

SISTEM REKOMENDASI COLLABORATIVE FILTERING PADA SMARTPHONE MENGUNAKAN K-MEANS

Reyhan Pratama¹, Donni Richasdy², Ramanti Dharayani³

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹reyhanpratama@student.telkomuniversity.ac.id, ²donnir@telkomuniversity.ac.id,

³dharayani@telkomuniversity.ac.id

Abstrak

Smartphone memenuhi kebutuhan *user* dengan menyediakan berbagai layanan komunikasi yang memungkinkan seperti transfer informasi dalam bentuk teks, grafik, suara, dan layanan Internet. Banyak dari masyarakat kebingungan untuk memilih dari banyak nya merk dan tipe yang beredar di pasar saat ini. Maka dari itu penelitian ini melakukan pemberian rekomendasi dengan perbandingan prediksi rating *smartphone* menggunakan metode *K-Means* dengan membandingkan tiga perhitungan *similarity* diantaranya *Pearson*, *Pearson Baseline* dan *Cosine*, dan penggunaan jumlah tetangga yang bervariasi. Dilakukan perbandingan tingkat kinerja antara skenario yang berbeda. Berdasarkan perhitungan dan analisis yang sudah dilakukan, didapatkan skenario antara penggunaan jumlah *trainset* 80% dan *testset* 20%, metode *similarity Pearson Baseline*, dan 90 jumlah tetangga menghasilkan nilai error terkecil dengan nilai RMSE 0.6599 yang merupakan skenario *K-Means* dengan kinerja paling tinggi dalam penelitian ini. Sedangkan skenario penggunaan jumlah *trainset* 70% dan *testset* 30%, metode *similarity Pearson*, dan 10 jumlah tetangga menghasilkan nilai error terbesar dengan nilai RMSE 0.7279 yang berarti skenario tersebut memiliki kinerja paling rendah.

Kata kunci: *Smartphone, K-Means, User-based Collaborative Filtering, Similarity, RMSE.*

Abstract

Smartphones fulfill the needs of users by providing a variety of possible communication services such as information transfer in the form of text, graphics, sound, and Internet services. Many people are confused to choose from the many brands and types circulating in the market today. Therefore, this research provides recommendations by comparing smartphone rating predictions using the K-Means method by comparing three similarity calculations including Pearson, Pearson Baseline and Cosine, and using a varied number of neighbors. A comparison of performance levels between different scenarios is carried out. Based on the calculation and analysis that has been done, it is found that the scenario between the use of the number of trainsets 80% and test set 20%, Pearson baseline similarity method, and 90 number of neighbors produces the smallest error value with an RMSE value of 0.6599 which is the K-Means scenario with the highest performance in this study. While the scenario using the number of trainsets 70% and test set 30%, Pearson similarity method, and 10 number of neighbors produces the largest error value with an RMSE value of 0.7279 which means the scenario has the lowest performance.

Keywords: *Smartphone, K-Means, User-based Collaborative Filtering, Similarity, RMSE.*

1. Pendahuluan

Latar Belakang

Pesatnya perkembangan teknologi di zaman sekarang ini membuat masyarakat haus akan informasi dan eksistensi, khususnya kaum remaja yang saat ini mengikuti perubahan teknologi yang berlangsung untuk berbagai kebutuhan penunjang keseharian mereka. *Smartphone* memenuhi kebutuhan user dengan menyediakan: layanan komunikasi yang memungkinkan transfer informasi dalam bentuk teks, grafik dan suara, layanan Internet nirkabel seperti browsing dan *e-mail*, dan layanan multimedia dan hiburan seperti layar berwarna, gerakan gambar, kamera, game, dan musik [1].

Smartphone sangat cepat berinovasi setiap tahunnya banyak *Smartphone* baru yang bermunculan, banyak dari masyarakat kebingungan untuk memilih dari banyak nya merk dan tipe yang beredar di pasar saat ini. Faktor yang paling mempengaruhi pilihan konsumen akan perangkat telepon seluler adalah fitur inovatif, rekomendasi pribadi dan harga dibandingkan dengan faktor lain seperti kualitas gambar, aspek daya tahan dan portabel, pengaruh media dan layanan purnajual [2].

Terdapat penelitian yang sudah berhasil menerapkan sistem rekomendasi *smartphone*, seperti yang dilakukan oleh Wahyu Henditya membandingkan metode content-based filtering, collaborative filtering dan mixed hybrid menggunakan nilai Accuracy, Precision dan Recall. Mendapatkan akurasi tertinggi dengan nilai rata-rata precision

33%, recall 73.17%, dan accuracy 95.11% menggunakan metode mixed hybrid[3]. Berikutnya penelitian yang dilakukan oleh Nugraha D, dkk mengembangkan sistem rekomendasi film menggunakan metode *user-based collaborative filtering*, dilakukan pengujian 4 preferences genre menghasilkan nilai error yaitu 0,73% [4]. Adapun penelitian yang dilakukan oleh Khusna A, pengujian RMSE sistem rekomendasi berbasis collaborative filtering yang diimplementasikan dapat menghasilkan akurasi rekomendasi sebesar 90.08%[5].

Oleh karena itu sistem yang dapat merekomendasikan smartphone kepada users dengan akurasi yang tinggi dibutuhkan, sudah banyak sistem yang dibangun untuk merekomendasikan items kepada users khususnya di dalam *e-commerce* di dunia dan di Indonesia, banyak nya sistem yang dibangun dengan bermacam-macam metode yang digunakan. Penulis disini membangun metode yang umum digunakan dalam membangun sistem rekomendasi, yaitu *K-Means* dengan parameter sebagai skenario yang berbeda untuk menentukan skenario dengan kinerja terbaik dalam penelitian ini.

Topik dan Batasannya

Topik yang dibahas pada tugas akhir ini adalah pemberian nilai prediksi rating *smartphone* sebagai rekomendasi, menggunakan pendekatan *User-based Collaborative Filtering*, dan dilakukan pengukuran tingkat kinerja berdasarkan skenario dari model *K-Means* menggunakan RMSE. Adapun batasan masalah dalam tugas akhir; Data yang diamati hanya data yang berada dalam dataset, pengumpulan dataset diambil dari kuesioner, hanya menggunakan tiga metode *similarity*, dan data *smartphone* diambil dari *GSMArena*.

Tujuan

Tujuan dari penelitian ini adalah untuk memberikan rekomendasi menggunakan metode *K-Means*, dengan penggunaan jumlah tetangga yang nilainya bervariasi, perbandingan jumlah *trainset* dan *testset* yang berbeda dan perhitungan *similarity* yang berbeda, diantaranya; *pearson correlation*, *pearson baseline*, dan *cosine*. Dalam membangun model prediksi menggunakan skenario yang berbeda, untuk mengetahui kinerja terbaik dari penggunaan skenario yang dilakukan, berdasarkan nilai error paling kecil menggunakan RMSE.

2. Studi Terkait

2.1 Recommender System

Recommender System melakukan pengumpulan data logis, pemeringkatan dan prosedur penyaringan dengan menangani yang ditentukan *user* pertanyaan [6]. Sistem rekomendasi bertujuan untuk menyarankan artikel kepada *user*. Sistem rekomendasi pada awal perkembangannya menjalankan algoritma yang menggunakan rekomendasi-rekomendasi yang diberikan oleh sekelompok pelanggan untuk memperoleh rekomendasi bagi seorang pelanggan tertentu [7]. Ada banyak sekali teknik yang sudah dipelajari tentang sistem rekomendasi, beberapa berdasarkan jumlah beratnya data dan juga berdasarkan dari preferensi *user* [8]. Algoritma ini dibagi menjadi dua yaitu *collaborative filtering* dan *content-based filtering*.

2.2 Collaborative Filtering

Collaborative filtering itu sendiri memprediksi preferensi *user* berdasarkan preferensi masa lalu mereka [9]. Ide utama dari *collaborative filtering* adalah menggunakan penilaian dari *user* lain yang ada untuk memprediksi *item* yang mungkin disukai atau bahkan yang akan diminati oleh target *user*. *Collaborative filtering* itu sendiri terbagi menjadi dua jenis yaitu *user-based* dan *item-based*.

2.3 User-Based Collaborative Filtering

User-based collaborative filtering adalah teknik yang digunakan untuk merekomendasikan kepada *user* aktif *item* yang sama yang disukai *user* lain di masa lalu [10]. Dengan menggunakan rating *user* sebelumnya pada *item* tertentu, akan memprediksi *item* yang akan direkomendasikan kepada target *user* [11]. Langkah pertama dalam *user-based collaborative filtering* adalah menemukan kesamaan *user* dengan target *user*. Kemiripan untuk dua *user* 'a' dan 'b' dapat dihitung menggunakan persamaan *similarity*.

2.4 Similarity

Similarity adalah masalah mendasar dalam tugas klasifikasi dan pengelompokan. Konsep *similarity* terkait dengan konsep jarak [12]. Perhitungan nilai *similarity* antara *users* dengan membandingkan rating dengan produk yang umum, kemudian memprediksi rating dari produk dengan rating rata-rata dari produk dengan *user* yang sama dengan *users* aktif. Pembobotan didefinisikan sebagai kesamaan *user* dengan target produk. Algoritma *Pearson Correlation* digunakan untuk mengukur nilai *similarity* antara dua atribut berurutan berhubungan satu sama lain [13] dapat didefinisikan sebagai: