Ad Deduplication with images

Description

Schibsted owns several online marketplaces such as e.g. *finn.no*. In these marketplaces

users can post classified ads for objects they wish to sell. For various reasons, users might

post multiple classified ads for the same item, i.e. creating duplicate classified ads. We’d like

to avoid this, as it makes the user experience more confusing for other users.

In this task, we ask that you write code to detect such duplicate classified ads just by

comparing the *images* from the classified ads. For simplicity we assume that each classified

ad only has a single image. We’ve included some examples of images that might be from

multiple such duplicate ads.

The task formulation is a bit vague on purpose, enabling you to make your own choices

where needed.

Instructions

● Download examples of images from duplicate ads from the link

https://drive.google.com/open?id=1GqnKQPjb\_4x5pL8WdIylr5fqfMsg-GCu

● For your reference, examples of images from non-duplicate ads can be downloaded from the

link https://drive.google.com/open?id=1Ok-yU50\_bzvCedRha6XwznpO621RiX-z

● Your task is to create code that takes as input two images and returns some value indicating

whether they are duplicates or not

● Please submit a running piece of code, documented as you please (in the code, separate

docs, document by tests, whatever makes the code understandable)

● We are interested in seeing your ideas, and your coding style/skills. Don’t worry if the results

are sub-optimal

● If you have ideas on how to improve the solution further, please write that down as well

● Spend approximately 3 hours on the task