

Recommendation System Using Knime

Business Understanding

Tujuan nya dibuat recommendation system ini adalah untuk memberikan rekomendasi film pilihan yang kira-kira akan disukai oleh pengguna berdasarkan data yang diperoleh dari Movie Lens. Dengan menggunakan sistem ini, pengguna akan lebih tertarik dengan personalized experience yang ditawarkan.

Data Understanding

Terdapat berbagai data yang diberikan oleh movie lens, yaitu

- Links
- Movies
- Rating
- Tags

Disini kita akan menggunakan data Movies yang berisi informasi film yang terdapat pada Movie Lens dan juga data Rating yang berisi riwayat rating yang diberikan oleh seluruh pengguna

Deskripsi Data Movies:

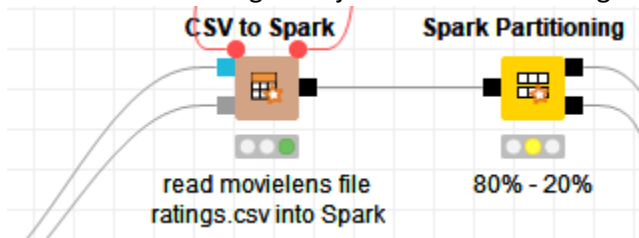
- File type: CSV
- File size: 1.33 MB
- Column : 3 (MovieId, Title, Genres)
- Row : 27,278

Deskripsi Data Ratings:

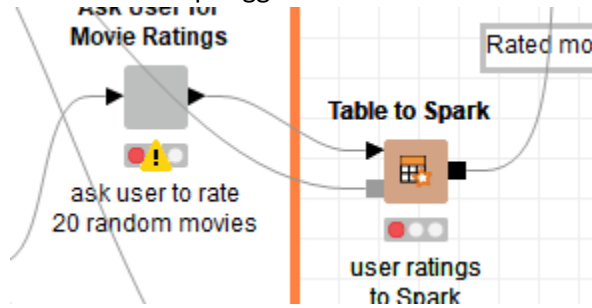
- File type: CSV
- File size: 508 MB
- Column: 4 (userId, MovieId, rating, timestamp)
- Row : 20,000,264

Data Preparation

1. Membaca data ratings menjadi table dan membagi training dan test set



2. Membuat data pengguna baru beserta informasi fil yang disukai



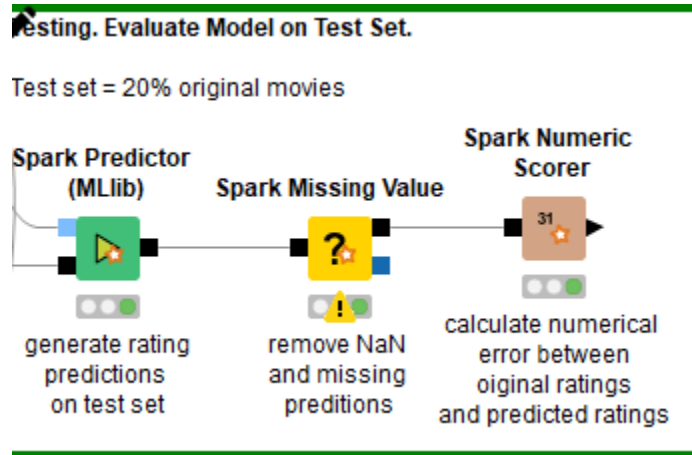
Data Modeling

Data modeling akan dilakukan dengan menggunakan metode collaborative filtering yang terdapat pada library spark MLlib, dengan menggunakan data training set yang telah digabung dengan data pengguna baru yang kita buat.

Row ID	I userId	I movieId	D rating	I timestamp	D Rating
Row0	232	2	3.5	1162148886	3.938
Row1	259	2	5	847280909	2.951
Row2	296	2	5	833870907	4.217
Row3	315	2	4	974695478	2.71
Row4	326	2	3	850438703	2.699
Row5	343	2	3	1088352634	2.634
Row6	427	2	0.5	1233531895	1.874
Row7	435	2	3	845455880	3.382
Row8	452	2	4	830833518	3.908
Row9	482	2	3.5	1375893470	3.816
Row10	536	2	3	944942962	3.22
Row11	546	2	3	1124362166	2.819
Row12	706	2	5	945898354	3.766
Row13	721	2	3	834359305	3.623
Row14	739	2	2.5	1214672577	3.183
Row15	766	2	4	848563304	4.099
Row16	827	2	4	1240559520	3.53
Row17	932	2	4	1173142496	2.645
Row18	982	2	2	1141510578	2.585
Row19	1002	2	4	1011472577	2.062

Figure 1 contoh hasil prediksi

Model Evaluation

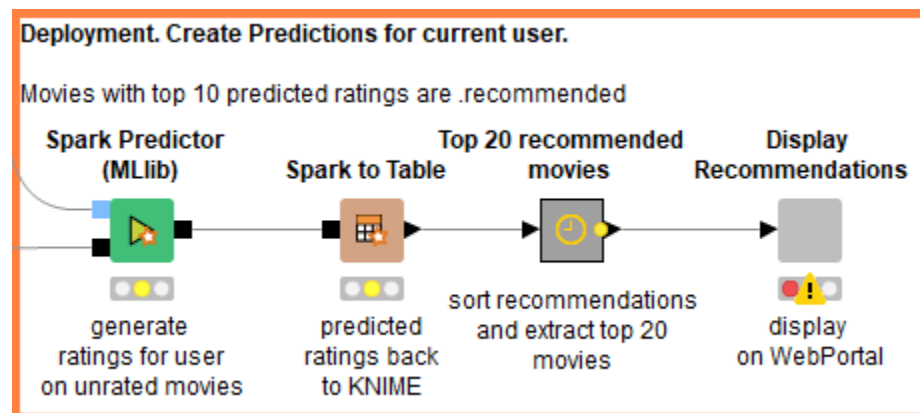


Dari model yang sudah di training, akan digunakan untuk memprediksi test set dan dibandingkan dengan rating aslinya. Namun perlu ada proses tambahan untuk menghilangkan data yang memiliki rating kosong.

Row ID	Prediction
R^2	0.415
mean absolute error	0.615
mean squared error	0.648
root mean squared deviation	0.805
mean signed difference	-0.072

Bisa dilihat dari gambar diatas merupakan hasil evaluasi dari model yang ada, kita bisa menyesuaikan variable saat training model nya dan mungkin mendapat hasil yang lebih baik atau lebih buruk.

Deployment



Untuk simulasi deployment dari model yang kita miliki, kita akan menggunakan data pengguna yang kita buat untuk dimasukkan ke dalam model dan mengambil 20 film dengan prediksi rating paling tinggi direkomendasikan ke pengguna.

Movie Recommendations for You

Based on your movie preferences, we recommend you the following movies.

Show10▼entries

Search:

	title⇅	genres⇅
	Pekka ja Päättökä puistotähteinä (1955)	Comedy
	Wrong Cops (2013)	Comedy Crime
	Birdemic: Shock and Terror (2010)	Romance Thriller
	Edison Kinetoscopic Record of a Sneeze (1894)	Documentary
	Rendez-vous (1985)	Drama
	Nã (1998)	Drama
	Naked Fear (2007)	Horror Thriller
	Palermo Shooting (2008)	Drama
	Two Days (2003)	Comedy Drama Thriller
	Thousand Clouds of Peace, A (Mil nubes de paz cercan el cielo, amor, jamás acabarás de ser amor) (2003)	Drama Romance

Showing 1 to 10 of 10 entries

Previous

1

Next