Flipkart GRID 4.0

2022 Campus Challenge

# Extract Trends from social media data

Team Name: **Winners**

Institute Name: **Rajiv Gandhi University of Knowledge Technologies, Srikakulam**

- **College**: Rajiv Gandhi University Of Knowledge Technologies, Srikakulam Campus
- **Name:** Rakesh Namadi

- **College**: Rajiv Gandhi University Of Knowledge Technologies, Srikakulam Campus
- **Name:** Durga Vadige

- **College**: Rajiv Gandhi University Of Knowledge Technologies, Srikakulam Campus
- **Name:** Chaithanya Vasantha Kumar Hanumanthu

# Team members details

| Team Name | Winners | | |
|---|---|---|---|
| Institute Name | Rajiv Gandhi University of Knowledge Technologies, Srikakulam | | |
| Team Members > | 1 (LEADER) | 2 | 3 |
| Name | **NAMADI RAKESH** | **VADIGE DURGA** | **HANUMANTHU CHAITHANYA VASANTHA KUMAR** |
| Batch | 2023 | 2023 | 2023 |

# Glossary

- WordCloud : It is an in-built python package to visulaize the words in the data. It plots the most occurrence word in larger in size.
- UNET : It is a Fully convolutional neural netowork (FCN) Which is used in semantic image segmentation.

# Use-cases

- This model generates trending products from the social media dataset and maps them to Flipkart category databases.
- Relevant keywords are derived from post metadata and used to locate the exact product on the Flipkart website.
- Trending list of products are generated based on time stamps of the posts.
- The trending score/ Ranking will be calculated based on the number of occurrences of the product keyword in the dataset.
- This model frames a relevant link based on the trending structed data of the respective product. This link will redirects to official flipkart website.
- This model offers some images and videos which are trending in respective social media platforms used to promote or advertise in flipkart website.
- In addition, a deep learning model will extract the types of the clothes used in images and generates the structed keywords to improve search results.
- From the images we detect famous logos using a deep layered network.
- we extracted the type of the clothes used in the image.

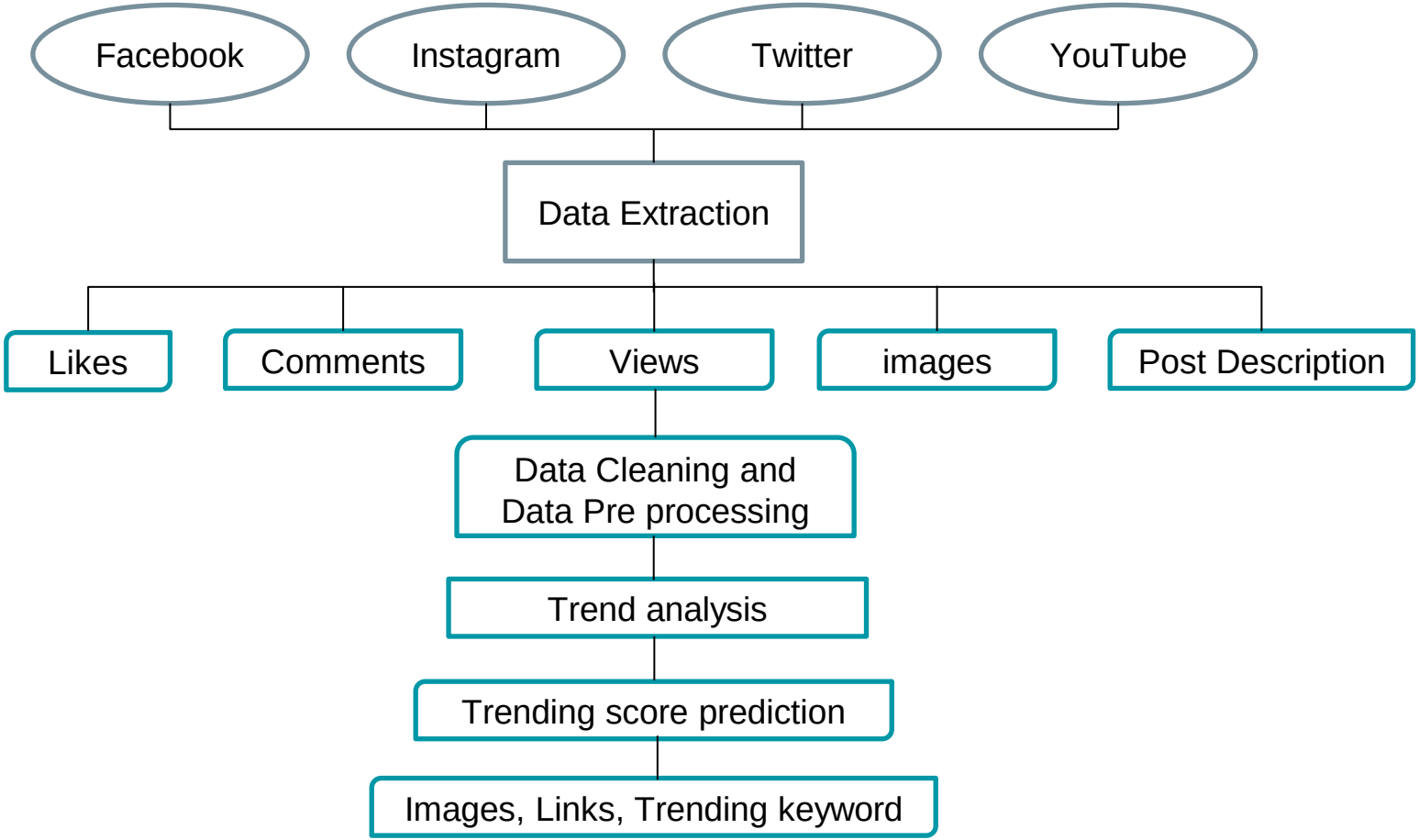
# Solution statement/ Proposed approach

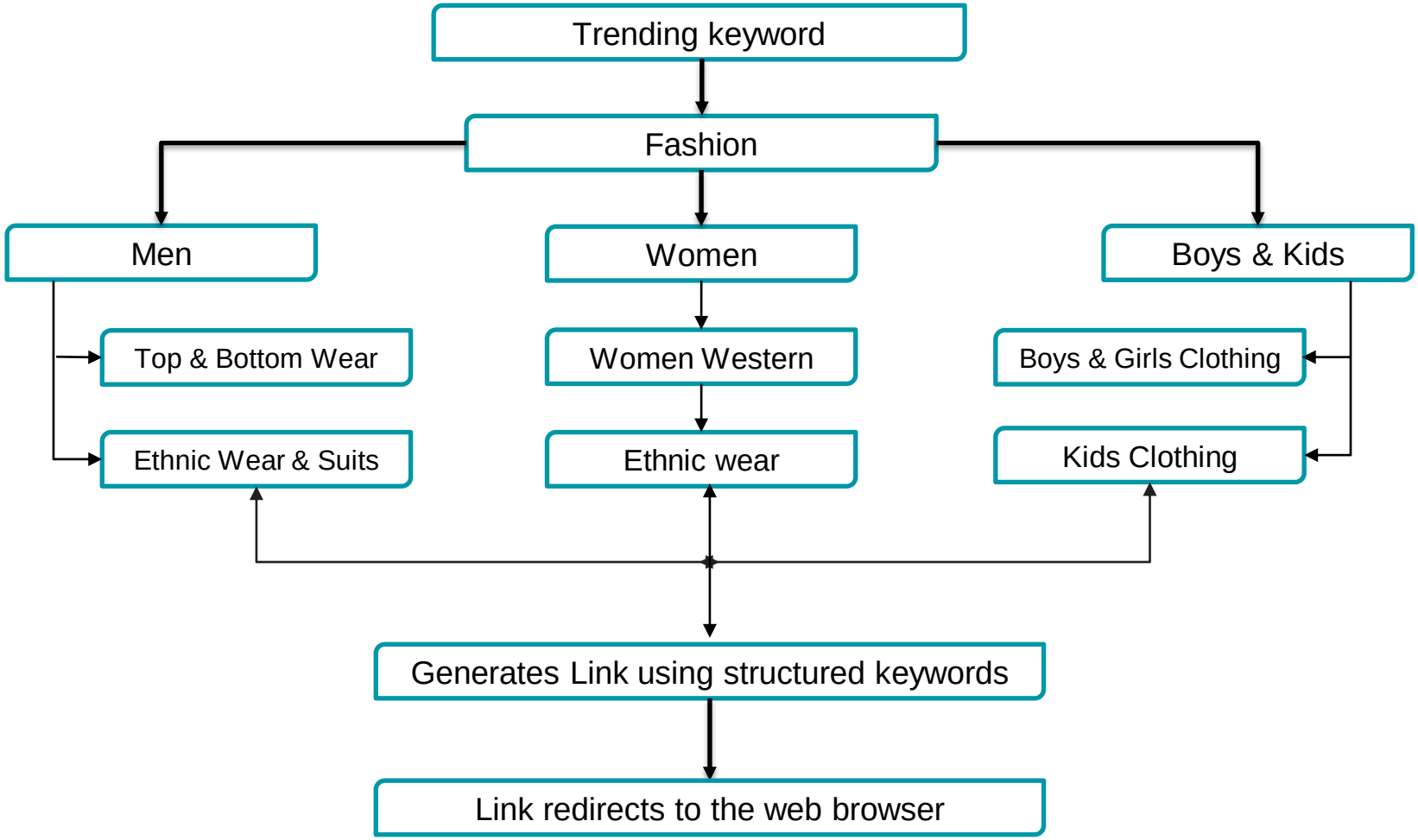
- Social media is a source which involves community to share their views, reactions, ideas, interests and it helps to participate in social networking. Social media holds different categories like political, fashion, technology etc.. Here our task is to identify the trending products from the social media data.
- We choose the popular social media platform Instagram to collect data for the analysis. Which has 1.21 billion active users around the world.
- We ran a model on the dataset; the model retrieved all the data from the post description and generated the keyword by most occurrences. That keyword is checked whether it is product or not and should available in flipkart.
- We used a module called "Word cloud," which displays the most trended products from the dataset. The larger the keyword, the more trended it was.
- For the trending products, we displayed the reach of the product as top likes, Average likes, top comments, Average comments, corresponded post links and images.
- Following that, we mapped the product to a Flipkart category and then linked it to the specific area where the product is offered in official Flipkart website.
- Once it is mapped, our model will automatically redirect to the category and open the product in the Flipkart website.
- The trending score will be obtained by

$$TS = \frac{\sum f_{\text{trending}}}{\sum_{i=0}^{k} f(i)} \times 100$$

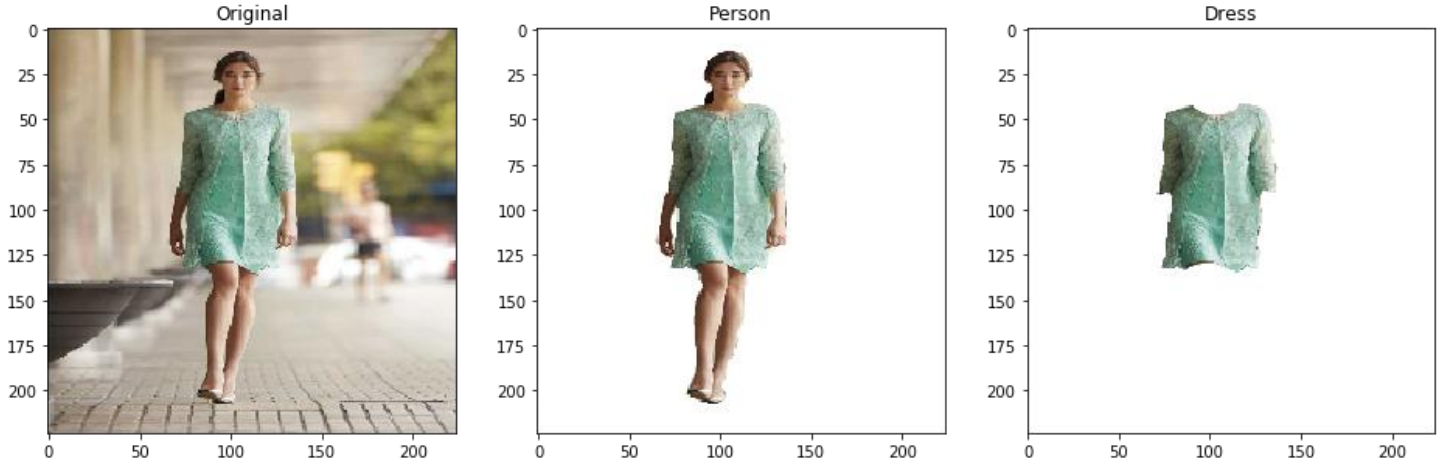
k = No.of. Words in post data

# BLOCK - DIAGRAM

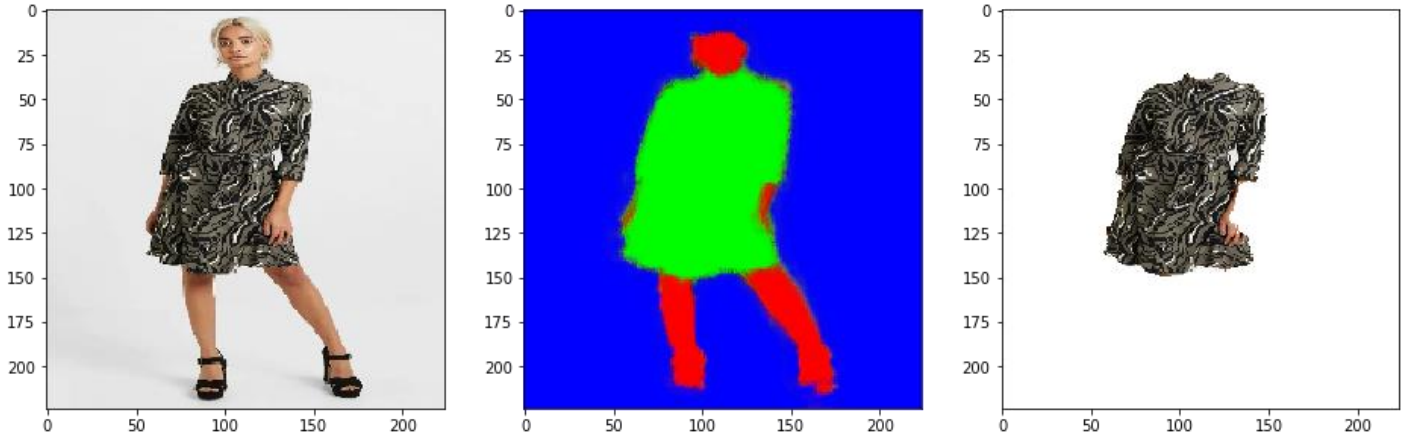


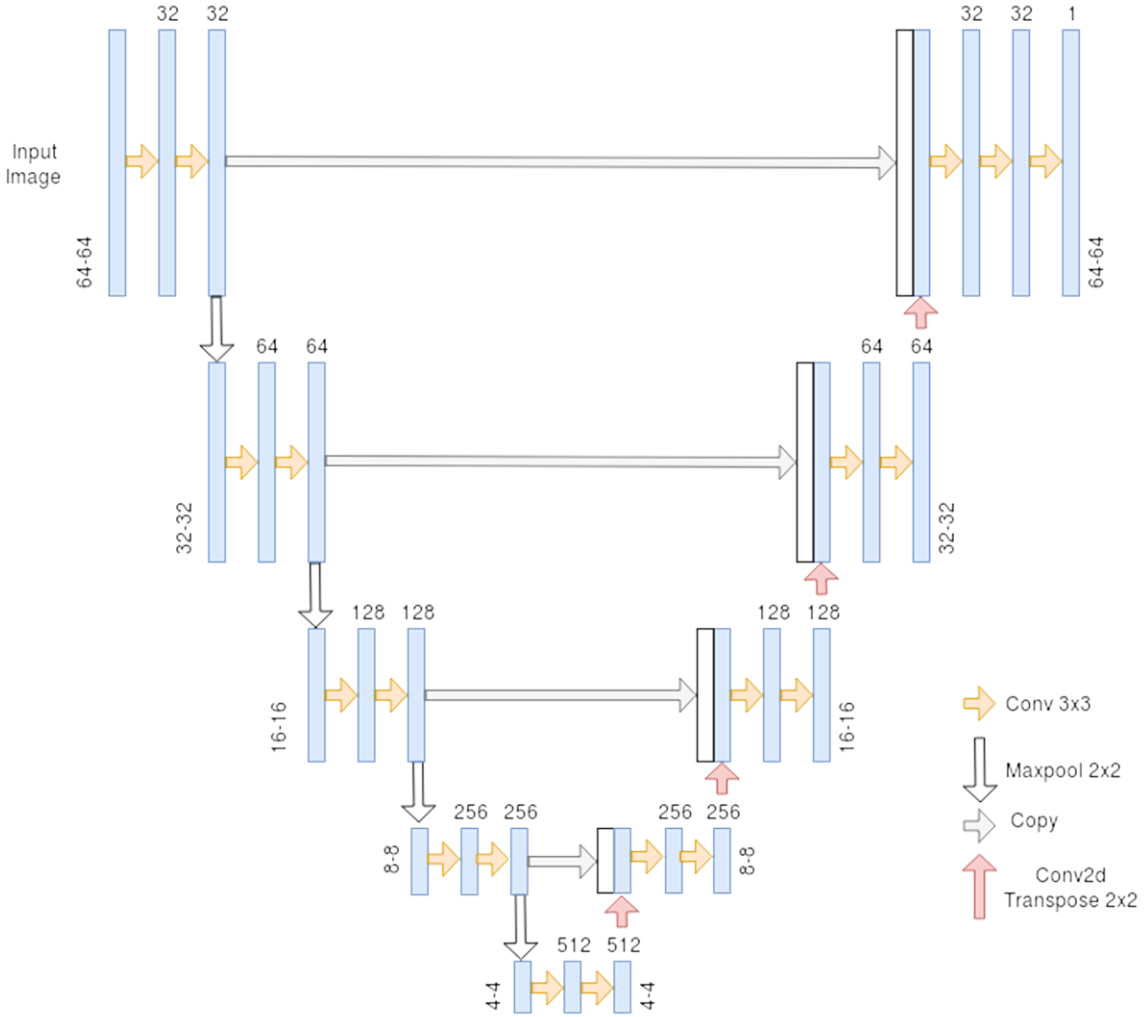

Trending keyword
Fashion
Men
Women
Boys & Kids
Top & Bottom Wear
Women Western
Boys & Girls Clothing
Ethnic Wear & Suits
Ethnic wear
Kids Clothing
Generates Link using structured keywords
Link redirects to the web browser

- Another approach to detect the trending keywords without the post metadata is to analyze the images which are posted in the social media platform.
- For this, we extracted the most liked and commented pictures from influencers, fashion vloggers, and bloggers in Instagram across India.
- We extracted all the objects from images to detect the products which are available in flipkart.This can achieve using R-CNN model. Through this model objects are extracted into different images. Then these individual images are compared with clothing photos using a pre-trained model. Then the model generates the patterns of clothes and detectes the type of the objects used in the images. These keywords are checked in the flipkart category list and generated a valid search link to match the products.
- We also generated the YouTube dataset. YouTube is also a popular social media platform having more than 2.6 billion active users. YouTube generated trending videos by default, and we collected thumbnails and photos from the videos based on the timestamps.
- To extract the text from the thumbnails we used OCR technique.
- Structured keywords are again generated from the extracted data from thumbnails and mapped to searchable query. Videos is analyzed to get the trending thumbnails.
- We can detect famous logos from the images using SSD: Single Shot Multi-box Detector is a method for detecting objects in images using a single deep neural network.
- We can detect clothes, brands, products, and brand logos instantly from a video or image without having any user data.




We make use of the **Grub Cut** Algorithm provided by OpenCV. This algorithm was implemented to separate the foreground from the background making use of the Gaussian Mixture Model.




Input Image
64-64
32
32
32
32
1
64-64
64
64
64
64
32-32
32-32
128
128
128
128
16-16
16-16
256
256
256
256
8-8
8-8
512
512
4-4
4-4
Conv 3x3
Maxpool 2x2
Copy
Conv2d Transpose 2x2

# How the Model works.?

- Using semantic image segmentation, a convolutional neural network model is trained to detect the types of clothes from images.
- We utilized discrimination because we want to distinguish between backgrounds, skin, and clothing. Because backgrounds and skins are the most important sources of noise in this type of problem, we try to eliminate them.
- We recreated the masks using the cuttings, by simply binarizing the image. The skin is obtained as a result of the difference between people and their clothing.
- We trained a model that takes a raw image as input and produces a three-dimensional mask that can recreate the desired separation between skin/background and dress from the original images. As a result, when a new raw image comes, we can divide it into three parts: background, skin, and dress. We take into consideration only the channel of our interest (dress). By using UNET segments the type of cloths.
- UNET uses convolution blocks followed by a maxpool down sampling to encode the input image into feature representations at multiple levels.

# Limitations

- This project is limited to proper images only. Extra noise, more colorful backgrounds, mass group photos may not give appropriate results.
- Misleading keywords in the post metadata may give inappropriate results. Somehow, those keywords will be verified with the text extracted from the image analysis.
- We have restricted our model only to the fashion categories but it works for all categories.
- For some products we can’t extract their brand logos as some products have their logo at different positions in the clothing.

# Future Scope

- This model will implement on OTT Platforms also. OTT Platforms have nearly 503 million active users all over India. It will detects the top trending products, objects, brands directly from the video snaps.
- Applying these techniques on top viewed movies, high rated TV shows to extract the brand logos or advertisements also helps to predict the trends.
  For ex:- Movie is in trend and the hero is holding the BoAt speaker. Model detect the BoAt speaker as the object and detects the logo of particular brand from the snap in the video at particular time interval. Then the model generates the valid search query which redirects to Flipkart official website.
- The results are mapped with the selling products at particular time stamps on flipkart and predict what exactly the product is going to sold.