

1. דירגו את המוצרים 5000 משתמשים ייחודיים, דורגו 500 מוצרים ייחודיים וקיימים בקובץ סה"כ 399,664 דירוגים.
2. מספר הדירוגים המינימלי שניתן למוצר הוא 737 והמקסימלי הוא 883.
3. מספר הדירוגים המינימלי שמשמש דירג הוא 53 והמקסימלי הוא 114.
- 4.

א. מערכת Collaborative Filtering היא מערכת המלצות המבוססת על המשתמשים עצמם למשל ע"י מציאת משתמשים הדומים והדירוג שלהם, ועל כך שמשתמשים עם מאפיינים דומים ככל הנראה גם יבחרו מוצרים דומים. לכן ברגע שיש למשתמש מסוים דירוגים מועטים, המערכת תתקשה למצוא משתמשים דומים אליו. בנוסף, המערכת לא תוכל להסיק מסקנות נכונות על מוצרים עם דירוגים מועטים (בעיקר כאלו חדשים או שאינם פופולריים) ולמצוא להם פריטים דומים (נקרא גם cold start).

ב. כאשר נרצה להמליץ על מוצרים חדשים או על מוצרים עליהם יש מיעוט נתוני דירוג, נשתמש במערכת המלצה מסוג Content based filtering כיוון שהיא מבוססת על attributes של המוצר, כלומר על התוכן שלו ולא על דירוגים שלו. המערכת מבצעת חישוב של פרופיל משתמש ע"י מוצרים שהוא דירג וכך יודעת להתאים לו מוצרים שמתאימים לו בתוכן. כך נימנע מ-cold start. כאשר נרצה להמליץ על מוצרים למשתמשים להם אין דירוגים בכלל, או למשתמשים חדשים, נשתמש ב-Non personalized RS כיוון שהיא לא תלויה בלקוח או בדירוגיו וכל המשתמשים מקבלים המלצות זהות. המערכת מבוססת על חישוב מוצרים פופולריים/טרנדיים/חדשים, המלצות מומחים, האתר, המיקום וכו'.

חלק ב – collaborating-filtering:

7. חמשת המוצרים המומלצים עבור משתמש "A3E00WA7R3LVBQ" עפ"י user-user cf יהיו:

['B0001Y74VS', 'B0002B0R14', 'B0009EX0NC', 'B0002JKQ8K', 'B0009PVV40']

8. חמשת המוצרים המומלצים עבור משתמש "A3E00WA7R3LVBQ" עפ"י item-item cf יהיו:

['B000052WYN', 'B000052YQ2', 'B000052ZSS', 'B0000534VO', 'B0000536M2']

חלק ג – הערכות:

9.

| | RMSE |
|------------------------|---------|
| user-based CF | 0.54751 |
| Item-based CF | 0.55211 |
| Mean based (benchmark) | 0.60155 |

10.

| | Precision@20 | Recall@20 |
|----------------------------|--------------|-----------|
| User-based CF | 0.04735 | 0.24685 |
| Highest-ranked (benchmark) | 0.04028 | 0.2128 |