

## Problem statement

Develop a machine learning model that, when provided with a random input image, can accurately identify and output the top five actors or actresses from a predefined dataset who most closely resemble the individual in the input image. The model should also provide the corresponding similarity scores for these top five matches.

## Approach

1. Collect images of various actors and actresses and label them.
2. Use a pre-trained CNN for feature extraction.
3. Then convert the images into feature vectors (embeddings) using the trained model.
4. Store these embeddings in an index for later use.
5. Implement a similarity search mechanism using KNN to retrieve most similar embeddings from the index.
6. For any given input image, preprocess the image and extract its embedding using the trained model.
7. Perform the similarity search against the indexed embeddings.
8. Output the top five matches along with the similarity scores.

