How to optimize product recommendations by using machine learning?

METHODOLOGY & DESIGN

SCIENTIFIC WRITING CLASS II

XIAOMIN JIN

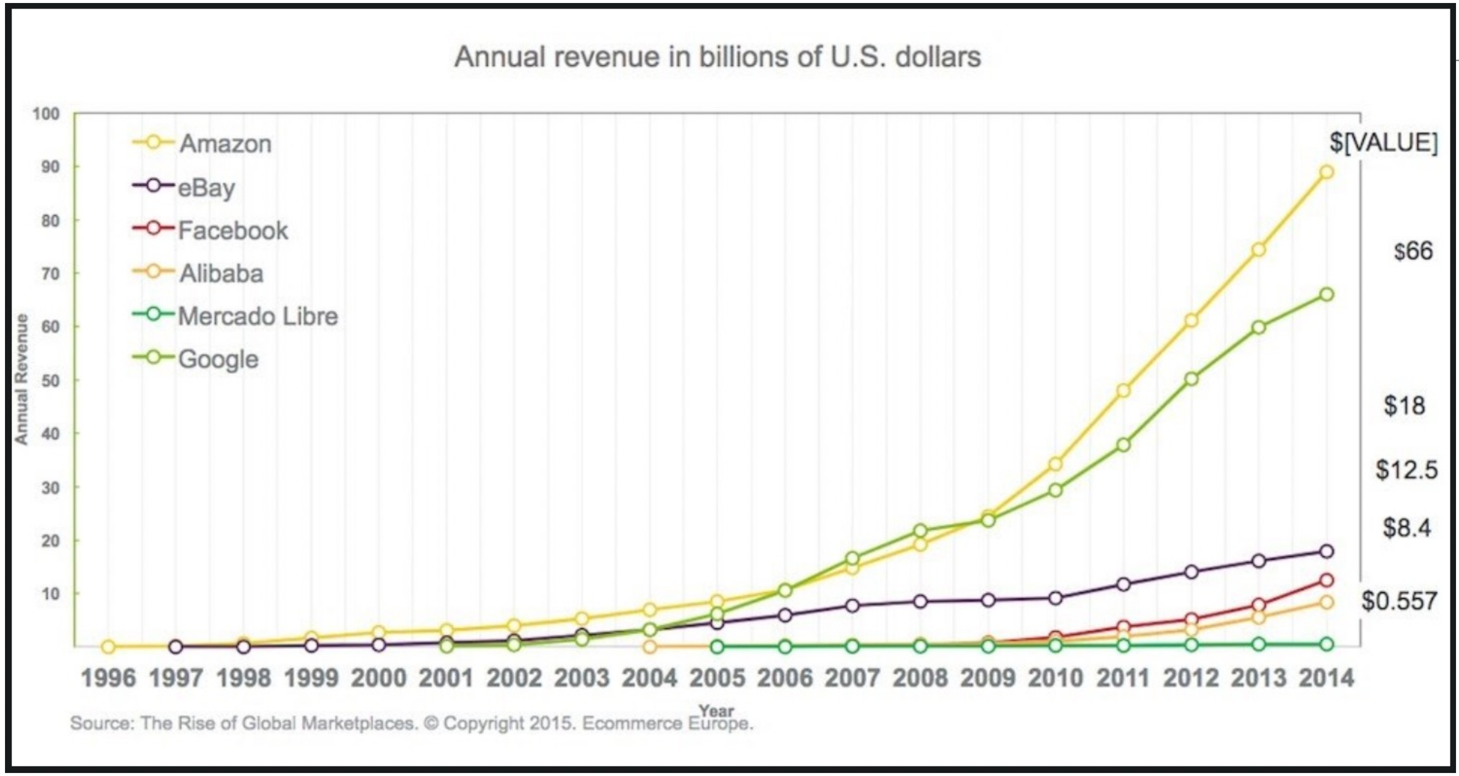
SUMMER TERM 2022

Introduction

What is a recommendation engine/system in general?

- uses a set of machine learning / deep learning algorithms
- gives customers a tailored experience
- used in areas like e-commerce, social media, streaming services, education, knowledge management

Based on user	Based on other users
purchase history	most-viewed items
search behavior	trending items
product preferences	feedback from other users
wishlists	most-sold-items
geographic location	
age	
gender	



Source: [6]

Motivation

- e-commerce exists for more than 40 years
- 1969 CompuServe was founded
- 1995 Amazon & ebay
- kicked off 2005
- Amazon + ebay = 15 billion$

Motivation

Source: [10]

Motivation

rise of consumers → companies want to provide more product → which product is suited for consumers interests?

- researchers [12]: connected customer loyality to quality of products

What is the goal of the paper?

- How to implement an optimized product recommendation engine
- recommend products with higher quality (4-5 stars)

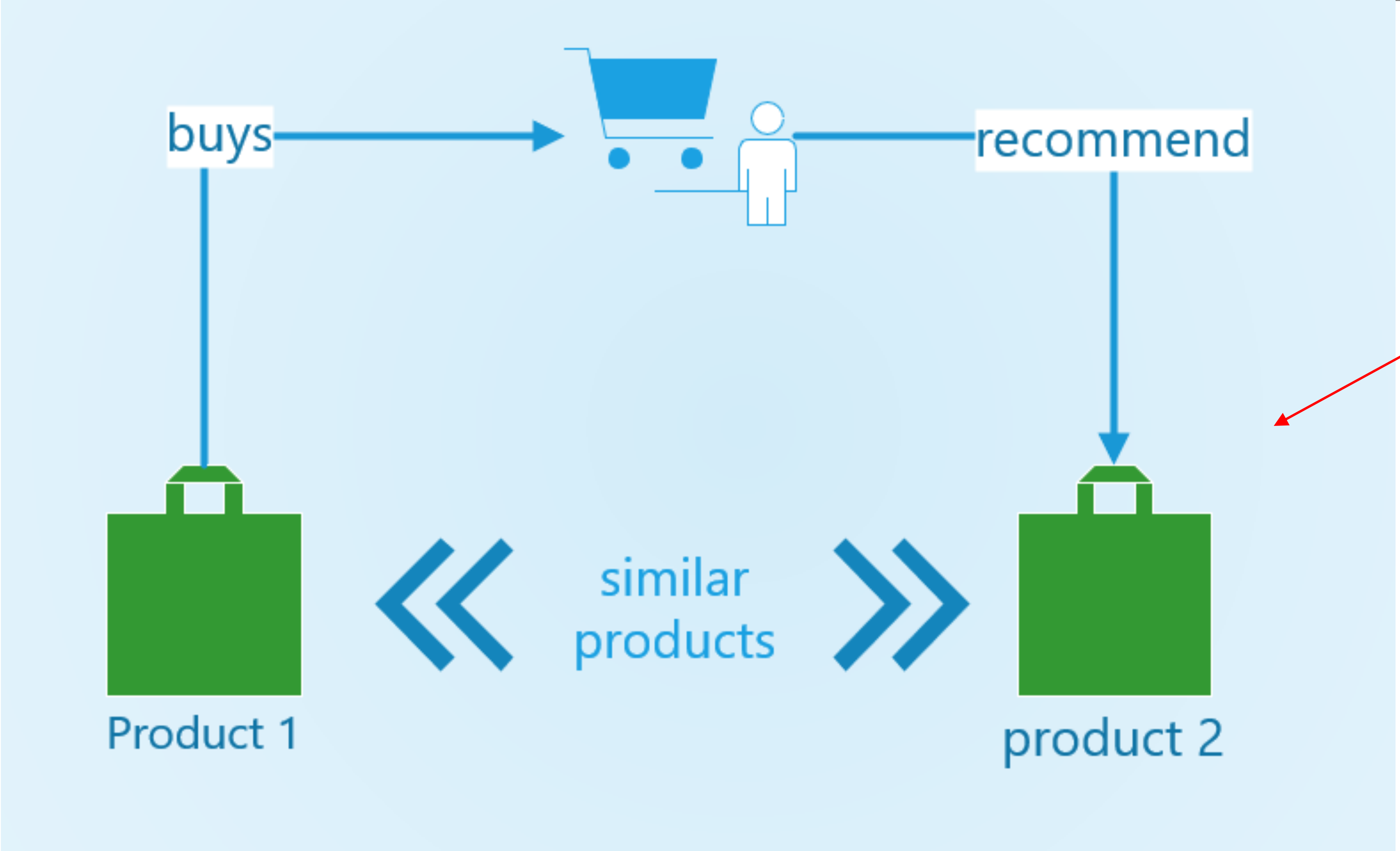
Product recommendation engines (PRE)

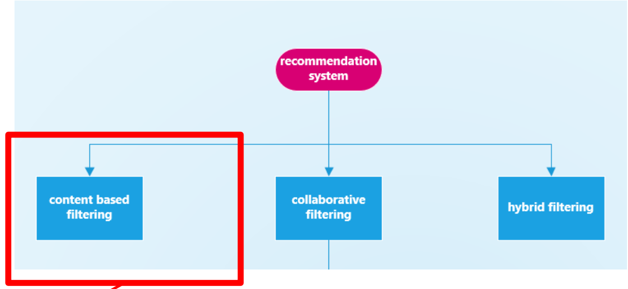
Common recommendation systems

Rating matrix: $user_i \times input_i$

	input1	input2	input3	input 4	input 5
user1	?	5	1	?	5
user2	2	4	?	2	?
user3	5	?	1	?	?
user4	1	?	?	5	2
user5	?	?	1	?	5

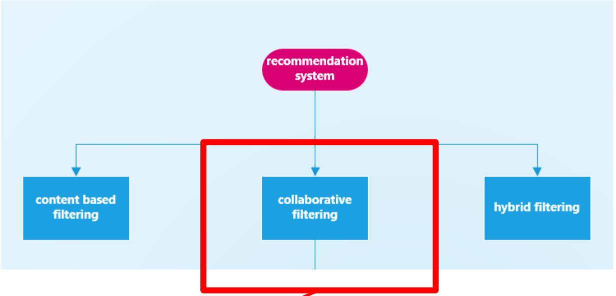
Content based filtering





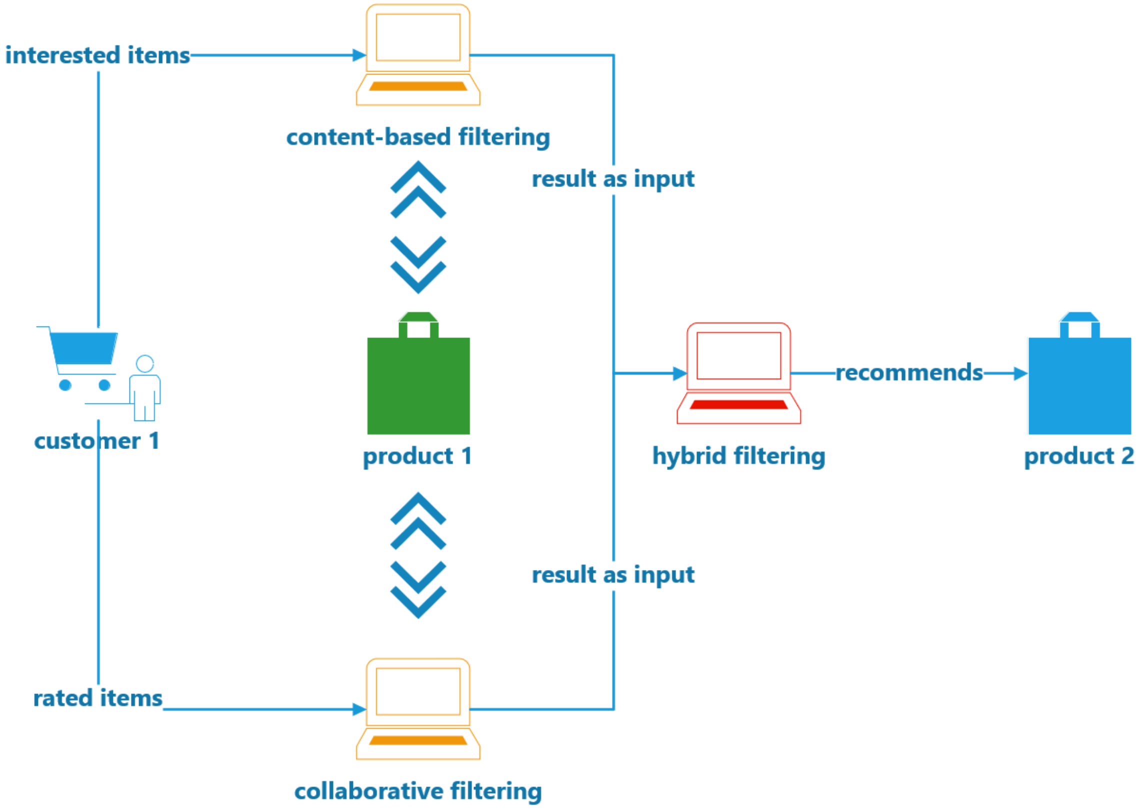
Source: [13]

Collaborative filtering



Source: [13]

Hybrid-based filtering



Rating matrix in collaborative filtering

How do humans behave?

	I1	I2	I3	I4	I5
Alice	3	1	1	3	1
Bob	1	2	4	1	3
Eve	3	1	1	3	1
Dave	4	3	5	4	4

- more realistic scenario:
- similarities between users
- similarities between items

In Style Source [14]: Serano Academy videos

How do humans behave?

	I1	I2	I3	I4	I5
Alice					
Bob	1	2	4	1	3
Eve	3	1	1	3	1
Dave	4	3	5	4	4

Bob + Eve = Dave

- Bob loves PC with face unlock
- Eve loves PC with biometric
- Dave loves both

In Style Source [14]: Serano Academy videos

How do humans behave?

	I1	I2	I3	I4	I5
Alice	3			3	
Bob	1			1	
Eve	3			3	
Dave	4			4	

I1 = I4

- Dell Latitude
- 14”
- Silver
- Biometric unlock

- M1 MacBook Pro
- 16”
- Silver
- Biometric unlock

In Style Source [14]: Serano Academy videos

Singular value decomposition

What is singular value decomposition (SVD)?

- a collaborative filtering algorithm

- utilises matrix factorization

Dimensional Reduction

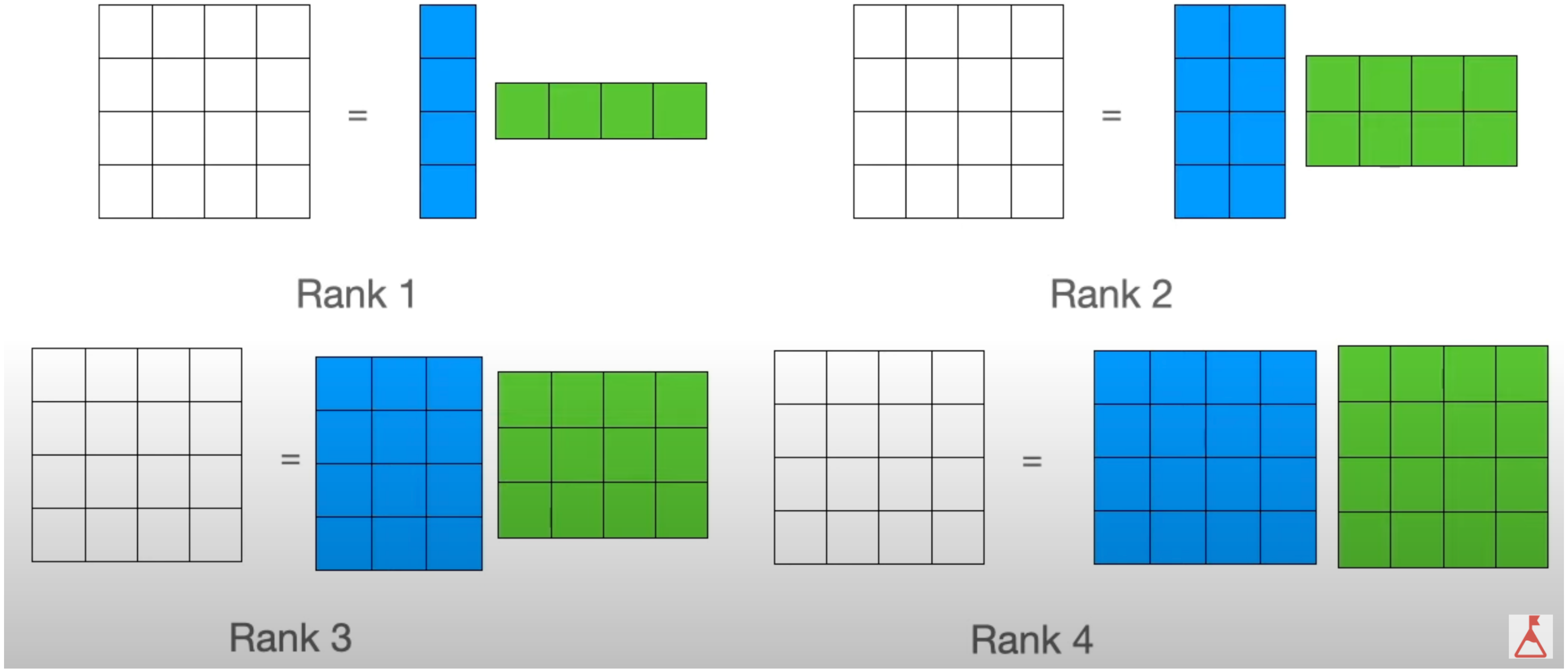
$$\text{Loss} = \Sigma_i \Sigma_j (A_{ij} - u_i v_j^T)^2$$

A_{ij} = original value in the matrix
u_i, v_j = values generated by the algorithm

In Style of Source [16]: Serano Academy videos

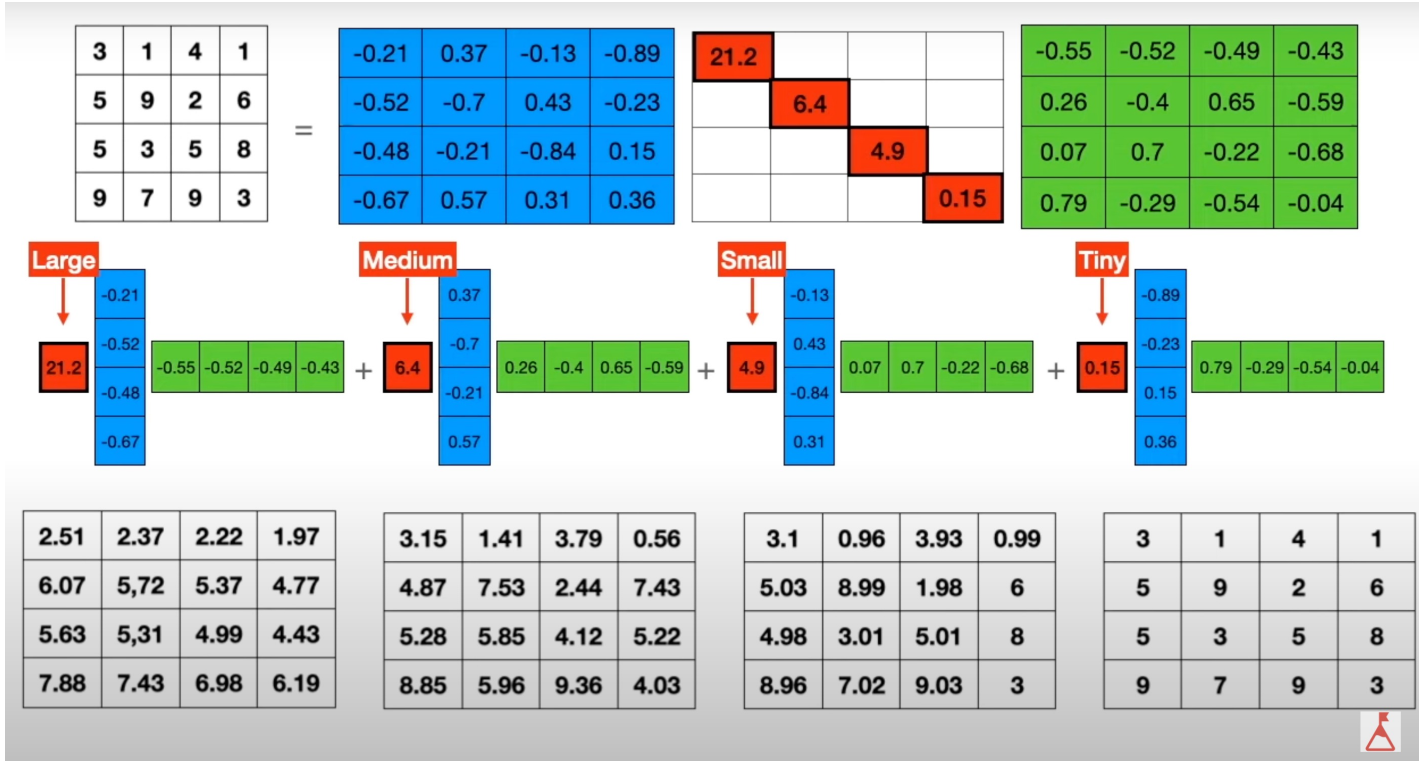
Dimensional reduction in SVD with known features

$$\sum_{i=1}^{n} a_i b_i =$$



Source [15]: Serano Academy videos

Dimensional reduction in SVD with known features



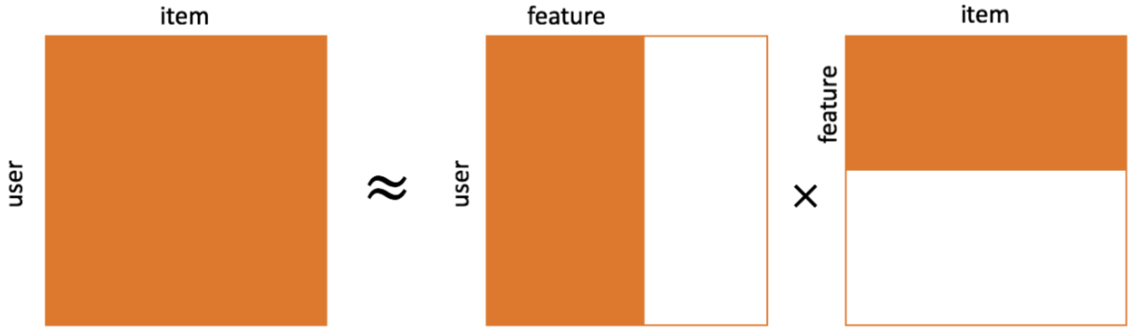
$$R = U\Sigma V$$

Source [15]: Serano Academy videos

Dimensional reduction in SVD with unknown features

	I1	I2	I3	I4	I5
Alice	3	5	4	4	2
Bob	0	0	0	0	0
Eve	0	0	0	0	0
Dave	4	3	0	0	0

(a) User-item table with unknown features and ratings



(b) Dimensional reduction for matrices with unknown features

Source [18]: picture (b)

$$r_{ui} = \sum_{k=1}^{K} w_{uk}\sigma_k z_{ki}$$

k = 1

$$r_{ui} z_{ki} \frac{1}{\Sigma} = w_{uk}$$

$$r_{ui} w_{uk} \frac{1}{\Sigma} = z_{ki}$$

$$e_{ui}^2 = (r_{ui} - \widetilde{r_{ui}})^2 = (r_{ui} - \sum_{k=1}^{K} w_{uk}\sigma_k z_{ki})$$

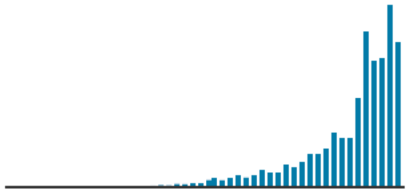
$$dw_{uk} = w_{uk} - \alpha^* \frac{\partial}{\partial w_{uk}}(e_{ui}^2) = w_{uk} + 2e_{ui}z_{ki}$$

$$dz_{ki} = z_{ki} - \alpha^* \frac{\partial}{\partial z_{ki}}(e_{ui}^2) = z_{ki} + 2e_{ui}w_{uk}$$

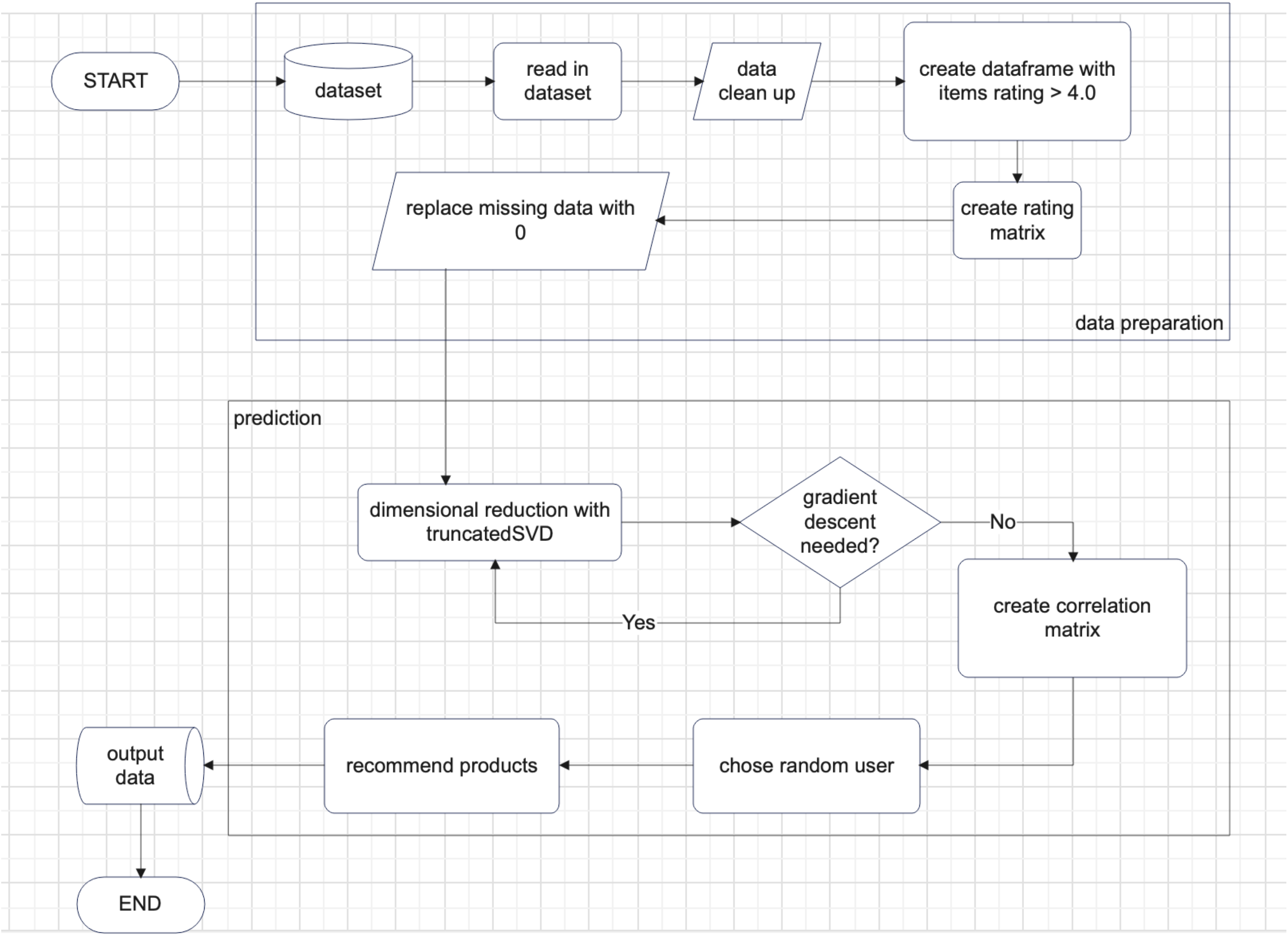
Design

Design of a PRE

Detail Compact Column 4

AKM1MP6P0OYPR	0132793040	# 5.0	# 1365811200
4201696 unique values	**476001** unique values	 1 5	 913m 1.41b
A2CX7LUOHB2NDG	0321732944	5.0	1341100800
A2NWSAGRHCP8N5	0439886341	1.0	1367193600
A2WNBOD3WNDNKT	0439886341	3.0	1374451200
A1GI0U4ZRJA8WN	0439886341	1.0	1334707200

source: [17]



Design of a PRE

Implementation

Choosing the right tools

- **Microsoft Visual Studio Code:**
  - Microsofts own multipurpose IDE
  - is very versatile for many languages
  - not good for displaying tables or charts
- **Jupyter Notebook:**
  - most used open-source scripting tool in ML
  - good at displaying plots
  - not good for RAM intensive calculations
- **Google Colab Notebook:**
  - Google’s version of the Jupyter Notebook.
  - has all the advantages of the Jupyter Notebook
  - has two runtime environments (computer, Google’s runtime environment)
  - is managed server-side.

Implementation: prepare data set

	UserId	ItemId	Rating	Timestemp
0	A2CX7LUOHB2NDG	0321732944	5.0	1341100800
1	A2NWSAGRHCP8N5	0439886341	1.0	1367193600
2	A2WNBOD3WNDNKT	0439886341	3.0	1374451200
3	A1GI0U4ZRJA8WN	0439886341	1.0	1334707200
4	A1QGNMC6O1VW39	0511189877	5.0	1397433600

display raw data set

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 4 columns):
 #   Column     Non-Null Count    Dtype
---  ------     --------------    -----
 0   UserId     1048575 non-null  object
 1   ItemId     1048575 non-null  object
 2   Rating     1048575 non-null  float64
 3   Timestemp  1048575 non-null  int64
dtypes: float64(1), int64(1), object(2)
memory usage: 32.0+ MB
```

display data type

Implementation: prepare data set

	Rating	Timestemp
count	1.048575e+06	1.048575e+06
mean	3.973379e+00	1.248822e+09
std	1.399329e+00	1.091615e+08
min	1.000000e+00	9.127296e+08
25%	3.000000e+00	1.169078e+09
50%	5.000000e+00	1.250035e+09
75%	5.000000e+00	1.355789e+09
max	5.000000e+00	1.406074e+09

caluclate:
min, max, mean, etc

Implementation: prepare data set

```
# create a data frame only with rating = (4 or 5)
df = pd.DataFrame(data)
df = df.loc[df['Rating'].isin([4, 5])]
#As data is huge so we take a fraction of the data so that we can create a user item matrix.
data1 = df.head(100000)
ratings_utility_matrix = data1.pivot_table(values='Rating', index='UserId', columns='ItemId')
ratings_utility_matrix.head()
```

ItemId	0321732944	0511189877	0528881469	059400232X	0594012015	0594033896	0594033926	0594033934	0594202442	0594287995	...	B0000632FV
UserId												
A00766851QZZUBOVF4JFT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
A00995931BE16NG4F52QC	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
A01255851ZO1U93P8RKGE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
A014623426J5CM7M12MBW	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
A01852072Z7B68UHLI5UG	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN

5 rows × 7973 columns

Implementation: prepare data set

UserId	A00766851QZZUBOVF4JFT	A00995931BE16NG4F52QC	A01255851ZO1U93P8RKGE	A014623426J5CM7M12MBW	A01852072Z7B68UHLI5UG	A0266076X6KPZ6CCHGVS	A0293130VT
ItemId							
0321732944	0.0	0.0	0.0	0.0	0.0	0.0	
0511189877	0.0	0.0	0.0	0.0	0.0	0.0	
0528881469	0.0	0.0	0.0	0.0	0.0	0.0	
059400232X	0.0	0.0	0.0	0.0	0.0	0.0	
0594012015	0.0	0.0	0.0	0.0	0.0	0.0	

fill unknown values with 0.0

Implementation: prediction phase

```
array([[ 1.        ,  0.02140177,  0.30060272, ...,  0.37926624,
        -0.14793125,  0.15077707],
       [ 0.02140177,  1.        ,  0.75425745, ...,  0.02838692,
         0.02519788, -0.10117692],
       [ 0.30060272,  0.75425745,  1.        , ..., -0.18943995,
        -0.52066005,  0.09321444],
       ...,
       [ 0.37926624,  0.02838692, -0.18943995, ...,  1.        ,
         0.68335819, -0.09329627],
       [-0.14793125,  0.02519788, -0.52066005, ...,  0.68335819,
         1.        , -0.36007566],
       [ 0.15077707, -0.10117692,  0.09321444, ..., -0.09329627,
        -0.36007566,  1.        ]])
```

- decomposed utility matrix with truncatedSVD
- created correlation matrix
- negative numbers = no similarities
- positive numbers = has similarities
- 1.0 = perfect match

Implementation: prediction phase

```
array([ 0.33975131,  0.64725252,  0.29503636, ...,  0.70086137,
        0.42684685, -0.04896438])
```

user-item correlation matrix for specific random user

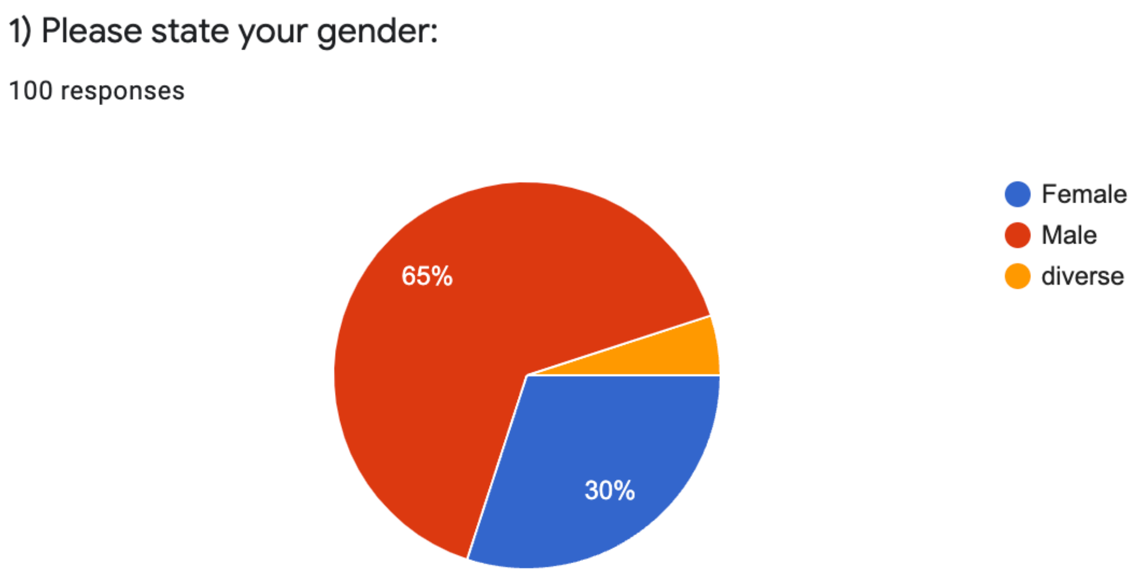
Implementation: prediction phase

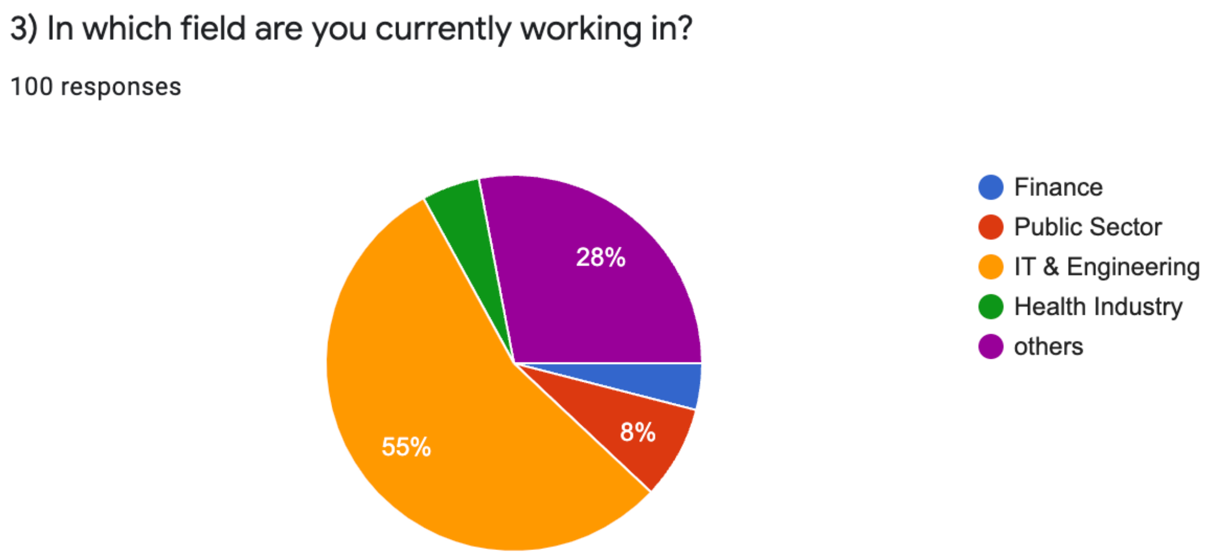
```
['1575839415',
 '1891747134',
 '6000006896',
 '8805002097',
 '9269807207',
 '9800359648',
 '9861023909',
 '986106172X',
 '9966286624',
 '9966541721']
```

Predicted items: most similar to users interests

Survey results

Survey results: general information

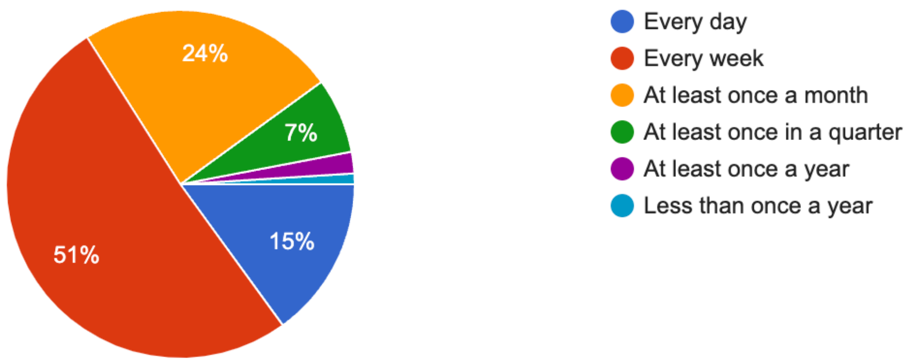




Survey results: shopping behavior

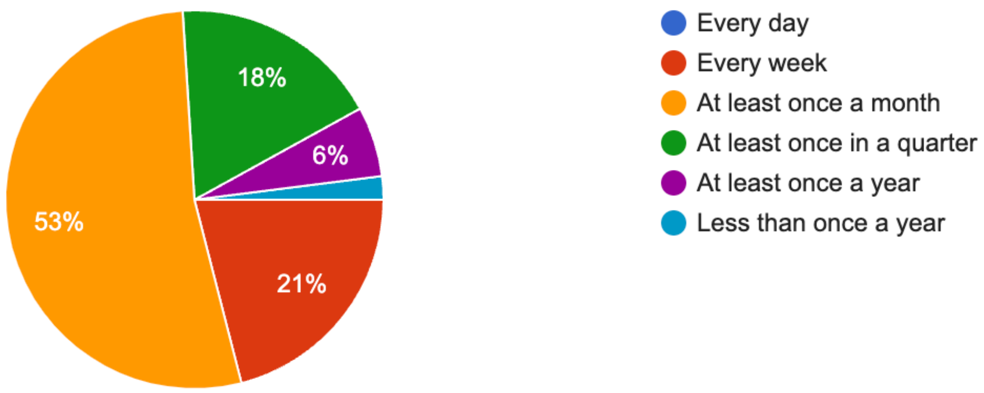
5) How often do you visit online shops?

100 responses



6) How often do you buy products online?

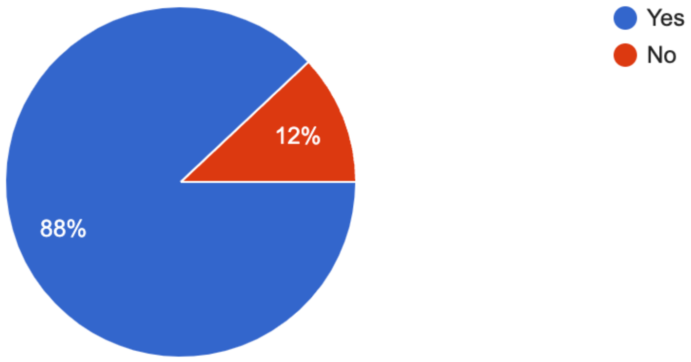
100 responses



Survey results: shopping behavior

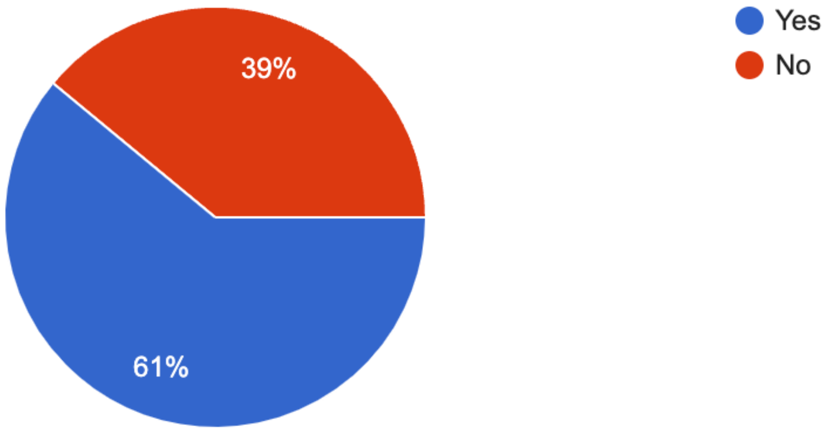
7) Have you ever clicked on a product that was recommended to you on a website / in an app?

100 responses

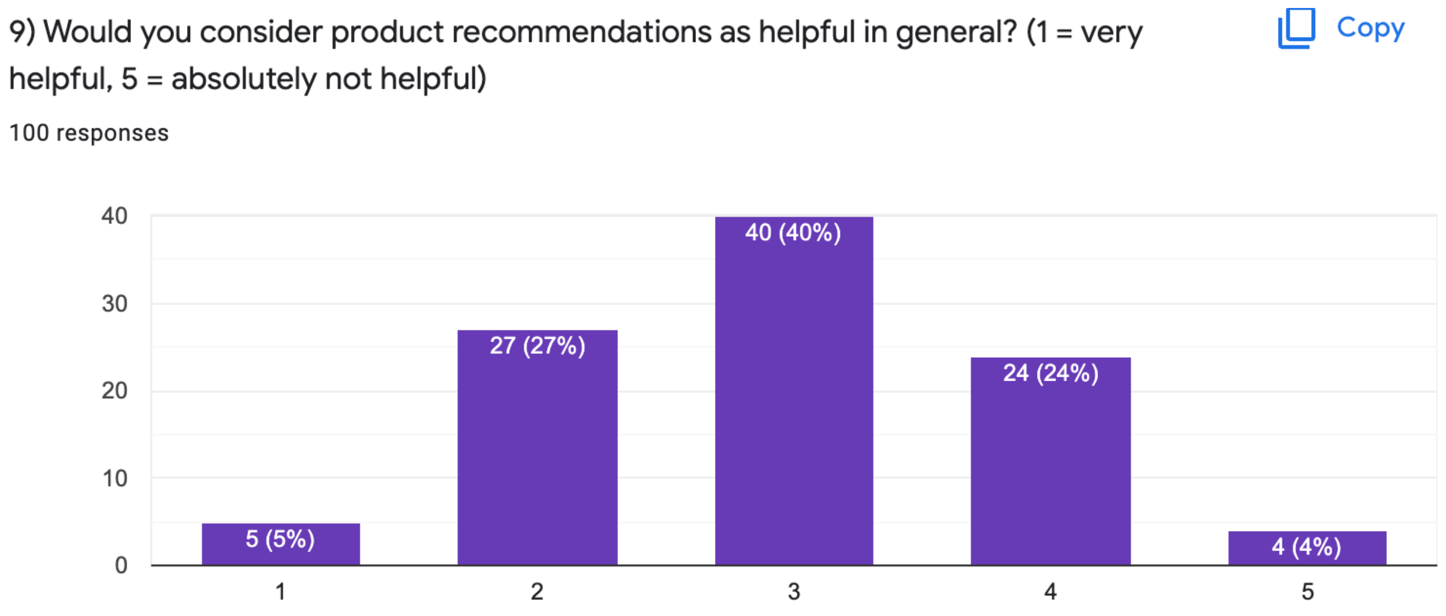


8) Have you ever bought a product that was recommended to you on a website / in an app?

100 responses



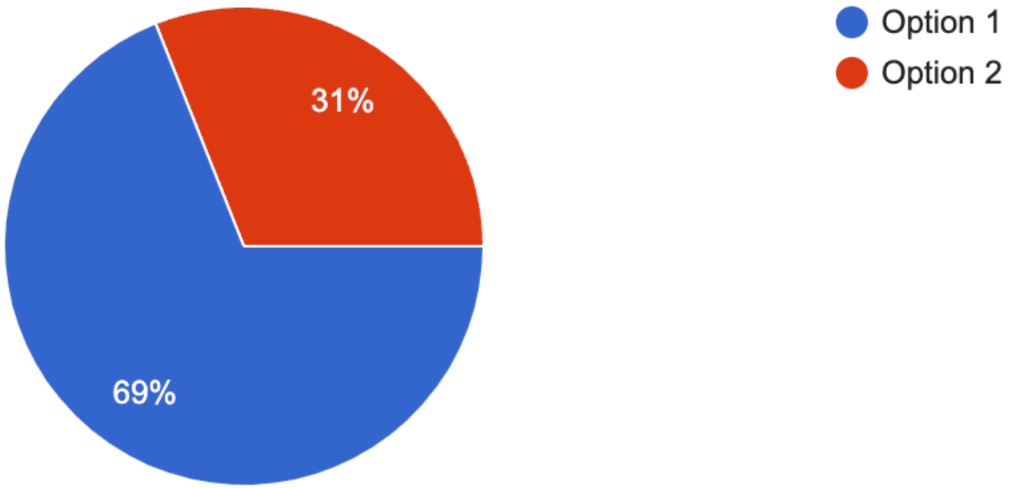
Survey results: product recommendation



Survey results: product recommendation

10.1) Imagine you want to buy new in-ear headphone. Would you rather see recommendations with only high ratings like 4-5 stars (option 1) or mixed ratings like 0-5 stars (option 2)?

100 responses

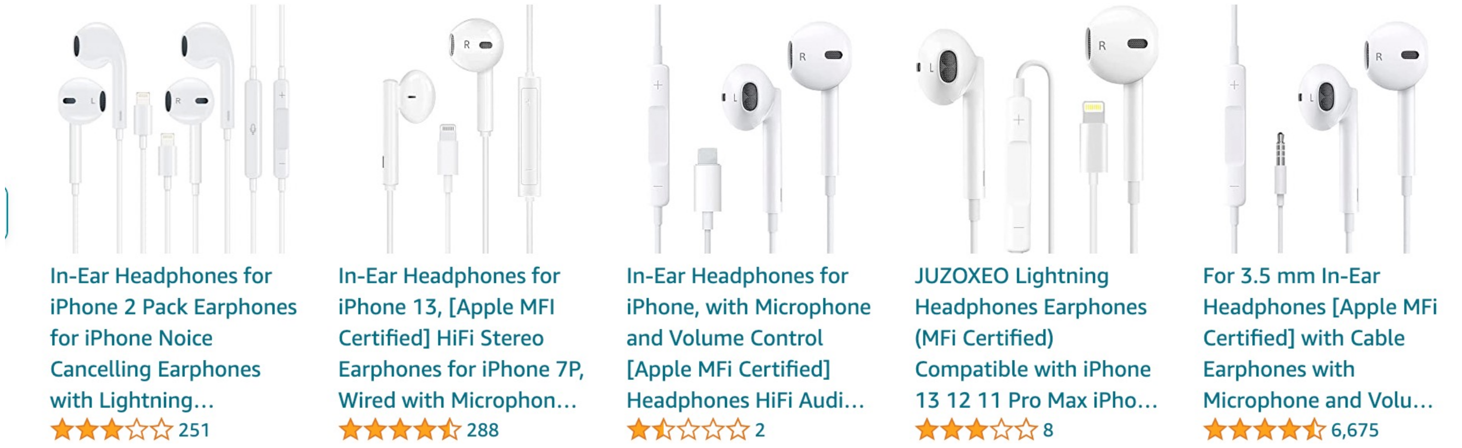


Survey results: product recommendation

Answers for option 2 (diverse ratings 0-5 stars):

- I trust in review by real users.
- It is an opportunity to investigate further by myself.
- It’s unrealistic no one has had issues. I like reading only the bad comments about a product.
- Diverse ratings, diverse experiences = find you own
- Find it hard to believe that a product can be that good so the range from 0-5 feels much more realistic.

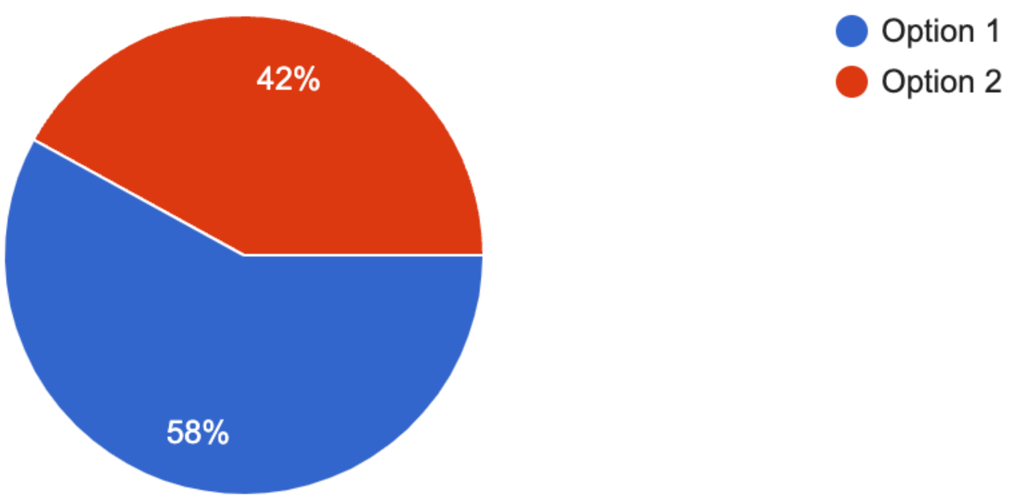
Survey results: product recommendation after improvements



Survey results: product recommendation after improvements

11.1) Imagine the following scenario. You want to buy new in-ear headphones. Compared to the picture below, which optimised recommendation would you consider the most helpful regarding product ratings? (Option 1 is 4-5 stars, Option 2 is 3-5 stars)

100 responses



Survey results: product recommendation after improvements

Four to five stars ratings:

- Higher ratings mostly leads me to the conclusion that the seller is reliable and I won't face problems while buying it.
- Because they're probably better anyway and I dont habe to look through bad headphones.
- Products that have legitimate higher ratings help avoid lesser quality products.
- I want a product which performs above average.
- I'm not interested in low rated products as it normally represents poor user experience.

Three to five stars ratings:

- More authentic.
- I would like to know why people dislike it an see if that a big deal for me or not.
- Some 4 to 5 stars can be fake. I would rather see what others have to offer.
- Higher range of opinions are always better.
- Products are different: some people like them, some people don't. I rather decide on good and bad reviews and my own experience than just on the number of stars.

Conclusion

Conclusion

- proposed recommending high rating by using an optimized PRE
- looked into different product recommendation engines
- decided to use SVD combined & gradient descent to reduce the error margin
- were able to recommend ten products to a randomly chosen user

Conclusion

- survey results with 100 consumers showed us
- 58% consider product recommendations with four to five stars:
  - as helpful for purchase process
  - high quality & user friendly experience
  - a time-saver while trying to find a suitable product for their usage.
- 100 participants is not representative and does not mirror the overall society.

Lessons learned

Lessons learned

- difficulties finding unbiased data towards e-commerce market development
- lost time through overconfidence
  - tunnel vision
  - not every problem can be solved with deep learning
  - only looked at first 300 rows
  - not checking if items are missing features or had none
- should be more open minded

Lessons learned

Future work

- Instagram, TikTok, Youtube testing live shopping services
- trend goes towards in-app purchases
- features should further include:
  - time spent on product / product type page
  - referrer
  - item search order
  - live location of customers
- use deep learning & neural networks for live product recommendation

Source

[1] A. C. Rivera, M. Tapia-Leon, and S. Lujan-Mora, “Recommendation systems in education: A systematic mapping study,” in Proceedings of the International Conference on Information Technology & Systems (ICITS 2018), A. Rocha and ´ T. Guarda, Eds. Springer International Publishing, 2018, pp. 937–947.

[2] M. E. Jennex and L. Olfman, Development Recommendations for Knowledge Management/ Organizational Memory Systems. Springer US, 2001, pp. 209– 222.

[3] B. Xiao and I. Benbasat, “E-commerce product recommendation agents: Use, characteristics, and impact,” MIS Quarterly, vol. 31, pp. 137–209, 03 2007.

[4] Salesforce, “What is a retail product recommendation engine – and why do you need one?” Available at https://www.salesforce.com/resources/articles/ what-is-product-recommendation-engine/ (2022/04/29).

[5] E. Eroglu, “Review of business information systems,” Review of Business Information Systems June 2014, vol. 18, no. 1, pp. 35–40, 2014.

[6] BigCommerce, “Ecomerce - frequently asked questions,” Available at https://www.bigcommerce.com/articles/ecommerce/ #ecommerce-frequently-asked-questions (2022/04/29).

Source

[7] M. A. Rahman, M. A. Islam, B. H. Esha, N. Sultana, and S. Chakravorty, “Consumer buying behavior towards online shopping: An empirical study on dhaka city, bangladesh,” Cogent Business & Management, vol. 5, no. 1, p. 1514940, 2018.

[8] UNCTAD, “Covid-19 has changed online shopping forever, survey shows,” Available at https://unctad.org/news/ covid-19-has-changed-online-shopping-forever-survey-shows (2022/04/30).

[9] Statista, “E-comerce worldwide,” Available at https://www.statista.com/ topics/871/online-shopping/#editorialPicks (2022/04/30).

[10] N. Suisse and UNCTAD, “Covid-19 and e-commerce: Findings from a survey of online consumers in 9 countrieds,” Available at https://unctad.org/system/ files/official-document/dtlstictinf2020d1 en.pdf (2022/04/30).

[11] Tageschau, “Im virtuellen supermarkt,” Available at https://www.tagesschau. de/wirtschaft/lieferdienste-onlinehandel-corona-boom-101.html (2022/04/30).

[12] E. Torres-Moraga, A. V́ asquez-Parraga, and J. Zamora, “Customer satisfaction and loyalty: Start with the product, culminate with the brand,” Journal of Consumer Marketing, vol. 25, pp. 302–313, 08 2008

[13] G. M P, “Research on recommendation systems using deep learning models,” 01 2022.

Source

[14] Serano Academy, “How does Netflix recommend movies? Matrix Factorization”, Available at https://www.youtube.com/watch?v=ZspR5PZemcs (2022/05/13)

[15] Serano Academy, ”Singular Value Decomposition (SVD) and Image Compression”, Available at https://www.youtube.com/watch?v=DG7YTlGnCEo&t=106s (2022/05/13)

[16] A. Roya, “Introduction to recommender systems- 1: Content-based filtering and collaborative filtering,” Available at https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421 (2022/04/30).

[17] OldMonk, “Kaggle: Amazon electronics data.” Available at https://www.kaggle.com/datasets/saurabhbagchi/amazon-electronics-data (2022/05/30)

[18] P. Sharma, “Comprehensive guide to build a recommendation engine from scratch (in python).” Availble at https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/ (2022/06/20).