

## **ABSTRACT**

### **IMPLEMENTATION OF FUZZY C-MEANS CLUSTERING AND PROXIMITY-IMPACT-POPULARITY ON USER-BASED COLLABORATIVE FILTERING**

Rubila Dwi Adawiyah

14/360054/PA/15759

Memory-based Collaborative Filtering (CF) is a widely used recommender system method due to its easy implementation. However, it suffers from sparsity, scalability, and cold start problems. These problems influence the performance of the recommender system. This research attempted to overcome scalability and sparsity problem on one type of memory-based CF which is user-based CF by implementing Fuzzy C-Means (FCM) clustering and to overcome cold-start problem by using Proximity-Impact-Popularity (PIP) as similarity measure. There are two defuzzification methods namely Best Cluster and All Cluster defuzzification. The system is implemented in MovieLens dataset with sparsity level 93,69 %.

The result is PIP gave higher accuracy and higher coverage (MAE = 0,7734 and coverage = 74.32%) than Pearson correlation (MAE = 0.849588 and coverage = 47%). However, it still suffers very poor scalability (throughput = 4239,113 rec/sec). The implementation of FCM both defuzzification methods improve the scalability of the system with higher throughput (throughput FCM Best Cluster defuzzification = 7386,41 rec/sec, throughput FCM All Cluster defuzzification = 15552,1 rec/sec). In increasing number of cold users, FCM with both defuzzification methods also result in higher accuracy and coverage.

**Keywords:** Collaborative Filtering, Recommender System, Sparsity, Cold Users, Fuzzy C-Means

## INTISARI

### IMPLEMENTASI FUZZY C-MEANS CLUSTERING DAN PROXIMITY- IMPACT-POPULARITY PADA USER-BASED COLLABORATIVE FILTERING

Rubila Dwi Adawiyah

14/360054/PA/15759

*Memory-based Collaborative Filtering* (CF) adalah metode sistem rekomendasi yang banyak digunakan karena implementasinya yang mudah. Namun, *memory-based CF* memiliki permasalahan *sparsity*, skalabilitas, dan *cold-start*. Masalah-masalah ini mempengaruhi performa sistem rekomendasi. Penelitian ini mencoba untuk mengatasi skalabilitas dan masalah *sparsity* pada salah satu tipe *memory-based CF* yaitu *user-based CF* dengan mengimplementasikan Fuzzy C-Means (FCM) dan mengatasi masalah *cold-start* dengan menggunakan Proximity-Impact-Popularity (PIP) sebagai perhitungan similaritas. Ada dua metode defuzzifikasi yang digunakan yaitu Best Cluster dan All Cluster defuzzification. Dataset yang digunakan yaitu *MovieLens* dataset dengan tingkat *sparsity* 93,69%.

Hasil dari penelitian ini adalah PIP menghasilkan akurasi dan *coverage* yang lebih tinggi ( $MAE = 0,7734$  dan  $coverage = 74,32\%$ ) daripada korelasi Pearson ( $MAE = 0,849588$  dan  $coverage = 47\%$ ). Namun, skalabilitas PIP sangat buruk ( $throughput = 4239,113$  rec/sec). Implementasi FCM pada kedua metode defuzzifikasi tersebut meningkatkan skalabilitas sistem dengan *throughput* yang lebih tinggi (*throughput* FCM Best Cluster defuzzification=7386,41 rec/sec, *throughput* FCM All Cluster defuzzification=15552.1 rec/sec). Dalam peningkatan jumlah *cold-user* (pengguna baru), FCM dengan kedua metode defuzzifikasi juga menghasilkan akurasi dan *coverage* yang lebih tinggi.

**Kata Kunci:** Collaborative Filtering, Recommender System, Sparsity, Cold User