

## INTISARI

### PARALELISASI METODE KOMBINASI *CONTENT BASED* DAN *COLLABORATIVE FILTERING* PADA SISTEM REKOMENDASI DI PLATFORM APACHE SPARK

RAKHMAD IKHSANUDIN

15/388496/PPA/04935

*Collaborative Filtering* sebagai metode yang populer dalam sistem rekomendasi. Improvisasi dilakukan dengan tujuan untuk meningkatkan akurasi dari hasil rekomendasi. Salah satu cara yang ditempuh adalah dengan mengkombinasikannya dengan metode *content based*. Namun teknik penggabungan tersebut memiliki kekurangan dalam hal skalabilitas. Penelitian ini berusaha untuk mengatasi masalah skalabilitas yang dihadapi oleh sistem rekomendasi yang menggunakan metode *hybrid collaborative filtering* dan *content based* dengan menerapkan paralelisasi pada *platform* Apache Spark.

Sistem rekomendasi dengan metode *hybrid collaborative filtering* dan *content based* ini diimplementasikan pada cluster Apache Spark yang di dalamnya terdapat sejumlah *node worker* yang saling terintegrasi dengan *master*. Dengan memanfaatkan API RDD pada Apache Spark, dataset dan pemrosesan dapat berjalan secara paralel di semua *node worker*. Dataset yang digunakan adalah dataset film dari MovieLens. Untuk mengukur skalabilitas sistem, dibuat berbagai ukuran *cluster*. Untuk kemudian dihitung perbandingan nilai *speedup* pada setiap ukuran *cluster*.

Berdasarkan hasil pengujian, didapatkan nilai *speedup* metode *hybrid collaborative filtering* dan *content based* pada *cluster* Apache Spark dengan 2 *node worker* adalah sebesar 1,003 yang kemudian meningkat menjadi 2,913 pada *cluster* yang mempunyai 4 *node worker*. Nilai *speedup* meningkat kembali menjadi 5,85 pada *cluster* yang memuat 7 *node worker*.

**Kata Kunci** : sistem rekomendasi, kombinasi *content based* dan *collaborative filtering*, Apache Spark, skalabilitas

## ABSTRACT

### ***PARALLELIZATION OF HYBRID CONTENT BASED AND COLLABORATIVE FILTERING METHOD IN RECOMMENDATION SYSTEM WITH APACHE SPARK PLATFORM***

RAKHMAD IKHSANUDIN

15/388496/PPA/04935

*Collaborative Filtering as a popular method that used for recommendation system. Improvisation is done in purpose of improving the accuracy of the recommendation. A way to do this is to combine with content based method. But the hybrid method has a lack in terms of scalability. The main aim of this research is to solve problem that faced by recommendation system with hybrid collaborative filtering and content based method by applying parallelization on the Apache Spark platform.*

*Recommendation system with hybrid collaborative filtering and content based method is implemented on the Apache Spark cluster. There are a number of node workers are integrated with the master. By using the Apache Spark RDD API, datasets and processes can run in parallel across all worker nodes. The dataset used is the movie dataset from MovieLens. A variety of cluster sizes are made to measure the scalability of the system. Then calculate the speedup ratio of each cluster size.*

*Based on the test results, the value of hybrid collaborative filtering method and content based on Apache Spark cluster with 2 node worker is 1,003 which then increased to 2,913 on cluster having 4 node worker. The speedup got more increased to 5,85 on the cluster that containing 7 node worker.*

**Keywords** : *recomendation system, hybrid content based and collaborative filtering method, Apache Spark, scalability*



UNIVERSITAS  
GADJAH MADA

**PARALELISASI METODE KOMBINASI CONTENT BASED DAN COLLABORATIVE FILTERING PADA  
SISTEM REKOMENDASI DI  
PLATFORM APACHE SPARK**

RAKHMAD IKHSANUDIN, Edi Winarko, M. Sc., Ph.D

Universitas Gadjah Mada, 2018 | Diunduh dari <http://etd.repository.ugm.ac.id/>