

The project Celebrity recogniser takes an input image and recognises whether it matches any of the celebrity present in our dataset. The dataset has six celebrity figures, each set having multiple images of that celebrity. The dataset of images is not preloaded from any external source and has been generated by the help of a python script using simple-images-download. This script helps us download a certain number of images of whatever string is passed to it. Using this we created a new dataset from scratch and then manually edited the incorrect entries to get a clean data set. In this way novelty has been shown in terms of data generation.

A new methodology of voting classifier is used in this project by me. A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. SVC, Logistic Regression, ExtraTreesClassifier and DecisionTreeClassifier are the various models used by me for voting classifier. In classification, a hard voting ensemble involves summing the votes for crisp class labels from other models and predicting the class with the most votes. A soft voting ensemble involves summing the predicted probabilities for class labels and predicting the class label with the largest sum probability. I have used soft voting in the project. Along with voting classifier I have used wavelet transformation function on the dataset before model training to extract the various features from the images. This dataset is then used to get good output results after model fitting.

Hyperparameter Tuning is used to get a good accuracy. I used GridSearchCV which takes a parameter grid and runs it to get the best parameter combination. The best parameters found by GridSearchCV is displayed by best\_params\_ attribute, and the best estimator is shown by best\_estimator\_ attribute: