

דו"ח בינה מלאכותית – תרגיל 3

אסף יוסף 322380387, אלעד ישראל 313448888

חלק א' - Non-personalized

שאלה 1:

מתוך הקבצים השתמשנו בקובץ rating.csv שם יש את העמודות user_id, book_id, rating. איתם חישבנו ממוצע של rating עבור כל ספר וספרנו את מספר המדרגים עבור כל ספר. כל זה כדי לחשב את ה-score על פי weighted average rating כמו שנלמד בכיתה:

weighted average ratings

$$WeightedRating(WR) = \left(\frac{v}{v+m} \cdot R \right) + \left(\frac{m}{v+m} \cdot C \right)$$

v – מספר הקולות לסרט

m – מספר ההצבעות המינימלי הנדרש כדי להיות ברשימה

R – הדירוג הממוצע לסרט

C – ההצבעה הממוצעת בכל המטריצה

שאלה 2:

10 הספרים המומלצים ביותר על פי האלגוריתם שמימשנו:

	book_id	title	score
24	25	Harry Potter and the Deathly Hallows (Harry Po...	4.338028
3	4	To Kill a Mockingbird	4.299843
101	102	Where the Wild Things Are	4.273212
84	85	The Giving Tree	4.240309
49	50	Where the Sidewalk Ends	4.239724
30	31	The Help	4.238851
143	144	Unbroken: A World War II Story of Survival, Re...	4.221864
26	27	Harry Potter and the Half-Blood Prince (Harry ...	4.213906
0	1	The Hunger Games (The Hunger Games, #1)	4.187383
132	133	Anne of Green Gables (Anne of Green Gables, #1)	4.181489

שאלה 3:

10 הספרים המומלצים ביותר על פי האלגוריתם שמימשנו עבור אדם הגר בOhio:

	book_id	title	score
125	126	Dune (Dune Chronicles #1)	4.367963
142	143	All the Light We Cannot See	4.317786
143	144	Unbroken: A World War II Story of Survival, Re...	4.266087
23	24	Harry Potter and the Goblet of Fire (Harry Pot...	4.249728
101	102	Where the Wild Things Are	4.226877
489	490	Maus I: A Survivor's Tale: My Father Bleeds Hi...	4.213664
1458	1462	The Orphan Master's Son	4.213664
981	983	Between the World and Me	4.213664
118	119	The Handmaid's Tale	4.199565
88	89	The Princess Bride	4.190062

שאלה 4:

10 הספרים המומלצים ביותר על פי האלגוריתם שמימשנו עבור אדם בן 28:

	book_id	title	score
24	25	Harry Potter and the Deathly Hallows (Harry Po...	4.326251
3	4	To Kill a Mockingbird	4.294203
84	85	The Giving Tree	4.289614
88	89	The Princess Bride	4.244702
132	133	Anne of Green Gables (Anne of Green Gables, #1)	4.224914
49	50	Where the Sidewalk Ends	4.216411
101	102	Where the Wild Things Are	4.204680
69	70	Ender's Game (Ender's Saga, #1)	4.204095
30	31	The Help	4.202891
20	21	Harry Potter and the Order of the Phoenix (Har...	4.196385

חלק ב' - Collaborative filtering(user based)

התוצאות הללו הם עבור $user_id = 1$

Cosine:

	title	book_id
100	Me Talk Pretty One Day	101
773	Just Kids	775
263	The Sun Also Rises	264
288	Watership Down (Watership Down, #1)	289
334	James and the Giant Peach	335
1082	To the Lighthouse	1084
467	Their Eyes Were Watching God	468
183	Matilda	184
82	A Tale of Two Cities	83
343	Naked	344

Euclidean:

	title	book_id
203	A Christmas Carol	204
246	The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change	247
3882	Immortality	3913
3528	I Was Told There'd Be Cake	3555
2943	The Blue Sword (Damar, #1)	2954
189	Wild: From Lost to Found on the Pacific Crest Trail	190
3213	The Story About Ping	3232
4464	How to Tell If Your Cat Is Plotting to Kill You	4511
1404	Lone Survivor: The Eyewitness Account of Operation Redwing and the Lost Heroes of SEAL Team 10	1408
175	It	176

Jaccard:

	title	book_id
118	The Handmaid's Tale	119
86	Night (The Night Trilogy #1)	87
435	The God of Small Things	436
36	The Lion, the Witch, and the Wardrobe (Chronicles of Narnia, #1)	37
171	Anna Karenina	172
224	East of Eden	225
49	Where the Sidewalk Ends	50
303	I Know This Much Is True	304
180	In Cold Blood	181
263	The Sun Also Rises	264

חלק ג' - Content-based filtering

שאלה 8:

הפיצ'רים שבהם בחרנו להשתמש הם:

מטבלת books.csv:

authors, language_code, title, original_title

מטבלת books_tags.csv ומטבלת tags.csv:

tag_name

שאלה 10:

עבור הספר 'Twilight' ו- $k = 10$ קיבלנו את רשימת הספרים הבאה:

	book_id	title
2015	2021	The Twilight Collection (Twilight, #1-3)
990	992	The Twilight Saga (Twilight, #1-4)
51	52	Eclipse (Twilight, #3)
4196	4234	Twilight (The Mediator, #6)
1613	1619	The Twilight Saga Complete Collection (Twilig...
3059	3075	Twilight: The Graphic Novel, Vol. 1 (Twilight:...
4052	4088	The Twilight Saga: The Official Illustrated Gu...
48	49	New Moon (Twilight, #2)
55	56	Breaking Dawn (Twilight, #4)
1392	1396	Dragons of Autumn Twilight (Dragonlance: Chro...

חלק ד' - Evaluation metrics:

	precision_k	ARHR	RMSE
cosine	0.080000000000000002	0.3234761904761905	0.9009798683709508
euclidean	0.008	0.026666666666666665	0.9191466001886901
jaccard	0.08	0.32176190476190475	0.904945056963301

עבור precision_k ו-ARHR סיננו את הטבלה test על ידי כך שמחקנו שורות שבהן rating קטן מ-4 ובנוסף לאחר מכן מחקנו יוזרים שכמות הספרים שהם דירגו קטנה מ-4.

כתוצאה מכך קיבלנו טבלה הרבה יותר קטנה ממקודם, מה שגם כנראה מסביר את האחוזים הדי נמוכים שקיבלנו עבור המדדים הללו.

עבור RMSE קיבלנו תוצאות הרבה יותר גבוהות וזאת מכיוון שהמדד הזה הרבה יותר גמיש מהמדדים האחרים, כלומר עבור מדד זה לא צריך לבצע סינון כלשהו, אלא רק לעבור על השורות של הטבלה ולחשב את ההבדלים בין הדירוג האמיתי לדירוג שחזינו, בריבוע, ולאחר מכן לחלק במספר השורות שיש בטבלה של test ולעשות שורש על הכל.