pembersihan data seperti menangani *missing value* dan *noisy data*. Transformasi data seperti *one-hot encoding* juga dilakukan agar algoritma tidak salah dalam menginterpretasi data yang dimiliki. Dalam proses membuat model klasifikasi, data yang dimiliki harus sesuai dengan *requirement input* algoritma yang dipakai. Algoritma klasifikasi yang dipakai meliputi *Naive Bayes*, *Support Vector Machine*, *Decision Tree* dan *Random Forest*.

Nilai dari *recall* merepresentasikan banyaknya *false negative* menjadi kunci karena nilai tersebut merupakan jumlah orang yang sebenarnya memiliki penyakit jantung tetapi diprediksi sebagai tidak. Hal tersebut fatal jika terjadi karena dapat menyebabkan kematian. Dilakukan dua tahap penelitian, dimana satu tahapannya berisi dua bagian. Bagian pertama merupakan percobaan keempat buah algoritma dengan *hyperparameter default* untuk melihat cara kerja algoritma terbaik. Bagian kedua dilakukan percobaan *hyperparameter tuning* dan *cross-validation* berdasarkan model terbaik. Tahapan selanjutnya, dilakukan bagian yang sama seperti tahap sebelumnya, yang berbeda hanyalah fitur yang dipakai. Berdasarkan hasil analisis, dapat disimpulkan bahwa faktor umur, jenis kelamin, adanya angina hingga hasil gradien elektrokardiogram pada titik tertinggi, menunjukkan adanya hubungan dengan ketepatan prediksi penyakit jantung. Eksplorasi data dan analisis model memberikan gambaran bahwa setiap faktor tersebut memiliki peran dalam menentukan kecenderungan seseorang untuk memiliki penyakit jantung.

Klasifikasi ini hanya akan melihat ya atau tidak penyakit jantung, dan berdasarkan penelitian yang telah dilakukan, aplikasi prediksi penyakit jantung berhasil menentukan seseorang memiliki penyakit jantung dengan tingkat akurasi senilai 76% dan *recall* senilai 93%. Proses ini dibantu dengan metode *data mining* yang dimulai dari pengambilan data hingga klasifikasi. Didapatkanlah bahwa algoritma klasifikasi *Random Forest* dengan *hyperparameter* yang telah dilakukan *tuning* merupakan algoritma terbaik dalam melakukan prediksi penyakit jantung. Dalam menampilkan hasil prediksinya, berhasil dibangun perangkat lunak berbasis web yang dapat menerima *input* kondisi kesehatan, lalu perangkat lunak akan menampilkan hasilnya.

**Kata-kata kunci:** Kesehatan, Penyakit Jantung, Prediksi, Klasifikasi, *Naive Bayes* (NBC), *Support Vector Machine* (SVM), *Decision Tree* (DTC), *Random Forest* (RF), *Recall*, *False Negative*

## ABSTRACT

Health is a vital aspect that needs attention from every individual. Various factors, such as poor eating habits, unhealthy lifestyles, and unstable mental conditions, can contribute to deteriorating human health and lead to heart disease. While heart disease can be prevented based on symptoms observed by patients and diagnosed by doctors, the cost associated with doctor consultations can be substantial. In an effort to reduce the expenses incurred during medical consultations, analysis can be performed using web-based software that leverages medical data to predict the likelihood of someone having heart disease. Data mining methods can be employed to help identify complex patterns in medical data, predicting heart disease and provide useful information to be displayed on the heart disease prediction application.

The exploration of medical data is conducted to select useful features in determining whether someone has heart disease or not. Feature selection is carried out using various techniques such as statistics, domain knowledge, and expertise. After feature selection, data cleaning is performed, addressing missing values and noisy data. Data transformation, such as one-hot encoding, is also carried out to ensure that algorithms interpret the data correctly. In the process of creating a classification model, the data must comply with the input requirements of the algorithm used. Classification algorithms employed include Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest.

The value of recall, representing the number of false negatives generated, becomes crucial as it indicates the number of people who actually have heart disease but are predicted not to have it. This situation is critical as it can lead to fatalities. The research is conducted in two stages, with each stage consisting of two parts. The first part involves experimenting with four algorithms using default hyperparameters to observe the best-performing algorithm. The second part involves hyperparameter tuning and cross-validation based on the best model. In the subsequent stage, the same parts are conducted, with the only difference being the features used. Based on the analysis results, it can be concluded that factors such as age, gender, the presence of angina, and the result of the highest point electrocardiogram gradient show a correlation with the accuracy of heart disease prediction. Data exploration and model analysis provide insight that each of these factors plays a crucial role in determining an individual's risk or tendency to experience heart disease.

This classification will only determine whether a person has heart disease or not, and based on the conducted research, this heart disease prediction application has successfully determined whether someone is likely to have heart disease with an accuracy rate of 76% and a recall rate of 93%. This process is assisted by the data mining method, starting from data collection to classification. It was found that the Random Forest classification algorithm with tuned hyperparameters is the best algorithm for predicting heart disease. To display the prediction results, a web-based software has been successfully developed that can receive health condition inputs, and the software will then present the prediction results.

**Keywords:** Health Care, Heart Disease, Prediction, Classification, *Naive Bayes* (NBC), *Support Vector Machine* (SVM), *Decision Tree* (DTC), *Random Forest* (RF), Recall, False Negative

*Skripsi ini saya persembahkan untuk papah (Stephanus), ibu (Eri),  
kaka (DJ), keluarga dan pacar (Sonia) yang tanpa lelah sudah  
mendukung semua keputusan dan pilihan dalam hidup saya serta  
tidak pernah berhenti mendoakan saya.*

## KATA PENGANTAR

Puji dan syukur peneliti panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat yang dilimpahkan oleh-Nya, sehingga penyusunan tugas akhir dapat diselesaikan. Tugas akhir ini merupakan sebuah penelitian dengan judul Aplikasi Prediksi Penyakit Jantung. Tugas akhir ini diajukan sebagai syarat untuk menempuh pendidikan di Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Dalam penyusunan dokumen tugas akhir ini tak lupa peneliti memberikan ucapan terima kasih kepada pihak-pihak yang terlibat baik dalam memberi bimbingan, bantuan dan dukungan secara langsung atau tidak ke dalam proses penyusunan penelitian, terutama untuk:

1. Kepada orang tua dan saudara Penulis yang selalu memberi doa, dukungan dan semangat;
2. Kepada Ci Vania Natali, M.T. selaku pembimbing. Terima kasih atas ilmu, pengalaman, semangat, dan bimbingan yang telah diberikan;
3. Kepada Ci Luciana Abednego, M.T.. Terima kasih atas masukan dan saran yang diberikan, sehingga skripsi ini dapat dikembangkan menjadi lebih baik;
4. Kepada seluruh dosen yang mengajar Penulis sejak pertama masuk kuliah. Terima kasih atas bekal ilmu dan pengalaman yang berharga;
5. Kepada teman-teman YEUUUU yang sudah berjuang bersama;
6. Kepada Sonia Putri Piwanda, terima kasih atas dukungan yang telah diberikan <3;
7. Kepada teman-teman mahasiswa angkatan 2017 sampai 2024. Terima kasih atas momen kebersamaannya. Semangat menjalani kuliahnya, bagi teman-teman adik tingkat;
8. Kepada seluruh pihak yang tidak Penulis cantumkan pada bagian ini. Terima kasih atas dukungannya.

Akhir kata peneliti berharap agar tugas akhir ini dapat memberikan sumbangan nyata ataupun pembelajaran bagi kemajuan Informatika dan bagi pihak yang memerlukannya.

Bandung, Januari 2024

Penulis

# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xix</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Tujuan . . . . .	3
1.4 Batasan Masalah . . . . .	3
1.5 Metodologi . . . . .	3
1.6 Sistematika Pembahasan . . . . .	4
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Penyakit Jantung . . . . .	5
2.1.1 Jenis Penyakit Jantung . . . . .	5
2.1.2 Penjelasan Istilah Lainnya . . . . .	6
2.2 Data . . . . .	7
2.3 Statistika . . . . .	9
2.3.1 Statistika Deskriptif . . . . .	9
2.3.2 Statistika Inferensi . . . . .	12
2.3.3 Korelasi . . . . .	16
2.4 <i>Data Mining</i> . . . . .	18
2.4.1 <i>Data Preprocessing</i> . . . . .	19
2.4.2 <i>Data Mining</i> . . . . .	22
2.4.3 <i>Pattern Evaluation</i> . . . . .	22
2.4.4 <i>Knowledge Presentation</i> . . . . .	23
2.5 Visualisasi Data . . . . .	23
2.6 Algoritma Klasifikasi Untuk Prediksi . . . . .	24
2.6.1 <i>Decision Tree Classifier</i> . . . . .	24
2.6.2 <i>Random Forest Classifier</i> . . . . .	26
2.6.3 <i>Naive Bayes Classifier</i> . . . . .	27
2.6.4 <i>Support Vector Machine (SVM)</i> . . . . .	28
2.6.5 Metrik Evaluasi . . . . .	30
<b>3 EKSPLORASI DAN PENYIAPAN DATA</b>	<b>33</b>
3.1 Ekplorasi Data . . . . .	33
3.1.1 <i>Heart Disease Data Set</i> . . . . .	34
3.1.2 Deskripsi Atribut <i>Dataset</i> Utama . . . . .	35
3.2 Penyiapan Data . . . . .	54
3.2.1 Analisis Statistika Inferensi . . . . .	54
3.2.2 <i>Data Selection</i> . . . . .	57

3.2.3	Transformasi Data . . . . .	61
3.2.4	<i>One-Hot Encoding</i> dan <i>Train-Test Split</i> . . . . .	65
<b>4</b>	<b>EVALUASI DAN ANALISIS MODEL</b>	<b>67</b>
4.1	Evaluasi Model Tahap Awal . . . . .	67
4.1.1	Evaluasi Model Dengan <i>Hyperparameter Default</i> . . . . .	67
4.1.2	Eksperimen <i>Hyperparameter</i> . . . . .	75
4.1.3	Cross Validation . . . . .	78
4.1.4	<i>Hyperparameter</i> yang Dipakai . . . . .	80
4.1.5	Evaluasi Model Dengan <i>Hyperparameter</i> yang Telah Ditentukan . . . . .	81
4.1.6	Evaluasi Model Khusus Fitur “ <i>merokok</i> ” . . . . .	82
4.2	Evaluasi Model Tahap Akhir . . . . .	87
4.2.1	Penyiapan Data . . . . .	88
4.2.2	Evaluasi Model Tahap Akhir Dengan <i>Hyperparameter Default</i> . . . . .	97
4.2.3	Eksperimen <i>Hyperparameter</i> . . . . .	104
4.2.4	Cross Validation . . . . .	106
4.2.5	<i>Hyperparameter</i> yang Dipakai . . . . .	107
4.2.6	Evaluasi Model Dengan <i>Hyperparameter</i> yang Telah Ditentukan . . . . .	108
4.2.7	Evaluasi Model Khusus Fitur “ <i>merokok</i> ” . . . . .	109
4.3	Model yang Dipakai Pada Aplikasi Prediksi Penyakit Jantung . . . . .	114
<b>5</b>	<b>PERANCANGAN <i>Website</i> DAN IMPLEMENTASINYA</b>	<b>119</b>
5.1	Fitur <i>Website</i> . . . . .	119
5.2	Diagram <i>Use Case</i> . . . . .	120
5.3	Implementasi <i>Website</i> . . . . .	121
<b>6</b>	<b>KESIMPULAN DAN SARAN</b>	<b>125</b>
6.1	Kesimpulan . . . . .	125
6.2	Saran . . . . .	126
	<b>DAFTAR REFERENSI</b>	<b>127</b>
	<b>A KODE PROGRAM</b>	<b>129</b>
	<b>B <i>DATASET</i></b>	<b>201</b>

## DAFTAR GAMBAR

2.1	Macam-macam data <sup>1</sup> . . . . .	8
2.2	Ilustrasi dari <i>bar chart</i> . . . . .	23
2.3	Ilustrasi dari <i>box plot</i> . . . . .	24
2.4	Ilustrasi <i>Decision Tree</i> <sup>2</sup> . . . . .	25
2.5	Ilustrasi Dari Algoritma <i>Random Forest</i> <sup>3</sup> . . . . .	26
2.6	Ilustrasi <i>Support Vector Machine</i> (SVM) <sup>4</sup> . . . . .	29
3.1	Contoh data ( <i>vector</i> ) pada “ <i>file</i> ” dengan jenis “. <i>txt</i> ”. . . . .	34
3.2	<i>Bar chart</i> : <i>P-value</i> menggunakan <i>chi-square</i> terhadap fitur kategorik sesudah transformasi data. . . . .	54
3.3	<i>Bar chart</i> : <i>P-value</i> menggunakan ANOVA terhadap fitur numerik sesudah transformasi data. . . . .	55
3.4	<i>Bar chart</i> : Korelasi Pearson terhadap fitur “ <i>num</i> ”. . . . .	56
3.5	<i>Bar chart</i> : Korelasi Kendall’s Tau terhadap fitur “ <i>num</i> ”. . . . .	56
3.6	<i>Bar chart</i> : Korelasi Spearman terhadap fitur “ <i>num</i> ” . . . . .	57
3.7	Boxplot: distribusi <i>outlier</i> pada atribut “ <i>chol</i> ” (sebelum dilakukan pembersihan data). . . . .	62
3.8	Boxplot: distribusi <i>outlier</i> pada atribut “ <i>chol</i> ” (setelah dilakukan pembersihan data). . . . .	63
3.9	Barchart: distribusi data pada atribut “ <i>chol</i> ” (sebelum dilakukan pembersihan data). . . . .	63
3.10	Barchart: distribusi data pada atribut “ <i>chol</i> ”(setelah dilakukan pembersihan data). . . . .	64
4.1	<i>Bar chart</i> : <i>P-value</i> menggunakan <i>chi-square</i> terhadap fitur kategorik sesudah transformasi data. . . . .	90
4.2	<i>Bar chart</i> : <i>P-value</i> menggunakan ANOVA terhadap fitur numerik sesudah transformasi data. . . . .	91
4.3	<i>Bar chart</i> : Korelasi Pearson terhadap fitur “ <i>num</i> ”. . . . .	91
4.4	<i>Bar chart</i> : Korelasi Kendall’s Tau terhadap fitur “ <i>num</i> ”. . . . .	92
4.5	<i>Bar chart</i> : Korelasi Spearman terhadap fitur “ <i>num</i> ” . . . . .	92
4.6	Boxplot: distribusi <i>outlier</i> pada atribut “ <i>chol</i> ” (sebelum dilakukan pembersihan data). . . . .	94
4.7	Boxplot: distribusi <i>outlier</i> pada atribut “ <i>chol</i> ” (setelah dilakukan pembersihan data). . . . .	94
4.8	Barchart: distribusi data pada atribut “ <i>chol</i> ” (sebelum dilakukan pembersihan data). . . . .	95
4.9	Barchart: distribusi data pada atribut “ <i>chol</i> ”(setelah dilakukan pembersihan data). . . . .	95
5.1	Diagram <i>use case</i> aplikasi untuk prediksi penyakit jantung. . . . .	120
5.2	Tampilan awal antarmuka <i>website</i> . . . . .	121
5.3	Tampilan awal antarmuka <i>website</i> setelah diisi pengguna. . . . .	122
5.4	Tampilan antarmuka <i>website</i> setelah sistem mengklasifikasikan sebagai memiliki kemungkinan terkena penyakit jantung. . . . .	123
5.5	Tampilan antarmuka <i>website</i> setelah sistem mengklasifikasikan sebagai tidak memiliki kemungkinan terkena penyakit jantung. . . . .	123

# BAB 1

## PENDAHULUAN

Pada bab ini akan dibahas tentang latar belakang penelitian yang mendeskripsikan gambaran besar permasalahan yang ada. Dilanjutkan dengan pembahasan rumusan masalah, tujuan penelitian, dan batasan penelitian. Pada tiga bagian tersebut, dijelaskan inti permasalahan, tujuan yang ingin dicapai sebagai solusi masalah, serta asumsi yang digunakan untuk membatasi ruang lingkup penelitian ini. Pembahasan tentang metodologi dan sistematika pembahasan menjadi penutup dari bab ini. Dua bagian tersebut menjelaskan tentang tahapan eksperimen dan penjabaran konten setiap bab pada buku ini.

### 1.1 Latar Belakang

Kesehatan merupakan hal esensial yang harus dijaga oleh setiap manusia. Terdapat beberapa faktor yang dapat memperburuk kesehatan manusia, seperti pola makan yang buruk, pola hidup yang tidak sehat dan faktor eksternal seperti depresi[1]. Pola makan yang buruk dapat disebabkan karena konsumsi garam, gula dan produk olahan daging yang berlebih<sup>1</sup>. Konsumsi garam berlebih tidak baik untuk kesehatan karena dapat menyebabkan hipertensi, sehingga kerja jantung akan lebih berat dibandingkan dengan kondisi tekanan darah normal<sup>2</sup>. Konsumsi gula berlebih juga akan meningkatkan risiko terkena penyakit diabetes. Kadar gula darah yang berlebih dapat merusak pembuluh darah dan saraf jantung, sehingga dapat meningkatkan peluang seseorang terkena penyakit jantung<sup>3</sup>. Konsumsi produk olahan daging berlebih dapat meningkatkan kadar kolesterol, sehingga dapat menyebabkan penyempitan pembuluh darah[2].

Selain pola makan yang buruk, pola hidup yang tidak sehat juga menjadi pendorong seseorang rentan terkena penyakit. Pola hidup yang tidak sehat dapat disebabkan dari jarang berolahraga atau merokok secara terus-menerus. Jarang berolahraga dapat membuat tubuh tidak bisa menangkal hipertensi dan kolesterol yang tinggi. Merokok dapat menyebabkan pembuluh darah menyempit dan kaku, sehingga berpengaruh pada sistem kardiovaskular[3][4].

Faktor eksternal seperti kesehatan mental dapat menjadi pendorong kesehatan seseorang terganggu. Depresi menjadi penyebab utama yang berkontribusi dalam penyakit serius seperti penyempitan pembuluh darah[5]. Seseorang yang mengalami depresi memiliki trombosit yang lengket/kental

---

<sup>1</sup>“*Bad Diets Are Responsible For More Deaths Than Smoking, Global Study Finds*”, Allison Aubrey, 2019, <https://www.npr.org/sections/thesalt/2019/04/03/709507504/bad-diets-are-responsible-for-more-deaths-than-smoking-global-study-finds>, diakses pada 27 Januari 2023.

<sup>2</sup>“*Salt and your blood pressure*”, Blood Pressure UK, <https://www.bloodpressureuk.org/your-blood-pressure/how-to-lower-your-blood-pressure/healthy-eating/salt-and-your-blood-pressure/>, diakses pada 27 Januari 2023.

<sup>3</sup>“*Diabetes and Your Heart*”, Centers for Disease Control and Prevention, 2022, <https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html#:~:text=If%20you%20have%20diabetes%2C%20you,are%20to%20have%20heart%20disease.>, diakses pada 27 Januari 2023.

sehingga dapat menyebabkan darah menggumpal dan menyumbat pembuluh darah<sup>4</sup>.

Akibat dari faktor-faktor tersebut, seseorang dapat terkena gangguan kesehatan seperti penyakit jantung (*heart disease*). *World Health Organization* (WHO) menyatakan bahwa penyakit jantung berada di deretan pertama dari 10 penyebab kematian tertinggi[6]. Menurut *Centers for Disease Control and Prevention* (CDC) pada tahun 2020, di Amerika Serikat, terdapat seseorang meninggal dunia karena penyakit jantung setiap 34 detik atau sekitar 697.000 orang per tahunnya. Penyakit jantung memakan biaya sekitar \$229 miliar per tahun (2017-2018), hal ini termasuk layanan kesehatan, obat-obatan dan kehilangan produktivitas<sup>5</sup>.

Penjelasan di atas mengenai penyakit jantung dapat dicegah berdasarkan gejala yang timbul dari pasien yang diduga mengalami penyakit jantung. Gejala tersebut dapat didiagnosis oleh dokter, tetapi diagnosis dokter membutuhkan pasiennya untuk mengeluarkan biaya, dan biaya tersebut terkadang tidaklah sedikit. Rincian biayanya biasanya meliputi tes laboratorium, konsultasi (analisis) dokter, rumah sakit/klinik, dan obat. Selain itu, terdapat potensi kesalahan manusia (*human error*) dan efek yang tidak terduga (*unpredictable effects*) dalam diagnosis dokter karena keterbatasan belajar dan pengalaman yang dimiliki[7].

Dalam upaya untuk mengatasi tantangan tersebut, aplikasi ini ditujukan untuk digunakan oleh tenaga kesehatan non-dokter dan siswa kedokteran. Aplikasi ini memanfaatkan analisis data pasien terdahulu, termasuk umur, jenis kelamin, diagnosis penyakit, tekanan darah, dan riwayat penyakit untuk memberikan prediksi awal terkait kemungkinan penyakit jantung. Penggunaan teknologi dan analisis data diharapkan dapat mengurangi biaya dan meningkatkan efisiensi dalam proses diagnosa awal. Dengan demikian, aplikasi ini diarahkan untuk memberikan solusi yang lebih terjangkau dan efektif bagi tenaga kesehatan dan siswa kedokteran dalam mendukung proses diagnosa penyakit jantung. Analisis dapat dilakukan menggunakan perangkat lunak berbasis web yang memanfaatkan data medis untuk memprediksi kemungkinan seseorang terkena penyakit jantung. Metode *data mining* dapat digunakan sebagai pendekatannya[7].

Penting untuk dicatat bahwa meskipun aplikasi ini dapat menjadi alat bantu yang berharga untuk diagnosa awal, namun tidak dapat sepenuhnya menggantikan peran dokter. Peran dokter tetap sangat penting dalam memberikan diagnosis, mengevaluasi hasil tes, dan menyusun rencana pengobatan yang sesuai. Aplikasi ini dirancang untuk membantu tenaga kesehatan non-dokter dan siswa kedokteran dengan memberikan prediksi awal, namun keputusan akhir dan tindakan medis tetap menjadi tanggung jawab dokter yang berkualifikasi. Oleh karena itu, peran kolaboratif antara teknologi dan tenaga medis sangat ditekankan untuk mencapai hasil yang terbaik dalam ranah kesehatan.

Penelitian ini terdiri dari beberapa tahap. Tahap pertama yaitu melakukan penyiapan data (*data preprocessing*). Data yang didapatkan bersumber dari berbagai industri kesehatan di beberapa kota di Amerika Serikat, maka dari itu perlu dilakukan integrasi data dan pembersihan data. Ketika data sudah disiapkan, tahapan selanjutnya yaitu pemilihan fitur. Pada penelitian ini, akan digunakan fitur yang berkaitan dengan penyakit jantung seperti tekanan darah, gula darah, dll. Tingkat keterkaitan tersebut dapat diukur menggunakan metode statistik. Tahapan ketiga yaitu pembuatan model klasifikasi sebagai metode analisis prediktif dalam *data mining*. Kemudian model tersebut dievaluasi untuk mendapatkan hasil yang terbaik. Tahapan terakhir yaitu melakukan integrasi model tersebut dengan perangkat lunak berbasis web untuk mempermudah penggunaanya dalam memprediksi apakah seseorang terkena penyakit jantung atau tidak berdasarkan data medis.

---

<sup>4</sup>"How Depression and Heart Disease Relate to Each Other", Roy Charles Ziegelstein, M.D., 2019, <https://www.hopkinsmedicine.org/health/conditions-and-diseases/depression-and-heart-disease#:~:text=People%20with%20depression%20may%20have,the%20chance%20of%20heart%20attack.>, diakses pada 1 Februari 2023.

<sup>5</sup>"Heart Disease Facts", Centers for Disease Control and Prevention, 2022, <https://www.shorturl.at/gzGI0>, diakses pada 27 Januari 2023.

## 1.2 Rumusan Masalah

Bagian ini akan diisi dengan penajaman dari masalah-masalah yang sudah diidentifikasi di bagian sebelumnya. Terdapat beberapa rumusan masalah dibawah ini yang nantinya akan diselesaikan permasalahannya yaitu:

1. Faktor apa saja yang berhubungan dengan penyakit jantung?
2. Bagaimana cara menerapkan tahapan *data mining* pada studi kasus?
3. Bagaimana cara melakukan analisis prediktif dengan presisi untuk memprediksi kemungkinan seseorang terkena penyakit jantung?
4. Bagaimana cara membuat perangkat lunak berbasis web untuk memprediksi apakah seseorang terkena penyakit jantung?

## 1.3 Tujuan

Akan dipaparkan secara lebih terperinci dan terstruktur apa yang menjadi tujuan pembuatan tugas akhir ini. Adapun tujuan pembuatan tugas akhir tersebut adalah:

1. Mempelajari faktor yang berhubungan seseorang memiliki penyakit jantung.
2. Mempelajari tahapan *data mining* pada studi kasus.
3. Melakukan prediksi dengan presisi terhadap data studi kasus.
4. Membuat perangkat lunak berbasis web untuk memprediksi apakah seseorang terkena penyakit jantung.

## 1.4 Batasan Masalah

Terdapat beberapa batasan masalah agar penelitian ini tidak terjadi penyimpangan maupun pelebaran dari pokok masalah sehingga penelitian ini terarah. Adapun batasan masalahnya yaitu:

1. Algoritma klasifikasi yang digunakan hanya meliputi *Naive Bayes*, *Support Vector Machine*, *Decision Tree* dan *Random Forest*.
2. Data yang dipakai bukanlah data dengan demografi Indonesia.
3. Aplikasi ini digunakan hanya digunakan untuk penyaringan penyakit jantung (*filtering*) saja, dan tidak menggantikan peranan dokter.

## 1.5 Metodologi

Metodologi yang digunakan dalam pembuatan tugas akhir ini adalah sebagai berikut:

1. Melakukan studi literatur tentang *domain knowledge* dari penyakit jantung.
2. Melakukan studi literatur tentang statistika deskriptif untuk menganalisa data penyakit jantung.
3. Melakukan studi literatur tentang *data mining* khususnya prediktif analisis.
4. Melakukan studi literatur tentang teknik klasifikasi.
5. Melakukan studi literatur tentang *tools* dan *library machine learning* (algoritma klasifikasi).
6. Melakukan eksplorasi data penyakit jantung.
7. Melakukan penyiapan data penyakit jantung.
8. Membuat model klasifikasi untuk memprediksi penyakit jantung.
9. Merancang *website* yang akan dibuat mulai dari *mock up*, arsitektur hingga integrasi dengan model.
10. Melakukan integrasi model dengan *interface* dari *website*.
11. Menulis dokumen tugas akhir.

## 1.6 Sistematika Pembahasan

Sistematika pembahasan dibuat untuk mempermudah dalam penyusunan dan pembacaan tugas akhir ini. Sistematika pembahasannya yaitu sebagai berikut:

1. Bab 1 Pendahuluan

Pada bab ini membahas tentang pengantar tugas akhir yang berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika pembahasan.

2. Bab 2 Landasan Teori

Pada bab ini membahas tentang studi literatur yang mencakup penjelasan dari konsep-konsep yang digunakan, seperti teori penyakit jantung, data, statistika, *data mining*, *data preprocessing*, transformasi data, visualisasi data dan algoritma klasifikasi.

3. Bab 3 Pengumpulan, Eksplorasi dan Penyiapan Data

Pada bab ini membahas mengenai bagaimana cara data didapatkan, isi dan informasi dari data tersebut, dan penyiapan data agar dapat diterima oleh algoritma klasifikasi dengan baik.

4. Bab 4 Evaluasi Model *Random Forest*

Pada bab ini membahas mengenai hasil dari klasifikasi dari algoritma *random forest*.

5. Bab 5 Perancangan *Website* dan Implementasinya

Pada bab ini membahas mengenai fitur yang tersedia pada perangkat lunak dan juga memperlihatkan cara kerja dari perangkat lunak yang dibangun beserta hasilnya.

6. Bab 6 Kesimpulan dan Saran

Pada bab ini membahas tentang kesimpulan dari keseluruhan penelitian yang sudah dilakukan dan saran yang membangun yang bisa diterapkan pada pengembangan selanjutnya.