

Pengklusteran Data Iris Menggunakan Metode Fuzzy C-Mean

Gde Krishna Sankya Yogeswara^{a1}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana, Bali
Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 08261, Bali, Indonesia
¹krishnasankya@gmail.com

Abstract

This study focuses on the application of the Fuzzy C-Means method for clustering the Iris dataset. Clustering is a widely used technique for grouping similar data objects together, and the Iris dataset, which consists of measurements of iris flowers, has been a popular choice for clustering analysis. The Fuzzy C-Means algorithm, based on fuzzy logic, allows for a more flexible and nuanced approach to clustering by assigning degrees of membership to data points, capturing the inherent uncertainty and ambiguity in the dataset. By utilizing fuzzy logic, the Fuzzy C-Means method aims to accurately classify iris flowers into distinct clusters based on their petal width, petal length, sepal width, and sepal length. The results of this study contribute to the understanding of fuzzy clustering techniques and their application in pattern recognition and data analysis.

Keywords: Iris dataset, clustering, Fuzzy C-Means, fuzzy logic.

1. Pendahuluan

Indonesia memiliki kekayaan sumber daya alam yang melimpah, sehingga negara ini memiliki beragam tumbuhan dan bunga yang tersebar di seluruh wilayahnya. Meskipun demikian, hanya sekitar 20% dari total tumbuhan di Indonesia yang telah diidentifikasi [1]. Secara umum, tanaman yang belum diidentifikasi sering diklasifikasikan atau dikelompokkan ke dalam beberapa kelompok. Pengklusteran atau pengelompokan adalah proses mengelompokkan objek atau kasus ke dalam kelompok-kelompok yang lebih kecil, di mana setiap kelompok berisi objek atau kasus yang memiliki kesamaan satu sama lain [2]. Data iris sering digunakan untuk melakukan pengklusteran berbagai jenis bunga berdasarkan lebar mahkota (petal width), panjang mahkota (petal length), lebar kelopak (sepal width), dan panjang kelopak (sepal length).

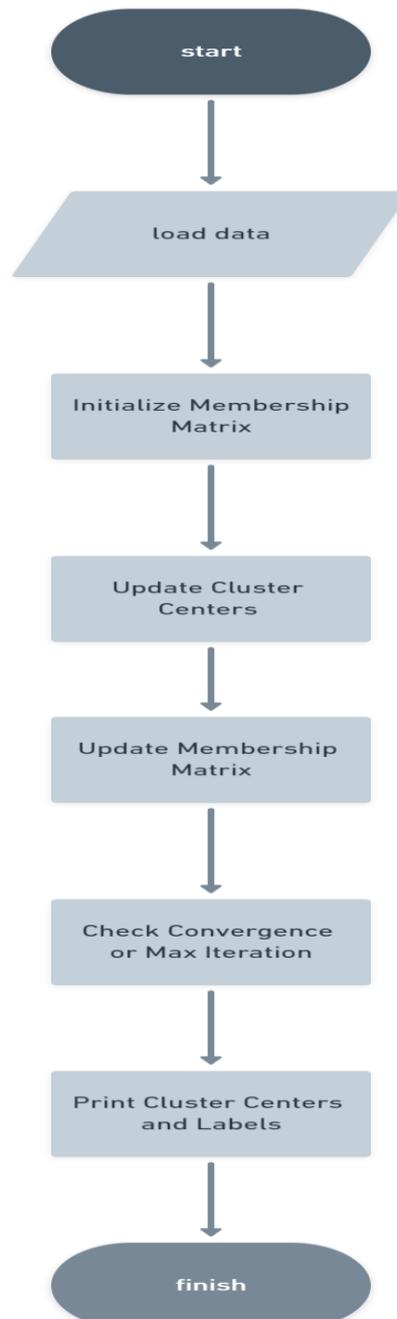
Data iris adalah kumpulan data yang terdiri dari bunga yang diidentifikasi berdasarkan ukuran panjang mahkota, lebar mahkota, panjang kelopak, dan lebar kelopaknya [3]. Sebelumnya, banyak peneliti telah menggunakan data iris sebagai sarana untuk menguji metode pengklusteran. Data iris dipilih karena sederhana dan mudah diperoleh. Terdapat beberapa metode yang digunakan untuk mengelompokkan data menjadi beberapa kelompok, salah satunya adalah dengan memanfaatkan cabang ilmu matematika seperti data mining dan logika fuzzy.

Logika fuzzy merupakan salah satu cabang ilmu matematika yang fokus pada konsep logika yang kabur atau tidak tegas. Logika fuzzy menggunakan rentang keanggotaan antara 0 dan 1, berbeda dengan logika klasik yang hanya memiliki nilai keanggotaan 0 atau 1. Logika fuzzy sering dianggap sebagai kotak hitam yang menghubungkan ruang input dengan ruang output. Penggunaan himpunan fuzzy membantu mengatasi keterbatasan himpunan crisp yang terkadang kurang adil. Dalam himpunan crisp, perubahan kecil dalam nilai batas dapat menyebabkan perbedaan kategori yang signifikan. Dengan menggunakan himpunan fuzzy, perubahan kecil tersebut dapat diantisipasi dan kategori yang lebih akurat dapat ditentukan [4].

Pada penelitian ini akan dilakukan pengklusteran data iris dengan cara melakukan pengukuran berdasarkan lebar mahkota (petal width), panjang mahkota (petal length), lebar kelopak (sepal width), dan panjang kelopak (sepal length). Dalam hal ini, pengklasteran dilakukan untuk mengelompokkan bunga-bunga berdasarkan karakteristik tersebut.

2. Metode Penelitian

Dalam jurnal ini, dilakukan pengklasteran data iris menggunakan metode, yaitu metode fuzzy c-means. Sebagaimana dijelaskan pada bab sebelumnya mengenai algoritma metode tersebut. Untuk memahami metode fuzzy c-means dapat dilihat dari flowchart algoritma, gambar 1. Tahapan penelitian merupakan ilustrasi langkah- langkah metode yang akan dikerjakan.



Gambar 1. Alur Logika Fuzzy

Gambar 1. Ini menggambarkan langkah-langkah utama dalam algoritma Fuzzy C-Means untuk pengklasteran data iris. Program dimulai dengan memuat data dari file CSV, kemudian melakukan inisialisasi matriks keanggotaan. Selanjutnya, program melakukan pembaruan pusat kluster dan matriks keanggotaan secara berulang sampai mencapai konvergensi atau mencapai batas iterasi maksimum yang ditentukan. Setelah itu, program mencetak pusat kluster dan label kluster untuk setiap sampel data iris sebelum akhirnya berakhir.

2.1. Fuzzy C-Means

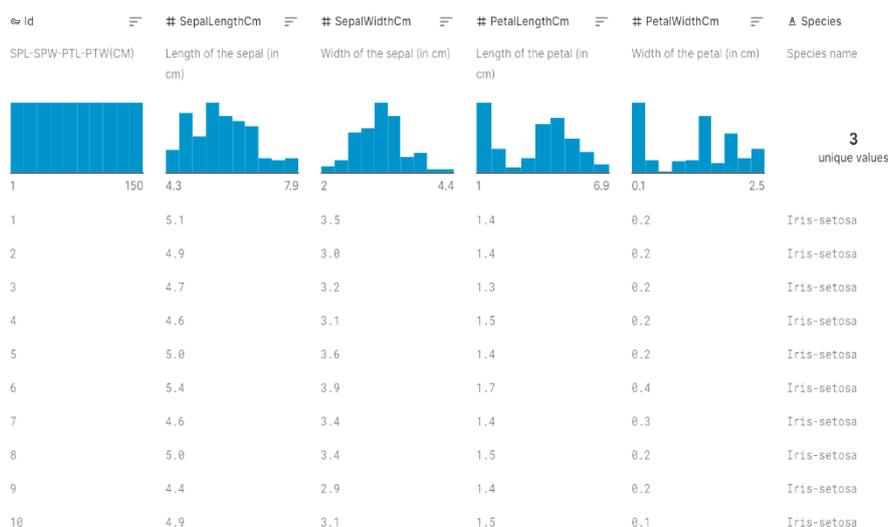
Logika fuzzy merupakan suatu logika yang memiliki nilai kekaburan atau kesamaran (fuzziness) antara benar atau salah. Logika Fuzzy pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dalam teori logika fuzzy suatu nilai bias bernilai benar atau salah secara bersama. Namun berapa besar keberadaan dan kesalahan suatu tergantung pada bobot keanggotaan yang dimilikinya. Logika fuzzy memiliki derajat keanggotaan dalam rentang 0 hingga 1. Berbeda dengan logika digital yang hanya memiliki dua nilai 1 atau 0. Logika fuzzy digunakan untuk menterjemahkan suatu besaran yang diekspresikan menggunakan bahasa (linguistic), misalkan besaran kecepatan laju kendaraan yang diekspresikan dengan pelan, agak cepat, cepat, dan sangat cepat. Dan logika fuzzy menunjukkan sejauh mana suatu nilai itu benar dan sejauh mana suatu nilai itu salah. Tidak seperti logika klasik (crisp)/ tegas, suatu nilai hanya mempunyai 2 kemungkinan yaitu merupakan suatu anggota himpunan atau tidak. Derajat keanggotaan 0 (nol) artinya nilai bukan merupakan anggota himpunan dan 1 (satu) berarti nilai tersebut adalah anggota himpunan [5].

2.2. Data dan Sumber Data

Penelitian ini menggunakan dataset *Pengelompokan pada Dataset Iris* terkait dengan latar belakang yang didapat dari platform kaggle.

2.3. Analisis Data

Tahap ini bertujuan untuk memastikan integritas data sehingga nantinya tidak menimbulkan masalah pada proses data training atau pelatihan data. data iris yang digunakan terbagi menjadi 4 kelas yaitu berdasarkan lebar mahkota (petal width), panjang mahkota (petal length), lebar kelopak (sepal width), dan panjang kelopak (sepal length).

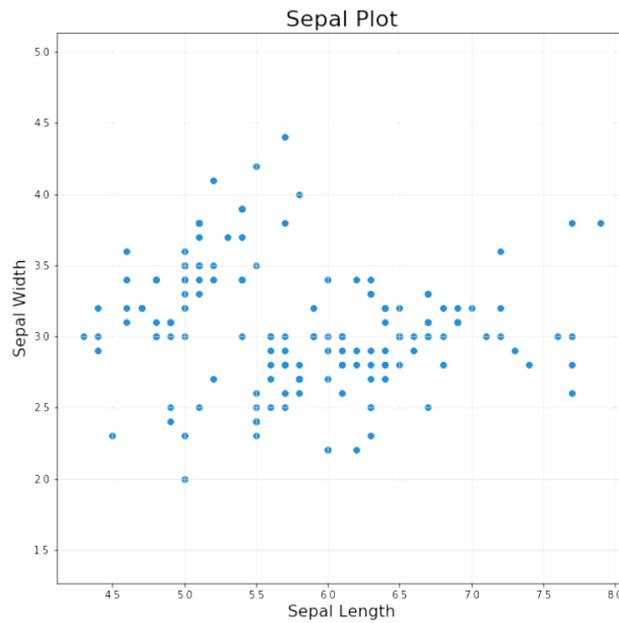


Gambar 2. Analisis Data

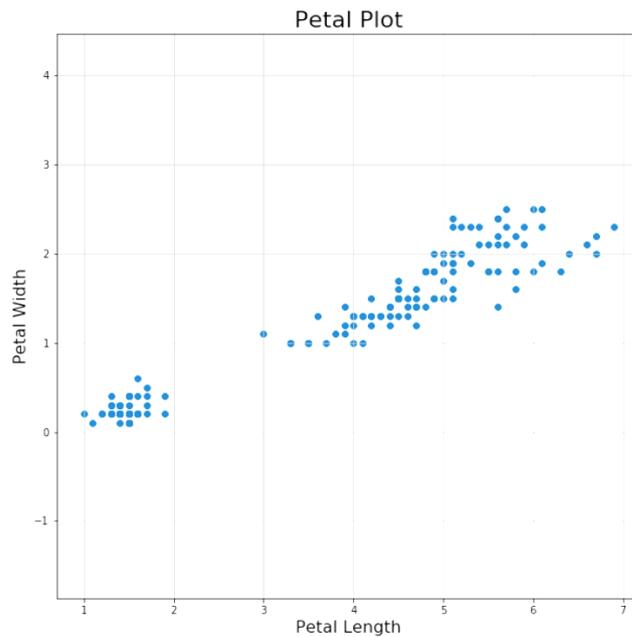
3. Hasil dan Pembahasan

3.1. Pengujian dan Evaluasi

Pada tahap ini, Setelah dijalankan, pengklasteran akan menghasilkan kelompok-kelompok data. Kelompok-kelompok data ini dapat direpresentasikan dalam bentuk grafik. Dengan demikian, dapat dilihat sebaran data pada setiap kluster berdasarkan kedekatannya dengan pusat kluster, hal tersebut terlihat seperti pada gambar 3. dan gambar 4.



Gambar 3. Sebaran Data 1



Gambar 4. Sebaran Data 2

- [5] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, New York: Plenum Press, 1981.