[2]:
```
data_dir = "../input/rice-image-dataset/Rice_Image_Dataset" # Datasets path
data_dir = pathlib.Path(data_dir)
data_dir
```

[2]: PosixPath('../input/rice-image-dataset/Rice_Image_Dataset')

Splitting Data on Classes

Inside the data folder there are several folders for different classes.

[3]:
```
arborio = list(data_dir.glob('Arborio/*'))[:600]
basmati = list(data_dir.glob('Basmati/*'))[:600]
ipsala = list(data_dir.glob('Ipsala/*'))[:600]
jasmine = list(data_dir.glob('Jasmine/*'))[:600]
karacadag = list(data_dir.glob('Karacadag/*'))[:600]
```

Changing size of the images:

Since the input dimensions of the MobileNet are (224,224,3). We have to resize our images in the same way.

Currently the size of images is (250,250,3).

```
resized_img = cv2.resize(img, (224, 224))
```

[6]:
```
img = cv2.imread(str(df_images['arborio'][0])) # Converting it into numerical arrays
img.shape # Its currently 250 by 250 by 3
```

[6]: (250, 250, 3)

Link images to different classes

Here we have 5 classes and the images need to be labelled with appropriate classes.

[5]:

```
# Contains the images path
df_images = {
    'arborio' : arborio,
    'basmati' : basmati,
    'ipsala' : ipsala,
    'jasmine' : jasmine,
    'karacadag': karacadag
}

# Contains numerical labels for the categories
df_labels = {
    'arborio' : 0,
    'basmati' : 1,
    'ipsala' : 2,
    'jasmine' : 3,
    'karacadag': 4
}
```

[7]:

```
X, y = [], [] # X = images, y = labels
for label, images in df_images.items():
    for image in images:
        img = cv2.imread(str(image))
        resized_img = cv2.resize(img, (224, 224)) #
        X.append(resized_img)
        y.append(df_labels[label])
```