Flipkart GRiD 5.0

Problem Statement Title 2: Personalized Product Recommendations

Team Name: 686157-UI3YH533

Team members details

Team Name	686157-UI3YH533		
Institute Name/Names	Maharaja Agrasen Institute of Technology (MAIT), New Delhi Harcourt Butler Technical University (HBTU), Kanpur		
Team Members >	**1(team leader)**	**2**	**3**
Name	Ujjwal Pathak	Manan Jain	Prakhar Singh
Batch	2025	2025	2025

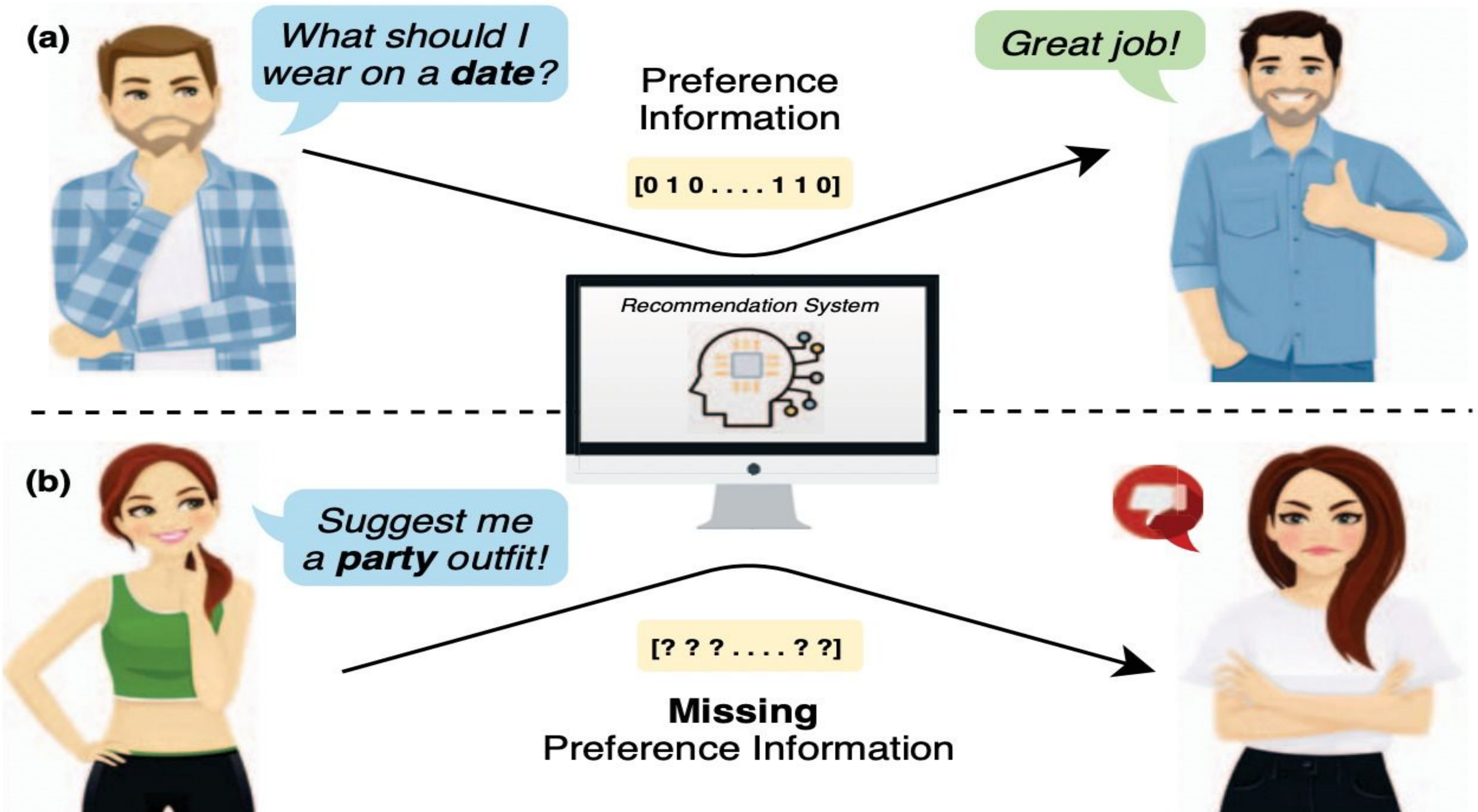
StyleSync

-Where Fashion Becomes Your Canvas

- In our busy lives, finding time for e-commerce shopping can be a daunting task. But the challenges don't end there. Many of us have unique and sometimes even peculiar tastes, making personalized recommendations a must-have.

- We are on a clear mission to revolutionize the way you experience fashion recommendations. Our aim is to overcome these challenges and deliver a truly personalized and hassle-free fashion journey.

- We're here to overcome these challenges and deliver a truly personalized and hassle-free fashion journey

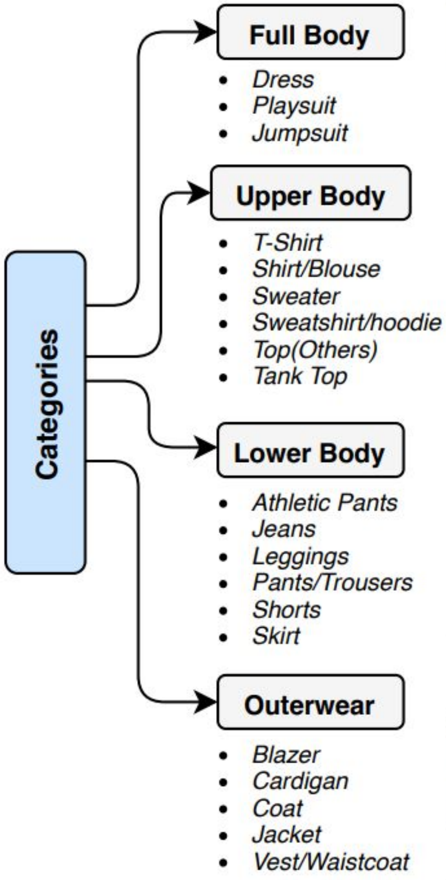

(a)
What should I wear on a **date**?
Preference Information
[0 1 0 . . . . 1 1 0]
Great job!
Recommendation System
(b)
Suggest me a **party** outfit!
[? ? ? . . . . ? ?]
Missing Preference Information

Dataset

- Introduction: The dataset we used to design StyleSync consists of valuable resource representing diverse fashion knowledge. It includes images collected from social media, annotated by experts. This dataset is crucial for training our system to provide personalized fashion recommendations.

- Data Collection: The dataset was collected from Instagram and Pinterest, featuring 2893 high-quality images. We used trending hashtags and keywords to categorize images into various occasions like Travel, Workout, Wedding, and more. This diverse collection forms the basis of occasion-oriented fashion knowledge.

- Data Filtering: We meticulously filtered the dataset to exclude low-quality images, selfies, and irrelevant content. This manual preprocessing ensured data quality and relevance for our project's accuracy.



Attribute	Attribute Value
Colour(Top/Bottom)	*multi, brown, gold, white, yellow, green, grey, purple, pink, black, blue, red, silver, navy, orange*
Sleeve Length	*Short, Long, Sleeveless, Three-quarter*
Upper Body Length	*Crop, Hip Length, Thigh length, Knee Length, Below Knee*
Lower Body Length	*Extra Short. Thigh length, Knee Length, Calf Length, Full Length*
Neckline	*Collar, Hooded,Turtleneck, Round Neck, V-neck, Halter Neck, Off-shoulder*
Closure	*Zip, Pullover, Buttoned*

Travel	Sports	Wedding	Date	Party	Prom	Office
500	438	335	423	362	370	465

Figure 2. Our dataset, with 20 categories, 39 attributes across 7 occasions

- In our dataset, we define a set of categories and attributes to best represent the fashion knowledge in terms of the clothing aspects for each image, as shown in the fig.
- Each category defines a particular region of clothing (e.g. t-shirt: upper body, athletic pants: lower body) and each attribute further characterises a particular category (e.g. sleeve length, neckline: t-shirt, lower body length: athletic pants).

Methodology

A) Visual Preference Modeling

- Fashion outfits consist of various clothing items like tops, bottoms, or full-body dresses. To analyze these outfits effectively, the first step is to identify different clothing items within an image. The Faster RCNN model with a ResNet-50 backbone is used for this task. It's a state-of-the-art technique for real-time object detection.

- T he Faster RCNN model is fine-tuned on the ModaNet dataset. Fine-tuning ensures that the model is specialized in detecting fashion-related elements in images and excludes irrelevant objects. ModaNet is used as a reference dataset for fashion-related object detection.

- The regions of interest detected by Faster RCNN represent different clothing items but are unlabeled and cannot directly convey fashion knowledge. These regions need to be transformed into a semantic space representation that encodes visual fashion knowledge in terms of categories and attributes. The goal is to align this representation with the labeled fashion knowledge in the dataset.

- For each region of interest within an image, the objective is to predict both a category label (e.g., "shirt," "jeans") and a set of attribute values (e.g., "short sleeves," "blue color"). Each region is characterized by a clothing category and a set of attribute values that are semantically similar to the labeled dataset.

- The representations of individual regions of interest are combined (pooled) to form a vector representation for each image. This vector summarizes the fashion elements present in the image and encodes them in a way that can be used for recommendation.

- Since, clothing categories and attribute values are not independent, Multi-Task learning is used to capture these dependencies and individual features effectively. A multi-task classifier, MobileNetV2 is used which is initially trained on the DeepFashion dataset to learn relevant high-level features for fashion-related tasks and is then fine-tuned on the specific dataset used in this study.

- Different classification heads within the classifier are responsible for optimizing category and attribute prediction tasks.

B) Personalised Outfit Recommendation

- The process starts with the extraction of vector representations for user input images, as described in the previous section. These vectors capture the fashion knowledge from the images.

- Content-based filters are applied to the fashion dataset to retrieve outfits that match the user's gender and the specific occasion. This filtering narrows down the dataset to relevant outfit images.

- To group outfits into high-level concept subgroups, an unsupervised learning approach is used, specifically a modified version of the k-modes clustering algorithm. Modifications include using a weighted dissimilarity measure (instead of simple hamming distance) and assigning different weights (α and β) to category and attribute features, respectively (as described below). The goal is to identify clusters of outfits that share similar fashion concepts.

$$D_m(X,Y) = \alpha \sum_{j \in cat} \delta(x_j, y_j) + \beta \sum_{j \in attr} \delta(x_j, y_j)$$

- Rather than simply drawing an equal number of samples from each cluster for the final recommendation (which would provide diversity but not consider user preferences), a more personalized approach is used.

- Each cluster is weighted based on the similarity between the semantic representations of outfits in the user's input images and the concept clusters. The semantic representation of cluster centroids is used for comparison. The number of outfit samples drawn for the recommendation is determined by the weights assigned to each cluster.

- For example, if a female user's input images indicate a preference for sleeveless dresses (category: dress, sleeve length: sleeveless), the recommended outfits are more likely to include a greater number of outfits with these features.

- This approach ensures a diversified set of recommendations while also incorporating the user's preferences in a scenario, where there may be limited historical data about the user's fashion choices.

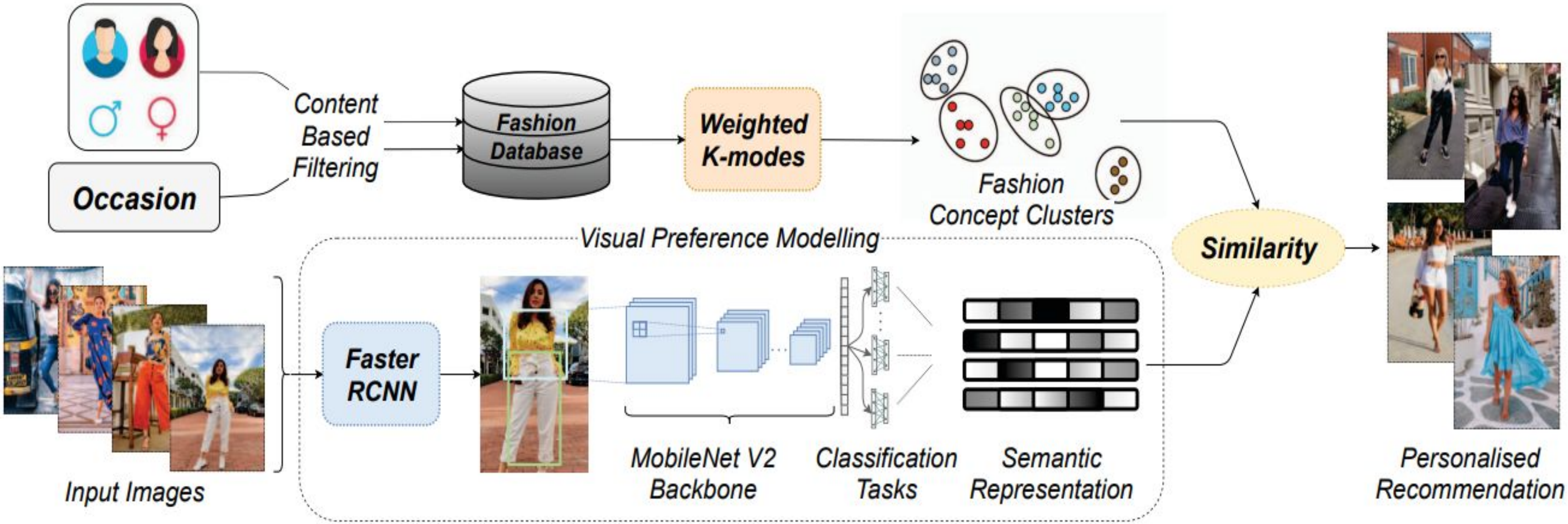


Figure 3. The proposed pipeline for personalised outfit recommendation in cold-start scenarios

Classification Results

- Several deep learning network architectures were explored as backbones for this prediction task. These architectures include MobileNet V2, ResNet-18, and ResNet-50. Both single-task configurations (predicting either clothing categories or attributes) and multi-task configurations (simultaneously predicting both categories and attributes) were examined.

- Among the various configurations and network architectures tested, the MobileNetV2 was identified as the best-performing model. This specific configuration involved three blocks of fine-tuning and empirically calculated weights ($\lambda1/\lambda2 = 2$, $\alpha/\beta = 2$). This model achieved an accuracy of 73.5% for clothing category classification and 89.1% for average attribute classification.

- The performance of the best-performing model was compared to baseline models, which were single classification task CNNs. The results showed that the model performed competitively with these systems in terms of category classification and outperformed them for attribute classification.

Limitations of Model

- Data Volume: The dataset used for training and testing may not cover all possible fashion scenarios and personal preferences, limiting the system's adaptability.

- Image Quality: The effectiveness of the system depends on the quality of user-provided images. Low-quality or ambiguous images can impact recommendation accuracy.

- Fashion Evolution: Rapid changes in fashion trends may lead to outdated recommendations, requiring regular updates to the system.

- Privacy Concerns: Collecting user images and preferences raises privacy concerns, which need to be addressed transparently.

Expanding Horizons

- Comprehensive Fashion Knowledge: To offer enhanced personalization, we recognize the need for a broader dataset. This dataset should encompass not only clothing but also accessories, footwear, and complete head-to-toe outfits.

- Advanced Algorithms: We aim to develop algorithms that will consider compatibility and style coherence within full outfits. By understanding how different elements complement each other, we can suggest outfits that truly resonate with the user's fashion sense.

- User Interaction Methods: Gathering preferences for entire outfits requires innovative user interaction methods. We're exploring intuitive interfaces and AI-driven conversations to collect detailed insights into users' style preferences.

Thank You