

NAMADI RAKESH

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES,
SRIKAKULAM

[mailto:- namadirakesh@gmail.com](mailto:namadirakesh@gmail.com)

8919652766

Problem statement: Extract trends from social media using data analysis.

- Explanation: Here I selected a problem to extract the products which are going to be in trend to help e-commerce websites to increase their product sales.
- Social media is one of the most important aspect in personal life and business life. Social media allows people to interact with each other. It is one of the way to trend anything within very short span. As part of my challenge I triggered social media to get the trending products. Usually famous people, influencers can make everything trend. Most of the people are following them their dressing styles, ornaments, wearables, gadgets, products, brand influencers etc. Some influencers are marketing some goods. This will create a major impact in crowds.

- As part of my challenge I recorded a dataset from Instagram. Instagram is one of the leading social media platform. Which have two billion active users across the world. It allows users to upload media that can be edited using filters and organised by hashtags and geolocational tagging. Using web scrappers and Facebook graph API we can get the real time data from Instagram and Facebook which is quite useful in real time working of this project. I personally collected images, user post metadata, their tags, likes count, comment count etc... to my dataset. Later on I performed exploratory data analysis on the dataset in order to conclude my key points.
- Key aspect of the project was to find the specific product which is in trending in social media and verify that product whether it is available in e commerce or not. As a ecommerce site I selected Flipkart which have more than 350 million users. This project finds the trending product and maps it to the Flipkart website by adding categories and subcategories to the link.

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 structured data of the respective product. This link will redirect 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 structured 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.

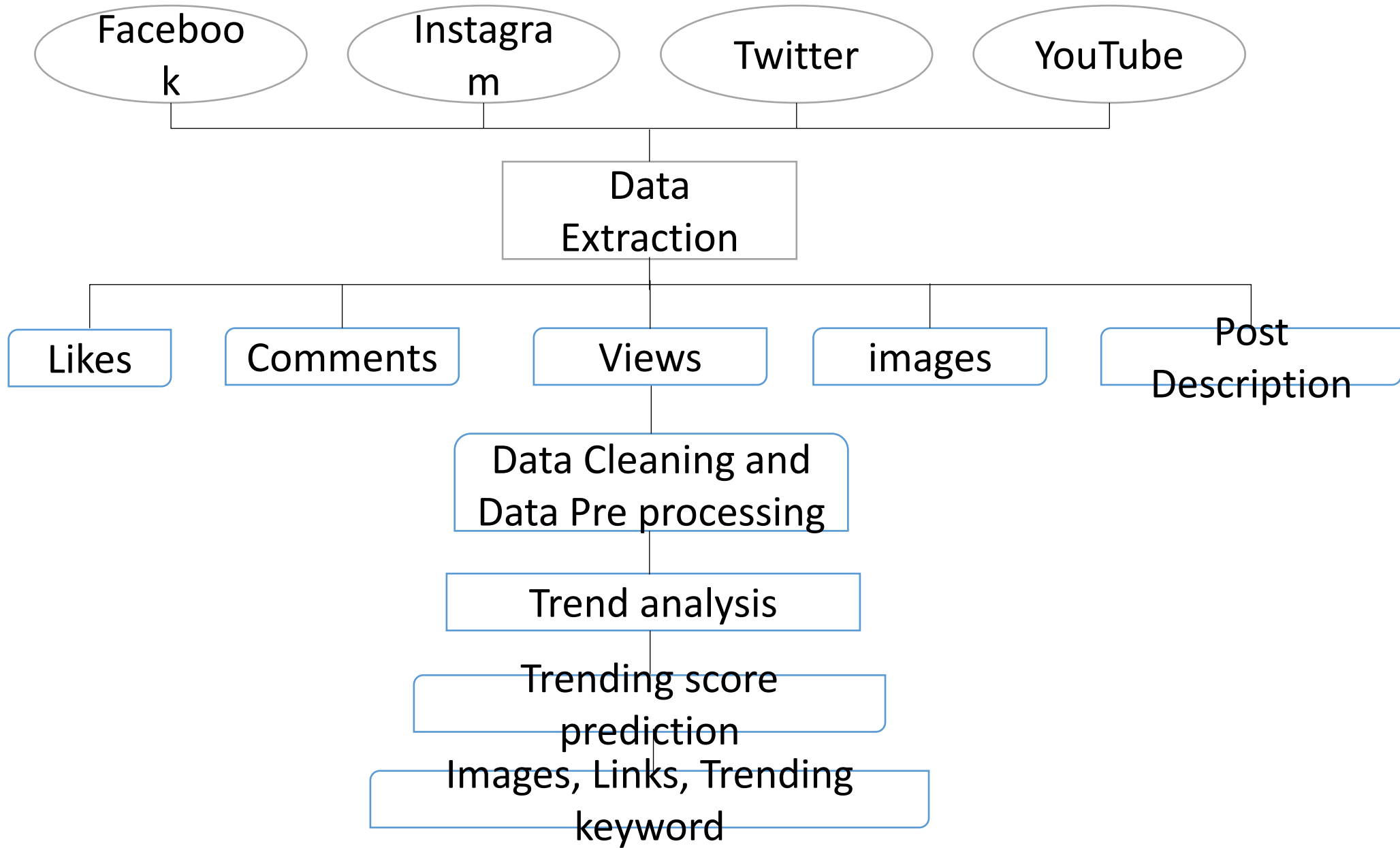
Solution statement/ Proposed approach

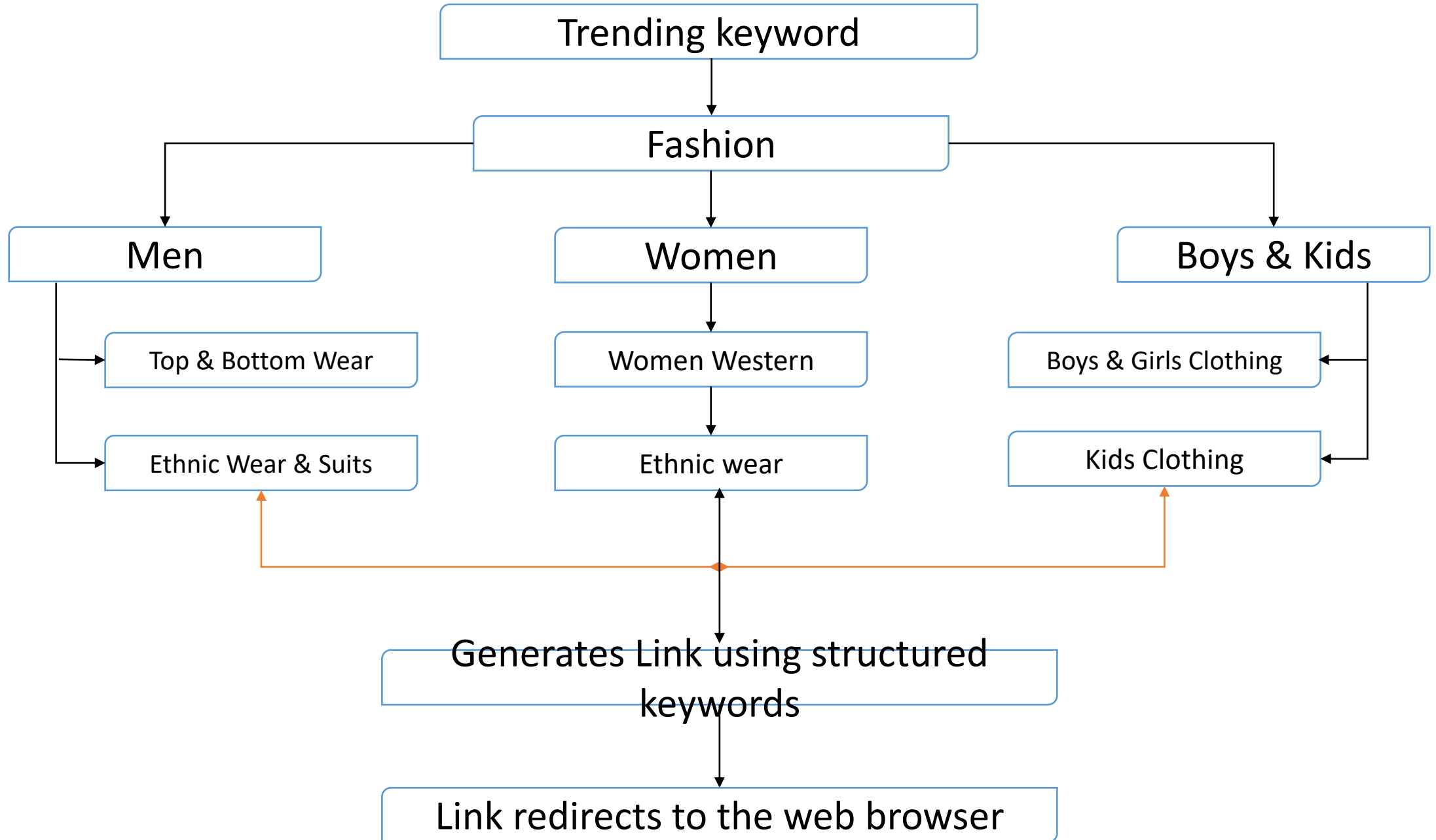
- Social media is a source which involves community to share their views, reactions, ideas, interests and it helps to participate in social networking. Social media holds different categories like political, fashion, technology etc.. Here our task is to identify the trending products from the social 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 be 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

k = No.of. Words in post data

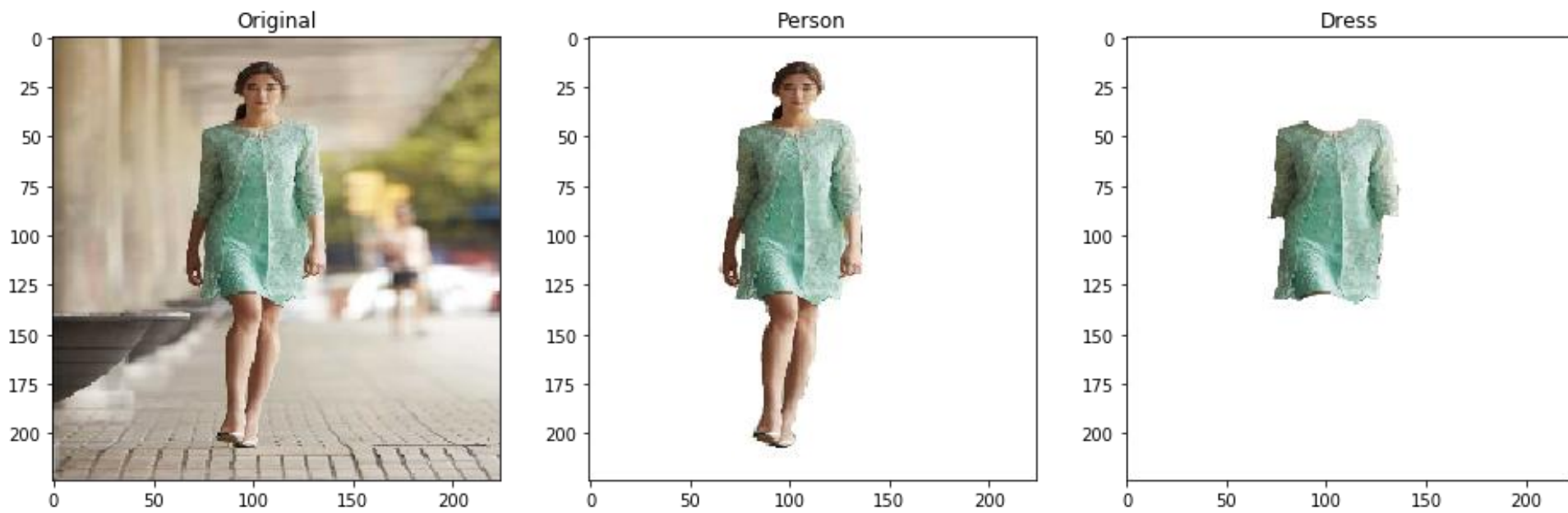
$$TS = \frac{\sum f_{\text{trending}}}{\sum_{i=0}^k f(i)} \times 100$$

BLOCK - DIAGRAM

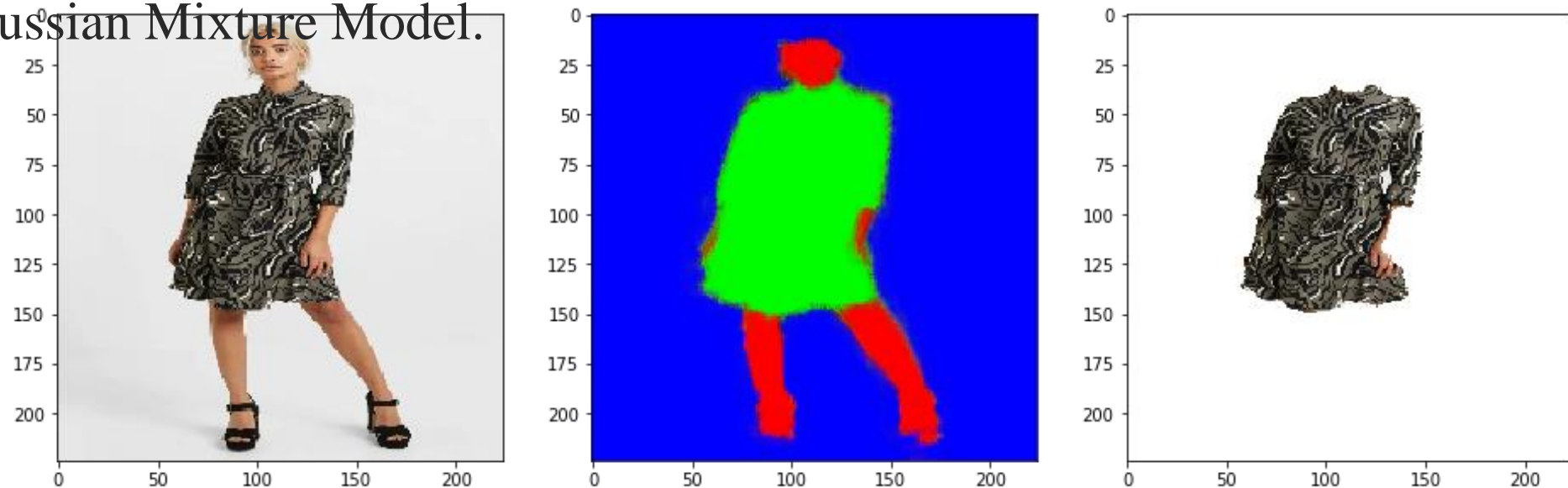


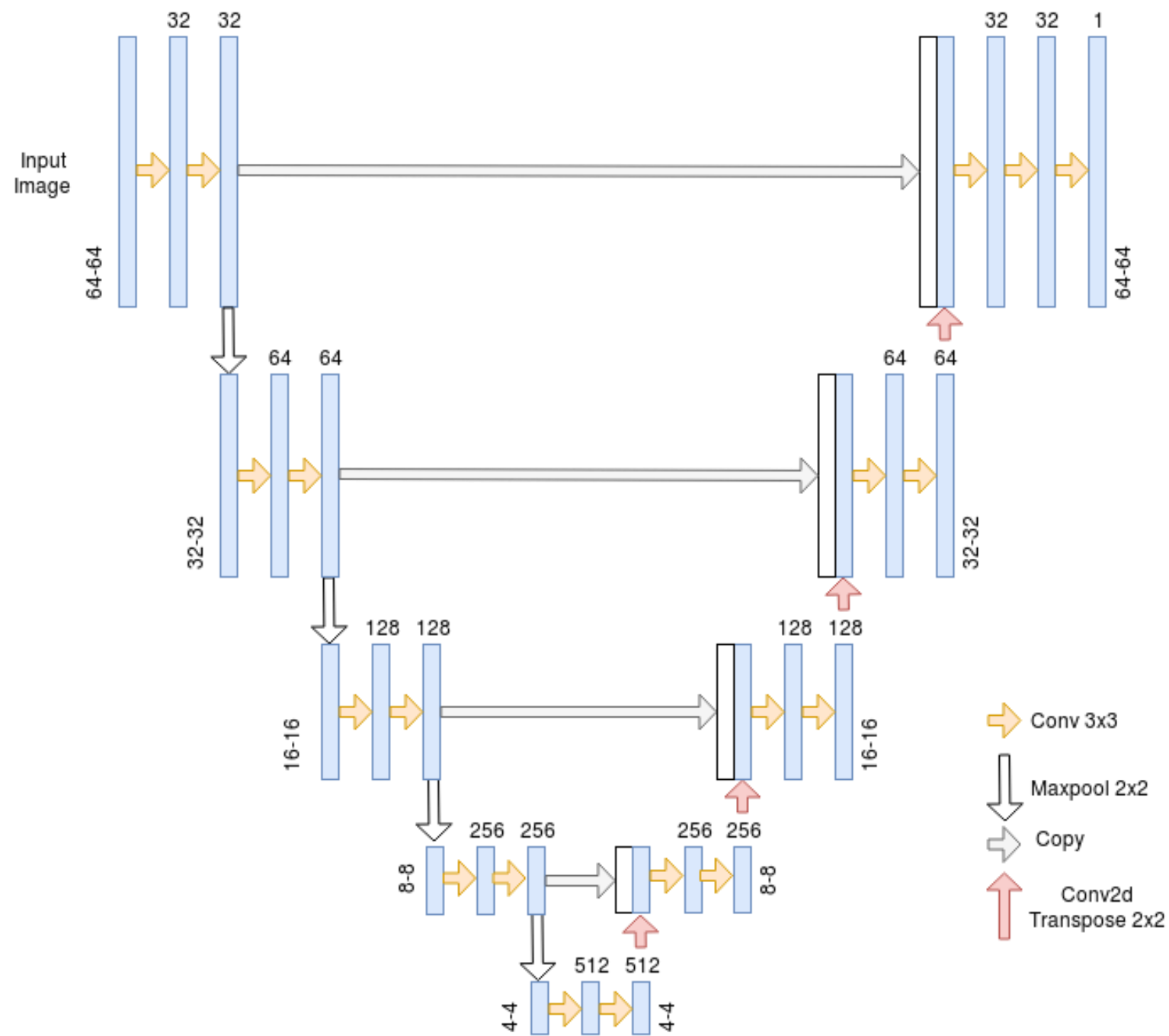


- Another approach to detect the trending keywords without the post metadata is to analyze the images which are posted in the social media platform.
- For this, we extracted the most liked and commented pictures from influencers, fashion vloggers, and bloggers in Instagram across India.
- We extracted all the objects from images to detect the products which are available in flipkart. This can be achieved 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 detects 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 are 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**.





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.

Outcomes:

1. *Saree and Lehanga* are the trending keywords on instagram. Hashtags will be like **#saree #lehanga**
2. Nearly 57 entries are there in 7th month of 2022
3. Top 10 keywords extracted from the post metadata are ['saree', 'lehenga', 'kurta', 'zara', '#bridal',
4. '#sumangali', '#lengha', '#suits', '#bridalmakeup', '#sareesoverseas']
4. Total words are 4818 in the dataset, trending keyword Saree count is 28, sum of top trending keyword occurrences 198
 1. Trending score **58%** for Saree
 - Tending score **56%** for Lehanga
6. Top trending analysis based on likes are:
 - Trending keyword: **Saree**
 - Top liked post: **1189129**
 - Total Likes: **4007068**
 - Average Likes: **97733**
 - Post Link : <https://www.instagram.com/p/CWnsbJMPJwJ/>

Note: Detailed Outcomes are in .ipynb notebook posted in Github.

Future Scope

- This model will implement on OTT Platforms also. OTT Platforms have nearly 503 million active users all over India. It will detect 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.