

---

---

# NEWS Recommender Engine

— For - Jhakaas News Vala —

---

---

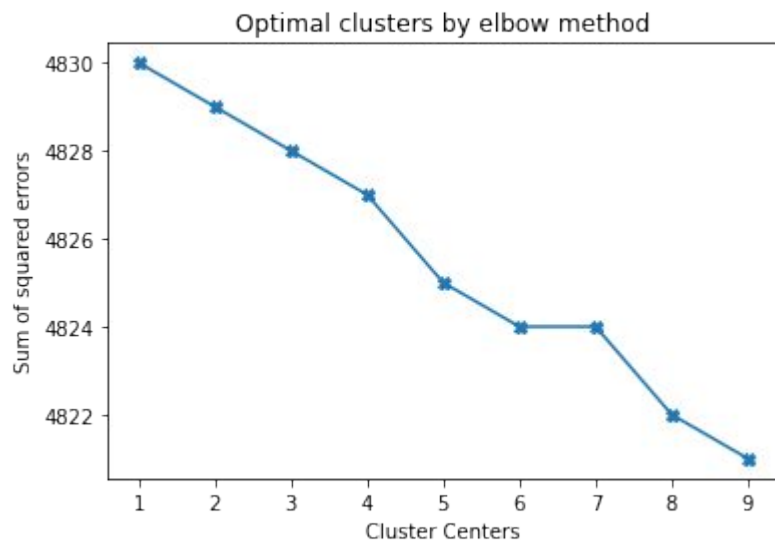
# The news articles

- Total number of articles = 4830
- Column headers : Article\_Id, Title, Author, Date, Content, URL
- Percentage of missing data:
  - URL = 0.2984
  - Author = 0.0012
  - Date = 0.001
  - Content = 0
  - Title = 0
- A few titles : <right column>

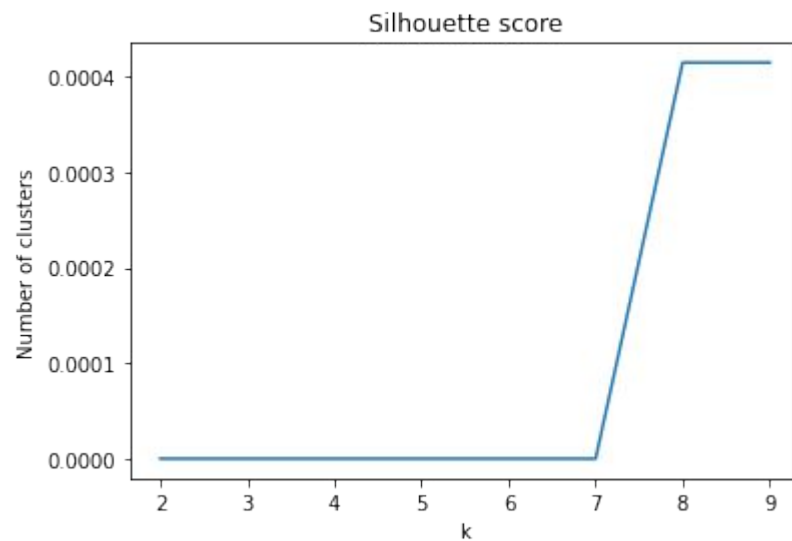
1. '14 dead after bus falls into canal in Telangana s Khammam district Andhra CM promises Rs 3 lakh compensation'
2. 'Pratibha Tiwari molested on busy road Saath Nibhana Saathiya actress drags accused to police station',
3. 'US South Korea begin joint military drill amid nuclear threat from North Korea',
4. 'Illegal construction in Bengaluru Will my house be demolished next ask worried residents',
5. 'Punjab Gau Rakshak Dal chief held for assaulting sodomising cattle traders',
6. 'Phillipines drug war 1 800 drug-related deaths since June-end'
7. 'Infosys shares likely to fall on Tuesday after company s client RBS scraps Williams Glyn project'
8. 'Dialogue crucial in finding permanent solution to Kashmir s crisis PM Modi'
9. 'School bus overturns in Jammu killing 1 and injuring 16 children',
10. 'Rajasthan Villagers rescue 50 kids after school bus falls into river in Bhilwara'

# Data clusters

**Elbow curve**



**Silhouette score**



# Recommendation Algorithms

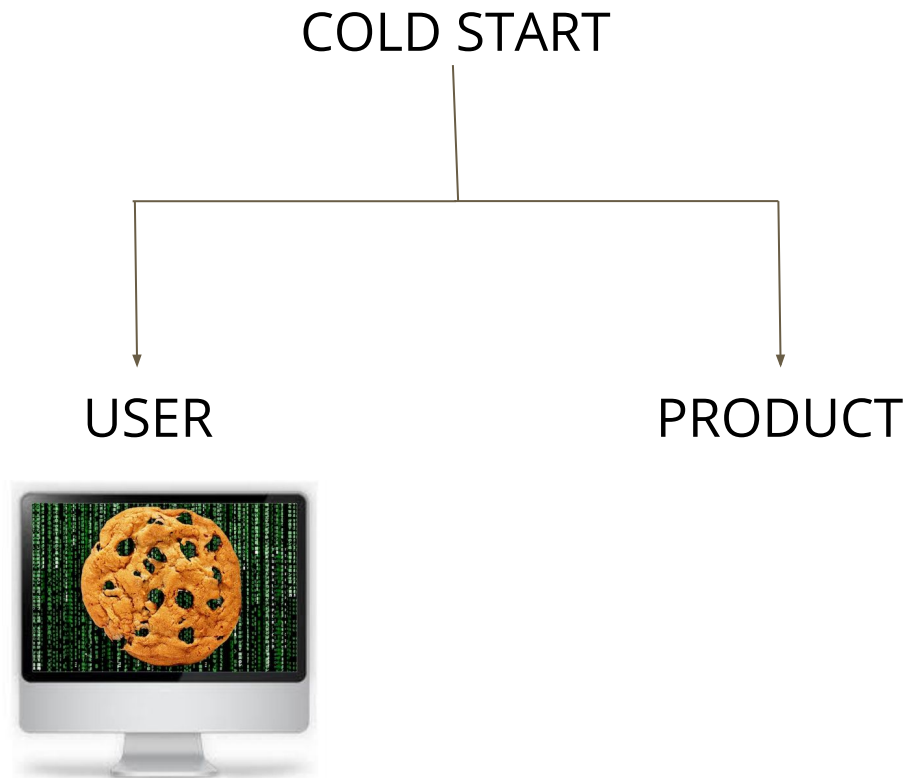
# Collaborative Filtering

- Ratings of other users are considered rather than attributes data
- Example - Synthetic control



# Cold Start Problem

- Circumstances are not yet optimal for the engine to provide best possible results
- User - The recommender engine meets the user for the first time and has no history about him / her
- Product -



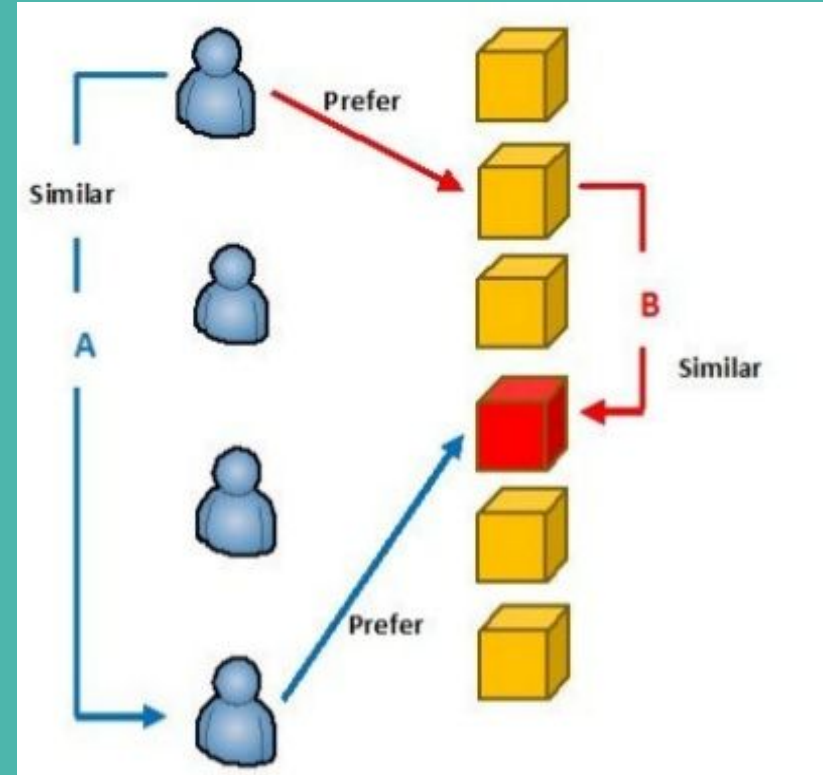
# Content Based

- Focuses on mapping user ratings against attributes of products i.e. recommend articles based on user profile
- Example - Facebook, LinkedIn



# Hybrid Approach

- Combines the two approaches to give more accurate results
- Example -Netflix





# Bottle Neck Problem

- A point of congestion that occurs when workloads arrive too quickly for the production process to handle
- Why is it a problem?
  - Increases the time and expense of production
  - Create delays and higher production costs
- Can be resolved easily by allocating more resources or people

# The Recommender Engine

# User Profile

Userld	Sessionld	ArticleID Served	Article Rank	Click	Time Spent (seconds)
1	1	1929	1	True	18
1	1	195	2	True	52
1	1	31	3	False	0
1	1	2814	4	True	67
1	1	26	5	False	0

Content Based

# Steps

1. Build user profile
2. Tfidf\_matrix
3. Cosine similarity using linear kernel
4. Obtaining a random input
  - a. Passing its index to the engine
5. Fetching the top ten articles

Article_I d	Title	Author	Date	Content	URL	Article Rank	Click	Time Spent (seconds)	
1	1	Pratibha Tiwari molested on busy road Saath ...	Suparno Sarkar	August 22, 2016 19:47 IST	TV actress Pratibha Tiwari who is best known ...	NaN	2	True	79
2	2	US South Korea begin joint military drill ami...	Namrata Tripathi	August 22, 2016 18:10 IST	The United States and South Korea began a join...	http://www.ibtimes.co.in/us-south-kor ea-begin-...	3	True	75
5	5	Phillipines drug war 1 800 drug-related death...	Pranshu Rathee	August 22, 2016 14:51 IST	Philippines police on Monday said that the num...	NaN	6	True	12
6	6	Infosys shares likely to fall on Tuesday after...	S V Krishnamachar i	August 15, 2016 16:03 IST	Infosys shares could fall sharply on Tuesday w...	http://www.ibtimes.co.in/infosys-share s-likely...	7	True	85
8	8	School bus overturns in Jammu killing 1 and in...	Devyani Sultania	August 10, 2016 11:51 IST	A school bus in the town of Bishnah in Jammu s...	http://www.ibtimes.co.in/school-bus-o verturns-...	9	True	35

## Recommendations

Article chosen by user -

**1555 Shah Rukh Khan s  
role in Aditya Chopra s  
next....**

1811 Will Aditya Chopra Bring Shah Rukh Khan and Ra..

1791 Shah Rukh Khan s Raees First Look Revealed ...

1060 Year 2014 For Amir Khan Shah Rukh Khan and S ...

1581 Shah Rukh Khan to romance Kareena Kapoor Khan

1763 Shah Rukh Khan s character in Gauri Shinde s n ...

1554 Katrina Kaif or Jacqueline Fernandez to replac ...

1257 Shah Rukh Khan Priyanka Chopra beat Salman Kh .

1582 Revealed Shah Rukh Khan to play a dwarf with ...

1712 Dangal release date changed Is this Amir K ...

1779 Shah Rukh Khan Anushka Sharma to star in Imti ...

# Collaborative Filtering



# Steps

1. Create ratings matrix for different users
2. Matrix factorization
  - a. User\_pivot matrix
  - b. Csr\_matrix
3. Making recommendation
  - a. Pass user id as a parameter
  - b. Find article id in news article dataset
  - c. Recommend top 10

# Ratings

Userld	Sessionld	Article Rank	Time Spent (seconds)	Ratings (0 - 5)
1	1	1	30	4
1	1	3	78	1
1	1	7	54	2
1	1	8	32	2
1	1	2	16	4
.....	.....	.....	.....	.....
4827	4812	8	55	1

## Matrix Factorization

- User Pivot matrix
- csr matrix -  
Compressed Sparse  
Row matrix represents  
a matrix by three one  
dimensional arrays  
containing non zero  
values

```
[[[4., 1., 2., ..., 0., 0., 0.],  
 [0., 0., 0., ..., 0., 0., 0.],  
 [0., 0., 0., ..., 0., 0., 0.],  
 ...,  
 [0., 0., 0., ..., 1., 0., 0.],  
 [0., 0., 0., ..., 0., 2., 0.],  
 [0., 0., 0., ..., 0., 0., 1.]]]
```

## Recommendations

User id : 1555

1137 Coolpad lanches Mega2 5D selfie smartphone f ..

116 Tollywood 2015 Top Telgu controversies relat ...

1651 Rohan Bopanna qualifies for Rio Olympics to d...

1648 Rio 2016 Leander Paes will team up with Rohan ..

1646 Wimbledon 2016 Mixed day for India as Sania a ..

1645 Sania Mirza and Martina Hingis crash out of Wi ...

1642 Manchester United and Jose Mourinho will give ...

1641 EXCLUSIVE MMA star Sascha Sharma dreams of  
ma ...

1640 Pro Wrestling League 2015 auction Full list o ...

1639 Pro Wrestling League 2015 Mumbai Garuda becom

# Work for future

- A mobile application for recommending news articles to readers can be developed