

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kanker diartikan sebagai pertumbuhan sel abnormal yang menyebar tidak terkendali dan berpotensi membahayakan komponen tubuh lainnya (Zafrial dan Amalia 2018). Salah satu kanker yang banyak terjadi adalah kanker payudara. Kanker payudara adalah kanker ganas pada payudara yang berasal dari sel kelenjar, saluran kelenjar, serta jaringan penunjang payudara, namun tidak termasuk kulit payudara. Usia di atas 50 tahun, riwayat keluarga (genetika), penggunaan *alcohol* dan tembakau, obesitas (kelebihan berat badan), penggunaan kontrasepsi hormonal jangka panjang, paparan radiasi, tidak pernah melahirkan pada usia di atas 35 tahun, tidak pernah menyusui, menstruasi pertama pada usia di bawah 12 tahun, dan menopause terlambat (di atas 50 tahun) adalah beberapa faktor yang dapat meningkatkan risiko kanker payudara (Direktorat Pengendalian Penyakit Tidak Menular 2014).

Penyakit kanker yang paling banyak terjadi di Indonesia menurut data *International Agency for Research on Cancer in Globocan (Globocan cancer Statistics)* tahun 2020 adalah kanker payudara. Dari 396.914 kasus dengan 22.430 (9,6%) kasus kematian akibat penyakit ini, terdapat sekitar 65.858 kasus baru (16,6%) kanker payudara yang terdeteksi secara nasional yang mana Sumatera Utara masuk menjadi 10 besar penyumbang tertinggi dari seluruh provinsi di Indonesia. Peningkatan angka terjadinya kanker payudara selalu meningkat setiap tahunnya di Sumatera Utara khususnya kota Medan. Pada tahun 2021, penderita kanker payudara di kota Medan mencapai 824 kasus (<https://sumutprov.go.id>). Pada tahun 2022 terjadi kurang lebih 100 kasus penyakit kanker payudara di Rumah Sakit Umum Provinsi (RSUP) Haji Medan.

Hal yang diperlukan saat menganalisis data di atas adalah suatu metode yang mampu menganalisis beberapa variabel dan juga dapat mengukur hubungan antar variabel. Metode statistik yang mampu dalam menganalisis beberapa variabel adalah statistic multivariat. Ciri khas tertentu dari analisis multivariat umumnya

menggunakan variabel yang banyak untuk variabel independen dan dependen (Herman 2012). Dalam statistik multivariat, setiap metode memiliki manfaatnya masing-masing. Analisis faktor adalah teknik statistic multivariate yang umum digunakan (Wiratmanto 2014).

Menurut Susetyo dalam (Sangila dan Jufri 2018) analisis statistika berdasarkan variabel terbagi menjadi analisis statistika univariat, multivariat, dan bivariat. Analisis statistik multivariat adalah suatu metode yang biasanya dilakukan untuk meneliti beberapa variabel secara bersamaan. Menganalisis pengaruh variabel yang satu dengan variabel yang lainnya secara simultan dapat menggunakan metode ini (Rumlawang dan Toamain 2007) Pada analisis multivariat terdapat 12 jenis teknik, termasuk *Principal Component Analysis* (PCA) dan analisis faktor. Sebuah komponen analisis statistik multivariat yaitu analisis faktor berupaya mengidentifikasi hubungan antara variabel-variabel yang sebelumnya tidak bergantung satu sama lain untuk menciptakan satu variabel atau sekelompok variabel kecil yang terkait (Simarmata dkk. 2015). Analisis faktor merupakan metode untuk mengkaji dan memahami suatu fenomena (Rumlawang dan Toamain 2007).

Menurut (Filik Iscen dkk. 2008) salah satu teknik statistik multivariat yang disebut analisis komponen utama mengubah sekumpulan variabel asli yang lebih besar menjadi sekumpulan variabel kecil yang tidak berhubungan dan tetap dapat mencerminkan informasi dari kumpulan variabel asli dalam garis lurus. *Principal component analisis* adalah statistik multivariat yang dapat digunakan untuk menggambarkan bagaimana sekumpulan data (parameter) yang tidak berkorelasi dapat bervariasi menjadi beberapa parameter independen (komponen utama) (Nyoman Radiarta dkk. 2012).

Analisis faktor menggunakan *principal component* adalah jenis analisis data yang dapat digunakan untuk mengekstraksi indikator penilaian guna menciptakan faktor-faktor baru yang tidak berkorelasi satu sama lain dan memperkuat suatu kategori penilaian. Derajat variansi data pada seluruh indikator diperhitungkan dalam komponen utama. (Wangge 2021).

Metode analisis *Principal Component Analysis* (PCA) digunakan oleh (Nasution dkk. 2019) untuk menyederhanakan dan menghilangkan sejumlah variabel skrining pada data kanker serviks yang kurang penting tanpa mengurangi maksud dan tujuan data aslinya. Sembilan variabel digunakan pada penelitian

ini. Namun, hanya tiga faktor yang paling signifikan (berpengaruh), yaitu faktor usia yang memiliki proporsi varians sebesar 58,50%, faktor kecanduan rokok yang memiliki varians 14,90% dan faktor penggunaan kontrasepsi dengan media suntik (*Hormonal Contraceptives*) yang memiliki varians 9,3%. Sembilan variabel skrining tersebut menghasilkan total varians sebesar 99%.

Penelitian sebelumnya yang menggunakan pemilihan (seleksi) fitur dengan metode PCA (*Principal Component Analysis*) adalah penelitian yang dilakukan oleh (Fadhila Salma dkk. 2021) yang mana maksud penelitian ini untuk memahami perbedaan dari pola agregasi eritrosit antara subjek normal dengan subjek penderita kanker payudara dengan menggunakan metode PCA dan juga untuk mengetahui proses terjadinya agregasi eritrosit. Dalam penelitian ini, model grafik klusterisasi 2 dimensi dan 3 dimensi dijadikan sebagai parameter pembeda. Penelitian ini tidak mendapatkan hasil yang cukup karena hanya dengan satu metode, penelitian ini tidak dapat membedakan pola yang diteliti.

Penelitian sebelumnya yang dilakukan oleh (Fauzi dkk. 2020) berjudul “Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis *Principal Component Analysis* dan *Random Forest*” bertujuan mencapai tingkat akurasi yang tinggi dalam deteksi kanker payudara. Penelitian ini menggunakan model klasifikasi yaitu *Principal Component Analysis* berbasis *Random Forest* untuk memprediksi permasalahan yang diteliti yang mana hasil evaluasi model tersebut akan dilihat nilai akurasinya. Temuan penelitian ini menunjukkan bahwa ketika pemilihan fitur *Random Forest* dan *logitboost* digabungkan, metode pemilihan fitur berbasis analisis komponen utama secara signifikan meningkatkan kinerja klasifikasinya. Model yang paling akurat adalah *Random Forest* yang memiliki akurasi 79.3103% dengan nilai AUC sebesar 0,843.

Metode PCA memiliki tujuan dalam untuk mengatasi masalah multikolinearitas. Metode ini efektif dalam mengatasi masalah multikolinearitas dan menghilangkan korelasi di antara variabel-variabel bebas sampai tidak berkorelasi. Kelebihan metode ini adalah dapat menghilangkan korelasi tanpa mengurangi atau menghilangkan variabel-variabel aslinya (Kusuma dan Wibowo 2017).

Oleh karena itu, peneliti tertarik untuk menggunakan metode PCA (*Principal Component Analysis*) untuk menganalisis beberapa faktor yang memengaruhi kanker payudara di RSUP Haji Medan dengan mempertimbangkan faktor-faktor

yang berpotensi menyebabkan Kanker Payudara. Maka peneliti akan melakukan penelitian dengan mengangkat judul “**Analisis Faktor yang Memengaruhi Kanker Payudara di RSUP Haji Medan Menggunakan Metode *Principal Component Analysis***”.

## **1.2 Identifikasi Masalah**

Setelah melakukan observasi di RSUP Haji Medan, terdapat beberapa masalah yang terjadi, yaitu:

1. Cenderung meningkatnya kasus penyakit Kanker terutama Kanker Payudara di RSUP Haji Medan Tahun 2022.
2. Belum dilakukan penelitian terkait aktor-faktor yang paling memengaruhi penyebab kanker payudara di RSUP Haji Medan Tahun 2022.

## **1.3 Ruang Lingkup**

Penelitian yang akan dilakukan ini menggunakan data sekunder yang dikumpulkan dari Rumah Sakit Umum Provinsi (RSUP) Haji Medan yang akan diolah dengan menggunakan metode *Principal Component Analysis* (PCA). Peneliti mengolah data dengan menggunakan *software Python*.

## **1.4 Batasan Masalah**

Batasan masalah sangat penting untuk memungkinkan penelitian lebih focus dan mendalam. Berikut batasan masalah penelitian:

1. Beberapa variabel yang digunakan dalam penelitian ini adalah usia, jenis kelamin, genetika (riwayat keluarga), kebiasaan merokok, status perkawinan, pendidikan terakhir, dan pekerjaan.
2. Data yang digunakan dan diolah pada penelitian ini adalah data pasien kanker payudara RSUP Haji Medan tahun 2022.

## **1.5 Rumusan Masalah**

Faktor apa yang paling besar pengaruhnya terhadap pasien kanker payudara di RSUP Haji Medan dengan menggunakan metode *Principal Component Analysis* (PCA)?

## 1.6 Tujuan Penelitian

Berdasarkan pada rumusan masalah diatas, tujuan dari penelitian ini adalah untuk mengetahui faktor yang dominan terhadap Penyakit Jantung Koroner (PJK) di RSUP H. Adam Malik dengan menggunakan metode *Principal Component Analysis* (PCA).

## 1.7 Tujuan Penelitian

Tujuan penelitian ini adalah mengetahui faktor yang paling besar pengaruhnya terhadap pasien kanker payudara di RSUP Haji Medan dengan menggunakan metode *Principal Component Analisis* (PCA).

## 1.8 Manfaat Penelitian

1. Mahasiswa mampu memahami metode *Principal Component Analisis* (PCA) agar mendapatkan hasil yang diharapkan pada penelitian ini. Mahasiswa mampu memahami menggunakan aplikasi (*software*) yang mendukung dalam mengerjakan penelitian ini.
2. Diharapkan agar penelitian ini dapat memberikan informasi tambahan kepada tim RSUP Haji Medan dalam mendeteksi (memprediksi) kanker payudara. Tidak hanya untuk RSUP Haji Medan saja, tetapi penelitian ini juga diharapkan dapat memberikan informasi kepada masyarakat luas dalam mendeteksi (memprediksi) kanker payudara agar penyakit kanker payudara dapat lebih cepat ditangani oleh pihak medis.