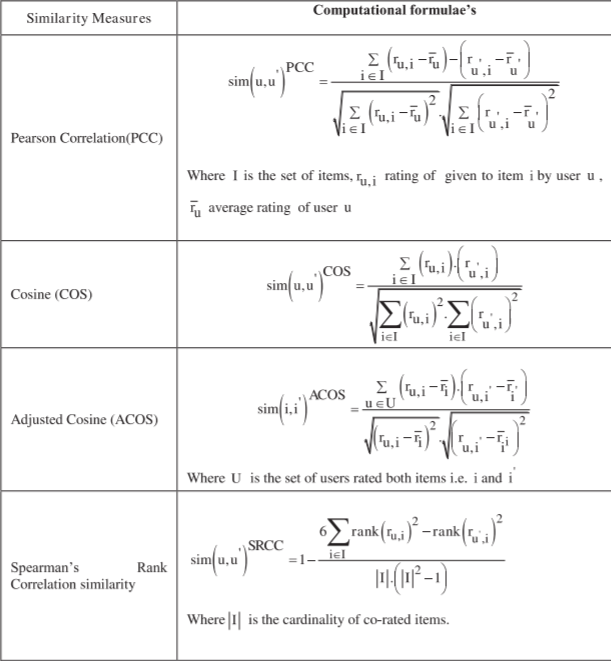
Báo cáo đánh giá hiệu quả của thuật toán collaborative filtering

Với thuật toán Collaborative Filtering chúng ta sử dụng similarity hay “độ tương đồng” để đưa ra gợi ý.

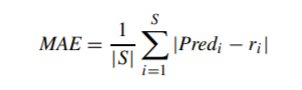
Với user-based Collaborative Filtering thì độ tương có thể tính bằng phương pháp đo lường cosine, Pearson, Euclidean, …



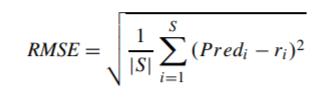
*Hình 1. Trích - A New Similarity Measure Based on Mean Measure of Divergence for Collaborative Filtering in Sparse Environment*

Hiệu quả và độ chính xác của các dự đoán được xác định dựa trên tham số đó chính là RMSE (Root Mean Squared Error) hoặc MAE (Mean Absolute Error).

MAE sẽ được tính bằng công thức sau:



RMSE sẽ được tính bằng công thức sau:



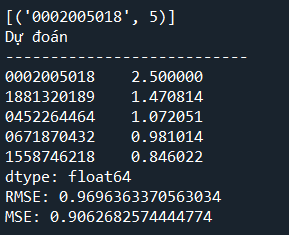
RMSE là một số thực không âm sẽ có giá trị từ 0.0 và 0.0 là trường hợp tốt nhất khi mà dự đoán không có sai lệch.

Để thực hiện tính toán 2 tham số này trong python chúng ta có thể sử dụng 2 thư viện là

**from** **sklearn.metrics** **import** mean\_absolute\_error để tính MAE

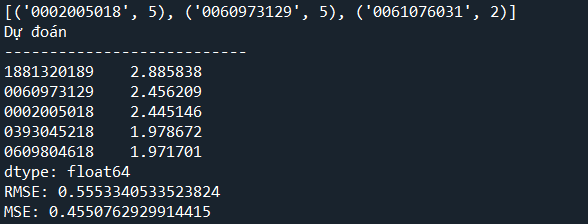
và **from** **sklearn.metrics** **import** mean\_squared\_error và set squared = False để tính RMSE

Bằng cách áp dụng vào bài demo chúng ta thu được kết quả như sau:



Với mỗi rating được dự đoán sẽ sai lệch khoảng 0.96 nếu người dùng chỉ đánh giá một cuốn sách điều nay cho chúng ta thấy hiệu suất ở lần dự đoán này là thấp.

Tương tự nếu chúng ta thực hiện dự đoán tiếp:



Với mỗi rating được dự đoán sẽ sai lệch khoảng 0.55, đây là một con số tốt đối với giải thuật Collaborative Filtering cho thấy hiệu suất của giải thuật là tương đối cao.

Kết luận: Collaborative Filtering là một giải thuật tốt trong một hệ thống recommend vừa và nhỏ.