

# 1. Pendahuluan

## 1.1 Latar Belakang

Dalam teori *fuzzy* dikenal adanya istilah *clustering* atau pengelompokan. [9] *Clustering* merupakan suatu proses pengelompokan data ke dalam kelas-kelas atau *cluster-cluster* sehingga data dalam suatu kelas tersebut memiliki tingkat persamaan yang tinggi antara satu dengan yang lainnya akan tetapi berbeda dengan data dalam kelas yang lainnya. [7] *Fuzzy clustering* adalah salah satu teknik untuk menentukan *cluster* optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar vektor.

Salah satu kasus yang berkaitan dengan pengelompokan data yaitu pengelompokan suatu tumor yang diderita pasien ke dalam dua kelas yaitu tumor jinak (*benign*) atau tumor ganas (*malignant*). Identifikasi seseorang terkena kanker payudara bisa diketahui dari jenis tumor payudara yang dideritanya. [2] Tumor merupakan kumpulan dari sel abnormal yang terbentuk dari sel-sel yang tumbuh secara terus menerus secara tidak terbatas. Pertumbuhan tumor tersebut dapat mengarah ke pertumbuhan jinak (*benign*) atau ke pertumbuhan ganas (*malignant*). Tumor pada payudara yang mengalami pertumbuhan ganas itu lah yang merupakan kanker payudara. Pada penelitian ini digunakan beberapa baris data yang terdiri dari beberapa atribut yang salah satu atributnya merupakan suatu kelas yang terdiri dari dua jenis kelas yaitu *benign* dan *malignant*. Atribut-atribut tersebut terdiri dari masukan *real-valued features* yang dihitung dari setiap sel nukleus dari setiap pasien yang menderita tumor pada payudaranya. Berdasarkan masukan *real-valued features* tersebut dapat diprediksi apakah tumor yang diderita pasien tersebut tergolong jinak (*benign*) atau ganas (*malignant* atau kanker).

Terdapat beberapa algoritma *clustering* data salah satunya yaitu *fuzzy c-means* (FCM). Pendekatan *clustering* menghasilkan partisi dimana dalam partisi tersebut setiap anggota hanya dimiliki oleh satu *cluster*. *Fuzzy clustering* mempunyai sebuah skema *soft clustering* yang mana setiap pola berasosiasi dengan seluruh *cluster* dengan menggunakan fungsi keanggotaan. Nilai keanggotaan yang besar mempunyai kemungkinan yang besar pula suatu pola menjadi bagian dari suatu *cluster*. [6] Algoritma *fuzzy c-means* merupakan salah satu algoritma *fuzzy clustering* yang paling banyak digunakan karena berusaha mengelompokkan setiap elemen-elemen yang merupakan anggota dari suatu *cluster* ke dalam suatu *cluster*. Algoritma *fuzzy c-means* dapat digunakan untuk mengelompokkan data yang sudah diketahui jumlah *cluster*-nya. Algoritma *fuzzy c-means* merupakan algoritma yang

sederhana, mudah diimplementasikan, dan memiliki kemampuan untuk mengelompokkan data dalam jumlah besar.

## 1.2 Perumusan Masalah

Rumusan masalah yang terdapat pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *fuzzy c-means* untuk memprediksi resiko terkena kanker payudara berdasarkan pengelompokan jenis tumor yang dihasilkan?
2. Bagaimana proses yang dilakukan untuk mendapatkan keluaran berupa status resiko terkena kanker payudara dengan menggunakan metode *fuzzy c-means*?
3. Bagaimana menganalisis tingkat akurasi yang dihasilkan oleh metode *fuzzy c-means* dalam memprediksi seseorang terkena kanker payudara atau tidak berdasarkan beberapa *real-valued features* yang diberikan?

Adapun batasan masalah dalam tugas akhir ini yaitu :

1. Menggunakan data *breast cancer wisconsin (diagnostic) data set* yang diperoleh dari *UCI repository*.
2. *Preprocessing* data dilakukan di luar sistem program.

## 1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dari tugas akhir ini adalah:

1. Mengimplementasikan metode *fuzzy c-means* untuk memprediksi resiko terkena kanker payudara berdasarkan karakteristik tumor yang dihasilkan.
2. Melakukan proses untuk mendapatkan keluaran berupa status resiko terkena kanker payudara dengan menggunakan metode *fuzzy c-means*.
3. Menganalisis tingkat akurasi yang dihasilkan oleh metode *fuzzy c-means* dalam memprediksi seseorang terkena kanker payudara atau tidak berdasarkan beberapa *real-valued features* yang diberikan.

## 1.4 Metodologi Penyelesaian Masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah :

1. Melakukan studi Literatur
  - a. mencari referensi dan sumber-sumber lain yang berhubungan dengan tugas akhir ini seperti *fuzzy logic*, *fuzzy clustering*, dan *fuzzy c-means*.
  - b. mempelajari dan memahami materi yang berhubungan dengan tugas akhir ini.
2. Mempelajari algoritma *fuzzy c-means* yang akan digunakan dalam penelitian ini.

3. Memahami dan mempelajari data *breast cancer wisconsin (diagnostic) data set* yang diperoleh dari *UCI repository*.
4. Melakukan perancangan dan pembuatan sistem menggunakan metode *fuzzy c-means* untuk memperoleh keluaran berupa prediksi seseorang terkena kanker payudara atau tidak.
5. Melakukan pengujian akurasi terhadap metoda *fuzzy c-means* sesuai dengan kasus yang sedang dikerjakan dalam tugas akhir ini.
6. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.