Flipkart GRID 4.0

2022 Campus Challenge

Extract Trends from social media data

Team Name: noobie_learner

Institute Name: National Institute of Technology, Patna

Team members details

Team Name	noobie_learner		
Institute Name	National Institute of Technology, Patna		
Team Members >	**1 (Leader)**	**2**	**3**
Name	Ayush Kumar	Ayush Gupta	Saurabh Kumar Patil
Batch	2019-2023	2019-2023	2019-2023

Experimental Set

- Python Version used – 3.7.13

- Python libraries used
    - numpy
    - pandas
    - sklearn
    - seaborn
    - matplotlib

Glossary

- Introduction
- Implementation Details
- Result and Observation
- Future Scope

Introduction

1.Recommendation System

- Recommendation system is an informative filtering technique,which provides users with information, which he/she may be interested in.

2.Best Discounts

- Which product categories have the best discounts?

Implementation Details

Areas of Use

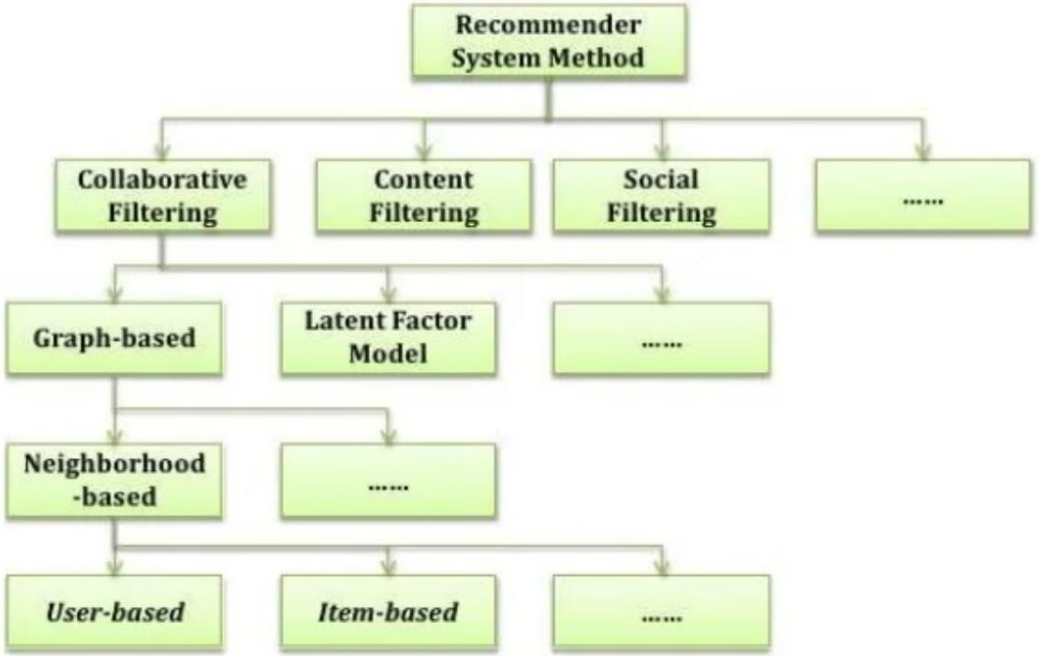
Why there is a need?

- “Getting information off the internet is like taking a drink from a fire hydrant” - Mitchell Kapor

- Information Overload
- User Experience
- Revenues

Recommendation systems help in addressing the information overload problem by retrieving the information desired by the user based on his or similar user’s preferences and interests.

Recommender Algorithms



Dataset

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	image	is_FK_Advantage_product	descri
0	c2d766ca982eca8304150849735ffef9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2FF9KEDEFGF	999.0	379.0	["http://img5a.flixcart.com/image/short/u/4/a/...	False	Key Featu Allisha Wo Cyclir
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY	32157.0	22646.0	["http://img6a.flixcart.com/image/sofa-bed/j/f...	False	FabHome Fabric D Sof (Finist
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUBJGZXE	999.0	499.0	["http://img5a.flixcart.com/image/shoe/7/z/z/r...	False	Key Featu AW E Sa Wedges I
3	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2F6HUZMQ6SJ	699.0	267.0	["http://img5a.flixcart.com/image/short/6/2/h/...	False	Key Featu Alisha Wo Cyclir
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMSYARJ5	220.0	210.0	["http://img5a.flixcart.com/image/pet-shampoo/...	False	Specific of Sicc Purpose

Preprocessing

Removal of unwanted data

- Punctuation marks
- Stop word removal
- Lemmatization
- Stemming

Collaborative Filtering (CF) Approaches

- K-nearest neighbor (KNN)
  - User based methods
    - Neighborhood formation phase
    - Recommendation phase

  - Item based methods
    - Recommendation phase
- Association rules based prediction

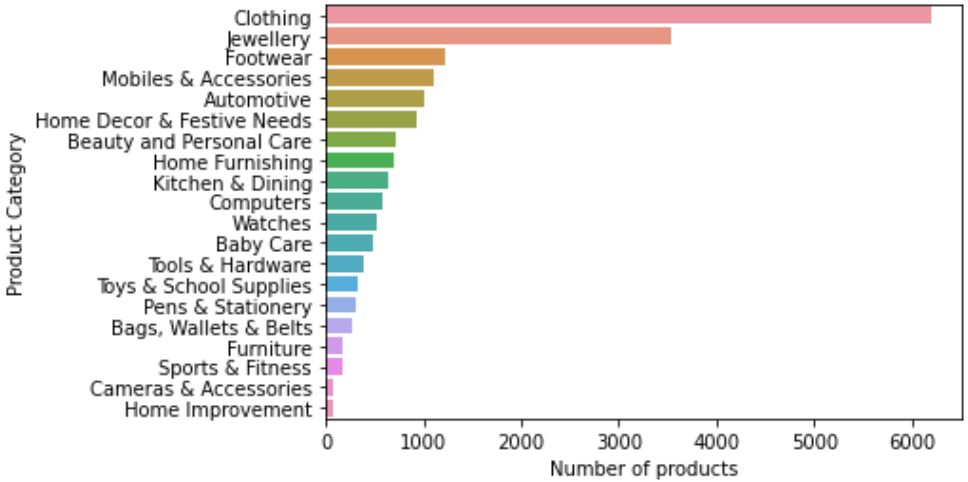
- Matrix factorization

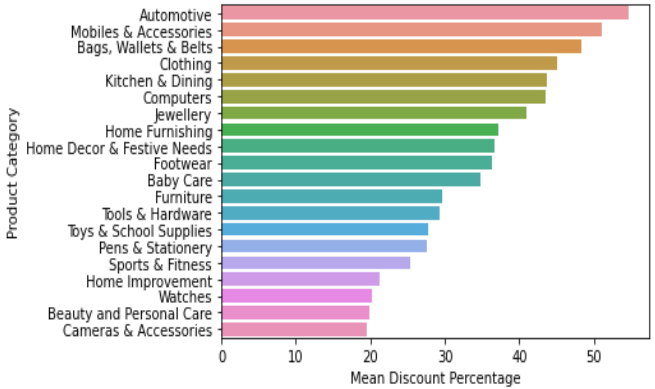
Result and Observations

primary_category	discount_percent_mean	primary_category_count
Clothing	45.070525	6198
Jewellery	40.889934	3531
Footwear	36.268531	1227
Mobiles & Accessories	50.994470	1099
Automotive	54.650991	1012
Home Decor & Festive Needs	36.649404	929
Beauty and Personal Care	19.829674	710
Home Furnishing	37.218050	700
Kitchen & Dining	43.771695	647
Computers	43.467316	578
Watches	20.170555	530
Baby Care	34.717297	483
Tools & Hardware	29.330164	391
Toys & School Supplies	27.743839	330
Pens & Stationery	27.651471	313
Bags, Wallets & Belts	48.272196	265
Furniture	29.721585	180
Sports & Fitness	25.282278	166
Cameras & Accessories	19.441131	82
Home Improvement	21.262268	81

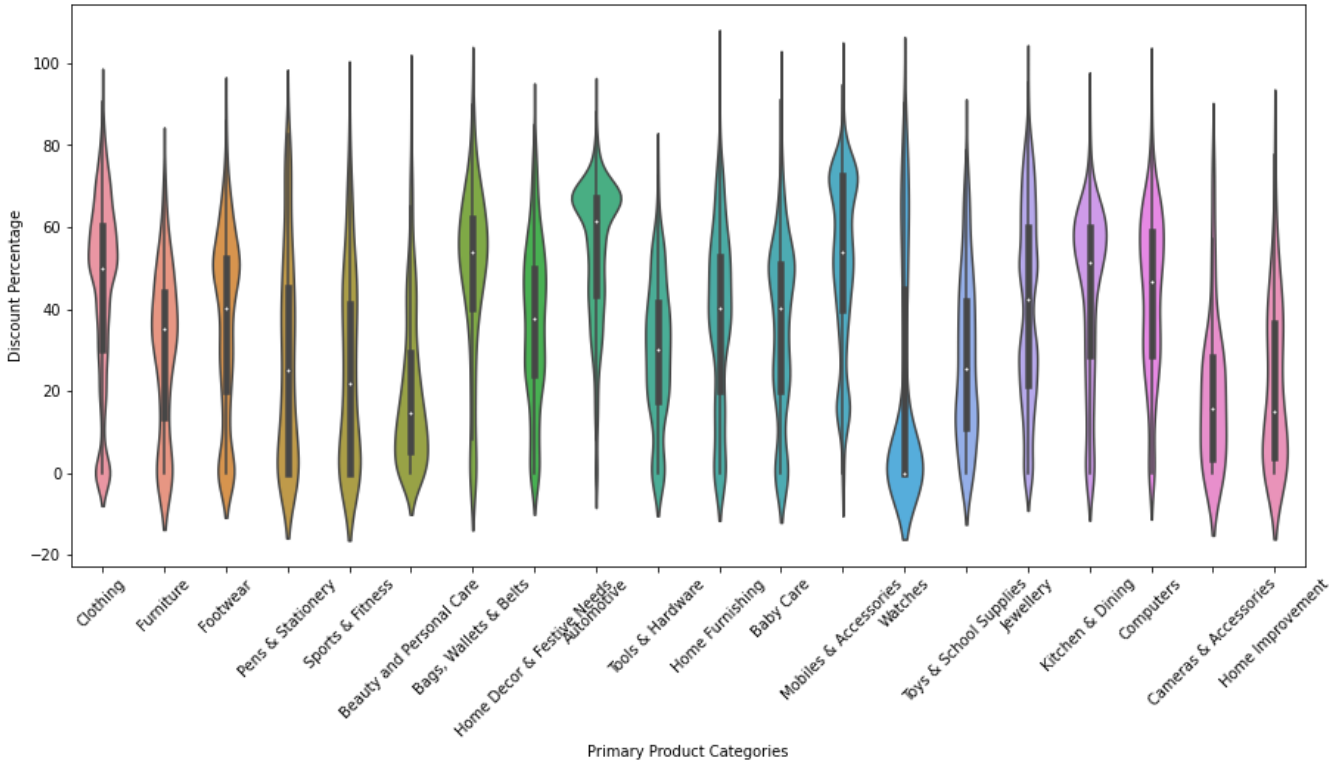
```
        primary_category  discount_percent_mean  primary_category_count
0               Clothing              45.070525                    6198
1              Jewellery              40.889934                    3531
2               Footwear              36.268531                    1227
3  Mobiles & Accessories              50.994470                    1099
4             Automotive              54.650991                    1012
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 3 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   primary_category       20 non-null     object
 1   discount_percent_mean  20 non-null     float64
 2   primary_category_count  20 non-null     int64
dtypes: float64(1), int64(1), object(1)
memory usage: 608.0+ bytes
None
       discount_percent_mean  primary_category_count
count              20.000000               20.000000
mean               34.620169              972.600000
std                10.885227             1438.728657
min                19.441131               81.000000
25%                27.059173              301.000000
50%                35.492914              554.000000
75%                43.543411              949.750000
max                54.650991             6198.000000
```

Cont…


Clothing
Jewellery
Footwear
Mobiles & Accessories
Automotive
Home Decor & Festive Needs
Beauty and Personal Care
Home Furnishing
Kitchen & Dining
Computers
Watches
Baby Care
Tools & Hardware
Toys & School Supplies
Pens & Stationery
Bags, Wallets & Belts
Furniture
Sports & Fitness
Cameras & Accessories
Home Improvement
Product Category
0
1000
2000
3000
4000
5000
6000
Number of products


Automotive
Mobiles & Accessories
Bags, Wallets & Belts
Clothing
Kitchen & Dining
Computers
Jewellery
Home Furnishing
Home Decor & Festive Needs
Footwear
Baby Care
Furniture
Tools & Hardware
Toys & School Supplies
Pens & Stationery
Sports & Fitness
Home Improvement
Watches
Beauty and Personal Care
Cameras & Accessories
Product Category
0
10
20
30
40
50
Mean Discount Percentage

Cont…


Discount Percentage
100
80
60
40
20
0
−20
Clothing
Furniture
Footwear
Pens & Stationery
Sports & Fitness
Beauty and Personal Care
Bags, Wallets & Belts
Home Decor & Festive Needs
Automotive
Tools & Hardware
Home Furnishing
Baby Care
Mobiles & Accessories
Watches
Toys & School Supplies
Jewellery
Kitchen & Dining
Computers
Cameras & Accessories
Home Improvement
Primary Product Categories

Future Scope

- The future of recommendation systems in e-commerce is likely to be more personalized and tailored to the individual.

- As data science advances, retailers are able to collect more data on customers' shopping habits and preferences. This allows them to create a more detailed profile of each customer, which can then be used to personalize the recommendations that are made.

- In addition, as machine learning becomes more sophisticated, recommendation systems will become better at understanding how individual customers interact with different products and what kind of products they are likely to be interested in. This will allow retailers to provide more accurate and relevant recommendations for each customer.