Product Recommender for online stores.

Capstone Project
Data Science Career Track, Springboard

Thanks to mentor Julian Jenkins III

With the Surprise library, I used below algorithms

BaselineOnly: Algorithm predicting the baseline estimate for a given user and item.

KNNBaseline:A basic collaborative filtering algorithm taking into account a *baseline* rating.
NMF:A collaborative filtering algorithm based on Non-negative Matrix Factorization.
Co-clustering:A collaborative filtering algorithm based on co-clustering.
SVD:When baselines are not used, this is equivalent to Probabilistic Matrix Factorization, it is as popularized by Simon Funk during the Netflix Prize

Metrics used are rmse and mae. And here is the result.

Algorithm	test_rmse	test_mae	fit_time	test_time
BaselineOnly()	1.291117	1.018094	5.764257	2.878079
SVD()	1.295390	1.018963	52.802452	3.046659
CoClustering()	1.445493	1.115695	83.513256	2.960003

BaselineOnly algorithm has given the best rmse. So I used this model to get the product recommendation based on ratings.

Top 10 best predictions

	reviewerId	productId	ratings	predicted Ratings	itemsCounts_byReviewer	reviwerCounts_forItem	err
107286	AUQKQ2Q0N5N79	B005DB6NG6	3.0	2.999966	1	27	0.000034
277341	A1JFEP3BI48Q33	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
242103	A2AKFOQKKC1M2G	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
307506	A1WQ0YSLJ06KFT	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
202183	AOAVQTWTYP8C0	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
40597	A2TQCR6GN8XSP6	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
302422	A1CKWE4BW40LLG	B0094Z6N9Y	3.0	3.000041	0	176	0.000041
175043	AC5GLGVJXDEFV	B000IF51UQ	4.0	4.000056	2	512	0.000056
9586	AMIWD0PA1H1YB	B000HGHMF8	4.0	4.000077	5	26	0.000077
66049	A2SZZB6Y5X9T8M	B0028E1E2Y	4.0	4.000124	1	41	0.000124

The above are the top 10 best Predictions. The product 'B000HGHMF8','B000IF51UQ' and 'B0028E1E2Y' are rated best with ratings 4.

Top 10 worst predictions

	reviewerId	productId	ratings	predicted Ratings	itemsCounts_byReviewer	reviwerCounts_forItem	err
119632	A195EZSQDW3E21	B004J3V90Y	1.0	4.870298	19	1221	3.870298
8116	A227IMN9BB71BO	B0019EHU8G	1.0	4.880704	2	9265	3.880704
299570	AGR1V15L6FLMA	B007WTAJTO	1.0	4.894745	16	10683	3.894745
293699	A2QIX73WKLLDRV	B004Z4FBE2	1.0	4.902412	3	1154	3.902412
167309	ARXYNHFIXEYO2	B00BEW8MVC	1.0	4.909606	3	172	3.909606
333029	A1D9V11QUHXENQ	B004LSNF04	1.0	4.926395	6	1448	3.926395
241191	AC8YM4BB5LVOM	B0045JCFLY	1.0	4.954445	6	109	3.954445
168002	A6KL17KKN0A5L	B000JE7GPY	1.0	4.970054	11	1462	3.970054
311480	A3S3R88HA0HZG3	B008Q3CCZE	1.0	4.974388	7	152	3.974388
64010	A1Z16630QMH8Q6	B000E8BGCE	1.0	5.058453	13	164	4.058453

Cross checking with productId B007WTAJTO, this has predicted rating 4.89 and 10k reviewers rated this.

B0019EHU8G ratings