E-Commerce Recommendation Engine

Group 5

Recommendation Engine

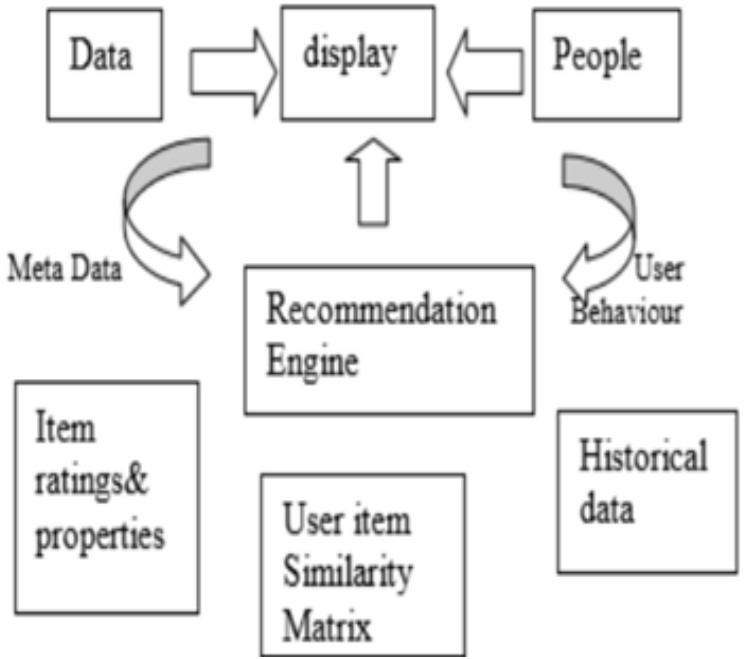


Fig : Architecture of a Recommendation Engine

Recommender systems can be stated as programs which try to recommend the most accurate suitable items to specific users by predicting a user's interest in an item, based on related information (By comparing all the features) about the items, the users and the interactions between items and users.

Recommendation engines are broadly classified into two categories:

- **Content based Filtering**
- **Collaborative Filtering**

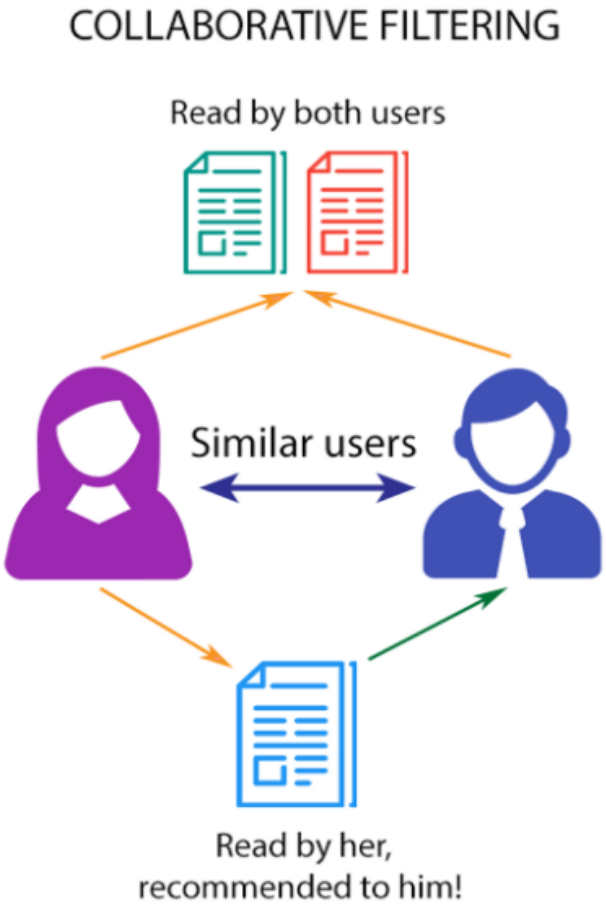


Fig : Collaborative Filtering

Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating).

We worked on 4 different models under the Collaborative Filtering approach.

1. **K - Nearest Neighbors**
2. **Stochastic Gradient Descent**
3. **Alternating Least Squares**
4. **Deep Learning based Neural Collaborative Filtering**

Related Work

- **Amazon uses an item-item collaborative filtering with matrix factorization to give personalized results.**
- **Amazon researchers found that using neural networks to generate movie recommendations for Amazon Prime performed better when the input data was organized chronologically and used to predict future movie preferences over a short period of time.**

Our approach consists of utilizing four different machine learning models and combining them using an ensemble approach to achieve a better performance.

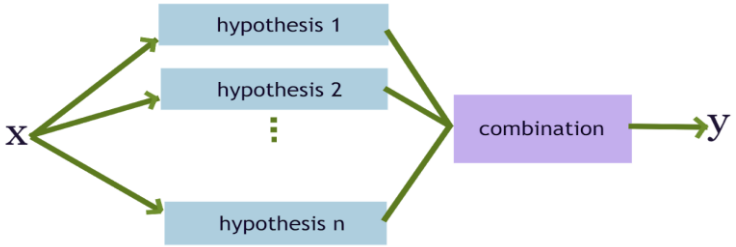


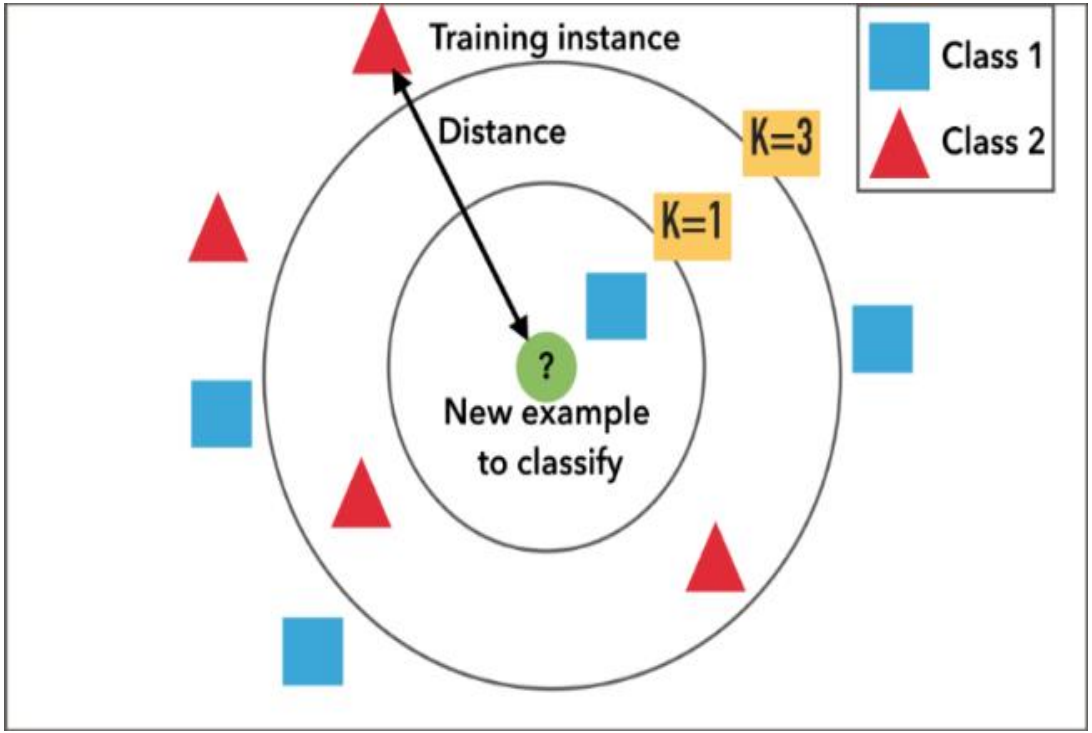
Figure: A general Ensemble architecture

Dataset

Amazon Review Data :

	overall	verified	reviewTime	reviewerID	asin	reviewerName	reviewText	summary	unixReviewTime	vote	style
7	3.0	False	03 15, 2006	A3JPFWKS83R49V	B00005N7PS	Bryan Carey	Details is a mildly interesting magazine that ...	Give Me the Details	1142380800	98	NaN
8	4.0	False	09 24, 2007	A3JPFWKS83R49V	B00005N7P0	Bryan Carey	Maximum PC is a magazine for electronics freak...	Maximizing Electronics Enjoyment	1190592000	2	NaN
131	4.0	False	12 23, 2011	A3JPFWKS83R49V	B00005N7T3	Bryan Carey	Texas Monthly Magazine is a magazine about a s...	Texas Monthly Keeps You Up to Speed with the L...	1324598400	NaN	{'Format:': ' Print Magazine'}
264	4.0	False	07 12, 2007	A3JPFWKS83R49V	B00005N7SS	Bryan Carey	Smart Money is a solid magazine about business...	Smart Money Combines Business and Personal Fin...	1184198400	21	NaN

Neighbourhood based collaborative filtering Using KNN



Neighborhood Based Collaborative Filtering leverages the behavior of other users to know what our user might enjoy.

It may find people similar to our user and recommend stuff they liked or recommend stuff that other people bought after buying what our user has bought.

Results

	K	Cosine	Euclidean	Manhattan
0	5	30.66	28.80	31.37
1	10	40.54	41.98	42.41
2	15	46.13	41.55	44.41
3	20	42.84	40.40	43.41
4	25	43.55	41.12	37.68
5	30	42.55	37.97	42.84
6	35	41.40	40.40	40.97
7	40	41.26	39.40	41.40
8	45	39.26	39.26	41.55
9	50	39.68	38.97	41.55

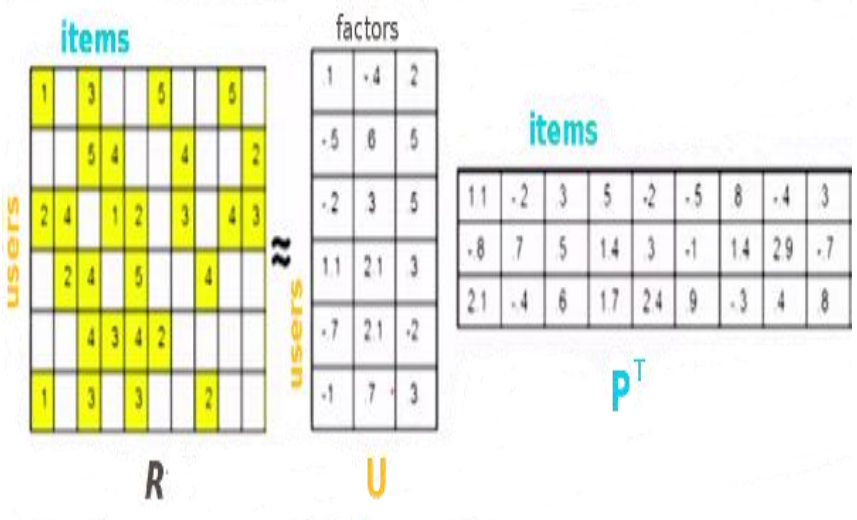
	Max Hit_Rate %	k
Cosine	46.13	15
Euclidean	41.98	10
Manhattan	44.41	15

Evaluation Metric: Hit Ratio

Top 10 Recommendations :

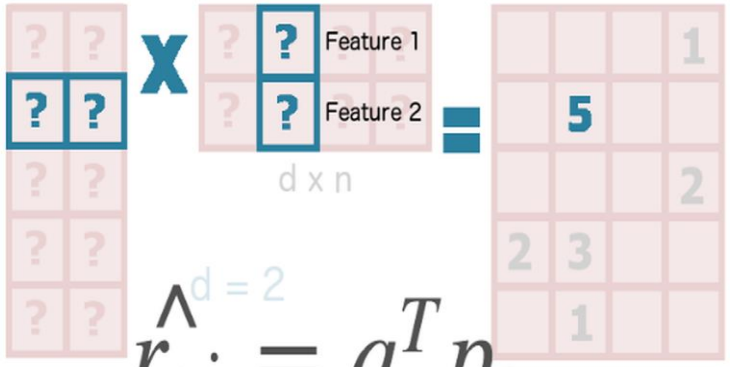
User Has Rated:
B00YMMOHU6 - Massage Oil Body Oil Fractionated Coconut Oil For Aromatherapy Relaxing Oil Massage - P
Items Recommended:
B019ERA2SY - SDB Brush Hair Straightener,Anion Hair Care, Anti Scald,Instant Magic Silky Straight Ha
B000LCETUO - Pure Instinct - #Jelique
B00YDUKVFC - Fan Makeup Brush - Powder Concealer Foundation Blush Blending Highlight Bronzer Contour
B001QY8QXM - Astra Platinum Double Edge Safety Razor Blades ,100 Blades (20 x 5) - #Astra
B00FP0HB1G - Milliard NON-GMO Emulsifying Wax Pastilles NF –16 OZ. Resealable Freshness Storag
B00FIUEIY6 - Beauty Aura 100% Pure Avocado Oil - 16 Fl Oz - For Healthy Hair, Skin & Nails. - #B
B003XGYTQE - Brahmi Amla Hair Oil (8 oz) by Vadik Herbs | Ayurvedic herbal hair growth oil and hair
B003K58UEA - Estee Lauder Advanced Night Repair Synchronized Recovery Complex, 3.4 Ounce - #Estee La
B015W8Y1G8 - Living Nature Pure Antibacterial Manuka Oil - #

Matrix Factorization



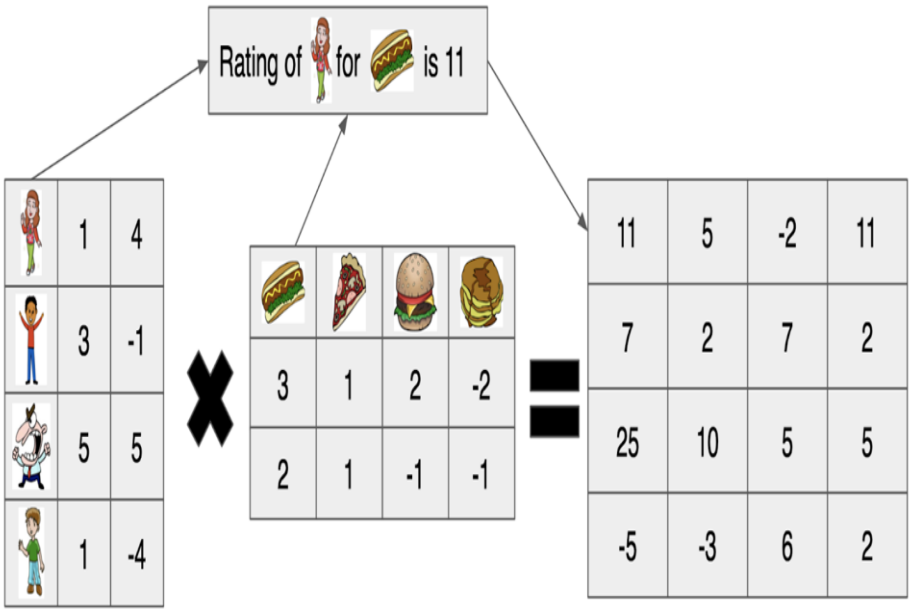
Matrix Factorization

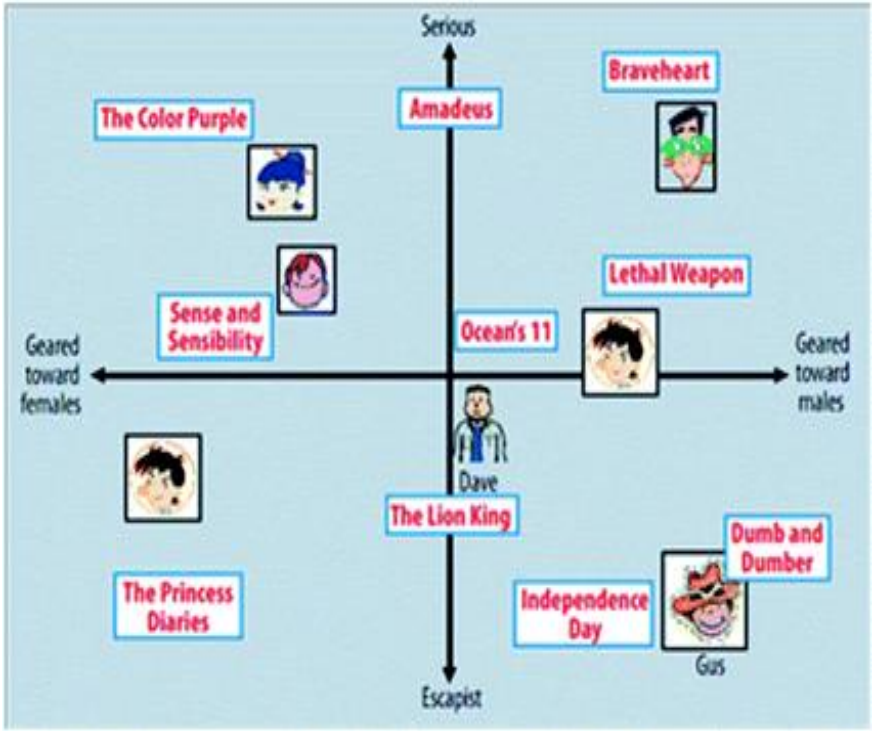
m = number of users, n = number of items
choose d, the number of features



$$\hat{r}_{ui} = q_i^T p_u$$

Matrix Factorization





Stochastic Gradient Descent (SGD)

$$\min_{q^*,p^*} \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

$$e_{ui} \overset{def}{=} r_{ui} - q_i^T p_u.$$

$$q_i \leftarrow q_i + \gamma \cdot (e_{ui} \cdot p_u - \lambda \cdot q_i)$$
$$p_u \leftarrow p_u + \gamma \cdot (e_{ui} \cdot q_i - \lambda \cdot p_u)$$

With SGD,

- **We take derivatives of the loss function, but we take the derivative with respect to each variable in the model.**
- **The “stochastic” aspect of the algorithm involves taking the derivative and updating feature weights one individual sample at a time.**
- **So, for each sample, we take the derivative of each variable, set them all equal to zero, solve for the feature weights, and update each feature until they converge.**

Alternating Least Squares (ALS)

For ALS minimization,

- **Holding one set of latent vectors constant, we then take the derivative of the loss function with respect to the other set of vectors (the user vectors).**
- **We set the derivative equal to zero (we're searching for a minimum) and solve for the non-constant vectors (the user vectors).**
- **We alternate back and forth and carry out this two-step dance until convergence.**

$$W_{n+1}^{\mathrm{T}} \leftarrow (H_n H_n^{\mathrm{T}})^{-1} H_n V^{\mathrm{T}},$$

$$H_{n+1} \leftarrow (W_{n+1}^{\mathrm{T}} W_{n+1})^{-1} W^{\mathrm{T}} V.$$

$$\lambda(\|W\|_{\mathrm{F}}^2 + \|H\|_{\mathrm{F}}^2)$$

$$W_{n+1}^{\mathrm{T}} \leftarrow (H_n H_n^{\mathrm{T}} + \lambda I)^{-1} H_n V^{\mathrm{T}}$$

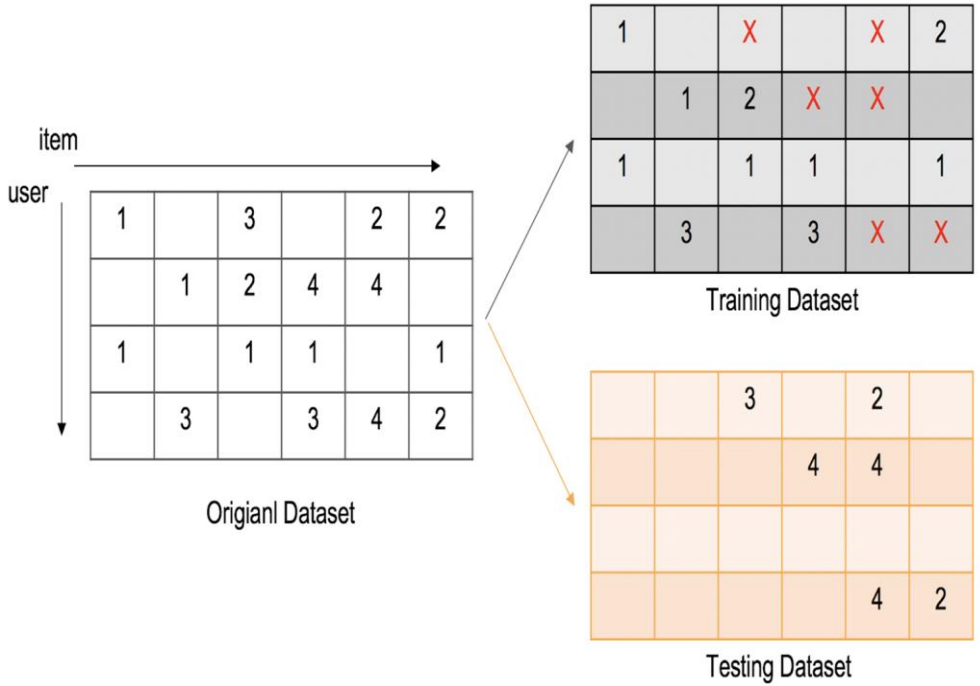
$$H_{n+1} \leftarrow (W_{n+1}^{\mathrm{T}} W_{n+1} + \lambda I)^{-1} W^{\mathrm{T}} V$$

Train Test Split for SGD and ALS

Dataset Format

	uid	iid	ratings	timestamp
0	A2HOI48JK8838M	B00004U9V2	5.0	1515110400
1	A1YIPEY7HX73S7	B00004U9V2	5.0	1491350400
2	A2QCGHIJ2TCLVP	B00004U9V2	5.0	1490572800
3	A2R4UNHFJBA6PY	B00004U9V2	5.0	1489968000
5	A1606LA683WZZU	B00004U9V2	5.0	1487980800

time: 11.5 ms (started: 2021-04-26 22:13:38 +00:00)



Results

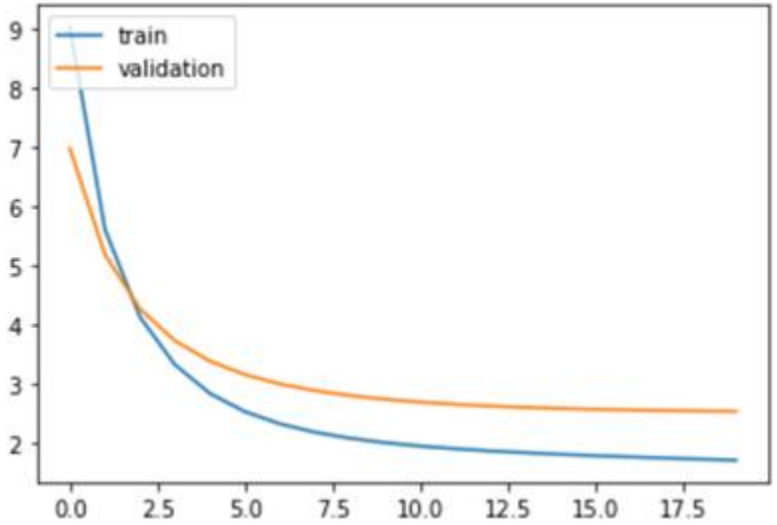


Fig: Learning Curve

Evaluation Metric: Mean Absolute Error

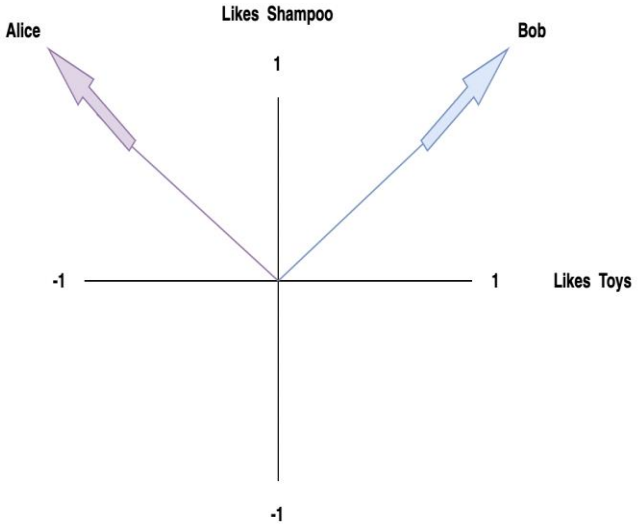
	uid	iid	ratings	timestamp	SGD_predicted
129604	A36XJ9R83M2KN8	B00WUK8H3K	3.0	1497830400	4.393713
63295	A2XQM08NJJDYKH	B00SH4ORI2	5.0	1457222400	4.640791
66250	A1EVYLO4NV4MH0	B00TIRORLE	5.0	1467676800	4.643406
50210	A1JFOYTNINKWAY	B00O4ZM0B4	5.0	1513296000	3.416471
22784	A1HVD85HGVA0DI	B001HWGMBG	3.0	1474070400	5.040785

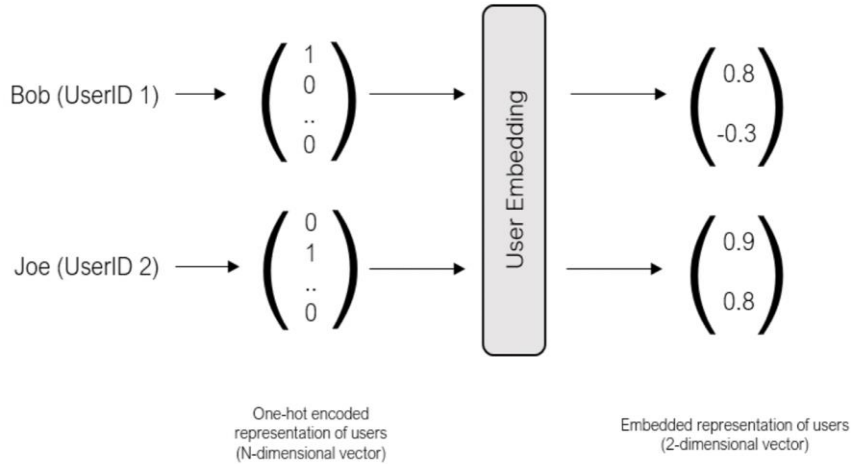
Top 10 Recommendations :

```
A85WY5ZDT8GXW
B0006PJRRG: ('ZOYA Nail Polish, 0.5 fl. oz.', 4.177538778133386)
B00R3PZK14: ('CND Shellac Nail Polish, Field Fox', 4.081247011111219)
B0006PJRVM: ('ZOYA Nail Polish, 0.5 fl. oz.', 4.022690431495288)
B008H3SW4I: ('Microsoft Windows 8 Pro - Upgrade [Old Version]', 3.894872889554204
B00QYX6F5G: ('HDE Non-Contact Infrared Thermometer Digital Temperature Gun with L
B00IRNQJYI: ('CND Shellac, Sultry Sunset', 3.846578013810981)
B004N2SQUC: ('CND Shellac Nail Polish, Hollywood, 0.25 fl. oz.', 3.81924497285823
B00MNSHGPE: ('Band-Aid Brand Tough-Strips Adhesive Bandage for Minor Cuts &amp;amp; S
B01EUII7OI: ('Lipton Green Tea Bags, Decaffeinated, 40 ct', 3.7968408424345617)
B005CEHTEY: ('3M 03142 5&amp;quot; Backed Adhesive Disc Pad', 3.770796660224828)
time: 3.67 s (started: 2021-04-26 20:57:40 +00:00)
```

Neural Collaborative Filtering

User/Item Embeddings



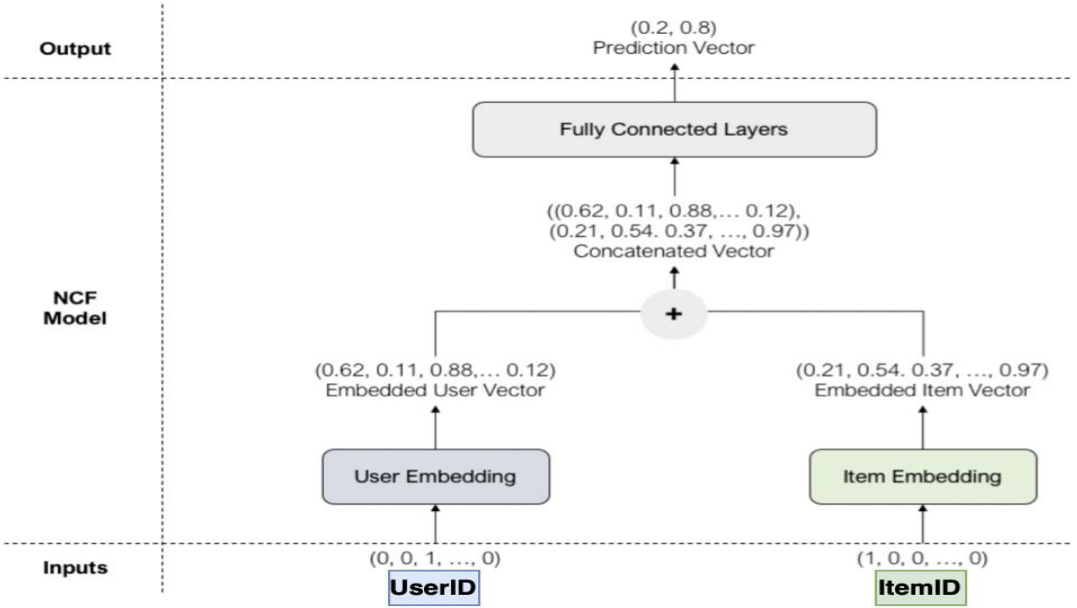


Neural Collaborative Filtering

Implicit Ratings

	reviewerID	asin	overall
0	74103	1	1

Model Architecture



Neural Collaborative Filtering

```
Actual item bought by user:  Avalon Grapefruit and Geranium Smoothing Shampoo, 11 Ounce
Top 10 recommended items for user:  974
item  1 :  Avalon Grapefruit and Geranium Smoothing Shampoo, 11 Ounce
item  2 :  Yardley By Yardley Of London Unisexs Lay It On Thick Hand &amp;amp; Foot Cream 5.3 Oz
item  3 :  Fruits &amp;amp; Passion Blue Refreshing Shower Gel – 6.7 fl. oz.
item  4 :  65
item  5 :  77
item  6 :  16
item  7 :  68
item  8 :  14
item  9 :  12
item  10 :  14
```

- For each user, randomly select 99 items that the user **has not interacted with.**
- Combine these 99 items with the test item (the actual item that the user last interacted with). We now have 100 items.
- Run the model on these 100 items, and rank them according to their predicted probabilities.
- Select the top 10 items from the list of 100 items. If the test item is present within the top 10 items, then we say that this is a hit.
- Repeat the process for all users. The Hit Ratio is then the average hits.

Evaluation Metric: Hit Ratio

Optimizers	Hit Ratio
Adam	0.82
Nesterov (beta =0.09)	0.80
RMSprop	0.91

Summary

Liked in the project:

- **The way different recommendation engines work.**
- **Working with the amazon dataset provided us insights into the user and item data along with their metadata.**

Learnt from the project:

- **Different evaluation measures for recommendation engines.**
- **Way to work with sparse matrix and deal with the train-test split for matrix factorization.**
- **The way implicit and explicit ratings work for the recommendation systems.**

Challenges faced:

- **In the case of large data, the user-item pivot table was taking a lot of space and it slowed down our model so used user and item mapping instead.**
- **Due to the unavailability of the user metadata we could not counter the cold start problem using content based filtering,so we are recommending the most popular items to the new users.**

References

- https://blog.insightdatascience.com/explicit-matrix-factorization-als-sgd-and-all-that-jazz-b00e4d9b21ea
- http://deepyeti.ucsd.edu/jianmo/amazon/index.html
- https://www.amazon.science/the-history-of-amazons-recommendation-algorithm
- https://stanford.edu/~rezab/classes/cme323/S16/projects_reports/baalbaki.pdf
- https://arxiv.org/abs/1708.05031
- https://datajobs.com/data-science-repo/Recommender-Systems-[Netflix].pdf
- l9.pdf (stanford.edu)
- lec14.pdf (stanford.edu)
- https://heartbeat.fritz.ai/recommender-systems-with-python-part-ii-collaborative-filtering-k-nearest-neighbors-algorithm-c8dcd5fd89b2

Thank You