stc tv

نظام تقديم التوصيات

بناء نظام توصية تعاوني (Collaborative Filtering) وعرض أعلى 5 توصيات
للمستخدمين الذين شاهدوا فيلم "Moana" (Jawwy Dataset)

تقرير تحليل بيانات وبناء نموذج — مؤسسة مسك / مهارات Skills

أولًا: نظرة عامة على البيانات

1. شكل البيانات

يحتوي ملف البيانات على **1,048,575** صفًا في **5** أعمدة: معرّف المستخدم، اسم البرنامج، تقييم المستخدم للبرنامج (من 1 إلى 4)، تاريخ المشاهدة، ونوع البرنامج (Genre). تغطي البيانات الفترة من **14-03-2017** إلى **30-04-2018**، وتشمل **11,578** مستخدمًا و**8,013** برنامجًا موزعة على **16** نوعًا (Genre).

user_id_maped	program_name	rating	date_	program_genre
26138	100 treets	1	2017-05-27	Drama
7946	Moana	1	2017-05-21	Animation
7418	The Mermaid Princess	1	2017-08-10	Animation
19307	The Mermaid Princess	2	2017-07-26	Animation
15860	Churchill	2	2017-07-07	Biography

2. فحص البيانات الفارغة

Column	Null count
user_id_maped	0
program_name	0
rating	0
date_	0
program_genre	0

لا توجد أي قيم فارغة في البيانات.

3. الوصف الإحصائي

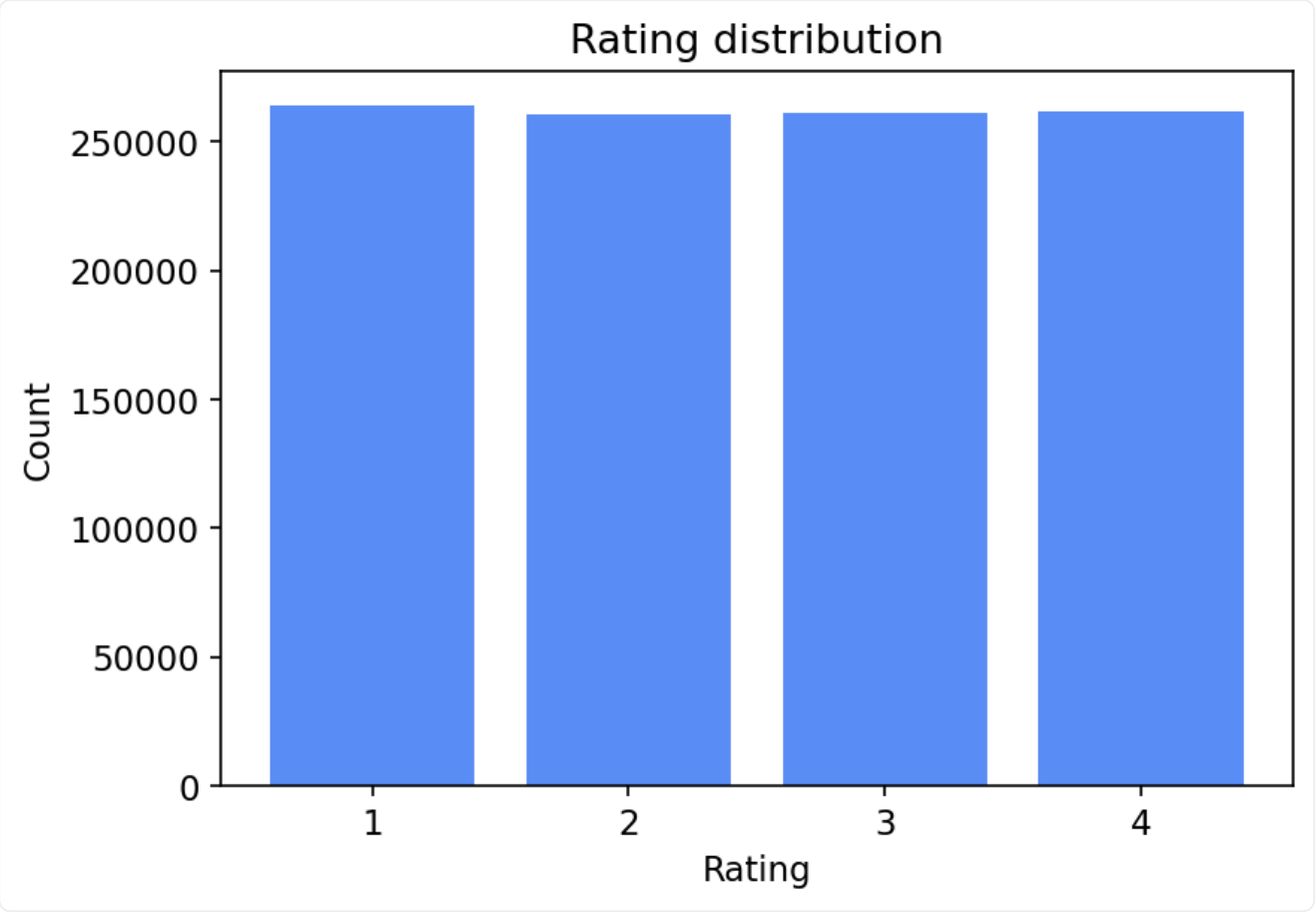
	user_id_maped	rating
count	1048575.00	1048575.00
mean	17092.66	2.50
std	10035.13	1.12

	user_id_maped	rating
min	1.00	1.00
25%	8253.00	1.00
50%	17149.00	2.00
75%	25665.00	3.00
max	34280.00	4.00

4. توزيع التقييمات (Rating)

توفر هذه النسخة من البيانات عمود `rating` جاهزًا (من 1 إلى 4)، وهو مؤشر مُشتق مسبقًا من سلوك المستخدم (مثل مدة المشاهدة) يعكس مدى إعجاب المستخدم بالبرنامج. تم استخدام هذا العمود مباشرة كمقياس "الإعجاب" لبناء نظام التوصية.

rating	count
1	264293
2	260664
3	261505
4	262113



التوزيع متقارب تقريبًا بين القيم الأربع (كل قيمة ≈25% من البيانات)، بدون انحياز واضح نحو تقييم معين.

ثانيًا: تحديد مجموعة البيانات وبناء نظام التوصية

1. معالجة التكرارات

لوحظ أن نفس المستخدم قد يشاهد نفس البرنامج أكثر من مرة (بتواريخ وتقييمات مختلفة قليلًا): عدد أزواج (مستخدم، برنامج) المكررة بلغ **608,338** من إجمالي **1,048,575** صف. تمت معالجة ذلك بحساب **متوسط التقييم** لكل زوج (مستخدم، برنامج)، فأصبحت مجموعة البيانات النهائية للعمل عليها بشكل **440,237** تقييم فريد.

2. بناء النموذج: نظام توصية تعاوني (Item-based Collaborative Filtering)

كما ورد في المصدر الداعم، هناك ثلاث طرق أساسية لأنظمة التوصية: توصية قائمة على المحتوى (Content-based)، توصية تعاونية (Collaborative)، وتوصية مختلطة (Hybrid). تم اختيار نظام **توصية تعاوني قائم على تشابه البرامج**: يحوّل النموذج البيانات إلى مصفوفة متفرقة (Sparse Matrix) بحجم **11,578 مستخدم × 8,013 برنامج** (440,237 تقييم معروف)، ثم يحسب **درجة التشابه (Cosine Similarity)** بين كل زوج من البرامج بناءً على تطابق أنماط تقييم جميع المستخدمين، فينتج عن ذلك مصفوفة تشابه بين البرامج بحجم 8,013×8,013.

هذه هي بالضبط فكرة "التوصية بناءً على مشاهدات وتفضيلات مستخدمين آخرين يشاركون نفس الاهتمامات": فإذا شاهد عدد كبير من المستخدمين برنامجين وأعطوهما تقييمات متقاربة، يعتبر النموذج أن هذين البرنامجين متشابهان، ويستخدم ذلك لتوصية من أحب أحدهما بالآخر.

3. تعميم النموذج: توصيات شخصية لأي مستخدم

إلى جانب التوصية بمشابهات برنامج واحد، تم بناء دالة تجمع توصيات شخصية لمستخدم بعينه بناءً على **كل** ما شاهده وقيّمه. مثال على المستخدم رقم 1:

البرامج التي شاهدها المستخدم:

Program	Rating
Baywatch	2.0
Going in tyle	2.0
Kidnap	2.0
Men In Black	2.5
Pingu	1.0
The Boss Baby	1.0

التوصيات المولّدة له:

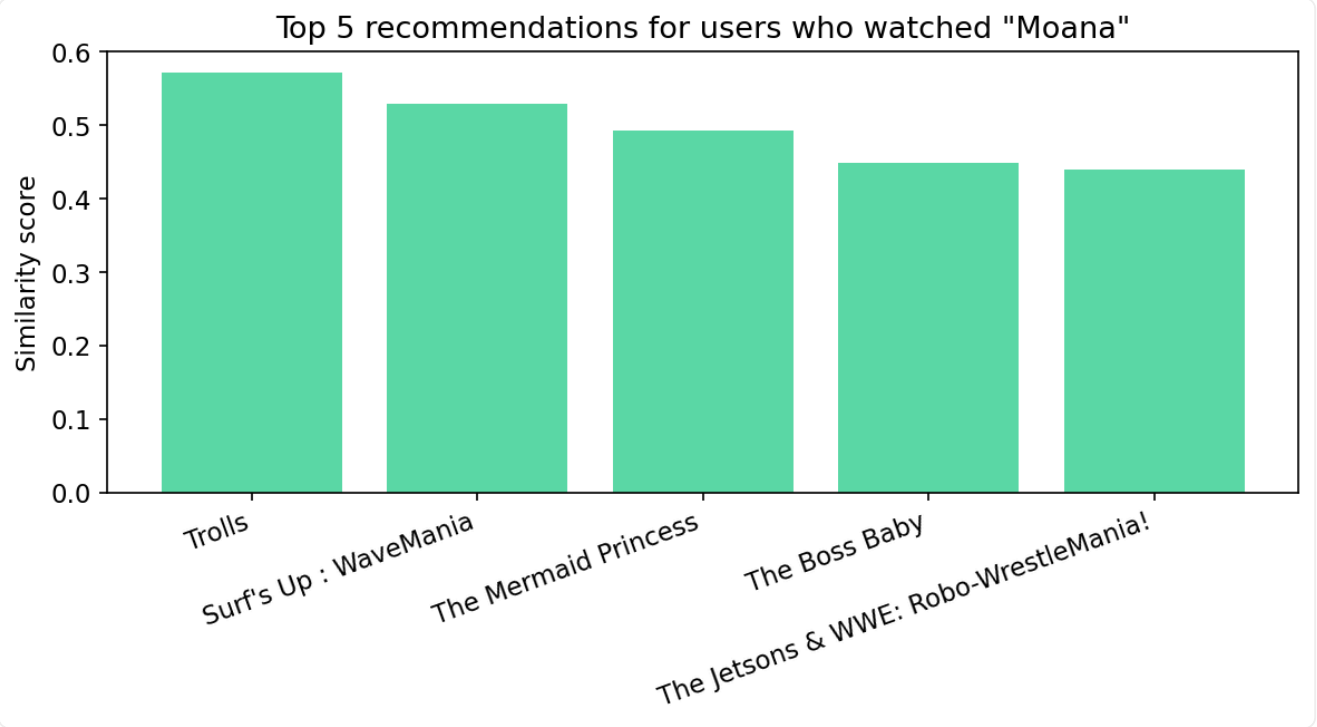
Recommended program	Score	Genre
Inside	2.629	Horror
Alien: Covenant	2.618	Horror
Table	2.515	Comedy
Snatched	2.503	Comedy
Dunkirk	2.494	NOT_DEFINED_IN_UMS

المستخدم شاهد أفلامًا من نوع Action/Comedy (مثل Men In Black وKidnap)، وكانت التوصيات الناتجة منطقية (Alien: Covenant، Snatched) أي ضمن أنواع مشابهة.

ثالثًا: أعلى 5 توصيات لمشاهدي فيلم "Moana"

باستخدام النموذج نفسه، تم استخراج أعلى 5 برامج مشابهة لفيلم Moana بناءً على أنماط مشاهدة وتقييم جميع المستخدمين الآخرين الذين شاهدوه:

Recommended program	Similarity score	Genre
Trolls	0.572	Animation
Surf's Up : WaveMania	0.529	Animation
The Mermaid Princess	0.493	Animation
The Boss Baby	0.449	Animation
The Jetsons & WWE: Robo-WrestleMania!	0.439	Animation



نتائج نظام التوصية

- بُني النموذج على **440,237 تقييم فريد** (مستخدم، برنامج) بعد تلخيص التكرارات، تغطي **11,578 مستخدمًا** و **8,013 برنامجًا**.
- أعلى 5 توصيات لمشاهدي فيلم **Moana** هي جميعها أفلام رسوم متحركة (Animation): Trolls، Surf's Up: WaveMania، The Boss Baby، The Mermaid Princess، وThe Jetsons & WWE: Robo-WrestleMania!، وهذا يتطابق منطقيًا مع نوع وجمهور فيلم Moana نفسه.
- الملفت أن هذا التطابق في النوع (Genre) **لم يُفرض على النموذج مسبقًا**؛ بل استُخرج بالكامل من سلوك المستخدمين الفعلي (من قام بمشاهدة وتقييم كل برنامج)، وهذا يؤكد أن نظام التوصية التعاوني ينتج نتائج منطقية ومتماسكة دون الاعتماد على خصائص المحتوى نفسه.

- الدالة العامة لتوصية أي مستخدم (وليس فقط برنامج واحد) أثبتت أيضًا نتائج منطقية، مما يعني أن النموذج قابل للتطبيق المباشر على كل قاعدة المستخدمين.

رابعًا: الخلاصة والتوصيات

- نظام التوصية التعاوني المبني هنا (Item-based Collaborative Filtering) يعتمد بالكامل على سلوك المستخدمين الفعلي (التقييمات)، وهو أسلوب فعّال ومُستخدم على نطاق واسع في المنصات الكبرى (Netflix، Amazon، YouTube) كما ورد في المصدر الداعم.
- يمكن تحسين النظام مستقبلًا بدمجه مع توصية قائمة على المحتوى (نوع البرنامج، الكلمات المفتاحية) للحصول على **نظام توصية مختلط (Hybrid)** يعالج مشكلة "البداية الباردة" (Cold start) للبرامج أو المستخدمين الجدد الذين لا تتوفر عنهم بيانات تقييم كافية.
- يُنصح بإعادة بناء مصفوفة التشابه بشكل دوري (مثلاً أسبوعيًا) كلما تجمّعت تقييمات جديدة، لضمان مواكبة التوصيات لأحدث اهتمامات المستخدمين.
- يمكن أيضًا تجربة تحسينات إضافية مثل "Mean-centering" (طرح متوسط تقييم كل مستخدم قبل حساب التشابه) لتقليل تأثير اختلاف معايير التقييم بين المستخدمين.

تم إعداد هذا التقرير بالاستعناد إلى تحليل بيانات stc TV Data Set T3 باستخدام Python (pandas, scikit-learn, scipy, matplotlib, plotly) ضمن بيئة Google Colab، وهو يلخّص الكود والنتائج الواردة في ملف stc_TV_T3_completed.ipynb.