



Algoritma *Alternating Least Squares* Untuk Mesin Rekomendasi Film

Balqis Dwian Fitri Zamzami¹, Aisyah Tiara Pratiwi², Della Septiani³, Nadia Fitri Yani⁴, Yosia Letare Banurea⁵, Luluk Muthoharoh⁶, Ardika Satria⁷, Rizty Maulida Badri⁸

^{1,2,3,4,5,6,7,8}Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

¹balqis.121450018@student.itera.ac.id

³rizty.badri@sd.itera.ac.id

Corresponding author email: luluk.muthoharoh@sd.itera.ac.id

Abstract: *The entertainment world is inseparable from the rapidly growing movie industry and is accompanied by huge data growth. The rapid growth of data has brought about a new era of information. These data are utilized to build innovative, efficient and more effective systems. This research implements a movie recommendation system using the Alternating Least Squares (ALS) algorithm from Apache Spark MLlib with the MovieLens 25M dataset. A collaborative filtering approach with matrix factorization is used to model user preferences and movie characteristics. The evaluation is done by calculating the Root Mean Squared Error (RMSE) on the test data. The first ALS model with default parameters produced the best RMSE of 0.811671, while the second model with different parameters produced an RMSE of 0.839577. The results show that the ALS model with default parameters is able to provide accurate and relevant movie recommendations according to user preferences*

Keywords: *Film, Recommendation, Big Data, Alternating Least Square, Pyspark*

Abstrak: Dunia hiburan tak terlepas dari industri film yang berkembang sangat pesat dan disertai dengan pertumbuhan data yang sangat besar. Pesatnya pertumbuhan data telah membawa era baru informasi. Data-data ini dimanfaatkan untuk membangun sistem-sistem yang inovatif, efisien dan lebih efektif. Penelitian ini mengimplementasikan sistem rekomendasi film menggunakan algoritma Alternating Least Squares (ALS) dari Apache Spark MLlib dengan dataset MovieLens 25M. Pendekatan collaborative filtering dengan matrix factorization digunakan untuk memodelkan preferensi pengguna dan karakteristik film. Evaluasi dilakukan dengan menghitung Root Mean Squared Error (RMSE) pada data uji. Model ALS pertama dengan parameter default menghasilkan RMSE terbaik sebesar 0.811671, sementara model kedua dengan parameter yang berbeda menghasilkan RMSE 0.839577. Hasil menunjukkan bahwa model ALS dengan parameter default mampu memberikan rekomendasi film yang akurat dan relevan sesuai preferensi pengguna

Kata kunci: *Alternating Least Square, Big Data, Film, Pyspark, Rekomendasi*

I. PENDAHULUAN

Dunia hiburan tak terlepas dari industri film yang berkembang sangat pesat dan disertai dengan pertumbuhan data yang sangat besar[1]. Industri film juga didukung dengan keberadaan platform streaming dan peningkatan penggunaan internet. Dengan hal tersebut, masyarakat umum dapat lebih leluasa dengan memiliki akses yang mudah untuk menikmati berbagai jenis film. Dengan disertai banyaknya film yang dihadirkan berbagai *platform streaming*, jumlah konten yang ada akan semakin berlimpah menjadikan tantangan yang muncul, salah satunya yaitu kaitannya bagaimana cara untuk menyajikan rekomendasi film yang sesuai dengan preferensi pengguna seiring pesatnya pertumbuhan data. Data-data ini dimanfaatkan untuk membangun sistem-sistem yang inovatif, efisien dan lebih efektif. Di masa data melimpah ini, mesin rekomendasi menjadi salah satu kategori penting dalam sistem penyaringan data yang berfungsi untuk memprediksi tingkat penilaian atau rating yang akan diberikan oleh pengguna terhadap objek-objek yang diminatinya di internet. Butuh hampir 15 tahun tanpa henti untuk menyaksikan seluruh koleksi Netflix, HBO Go, Prime Video, dan Disney Hotstar saja. Mayoritas pelanggan tidak sanggup menyaksikan sebagian kecil saja koleksi tontonannya[2]. Sehingga perlu suatu mesin pemberi rekomendasi. Dalam konteks ini, rekomendasi film dapat



dianggap sebagai sebuah implementasi dari pendekatan pengolahan data besar. Pada umumnya, mesin rekomendasi bekerja dengan menyaring kualitas hasil pencarian dan kemudian menampilkan item-item yang paling relevan dengan minat pengguna berdasarkan riwayat pencarian mereka[3].

Ledakan data dan informasi memunculkan permasalahan penemuan kembali informasi (*information retrieval*). Semakin banyak bisnis yang hadir secara online untuk melacak keunggulan kompetitif mereka karena pertumbuhan besar-besaran jaringan internet. Agar organisasi mana pun bisa sukses secara domestik atau internasional, internet telah menjadi instrumen yang sangat penting. Pertumbuhan Internet yang eksponensial telah mengubah cara kerja sebagian besar perusahaan. Internet menyediakan platform unik bagi perusahaan *e-Commerce* untuk menawarkan dan membeli produk dan layanan bagi konsumen. Semakin banyak bisnis yang hadir secara online untuk melacak keunggulan kompetitif mereka karena pertumbuhan besar-besaran jaringan internet. Agar organisasi mana pun bisa sukses secara domestik atau internasional, internet telah menjadi instrumen yang sangat penting. Pertumbuhan Internet yang eksponensial telah mengubah cara kerja sebagian besar perusahaan. Internet menyediakan platform unik bagi perusahaan *e-Commerce* untuk menawarkan dan membeli produk dan layanan bagi konsumen. Mesin pencari (*search engine*) memecahkan sebagian masalah itu, namun personalisasi informasi tidak diberikan karena mesin pencari terbatas pada kata/*query* yang dimasukkan pengguna, persoalan muncul ketika pengguna mengetahui produk yang dibutuhkan tapi tidak mengetahui nama produk tersebut atau kata yang mewakili produk tersebut, seperti berkeliling di salah satu mall besar atau perpustakaan besar dimana setelah berputar seharian tidak menemukan barang yang dicari. Rekomendasi dari seorang ahli yang mengerti kebutuhan dan profil pengguna menjadi sangat dibutuhkan. Disinilah era digital membawa era Big Data meninggalkan popularitas mesin pencari menuju mesin rekomendasi[4].

Belakangan ini, sistem rekomendasi memegang peranan penting dalam berbagai bidang penelitian dan implementasikan pada banyak platform pengambilan informasi seperti Amazon, Netflix, dan YouTube. Sistem pemberi rekomendasi memperoleh informasi dari klien dan merekomendasikan barang yang menurutnya paling bernilai di antara produk yang ada[5]. Sistem rekomendasi ini bekerja dengan menghasilkan rekomendasi berdasarkan riwayat pencarian pengguna dan data dari platform-platform seperti LinkedIn, Facebook, Amazon, dan Netflix. Tujuan utama dari platform tersebut adalah untuk memberikan pengalaman yang lebih baik kepada penggunanya[6]. Sistem rekomendasi adalah salah satu yang paling penting aplikasi dalam analisis data besar dan telah berkinerja sangat baik untuk banyak bisnis[7], [8].

Salah satu pendekatan yang paling populer dalam sistem rekomendasi adalah *collaborative filtering*, dimana rekomendasi dihasilkan dengan menganalisis pola preferensi pengguna yang memiliki selera serupa. Algoritma *Alternating Least Squares* (ALS) telah muncul sebagai metode *collaborative filtering* berbasis model yang kuat, terutama dalam hal akurasi rekomendasi. ALS bekerja dengan memodelkan preferensi pengguna dan karakteristik item sebagai vektor laten dalam ruang dimensi yang sama, dan kemudian memprediksi rating pengguna terhadap item yang belum dinilai dengan mengalikan vektor laten tersebut[9]. Apache Spark, sebuah mesin analitik yang terdistribusi mengeksekusi pekerjaan secara in-memory, muncul sebagai solusi untuk mengatasi tantangan skalabilitas Algoritma *Alternating Least Squares* (ALS). Dimana Spark memungkinkan komputasi terdistribusi efisien pada data dalam melalui fitur *Resilient Distributed Dataset* (RDD).

Ada banyak penelitian untuk sistem rekomendasi menggunakan berbagai pendekatan dengan metode *machine learning* seperti *Singular Value Decomposition* (SVD), *Alternating Least Squares* (ALS), dll. Penelitian oleh Adyatama dan Baizal (2023) membuat sistem rekomendasi buku



menggunakan SVD didapatkan hasil *Root Mean Square Error* (RMSE) sebesar 0.868 sedangkan menggunakan ALS hasil RMSE sebesar 109.320. Kesimpulan dari penelitian ini metode SVD lebih baik dibandingkan dengan metode ALS dalam implementasi sistem rekomendasi[9]. Sedangkan penelitian lainnya oleh Martin, dkk. (2020) dalam membuat sistem rekomendasi pada Goodreads menunjukkan hasil RMSE metode ALS lebih baik dibandingkan dengan hasil ALS pada penelitian sebelumnya yaitu sebesar 0.67[10]. Dari hasil penelitian sebelumnya, mereka telah berhasil mengembangkan sebuah model sistem rekomendasi yang efektif dengan menggunakan metode ALS. Model ini mampu mencapai nilai RMSE yang rendah, yang mengindikasikan tingkat akurasi yang baik dalam memberikan rekomendasi. Oleh karena itu, penulis dalam penelitian ini bermaksud untuk mengimplementasikan, menguji, dan mengevaluasi mesin sistem rekomendasi serupa, namun fokus pada mesin sistem rekomendasi film menggunakan data dari situs MovieLens[11]. Sehingga, penulis memilih judul *Algoritma Alternating Least Squares Untuk Mesin Rekomendasi Film Dengan mengadopsi pendekatan yang telah terbukti efektif dalam penelitian sebelumnya, penulis berharap dapat mengembangkan sistem rekomendasi film yang andal dan akurat.*

II. METODE

2.1. Deskripsi Data

Penelitian ini menggunakan dataset MovieLens 25M yang dirilis oleh GroupLens Research di Universitas Minnesota. Dataset ini berisi 25.000.095 rating dan 1.093.360 tag aplikasi dari 162.541 pengguna terhadap 62.423 film, yang dihasilkan antara 9 Januari 1995 hingga 21 November 2019. Dataset MovieLens 25M dipilih karena ukurannya yang besar, sehingga menjadi tantangan yang ideal untuk menguji kinerja dan skalabilitas sistem rekomendasi yang dikembangkan.

Tabel 1. Dataset Ratings

userId	movieId	rating	timestamp
1	296	5.0	1147880044
1	306	3.5	1147868817
1	307	5.0	1147868828
1	665	5.0	1147878820
1	899	3.5	1147868510

Tabel 2. Dataset Links

movieId	imdbId	tmdbId
1	114709	862
2	113497	8844
3	113228	15602
4	114885	31357
5	113041	11862

Tabel 3. Dataset Movies

movieId	title	genres
1	Toy Story (1995)	Adventure Animation
2	Jumanji (1995)	Adventure Children
3	Grumpier Old Men ...	Comedy Romance
4	Waiting to Exhale...	Comedy Drama Romance
5	Father of the Bri...	Comedy



Tabel 4. Dataset Tags

userId	movieId	tag	timestamp
3	260	classic	1439472355
3	260	sci-fi	1439472256
4	1732	dark comedy	1573943598
4	1732	great dialogue	1573943604
4	7569	so bad it's good	1573943455

Tabel 5. Final Dataset

userId	movieId	rating	title	genres
1	307	5.0	Three Colors: Blue	Drama
1	899	3.5	Singin' in the Rain	Comedy Musical Romance
1	1088	4.0	Dirty Dancing (1987)	Drama Musical, Romance
1	1237	5.0	Seventh Seal, The..	Drama
1	1250	4.0	Bridge On the River	Adventure, Drama War
1	1260	3.5	M(1931)	Crime Film-Noir Thriller
32906	1188	5.0	Strictly Ballroom	Comedy Romance
32906	1193	5.0	One Flew Over the...	Drama
32906	1198	5.0	Raiders of the Lo..	Action Adventure

2.2. Collaborative filtering dengan Matrix Factorization pada Algoritma Alternating Least Squares (ALS)

Dalam upaya membangun sebuah sistem rekomendasi film yang akurat dan juga efisien, penelitian ini mengimplementasikan pendekatan *Collaborative Filtering* dengan algoritma ALS dari Apache Spark MLlib. *Collaborative Filtering* merupakan sebuah teknik rekomendasi yang memanfaatkan informasi preferensi atau perilaku pengguna lainnya yang memiliki kemiripan dengan pengguna sasarannya. Teknik ini bekerja dengan menganalisa poin-poin rating atau interaksi pengguna terhadap item-item tertentu, dan kemudian memanfaatkan informasi tersebut untuk memprediksi preferensi pengguna sasaran terhadap item-item yang belum dinilai atau interaksi sebelumnya.

Algoritma *Alternating Least Squares* (ALS) merupakan metode umum dan efisien untuk dekomposisi matriks data ulasan pengguna artikel. Faktorisasi matriks ini bertujuan untuk memetakan pengguna dan item ke dalam ruang vektor laten yang sama sehingga kemiripan antara pengguna dan item dapat dihitung dengan mudah. Faktorisasi matriks merupakan teknik yang digunakan untuk menemukan pola tersembunyi dalam data rating dengan cara memfaktorkan matriks rating pengguna-item menjadi perkalian dua matriks, yaitu matriks faktor pengguna dan matriks faktor item. Hal ini meminimalkan nilai kesalahan kuadrat antara skor aktual dan skor prediksi. Peringkat aktual adalah elemen matriks R yang berisi data penilaian pengguna tentang item tersebut, dan peringkat prediksi diperoleh dengan mengalikan matriks U dan V [12]. Secara matematis, jika kita memiliki matriks rating R dengan dimensi $m \times n$, dimana m adalah jumlah pengguna dan n adalah jumlah item, faktorisasi matriks ini berusaha untuk mendekati matriks R dengan perkalian dua matriks berdimensi lebih kecil:

$$\widehat{R}_{ij} \approx U_i \times V_j^T \quad (1)$$

Keterangan:

\widehat{R}_{ij} : merupakan matriks rating yang diprediksi dengan dimensi m (jumlah item film) \times n (jumlah pengguna)



U_i : matriks faktor pengguna dengan dimensi m (jumlah pengguna) \times k (jumlah faktor laten yang digunakan untuk mengatasi masalah sparsity)
 V_j : matriks faktor item dengan dimensi $n \times k$

Faktor laten ini digunakan untuk mengatasi masalah *sparsity* dalam data rating, yaitu situasi di mana sebagian besar elemen dalam matriks rating tidak memiliki nilai (kosong). Disebutkan bahwa jaringan preferensi dapat menangkap preferensi pengguna lebih baik dalam matriks fitur laten pengguna dibandingkan dengan jejaring sosial. Nilai yang hilang dicari sehingga rekomendasi dapat dipersonalisasi dengan tepat menggunakan model variabel laten[13].

2.3. Optimasi Algoritma Alternating Least Squares (ALS)

Dalam menemukan matriks faktor pengguna U dan matriks faktor item V yang optimal, perlu dilakukan optimasi pada algoritma Algoritma *Alternating Least Squares* (ALS). Optimasi ini bekerja dengan secara iteratif memperbaiki matriks U dan V untuk meminimumkan kesalahan dari prediksi rating terhadap rating sebenarnya. Pada setiap iterasinya, ALS memperbarui matriks V dengan menganggap matriks U konstan. Proses ini akan diulang hingga diperoleh konvergensi nilai matriks U dan V yang menghasilkan kesalahan prediksi paling minimum. Adapun proses optimasi ini dilakukan dengan menggunakan rumus sebagai berikut:

$$V_i = R_i U (U^T U + \lambda I)^{-1} \quad (2)$$

$$U_j = R_j V (V^T V + \lambda I)^{-1} \quad (3)$$

Keterangan:

V_i : matriks pengguna terbaru hasil optimasi
 R_i : vektor rating yang dihasilkan oleh pengguna (i)
 U : matriks faktor laten pengguna
 λ : parameter regulasi untuk menghindari *overfitting*
 I : matriks identitas
 U_j : matriks item terbaru hasil optimasi
 R_j : vektor rating yang diterima oleh item (j)
 V : matriks faktor laten item

Proses ini diulang hingga diperoleh konvergensi nilai matriks U dan V yang menghasilkan kesalahan prediksi paling minimum.

2.4. Evaluasi Model dengan Root Mean Squared Error (RMSE)

Untuk mengevaluasi performa dan keakuratan dari model rekomendasi film yang dihasilkan, matriks *Root Mean Squared Error* (RMSE) digunakan sebagai ukuran kesalahan prediksi rating. *Root Mean Squared Error* (RMSE) merupakan matriks evaluasi yang paling umum digunakan dalam sistem rekomendasi berbasis rating. Matriks ini yang dapat mengukur akar rata-rata kuadrat dari selisih antara nilai rating yang sebenarnya (R_{ij}) yang diberikan oleh pengguna dengan nilai rating prediksi (\hat{R}_{ij}) yang dihasilkan oleh model yang dibangun.

Secara matematis, *Root Mean Squared Error* (RMSE) didefinisikan sebagai:

$$RMSE = \sqrt{\frac{1}{|R|} \sum_{i=1}^R (R_{ij} - \hat{R}_{ij})^2} \quad (4)$$

Keterangan:

R : himpunan pasangan pengguna dan item (i,j) yang memiliki nilai rating sebenarnya

R_{ij} : nilai rating sebenarnya yang diberikan oleh pengguna (i) untuk item (j)

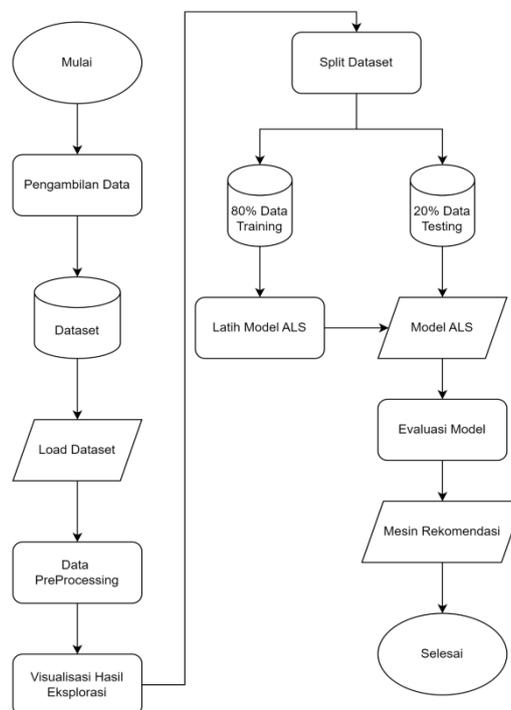
\widehat{R}_{ij} : nilai rating yang diprediksi oleh model ALS untuk pengguna (i) dan item (j)

$|R|$: jumlah total pasangan pengguna dan item yang memiliki rating

Nilai RMSE biasanya berkisar antara 0 hingga 1. Semakin rendah nilai RMSE, semakin akurat performa model dalam memprediksi rating yang mendekati nilai sebenarnya.

2.5. Desain Penelitian

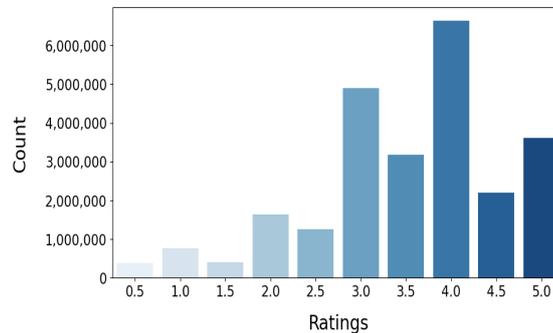
Desain penelitian berikut menjelaskan berbagai tahapan dalam membangun sistem rekomendasi film menggunakan algoritma ALS pada dataset MovieLens.



Gambar 1. Tahapan Alur Penelitian

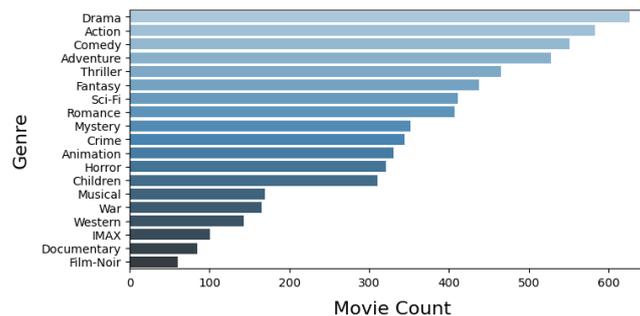
III. HASIL DAN PEMBAHASAN

Peneliti melakukan eksplorasi pada dataset dan didapatkan grafik yang menunjukkan distribusi jumlah rating pada skala 0 hingga 5 untuk sebuah dataset rating film. Hasil grafik ini ditunjukkan pada **Gambar 2**. Dari gambar 2 ini, terlihat bahwa rating tertinggi yaitu 4, mendominasi dengan jumlah sekitar 6 juta. Rating terendah, yaitu 0.5 dan 1, memiliki jumlah yang sangat rendah, sekitar 1 juta atau kurang, menunjukkan bahwa pengguna jarang memberikan rating yang sangat rendah. Secara keseluruhan, distribusi rating ini menunjukkan kecenderungan positif di mana pengguna lebih sering memberikan rating tinggi daripada rendah, mengindikasikan bahwa film-film dalam dataset ini umumnya dianggap baik oleh pengguna. Distribusi ini asimetris dengan puncak pada rating 4 dan penurunan tajam setelah itu, memberikan wawasan tentang preferensi penonton yang dapat membantu dalam pengembangan model rekomendasi yang lebih akurat. Seperti promosi dari mulut ke mulut, *collaborative filtering* memberikan prediksi rating dan personal rekomendasi berdasarkan yang disukai pengguna lain yang mempunyai selera yang sama[14].



Gambar 2. Visualisasi Distribusi Ratings

Grafik selanjutnya pada **Gambar 3**, menunjukkan jumlah film berdasarkan genre dalam dataset. Genre drama mendominasi dengan lebih dari 600 film, menunjukkan popularitas tinggi dari genre lainnya. Di sisi lain, genre dengan jumlah film terendah yaitu film-noir, dengan kurang dari 100 film, menunjukkan bahwa genre ini adalah yang paling jarang ditemukan dalam dataset. Analisis ini memberikan gambaran tentang variasi dan preferensi genre dalam koleksi film tersebut, dengan dominasi drama, action, dan comedy serta keberagaman yang mencakup berbagai genre lainnya.



Gambar 3. Visualisasi Jumlah Film Berdasarkan Genre

Setelah melakukan eksplorasi penulis membangun model rekomendasi film dengan menggunakan algoritma ALS dari pustaka Apache Spark. Model ini dilatih menggunakan data pelatihan dan diuji menggunakan data uji dengan perbandingan 80:20. Pembagian data ini penting untuk memastikan bahwa model yang dibangun tidak terlalu *overfitting* terhadap data latih dan dapat memberikan performa yang baik pada data baru yang belum pernah dilihat sebelumnya [15].

Model ALS pertama yang dibangun dengan menggunakan parameter default dari library Spark MLlib. mesin pembelajaran spark berusaha mencari model terbaik dari kombinasi parameter yang telah ditetapkan. Beberapa parameter utama antara lain *userCol* (kolom yang berisi ID pengguna), *itemCol* (kolom yang berisi ID film), *ratingCol* (kolom yang berisi rating), *coldStartStrategy* (strategi untuk menangani pengguna atau item baru yang tidak memiliki data rating), *nonnegative* (apakah faktor laten harus bernilai non-negatif), dan *implicitPrefs* (apakah rating dianggap sebagai preferensi implisit atau eksplisit). Pada model ini, parameter yang dipilih sebagai parameter default adalah *rank=10*, *regParam=0.1*, dan *maxIter=10*. Kemudian, dilakukan evaluasi terhadap model yang dibangun dengan menghitung RMSE pada data uji menggunakan fungsi *RegressionEvaluator* dari Spark MLlib. RMSE disini digunakan untuk mengevaluasi performa model rekomendasi. Semakin kecil nilai RMSE, maka semakin baik performa model yang dibangun dalam memprediksi rating.



Model yang dibangun menghasilkan nilai *Root Mean Square Error* (RMSE) sebesar 0.811 untuk model dengan parameter default.

Untuk dapat meningkatkan performa yang lebih baik, dilakukan optimasi parameter model dengan mencoba beberapa kombinasi nilai parameter yang berbeda. Proses optimasi ini dilakukan dengan menggunakan parameter yang lebih besar yaitu menggunakan parameter $rank=50$, $regParam=0.15$, dan $maxIter=10$ dari parameter default sebelumnya dengan nilai $rank=10$, $regParam=0.1$, dan $maxIter=10$. Setelah percobaan kedua dengan kombinasi nilai parameter yang berbeda, diperoleh model dengan nilai RMSE sebesar 0.839. Didapatkan hasil RMSE sedikit meningkat dibandingkan dengan model awal.

Didapatkan hasil evaluasi model seperti yang disajikan pada Tabel 6. Terlihat bahwa terdapat perbedaan hasil pengujian dengan perbandingan kedua algoritma tersebut. Untuk algoritma ALS dengan parameter default, nilai RMSE sebesar 0.811671. Untuk algoritma ALS kedua yang tidak menggunakan parameter default, nilai RMSE didapatkan sebesar 0.839577. Dari Tabel 6 juga terlihat bahwa paramater yang berbeda akan menghasilkan RMSE yang berbeda. semakin banyak iterasi data juga semakin terlatih yang menyebabkan RMSE menjadi semakin kecil. Dari hasil kedua tersebut model ALS pertama atau dengan parameter default akan digunakan untuk pemrosesan selanjutnya.

Tabel 6. Nilai RMSE Model ALS.

No	Metode	RMSE
1	ALS model 1	0.811671
2	ALS model 2	0.839577

Pada **Tabel 7**, terlihat bahwa nilai prediksi untuk rating film berkisar antara 3.128 hingga 3.997, yang cukup dekat dengan rating aktual yang diberikan oleh pengguna. Secara keseluruhan, hasil ini menunjukkan bahwa model ALS pertama mampu menangkap preferensi pengguna secara efektif dan memberikan prediksi yang masuk akal untuk film yang belum diberi rating oleh pengguna[16].

Tabel 7. Representasi Hasil Perbandingan dari Prediksi Set Pelatihan dan Pengujian pada model ALS pertama.

User Id	Movie Id	Actual Rating	Predicted Rating
1	4144	5.0	3.997
1	8360	4.0	3.128
1	8327	5.0	3.787
1	8973	4.0	3.739
1	899	3.0	3.680

Dapat dilihat pada Tabel 8, hasil dari mesin sistem rekomendasi film untuk pengguna yang dimasukkan dengan ID 1 menunjukkan bahwa model ALS berhasil menghasilkan daftar film dengan prediksi rating tertinggi. Beberapa film yang direkomendasikan mencakup berbagai genre, dengan banyak film yang memiliki rating prediksi sempurna (5.0). Urutan rekomendasi diatur berdasarkan rating prediksi tertinggi, diikuti oleh judul dan genre, memastikan bahwa film dengan prediksi rating yang sama diurutkan secara konsisten. Hasil ini menunjukkan bahwa model ALS mampu memberikan rekomendasi yang relevan dan beragam berdasarkan preferensi pengguna, memberikan pengalaman yang dipersonalisasi untuk pengguna dengan ID yang dimasukkan pada mesin userID 1.

Permasalahan skalabilitas juga berhasil dipecahkan, setelah training dilakukan, *best model* diciptakan untuk melakukan prediksi dan pemeringkatan, ketika data baru muncul, prosesnya hanya mencocokkan dan mencari rating yang mirip antar item masukan dengan item yang telah diprediksi, hal ini menguntungkan karena proses *generate* menjadi lebih cepat dan ringan dibandingkan melakukan semuanya dari awal seperti pada memori-based[17]. Sejak penelitian sistem rekomendasi dimulai, evaluasi prediksi dan rekomendasi menjadi penting[18], [19]. Penelitian di bidang



kepercayaan dan reputasi dapat memberikan titik awal yang cocok untuk menciptakan interaksi sosial di antara pengguna Sistem Rekomendasi, namun, penelitian yang paling relevan mengenai subjek ini terbatas pada penggunaan hubungan kepercayaan untuk meningkatkan kualitas layanan rekomendasi. Bab buku O'donovan mengkaji keragaman sumber dari mana informasi kepercayaan dapat dimanfaatkan dalam aplikasi web sosial[20].

Tabel 8. Hasil mesin sistem rekomendasi film untuk pengguna dengan ID 1 berdasarkan prediksi tertinggi.

No	Judul Film	Genre	Predicted Rating
1	.hack Liminality In the Case of Kyoko Tohno	Sci-Fi Adventure Fantasy	5.0
2	.hack Liminality In the Case of Yuki Aihara	Sci-Fi Adventure Fantasy	5.0
3	A Dirty Story (1977)	Drama	5.0
4	Acı Aşk (2009)	Drama	5.0
5	Adrenaline (1990)	Action	5.0
6	Familie Brasch (2018)	Documentary	5.0
7	Head Trauma (2006)	Horror Mystery Thriller	5.0
8	I'll Take You There (1999)	Comedy Drama Romance	5.0
9	NOFX Backstage Passport 2	Documentary Music	5.0
10	Sing (1989)	Drama Musical	5.0
11	The Carrier (1988)	Drama Horror Thriller	5.0
12	The Law and the Fist (1964)	Drama War	5.0
13	The Thorn (1971)	Comedy	5.0
14	Vergeef me	Comedy	5.0
15	Mickybo and Me (2005)	Comedy Drama	4.899432
16	Ο Θανάσις στη χώρα της σφαλιάρας (1976)	Drama	4.883376
17	Marlene (1984)	Documentary	4.8807483
18	Once Upon a Ladder (2016)	Comedy Romance	4.8663144
19	National Theatre Live: One Man, Two Guvnors (2011)	Comedy	4.8603783
20	Civilisation (1969)	Documentary History	4.8388166

IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi sistem rekomendasi film menggunakan algoritma *Alternating Least Squares* (ALS) dari Apache Spark MLlib menghasilkan performa yang baik dalam memprediksi rating film. Model ALS pertama dengan parameter default memberikan nilai *Root Mean Squared Error* (RMSE) terbaik sebesar 0.811671, menunjukkan kemampuan yang akurat dalam memprediksi rating film mendekati nilai sebenarnya. Hasil rekomendasi film yang dihasilkan relevan dengan preferensi pengguna dan mencakup berbagai genre. Dengan demikian, pendekatan *collaborative filtering* yang diterapkan dalam penelitian ini terbukti efektif untuk membangun mesin sistem rekomendasi film yang andal dan akurat. Kemampuan Apache Spark dalam memproses data skala besar secara terdistribusi dan paralel memungkinkan penerapan algoritma ALS yang kompleks pada dataset rating film yang besar, menunjukkan peran penting teknologi big data dalam menghasilkan sistem rekomendasi berkualitas tinggi.

REFERENSI

1. I. P. Putri, “Industri Film Indonesia Sebagai Bagian Dari Industri Kreatif Indonesia,” *J. Ilm. LISKI (Lingkar Stud. Komunikasi)*, vol. 3, no. 1, p. 24, 2017, doi: 10.25124/liski.v3i1.805.
2. Mada Kris, “Koleksi Film Melimpah, Pelanggan Kewalahan Menonton,” *Kompas.id*, 2023.
3. A. A. Rehman, M. J. Awan, and I. Butt, “Comparison and Evaluation of Information Retrieval Models,” *VFAST Trans. Softw. Eng.*, vol. 6, no. 1, pp. 7–14, 2018, doi: 10.21015/vtse.v13i1.496.
4. R. K. Sorde and S. N. Desmukh, “Comparative Study on Approaches of Recommendation Systems,”



- Adv. Intell. Syst. Comput.*, vol. 118, no. 2, pp. 10–14, 2015, doi: 10.1007/978-981-15-0947-6_72.
5. H. Hwangbo, Y. S. Kim, and K. J. Cha, “Recommendation system development for fashion retail e-commerce,” *Electron. Commer. Res. Appl.*, vol. 28, pp. 94–101, 2018, doi: 10.1016/j.elerap.2018.01.012.
 6. F. Amato, V. Moscato, A. Picariello, and G. Sperli, “KIRA: A System for Knowledge-Based Access to Multimedia Art Collections,” 2017, p. 1.
 7. Q. Zhao, Y. Zhang, D. Friedman, and F. Tan, “E-commerce recommendation with personalized promotion,” *RecSys 2015 - Proc. 9th ACM Conf. Recomm. Syst.*, no. August, pp. 219–225, 2015, doi: 10.1145/2792838.2800178.
 8. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013, doi: 10.1016/j.knosys.2013.03.012.
 9. H. A. Adyatma and Z. K. A. Baizal, “Book Recommender System Using Matrix Factorization with Alternating Least Square Method,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1286–1292, 2023, doi: 10.47065/josh.v4i4.3816.
 10. S. Martin, J. I. Sihotang, and B. Jonathan, “Mesin Rekomendasi Menggunakan Algoritma Alternating Least Square (ALS) pada Goodreads,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, p. 79, 2020, doi: 10.24014/coreit.v6i2.11578.
 11. F. M. Harper and J. A. Konstan, “The movielens datasets: History and context,” *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, 2015, doi: 10.1145/2827872.
 12. M. J. Awan *et al.*, “A recommendation engine for predicting movie ratings using a big data approach,” *Electron.*, vol. 10, no. 10, 2021, doi: 10.3390/electronics10101215.
 13. R. Lumbantoruan, P. Simanjuntak, I. Aritonang, and E. Simaremare, “TopC-CAMF: Sistem Rekomendasi Matrix Factorization Berbasis Top Context,” *J. Nas. Tek. Elektro dan Teknol. Inf.* |, vol. 11, no. 4, pp. 258–266, 2022.
 14. T. Setiyaningsih and M. N. Afifah, “Data Mining Untuk Sistem Rekomendasi Pemilihan Buku Pada Perpustakaan,” *J. Sains Teknol. Fak. ...*, vol. IX, no. 2, pp. 56–67, 2019, [Online]. Available: <http://repository.unsada.ac.id/id/eprint/1381>
 15. K. McNally, M. P. O’Mahony, M. Coyle, P. Briggs, and B. Smyth, “A case study of collaboration and reputation in social web search,” *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 1, pp. 1–29, 2011, doi: 10.1145/2036264.2036268.
 16. C. Kaleli and H. Polat, “Privacy-preserving SOM-based recommendations on horizontally distributed data,” *Knowledge-Based Syst.*, vol. 33, pp. 124–135, 2012, doi: 10.1016/j.knosys.2012.02.013.
 17. I. S. Wahyudi, “Mesin rekomendasi film menggunakan metode kemiripan genre berbasis collaborative filtering,” 2017. [Online]. Available: <https://core.ac.uk/download/pdf/291462920.pdf>
 18. J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4321 LNCS, no. 1, 2007, pp. 291–324. doi: 10.1007/978-3-540-72079-9_9.
 19. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Analysis of recommendation algorithms for e-commerce,” *EC 2000 - Proc. 2nd ACM Conf. Electron. Commer.*, pp. 158–167, 2000, doi: 10.1145/352871.352887.
 20. C. Christakou, S. Vrettos, and A. Stafylopatis, “A hybrid movie recommender system based on neural networks,” *Int. J. Artif. Intell. Tools*, vol. 16, no. 5, pp. 771–792, 2007, doi: 10.1142/S0218213007003540.