

Instagram Like Classification mini project

Name: Gokulraj Varatharajan

Email: gvaratha@usc.edu

Introduction:

This project aims to classify Instagram posts into three categories (Low, Medium, High) based on the number of likes, using both meta-data and image-based features. The likes are classified based on quartiles (0 to 33, 33 to 66 and 66 to rest)

Methods:

I have used Random Forest as the common model across all variations to get standardized results which can be compared. For the CLIP based model alone I have used multiple classification models to compare the accuracy of various models.

The different variations in how models are trained:

Experiment 1 - Meta data only

Experiment 2- Meta data + features extracted from images using Resnet 50

Experiment 3 - Meta data + extracting color histogram from images

Experiment 4 – Meta data + SIFT features extracted from the images

Experiment 5 - Meta data + CLIP

Results:

| Experiments | Precision (Low) | Precision (Medium) | Precision (High) | Accuracy |
|-------------|-----------------|--------------------|------------------|----------|
| 1 | 0.84 | 0.85 | 0.64 | 0.78 |
| 2 | 0.62 | 0.61 | 0.48 | 0.58 |
| 3 | 0.67 | 0.64 | 0.48 | 0.61 |
| 4 | 0.37 | 0.37 | 0.33 | 0.36 |
| 5 | 0.76 | 0.70 | 0.63 | 0.71 |

Conclusion:

The classifier trained only using the meta data performed very well compared to any classifier using image features as an input. Among the models that use image features, CLIP model performs the best and comes close to the meta data only model.