

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro



## Actividad 2.2 Preprocesado de los datos

Desarrollo de aplicaciones avanzadas de ciencias computacionales (Gpo 301)

Santiago de Querétaro, Querétaro, 19 de mayo del 2024

Aldo Tena García - A01275222

[A01275222@tec.mx](mailto:A01275222@tec.mx)

Enlace al repositorio de GitHub: <https://github.com/ATG-5222/ImageClassification>



The screenshot shows a GitHub repository page for 'ImageClassification'. At the top, there are tabs for 'README' and 'MIT license'. The main heading is 'ImageClassification'. Below it, a subtitle reads 'Image Classification (Pizza and Ice cream)'. The section 'Selección del dataset' contains two paragraphs. The first paragraph states that the dataset 'Pizza vs Ice Cream' was downloaded from Kaggle on May 14, 2024, for the purpose of comparing two different food items. It mentions the dataset structure with train and test folders and an initial count of 500 training and 50 testing images. The second paragraph describes how the image counts were adjusted to 200 for both train and test sets to ensure proper model training, and that data augmentation techniques were used to provide necessary resources for training.

# ImageClassification

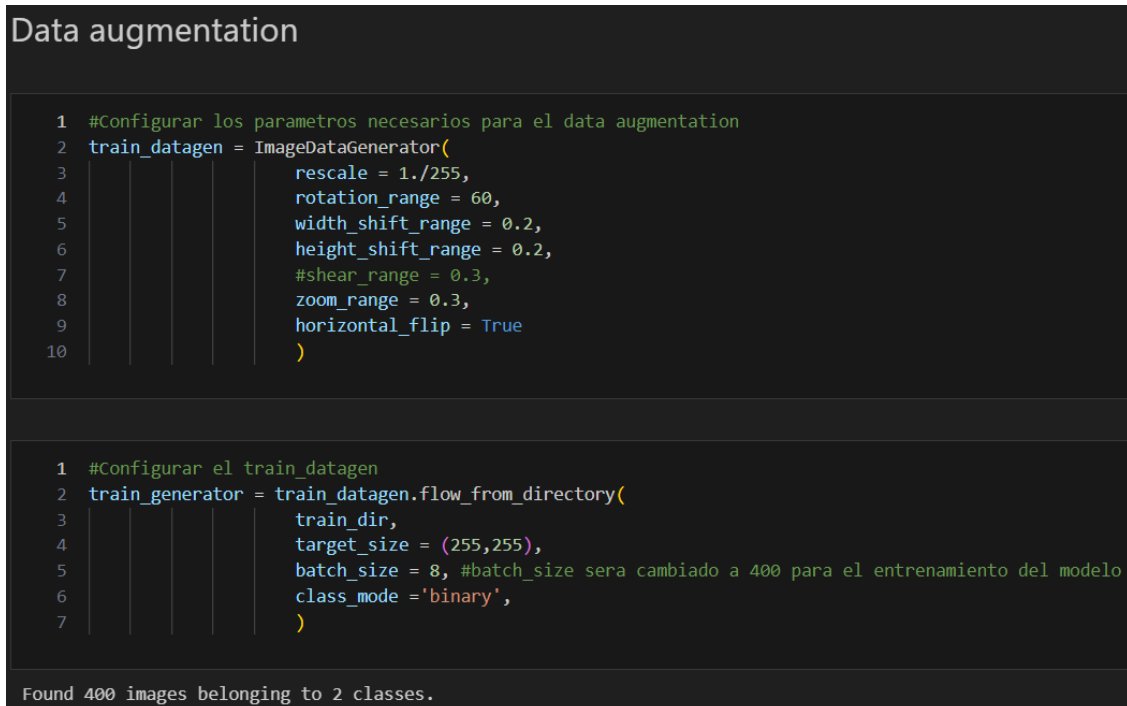
Image Classification (Pizza and Ice cream)

## Selección del dataset

El dataset seleccionado para este modelo es "Pizza vs Ice Cream", fue descargado de la plataforma kaggle ([Pizza vs Ice Cream](#)) el 14 de mayo del 2024. La razón principal para la selección de este dataset fue el objetivo de comparar 2 elementos diferentes usando el modelo a desarrollar, en este caso 2 alimentos diferentes, además de la estructura ya definida del set (con las carpetas train y test con sus respectivos tags) y la cantidad inicial de imágenes contenidas en el dataset (aproximadamente 500 imágenes de train y 50 de test).

Después de ver nuevamente la cantidad de imágenes necesarias para entrenar y verificar el correcto funcionamiento del modelo se modificó la cantidad de imágenes presentes dentro de los datasets para dejar 200 imágenes dentro de test y de train respectivamente. Para poder entrenar de la mejor posible el modelo se hará uso de técnicas de data augmentation para poder suministrar los recursos necesarios para el entrenamiento del modelo.

Ya que se necesita de una gran cantidad de elementos para poder entrenar el modelo de forma adecuada se va a hacer uso de técnicas de “data augmentation” para poder generar los recursos necesarios en el momento del entrenamiento.



The screenshot shows a code editor with Python code for data augmentation. The code is organized into two sections. The first section configures the 'train\_datagen' with parameters for rescaling, rotation, width and height shifts, shearing, zooming, and horizontal flipping. The second section configures the 'train\_generator' by flowing from the directory, setting the target size to (255, 255), batch size to 8, and class mode to 'binary'. A final line of output text at the bottom states 'Found 400 images belonging to 2 classes.'

## Data augmentation

```
1 #Configurar los parametros necesarios para el data augmentation
2 train_datagen = ImageDataGenerator(
3     |         |         |         |         |
4     |         |         |         |         |
5     |         |         |         |         |
6     |         |         |         |         |
7     |         |         |         |         |
8     |         |         |         |         |
9     |         |         |         |         |
10    |         |         |         |         |
11    )

1 #Configurar el train_datagen
2 train_generator = train_datagen.flow_from_directory(
3     |         |         |         |         |
4     |         |         |         |         |
5     |         |         |         |         |
6     |         |         |         |         |
7     |         |         |         |         |
8     |         |         |         |         |
9     |         |         |         |         |
10    |         |         |         |         |
11    )

Found 400 images belonging to 2 classes.
```