Sistemes d’Informació en les Organitzacions

Filtratge col·laboratiu: Prediccions de puntuacions dels restaurants

Grup 4

Aleix Sancho Pujals

Francesc Ferré Tarrés

Víctor López Romero

Índex

[Mètodes emprats 3](#_Toc40561679)

[Instruccions sobre el codi 3](#_Toc40561680)

# Mètodes emprats

Inicialment vam fer ús de diferents tècniques que consistien en repartir els usuaris en funció dels que van puntuar més restaurants (en una escala de 10). Per tant teníem els restaurants repartits entre els que van anar de 0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100. Llavors, un cop realitzat això, el que vam fer va ser utilitzar les tècniques presents en el paquet de python anomenat scikit-learn, destinat a predir valors desconeguts (<https://scikit-learn.org/stable/modules/impute.html>). Els resultats d’aquestes prediccions, van ser nefastes, obteníem un MAE molt elevat (anava entre 4.48 a 6.68).

Llavors, veient aquests mal resultats, vam decidir cercar altres formes i vam veure un altre paquet de python, que es basa amb algorismes d’entrenament i que posteriorment, un cop entrenat l’algorisme, realitza prediccions. Aquest paquet s’anomena Surprise (<https://surprise.readthedocs.io/en/stable/getting_started.html>).

Aquest és el que ens ha permès arribar segons en el rànquing i passar d’estar suspesos a un 8.58, amb un MAE de 3.3862. En aquest cas, hem decidit que per entrenar l’algorisme el que li passem és aquells usuaris que han estat en més restaurants (més de 90) i aquells restaurants que han estat visitats més de 68000 vegades, per tal que les dades siguin el més fiables possibles.

Un cop fet això, vam estar un temps prudencial provant cadascun dels algorismes que el Surprise és capaç de realitzar a l’hora de realitzar prediccions amb una funció anomenada *cross\_validate*. Amb aquesta es fa com una prova per saber si va bé o malament cadascun dels algorismes. Per escollir l’algorisme bo, hem mirat aquell que ha donat més bon RMAE (segons les seves prediccions). En el nostre cas, ha estat l’algorisme [SVD](https://surprise.readthedocs.io/en/stable/matrix_factorization.html#surprise.prediction_algorithms.matrix_factorization.SVD). I ja a partir d’aquí simplement hem recollit la seva predicció, a través de la funció *predict*.

# Instruccions sobre el codi

L’execució en un ordinador no funciona del tot bé perquè dóna errors sobre la memòria. Per tant, hem hagut de realitzar aquesta pràctica amb l’ús del poder d’execució del cloud. En aquest cas, hem fet ús del python que hi ha a Anacondas i IDE Spyder.