

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit *Cardiovascular Disease* (CVD) atau *kardiovaskular* merupakan penyakit yang paling umum dan menjadi penyebab utama kematian di negara-negara maju. Penyakit *kardiovaskular* (PKV) menyerang jantung dan pembuluh darah. Di Indonesia pada tahun 1992, Kesehatan Rumah Tangga Nasional (SKRT) menegaskan bahwa Penyakit *Kardiovaskular* (PKV) menjadi penyebab utama kematian pada individu yang berusia di atas 40 tahun (Gusti Ayu Sri Ekayanti, 2019). Berdasarkan data yang dirilis oleh Organisasi Kesehatan Dunia (WHO), angka prevalensi penyakit *kardiovaskular* akibat tingginya kadar kolesterol dalam tubuh terus meningkat. Penyakit *kardiovaskular* yang disebabkan oleh kolesterol tinggi juga mengalami peningkatan di Indonesia, dengan 30% kasus komplikasi penyakit *kardiovaskular* (Husen dkk., 2022).

Kolesterol tinggi merupakan masalah kesehatan masyarakat utama di seluruh dunia. Beberapa faktor yang dapat mempengaruhi kadar kolesterol di dalam darah meliputi faktor keturunan, usia, jenis kelamin, kebiasaan merokok, konsumsi alkohol, kurangnya asupan sayuran dan buah, obesitas, diabetes melitus, tingkat stres, dan kebiasaan minum kopi berlebihan menurut (Permatasari et al., 2022). Penyebab kolesterol tinggi salah satunya adalah diabetes, karena Diabetes melitus tipe 2 bisa menyebabkan gangguan metabolisme lipid (dyslipidemia), termasuk peningkatan kadar kolesterol total (Saptaningtyas et al., 2022). Glukosa, sebagai gula darah utama, merangsang aktivasi AMP-activated protein kinase (AMPK) oleh insulin. AMPK menekan SREBP2, mengurangi produksi enzim HMGCR dan HMGCS yang terlibat dalam pembentukan kolesterol (Nurlita et al., 2020). Kolesterol berperan penting dalam mengatur fluiditas dan permeabilitas membran. Fluiditas membran merupakan kemampuan membran untuk bergerak atau bersifat fleksibel, serta permeabilitas membran, mengukur kemampuan membran untuk

memungkinkan zat, seperti molekul kecil, ion, atau molekul lainnya, untuk melintasi atau keluar dari sel, merupakan dua sifat penting yang memengaruhi struktur dan fungsi sel. Selain itu, darah juga mengandung lipoprotein plasma, senyawa lipid amfipatik yang berperan penting dalam tubuh dan diproduksi oleh hati. Lipoprotein ini memiliki dua bagian, satu yang menyukai air dan satu yang tidak, memungkinkannya membawa lemak dan kolesterol dalam darah. Peningkatan kadar kolesterol dalam darah disebut hiperkolesterolemia (Dana dkk., 2022). Kadar kolesterol darah yang tinggi atau hiperkolesterolemia dapat menyebabkan tekanan darah tinggi (Leonita suci, 2020) dan tingginya tingkat kolesterol dalam tubuh dapat meningkatkan risiko penyakit *kardiovaskular*, seperti serangan jantung dan stroke. Oleh karena itu, identifikasi dini dan prediksi risiko kolesterol tinggi pada individu dapat menjadi langkah krusial dalam upaya pencegahan penyakit tersebut.

Tujuan penelitian ini adalah untuk mengembangkan sebuah sistem prediksi risiko kolesterol tinggi yang dapat membedakan antara individu yang beresiko mengalami kadar kolesterol tinggi (hiperkolesterolemia) dan individu yang tidak beresiko. Penelitian ini diharapkan dapat memberikan kontribusi pada identifikasi dini individu yang memerlukan perhatian lebih dalam upaya pencegahan penyakit *kardiovaskular*.

Seiring dengan kemajuan teknologi informasi dan perkembangan dalam bidang ilmu data, pendekatan klasifikasi menggunakan metode *machine learning* semakin menjadi populer dilakukan dalam dunia kesehatan. Salah satu metode *machine learning* yang terbukti efektif untuk prediksi adalah Metode *Naive bayes*, yang diperbaiki dengan menerapkan teknik Laplacian Correction. Penelitian yang sama juga pernah dilakukan oleh Sekar Arum Nurhusni, Riza Ibnu Adam, dan Carudin pada Desember 2021 membahas klasifikasi kadar kolesterol dengan menerapkan ekstraksi ciri menggunakan *Moment Invariant* dan menggunakan algoritme *K-Nearest Neighbor* (KNN) untuk proses klasifikasi. Hasil penelitian ini menunjukkan akurasi sebesar 84.8485%. Meskipun demikian, perlu diperhatikan bahwa metode yang digunakan dalam penelitian ini memiliki kekurangan yang perlu diakui.

Kekurangan algoritma KNN melibatkan kebutuhan untuk menentukan jumlah tetangga terdekat dari data target, yang direpresentasikan oleh parameter K . Selain itu, proses pemilihan, eksperimen, dan penentuan jenis jarak yang digunakan, serta atribut mana yang digunakan untuk mendapatkan perhitungan jarak yang optimal, dapat menyebabkan kurangnya akurasi. Selain itu, biaya komputasi yang tinggi juga menjadi masalah karena memerlukan perhitungan jarak dari setiap contoh data pelatihan untuk setiap *query instance* (Saifur Rohman Cholil., 2021).

Selain metode *K-Nearest Neighbors* (KNN) yang telah dipertimbangkan dalam penelitian sebelumnya, salah satu alternatif yang dapat dijelajahi untuk klasifikasi kadar kolesterol adalah metode *Naive bayes*. Meskipun penelitian sebelumnya menggunakan KNN dengan hasil akurasi yang signifikan, *Naive bayes* memiliki keunggulan tersendiri yang meliputi penanganan data kuantitatif dan diskrit, ketahanan terhadap titik *noise* yang terisolasi, seperti pada titik yang di-rata-ratakan saat mengestimasi probabilitas bersyarat data. Metode ini hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter yang diperlukan untuk klasifikasi, seperti rata-rata, variasi, dan variabel. Selain itu, metode ini mampu menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang, memiliki kinerja cepat dan efisien dalam penggunaan ruang, serta stabil terhadap atribut yang tidak relevan (Cahya, 2022). Oleh karena itu, penentuan metode klasifikasi antara KNN dan *Naive bayes* sebaiknya dipertimbangkan dengan cermat, dan eksperimen empiris dapat membantu mengevaluasi kinerja relatif keduanya pada dataset dan kasus klasifikasi yang spesifik.

Naive bayes merupakan suatu pendekatan klasifikasi yang berdasarkan teorema probabilitas Bayes. Walaupun disebut "*Naive*" karena mengasumsikan independensi antar-fitur, metode ini tetap efektif dalam menggambarkan hubungan antara fitur-fitur tersebut. Metode *Naive bayes* menggunakan teknik klasifikasi dengan memanfaatkan konsep probabilitas sederhana untuk menghitung sejumlah probabilitas. Hal ini dilakukan dengan menggabungkan frekuensi dan kombinasi nilai dari suatu dataset tertentu (Tri dkk., 2021). Dalam

implementasinya, *Naïve bayes* menggunakan perhitungan probabilitas matematis untuk menentukan probabilitas keputusan yang benar berdasarkan informasi objek (Novaldy & Herliana, 2021). Dalam konteks penelitian ini, *Naive bayes* digunakan untuk meramalkan kemungkinan seseorang mengalami kolesterol tinggi berdasarkan data kesehatan yang tersedia. *Naive bayes Classifier* terbukti mampu melakukan pengklasifikasian secara efektif dibandingkan dengan beberapa algoritme pengklasifikasian lainnya, sebuah temuan yang diungkapkan dalam jurnal oleh Xhemali, Daniela, Chris J. Hinde, dan Roger G. Stone berjudul "*Naive bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages.*" Temuan tersebut menegaskan bahwa performa *Naive bayes* dalam klasifikasi halaman web pelatihan dapat bersaing dan bahkan mengungguli algoritme pengklasifikasian lain seperti decision trees dan neural networks (Rizki dkk., 2021).. Dengan demikian, penelitian ini memberikan dukungan terhadap keefektifan *Naive bayes Classifier* sebagai pilihan yang kompetitif dalam konteks klasifikasi halaman web pelatihan, menjadi alternatif metode untuk eksplorasi dan pemahaman data (Afif, 2020). Adapun kelebihan lain dari penggunaan *Naive bayes* untuk teknik klasifikasi terletak pada kemampuannya untuk menghasilkan estimasi parameter yang dibutuhkan dalam proses pengklasifikasian dengan hanya memerlukan jumlah data pelatihan yang relatif kecil (Tri dkk., 2021).

Untuk meningkatkan ketepatan prediksi, diperlukan tindakan perbaikan atau penyesuaian pada metode *Naive bayes*. Beberapa langkah perbaikan lain yang dapat diterapkan pada kinerja *Naive bayes* melibatkan teknik seperti *Laplacian correction* (Add-One Smoothing), seleksi fitur, normalisasi data, penanganan missing values, dan pemilihan model yang sesuai dengan jenis data. Keputusan menggunakan *Laplacian correction*, karena teknik tersebut dapat mengatasi masalah nol frekuensi pada data latih, yang dapat muncul ketika suatu nilai fitur tidak muncul pada kelas tertentu. Untuk mengatasi masalah ini dan memastikan bahwa penghitungan probabilitas menggunakan pengklasifikasi *Naive bayes* (NBC) tidak mengembalikan nilai 0 karena data yang hilang untuk kategori

tertentu dalam suatu kelas, dapat menggunakan teknik yang disebut Koreksi Laplacian atau Koreksi Laplacian (Akhyar, 2023) (Rizki dkk., 2021).

Sebuah penelitian berjudul "Eksplorasi Algoritme *Naive bayes Classifier* dengan Penggunaan Teknik *Laplacian correction*" yang dilakukan oleh Muhammad Rizki, Muhammad Arhami, dan Huzeni menggambarkan berhasilnya penerapan teknik *Laplacian correction* dalam meningkatkan kinerja algoritme *Naive bayes Classifier*. Dalam penelitian ini, algoritme *Naive bayes Classifier* awalnya menunjukkan tingkat akurasi sebesar 94.44%. Namun, setelah diterapkan teknik *Laplacian correction*, yang artinya tingkat akurasi meningkat, menegaskan keberhasilan teknik tersebut dalam memperbaiki dan membangkitkan nilai probabilitas akhir yang sebelumnya bernilai 0. Temuan ini menunjukkan bahwa penggunaan *Laplacian correction* dapat memberikan kontribusi signifikan dalam meningkatkan performa algoritme *Naive bayes Classifier*, sehingga relevan untuk diterapkan dalam konteks pengembangan model klasifikasi.

Penelitian ini bertujuan untuk mengembangkan suatu sistem prediksi dengan klasifikasi risiko kolesterol tinggi berdasarkan data kesehatan menggunakan metode Improved *Naive bayes* dengan penerapan teknik *Laplacian correction*. Dengan memanfaatkan dataset kesehatan yang komprehensif, diharapkan penelitian ini dapat memberikan kontribusi pada identifikasi dini individu yang berpotensi mengalami kolesterol tinggi, sehingga intervensi pencegahan dapat dilakukan lebih efektif.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru dalam penanganan risiko kolesterol tinggi dan menjadi landasan untuk pengembangan lebih lanjut dalam bidang prediksi kesehatan menggunakan pendekatan ilmu data. Selain itu, aplikasi praktis dari metode ini dapat memberikan manfaat nyata dalam konteks pelayanan kesehatan masyarakat.

1.2. Rumusan Masalah

Dengan dasar informasi yang telah diuraikan, penelitian ini bertujuan untuk menjawab sejumlah pertanyaan penelitian yang meliputi:

1. Bagaimana cara implementasi metode *Naïve Bayes* biasa tanpa improvisasi dalam memprediksi tingkat kolesterol tinggi berdasarkan data kesehatan dari 200 tanggapan survei BRFSS (*Behavioral Risk Factor Surveillance System*) 2015 yang telah diolah dengan variabel yang relevan yang dimasukkan dalam perhitunga sesuai dengan rekomendasi dari Dr. M Fakhruddin Fakhry, sp. Pd, sebagai berikut: Usia, Jenis Kelamin, Indeks Massa Tubuh, Status Perokok, Tingkat Aktivitas Fisik, Konsumsi Buah-buahan, Konsumsi Sayuran, Indikator Konsumsi Alkohol Berat, Penilaian Umum tentang Kesehatan, Kesehatan Mental, Kesehatan Fisik, dan Status Diabetes responden?
2. Bagaimana cara implementasi *Naïve Bayes* yang sudah diimprove atau ditingkatkan berupa penerapan *Laplace Correction* dalam prediksi kolesterol tinggi berdasarkan data kesehatan?
3. Bagaimana perbandingan hasil evaluasi prediksi kolesterol tinggi berdasarkan data kesehatan antara penggunaan *Naïve Bayes* yang ditingkatkan dengan *Laplace Correction* dan *Naïve Bayes* tanpa peningkatan ?
4. Bagaimana rancang dan bangun sistem prediksi kolesterol tinggi berdasarkan data kesehatan menggunakan *improved naïve bayes* dengan metode *Extreme Programming*, yaitu pendekatan pengembangan perangkat lunak yang menekankan pada fleksibilitas, komunikasi intensif, pengujian berkelanjutan, dan iterasi pendek untuk memastikan sistem yang adaptif dan sesuai dengan kebutuhan pengguna?

1.3. Tujuan Penelitian

Dengan merumuskan masalah tersebut, tujuan yang ingin dicapai adalah:

1. Menerapkan metode *Naïve Bayes* biasa tanpa improvisasi untuk memprediksi risiko kolesterol tinggi berdasarkan data kesehatan dari 200 responden survei BRFSS 2015. Variabel yang digunakan dalam analisis mencakup usia, jenis kelamin, indeks massa tubuh, status merokok, tingkat aktivitas fisik, konsumsi buah-buahan, konsumsi sayuran, indikator konsumsi alkohol berat, penilaian umum tentang kesehatan, kesehatan mental, kesehatan fisik, dan status diabetes responden. Hanya variabel yang relevan yang dimasukkan dalam perhitungan sesuai dengan rekomendasi dari Dr. M Fakhruddin Fakhry, sp. Pd
2. Mengimplementasikan *Naive Bayes* yang sudah diimprove atau ditingkatkan berupa penerapan *Laplace Correction* dalam prediksi kolesterol tinggi berdasarkan data kesehatan.
3. Membandingkan hasil prediksi antara *Naive Bayes* yang sudah diimprove dengan *Naive bayes* tanpa improvisasi.
4. Merancang dan membangun sistem prediksi kolesterol tinggi dengan *Improved Naive bayes* menggunakan metode *Extreme Programming*, yaitu pendekatan pengembangan perangkat lunak yang menekankan pada fleksibilitas, komunikasi intensif, pengujian berkelanjutan, dan iterasi pendek untuk memastikan sistem yang adaptif dan sesuai dengan kebutuhan pengguna.

1.4. Manfaat Penelitian

Penelitian ini memiliki manfaat signifikan dalam pengembangan aplikasi prediksi risiko kolesterol tinggi menggunakan metode *Naive bayes*. Aplikasi ini dirancang untuk memberikan informasi langsung kepada pengguna apakah mereka "berisiko kolesterol" atau "aman dari risiko kolesterol" berdasarkan analisis data kesehatan. Dengan pendekatan ini, pengguna dapat dengan cepat mengetahui status kesehatan mereka terkait risiko kolesterol tinggi tanpa memerlukan pemeriksaan kesehatan yang rumit. Selain memberikan manfaat

individual, aplikasi ini juga memiliki potensi untuk kontribusi pada penelitian lebih lanjut tentang efektivitas metode *Naive bayes* dalam prediksi kesehatan kolesterol.

1.5. Batasan Masalah

Batasan-batasan dalam penelitian ini diantaranya:

- a. Data latih yang dipakai berasal dari 200 tanggapan survei BRFSS (Behavioral Risk Factor Surveillance System) 2015 yang telah melalui proses pembersihan data, dan variabel serta nilainya telah divalidasi sesuai saran Dr. M. Fakhruddin Fakhry, sp. PD. Data tersebut dapat diakses melalui tautan berikut:
<https://www.kaggle.com/datasets/prosperchuks/health-dataset>.
- b. Variabel yang digunakan yakni Age, Sex, BMI, Smoker, PhysActivity, Fruits, Veggies, HvyAlcoholConsump, GenHlth, MentHlth, PhysHlth, dan Diabetes.
- c. Teknik *Laplace Correction* akan diterapkan pada model Naïve Bayes dengan menggunakan dataset yang disediakan oleh Kaggle sebagai sumber data. Dataset tersebut telah dimodifikasi untuk mengukur efek dari hasil improvisasi, termasuk penerapan *Improved Naïve Bayes* dengan Laplace correction. Modifikasi pada dataset mencakup penambahan informasi atau penyesuaian tertentu yang diperlukan untuk menguji efektivitas peningkatan performa algoritme.
- d. Dalam klasifikasi resiko kolesterol tinggi terdiri dari beresiko kolesterol tinggi dan tidak beresiko kolesterol tinggi
- e. Aplikasi yang akan dikembangkan akan menggunakan bahasa pemrograman PHP dan menggunakan database MySQL.

1.6. Metodologi Penelitian

Metode yang diterapkan dalam perancangan sistem mencakup:

1. Tahap Pengumpulan Data

Dalam fase ini, penulis melakukan studi literatur terkait konsep dan teori metode Naïve Bayes dengan fokus pada penerapan bahasa pemrograman PHP dan database MySQL. Selain itu, dilakukan observasi guna mengumpulkan data yang dibutuhkan untuk penelitian ini.

2. Studi Literatur

Dalam langkah ini, penulis berusaha memanfaatkan literatur yang telah tersedia guna memperoleh informasi terkait pembuatan dan struktur sistem. Sumber-sumber literatur yang digunakan mencakup buku, karya ilmiah, dan referensi lain yang relevan dengan permasalahan yang dibahas, sebagaimana diindikasikan oleh (Riani dkk,2019).

3. Analisis Kebutuhan dan Desain Perangkat Lunak

Proses analisis kebutuhan dan perancangan perangkat lunak dilakukan untuk menetapkan kebutuhan pembangunan perangkat lunak dan merancang struktur data serta aktivitas perangkat lunak yang akan dibangun dengan pendekatan berorientasi objek.

4. Implementasi

Tahap implementasi ini mewakili langkah dari fase perancangan ke dalam aplikasi yang konkret, menggunakan bahasa pemrograman PHP dan basis data MySQL. Dalam tahap ini, penulis menerapkan desain yang telah disusun, mengubahnya menjadi kode, dan mengintegrasikan berbagai komponen sistem. Proses implementasi ini menitikberatkan pada kesesuaian antara rencana desain dan aplikasi yang dihasilkan, serta memastikan bahwa aplikasi beroperasi sesuai dengan harapan yang telah ditetapkan.

5. Pengujian Sistem

Tahap ini dilakukan dengan tujuan untuk menguji coba program yang telah dibangun, mengevaluasi kinerja sistem, serta mengukur tingkat keakuratan metode yang digunakan. Hal ini bertujuan agar hasil pengujian dapat memberikan informasi sesuai dengan harapan yang telah ditetapkan..

6. Penyusunan Laporan Akhir

Pada langkah ini, laporan penelitian disusun mengikuti kerangka penulisan yang telah ditetapkan sebelumnya.

1.7. Sistematika Penulisan

Struktur penyusunan laporan ini dijelaskan sebagai berikut:

BAB I: PENDAHULUAN

Pada bagian ini, akan dijabarkan mengenai konteks awal, tujuan penelitian, isu yang dihadapi, batasan permasalahan, metode penelitian, dan struktur penyusunan laporan.

BAB II: LANDASAN TEORI

Dalam bagian ini, akan dibahas teori-teori pendukung yang relevan dalam rangka perancangan, pembuatan, implementasi, dan pengujian sistem..

BAB III: ANALISIS DAN PERANCANGAN SISTEM

Dalam bagian ini, akan dijelaskan tahapan-tahapan dalam proses perancangan dan pembuatan sistem, termasuk kebutuhan sistem (input dan output), flowchart, diagram konteks sistem, serta struktur sistem untuk tahap implementasi..

BAB IV: IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada bagian implementasi sistem, termasuk dalamnya adalah penulisan kode dan antarmuka yang dihasilkan sebagai komponen pendukung sistem. Sementara itu, tahap pengujian akan fokus pada evaluasi validitas dan kesesuaian sistem dengan kebutuhan yang telah ditetapkan..

BAB V: PENUTUP

Bab penutup akan berisikan kesimpulan dan saran sebagai rangkuman dari laporan ini.