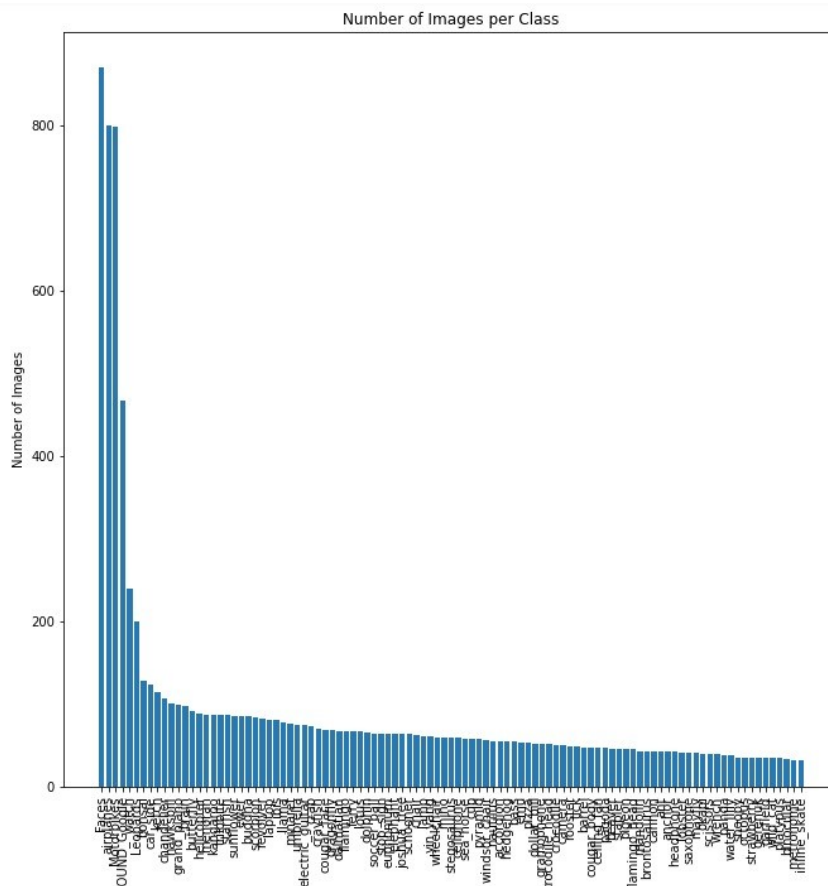# Introduction:

The report is to provide an analysis of the performance of K-Nearest Neighbour Classifier on the Caltech 101 image collection. The image collection contains images from 101 different classes and each image is represented by a feature called Edge Histogram which is given in the file EdgeHistogram.csv. We use this data for our analysis.


Methodology:

For this analysis, A KNN classifier was implemented using the Scikit learn library in python. This implementation consists of following steps:

Step-1 : <u>Data Preparation:</u> The data was split into training and test sets, where different amounts of training images per class (3,5,10,15) were tried out. The training and test data were selected using a function that is train _test_split() which was imported from Sklearn.model_selection.

Step-2 : <u>Model Training:</u> The KNN classifier was trained using training data that was obtained from Step-1. Hyperparameters were tuned to find the best performance. GridSearchCV was used to get the best Hyperparameters. We got K=11 after running GridSearchCV.

Results: After training the KNN algorithm with the data provided we got the accuracy of 35.59 percent for training amount of 3 images per class followed by accuracy of 36.68 percent for the training amount of 5 images per class. While the training amount is 10, we got the accuracy of 41.19 percent which is better than the remaining. Lastly, for the training amount of 15 images per class we got the accuracy of 35.59 percent.

The below diagram shows the Relationship between Training set size, Classification Parameters and Accuracy.
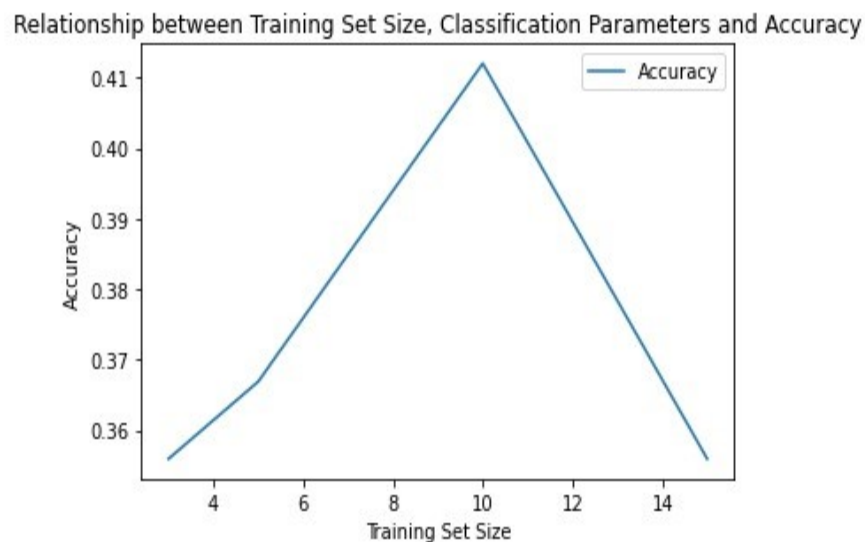
This implementation is done by the team of two.
 Bhavanishankar R Hakari
Nandavardhan Reddy Ramireddy