ALGORITHMS FOR WEB AND INFORMATION RETRIEVAL

PROJECT TITLE :
FASHION RECOMMENDER SYSTEM

TEAM DETAILS :

ID	NAME	SRN	SECTION
1	PRIYA MOHATA	PES2UG19CS301	E
2	PRIYANSH GAURAV	PES2UG19CS302	E
3	R SHARMILA	PES2UG19CS309	E
4	RITIK	PES2UG19CS332	E

SYNOPSIS :

In recent years, the textile and fashion industries have witnessed an enormous amount of growth in fast fashion. On e-commerce platforms, where numerous choices are available, an efficient recommendation system is required to sort, order, and efficiently convey relevant product content or information to users. Image-based fashion recommendation systems (FRSs) have attracted a huge amount of attention from fast fashion retailers as they provide a personalized shopping experience to consumers. With the technological advancements, this branch of artificial intelligence exhibits a tremendous amount of potential in image processing, parsing, classification, and segmentation.

An effective recommendation system is a crucial tool for successfully conducting an e-commerce business. Fashion recommendation systems (FRSs) generally provide specific recommendations to the consumer based on their browsing and previous purchase history. Social-network-based FRSs consider the user's social circle, fashion product attributes, image parsing, fashion trends, and consistency in fashion styles as important factors since they impact upon the user's purchasing decisions. FRSs have the ability to reduce transaction costs for consumers and increase revenue for retailers.

FUNCTIONALITY :

The project makes use of various concepts like DeepLearning [CNN] , Transfer learning using ResNet50 and in order to find similarity between the vectors we make use of KNN [with both cosine similarity and Euclidean distance].
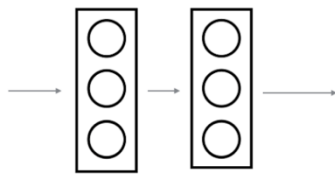
Transfer Learning :
- Transfer Learning is a machine learning method where we reuse a pre-trained model as the starting point for a model on a new task.
- By applying transfer learning to a new task, one can achieve significantly higher performance than training with only a small amount of data.
- To put it simply—a model trained on one task is repurposed on a second, related task as an optimization that allows rapid progress when modeling the second task.
- ImageNet, AlexNet, and Inception are typical examples of models that have the basis of Transfer learning.
- So for our project we would be making use of the ResNet50 [ImageNet] Model.
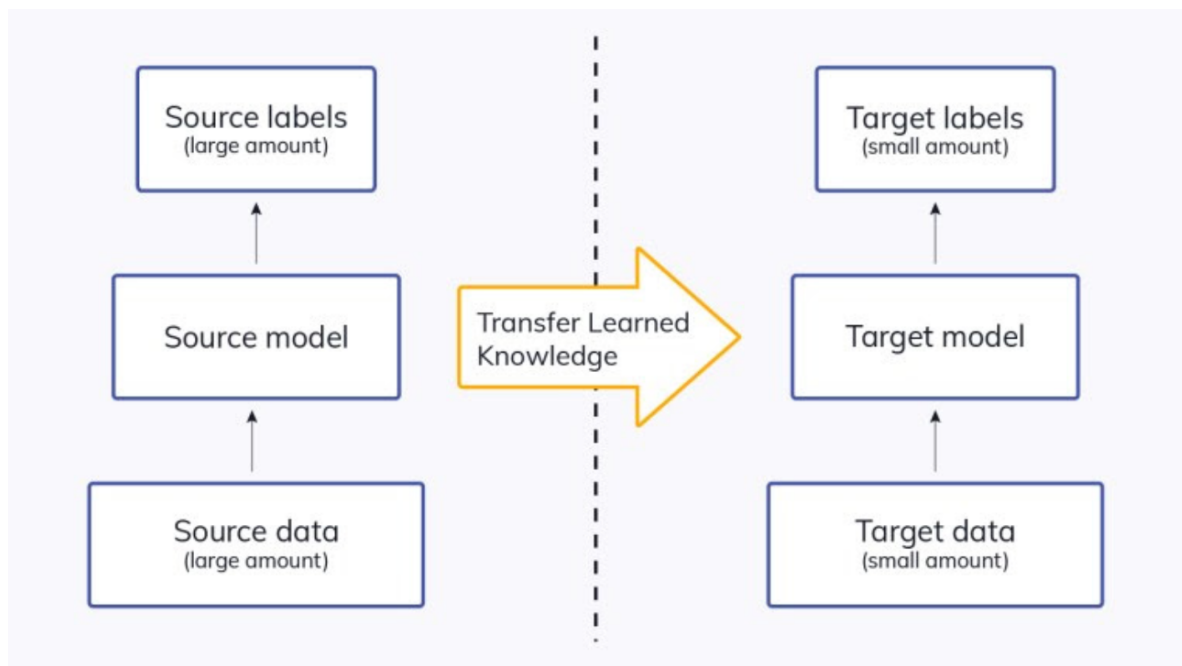

ResNet :

- <u>ResNet</u>, short for Residual Networks is a classic neural network used as a backbone for many computer vision tasks
- The fundamental breakthrough with ResNet was it allowed us to train extremely deep neural networks with 150+layers successfully. Prior to ResNet training very deep neural networks was difficult due to the problem of vanishing gradients.
- ResNet uses skip connection to add the output from an earlier layer to a later layer. This helps it mitigate the vanishing gradient problem
- We make use of the ResNet50 present in Keras that is pretrained.

without skip connection   with skip connection
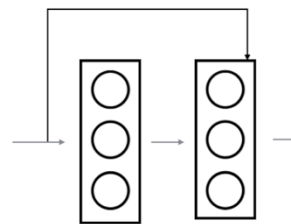
Skip Connection Image from DeepLearning.AI

<u>K-Nearest Neighbours :</u>
- The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.
- The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.
- "Birds of a feather flock together.

<u>Advantages</u>
- The algorithm is simple and easy to implement.
- There's no need to build a model, tune several parameters, or make additional assumptions.
- The algorithm is versatile. It can be used for classification, regression, and search (as we will see in the next section)

<u>Disadvantages</u>
- The algorithm gets significantly slower as the number of examples and/or predictors/independent variables increase.

The KNN Algorithm

1. Load the data

2. Initialize K to your chosen number of neighbors

3. For each example in the data

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection

4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. If regression, return the mean of the K labels

8. If classification, return the mode of the K labels

Cosine Similarity :

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.

Euclidean Distance :

In mathematics, the Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. It can be calculated from the Cartesian coordinates of the points using the Pythagorean theorem, therefore occasionally being called the Pythagorean distance.

1. *Cosine similarity*: This measures the similarity using the cosine of the angle between two vectors in a multidimensional space. It is given by:

$$\text{similarity}\,(x,y) = \cos\,(\theta) = \frac{x \cdot y}{|x||y|} \qquad (8.2)$$

2. *Euclidean distance*: This is the most common similarity distance measure and measures the distance between any two points in a euclidean space. It is given by:

$$d\,(x,y) = \sqrt{\sum_{i=1}^{n}\,(x_i - y_i)^2} \qquad (8.3)$$

About the dataset :

Source of the dataset

Has more than 44,000 images of various product images , like shirts , tshirts , watches , sneakers , sarees.

The images are in high resolution but due to the limitations in computation power we would be making use of the low resolution images for faster computation and less storage.

SOFTWARE REQUIRMENTS :

Tools and Technologies Used :
- Python 3.9
- Tensorflow
- Keras
- Scikit-Learn
- ResNet
- Streamlit
- Pycharm IDE
- Tqdm
- Numpy
- Pickle

Any operating system : Linux , Windows , MacOS.

RAM : Minimum 2GB

EXPECTED OUTPUT :

The given recommender system would recommend 5 top items to the user based on the given image by the user.