

## **Extract Trends from social media data**

Team Name: The Smurfs

Institute Name: Maharaja Agrasen Institute of Technology

# Team members details

Team Name	The Smurfs		
Institute Name	Maharaja Agrasen Institute of Technology , MAIT, Delhi		
Team Members >	1 (Leader)	2	3
Name	RITIK SINGLA	KUSHAGRA SHARMA	RAVI GUSAIN
Batch	IT-2023	IT-2023	IT-2023

# 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.

# Glossary

- Describe/ Expand abbreviations if you have used any in the slides below

**Taxonomy:** It defines the things that are most important to the business related to people, products, services, customers, processes and the value-creating mechanisms at the heart of the enterprise.

**Product Taxonomy :** Product attributes are **the properties that describe a product**.  
They include details that are tangible and intangible, subjective and objective.  
All of this information enables shoppers to find, compare, and choose products.

**PIP:** is a web page on an eCommerce site that presents the description of a specific product in view.  
The details displayed often include size, color, price, shipping information, reviews, and other relevant information customers may want to know before making a purchase.

# Instructions (You Can Delete this Slide)

Dear Team,

Congratulations on reaching this stage - We look forward to some amazing & innovative solutions.

Please find some important instructions before you begin to prepare your submission decks.

Slide Limit : 10 Slides of Content **post (after)** this Slide  
Saving Format : Save the file as a PDF to ensure your formatting remains intact  
Submission Guide: Only the '**Team Leader**' will be able to submit the Deck.  
Only the latest submission will be considered as final  
(You can keep updating your deck within the deadline)

Wishing you all the very best !

**Team Flipkart GRiD**

## Hierarchical Structure

OF TRENDING FASHIONABLE ITEMS



For the purpose of predicting fashion trends we studied E-commerce and social media websites. Fashion data is divided in a hierarchical structure or Category Taxonomy.

### Root Node – Category

**Level 1 Children** - Male clothing, Female Clothing and Kids Clothing

**Level 2 Children** - All the products .Hence in this way data is stored and traversed

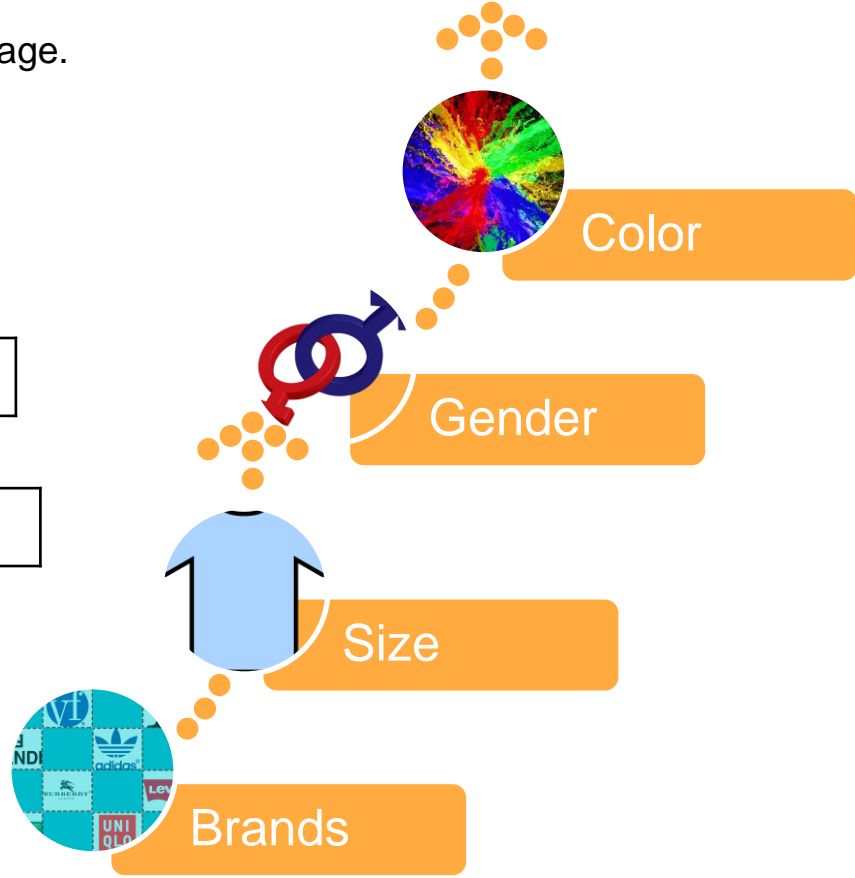
# Attribute Taxonomy

After the category taxonomy we reach the PIP Page.  
Here there is an attribute taxonomy . Here the  
details of the product is ingrained.

Relational Database:

Parent ID	Child ID
-----------	----------

Child ID	Name	Root
----------	------	------



With Respect to the aside picture data  
Here Category Taxonomy works  
In following manner  
Clothing (root Node)

Men Clothing (Level 1)

Top-Wear (Level2)

Tshirts (Level 3)

Home / Clothing / Men Clothing / Tshirts / Difference Of Opinion Tshirts > More By Difference Of Opinion



### Difference of Opinion

Men Mint Green Round Neck Drop-Shoulder Sleeves T-shirt

4.1 ★ | 691 Ratings

Rs. 571 ~~Rs. 1299~~ (56% OFF)

inclusive of all taxes

SELECT SIZE [SIZE CHART >](#)

S M L XL XXL

ADD TO BAG

WISHLIS

DELIVERY OPTIONS

Enter pincode [Check](#)

Please enter PIN code to check delivery time & Pay on Delivery Availability

After this Attribute Taxonomy comes into the picture  
Here Attribute Taxonomy depends on category taxonomy to categorize product in more detail  
It works in following manner

Attribute taxonomy has the detail of the product

For example

Size Ratings Brands Color

According to the given picture data

Here sizes available are S,M,L,XL,XXL

Rating is 4.1 ★ ,Brand is DoO , Color is Mint Green



# Trending Depends on two Criteria's

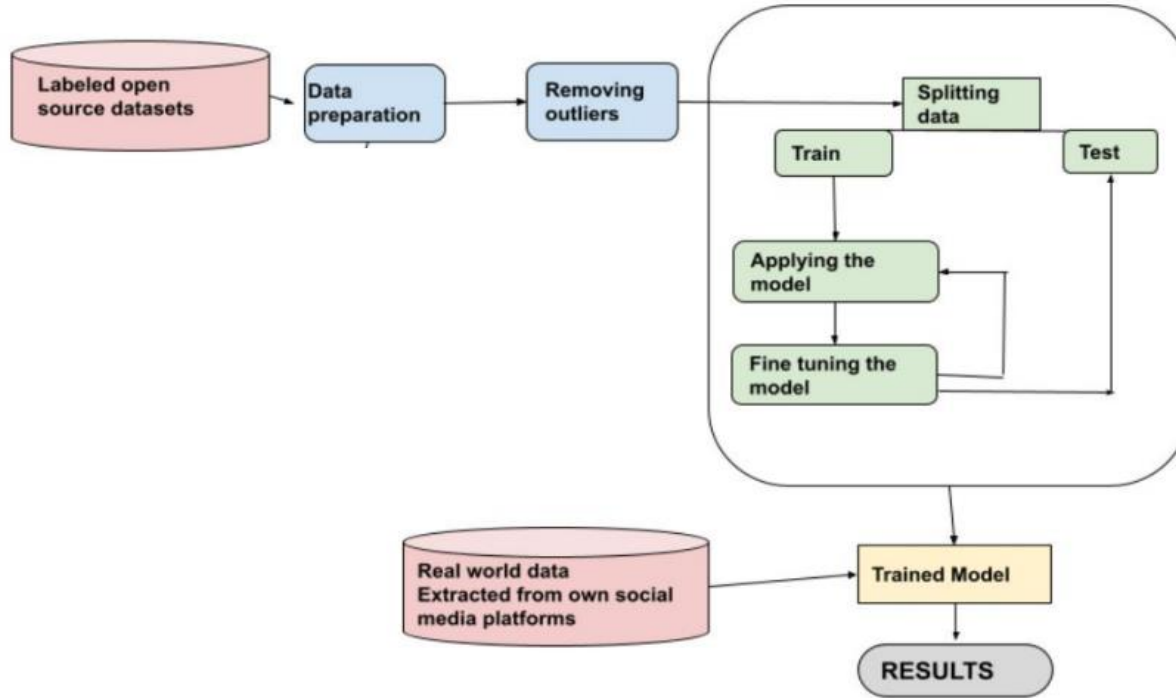
## Browse Ranking

- Also called Landing Page optimization
- it is the process of enhancing or improving each element displayed to the user in such a way as to increase the page's likability.
- Instead of designing the page on the bases of static inputs and variables, the data is dynamically chosen according to the audience and statistical data.
- By observing what the users are interested in seeing and catering to those needs, the website's conversion rate can be improved.

## Searching Querying

- It is the process of improving the in-site results/data in such a way as to be in accordance with the data/keywords provided by the user.
- This allows the website to be more easily accessible to the user.
- This can be done for example by improving autocomplete, prioritizing attractiveness as well as relevance, using AI tools to understand search language better, etc.

# Solution statement/ Proposed approach flow diagram



```
graph TD; A[Extracting Image Links from data frame] --> B[Object Detection by a res net model and annoated the images]; B --> C[Applied feature learning on the images and extracted the features];
```

**Extracting Image Links from  
data frame**

**Object Detection by a res net  
model and annoated the images**

**Applied feature learning on  
the images and extracted  
the features**

# Dataset that is used to train our model

Extraction of data from websites : Data extraction is done by web-scraping.

In our case using Python Selenium . We selected a few websites such as Flipkart, Amazon and Pintrest to scrape

the latest fashion trend data from.

The Base URL was obtained manually by going onto the website and searching for keywords such as "Men's Wear", "Sports Wear", etc.

These URLs are then put into Python scripts that then extract the Name, Category, Rating, Number of reviews and

Image Link from each of the recommended choices on the page.

From this data, a .csv file is created which is then further processed by an ML script.

# Use-cases

- P0: Searching Optimization and Landing Page Optimization.
- P1: Feature learning for the fashion trends
- P2: hierarchical feature representation learning
- P3: Image description retrieval

# Limitations

- ★ In using the social media data or social data a lot of data extraction is required as the social data contains a lot of excessive data or noise which is not useful.
- ★ Feature Encoding can be a bit more optimized. Also a wide variety of dataset can be used to further increase the accuracy of the model and rank the fashion trends.
- ★ Presently we are only using clothing from various e-commerce and social media websites . Fashion data from Instagram could have been used but to keep data uniform it was not used

# Future Scope

**Keyword based search and analysis** ★

**Manual Search and Analysis** ★

**Author CO-citation Analysis** ★      The underlying assumption in ACA is that authors who are frequently cited together tend to work on similar concepts. Thus, frequently co-cited authors are likely to cluster together when an ACA is performed

**Topic Modelling** ★      Topic modeling relies on algorithms and statistical methods to elicit the topics latent in a large corpus