Flipkart

GRiD 4.0

2022 Campus Challenge

Extract Trends from social media data

Team Name: Unstoppable

Institute Name: JSS Academy of Technical Education, Noida

Team members details

Team Name	Unstoppable		
Institute Name	JSS Academy of Technical Education, Noida		
Team Members >	1 (Leader)	2	3
Name	VANSH CHOPRA	DEVANG MISHRA	WAMA NIGAM
Batch	2024	2024	2024

Deliverables/Expectations for Level 2 (Idea + Code Submission)

Deliverable 1:

Identification of trends from social media

1. Identify trends on social media based on category. Can restrict to Fashion as a category for the project. Ex: Polka dots dresses are trending on twitter.
2. Ranking/scoring logic for trends extracted.
3. Outcome format:
   a. Option1: List of trending keyword(s) along with list of sample images and respective links from which the trend is derived with most trending first:
      Example: Trends:[{Polka dot dresses, <list of links/images>,trending score}, {Bellbottom Jeans, <list of links/images>,trending score}..]
   b. Option 2: structured data according to flipkart category, sub category, vertical and product attributes
      Example: {category: Fashion, Sub-category: Women Western, vertical: Women dresses, trending attribute type: Pattern, trending attribute value: Polka Print, list of sample images and links from which the trend is derived}.
      Outcome with Option 2 format will be given bonus points.

Deliverable 2:

Mapping trends with Flipkart products:

1. Create mapping of extracted trending keyword(s) with Flipkart category, sub category, vertical and product attribute(s), search page links.
   Example:{category: Fashion, Sub-category: Women Western, vertical: Women dresses, trending attribute type: Pattern, trending attribute value: Polka Print}
   Note: Use category, Subcategory combination from the Flipkart Website
2. From a trending keyword, creating a corresponding searchable term on Flipkart which will lead to matching products.
   Example: Tropical Tops keywords will not give right results directly on Flipkart but we can construct search query for it using some intelligence.
3. Points will be given based on similarity between sample images for trends and product results on Flipkart.

Use-Cases:
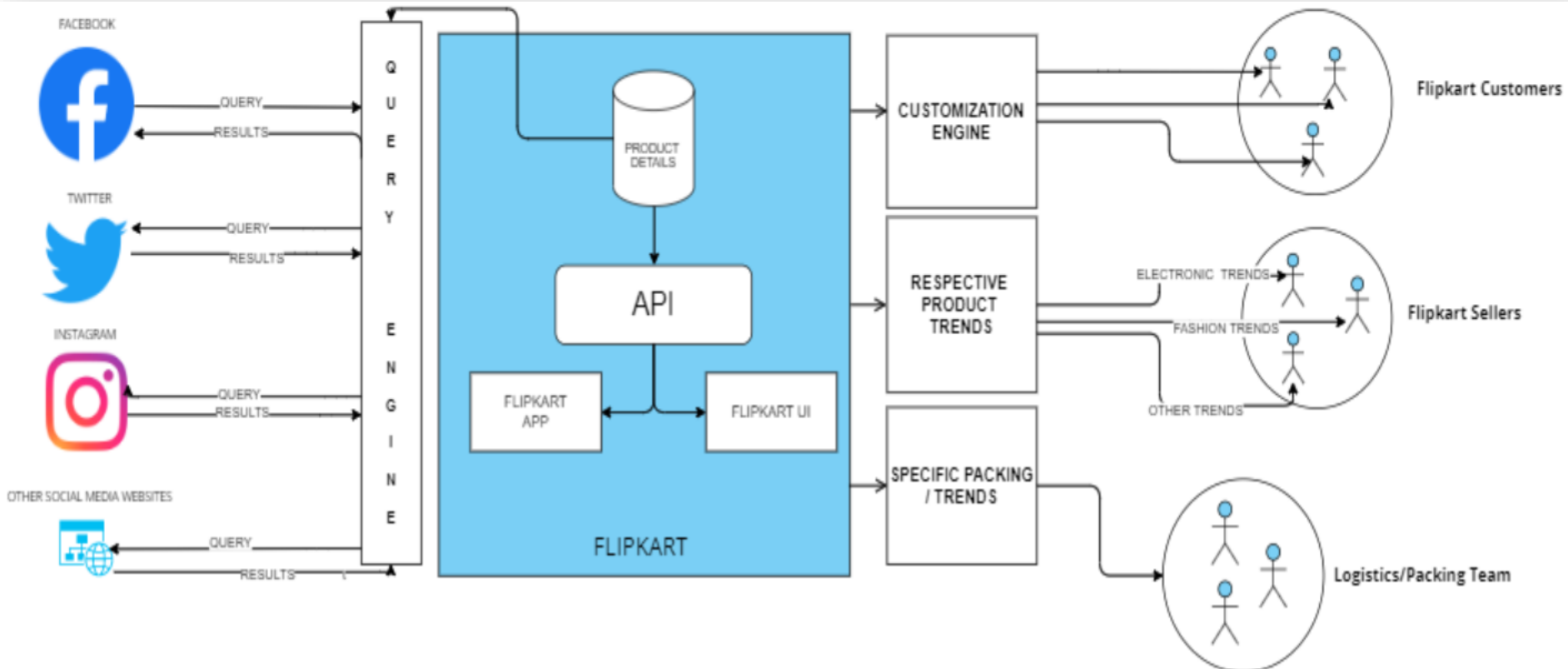
Use-cases

Customer

- Customer will stay updated with the trend.
- Location specific trends will be shown.
- Easy navigation to different sub-category assist the customer to make an easy choice.
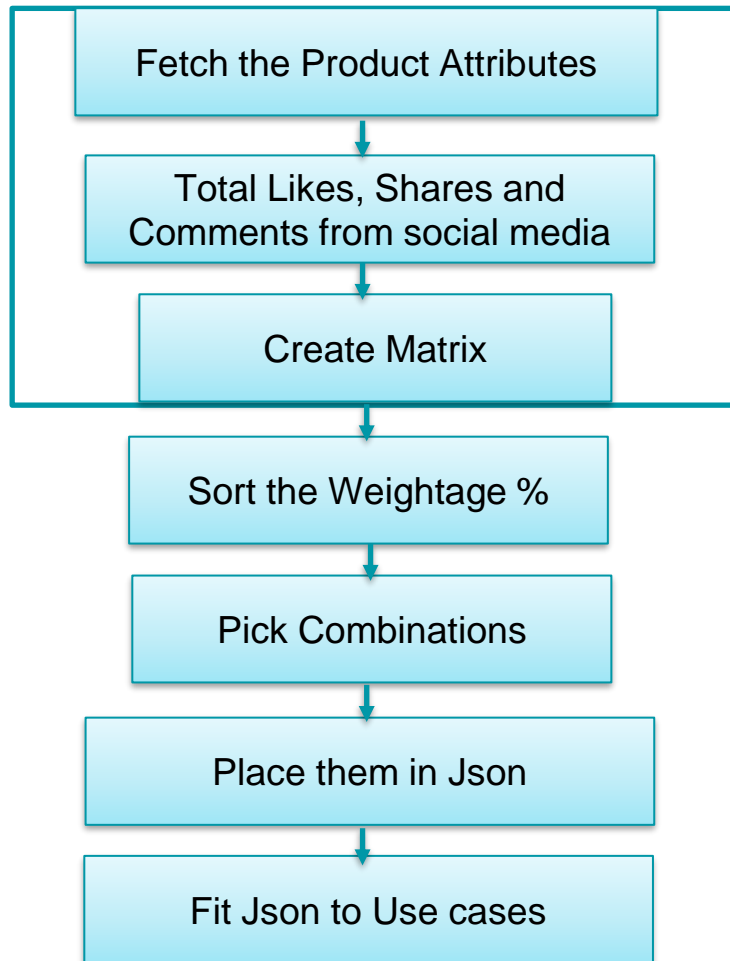
Seller

- Trending data will be shared to sellers so as to increase their sale.
- Location specific trends help the sellers to cover more market in their vicinity.
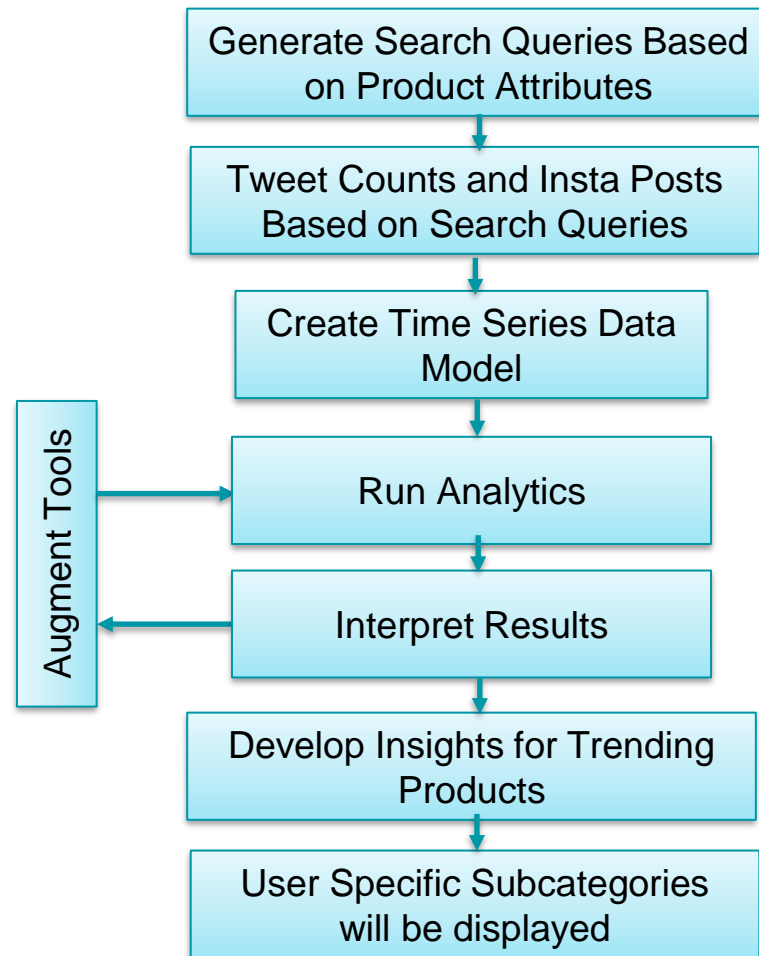- Product images assist the seller to quickly identify the products.

Flipkart

- Easy to search trending products on flipkart.
- Prediction of trending products for the next day based on model.
- Default trending subcategories will be there for new users.

Approach-1:

- Fetch the Product Attributes
- Total Likes, Shares and Comments from social media
- Create Matrix
- Sort the Weightage %
- Pick Combinations
- Place them in Json
- Fit Json to Use cases

Approach-2:

- Generate Search Queries Based on Product Attributes
- Tweet Counts and Insta Posts Based on Search Queries
- Create Time Series Data Model
- Augment Tools
- Run Analytics
- Interpret Results
- Develop Insights for Trending Products
- User Specific Subcategories will be displayed

HOW THE MATRIX WILL BE GENERATED?

Subcategory- Top Wear   Most Trending Query:- Pattern: Printed, Color: Multi-Color, Brand: Allen Solly

Attributes / Trend Count	PATTERN			COLOR			BRAND			
	Solid	Polka Print	Printed	Black	White	Multi color	Puma	Nike	Allen Solly	Van Heusen
Twitter	(1000/10000)*100 =10	(1867/8000)*100= 23.33	(5670/10000)*100 =56.7	(5067/9000)*100= 56.3	(2956/4000)*100= 73.9	(1999/3000)*100= 66.63	(2967/5000)*100= 59.34	(567/3000)*100= 18.9	(3567/6000)*100= 59.45	(1567/2000)*100= 78.35
Instagram	(1156/5000)*100= 23.12	(1567/6000)*100= 23.11	(2767/7000)*100= 39.52	(4567/8000)*100= 57.08	(5467/8000)*100= 68.33	(1674/5000)*100= 33.48	(3577/7000)*100= 51.1	(957/4000)*100= 23.92	(4567/7000)*100= 65.24	(1597/3000)*100= 53.23
Facebook	(189/1000)* 100= 18.9	(1967/5000)*100= 39.34	(2526/5000)*100= 50.52	(3567/7000)*100= 50.95	(567/10000)*100= 5.67	(5627/8000)*100= 70.33	(1967/6000)*100= 32.78	(1587/5000)*100= 31.74	(6567/8000)*100= 82.08	(507/3000)*100= 16.9
Average %	17.3	28.59	**48.91**	54.77	49.3	**56.81**	47.74	24.85	**68.92**	49.49

Solution statement/Proposed approach

APPROACH-1:
- For each category and subcategory, men top wear, fetch the product attributes / filter from flipkart. Example: Pattern, Color and Brand
- Search Twitter data by these attributes with below function:
  Tweepy module/client.get_recent_tweets_count(query=query, granularity='day')
- Similarly fetch the data from other social media platforms like Instagram, facebook, youtube, etc.,
- Get the Trend Count by summing the total number of likes, shares and comments on this data.
- Make the matrix consisting of Trend Count as rows and Product Attributes as columns.
- Calculate the percentage of trending products for each social media platform.
- Take the weightage for each social media platform.
- Calculate the final avg % and sort them in each category.
- Pick those combinations and place them in a json.
- This json data can be fit into use cases mentioned above.

APPROACH-2 (PREDICTIVE MODEL):

- For each category and subcategory, men top wear, fetch the product attributes / filter from flipkart. Example: Pattern, Color and Brand
- From Twitter, we get the tweet counts of 1 week (configurable) for each search query using below function:
  Tweepy module/client.get_recent_tweets_count(query=query, granularity='day')
- Using the time series model we plot the tweet counts and predict the tweet counts of the next day.
- Map the tweet counts with the product query using the hashtable.
- Sort the tweet counts and then add the product corresponding to it in the trending list from hashtable.
- This way we get the trending products in each subcategory like top wear from Twitter.
- From Instagram, we get total posts from the accounts of famous fashion designers and celebrity for each search query.
- Similarly data can be extracted from multiple social media platforms.
- If there are n social media platforms, take (100/n)% weightage of trending data from them.
- Display the subcategories and its trending list of products.

HOW SUBCATEGORIES WILL BE DISPLAYED?

- In case of existing user, based on the user history, cart and wish list subcategories(like top wear, bottom wear, etc.,) will be displayed.
- In case of new user, we will ask preferences from the user and display the subcategories accordingly.

Limitations

- Data collection from Twitter, Instagram, Facebook can be made more effective with higher level of access.
- New social media platforms needs to be included time to time to make the model more robust.
- Associated cost of data access from different social media platforms

Future Scope

- We will try to integrate more data from different sources like different trending fashion blogs to make our model more robust.
- Only those images will be displayed that matched with those from flipkart database.
- Videos and gifs of the trending products can also be shown.