Problem Description:

The volunteers are asked to upload the pictures they took to a shared Google Drive. There are multiple folders inside the shared folder 1.1 CatsDogs and each folder further contains the pictures of dogs and cats and also contain folder with files inside it and so on. It's like a directory tree.

Juan has uploaded some pictures he took on the 100ANDRO, which is already inside the DCIM and DCIM is further inside the folder 'Picture'. Similarly, Maria's collection is in Pictures/Maria/Pics/CatsAndDogs. John has copied and pasted the pictures on the root folder. So, in general, pictures are randomly distributed and mixed and located in many folders one after another. In order to tackle this situation, someone in the group has decided to use deep learning to classify the images. I have been asked to work this messy directory tree. I have to implement the method 'process_dir', which is supposed to traverse the directory tree and return the two distinct lists: dog_list and cat_list. They contain the paths to all dogs and cats' pictures respectively.

Purposed solution:

In order to resolve the messiness in the shared folder, I worked on the given starter code. Recursion is the key here. Recursion allows a function to call itself where fixed steps of code get executed again and again for new values. We also have to set criteria for deciding where the recursive call ends, which is known as the base case. In this case, new items are added in each recursive call till all the directories and files are traversed.

In this work she implements a method called classify_image that receives the path to an image and return a number between 0 and 1. The closer the number is to 1, the more likely it is to be a dog picture, the closer the number is to 0, the more likely it is to be a cat picture. Since we need the path to an image, absolute path is generated for each one. In each recursive call, each directory is traversed, and pictures of dogs and cats are appended and extended in their respective lists. The method of recursion continues till each folder/file are searched. When all of the folders are traversed, cat_list and dog_list are returned back. Each list contains full paths pictures of dogs and cats respectively.

Test of Algorithm:

In order to test if the algorithm, I ran the code multiple times. Firstly, I ran the code to check the current directory that is given as the startup directory. I worked on the function 'process_dir()' where the function 'get_dirs_and_files()' is called and after that dir_list and file_list which contains absolute paths for both directories and files generated. Next important thing here is the method of recursion. The function process_dir() is called each time when directory contains files inside it. When it is called it traverses over the directory and search for pictures of cats and dogs. This process continues for all the directories which contain file and when it is done cat_list and dog_list are returned back.

Further, I created some test folders as well which may or may not contains the files inside them and ran the code again. It clearly showed that test folders are also traversed like others and pictures are extended in the lists if they are either cats or dogs.

Pictures of input and output:


```
gvin@Govinda:~/Desktop/CS2302_lab1$ python Main.py
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs']
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures']
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Test', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics', '/U
sers/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria']
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Test/Test2']
[]
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/All']
[]
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO']
[]
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics']
['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics/CatsAndDogs']
[]
gvin@Govinda:~/Desktop/CS2302_lab1$
```

Fig1: Showing all the directories.


```
gvin@Govinda:~/Desktop/CS2302_lab1$ python Main.py
CATS_list:  ['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/cat.11.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/P
ictures/cat.12.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/cat.13.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 Ca
tsDogs/Pictures/cat.14.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Test/cat.00.jpg', '/Users/gvin/Desktop/CS230
2_lab1/1.1 CatsDogs/Pictures/Pics/cat.104.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/cat.105.jpg', '/User
s/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/cat.102.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/
cat.103.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/All/cat.129.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1
 CatsDogs/Pictures/Pics/All/cat.128.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/All/cat.130.jpg', '/Users/
gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/cat.19.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/cat
.20.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO/cat.25.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1
 CatsDogs/Pictures/DCIM/100ANDRO/cat.26.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/cat.31.jpg', '/Users/
gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/cat.33.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/c
at.32.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/cat.34.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsD
ogs/Pictures/Maria/Pics/cat.59.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics/cat.58.jpg', '/Users/gvin
/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics/CatsAndDogs/cat.84.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pic
tures/Maria/Pics/CatsAndDogs/cat.85.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics/CatsAndDogs/cat.86.j
pg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/Pics/CatsAndDogs/cat.82.jpg', '/Users/gvin/Desktop/CS2302_lab1
/1.1 CatsDogs/Pictures/Maria/Pics/CatsAndDogs/cat.83.jpg']

DOGS_list:  ['/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/dog.367.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/
Pictures/dog.368.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/dog.310.jpg', '/Users/gvin/Desktop/CS2302_lab
1/1.1 CatsDogs/Pictures/Pics/dog.309.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/dog.308.jpg', '/Users/gvi
n/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/All/dog.344.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/A
ll/dog.345.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Pics/All/dog.347.jpg', '/Users/gvin/Desktop/CS2302_lab1/
1.1 CatsDogs/Pictures/Pics/All/dog.346.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/dog.20.jpg', '/Users/gv
in/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/dog.7.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/dog.6.
jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/dog.8.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pict
ures/DCIM/dog.9.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO/dog.104.jpg', '/Users/gvin/Desktop/CS
2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO/dog.105.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO
/dog.102.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/DCIM/100ANDRO/dog.103.jpg', '/Users/gvin/Desktop/CS2302_la
b1/1.1 CatsDogs/Pictures/DCIM/100ANDRO/dog.101.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/dog.164.jpg',
'/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/dog.159.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Picture
s/Maria/dog.165.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/dog.161.jpg', '/Users/gvin/Desktop/CS2302_lab
1/1.1 CatsDogs/Pictures/Maria/dog.160.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/dog.162.jpg', '/Users/g
vin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/dog.163.jpg', '/Users/gvin/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures/Maria/P
```

Fig2: Showing the separate lists of dogs and cats with their respective full paths.


```
gvin@Govinda:~/Desktop/CS2302_lab1/1.1 CatsDogs/Pictures$ ls
DCIM/        Pics/        cat.11.jpg    cat.13.jpg    dog.367.jpg
Maria/       Test/        cat.12.jpg    cat.14.jpg    dog.368.jpg
```

Fig 3: Showing that I have added the Test folder to check the program works properly.

Conclusions:
From this lab, I learned to work on some messy directories, that may be directory tree, to separate the files on each directory. The method of recursion is well learned that helped to tackle this situation. I also got some idea about the machine learning while going through this problem. Images are classified either 1 or 0, that is used to train the model and the model could be used to predict the pictures coming from other source outside of the data collected. Some other aspects of the programming are also learned and got familiar with such as Lists, append(), extend() methods, absolute paths, etc.

Appendix:
Source code
****

```python
# Govinda KC
# CS 2302 10:30 am class
# Lab 1

import os
import random


def get_dirs_and_files(path):
    #print(path)
    #exit()
    dir_list = [path + '/' + directory for directory in os.listdir(path) if os.path.isdir(path + '/' +
directory)]
    file_list = [path+ '/' + directory for directory in os.listdir(path) if not os.path.isdir(path + '/' +
directory)]
    print(dir_list)
    #print(file_list)
    #exit()
    return dir_list, file_list


def classify_pic(path):
    # To be implemented by Diego: Replace with ML model
    if "dog" in path:
        return 0.5 + random.random() / 2

    return random.random() / 2


def process_dir(path):
    path = os.path.abspath(path)
    cat_list = []
```

```python
    dog_list = []
    dir_list, file_list = get_dirs_and_files(path)
    dir_list = [os.path.abspath(d) for d in dir_list]
    file_list = [os.path.abspath(f) for f in file_list]
    #print(dir_list, file_list)
    for _file in file_list:
        if classify_pic(_file) > 0.5:
            dog_list.append(_file)
        elif 'cat' in _file:
            cat_list.append(_file)
    if len(dir_list) > 0:
        for _dir in dir_list:
            #print(_dir)
            new_cats, new_dogs = process_dir(_dir)
            cat_list.extend(new_cats)
            dog_list.extend(new_dogs)
    return cat_list, dog_list


def main():
    start_path = './' # current directory

    cats, dogs = process_dir(start_path)
    print("CATS_list: ", cats)
    print()
    print("DOGS_list: ", dogs)

main()
```