

BAB I

PENDAHULUAN

1.1. Latar Belakang

Di dunia angka kematian pada wanita semakin meningkat yang diakibatkan oleh kanker payudara, kanker payudara disebut sebagai penyakit yang sangat ganas. Namun, kanker payudara bukanlah penyakit yang menular. Dengan demikian, kanker payudara memerlukan penanganan secara khusus, sehingga dalam hal tersebut dapat dilakukan pencegahan kanker payudara secara dini.

Jika tidak dilakukan pencegahan kanker payudara secara dini atau bahkan tidak mendapatkan penanganan yang lebih optimal, maka kasus penyakit kanker payudara dan tingkat kematian akan terus meningkat. Pada lingkungan medis pemeriksaan atau deteksi kanker payudara menggunakan mamografi, mamografi adalah metode skrining yang mampu mengidentifikasi kanker berdasarkan variable gejala-gejala fisik dari penyakit yang muncul. Namun, hasil mamografi dianggap kurang memuaskan, dikarenakan deteksi ini hanya mampu memberikan gambaran area *abnormal* pada jaringan payudara. Sehingga, tenaga medis perlu melakukan skrining tambahan untuk mendapatkan hasil lanjutan (Susilowati et al., 2019). Oleh karena itu, untuk mengidentifikasi diagnosa kanker perlu ditambahkan perangkat tambahan yang mampu membantu mengidentifikasi gejala kanker payudara, apakah hasil dari mamografi tersebut pasien tergolong kanker ganas atau kanker jinak.

Clustering sangat penting dalam pengelompokan data, karena membantu mengelompokkan data berdasarkan obyek yang ada pada data. *Clustering* dengan menggunakan *K-Means* sudah sering digunakan dalam pengelompokan data (Sani, 2020). Pada penelitian mengenai perbandingan algoritma *K-Means* dan *Fuzzy C-Means* untuk *clustering* puskesmas berdasarkan gizi balita di Surabaya mendapatkan hasil bahwa nilai normalisasi *K-Means* sebesar 0.518 sedangkan nilai normalisasi fuzzy c-means 0.496, maka dapat disimpulkan bahwasannya metode *K-Means* lebih optimal daripada metode *Fuzzy C-Means* (Kurnia, 2023). Pada penelitian mengenai perbandingan algoritma k-means dan fuzzy c-means dalam pengelompokan kabupaten atau kota di Kalimantan berdasarkan indikator IPM tahun 2019, mendapatkan hasil bahwa metode *Fuzzy c-means* sangat cocok untuk digunakan pada penelitian tersebut

dan metode *Fuzzy C-Means* merupakan metode yang lebih optimal daripada *K-Means* (Goejantoro & Deny Tisna Amijaya, 2021). Penelitian yang dilakukan oleh Manik et al. (2023) mengenai perbandingan algoritma *Fuzzy C-Means* dan *K-Means clustering* data penggunaan obat di RS *national hospital* Surabaya yang mendapatkan hasil bahwa *K-Means* membutuhkan 2 iterasi saja untuk menentukan banyak, sedang dan sedikitnya penggunaan obat sedangkan *Fuzzy C-Means* membutuhkan 6 kali iterasi untuk menentukan banyak, sedang dan sedikitnya penggunaan obat, maka dapat disimpulkan bahwa *K-Means* merupakan metode yang lebih optimal. Penelitian selanjutnya, dilakukan oleh Jajang et al. (2021) mengenai *Fuzzy C-Means* dan *K-Means* dalam pengelompokan kecamatan di Banyumas, mendapatkan hasil bahwa keduanya mendapatkan hasil *cluster* yang sama, namun rasio variansinya berbeda *K-means* dengan skenario 1 sebesar 69% dan skenario 2 sebesar 85%. Pada penelitian yang berjudul perbandingan metode *Fuzzy C-means* dan *K-Means* untuk pemetaan daerah rawan kriminalitas di kota Semarang, mendapatkan hasil bahwa metode *Fuzzy C-Means* lebih baik dibandingkan metode *K-Means*, dengan presentase hasil *K-Means* sebesar 0.569 dan *Fuzzy C-Means* sebesar 71.23. Penelitian mengenai deteksi penyakit diabetes dengan metode *Fuzzy C-Means Clustering* dan *K-Means Clustering* mendapatkan hasil bahwa metode *Fuzzy C-Means clustering* merupakan metode terbaik (A. Praja et al., 2021).

Berdasarkan latar belakang tersebut, penelitian ini akan menerapkan perbandingan metode untuk mencari metode mana yang lebih optimal dalam melakukan *clustering* data kanker payudara. Tujuan dari penelitian ini yaitu untuk menerapkan metode *K-Means* dan *Fuzzy C-Means* untuk mengetahui hasil akurasi metode mana yang lebih optimal, hasil penelitian nantinya dapat dijadikan referensi untuk penelitian selanjutnya. Pada penelitian sebelumnya yang dilakukan oleh (Susilowati et al., 2019) mengenai diagnosa kanker payudara dengan *Clustering*, hasil pengelompokan data kanker payudara, berdasarkan penelitian tersebut mendapatkan hasil bahwa yang tergolong kelompok ganas terdapat 11 data dan kanker payudara jinak terdapat 30 data, Saran yang disampaikan oleh peneliti terdahulu kepada peneliti selanjutnya untuk mengembangkan penelitian mengenai diagnosa kanker payudara dengan menambahkan metode pemecah lainnya yang lebih optimal dan memperbanyak sampel datanya.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, maka permasalahan yang dapat diambil adalah “Bagaimana cara menerapkan perbandingan metode *Fuzzy C-Means* dan *K-Means* untuk pengelompokan diagnosa jenis penyakit kanker payudara?”

1.3. Batasan Masalah

Batasan masalah untuk penelitian ini yaitu;

1. Data yang digunakan dalam penelitian ini diperoleh dari *dataset UCI Machine Learning* pada *google dataset*, dengan total jumlah data 699 data dan 10 variabel gejala.
2. Tidak membahas dan menjelaskan secara rinci mengenai tampilan, bahasa pemrograman, kepraktisan dan keamanan dari aplikasi yang dibuat.
3. Aplikasi ini menampilkan hasil cluster dan akurasi dari software R mengenai pengelompokkan data kanker payudara berdasarkan gejalanya.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan diatas, maka tujuan dari penelitian ini yaitu “Mengetahui perbandingan metode *Fuzzy C-Means* dan *K-Means* untuk pengelompokkan diagnosa jenis penyakit kanker payudara”

1.5. Manfaat Penelitian

1. Manfaat praktis, sebagai pembantu pengambil keputusan dalam menangani pasien, apakah pasien tersebut termasuk golongan pasien dengan kanker ganas, pra kanker ataupun kanker jinak.
2. Manfaat teoritis, sebagai informasi. Pengelompokkan kanker payudara menggunakan algoritma k-means dan fcm lebih optimal mana.