

BAB I

PENDAHULUAN

1.1. Latar Belakang

Meningkatnya kasus penyakit jantung berkaitan erat dengan perubahan gaya hidup masyarakat modern. Penyakit jantung masih menjadi salah satu penyebab kematian tertinggi di dunia dan merupakan tantangan besar dalam bidang kesehatan. Penyakit kardiovaskular tetap menjadi salah satu penyumbang angka kematian tertinggi secara global. Kondisi ini menunjukkan bahwa meskipun perkembangan teknologi medis semakin maju, penyakit jantung masih menjadi ancaman serius yang memerlukan perhatian lebih [1].

Saat ini, penyakit jantung tidak hanya dialami oleh kelompok usia lanjut, tetapi juga mulai banyak ditemukan pada kelompok usia produktif. Peningkatan kasus penyakit kardiovaskular pada kelompok usia muda menjadi perhatian karena dapat menurunkan kualitas hidup masyarakat, sekaligus berdampak pada produktivitas dan beban pembiayaan kesehatan [2].

Kurangnya aktivitas fisik, kebiasaan duduk dalam waktu yang lama, pola makan yang tidak sehat, kebiasaan merokok, konsumsi alkohol, serta stres merupakan beberapa faktor yang dapat meningkatkan risiko penyakit jantung. Hasil penelitian *Scientific Reports* menunjukkan bahwa kombinasi faktor-faktor tersebut dapat berkontribusi terhadap kerusakan pembuluh darah dan meningkatkan risiko terjadinya kardiovaskular [3].

Gaya hidup yang tidak sehat dalam jangka panjang dapat memicu berbagai faktor risiko, seperti obesitas, hipertensi, dislipidemia, dan peningkatan

kadar kolesterol. kondisi tersebut berkembang secara bertahap dan sering kali tidak menimbulkan gejala pada tahap awal. Oleh karena itu, deteksi dini terhadap faktor-faktor risiko menjadi langkah penting untuk mencegah terjadinya penyakit jantung yang lebih serius [4].

Seiring berkembangnya pendekatan *preventif* dalam bidang kesehatan, analisis risiko sejak dini menjadi salah satu strategi yang dapat digunakan untuk meningkatkan kesadaran masyarakat terhadap kondisi kesehatannya. Pemetaan tingkat risiko berdasarkan karakteristik individu dan pola gaya hidup dapat membantu mengidentifikasi tingkat risiko sebelum munculnya kondisi yang lebih serius [5].

Salah satu pendekatan yang dapat digunakan untuk melakukan analisis tersebut adalah *Machine Learning*, yaitu metode yang memungkinkan sistem mempelajari pola dari data untuk menghasilkan prediksi atau *Klasifikasi*. Dalam penelitian ini, algoritma *Random Forest* digunakan untuk melakukan *Klasifikasi* tingkat risiko penyakit jantung berdasarkan karakteristik pengguna dan faktor gaya hidup. *Random Forest* merupakan algoritma *ensemble learning* yang menggabungkan beberapa *decision tree* dan menentukan hasil *Klasifikasi* berdasarkan penggabungan prediksi dari sejumlah pohon keputusan [6]. Penerapan algoritma *Random Forest* dalam *Klasifikasi* risiko penyakit jantung juga telah dilakukan pada penelitian terdahulu dan menunjukkan bahwa algoritma tersebut dapat digunakan untuk mengolah berbagai faktor yang berkaitan dengan risiko penyakit jantung [7].

Pentingnya upaya pencegahan juga didukung oleh penelitian yang

menyatakan bahwa sekitar 80% kasus penyakit jantung koroner dapat dicegah melalui perubahan gaya hidup yang sehat [2]. Hal ini menunjukkan bahwa penerapan pola hidup sehat memiliki peran yang sangat besar dalam mengurangi risiko terjadinya penyakit jantung.

Kemajuan teknologi digital memberikan peluang dalam pengembangan aplikasi kesehatan untuk mendukung deteksi dini dan pencegahan penyakit. Melalui perangkat seluler, pengguna dapat mengakses informasi kesehatan dan melakukan pemantauan terhadap faktor-faktor yang berkaitan dengan kondisi kesehatannya. Pemanfaatan teknologi tersebut dapat menjadi salah satu sarana untuk membantu pengguna memahami kondisi kesehatan serta menerapkan pola hidup yang lebih sehat [8].

Berdasarkan uraian tersebut, diperlukan suatu aplikasi yang mampu melakukan *Klasifikasi* risiko penyakit jantung berdasarkan faktor gaya hidup serta memberikan rekomendasi pencegahan yang sesuai dengan kondisi pengguna. Oleh karena itu, penelitian ini mengembangkan aplikasi *Heartify* yang menerapkan *Machine Learning* dengan algoritma *Random Forest* untuk melakukan *Klasifikasi* tingkat risiko penyakit jantung berdasarkan karakteristik pengguna dan faktor gaya hidup. Hasil *Klasifikasi* digunakan sebagai dasar dalam memberikan informasi dan rekomendasi pencegahan yang sesuai dengan kondisi pengguna. Sistem ini diharapkan dapat membantu masyarakat mengenali tingkat risiko penyakit jantung sejak dini serta menjadi media edukasi dalam menerapkan pola hidup sehat sebagai upaya pencegahan penyakit jantung.

1.2. Rumusan Masalah dan Batasan Masalah

1.2.1 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang dan membangun aplikasi *Heartify* berbasis mobile yang dapat melakukan analisis risiko penyakit jantung berdasarkan pola gaya hidup pengguna?
2. Bagaimana menerapkan metode Machine Learning dengan algoritma *Random Forest* dalam melakukan *Klasifikasi* tingkat risiko penyakit jantung berdasarkan karakteristik pengguna dan faktor-faktor gaya hidup?
3. Bagaimana memberikan rekomendasi pencegahan yang sesuai dengan tingkat risiko dan faktor risiko utama pengguna?
4. Bagaimana mengetahui tingkat kelayakan dan usability aplikasi *Heartify* dalam membantu pengguna melakukan analisis risiko dan pencegahan penyakit jantung?

1.2.2 Batasan Masalah

Agar penelitian lebih terarah dan tidak terlalu luas, batasan masalah dalam penelitian ini adalah:

1. Aplikasi *Heartify* berfokus pada analisis risiko dan pencegahan penyakit jantung berdasarkan pola gaya hidup, bukan sebagai alat diagnosis medis.
2. Analisis risiko dilakukan berdasarkan beberapa faktor, yaitu usia, jenis kelamin, indeks massa tubuh (BMI), aktivitas fisik, kebiasaan merokok, lingkungan merokok, pola tidur, tingkat stres, dan pola makan.
3. Penentuan tingkat risiko dilakukan menggunakan metode Machine Learning

dengan algoritma *Random Forest* yang dilatih menggunakan dataset penelitian. Proses pengembangan model meliputi preprocessing data, pembagian data, pelatihan, dan pengujian model.

4. Hasil analisis digunakan untuk memberikan tingkat risiko (rendah, sedang, atau tinggi), faktor risiko utama, serta rekomendasi pencegahan yang sesuai dengan kondisi pengguna.
5. Pengujian aplikasi dibatasi pada pengujian fungsional dan usability untuk mengetahui kesesuaian fungsi serta kemudahan penggunaan aplikasi oleh pengguna.

1.3. Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan dari penelitian ini adalah:

1. Merancang dan membangun aplikasi *Heartify* berbasis mobile untuk melakukan penilaian risiko gaya hidup sebagai upaya pencegahan penyakit jantung.
2. Menerapkan algoritma *Random Forest* untuk mempelajari pola dari data faktor risiko penyakit jantung dan menghasilkan prediksi tingkat risiko pengguna.
3. Menyajikan rekomendasi pencegahan yang dipersonalisasi (personalized) serta pengingat (reminder) berdasarkan Faktor Risiko Utama (Top Factor) gaya hidup masing-masing pengguna.
4. Melakukan pengujian sistem (uji kelayakan dan usability) kepada pengguna untuk memastikan aplikasi berjalan dengan baik, ramah pengguna, dan

efektif digunakan dalam memantau kesehatan.

1.3.2 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak, antara lain:

1. Manfaat Bagi Peneliti

- a. Menambah pengetahuan dan pengalaman dalam mengembangkan aplikasi kesehatan berbasis Machine Learning dengan algoritma *Random Forest*.
- b. Menerapkan ilmu yang diperoleh selama perkuliahan, khususnya dalam bidang pengembangan aplikasi, pengolahan data, dan penerapan Machine Learning.
- c. Menjadi pengalaman dalam mengimplementasikan teknologi informasi pada bidang kesehatan..

2. Manfaat Bagi Masyarakat

- a. Membantu masyarakat mengetahui tingkat risiko penyakit jantung sejak dini berdasarkan faktor gaya hidup.
- b. Memberikan rekomendasi pencegahan yang sesuai dengan kondisi dan tingkat risiko pengguna.
- c. Meningkatkan kesadaran masyarakat untuk menerapkan pola hidup sehat sebagai upaya mencegah penyakit jantung.

3. Manfaat Bagi Institusi

- a. Menambah koleksi referensi ilmiah mengenai pengembangan aplikasi kesehatan berbasis Machine Learning dengan algoritma *Random*

Forest.

- b. Menjadi bahan referensi bagi mahasiswa dalam penelitian yang berkaitan dengan analisis risiko penyakit jantung, penerapan Machine Learning, algoritma *Random Forest*, maupun pengembangan aplikasi kesehatan.
- c. Mendukung pengembangan ilmu pengetahuan dan teknologi, khususnya dalam penerapan teknologi informasi di bidang kesehatan.

1.4. Tinjauan Pustaka

Penelitian ini bertujuan mengembangkan aplikasi mobile berbasis Android yang dirancang untuk membantu masyarakat melakukan analisis risiko penyakit jantung berdasarkan pola gaya hidup. Aplikasi ini diharapkan dapat menjadi media yang memudahkan pengguna dalam mengenali tingkat risiko penyakit jantung sejak dini sekaligus memperoleh informasi mengenai langkah-langkah pencegahan yang sesuai dengan kondisi masing-masing. Aplikasi yang dikembangkan menerapkan metode Machine Learning dengan algoritma *Random Forest* untuk melakukan *Klasifikasi* tingkat risiko penyakit jantung berdasarkan karakteristik pengguna dan faktor gaya hidup. *Random Forest* merupakan algoritma ensemble learning yang menggabungkan sejumlah decision tree untuk menghasilkan prediksi berdasarkan pola yang dipelajari dari data [7]. Proses analisis dilakukan dengan memanfaatkan data yang dimasukkan oleh pengguna, seperti usia, jenis kelamin, indeks massa tubuh (BMI), aktivitas fisik, pola makan, kebiasaan merokok, kualitas tidur, dan tingkat stres.

Aplikasi juga menyediakan fitur pemantauan aktivitas, riwayat hasil

analisis, pengingat, serta pemantauan tingkat stres untuk membantu pengguna menerapkan gaya hidup sehat secara berkelanjutan. Dengan adanya fitur-fitur tersebut, aplikasi tidak hanya berfungsi sebagai media deteksi dini, tetapi juga sebagai sarana edukasi dan pendamping dalam upaya pencegahan penyakit jantung melalui penerapan pola gaya hidup yang lebih sehat.

Penelitian [9] dalam artikel *A Utility-Based Machine Learning-Driven Personalized Lifestyle Recommendation for Cardiovascular Disease Prevention* bertujuan mengembangkan model rekomendasi gaya hidup yang dipersonalisasi untuk membantu pencegahan penyakit kardiovaskular. Penelitian ini mengintegrasikan analisis faktor risiko dengan rekomendasi perubahan gaya hidup sehingga setiap individu memperoleh saran yang sesuai dengan kondisi kesehatannya. Hasil penelitian menunjukkan bahwa rekomendasi yang disesuaikan dengan karakteristik pengguna mampu meningkatkan efektivitas upaya pencegahan penyakit kardiovaskular dibandingkan rekomendasi yang bersifat umum. Penelitian ini juga menekankan pentingnya identifikasi faktor risiko sejak dini sebagai dasar dalam menentukan strategi pencegahan yang tepat.

Penelitian [10] dalam artikel penelitian *Data Analysis and Predictive Modelling on Heart Disease based on People's Lifestyle*, faktor gaya hidup memiliki hubungan yang signifikan dengan risiko penyakit jantung. Penelitian tersebut memanfaatkan teknik analisis data dan *machine learning* untuk mengidentifikasi pengaruh berbagai faktor gaya hidup terhadap kemungkinan terjadinya penyakit jantung. Hasil penelitian menunjukkan bahwa informasi mengenai kebiasaan hidup seseorang dapat dimanfaatkan sebagai dasar dalam

membangun sistem prediksi risiko penyakit jantung.

Penelitian yang berjudul *Predicting Heart Disease Risk with Lifestyle Factors: Uncovering Vital Predictors via Feature Selection and Machine Learning Techniques* membahas penerapan metode *machine learning* dalam memprediksi risiko penyakit jantung berdasarkan faktor gaya hidup. Penelitian tersebut menggunakan teknik *feature selection* untuk menentukan variabel yang paling berpengaruh terhadap risiko penyakit jantung, kemudian memanfaatkan algoritma *machine learning* untuk meningkatkan akurasi proses prediksi. Hasil penelitian menunjukkan bahwa pemilihan variabel yang tepat mampu meningkatkan performa model dalam mengidentifikasi individu yang memiliki risiko tinggi terhadap penyakit jantung. Penelitian tersebut menunjukkan bahwa faktor gaya hidup memiliki peran penting dalam proses analisis risiko penyakit jantung. Namun, hasil prediksi yang dihasilkan masih berfokus pada tingkat akurasi model dan belum menjelaskan proses pengambilan keputusan secara rinci kepada pengguna [11].

Penelitian yang berjudul *Pemeriksaan Kesehatan Jantung dalam Rangka Pengembangan Aplikasi Screening Jantung Berbasis Android* menjelaskan proses pengembangan aplikasi screening jantung yang memanfaatkan data hasil pemeriksaan kesehatan masyarakat. Penelitian tersebut diawali dengan pengumpulan data pemeriksaan kesehatan terhadap 500 partisipan yang memiliki berbagai faktor risiko, seperti hipertensi, diabetes melitus, obesitas, dan kebiasaan merokok. Data tersebut kemudian digunakan sebagai dasar dalam pengembangan aplikasi *Android* yang dirancang untuk membantu proses deteksi dini serta

memberikan prediksi dan rekomendasi kesehatan jantung menggunakan algoritma machine learning. Selain itu, aplikasi juga dirancang memiliki fitur pengingat pemeriksaan kesehatan sebagai upaya meningkatkan kepatuhan pengguna terhadap pemantauan kesehatan secara berkala. Penelitian tersebut menunjukkan bahwa aplikasi berbasis *Android* memiliki potensi yang besar dalam mendukung deteksi dini penyakit jantung. Namun, aplikasi yang dikembangkan masih berorientasi pada hasil pemeriksaan kesehatan dan pemanfaatan machine learning sebagai metode prediksi [12].

Penelitian yang berjudul *Logistic Regression to Predict Heart Disease* mengembangkan sistem prediksi penyakit jantung menggunakan algoritma Logistic Regression dengan memanfaatkan data klinis dan demografis pasien. Tahapan penelitian meliputi data *preprocessing*, pemilihan fitur (*feature selection*), pelatihan model, serta evaluasi performa model untuk menghasilkan prediksi risiko penyakit jantung. Hasil penelitian menunjukkan bahwa *algoritma Logistic Regression* mampu memberikan tingkat akurasi yang baik sehingga berpotensi digunakan sebagai alat bantu dalam mendukung deteksi dini dan pengambilan keputusan di bidang kesehatan [13].

Penelitian yang berjudul "*Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung*" membandingkan performa algoritma *K-Nearest Neighbors* dan *Random Forest* dalam menghasilkan rekomendasi gaya hidup sehat untuk mencegah penyakit jantung. Penelitian tersebut menggunakan data faktor risiko penyakit jantung sebagai dasar proses *Klasifikasi*, kemudian membandingkan

tingkat akurasi kedua algoritma dalam memberikan rekomendasi kepada pengguna. Hasil penelitian menunjukkan bahwa kedua algoritma mampu digunakan untuk mendukung rekomendasi gaya hidup sehat, namun *Random Forest* memiliki tingkat akurasi yang lebih tinggi dibandingkan *K-Nearest Neighbors* dalam mengKlasifikasikan data penyakit jantung. penelitian tersebut berhasil menghasilkan rekomendasi berdasarkan metode *machine learning*, proses pengambilan keputusan yang dilakukan oleh algoritma masih sulit dipahami oleh pengguna karena hasil rekomendasi diperoleh melalui proses komputasi yang kompleks [14].

Penelitian yang berjudul *Aplikasi Prediksi Serangan Jantung Untuk Warga Kelurahan Pelawi Utara* mengembangkan aplikasi berbasis *Android* yang digunakan untuk memprediksi kemungkinan terjadinya serangan jantung pada masyarakat. Aplikasi tersebut memanfaatkan data kesehatan pengguna sebagai dasar dalam proses prediksi sehingga dapat membantu masyarakat mengetahui tingkat risiko penyakit jantung secara lebih cepat. Hasil penelitian menunjukkan bahwa aplikasi mampu mendukung proses deteksi dini dan memberikan informasi mengenai risiko serangan jantung sehingga dapat membantu pengguna mengambil langkah pencegahan lebih awal. Penelitian tersebut menunjukkan bahwa aplikasi mobile memiliki potensi yang besar sebagai media deteksi dini penyakit jantung. Namun, aplikasi yang dikembangkan masih berfokus pada proses prediksi risiko berdasarkan data kesehatan pengguna [15].

Berikut adalah tabel gap analisis yang membandingkan penelitian yang akan dilakukan dengan penelitian-penelitian sebelumnya. Tabel ini disusun berdasarkan

tahun penelitian.

Tabel 1. 1 Gap Penelitian

No.	Hasil Penelitian Terdahulu	Pembeda dengan Penelitian Saat Ini
1	Judul: <i>A Utility-Based Machine Learning-Driven Personalized Lifestyle Recommendation for Cardiovascular Disease Prevention</i> Hasil: Penelitian mengembangkan model rekomendasi gaya hidup yang dipersonalisasi berdasarkan analisis risiko penyakit kardiovaskular.	Penelitian berfokus pada pengembangan model rekomendasi gaya hidup, sedangkan penelitian ini mengembangkan aplikasi mobile <i>Heartify</i> yang menggunakan algoritma <i>Random Forest</i> untuk mengKlasifikasikan tingkat risiko penyakit jantung berdasarkan faktor gaya hidup serta memberikan rekomendasi pencegahan.
2.	Judul: <i>Data Analysis and Predictive Modelling on Heart Disease based on People's Lifestyle</i> Hasil: Penelitian mengembangkan model prediksi penyakit jantung berdasarkan faktor gaya hidup menggunakan teknik <i>machine learning</i> dan visualisasi data. Faktor seperti merokok, pola makan, aktivitas fisik, obesitas, dan konsumsi alkohol digunakan untuk mengidentifikasi risiko penyakit jantung.	Penelitian tersebut berfokus pada analisis data, pemodelan prediksi, dan visualisasi, sedangkan penelitian ini mengimplementasikan algoritma <i>Random Forest</i> ke dalam aplikasi mobile <i>Heartify</i> . Aplikasi menyediakan analisis risiko, rekomendasi pencegahan, riwayat analisis, pemantauan aktivitas fisik, pengingat, dan manajemen stres.
3	Judul: <i>Predicting Heart Disease Risk with Lifestyle Factors: Uncovering Vital Predictors via Feature Selection and Machine Learning Techniques</i> Hasil: Penelitian menerapkan <i>feature selection</i> dan metode <i>machine</i>	Penelitian tersebut berfokus pada pemilihan fitur dan peningkatan performa model prediksi, sedangkan penelitian ini menggunakan algoritma <i>Random Forest</i> untuk mengKlasifikasikan tingkat risiko penyakit jantung berdasarkan faktor gaya

No.	Hasil Penelitian Terdahulu	Pembeda dengan Penelitian Saat Ini
	<i>learning</i> untuk memprediksi risiko penyakit jantung berdasarkan faktor gaya hidup. Pemilihan fitur yang relevan digunakan untuk meningkatkan kemampuan model dalam mengidentifikasi individu yang berisiko.	hidup dan mengintegrasikannya ke dalam aplikasi mobile yang dilengkapi rekomendasi pencegahan.
4	Judul: Pemeriksaan Kesehatan Jantung dalam Rangka Pengembangan Aplikasi <i>Screening</i> Jantung Berbasis Android Hasil: Penelitian mengembangkan aplikasi <i>screening</i> jantung berbasis Android dengan memanfaatkan data hasil pemeriksaan kesehatan dari 500 partisipan. Aplikasi digunakan untuk membantu deteksi dini, memberikan prediksi risiko, rekomendasi kesehatan, dan pengingat pemeriksaan.	Penelitian tersebut menggunakan data hasil pemeriksaan kesehatan untuk proses <i>screening</i>, sedangkan penelitian ini berfokus pada analisis risiko penyakit jantung berdasarkan pola gaya hidup, seperti aktivitas fisik, pola makan, merokok, tidur, dan stres, menggunakan algoritma <i>Random Forest</i> dalam aplikasi <i>Heartify</i>.
5	Judul: <i>Logistic Regression to Predict Heart Disease</i> Hasil: Penelitian menunjukkan bahwa metode <i>Logistic Regression</i> dapat digunakan untuk membangun model prediksi penyakit jantung dengan tingkat akurasi yang baik sebagai alat bantu deteksi dini.	Penelitian tersebut menggunakan <i>Logistic Regression</i> untuk melakukan prediksi, sedangkan penelitian ini menggunakan algoritma <i>Random Forest</i> untuk mengKlasifikasikan tingkat risiko penyakit jantung berdasarkan faktor gaya hidup dan mengintegrasikannya ke dalam aplikasi mobile yang menyediakan rekomendasi pencegahan.
6	Judul: Perbandingan Algoritma K-	Penelitian tersebut berfokus pada

No.	Hasil Penelitian Terdahulu	Pembeda dengan Penelitian Saat Ini
	Nearest Neighbors dan <i>Random Forest</i> untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung Hasil: Penelitian membandingkan algoritma K-Nearest Neighbors dan <i>Random Forest</i> dalam menghasilkan rekomendasi gaya hidup sehat berdasarkan faktor risiko penyakit jantung.	perbandingan algoritma untuk menghasilkan rekomendasi gaya hidup, sedangkan penelitian ini berfokus pada implementasi algoritma <i>Random Forest</i> dalam aplikasi mobile <i>Heartify</i> untuk mengKlasifikasikan tingkat risiko penyakit jantung berdasarkan pola gaya hidup dan memberikan rekomendasi pencegahan.
7	Judul: Aplikasi Prediksi Serangan Jantung Untuk Warga Kelurahan Pelawi Utara Hasil: Penelitian mengembangkan aplikasi yang membantu masyarakat melakukan deteksi dini risiko serangan jantung dan memberikan informasi mengenai tingkat risiko berdasarkan data pengguna.	Penelitian tersebut berfokus pada prediksi risiko serangan jantung, sedangkan penelitian ini mengembangkan aplikasi <i>Heartify</i> untuk melakukan <i>Klasifikasi</i> risiko penyakit jantung berdasarkan pola gaya hidup menggunakan algoritma <i>Random Forest</i>, serta menyediakan rekomendasi pencegahan, riwayat analisis, pemantauan aktivitas, pengingat, dan manajemen stres.

1.5. Data Penelitian

Pengumpulan data dalam penelitian ini dilakukan melalui studi literatur, pengumpulan dataset, serta proses konsultasi dengan tenaga kesehatan di Rumah Sakit Kardinah Kota Tegal. Studi literatur digunakan untuk mengidentifikasi faktor-faktor yang berkaitan dengan risiko penyakit jantung, seperti usia, jenis kelamin, indeks massa tubuh (BMI), aktivitas fisik, pola makan, kebiasaan merokok, paparan asap rokok, tingkat stres, dan durasi tidur.

Dataset digunakan untuk melatih model dalam mengenali pola hubungan antara faktor-faktor risiko dengan tingkat risiko penyakit jantung. Model yang telah dilatih kemudian digunakan untuk melakukan *Klasifikasi* terhadap data pengguna baru pada aplikasi *Heartify*.

Selain pengumpulan dataset, peneliti juga melakukan konsultasi dengan tenaga kesehatan di Rumah Sakit Kardinah Kota Tegal. Konsultasi dilakukan untuk memperoleh masukan mengenai kesesuaian variabel yang digunakan, pemahaman terhadap faktor risiko penyakit jantung, interpretasi hasil *Klasifikasi*, serta penyusunan rekomendasi pencegahan yang ditampilkan oleh aplikasi. Hasil konsultasi digunakan sebagai bahan pendukung dalam pengembangan dan penyempurnaan aplikasi agar informasi yang diberikan tetap sesuai dengan konteks kesehatan dan tidak digunakan sebagai pengganti diagnosis medis.

1. Spesifikasi Data yang Digunakan

a. Data Faktor Risiko Penyakit Jantung (Input Analisis)

Data faktor risiko penyakit jantung merupakan data yang digunakan sebagai masukan dalam proses pelatihan dan pengujian model *Random Forest*. Variabel yang digunakan meliputi usia, jenis kelamin, tinggi badan, berat badan, indeks massa tubuh (BMI), aktivitas fisik, pola makan, kebiasaan merokok, paparan asap rokok, tingkat stres, dan durasi tidur.

Variabel-variabel tersebut digunakan untuk menggambarkan kondisi dan pola gaya hidup yang berkaitan dengan risiko penyakit jantung. Data tinggi badan dan berat badan digunakan untuk menghitung nilai BMI apabila nilai BMI belum tersedia dalam dataset. Selanjutnya, data yang telah diproses

digunakan sebagai fitur masukan atau *input features* bagi model *Random Forest*.

Adapun variabel yang digunakan dalam proses analisis dapat dilihat pada tabel berikut.

b. Sumber Data Set

Dataset yang digunakan dalam penelitian ini diperoleh dari sumber dataset yang relevan dengan penelitian penyakit jantung dan faktor gaya hidup. Dataset digunakan sebagai dasar dalam proses pelatihan dan pengujian algoritma *Random Forest*. Pemilihan dataset dilakukan dengan mempertimbangkan kesesuaian variabel dengan kebutuhan aplikasi *Heartify*, kelengkapan data, jumlah data, serta ketersediaan label tingkat risiko penyakit jantung.

Apabila terdapat perbedaan nama atau format variabel antara dataset dan variabel yang digunakan dalam aplikasi, dilakukan proses penyesuaian dan pemetaan data. Proses tersebut bertujuan agar data yang digunakan pada saat pelatihan model memiliki format yang sama dengan data masukan dari pengguna aplikasi.

Sumber dataset yang digunakan dalam penelitian ini dicantumkan secara jelas berdasarkan nama dataset, sumber atau platform penyedia, jumlah data, variabel yang tersedia, serta label yang digunakan dalam proses *Klasifikasi*.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Umur	Jenis_Kelamin	BMI	Merokok	Lingkungan_Perokok	Olahraga_Menit	Tidur_Jam	Stres	Pola_Makan	Target_Risiko			
1	56	Perempuan	16.9	Perokok Aktif	Ya	186	4	Sedang	Cukup Sehat	Risiko Sedang			
2	69	Laki-laki	26.8	Perokok Aktif	Ya	292	10	Rendah	Cukup Sehat	Risiko Tinggi			
3	46	Laki-laki	19.0	Tidak Merokok	Ya	60	7	Sedang	Tidak Sehat	Risiko Sedang			
4	32	Perempuan	18.3	Tidak Merokok	Tidak	197	4	Sedang	Sehat	Risiko Rendah			
5	60	Laki-laki	22.5	Tidak Merokok	Tidak	294	9	Sedang	Cukup Sehat	Risiko Sedang			
6	25	Laki-laki	17.7	Perokok Aktif	Ya	10	8	Sedang	Sehat	Risiko Sedang			
7	38	Perempuan	17.8	Tidak Merokok	Tidak	61	10	Tinggi	Cukup Sehat	Risiko Sedang			
8	56	Laki-laki	21.9	Perokok Pasif	Tidak	225	5	Sedang	Cukup Sehat	Risiko Sedang			
9	36	Laki-laki	34.6	Perokok Pasif	Tidak	152	8	Tinggi	Tidak Sehat	Risiko Tinggi			
10	40	Perempuan	19.3	Perokok Aktif	Tidak	198	7	Rendah	Cukup Sehat	Risiko Rendah			
11	28	Perempuan	16.3	Tidak Merokok	Ya	51	8	Tinggi	Sehat	Risiko Sedang			
12	28	Laki-laki	30.5	Perokok Aktif	Ya	259	7	Sedang	Tidak Sehat	Risiko Sedang			
13	41	Laki-laki	31.3	Tidak Merokok	Tidak	58	7	Rendah	Cukup Sehat	Risiko Sedang			
14	70	Perempuan	22.6	Tidak Merokok	Ya	71	7	Rendah	Sehat	Risiko Rendah			
15	53	Perempuan	24.8	Perokok Aktif	Ya	147	9	Sedang	Sehat	Risiko Sedang			
16	57	Perempuan	28.3	Tidak Merokok	Ya	62	7	Rendah	Sehat	Risiko Sedang			
17	41	Perempuan	16.9	Perokok Aktif	Tidak	242	7	Rendah	Sehat	Risiko Rendah			

Gambar 1. 1 Data Set

Gambar 1.1 menunjukkan dataset yang digunakan dalam penelitian untuk proses pelatihan dan pengujian model *Random Forest*. Dataset terdiri dari beberapa variabel yang berkaitan dengan faktor risiko penyakit jantung, yaitu umur, jenis kelamin, BMI, status merokok, lingkungan perokok, aktivitas olahraga, durasi tidur, tingkat stres, dan pola makan. Dataset juga memiliki variabel Target_Risiko yang digunakan sebagai label atau target *Klasifikasi*, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi.

c. Data Hasil *Klasifikasi*

Data hasil *Klasifikasi* merupakan label yang digunakan sebagai target dalam proses pelatihan model *Random Forest*. Label tersebut menggambarkan tingkat risiko penyakit jantung yang terdiri atas Risiko Rendah, Risiko Sedang, dan Risiko Tinggi.

Model *Random Forest* mempelajari hubungan antara fitur-fitur faktor risiko dengan label tingkat risiko berdasarkan pola yang terdapat pada dataset. Setelah proses pelatihan selesai, model digunakan untuk memprediksi tingkat

risiko berdasarkan data pengguna baru yang dimasukkan melalui aplikasi *Heartify*.

Hasil *Klasifikasi* yang diperoleh dari model ditampilkan kepada pengguna sebagai informasi mengenai tingkat risiko penyakit jantung. Hasil tersebut digunakan sebagai sarana deteksi dini dan edukasi kesehatan, bukan sebagai diagnosis medis.

- d. Data Penanganan dan Edukasi: Data ini berisi informasi mengenai pola hidup sehat, aktivitas fisik yang dianjurkan, pola makan, pengelolaan stres, kualitas tidur, serta kebiasaan lain yang dapat membantu mengurangi risiko penyakit jantung. Informasi tersebut diperoleh dari literatur ilmiah dan telah dikonsultasikan dengan tenaga kesehatan agar sesuai dengan rekomendasi medis.

2. Metode dan Prosedur Pengumpulan Data

a. Permohonan Penelitian

Peneliti mengajukan surat permohonan penelitian resmi kepada pihak Rumah Sakit Kardinah Kota Tegal sebagai dasar pelaksanaan konsultasi dengan tenaga kesehatan. Konsultasi dilakukan untuk memperoleh masukan mengenai variabel faktor risiko penyakit jantung, interpretasi hasil *Klasifikasi*, serta kesesuaian informasi rekomendasi pencegahan yang ditampilkan dalam aplikasi.

b. Studi Literatur

Peneliti mengumpulkan referensi yang berkaitan dengan penyakit jantung, faktor risiko penyakit jantung, Machine Learning, algoritma *Random Forest*,

pengolahan dataset, evaluasi model *Klasifikasi*, serta pencegahan penyakit jantung. Referensi diperoleh dari jurnal ilmiah, buku, pedoman kesehatan, dan sumber terpercaya lainnya.

Studi literatur digunakan untuk menentukan variabel penelitian, memahami hubungan antara pola gaya hidup dan risiko penyakit jantung, menentukan tahapan pengolahan data, serta mendukung penyusunan informasi edukasi dan rekomendasi pencegahan.

c. Pengumpulan Dataset

Dataset dikumpulkan dari sumber yang telah ditentukan sesuai dengan kebutuhan penelitian. Dataset yang diperoleh kemudian diperiksa untuk mengetahui jumlah data, jenis variabel, kelengkapan data, keberadaan nilai kosong, data duplikat, serta kesesuaian label tingkat risiko.

Data yang tidak sesuai dengan kebutuhan penelitian dapat dilakukan proses pembersihan atau penghapusan berdasarkan kriteria yang telah ditentukan. Tahap ini dilakukan agar dataset yang digunakan dalam proses pelatihan dan pengujian memiliki kualitas data yang baik.

d. Preprocessing Data

Preprocessing data dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pelatihan model *Random Forest*. Tahapan preprocessing meliputi pemeriksaan data kosong, penghapusan data duplikat, penyesuaian format data, pengubahan data kategorikal menjadi bentuk numerik, serta penyesuaian nama dan struktur variabel.

Data tinggi badan dan berat badan juga dapat digunakan untuk menghitung nilai BMI. Setelah seluruh data memiliki format yang sesuai, dataset dipisahkan menjadi fitur masukan dan label target. Fitur masukan digunakan untuk menggambarkan faktor risiko pengguna, sedangkan label target digunakan sebagai kelas yang akan diprediksi oleh model.

e. Pembagian Data Training dan Testing

Dataset yang telah melalui tahap preprocessing dibagi menjadi data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun dan melatih model *Random Forest*, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan *Klasifikasi* terhadap data yang belum pernah digunakan selama proses pelatihan.

Pembagian dataset dilakukan menggunakan perbandingan yang telah ditentukan, misalnya 80% data latih dan 20% data uji. Pembagian data dilakukan secara terstruktur agar data latih dan data uji dapat digunakan untuk memperoleh gambaran kemampuan model dalam melakukan prediksi terhadap data baru.

f. Pelatihan Model *Random Forest*

Data latih digunakan untuk melatih model *Random Forest*. Algoritma *Random Forest* membentuk sejumlah pohon keputusan (*decision tree*) yang bekerja secara bersama-sama untuk melakukan *Klasifikasi*. Setiap pohon menghasilkan prediksi berdasarkan pola data yang dipelajari, kemudian hasil prediksi dari seluruh pohon digabungkan untuk menentukan kelas akhir melalui mekanisme suara terbanyak.

Dalam penelitian ini, model *Random Forest* dilatih menggunakan variabel faktor risiko penyakit jantung sebagai fitur masukan dan tingkat risiko penyakit jantung sebagai label target. Model yang telah selesai dilatih kemudian disimpan dan diintegrasikan ke dalam backend aplikasi *Heartify* agar dapat digunakan untuk melakukan *Klasifikasi* terhadap data pengguna.

g. Pengujian dan Evaluasi Model

Pengujian model dilakukan menggunakan data uji yang tidak digunakan pada saat proses pelatihan. Pengujian bertujuan untuk mengetahui kemampuan model *Random Forest* dalam meng*Klasifikasi*kan tingkat risiko penyakit jantung berdasarkan faktor risiko pengguna.

Evaluasi model dilakukan menggunakan metrik evaluasi *Klasifikasi*, seperti accuracy, precision, recall, dan F1-score. Selain itu, confusion matrix dapat digunakan untuk melihat jumlah prediksi yang benar dan salah pada setiap kelas risiko. Hasil evaluasi digunakan untuk mengetahui performa model sebelum diimplementasikan ke dalam aplikasi *Heartify*.

h. Konsultasi dan Validasi Informasi

Konsultasi dengan tenaga kesehatan dilakukan untuk memperoleh masukan mengenai kesesuaian variabel faktor risiko, interpretasi kategori risiko, serta informasi rekomendasi pencegahan yang ditampilkan dalam aplikasi.

1.6. Alat Penelitian

Untuk memastikan bahwa faktor-faktor yang digunakan dalam *Klasifikasi* risiko sesuai dengan kondisi dan kaidah kesehatan, penelitian ini melibatkan tenaga kesehatan dari Rumah Sakit Kardinah Kota Tegal sebagai validator. Proses

validasi dilakukan terhadap faktor-faktor risiko yang digunakan sebagai variabel masukan, kategori tingkat risiko sebagai target *Klasifikasi*, serta rekomendasi pencegahan yang diberikan oleh aplikasi. Hasil validasi digunakan sebagai dasar untuk memastikan kesesuaian variabel dan informasi kesehatan sebelum digunakan dalam proses *Klasifikasi* menggunakan algoritma *Random Forest* dan diimplementasikan ke dalam sistem.

a. Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.5

Tabel 1. 2 Hardwer

No	Perangkat	Fungsi
1	Laptop	Digunakan untuk pengembangan aplikasi dan <i>backend</i> .
2	Smartphone <i>Android</i>	Digunakan untuk pengujian aplikasi.
3	Jaringan Internet	Digunakan untuk menghubungkan aplikasi dengan <i>server</i> dan <i>MongoDB Atlas</i> .

b. Perangkat Lunak Software

Perangkat lunak yang digunakan dalam penelitian ini ditunjukkan pada tabel 1.6

Tabel 1. 3 Software

No	Software	Fungsi
1	Flutter	Digunakan untuk mengembangkan aplikasi mobile berbasis <i>Android</i> .
2	Visual Studio	Digunakan sebagai editor kode dalam

No	Software	Fungsi
	Code	pengembangan aplikasi dan <i>backend</i> .
3	Python	Digunakan sebagai bahasa pemrograman <i>backend</i> . <i>Flask</i>
4	<i>Flask</i>	Digunakan untuk membangun <i>REST API</i> sebagai penghubung aplikasi dengan basis data.
5	Postman	Digunakan untuk menguji endpoint <i>API</i> sebelum diintegrasikan dengan aplikasi.
6	<i>MongoDB Atlas</i>	Digunakan sebagai basis data untuk menyimpan data pengguna, hasil analisis, dan riwayat penggunaan aplikasi.