Sistem Rekomendasi Collaborative Filtering Pada Smartphone Menggunakan K-Means

1st Reyhan Pratama
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
aldillarafi@student.telkomuniversity.ac.id

2nd Donni Richasdy
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
mirakania@telkomuniversity.ac.id

3rd Ramanti Dharayani
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
monterico@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 bervariatif. 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.**

I. PENDAHULUAN

a. 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.

b. 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.

c. Tujuan

Tujuan dari penelitian ini adalah untuk memberikan rekomendasi menggunakan metode K-Means, dengan penggunaan jumlah tetangga yang nilainya bervariatif, 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.

II. KAJIAN TEORI

A. Studi Terkait

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. 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.

3. User-Based Collaborative Filtering

User-based collaborative filtering adalah teknik yang digunakan untuk merekomendasikan kepada user aktif item yang 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.

4. Similarity

Similarity adalah masalah mendasar dalam tugas klasifikasi dan pengelompokan. Konsep similarity terkait dengan konsep jarak [12]. Perhitungan nilai similarity diantara 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:

$$pearson\_sim(a,u) = \frac{\sum_{i=1}^{n}(r_{a,i}-\bar{r}_a)(r_{u,i}-\bar{r}_u)}{\sqrt{\sum_{i=1}^{n}(r_{a,i}-\bar{r}_a)^2}\sqrt{\sum_{i=1}^{n}(r_{u,i}-\bar{r}_u)^2}} \quad (1)$$

Keterangan:
n = jumlah *item* yang diberikan rating oleh *user* a dan u secara bersamaan.
(a, u) = *similarity* antara *user* a dan *user* u.
ra,i = rating yang diberikan oleh *user* a untuk *item* i.
ra = rata-rata rating dari *user* a untuk seluruh *item* yang *user* a berikan rating.
ru,i = rating yang diberikan oleh *user* u untuk *item* i.
ru = rata-rata rating dari *user* u untuk seluruh *item* yang *user* u berikan rating.

Pearson Baseline adalah metode perhitungan dasar yang digunakan dalam sistem rekomendasi untuk mengevaluasi tingkat preferensi user terhadap item tertentu. Metode ini didasarkan pada *Pearson Correlation* antara nilai prediksi dan nilai aktual.

$$pearson\_baseline\_sim(a,u) = \frac{\sum_{i\in I_{au}}(r_{ai}-b_{ai})\cdot(r_{ui}-b_{ui})}{\sqrt{\sum_{i\in I_{au}}(r_{ai}-b_{ai})^2}\cdot\sqrt{\sum_{i\in I_{au}}(r_{ui}-b_{ui})^2}} \quad (2)$$

Keterangan:
(a, u) = *similarity* antara *user* a dan *user* u.
rai = rating dari *item* i yang diberikan oleh *user* a.
rui = rating dari *item* i yang diberikan oleh *user* u.
bai = rata-rata rating dari semua *item* yang diberikan oleh *user* a.
bui = rata-rata rating dari semua *item* yang diberikan oleh *user* u.
Iau = himpunan *item* yang diberi rating oleh kedua *user* a dan u.

Cosine similarity adalah metode yang digunakan untuk mengukur kesamaan antara dua vektor dalam ruang vektor. Metode ini mengukur kesamaan antara dua vektor dengan menghitung kosinus dari sudut antara kedua vektor tersebut. *Cosine similarity* dapat digunakan dalam berbagai aplikasi, seperti pengelompokan data, rekomendasi sistem, dan analisis teks.

$$cosine\_sim(a,u) = \frac{\sum_{i\in I} r_{ai}\cdot r_{ui}}{\sqrt{\sum_{i\in I_{au}} r_{ai}^2}\sqrt{\sum_{i\in I_{au}} r_{ui}^2}} \quad (3)$$

Keterangan:
(a, u) = *similarity* antara *user* a dan *user* u.
rai = rating dari *item* i oleh *user* a.
rui = rating dari *item* i oleh *user* u.
i ∈ Iau = menghitung keseluruhan *item* dalam himpunan I.

5. K-Means

K-Means adalah metode untuk menemukan struktur klaster dalam kumpulan data yang ditandai dengan kesamaan terbesar dalam cluster yang sama dan ketidaksamaan terbesar di antara cluster yang berbeda[14]. Algoritma k-means bergantung pada nilai k; yang selalu perlu ditentukan untuk untuk melakukan analisis pengelompokan apa pun. Pengelompokan dengan nilai k yang berbeda pada akhirnya akan menghasilkan hasil yang berbeda [15].

Pendekatan ini merupakan algoritma pengelompokan parsial yang mempartisipasi seluruh dataset dalam k disjoint cluster. Algoritma ini menangani kumpulan dataset dan dengan cepat meng konvergen ke local yang optimum. Dalam K-Means langkah pertama k yang merupakan poin awal dipilih dimana k adalah parameter yang menentukan jumlah kelompok yang akan dicari, dan parameter ini ditentukan di awal. Dipilih secara acak menggunakan titik awal sebagai pusat kelompok, lalu memindai seluruh dataset yang tersisa dan menetapkan ke setiap titik data pengelompokan terdekat berdasarkan nilai similarity.

Sementara rata-rata dari semua kelompok dihitung dan pusat kelompok diperbaharui ke nilai rata-rata, selanjutnya seluruh proses ini diulang dengan nilai pusat yang baru dan semua titik dipindahkan ke kelompok yang baru. Pembaharuan dalam nilai pusat didasarkan pada penugasan setiap titik baru ke kelompok atau dihapusnya setiap titik dari kelompok. Nilai pusat tetap dilakukan pembaharuan setiap dilakukan iterasi dan proses ini berlanjut sampai tidak ada pembaharuan di dalam kelompok pusat.

$$r_{ui} = \mu_u + \frac{\sum_{v \in N_i^k(u)} sim(a,u) \cdot (r_{vi} - \mu_v)}{\sum_{v \in N_i^k(u)} sim(a,u)} \quad (4)$$

Keterangan

rui = Perkiraan rating yang akan diberikan *user* u untuk *item* i

μu = Rata-rata rating yang diberikan *user* u

$N^k(u)$ = K tetanggą terdekat dari *user* u yang memberikan rating pada *item* i

$sim(a, u)$ = *Similarities* antara *user* a dan *user* u

rvi = Rating yang diberikan oleh *user* v pada *item* i

μv = Rata-rata rating dari *user* v

6. Root Mean Square Error

Root Mean Square Error (RMSE) merupakan rumus yang diterapkan dalam mengetahui tingkat akurasi suatu model. Hasil tingkat akurasi didapat melalui data penguji untuk mendapatkan nilai yang optimum [16]. *Root mean square error* (RMSE) telah digunakan sebagai metrik statistik standar standar untuk mengukur kinerja model dalam penelitian meteorologi, kualitas udara, dan iklim, dalam bidang geosains, banyak yang menggunakan RMSE sebagai metrik standar untuk kesalahan model[17]. *Root Mean Square Error* (RMSE) sendiri merupakan besaran tingkat kesalahan hasil prediksi, dimana semakin besar nilai RMSE maka semakin tidak akurat hasil prediksi, sebaliknya semakin kecil nilai RMSE maka hasil prediksi akan semakin akurat [18]. Berikut adalah rumus RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Yref - Ypred)^2}{n}} \quad (5)$$

Keterangan :

$Ypred$ = Nilai prediksi output ulangan ke-1

$Yref$ = Nilai aktual output ulangan ke-1

n = Banyak data yang digunakan

III. METODE

A. Sistem yang Dibangun

Berikut adalah *flowchart* sistem rekomendasi collaborative filtering pada smartphone *K-Means*.

GAMBAR 1.
Flowchart umum gambaran sistem yang dibangun.

Gambar 1 merupakan *flowchart* gambaran penggerjaan dalam penelitian ini. Pada penelitian ini, akan membandingakn beberapa skenario yang akan dijabarkan dan akan dilakukan analisis terhadap skenario yang telah dilakukan.

1. Pengumpulan Data

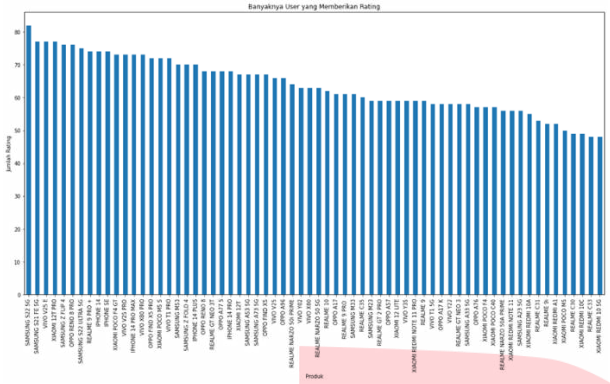
Pada penelitian ini dataset yang didapat berasal dari penyebaran kuesioner menggunakan Google Form, target dari penyebaran kuesioner ini adalah teman sebaya dan keluarga. Data yang sudah tersedia merupakan data smartphone keluaran rentan tahun 2022 berjumlah 66 produk smartphone. Pada kuesioner user diminta untuk memberikan rating kepada setiap produk yang mereka ketahui, tidak setiap produk diberi rating oleh users, kuesioner ini mendapatkan total 100 user yang telah memberikan rating terhadap produk smartphone yang mereka ketahui.

TABEL 1.
Sample Dataset Penelitian

date	Score	score_ max	User	product
5/12/2022 17:32	5.0	5.0	Luthfi Goldiansyah Kusumajadi	Samsung Z Flip 4
5/12/2022 17:32	4.0	5.0	Tandya Rizky Pratama	Samsung Z Flip 4
5/12/2022 17:35	4.0	5.0	M. Emir Sidiq	Samsung Z Flip 4
14/01/2023 13:45	2.0	5.0	Devi Aridaini	Xiaomi Redmi A1
14/01/2023 15:14			Intan Maharani	Xiaomi Redmi A1
15/01/2023 09:17	2.0	5.0	Mega Fitriani	Xiaomi Redmi A1

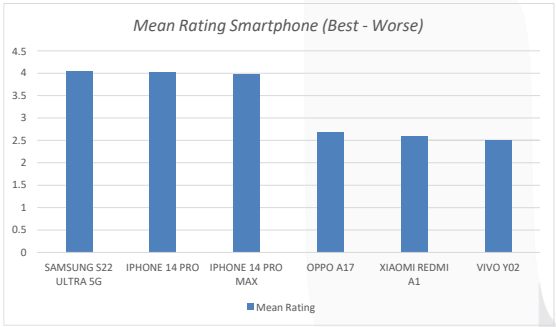
Tabel 1 merupakan hasil kuesioner yang telah disebarkan, dataset ini dibagi menjadi 6600 *rows* dan 5 *columns*, terdapat 5 *columns* diantaranya adalah; *date, score, score_max, user, dan product*. Dimana *date* adalah *timestamp* user memberikan rating, *score* adalah rating

yang diberikan oleh users kepada *product, score_max* merupakan nilai rating tertinggi pada setiap *product*, dan *user* adalah nama *users*, dan *product* adalah nama produk.



GAMBAR 2.
Banyaknya User yang Memberikan Rating.

Gambar 2, yang menggambarkan penyebaran jumlah rating yang diberikan oleh users kepada produk. Berdasarkan dataset yang digunakan dalam penelitian ini yaitu berasal dari kuesioner yang disebarkan, produk dengan nama 'SAMSUNG S22 5G' merupakan produk yang paling banyak diberikan rating oleh *user*. Sedangkan produk dengan nama 'XIAOMI REDMI 10 5G' merupakan produk yang paling sedikit diberikan rating oleh *user*.



GAMBAR 4.
Rating Rata-Rata Smartphone Terbaik dan Terburuk

Gambar 4, menampilkan *smartphone* dengan rata-rata rating terbaik dan terburuk. Smartphone dengan rata-rata rating terbaik adalah 'SAMSUNG S22 ULTRA 5G' dan smartphone dengan rata-rata rating terburuk adalah 'VIVO Y02'.

2. Processing Data

Gambar 5.
Flowchart Processing Data.

Pada *Flowchart Processing* data merujuk kepada Gambar 5, dataset dilakukan *Preprocessing* data, setelah itu dilakukan perhitungan prediksi rating dilakukan perhitungan *similarity* menggunakan *Pearson, Pearson Baseline,* dan *Cosine.* Kemudian dilakukan perhitungan nilai error menggunakan RMSE.

a. Preprocessing Data

Dalam proses ini dilakukan pengecekan pada dataset apabila ada data yang kosong, data kosong dalam *rows score* dan *score_max* akan diisikan nilai *mean*, dilakukan juga *drop columns* yang tidak diperlukan, dataset juga telah diacak menggunakan *random_state* = 1, hingga akhir nya menampilkan dataset yang siap diproses, dataset setelah dilakukan *preprocessing* berjumlah 6600 *rows* dan 3 *columns*.

TABEL 3.
Sample Dataset Setelah Preprocessing

score	user	product
3.0	Rangga	Realme C33
5.0	Devita	Samsung S22 Ultra 5G
4.0	Jalza	Samsung S22 Ultra 5G
4.0	Dea	Samsung Z Flip 4
3.0	Intan	Xiaomi Redmi Note 11 Pro

Pada Tabel 2, data yang sudah dilakukan preprocessing meninggalkan 3 columns yang akan diproses diantaranya; *score* berisikan ranting oleh *user* kepada *product*, *user* merupakan user, dan *product* adalah produk smartphone.

Langkah selanjutnya adalah melakukan *splitting*, data dibagi menjadi dua yaitu *traintes* dan *testset*. *Trainset* digunakan untuk melatih model *machine learning*, model akan mempelajari pola dan korelasi antara *item* dan rating yang diberikan oleh *users*. Kemudian *testset* digunakan untuk mengetahui seberapa baik kinerja dari model yang telah dilatih sebelumnya menggunakan *trainset*, dan sebagai skenario pembagian jumlah persentase data adalah *trainset* 70%, *testset* 30%, dan *trainset* 80%, *testset* 20%.

b. Perhitungan Similarity

Dataset yang sebelumnya sudah dilakukan splitting menjadi trainset, dan testset digunakan untuk perhitungan similarity, sebagai skenario pada penelitian ini akan membandingkan tiga metode perhitungan similarity yaitu menggunakan Pearson, Pearson Baseline, dan Cosine dengan library surprise.

c. Prediksi Rating (K-Means)

Setelah dilakukan perhitungan nilai similarity hasil nilai similarity akan dilakukan perhitungan nilai prediksi menggunakan K-Means dengan library surprise. Trainset akan dilakukan perhitungan nilai prediksi, setelah model didapatkan kemudian dilakukan validasi model menggunakan testset setelah didapatkan nilai prediksi rating. Dengan skenario membandingkan jumlah tetangga menjadi 9 bagian (10, 20, 30, 40, 50, 60, 70, 80, 90) untuk menentukan jumlah tetangga yang optimal dalam penggunaan model K-Means sebagai metode rekomendasi dalam penelitian ini. Model akan dinilai tingkat kinerjanya menggunakan RMSE berdasarkan nilai error terkecil.

d. Nilai Kinerja

Setelah model yang sudah dilakukan validasi dengan testset, hasil dari model K-Means akan dilakukan penilaian kinerja menggunakan RMSE, sebagai pengukur nilai error daripada skenario yang dilakukan dalam peneitian ini.

IV. HASIL DAN EVALUASI

A. Hasil Analisis

Hasil analisis dalam penelitian ini, adalah membandingkan nilai RMSE antara penggunaan jumlah trainset dan testset yang berbeda, penggunaan tiga metode similarity yang berbeda, dan perbedaan jumlah tetangga *K-Means* yang berbeda.

GAMBAR 6.
Perbandingan Hasil RMSE pada Trainset 80% dan Testset 20%.

Merujuk kepada Gambar 6, hasil perhitungan skenario yang dilakukan menunjukan bahwa skenario yang menghasilkan nilai RMSE paling kecil adalah penggunaan *Pearson Baseline* dengan jumlah tentangga 90 dan jumlah *trainset* 80% banding *testset* 20%, dapat diartikan bahwa penggunaan skenario ini memiliki kinerja yang paling tinggi. Hal ini dikarenakan metode *Pearson Baseline* menggabungkan teknik perhitungan *Pearson Correlation* dengan *Baseline* rating rata-rata dari *user* atau *item*. Memungkinkan metode ini untuk mengatasi masalah yang dihadapi oleh metode *Pearson Correlation* dan *Cosine*, seperti perbedaan skala rating antara *user* atau *item* dan masalah bias pada perhitungan *Pearson*. Karena itu, metode *Pearson Baseline* lebih akurat dibandingkan metode *Pearson* dan *Cosine*. Adapun penggunaan jumlah tetangga yang cukup banyak membuat model dapat menangkap lebih banyak informasi dari tetangga-tetangga yang berdekatan dengan *user* yang akan diberikan rekomendasi.

GAMBAR 7.
Perbandingan Hasil RMSE pada Trainset 70% dan Testset 30%.

Sedangkan merujuk kepada Gambar 7, penggunaan metode *Pearson* dengan jumlah tetangga 10 dan perbandingan antara *trainset* 70% dan *testset* 30% menghasilkan nilai RMSE yang paling besar yaitu 0.7279 yang berarti skenario ini memiliki kinerja terburuk dalam penelitian ini. Karena penggunaan jumlah tetangga sejumlah 10 dalam skenario ini menyebabkan *overfitting*, yaitu ketika model terlalu menyesuaikan dengan data *training* menyebabkan kinerjanya menurun pada data *testing*.

B. Hasil Rekomendasi

Hasil rekomendasi menggunakan skenario yang memiliki kinerja paling tinggi dalam penelitian ini. Perhitungan *similarity* antar user menggunakan *Pearson Baseline*, dengan jumlah tetangga K = 90, dan perbandingan *trainset* 80% dan *testset* 20%. Didapatkan hasil rekomendasi untuk setiap *user*, hasil rekomendasi setiap *user* berdasarkan preferensi setiap *user* atas pemberian rating *item*, kemudian setiap *user* akan diberikan rekomendasi *item* berdasarkan *user* lain yang memiliki preferensi yang serupa. Berikut adalah hasil rekomendasi tiga produk *smartphone* untuk *user* Reyhan Pratama:

TABEL 4.
Hasil Rekomendasi *Smartphone* Terhadap Reyhan Pratama.

	Produk	Prediksi Rating
Rekomendasi ke - 1	SAMSUNG S22 ULTRA 5G	4.473275
Rekomendasi ke - 2	IPHONE 14 PRO	4.337610
Rekomendasi ke - 3	SAMSUNG A73 5G	4.003942

Pada Tabel 4, menunjukan hasil estimasi rating produk sebagai rekomendasi produk kepada target *user* yaitu Reyhan Pratama, target *user* mendapatkan tiga rekomendasi produk yang memiliki estimasi rating paling tinggi. Rekomendasi pertama adalah produk dengan nama 'SAMSUNG S22 ULTRA 5G' dengan estimasi rating 4.473275, rekomendasi produk kedua adalah 'IPHONE 14 PRO' dengan estimasi rating 4.337610, dan rekomendasi produk trakhir adalah 'SAMSUNG A73 5G' dengan estimasi rating 4.003942.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan analisis yang telah dilakukan dalam penelitian ini, penggunaan metode K-Means dalam sistem rekomendasi Collaborative Filtering pada smartphone mampu memperhitungkan nilai prediksi rating produk kepada target user sebagai hasil rekomendasi. Penelitian ini memiliki perbedaan dengan penelitian sebelumnya, diantaranya dataset yang berbeda, penggunaan metode yang berbeda, dan juga menentukan parameter yang memiliki kinerja paling tinggi berdasarkan nilai error yang paling kecil pada metode K-means. Membandingkan kinerja model atas jumlah trainset dan testset yang berbeda, penggunaan metode similarity yang berbeda, dan penggunaan jumlah tetangga yang berbeda.

Penggunaan jumlah trainset 70% dan testset 30%, metode similarity pearson, dan tetangga sejumlah 10, menghasilkan nilai error yang paling besar dengan nilai

RMSE 0.7279. Sedangkan penggunaan jumlah trainset 80% dan testset 20%, metode perhitungan similarity menggunakan pearson baseline, dan tetangga sejumlah 90 menghasilkan nilai error paling kecil dengan nilai RMSE 0.6599, dapat dikatakan bahwa kombinasi tersebut memiliki kinerja paling tinggi dalam penelitian ini. Saran yang dapat dilakukan selanjutnya adalah penggunaan metode yang berbeda untuk menghasilkan nilai error yang lebih rendah, dan pengembangan selanjutnya seperti implementasi sistem ini diatas website.

REFERENSI

[1] A. Y. Yaakop and S. Mokhlis, "Consumer Choice Criteria in Mobile Phone Selection: An Investigation of Malaysian University Students," International Review of Social Sciences and Humanities, vol. 2, no. 2, pp. 203–212, 2012, [Online]. Available: www.irssh.com

[2] H. Karjaluoto et al., "Factors affecting consumer choice of mobile phones: Two studies from Finland,"

[3] Journal of Euromarketing, vol. 14, no. 3, pp. 59–82, 2005, doi: 10.1300/J037v14n03_04.

[4] W. HENDITYA, "2 BAB II DASAR TEORI 2.1 Pengertian Smartphone," Bandung, 2016. Accessed: May 20, 2022. [Online]. Available: https://openlibrary.telkomuniversity.ac.id/pustaka/117014/sistem- rekomendasi-pembelian-smartphone-menggunakan-metode-collaborative-filtering-dan-content-based- filtering.html

[5] D. Nugraha, T. W. Purboyo, and R. A. Nugrahaeni, "LAMPIRAN V-Format Jurnal TA SISTEM REKOMENDASI FILM MENGGUNAKAN METODE USER BASED COLLABORATIVE FILTERING (MOVIE RECOMMENDATION SYSTEM USING USER BASED COLLABORATIVE FILTERING METHOD)."

[6] A. N. Khusna, K. P. Delasano, and D. C. E. Saputra, "Penerapan User-Based Collaborative Filtering Algorithm," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 20, no. 2, pp. 293–304, May 2021, doi: 10.30812/matrik.v20i2.1124.

[7] S. Chakrabarti, H. N. Saha, University of British Columbia, Institute of Electrical and Electronics Engineers. Vancouver Section, and Institute of Electrical and Electronics Engineers, IEEE IEMCON - 2016 : the 7th IEEE Annual Information Technology, Electronics & Mobile Communication

[8] Conference : 13-15 October 2016, University of British Columbia, Vancouver, Canada.

[9] S. Sari and D. Tri Hendra, "Aplikasi Rekomendasi Film menggunakan Pendekatan Collaborative Filtering dan Euclidean Distance sebagai ukuran kemiripan rating."

[10] A. Chauhan, D. Nagar, and P. Chaudhary, "Movie recommender system using sentiment analysis," in Proceedings of International Conference on Innovative Practices in Technology and Management, ICIPTM 2021, Feb. 2021, pp. 190–193. doi: 10.1109/ICIPTM52218.2021.9388340.

[11] S. Shaposhnikov et al., Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference (2017 ElConRus) : February 1-3, 2017, St. Petersburg, Russia, 2017.

[12] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in

[13] Recommender Systems Handbook, Springer US, 2011, pp. 1–35. doi: 10.1007/978-0-387-85820-3_1.

[14] H. Koohi and K. Kiani, "User based Collaborative Filtering using fuzzy C-means," Measurement (Lond), vol. 91, pp. 134–139, Sep. 2016, doi: 10.1016/j.measurement.2016.05.058.

[15] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," Inf Sci (N Y), vol. 307, pp. 39–52, Jun. 2015, doi: 10.1016/j.ins.2015.02.024.

[16] F. Mansur, V. Patel, and M. Patel, "A Review on Recommender Systems."

[17] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," IEEE Access, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[18] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Electronics (Switzerland), vol. 9, no. 8. MDPI AG, pp. 1–12, Aug. 01, 2020. doi: 10.3390/electronics9081295.

[19] M. Nilashi, K. Bagherifard, O. Ibrahim, H. Alizadeh, L. A. Nojeem, and N. Roozegar, "Collaborative filtering recommender systems," Research Journal of Applied Sciences, Engineering and Technology, vol. 5, no. 16, pp. 4168–4182, 2013, doi: 10.19026/rjaset.5.4644.

[20] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?," Geosci. Model Dev. Discuss, vol. 7, pp. 1525–1534, 2014, doi: 10.5194/gmdd-7-1525-2014.

[21] W. Wang and Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," in IOP Conference Series: Materials Science and Engineering, Apr. 2018, vol. 324, no. 1. doi: 10.1088/1757-899X/324/1/012049.