GIS Assignment for Fellow Candidates October 2024

Introduction

This document explains the task assignment for a GIS fellow applicant for SDG AI Lab.

1. **Task 1** is checking the candidate's ability to work with Google Earth Engine, data collection and processing.
2. **Task 2** tests the candidate's ability to perform geo-analysis.
3. **Task 3** expects the candidate to work with pretrained models to perform scene classification and evaluation.

The candidate is supposed to use Python scripts with any libraries desired. The delivered code is expected to be commented on as much as possible. Using other methods than described may result in point reductions. Please wrap repeating methods in functions.

Setting-up development environment

1. Create a private GitHub repository under your own account.
2. Please add the users *Skerre* and *sdgailabtest* to your repository with read permission.
3. Set up an Anaconda Virtual Environment for this exercise and submit the solutions as part of the environment. Clean folder structure, file and coding styles will be rewarded.
4. Add your requirements files for us to reproduce your environment.
   Not doing so will deduct points.
5. Datasets for Task 3 are available here: LINK

Task 1 – Cloud service usage and data processing (10 points)

1. Write a script in **Google Earth Engine** that downloads nitrogen dioxide emissions for the country of Moldova and outputs 12 files with mean monthly values for the year 2023. **(4p)**
2. Upload these files in a folder of your GitHub repository and write a python script to
   a. Find the files. **(0.5p)**
   b. Calculate their mean value over the period of 12 months. **(1p)**
   c. Return a resulting geotiff file with a different resolution than its inputs and tightly clip the boundaries to the country outline, if not already done. **(1p)**
   d. Make sure that the file can be run from the terminal and is working on any machine. **(0.5p)**
3. Use a jupyter notebook to load the resulting file and plot it with a library of your choice. **(1p)**
4. Create a gif file for the 12 input files and visualize the change of NO2 emissions over the period of 12 months. Explain what you see. **(2p)**

Task 2 – Data acquisition and analysis (10 points)

1. Acquire a nightlight intensity image of Greater London for the following area: Extent: [-0.91,51.11,0.75,51.84] **(2pt)**
2. Transform that image to geotiff. **(1pt)**
3. Make sure the nightlight image is clipped to the extent. **(1pt)**
4. Download street network data for the area (dataset should include road types, such as 'highway' or similar) **(1pt)**
5. Analyze the correlation between road density (per unit area) and the nightlight intensity in the area. What do you observe? **(2pt)**
6. Create a model that predicts nightlight intensity using road types and road density. Describe which model you use and evaluate the accuracy, results and ways to improve it. **(3pt)**

Task 3 – Binary scene classification using deep learning (15 points)

The files provided in the Google Drive folder for this task are

1. Weights file of the trained model
2. A dataset folder consisting of images with 2 classes.

In this task, we would like to ask the candidates to utilize the trained deep learning model to have predictions for the images given. For this task, please follow the steps below.

1. Import necessary libraries, load the dataset, and visualize a sample image belonging to one of the land use classes in your code file. **(1+1+1 = 3pts)**
2. Initiate a ResNet-50 model and load your model with the provided model weights file. (Hint: You can use CPU as a device instead of GPU if your computer has no GPU. The model needs to be modified to be adapted for a binary classification task). **(3 pts)**
3. Apply necessary transformations to the images before using your model to classify the scenes. Please explain why you made these transformations. **(1.5+1.5 = 3 pts)**
4. Pass the transformed image to the model, obtain the predicted classes and visualize 10 images with corresponding classes. **(1+1+1 = 3 pts)**
5. Report accuracies/confusion matrices of your predictions (accuracy scores are not judged, but the interpretation and commentary) **(3 pts)**