Problem Statement

For HackerEarth's ML Challenge – Pride Month (July, 2020)

You can see the problem statement on the HackerEarth website and then solve it yourself and submit your work in order to get your score (I highly recommend you to do this) using the following link:

https://www.hackerearth.com/problem/machine-learning/detect-the-sentiment-of-a-quote-2-ca749be7/

Problem Statement:

You work as a social media moderator for your firm. Your key responsibility is to tag uploaded content (images) during Pride Month based on its sentiment (positive, negative, or random) and categorize them for internal reference and SEO optimization.

Task

Your task is to build an engine that combines the concepts of OCR and NLP that accepts a .jpg file as input, extracts the text, if any, and classifies sentiment as **positive** or **negative**. If the text sentiment is neutral or an image file does not have any text, then it is classified as **random**.

Data

You must use an external dataset to train your model. The attached dataset link contains the sample data of each category [Positive | Negative | Random] and test data.

Data files

File name	Description
Test.zip	Contains image files to be classified
Sample.zip	Contains sample image files belonging to each category
Test.csv	Predictions file containing indices of test data and a blank target column
sample_submission.csv	Submission format to be followed for uploading predictions

Data description

Column name	Description
Filename	File name of test data image
Category	Target column [values: 'Positive'/'Negative'/'Random']

Submission format

Please refer to sample_submission.csv for more details (check the website using the link given at the start of this file in order to download this)

Evaluation criteria

$$score = 100 * recall\_score(actual\_values, predicted\_values)$$