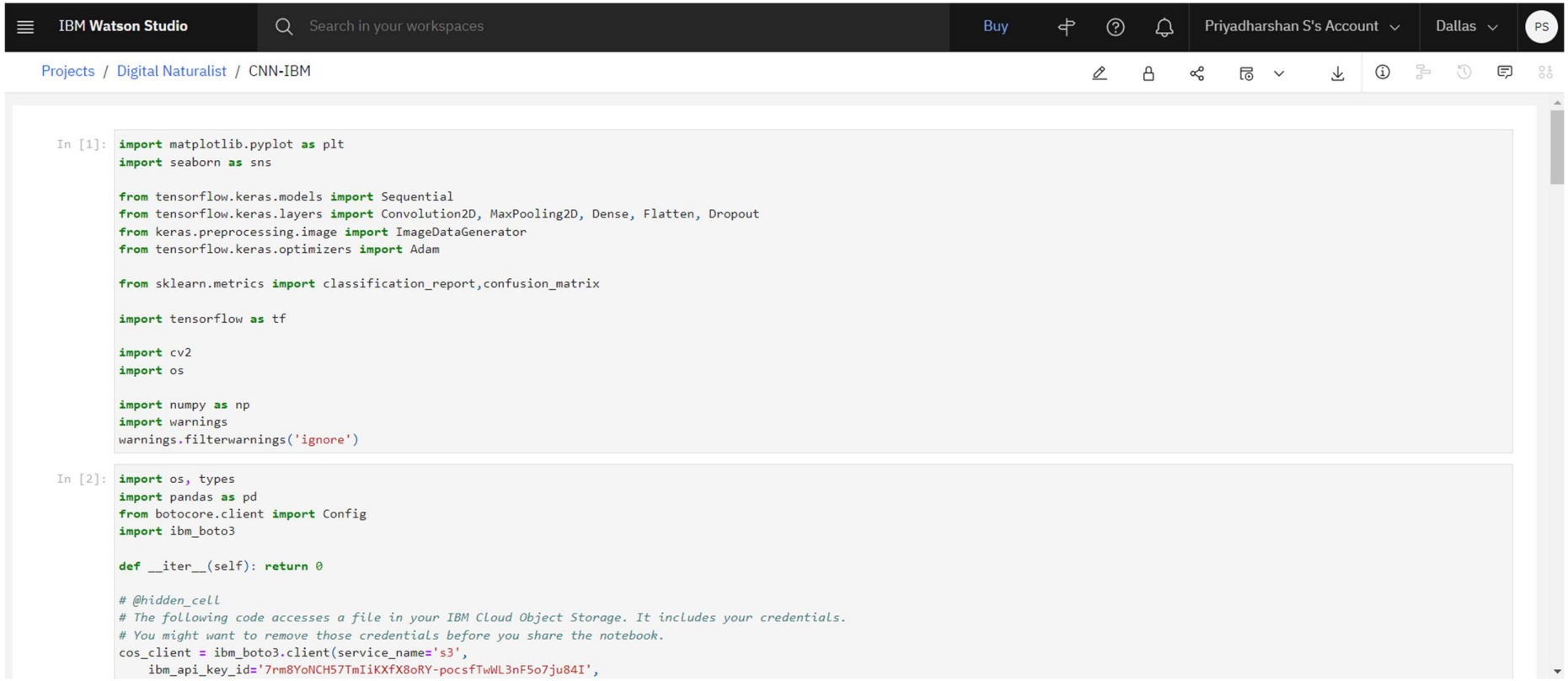
Project Development Phase
Sprint - I

Date	29 October 2022
Team ID	PNT2022TMID35899
Project Name	Digital Naturalist - AI Enabled Tool For Biodiversity Researchers
Maximum Marks	4 Marks

Jupyter Notebook asset in IBM Watson Studio feature of IBM Cloud:



Sprint - I Milestones:

1. Import the Necessary Libraries

In [1]:
```
import matplotlib.pyplot as plt
import seaborn as sns

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Convolution2D, MaxPooling2D, Dense, Flatten, Dropout
from keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.optimizers import Adam

from sklearn.metrics import classification_report,confusion_matrix

import tensorflow as tf

import cv2
import os

import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

2. Upload and Connect Dataset with notebook

```
import os, types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='7rm8YoNCH57TmIiKXfX8oRY-pocsfTwWL3nF5o7ju84I',
    ibm_auth_endpoint="https://iam.cloud.ibm.com/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.private.us.cloud-object-storage.appdomain.cloud')

bucket = 'digitalnaturalist-donotdelete-pr-omp7j8klw6rzuh'
object_key = 'Final Dataset.zip'

streaming_body_5 = cos_client.get_object(Bucket=bucket, Key=object_key)['Body']
```

3. Extracting the Dataset using BytesIO unzip function

In [4]:
```
pwd
```

Out[4]:
```
'/home/wsuser/work'
```

In [3]:
```
from io import BytesIO
import zipfile

unzip = zipfile.ZipFile(BytesIO(streaming_body_5.read()),'r')
file_paths = unzip.namelist()
for path in file_paths:
    unzip.extract(path)
```

In [5]:
```
ls
```

```
'Final Dataset'/
```

4. The dataset directory is divided into training, testing and validation folders

|:
```
direct_of_final_dataset_folder = '/home/wsuser/work/Final Dataset'
```

|:
```
os.listdir(direct_of_final_dataset_folder)
```

[7]:
```
['Testing data', 'Validation data', 'Training data']
```

:
```
train_dir = '/home/wsuser/work/Final Dataset/Training data'
val_dir = '/home/wsuser/work/Final Dataset/Validation data'
test_dir = '/home/wsuser/work/Final Dataset/Testing data'
```

:
```
print(os.listdir(train_dir))
```

```
['Trout', 'Sea Bass', 'rose', 'monkey', 'BARN OWL', 'Red Sea Bream', 'Butterfly', 'squirrel', 'Elephant', 'tulip', 'sunflower', 'pa
nda', 'Black Sea Sprat', 'PINK ROBIN', 'AMERICAN KESTREL', 'PEACOCK', 'dandelion', 'PEREGRINE FALCON', 'BLACK SWAN', 'Dogs', 'Cow',
'Cats', 'Striped Red Mullet', 'Gilt-Head Bream', 'hen', 'OSTRICH', 'Red Mullet', 'horse', 'ALPINE CHOUGH', 'Hourse Mackerel', 'shee
p', 'spider', 'Shrimp', 'daisy', 'BALD EAGLE']
```

5. **Configuring ImageDataGenerator Class**

```
train_datagen = ImageDataGenerator(rescale=1./255,
                                   rotation_range = 20,
                                   horizontal_flip = True)
valid_datagen = ImageDataGenerator(rescale=1./255)
```

6. **Training, Testing and Validation Split-up**

   70% data is split to Training data, 20% data is split to Validation data, 10% data is split to Testing data. Each dataset contain 35 classes. And in each class the Training dataset consist of 150 images, Validation dataset consists of 30 images and the Test dataset consists of 15 images.

7. **Apply ImageDataGenerator Functionality to train set and validation set**

```
train_generator = train_datagen.flow_from_directory(train_dir,
                                                    target_size=(299,299),
                                                    color_mode="rgb",
                                                    class_mode='categorical',
                                                    batch_size=64)
valid_generator = valid_datagen.flow_from_directory(val_dir,
                                                    target_size=(299,299),
                                                    color_mode='rgb',
                                                    class_mode='categorical',
                                                    batch_size=64)
```

```
Found 5250 images belonging to 35 classes.
Found 1050 images belonging to 35 classes.
```

Notebook Link: https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/d8cf79ed-e0b3-44dd-b150-cf2ececc025d?projectid=9f99f93d-6e5b-44eb-ad3d-f2133b037f06&context=cpdaas