

ABSTRAK

Bayu Wiradarma. 2014. Implementasi Algoritma *Fuzzy C-Means* untuk Pengelompokan Tingkat Penyakit Anemia. Skripsi Program Studi Ilmu Komputer, Program Teknologi Informasi Dan Ilmu Komputer Universitas Brawijaya. Pembimbing : Drs. Marji, MT. dan Nurul Hidayat, S.Pd., MSc.

Penelitian ini membahas penerapan algoritma *Fuzzy C-Means* (FCM) dalam mengelompokkan data tingkat penyakit anemia. Anemia merupakan keadaan dimana jumlah sel darah merah atau jumlah *hemoglobin* (protein pembawa oksigen), leukosit, trombosit dan eritrosit dibawah normal atau kekurangan sel darah merah, sehingga darah tidak dapat mengangkut oksigen dalam jumlah sesuai yang diperlukan tubuh. *Clustering* merupakan metode mengelompokkan suatu objek ke dalam sejumlah kelompok (*cluster*) yang sesuai. Prinsip dari *clustering* adalah meminimumkan kesamaan antar anggota satu *cluster* dan memaksimalkan kesamaan antar anggota *cluster* yang berbeda. *Fuzzy C-Means* merupakan (FCM) merupakan teknik pengklusteran dimana tiap-tiap data ditentukan oleh derajat keanggotaannya. Semakin besar nilai derajat keanggotaan data dalam suatu *cluster* maka semakin besar pula data tersebut menjadi anggota *cluster* tersebut. Hasil *clustering* kemudian diproses atau dievaluasi menggunakan nilai *F-measure* dimana sebelumnya harus diketahui nilai *precision* dan nilai *recall* terlebih dahulu. Pengujian dilakukan dengan sebaran data yang berbeda dengan jumlah dataset yaitu 150 dataset, 180 dataset dan 210 dataset. Dari ketiga pengujian yang dilakukan menunjukkan bahwa nilai *F-measure* dari 150 dataset yang tertinggi pada *cluster* ke 2 yaitu 0.61013, *F-measure* dari 180 dataset yang tertinggi pada *cluster* ke 2 yaitu 0.58547 dan *F-measure* dari 210 dataset yang tertinggi pada *cluster* ke 2 yaitu 0.55989 yang dilakukan sebanyak 10 kali. Perbedaan nilai *f-measure* dikarenakan pada proses awal pengelompokkan terdapat pembangkitan bilangan random untuk pembentukan matriks awal dan karena sebaran data yang digunakan berbeda-beda pada data latih dan data uji. Sehingga pada pengelompokkan data tingkat penyakit anemia ini nilai *f-measure* yang tertinggi berada pada cluster 2.

Kata kunci : pengelompokkan data, penyakit anemia , *fuzzy c-means*



ABSTRACT

Bayu Wiradarma. 2014. The Implementation Of *Fuzzy C-Means* Algorithm To Group Of Anemia Disease Rate. Skripsi Program Studi Ilmu Komputer, Program Teknologi Informasi Dan Ilmu Komputer Universitas Brawijaya. Advisor : Drs. Marji, MT. and Nurul Hidayat, S.Pd., MSc.

Research discusses the application of the algorithm of *Fuzzy C-Means* (FCM) to group the data of anemia disease rate. Anemia is a condition when red blood cell rate or hemoglobin rate (oxygen carrying protein), leucocyte, thrombocyte and erythrocyte are below normal or when red blood cell is lacking such that blood cannot transport oxygen in necessary amount required by the body. *Clustering* is a method to group an object into a number of appropriate groups (*cluster*). The principle of *clustering* is to minimize the similarity between members of a *cluster* and to maximize the similarity between members of different *cluster*. *Fuzzy C-Means* (FCM) is a clustering technique where each data is determined by the degree of membership. Higher degree of membership of the data in a *cluster* is meant as higher possibility that the data will be the member of the *cluster*. Result of *clustering* is processed or evaluated using *F-Measure* rate after their *precision* or *recall* rates are understood. Three data sets such as 150 data sets, 180 data sets, and 210 data sets, which represent the distribution of data, are tested. Based on result of three tests, the highest *F-measure* rate from 150 data sets is 0.61013 found at 2nd cluster, the highest *F-measure* rate from 180 data sets is 0.58547 found at 2nd cluster, and the highest *F-measure* rate from 210 data sets is 0.55989 found at 2nd cluster. Each test is repeated 10 times. Different *F-measure* rates are obtained because early grouping process has generated random numbers to establish early matrix and also because data distribution shall be different at practice data and test data. Therefore, in the grouping of anemia disease rate data, the highest *F-measure* rate remains in the 2nd cluster.

Keywords: data grouping, anemia disease, *Fuzzy C-Means*