

Rekomendasi Data Toko Buku

Laporan Tugas Besar

Kelas MK Penambangan Data (CII4I3)

1301184245

Muhammad Raihan Muhith



Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2021

Daftar isi

1. Formulasi Masalah	3
2. Eksplorasi Data.....	3
3. Preprocessing dan Pemodelan	4
3.1. Preprocessing Berdasarkan Kolom Order	5
3.2. Pemodelan Rekomendasi Berdasarkan Kolom Order	6
3.3. Preprocessing Berdasarkan Kolom Basket	7
3.4. Pemodelan Rekomendasi Berdasarkan Kolom Basket	8
3.5. Preprocessing Berdasarkan Kolom Click.....	9
3.6. Pemodelan Berdasarkan Kolom Click	9
4. Hasil dan Kesimpulan.....	11

1. Formulasi Masalah

Tujuan dari pengerjaan tugas ini adalah pembuatan model rekomendasi berdasarkan riwayat transaksi dan karakteristik item. Untuk produk apapun, model diminta untuk menghasilkan lima rekomendasi teratasnya. Dalam menunjang pembuatan model yang diusulkan, mahasiswa diberikan data riwayat transaksi, dan data deskripsi produk dalam bentuk file teks terstruktur (.csv). Data disediakan dalam tiga file individual. Satu file berisi Riwayat transaksi (“transactions.csv”), satu file data deskriptif (“items.csv”) dan file terakhir (“evaluation.csv”) berisi template untuk pengiriman hasil. Namun pada pengerjaan ini, saya tidak menggunakan file terakhir dikarenakan hasil rekomendasi yang dihasilkan oleh model tidak sesuai yang diharapkan.

Pendekatan yang digunakan dalam pengerjaan ini adalah dengan analisis asosiasi. Analisis asosiasi merupakan sebuah metodologi untuk mencari relasi istimewa/menarik yang tersembunyi dalam himpunan data (data set) yang besar. Relasi yang tersembunyi ini dapat direpresentasikan dalam bentuk aturan asosiasi (association rules) atau himpunan barang yang seringkali muncul (frequent itemset). Garis utama dari ide pengerjaan ini adalah dengan membuat model rekomendasi hanya dari data riwayat transaksi saja, sedangkan data deskripsi item hanya digunakan untuk data *lookup value* dari hasil rekomendasi.

2. Eksplorasi Data

Mengenali dataset merupakan tahap awal proses penambangan data. Tahap ini merupakan hal yang penting bagi seorang *data miner* sebelum melakukan proses pemodelan. Hal pertama yang saya lakukan adalah mengecek dataset itu sendiri dari segi bentuk, ukuran, jumlah kolom, deskripsi kolom, dan *missing value*. Berikut ditampilkan bentuk, ukuran, dan jumlah kolom dari dataset pada Gambar 1.

	sessionID	itemID	click	basket	order
0	0	21310	1	0	0
1	1	73018	1	0	0
2	2	19194	1	0	0
3	3	40250	1	0	0
4	4	46107	1	0	0
...
365138	279351	70183	1	0	0
365139	279352	39716	1	0	0
365140	279353	35260	1	0	0
365141	279353	18805	4	0	0
365142	279354	3083	1	0	0
365143 rows × 5 columns					

Gambar 1. Dataset

Dapat dilihat dari Gambar 1., dataset yang digunakan memiliki jumlah baris sebanyak 365143 baris dan jumlah kolom sebanyak 5 kolom dengan deskripsi sebagai berikut:

Nama Kolom	Deskripsi Kolom
sessionID	Kode unik dari setiap sesi transaksi
itemID	Kode unik dari setiap produk
click	Jumlah klik dari produk dalam sebuah sesi transaksi
basket	Kuantitas produk yang ditambahkan ke keranjang dalam sebuah sesi transaksi
order	Jumlah order dari produk dalam sebuah sesi transaksi

Sedangkan upaya yang dilakukan untuk mengecek missing value dari dataset tersebut adalah tentu saja dengan mengecek eksistensi nilai NaN dan dilanjutkan dengan mem-*filter* baris yang memiliki nilai 0 secara bersamaan pada kolom click, basket, dan order. Dari hasil proses tersebut, tidak ditemukan *missing value* dari dataset, sehingga dapat disimpulkan bahwa dataset ini terhindar dari noise.

Ide acuan pemberian rekomendasi yang saya gunakan adalah dengan membedakan sesi transaksi berdasarkan kolom order, basket, dan klik. Pada proses ini, saya menetapkan bahwa pemodelan rekomendasi berdasarkan sesi transaksi dengan kolom order dilakukan pertama, sedangkan sesi transaksi dengan kolom basket dilakukan kedua, dan dilanjutkan sesi transaksi dengan kolom click. Hal ini dilatar belakangi oleh bobot nilai kolom dari suatu item apakah item tersebut hanya diklik, dimasukkan ke keranjang belanja, atau dilanjutkan ke pembelian. Contoh kasusnya adalah jika terdapat dua item atau lebih yang di order, maka pemilik toko akan lebih merekomendasikan barang tersebut dibandingkan dengan yang hanya dimasukkan kekeranjang atau bahkan yang hanya diklik saja.

3. Preprocessing dan Pemodelan

Pada tahap ini, dilakukan penyaringan dan transformasi data sehingga dataset dapat diteruskan ke tahap pemodelan. Pada tahap pemodelan, saya menggunakan algoritma apriori dalam upaya menghasilkan *frequent itemset* dari data transaksi. algoritma ini merupakan algoritma yang cocok digunakan untuk melakukan analisis asosiasi terhadap data yang besar, serta metode ini lebih sederhana dan lebih mudah dipahami struktur kerja dan implementasinya. Seperti yang sudah dijelaskan sebelumnya, proses pengerjaan pemberian rekomendasi ini dibagi menjadi tiga bagian, sehingga proses preprocessing dan pemodelan dilakukan sebanyak tiga kali. Berikut ditampilkan ilustrasi dataset yang telah dibedakan menjadi tiga bagian pada Gambar 2.

sessionID	itemID	order	sessionID	itemID	basket	sessionID	itemID	click		
0	0	21310	0	0	21310	0	0	21310	1	
1	1	73018	1	1	73018	0	1	73018	1	
2	2	19194	2	2	19194	0	2	19194	1	
3	3	40250	3	3	40250	0	3	40250	1	
4	4	46107	4	4	46107	0	4	46107	1	
...	
365138	279351	70183	365138	279351	70183	0	365138	279351	70183	1
365139	279352	39716	365139	279352	39716	0	365139	279352	39716	1
365140	279353	35260	365140	279353	35260	0	365140	279353	35260	1
365141	279353	18805	365141	279353	18805	0	365141	279353	18805	4
365142	279354	3083	365142	279354	3083	0	365142	279354	3083	1
365143 rows × 3 columns			365143 rows × 3 columns			365143 rows × 3 columns				

Gambar 2. Pembagian Dataset

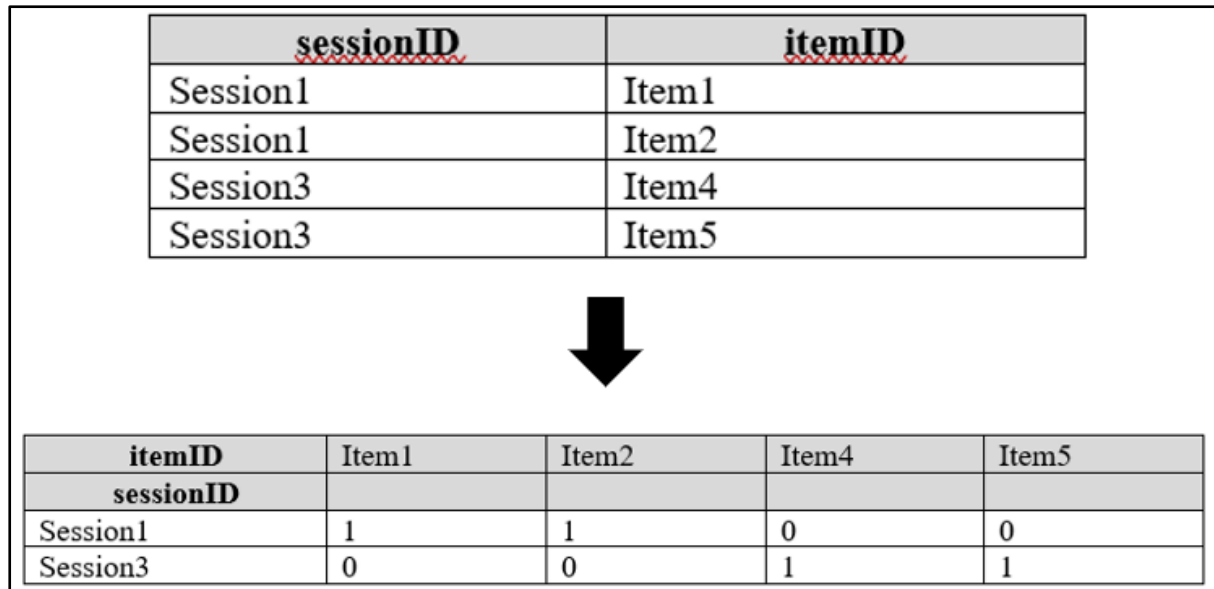
3.1. Preprocessing Berdasarkan Kolom Order

Langkah pertama yang dilakukan yaitu menyaring data yang memiliki nilai order lebih besar dari sama dengan satu. Hal ini dilakukan karena terdapat nilai nol pada kolom tersebut yang menandakan bahwa pada sesi transaksi tersebut, item yang bersangkutan tidak diorder. Setelah data tersaring, saya mendapati bahwa terdapat dua atau lebih sesi transaksi yang sama dengan produk yang berbeda, artinya pada dataset terdapat multitransaksi. Sedangkan sistem rekomendasi membutuhkan data dengan multitransaksi, Sehingga saya melakukan penyaringan lagi dengan mengambil data dengan sesi transaksi yang lebih dari satu. Berikut ditunjukkan ilustrasi penyaringan data pada Gambar 3.

Sebelum Penyaringan	
sessionID	itemID
Session1	Item1
Session1	Item2
Session2	Item3
Session3	Item4
Session3	Item5
<div> <div> Multitransaksi, karena jumlah sesi transaksi dengan id 'Session1' lebih dari 1 </div> <div> Multitransaksi, karena jumlah sesi transaksi dengan id 'Session3' lebih dari 1 </div> </div>	
Setelah Penyaringan :	
sessionID	itemID
Session1	Item1
Session1	Item2
Session3	Item4
Session3	Item5

Gambar 3. Ilustrasi penyaringan data multitransaksi

Setelah proses penyaringan selesai, dataset perlu di transformasikan ke dalam bentuk matriks ketetangaan antara sessionID dan itemID karena dalam pengerjaan tugas ini, saya menggunakan package yang disediakan oleh mlxtend.frequent_pattern dimana inputan yang dapat dibaca oleh package tersebut merupakan data dalam bentuk matriks ketetangaan. Untuk dapat memahami proses transformasi tersebut, lihat Gambar 4.



Gambar 4. ilustrasi transofrmasi data

Pada gambar diatas, angka 1 dan 0 menunjukkan eksistensi dari setiap produk didalam sebuah sesi transaksi. Contohnya, Session1 memiliki nilai 1 pada Item1 dan Item2, sedangkan Session3 memiliki nilai 1 pada Item4 dan Item5.

3.2. Pemodelan Rekomendasi Berdasarkan Kolom Order

Pada tahap ini, saya menggunakan algoritma apriori yang disediakan oleh package mlxtend.frequent_pattern dalam menghasilkan *frequent itemset* berserta rulesnya. Nilai minimal support yang ditetapkan yaitu sebesar 0,001 , hal ini dilakukan karena dataset sangat bervariasi sehingga perbandingan terjadinya sebuah itemset terhadap jumlah seluruh itemset pada dataset sangatlah kecil. Berikut merupakan *frequent itemset* yang dihasilkan.

support	itemsets
0 0.001394	(125)
1 0.001394	(194)
2 0.001859	(245)
3 0.001394	(616)
4 0.001394	(685)
...	...
1030 0.001394	(27041, 14093, 60430, 69073, 47221)
1031 0.001394	(27041, 60430, 69073, 47221, 26719)
1032 0.001394	(8960, 19458, 36098, 4626, 61335, 48856)
1033 0.001859	(8960, 36098, 23654, 4626, 61335, 48856)
1034 0.001394	(27041, 14093, 60430, 69073, 47221, 26719)

1035 rows × 2 columns

Gambar 5. *frequent itemset* berdasarkan kolom order

Setelah *frequent itemset* didapatkan, maka dilanjutkan ke proses pembuatan rules dengan minimum confidence sebesar 0,1 , hal ini dilakukan karena frekuensi munculnya item-item dalam sebuah itemset dalam transaksi sangatlah sedikit, akibatnya guna mendapatkan lima rekomendasi dari sebuah itemset sekaligus, dibutuhkan batas minimal confidence yang sangat kecil. Rules yang didapat ditunjukkan pada Gambar 6.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(125)	(48678)	0.001394	0.001394	0.001394	1.000000	717.333333	0.001392	inf
1	(48678)	(125)	0.001394	0.001394	0.001394	1.000000	717.333333	0.001392	inf
2	(29485)	(782)	0.001859	0.001859	0.001394	0.750000	403.500000	0.001391	3.992565
3	(782)	(29485)	0.001859	0.001859	0.001394	0.750000	403.500000	0.001391	3.992565
4	(41582)	(782)	0.001859	0.001859	0.001859	1.000000	538.000000	0.001855	inf
...
2955	(14093)	(27041, 60430, 69073, 47221, 26719)	0.007435	0.001394	0.001394	0.187500	134.500000	0.001384	1.229053
2956	(60430)	(27041, 14093, 69073, 47221, 26719)	0.006970	0.001859	0.001394	0.200000	107.600000	0.001381	1.247677
2957	(69073)	(27041, 14093, 60430, 47221, 26719)	0.006506	0.001394	0.001394	0.214286	153.714286	0.001385	1.270953
2958	(47221)	(27041, 14093, 60430, 69073, 26719)	0.007900	0.001394	0.001394	0.176471	126.588235	0.001383	1.212593
2959	(26719)	(27041, 14093, 60430, 69073, 47221)	0.006970	0.001394	0.001394	0.200000	143.466667	0.001384	1.248257

2960 rows x 9 columns

Gambar 6. Rules berdasarkan kolom order

Dari rules yang sudah dihasilkan, saya memfilter data dengan jumlah consequents sebanyak 5, setelah itu saya mendapati bahwa terdapat dua atau lebih antecedents yang sama dengan value consequents yang berbeda, hal ini saya tangani dengan menghapus data duplikasi tersebut yang memiliki nilai lift yang lebih kecil. Kolom lift menunjukkan adanya tingkat kekuatan rule atas kejadian acak dari antecedent dan consequent berdasarkan pada supportnya masing-masing. Berikut ditunjukkan hasil rekomendasi berdasarkan kolom order pada gambar 7.

	itemID	recomendations
0	8960	19458, 36098, 4626, 61335, 48856
1	19458	8960, 36098, 4626, 61335, 48856
2	36098	8960, 19458, 4626, 61335, 48856
3	4626	8960, 19458, 36098, 61335, 48856
4	61335	8960, 19458, 36098, 4626, 48856
5	48856	8960, 19458, 36098, 4626, 61335
6	23654	8960, 36098, 4626, 61335, 48856
7	27041	14093, 60430, 69073, 47221, 26719
8	14093	27041, 60430, 69073, 47221, 26719
9	60430	27041, 14093, 69073, 47221, 26719
10	69073	27041, 14093, 60430, 47221, 26719
11	47221	27041, 14093, 60430, 69073, 26719
12	26719	27041, 14093, 60430, 69073, 47221

Gambar 7. Hasil rekomendasi berdasarkan kolom order

3.3. Preprocessing Berdasarkan Kolom Basket

Setelah mendapatkan rekomendasi item berdasarkan kolom order, langkah selanjutnya membuat model rekomendasi berdasarkan kolom basket. Untuk tahap preprocessingnya sendiri tidak jauh berbeda dari tahap preprocessing sebelumnya, hanya saja diawal tahap ini, saya memfilter data dengan itemID yang tidak ada di itemID hasil rekomendasi berdasarkan kolom

order, dengan harapan antecedents yang dihasilkan oleh model ini, tidak sama dengan antecedents yang dihasilkan oleh model rekomendasi berdasarkan kolom order.

3.4. Pemodelan Rekomendasi Berdasarkan Kolom Basket

Pemodelan ini juga dilakukan dengan metode yang sama seperti pada pemodelan sebelumnya, yaitu dengan algoritma apriori. Namun, nilai minimal support yang digunakan untuk menghasilkan *frequent itemset* adalah 0,0013, hal ini dikarenakan ketika menggunakan nilai yang sama seperti sebelumnya atau lebih kecil, media penyimpanan sementara (RAM) yang dimiliki oleh *device* saya tidak memumpuni. Karena pada dasarnya proses menggenerasi *frequent itemset* merupakan tahapan yang berat dari sudut pandang komputasi. Salah satu contoh error yang dijumpai saat melakukan running dengan nilai minimum support yang lebih kecil dari yang sudah ditentukan ditampilkan pada bagian lampiran. Berikut *frequent itemset* yang dihasilkan berdasarkan kolom basket.

	support	itemsets
0	0.001402	(283)
1	0.001823	(616)
2	0.001542	(782)
3	0.004627	(1193)
4	0.005889	(1607)
...
892	0.001402	(62913, 58851, 11525, 1607, 57800, 74654, 14207)
893	0.001402	(62913, 58851, 11525, 1607, 57800, 42353, 74654)
894	0.001402	(62913, 58851, 1607, 57800, 42353, 74654, 14207)
895	0.001402	(62913, 58851, 11525, 57800, 42353, 74654, 14207)
896	0.001402	(62913, 58851, 11525, 1607, 57800, 42353, 7465...

897 rows × 2 columns

Gambar 8. *frequent itemset* berdasarkan kolom basket

Nilai confidence yang digunakan untuk menghasilkan rules juga sama yaitu 0,1. Berikut merupakan rules yang dihasilkan berdasarkan kolom basket.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(29485)	(782)	0.001823	0.001542	0.001402	0.769231	498.741259	0.001399	4.326650
1	(782)	(29485)	0.001542	0.001823	0.001402	0.909091	498.741259	0.001399	10.979950
2	(1193)	(74094)	0.004627	0.006450	0.001823	0.393939	61.077734	0.001793	1.639358
3	(74094)	(1193)	0.006450	0.004627	0.001823	0.282609	61.077734	0.001793	1.387490
4	(11525)	(1607)	0.005048	0.005889	0.001963	0.388889	66.037037	0.001933	1.626727
...
7031	(1607)	(62913, 58851, 11525, 57800, 42353, 74654, 14207)	0.005889	0.001402	0.001402	0.238095	169.809524	0.001394	1.310660
7032	(57800)	(62913, 58851, 11525, 1607, 42353, 74654, 14207)	0.004487	0.001402	0.001402	0.312500	222.875000	0.001396	1.452506
7033	(42353)	(62913, 58851, 11525, 1607, 57800, 74654, 14207)	0.005889	0.001402	0.001402	0.238095	169.809524	0.001394	1.310660
7034	(74654)	(62913, 58851, 11525, 1607, 57800, 42353, 14207)	0.004627	0.001402	0.001402	0.303030	216.121212	0.001396	1.432771
7035	(14207)	(62913, 58851, 11525, 1607, 57800, 42353, 74654)	0.004487	0.001402	0.001402	0.312500	222.875000	0.001396	1.452506

7036 rows × 9 columns

Gambar 9. Rules berdasarkan kolom basket

Tahapan setelah peng-*generate*an rules juga sama seperti sebelumnya yaitu, memfilter data dengan jumlah consequents sebanyak 5, dan dilanjutkan dengan menghapus data duplikasi yang memiliki nilai lift yang lebih kecil. Berikut rekomendasi yang dihasilkan berdasarkan kolom basket.

	itemID	recomendations
0	57800	62913, 58851, 42353, 74654, 14207
1	14207	58851, 11525, 1607, 42353, 74654
2	74654	62913, 11525, 1607, 57800, 14207
3	11525	62913, 58851, 57800, 42353, 14207
4	58851	62913, 1607, 57800, 42353, 74654
5	62913	58851, 11525, 57800, 42353, 14207
6	1607	62913, 11525, 42353, 74654, 14207
7	42353	62913, 58851, 11525, 1607, 57800

Gambar 10. Rekomendasi berdasarkan kolom basket

3.5. Preprocessing Berdasarkan Kolom Click

Setelah mendapatkan rekomendasi berdasarkan kolom order dan basket, selanjutnya dilakukan pembuatan model rekomendasi berdasarkan kolom click. Tahap preprocessing dataset ini juga tidak jauh berbeda dari tahap preprocessing sebelumnya, hanya saja diawal tahap ini, saya memfilter data dengan itemID yang tidak ada di itemID hasil rekomendasi berdasarkan kolom order dan basket, dengan harapan antecedents yang dihasilkan oleh model ini, tidak sama dengan antecedents yang dihasilkan oleh model rekomendasi berdasarkan kolom-kolom sebelumnya.

3.6. Pemodelan Berdasarkan Kolom Click

Tentu saja pemodelan ini juga menggunakan metode yang sama seperti sebelumnya, namun nilai minimal support yang digunakan untuk menghasilkan *frequent itemset* oleh model ini adalah 0.0022, hal ini didasari oleh alasan yang sama seperti di tahapan pemodelan berdasarkan kolom basket. Berikut *frequent itemset* yang dihasilkan berdasarkan kolom click.

	support	itemssets
0	0.002345	(616)
1	0.002028	(1142)
2	0.002535	(1193)
3	0.004753	(1713)
4	0.002028	(1969)
...
272	0.003137	(32266, 5291, 29711)
273	0.002091	(25451, 33908, 31995)
274	0.002471	(25451, 72468, 31995)
275	0.002250	(33908, 25451, 72468)
276	0.002281	(33908, 31995, 72468)
277 rows × 2 columns		

Gambar 11. *frequent itemset* berdasarkan kolom click

Nilai confidence yang digunakan untuk menghasilkan rules juga sama seperti tahap sebelumnya yaitu 0,1. Berikut merupakan rules yang dihasilkan berdasarkan kolom basket.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(1713)	(36071)	0.004753	0.007890	0.003200	0.673333	85.342972	0.003163	3.037072
1	(36071)	(1713)	0.007890	0.004753	0.003200	0.405622	85.342972	0.003163	1.674436
2	(2253)	(11743)	0.011597	0.007129	0.004753	0.409836	57.486339	0.004670	1.682364
3	(11743)	(2253)	0.007129	0.011597	0.004753	0.666667	57.486339	0.004670	2.965209
4	(2253)	(17727)	0.011597	0.012833	0.002376	0.204918	15.968427	0.002228	1.241592
...
119	(33908, 72468)	(31995)	0.002757	0.006812	0.002281	0.827586	121.481957	0.002263	5.760488
120	(31995, 72468)	(33908)	0.003802	0.003739	0.002281	0.600000	160.474576	0.002267	2.490653
121	(33908)	(31995, 72468)	0.003739	0.003802	0.002281	0.610169	160.474576	0.002267	2.555464
122	(31995)	(33908, 72468)	0.006812	0.002757	0.002281	0.334884	121.481957	0.002263	1.499352
123	(72468)	(31995, 33908)	0.005862	0.002725	0.002281	0.389189	142.823382	0.002265	1.632707

124 rows x 9 columns

Gambar 12. Rules berdasarkan kolom click

Dapat dilihat pada Gambar 12. Rules yang dihasilkan oleh model ini lebih sedikit dibanding dengan rules-rules sebelumnya, sehingga saya memutuskan untuk menyaring antecedents yang setidaknya memiliki dua rekomendasi, dan dilanjutkan dengan menghapus data duplikasi yang memiliki nilai lift yang lebih kecil. Berikut rekomendasi yang dihasilkan berdasarkan kolom click.

	itemID	recomendations
0	25451	33908, 72468
1	33908	25451, 72468
2	72468	25451, 33908
3	12723	5226, 50523
4	29711	32266, 5291
5	31995	25451, 33908
6	50523	5226, 12723
7	32266	5291, 29711
8	5226	50523, 12723
9	11743	2253, 46998
10	5291	32266, 29711
11	2253	46998, 11743
12	46998	2253, 11743

Gambar 13. Rekomendasi berdasarkan kolom click

4. Hasil dan Kesimpulan

Setelah mendapatkan hasil rekomendasi berdasarkan ketiga kolom, yaitu order, basket, dan click, ketiga hasil tersebut digabungkan, ditransformasikan, dan nilai itemID diubah menjadi judul buku dari itemID tersebut yang didapat dari dataset deskripsi item. Berikut merupakan hasil akhir dari proses pengerjaan tugas ini.

	Items	rec1	rec2	rec3	rec4	rec5
0	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 6 und der Halbblutprinz	Harry Potter 5 und der Orden des Phönix	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 7 und die Heiligtümer des Todes
1	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 6 und der Halbblutprinz	Harry Potter 5 und der Orden des Phönix	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 7 und die Heiligtümer des Todes
2	Harry Potter 6 und der Halbblutprinz	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 5 und der Orden des Phönix	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 7 und die Heiligtümer des Todes
3	Harry Potter 5 und der Orden des Phönix	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 6 und der Halbblutprinz	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 7 und die Heiligtümer des Todes
4	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 6 und der Halbblutprinz	Harry Potter 5 und der Orden des Phönix	Harry Potter 7 und die Heiligtümer des Todes
5	Harry Potter 7 und die Heiligtümer des Todes	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 4 und der Feuerkelch. Taschenbuch	Harry Potter 6 und der Halbblutprinz	Harry Potter 5 und der Orden des Phönix	Harry Potter 3 und der Gefangene von Askaban
6	Harry Potter 1 und der Stein der Weisen	Harry Potter 2 und die Kammer des Schreckens	Harry Potter 6 und der Halbblutprinz	Harry Potter 5 und der Orden des Phönix	Harry Potter 3 und der Gefangene von Askaban	Harry Potter 7 und die Heiligtümer des Todes
7	Throne of Glass 5 - Die Sturmbezwingerin	Throne of Glass 06 - Der verwundete Krieger	Throne of Glass 2 - Kriegerin im Schatten	Throne of Glass 4 - Königin der Finsternis	Throne of Glass 7 - Herrscherin über Asche und...	Throne of Glass 3 - Erbin des Feuers
8	Throne of Glass 06 - Der verwundete Krieger	Throne of Glass 5 - Die Sturmbezwingerin	Throne of Glass 2 - Kriegerin im Schatten	Throne of Glass 4 - Königin der Finsternis	Throne of Glass 7 - Herrscherin über Asche und...	Throne of Glass 3 - Erbin des Feuers
.						
.						
.						
22	Die Chroniken von Rotkäppchen - Allein im tief...	Die Chroniken der Meerjungfrau - Der Fluch der...	Die Chroniken von Peter Pan - Albraum im Nimm...	NaN	NaN	NaN
23	Die Chroniken von Peter Pan - Albraum im Nimm...	Die Chroniken der Meerjungfrau - Der Fluch der...	Die Chroniken von Rotkäppchen - Allein im tief...	NaN	NaN	NaN
24	Der Palast im Himmel	Das Haus der tausend Räume	Das wandelnde Schloss	NaN	NaN	NaN
25	Sister of the Moon	Sister of the Stars	Sister of the Night	NaN	NaN	NaN
26	Die Chroniken von Alice - Dunkelheit im Spiege...	Die Chroniken der Meerjungfrau - Der Fluch der...	Die Chroniken von Rotkäppchen - Allein im tief...	NaN	NaN	NaN
27	Das wandelnde Schloss	Das Haus der tausend Räume	Der Palast im Himmel	NaN	NaN	NaN
28	Sister of the Stars	Sister of the Night	Sister of the Moon	NaN	NaN	NaN
29	Das Haus der tausend Räume	Das wandelnde Schloss	Der Palast im Himmel	NaN	NaN	NaN
30	Lodernde Schwingen	Eisige Wellen	Goldene Flammen	NaN	NaN	NaN
31	Sister of the Night	Sister of the Stars	Sister of the Moon	NaN	NaN	NaN
32	Eisige Wellen	Goldene Flammen	Lodernde Schwingen	NaN	NaN	NaN
33	Goldene Flammen	Eisige Wellen	Lodernde Schwingen	NaN	NaN	NaN

Gambar 14. Hasil Rekomendasi Keseluruhan

Dari hasil yang didapat, dapat dikatakan bahwa ide yang digunakan untuk menyelesaikan tugas ini tidak cukup untuk memberikan setiap produk memiliki produk rekomendasinya, dimana total produk yang terdapat pada dataset deskripsi item sangat banyak. Hal ini dikarenakan beberapa faktor, yang pertama tentunya terdapat kemungkinan bahwa produk tidak mempunyai riwayat transaksi atau pada dataset transaksi, produk terfilter karena sesi transaksinya tidak termasuk kedalam kategori multitransaksi, sehingga tidak ada acuan pemberian rekomendasi untuk item tersebut. Lalu faktor yang kedua adalah kurangnya *frequent itemset* yang tergenerasi karena keterbatasan perangkat keras. Ketika suatu transaksi memiliki transaksi yang bervariasi, ditambah dengan jumlah item yang sangat banyak, maka support

yang dihasilkan untuk setiap item dalam menggenerasi *frequent itemset* sangatlah kecil, dan semakin kecil nilai minimal support yang ditentukan, maka semakin besar pula media penyimpanan sementara (RAM) yang diperlukan. Lalu faktor yang terakhir yaitu, untuk mendapatkan consequent sebanyak 5 sekaligus, memerlukan frequent itemset yang banyak, karena frequent itemset yang tergenerasi pada pengerjaan tugas ini terbatas, sehingga hasil rekomendasi item yang memiliki 5 item rekomendasi sekaligus sangat sedikit.

Lampiran

```
-----  
MemoryError                                Traceback (most recent call last)  
~\AppData\Local\Temp\ipykernel_28540\3562207434.py in <module>  
    1 # menggenerasi frequent itemset dengan nilai minimum support = 0.002  
    2  
----> 3 frequent_itemsets = apriori(clicked_transformed_sets, min_support=0.002, use_colnames=True)  
    4 frequent_itemsets  
  
~\AppData\Local\Programs\Python\Python310\lib\site-packages\mlxtend\frequent_patterns\apriori.py in apriori(df, min_s  
    300         _bools = _bools & (X[:, combin[:, n]] == all_ones)  
    301     else:  
--> 302         _bools = np.all(X[:, combin], axis=2)  
    303  
    304         support = _support(np.array(_bools), rows_count, is_sparse)  
  
MemoryError: Unable to allocate 12.3 GiB for an array with shape (26106, 2, 31560) and data type int64
```