Capstone III – Project Report

Problem Statement

The objective of Capstone project is to create an Image Classification Model. The model should be able to take an input/test image and accurately classify it in the correct category.

It is going to use a Supervised Learning algorithm and the method used for clustering is K Means clustering. The metric that is going to be used for assessing the model is Mean Absolute Error (MAE).

There are broadly 4 steps in the Project:

1. Data Wrangling
2. EDA (Exploratory Data Analysis)
3. Model building
4. Model validation

Data Wrangling & EDA

For the purpose of the project, there is an image downloader downloaded in Google. Then approx. 100 images are saved from Google Images are saved in different folders for Sunrise, Noon and Sunset times.

All the images are then resized to a single dimension – (300\*300). Then we extract the ‘Red’, ‘Green’ and ‘Blue’ pixel values for the images in a csv file.

Then separate histograms are drawn for all 3 types of pixels.

My mentor observed that the histogram for ‘Red’ pixels were similar for both Sunrise and Sunset Images. So, we decided to remove the set of Sunrise images from scope of the project.

Model Building

The method chosen for building the model was K Means clustering. I started off by assigning a cluster value 1 to set of Sunset pixels and cluster value 0 to set of Noon pixels. Combined the dataset for both the pixels.

Then did a Train-Test split for the entire dataset in the ratio of 75:25. I then fit a K Means model on the Training dataset and predicted values for Test dataset.

Model Validation

Model validation was done in 2 steps:

1. Error Metric – Mean Absolute Error for the model came to around 0.32 which is quite low and thus the model is good.
2. 2 test images clicked at Noon time and Sunset are used as input to the model and model predicted the cluster accurately as 0 and 1– (Noon Image & Sunset Image).