

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Matematika sebagai bahasa universal yang mendasari ilmu pengetahuan dan teknologi (Ahmad dan Siregar, 2022). Dari penghitungan sederhana hingga perhitungan yang rumit, matematika membantu kita memahami pola, memecahkan masalah, dan menjelajahi dunia di sekitar kita. Matematika lebih dari sekedar angka dan perhitungan, matematika juga menawarkan konsep abstrak dan struktur logis yang kuat dalam membantu kita berpikir kritis, menganalisis informasi, dan memecahkan masalah secara sistematis.

Menurut Simangunsong tahun (2021) Secara etimologi matematika berasal dari bahasa Latin yaitu “mathematike” atau “mathenein” yang artinya adalah belajar atau mempelajari. Pada awalnya kata tersebut diambil dari bahasa Yunani kuno “mathema” yang memiliki arti ilmu pengetahuan (M. Sari dan Hasanudin, 2023). Aritmatika, aljabar, geometri, kalkulus, statistika, dan probabilitas adalah cabang utama matematika yang mendukung berbagai aspek kehidupan. Masing-masing cabang matematika ini memiliki peran unik, namun saling terkait dan melengkapi satu sama lain.

Statistika merupakan bidang ilmu yang berfokus pada pengembangan dan penerapan metode yang efisien untuk mengumpulkan, menganalisis, dan menginterpretasikan data kuantitatif. Dengan demikian, kesalahan dalam kesimpulan dan perkiraan dapat dihitung melalui penalaran induktif yang didasarkan pada konsep probabilitas matematika (Huwaida, 2019) . Secara etimologi, asal kata *statistic* sendiri adalah ‘status’ yang jika diterjemahkan dalam bahasa Indonesia memiliki arti negara. Ilmu statistika sendiri identik dengan kata probabilitas, sifat probabilistiknya menjadi pembeda antara ilmu statistika dengan induknya yaitu matematika yang cenderung bersifat deterministik (Mardhotillah *et al.*, 2022)

Penyakit jantung atau dikenal juga dengan penyakit kardiovaskular merupakan penyakit yang disebabkan oleh gangguan fungsi jantung. Penyakit jantung disebabkan oleh penumpukan plak di arteri koroner, yang menghambat aliran darah ke jantung dan meningkatkan risiko serangan jantung serta komplikasi lainnya (Puji, 2022). Menurut Kurniadi (2013) penyakit jantung koroner merupakan penyakit yang di akibatkan tersumbatnya pembuluh darah yang menyuplai nutrisi dan oksigen ke otot jantung (D. C. Rahayu *et al.*, 2021). Menurut WHO (2013) menjelaskan kondisi yang menyebabkan penyakit jantung koroner adalah tingginya jumlah plak/lipid/kolesterol yang menumpuk di dalam arteri koroner. Plak yang menumpuk ini dapat mengganggu kemampuan arteri koroner dalam menyuplai oksigen ke otot jantung (Rachmawati *et al.*, 2020).

Menurut Atika *et al* tahun 2021, faktor yang mempengaruhi kejadian penyakit jantung antara lain adalah faktor yang tidak dapat dimodifikasi seperti riwayat keluarga, usia dan jenis kelamin (Tampubolon *et al.*, 2023). Dalam Husni *et al* tahun 2018 menjelaskan faktor lain yang menyebabkan penyakit jantung ada faktor yang dapat dimodifikasi meliputi dislipidemia, aktivitas fisik, merokok, obesitas, diabetes melitus, dan hipertensi (Tampubolon *et al.*, 2023).

Menurut data dari Pusat Pengendalian dan Pencegahan Penyakit Kementerian Kesehatan RI pada tahun 2010, Penyakit Jantung Koroner (PJK) merupakan penyebab kematian keenam terbesar dengan proporsi 4% dari total kematian di Indonesia. Sementara itu, pada tahun 2017, berdasarkan laporan Pusat Data dan Informasi Kementerian Kesehatan RI, PJK telah naik menjadi penyebab kematian kedua setelah stroke, dengan angka mencapai 12,9% (Johanis *et al.*, 2020). Dalam data Riskesdes 2018 yang di rujuk dari jurnal (D. C. Rahayu *et al.*, 2021) data prevalensi penyakit jantung berdasarkan diagnosis dokter di Indonesia menunjukkan angka 1,5%. Provinsi Kalimantan Utara memiliki prevalensi tertinggi sebesar 2,8%, diikuti oleh DIY dengan 2,5% dan Gorontalo dengan 2%. Selain ketiga provinsi tersebut, terdapat 8 provinsi lain yang memiliki prevalensi lebih tinggi dibandingkan angka nasional, yaitu Aceh, Sumatera Barat, DKI Jakarta, Jawa Barat, Jawa Tengah, Kalimantan Timur, Sulawesi Utara, dan Sulawesi Tengah.

Principal Component Analysis (*PCA*) atau sering disebut analisis komponen utama adalah teknik yang mengambil data berdimensi tinggi dan menggunakan ketergantungan antar variabel untuk merepresentasikannya secara lebih sistematis. Hal ini menciptakan dimensi yang lebih rendah tanpa kehilangan banyak informasi dalam kumpulan data (Muningsih et al., 2020). Menurut (Wangge, 2021), Principal Component Analysis (*PCA*) merupakan teknik statistik multivariat yang mengubah sekumpulan variabel asli secara linier menjadi sejumlah variabel baru yang lebih sedikit dan tidak berkorelasi, tetapi tetap mampu mewakili informasi dari variabel-variabel asli tersebut. Metode *PCA* dapat digunakan untuk menyederhanakan data dengan menghilangkan faktor atau indikator yang kurang signifikan atau relevan, tanpa mengurangi esensi serta tujuan dari data asli yang dianalisis (Nasution et al., 2019). *PCA* menghasilkan performa yang optimal ketika diterapkan pada atribut yang saling berkorelasi. Pada penelitian ini, *PCA* digunakan untuk melatih dan menguji karakteristik dari kumpulan data faktor yang menyebabkan penyakit jantung.

K-Nearest Neighbor (*KNN*) adalah algoritma pengelompokan yang sederhana namun efektif dalam memproses data. Algoritma ini bekerja dengan mengelompokkan data berdasarkan pola dan kemiripan dengan data pelatihan yang sudah ada. Dalam penerapannya pada pengelompokan penyakit jantung, *KNN* dapat digunakan untuk melakukan pengelompokan pasien berdasarkan faktor risiko yang berkaitan dengan penyakit jantung (Sularno et al., 2023). Menurut Yani et al., 2020 dalam jurnal (Nuraeni et al., 2023) *KNN* merupakan metode yang termasuk dalam pengaplikasian data yang sederhana dan sangat mudah dalam pengimplementasiannya, efektif pada data yang lebih besar, serta dapat mengelompokkan data dengan cepat. Menurut (Argina, 2020) *KNN* memiliki kemampuan yang baik terhadap data *training* yang *noise*, dan dapat menghasilkan data yang lebih akurat. Algoritma ini dapat mengelompokkan objek berdasarkan data testing yang jaraknya paling dekat dengan objek tersebut. *KNN* termasuk dalam kelompok *instance based learning*, algoritma ini juga sering disebut *lazy learning*. *KNN* dilakukan dengan mencari kelompok *k* objek pada data latih yang terdekat (serupa) dengan objek pada data baru atau data uji, setelah tetangga terdekat

ditemukan, metode *KNN* melakukan prediksi dengan melihat mayoritas kelas atau kelompok dari tetangga tersebut. Jika mayoritas tetangga masuk ke dalam satu kategori tertentu, maka data baru diprediksi akan masuk ke kategori yang sama.

Pada Penelitian (Dinanti dan Purwadi, 2023), yang membahas mengenai “*Analisis Performa Algoritma K-Nearest Neighbor dan Reduksi Dimensi Menggunakan Principal Component Analysis*”, pada penelitian tersebut menunjukkan bahwa reduksi dimensi menggunakan *PCA* menghasilkan tiga komponen utama dari delapan variabel pada data asli, yaitu *PC1*, *PC2*, dan *PC3*. Hasil klasifikasi menggunakan algoritma *KNN* menunjukkan hasil jika menggunakan tiga parameter tetangga terdekat (*K*): $K = 3$, $K = 5$, dan $K = 7$. Untuk $K = 3$ diperoleh hasil akurasi sebesar 67,53%, untuk $K = 5$ sebesar 72,72%, dan untuk $K = 7$ diperoleh hasil akurasi sebesar 77,92%.

Dalam penelitian yang dilakukan oleh (Akbar et al., 2022) mengenai penerapan metode “*Principal Component Analysis (PCA) dan K-Nearest Neighbors (KNN) untuk klasifikasi data kanker paru-paru*”, dataset digunakan untuk mengklasifikasikan sampel kanker dan non-kanker dengan menggunakan metode ekstraksi fitur *PCA* serta klasifikasi *KNN* dengan nilai $K = 1, 3, 5, 7$, dan 9. Penelitian ini menghasilkan akurasi terbaik sebesar 98% pada $K = 9$. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode *PCA* dan *KNN* sangat cocok untuk pengolahan dataset kanker paru-paru.

Mengingat tingginya prevalensi penyakit jantung di Indonesia dan kurangnya pemahaman masyarakat mengenai faktor-faktor penyebabnya, peneliti tertarik untuk melakukan penelitian yang berfokus pada analisis faktor penyebab penyakit jantung. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap penyakit jantung serta memprediksi pasien yang memiliki riwayat penyakit jantung, dengan memperhatikan berbagai faktor penyebab. Penyakit jantung merupakan salah satu penyebab kematian dan ancaman serius bagi kesehatan masyarakat, tidak hanya mengakibatkan jumlah kematian yang signifikan tetapi juga menimbulkan biaya pengobatan yang tinggi, sering kali tanpa diagnosis pada tahap awal. Fokus penelitian ini adalah pada pasien rawat inap di Rumah Sakit Umum Haji. Melalui penelitian ini, diharapkan dapat membantu

dalam memberikan intervensi medis yang lebih cepat dan tepat, sehingga meningkatkan prognosis pasien dan mengurangi komplikasi yang lebih serius. Pada penelitian ini, metode Principal Component Analysis (PCA) dan K-Nearest Neighbors (KNN) akan diterapkan untuk memprediksi dan menganalisis penyakit jantung berdasarkan faktor-faktor penyebabnya. PCA dipilih karena kemampuannya dalam mereduksi dimensi data, sehingga memungkinkan identifikasi faktor risiko utama tanpa menghilangkan informasi penting. Di sisi lain, KNN digunakan untuk memprediksi pasien berdasarkan faktor-faktor yang terkait dengan penyakit jantung, sehingga memungkinkan prediksi dengan tingkat akurasi tinggi bagi pasien yang memiliki riwayat penyakit jantung. KNN juga terbukti handal dalam menangani data yang mengandung noise dan efektif untuk dataset besar (Akbar *et al.*, 2022). Dengan menggabungkan kedua metode ini, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai faktor-faktor penyebab penyakit jantung. Berdasarkan latar belakang tersebut, penulis berminat untuk melakukan penelitian berjudul ***“Pengelompokan Pasien Yang Memiliki Faktor Penyebab Penyakit Jantung Dengan Menggunakan Metode Principal Component Analysis (PCA) Dan K Nearest Neighbors (KNN) ”***

1.2. Identifikasi Masalah

Berdasarkan latar belakang permasalahan yang ada maka di dapat identifikasi masalah sebagai berikut :

1. Penyakit jantung merupakan salah satu penyebab kematian utama di Indonesia, dengan prevalensi yang tinggi di berbagai provinsi.
2. Faktor yang mempengaruhi kejadian penyakit jantung antara lain adalah faktor yang tidak dapat dimodifikasi dan yang dapat dimodifikasi.
3. Metode *PCA* dapat digunakan untuk menyederhanakan data dengan mengeliminasi faktor atau indikator yang kurang signifikan atau relevan, tanpa mengubah makna dan tujuan dari data aslinya.
4. Algoritma *KNN* dapat digunakan untuk mengklasifikasikan pasien berdasarkan gejala, riwayat kesehatan, dan faktor risiko terkait penyakit jantung

1.3. Ruang Lingkup

Ruang lingkup penelitian skripsi dengan judul "Pengelompokan Pasien yang Memiliki Faktor Penyebab Penyakit Jantung dengan Menggunakan Metode Principal Component Analysis (*PCA*) dan K Nearest Neighbors (*KNN*)" mencakup beberapa aspek penting. Penelitian ini berfokus pada pengelompokan pasien yang memiliki faktor penyebab penyakit jantung menggunakan metode *PCA* untuk reduksi dimensi dan *KNN* untuk pengelompokan. Data yang akan dianalisis diperoleh dari rumah sakit RSUD Umum Haji dengan data yang di pakai dari Januari 2021 sampai dengan Desember 2023. Objek penelitian ini adalah data pasien yang di rawat inap di RSUD Umum Haji Medan.

Variabel yang akan diteliti meliputi Nama, Usia, Jenis Kelamin, Kolestrol, Hipertensi, Kadar Gula Darah, Heart Rate dan Hemoglobin. Data akan dikumpulkan dari rekam medis elektronik atau database rumah sakit. Penelitian ini menggunakan konsep-konsep dari ilmu statistik, data mining, dan machine learning, khususnya yang berkaitan dengan *PCA* dan *KNN*. Selain itu, teori-teori yang relevan tentang faktor risiko penyakit jantung juga akan digunakan sebagai kerangka acuan.

Analisis akan difokuskan pada pengelompokan pasien berdasarkan faktor risiko dan mencari pola-pola signifikan dalam data, yang diharapkan dapat memberikan wawasan baru tentang faktor-faktor yang paling berpengaruh terhadap risiko penyakit jantung. Penelitian ini dibatasi pada pasien yang berusia lebih dari 20 tahun. Selain itu, hasil pengelompokan hanya berlaku untuk populasi yang diteliti dan mungkin tidak dapat digeneralisasikan untuk populasi yang lebih luas. Dengan ruang lingkup yang jelas ini, penelitian dapat dilakukan secara terarah dan fokus pada tujuan yang ingin dicapai.

1.4. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah

1. Bagaimana metode *PCA* dapat mereduksi dimensi data faktor-faktor penyebab penyakit jantung?
2. Apa saja faktor-faktor penyebab penyakit jantung yang paling signifikan?
3. Bagaimana metode *KNN* dapat memprediksi pasien penyakit jantung berdasarkan faktor-faktor penyebabnya?

1.5. Batasan Masalah

Dalam penelitian ini, cakupan masalah yang dibatasi adalah sebagai berikut:

1. Penelitian ini dilakukan untuk memprediksi pasien penyakit jantung dengan menggunakan faktor penyebab penyakit jantung yang sudah di reduksi dengan menggunakan metode *PCA*.
2. Untuk data yang digunakan, dilakukan studi kasus di RSUD HAJI MEDAN dengan melihat data penyakit pasien secara umum.
3. Data yang digunakan data 2021 Bulan Januari sampai dengan Bulan Desember 2023
4. Data yang di ambil ialah data pasien yang umurnya lebih atau sama dengan 20 tahun.

1.6. Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dalam penelitian ini adalah

1. Mereduksi faktor-faktor penyebab penyakit jantung dengan menggunakan metode *PCA*
2. Menganalisis faktor-faktor penyebab penyakit jantung yang paling signifikan
3. Memprediksi pasien yang mengalami penyakit jantung dengan menggunakan metode *KNN*

1.7. Manfaat Penelitian

1. Bagi Peneliti

Sebagai Sarana Untuk Meningkatkan Pemahaman Tentang Metode *PCA* Dan *KNN* dalam Mengelompokkan pasien yang memiliki faktor penyebab penyakit jantung.

2. Bagi Pembaca

Peneliti atau akademisi yang tertarik dengan topik penerapan Metode *PCA* Dan *KNN* dalam mereduksi faktor faktor penyebab penyakit jantung serta memprediksi individu yang memiliki faktor penyebab penyakit jantung dapat menggunakan temuan dan metodologi dari penelitian ini sebagai dasar untuk penelitian lebih lanjut atau eksplorasi topik terkait.

3. Bagi Masyarakat Umum

Sebagai sarana untuk mengedukasi untuk menjaga kesehatan tubuh terkhusus pada jantung serta untuk memberikan pengetahuan untuk menjaga kesehatan jantung dengan memahami faktor resiko yang di bahas pada penelitian ini.

