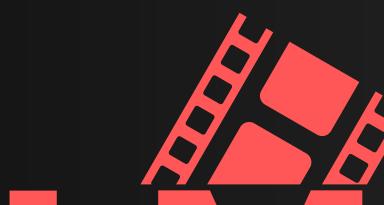




ALGORITMA ALTERNATING LEAST SQUARES UNTUK MESIN REKOMENDASI **FILM**



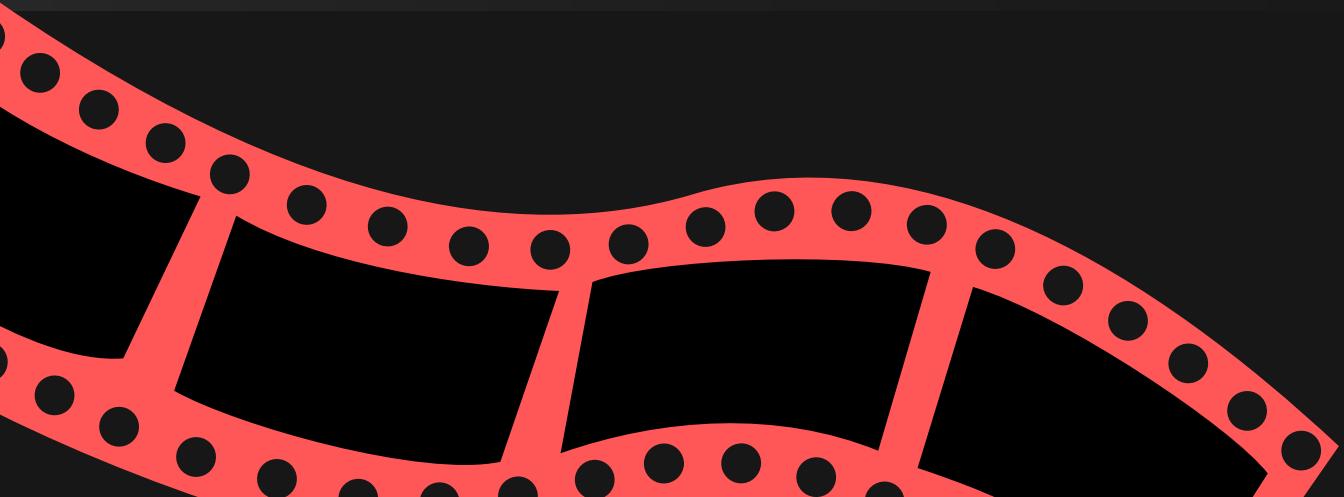
Balqis Dwian Fitri Zamzami, Aisyah Tiara Pratiwi, Della Septiani, Nadia Fitri Yani, Yosia Letare
Banurea, Luluk Muthoharoh, Ardika Satria, Rizty Maulida Badri

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera



Outline

- 1 Pendahuluan**
- 2 Metode**
- 3 Hasil dan Pembahasan**
- 4 Kesimpulan**



Pendahuluan



Dunia hiburan tak terlepas dari industri film yang berkembang sangat pesat dan disertai dengan pertumbuhan data yang sangat besar.



Tujuan penelitian ini untuk mengimplementasikan, menguji, dan mengevaluasi mesin sistem rekomendasi film menggunakan data dari situs MovieLens.



Algoritma Alternating Least Squares (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.

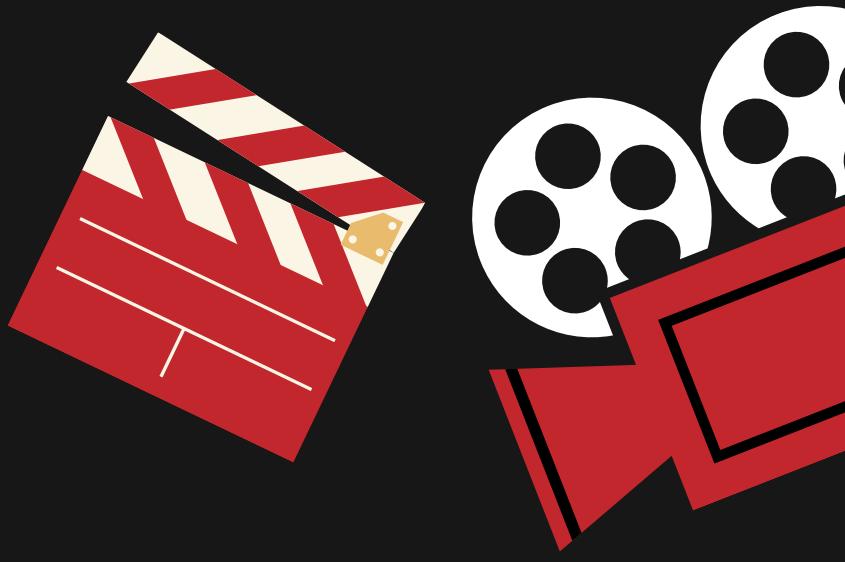




Metode

grouplens

Penelitian ini menggunakan dataset MovieLens 25M yang dirilis oleh GroupLens Research di Universitas Minnesota, mencakup data dari 9 Januari 1995 hingga 21 November 2019.





Metode

Faktorisasi matriks:

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

Keterangan:

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

U_i : matriks faktor pengguna dengan dimensi m (jumlah pengguna) x k (jumlah faktor laten yang digunakan untuk mengatasi masalah sparsity)

V_j : matriks faktor item dengan dimensi n x k

Optimasi ALS:

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

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

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_i : matriks item terbaru hasil optimasi

R_i : vektor rating yang diterima oleh item (j)

V : matriks faktor laten item

Root Mean Squared Error (RMSE):

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

Keterangan:

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

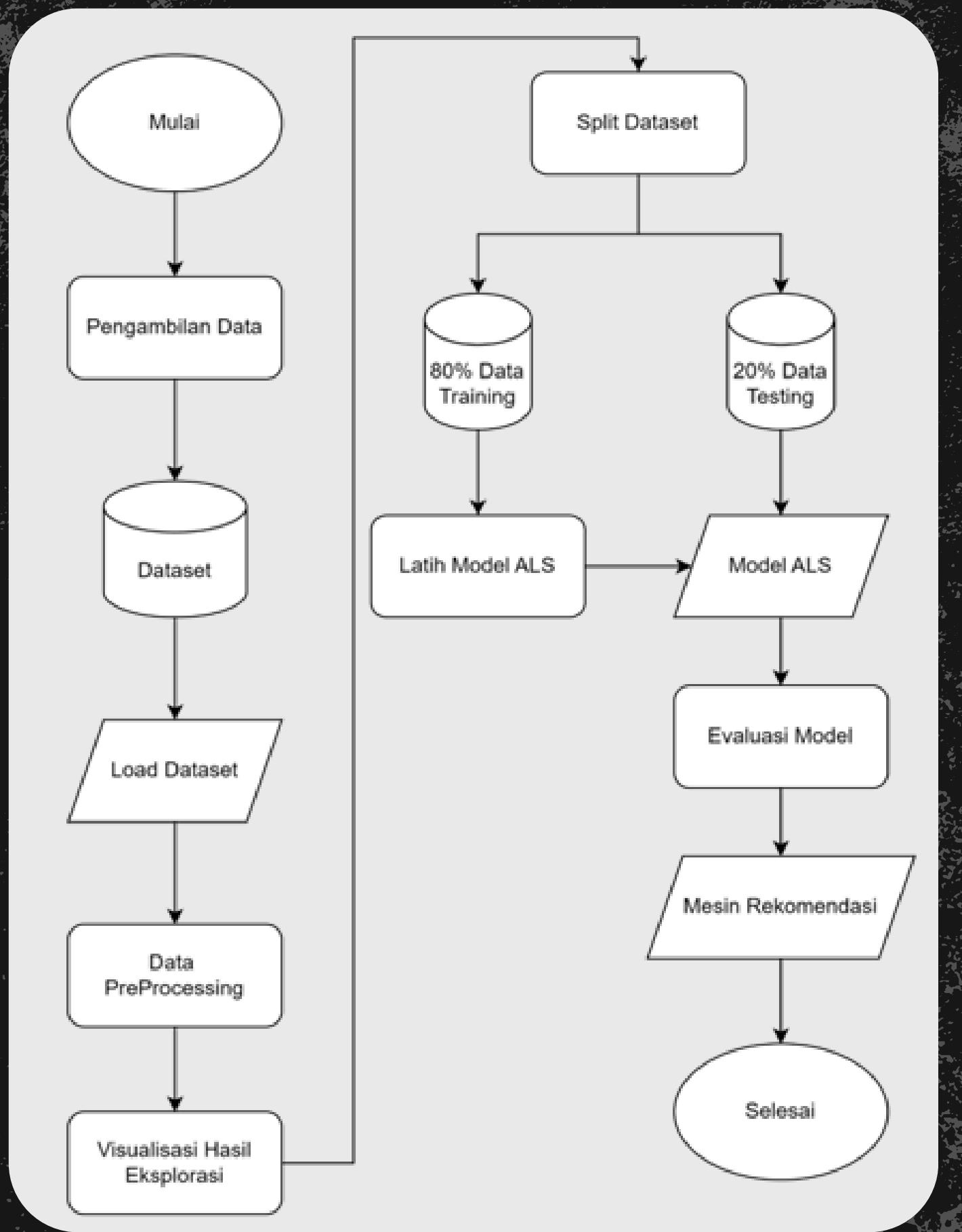
R_{ij} : nilai rating sebenarnya yang diberikan oleh pengguna (i) untuk item (j)

\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

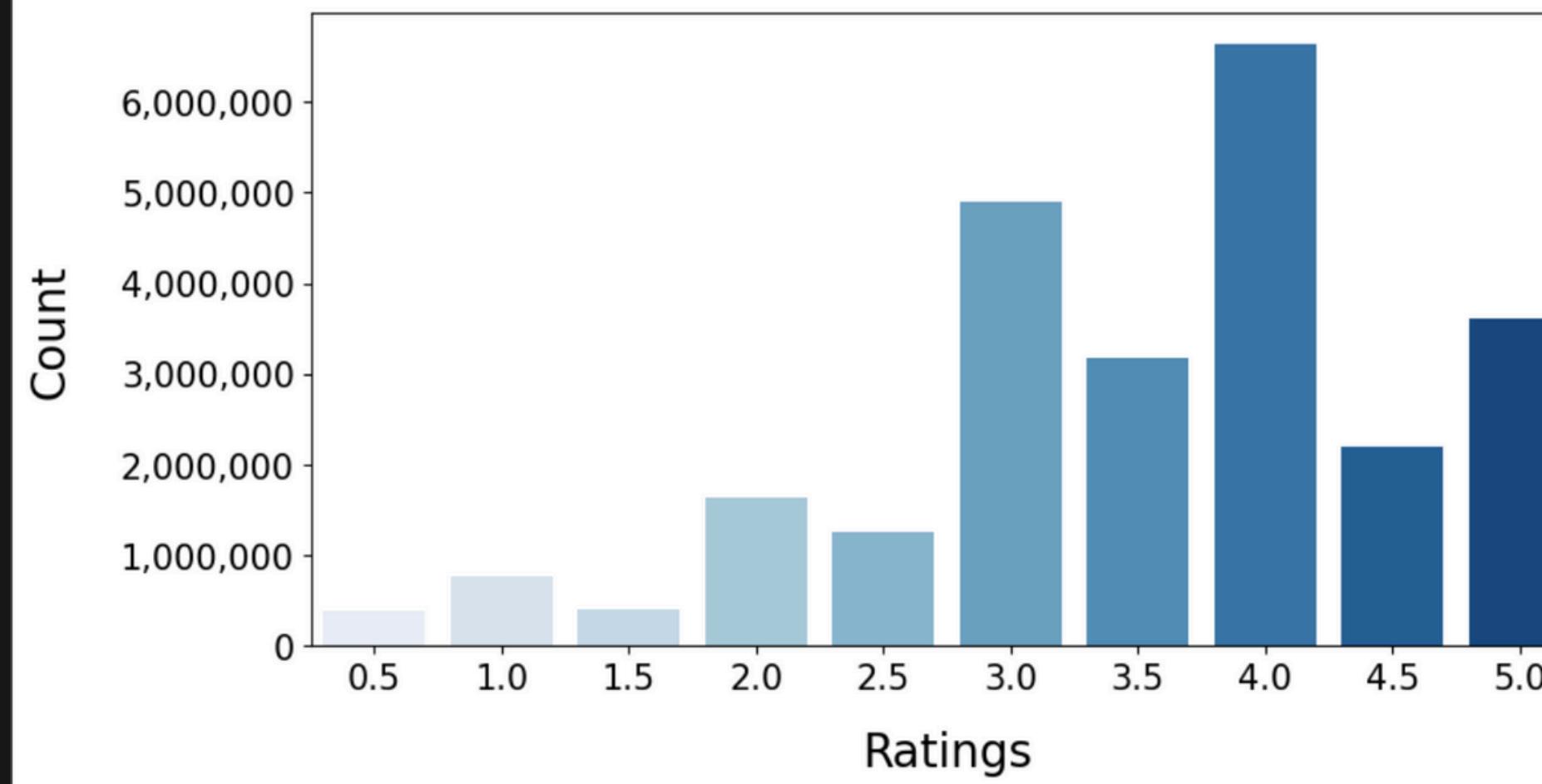


Tahapan Alur Penelitian



Hasil dan Pembahasan





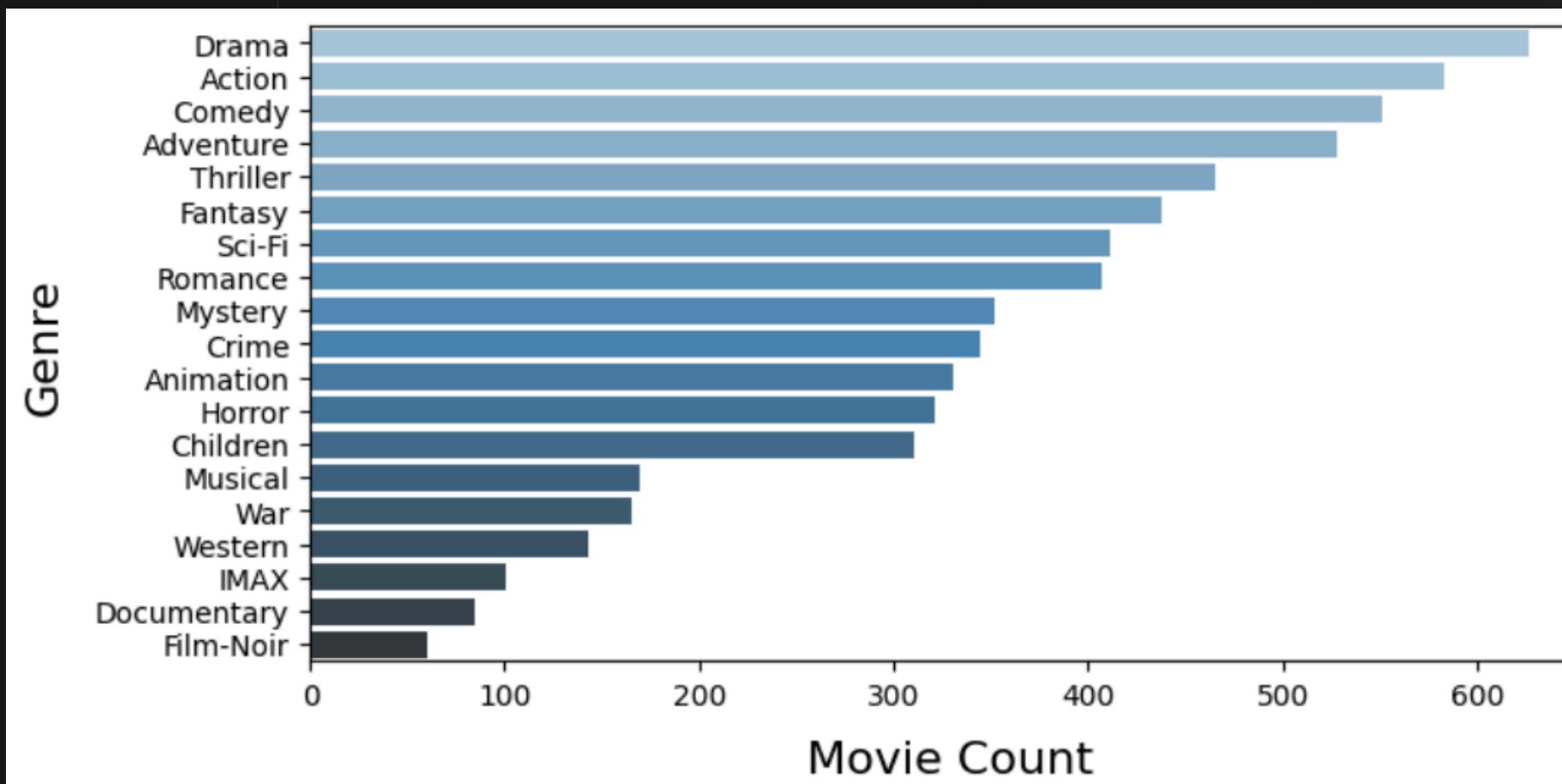
Visualisasi Distribusi Ratings

Dari gambar 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.



Visualisasi Jumlah Genre Pada Film



Gambar tersebut menunjukkan jumlah film berdasarkan genre dalam dataset.

- Genre drama mendominasi dengan lebih dari 600 film, menunjukkan popularitas tinggi dari genre lainnya.
- Genre dengan jumlah film terendah yaitu film-noir, dengan kurang dari 100 film, menunjukkan bahwa genre ini adalah yang paling jarang ditemukan dalam dataset



Evaluasi RMSE

NO	Metode	RMSE
1	ALS Model 1	0.811671
2	ALS Model 2	0.839577

Disimpulkan dari 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.



UserId	Movield	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

Perbandingan Actual dan Predicted Rating

Dapat dilihat bahwa nilai prediksi untuk rating film berkisar antara 3.128 hingga 3.997, yang cukup dekat dengan rating aktual yang diberikan oleh pengguna.



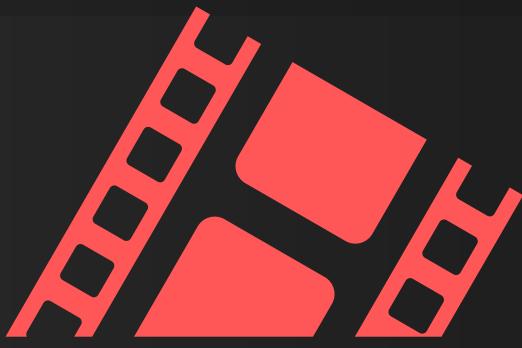
Hasil mesin sistem rekomendasi film untuk pengguna dengan ID 1 berdasarkan prediksi tertinggi.

- 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.

Kesimpulan

Implementasi sistem rekomendasi film menggunakan algoritma *Alternating Least Squares* (ALS) dari Apache Spark MLlib terbukti efektif dengan RMSE terbaik sebesar 0.811671. Model ini menghasilkan prediksi rating film yang akurat dan relevan dengan preferensi pengguna, mencakup berbagai genre. Teknologi big data Apache Spark memungkinkan pemrosesan data skala besar secara terdistribusi, dan algoritma ALS menunjukkan peran penting teknologi big data dalam menghasilkan sistem rekomendasi berkualitas tinggi.





TERIMA KASIH