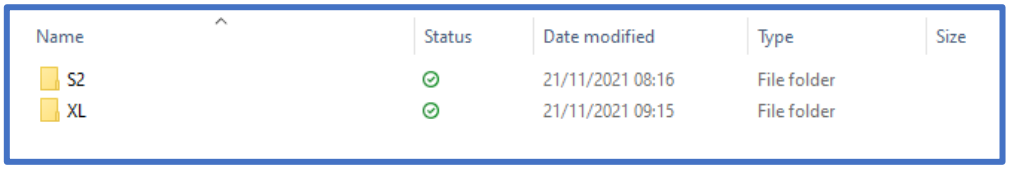
ASSAZZIN Solution-Radiant Earth XL / S2 competitions

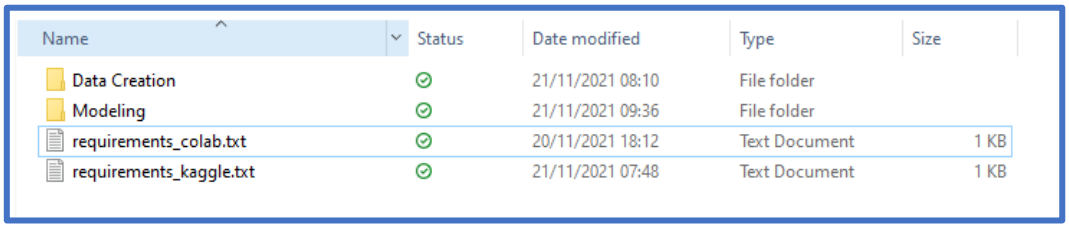
I'll try to add facilitate the reviewer mission with this document

1. **Solution Overview :**
   there is 2 folders XL and S2



   a. **XL : contains notebook that work on both S1+S2 data**
   b. **S2 : contains notebook that work on both S2 data**

2. **S2 Folder**



   a. **Data Creation folder :** contains notebooks that generate the data

More details :

- I provided the necessary code to generate the data that will be used in the training step.
- I tried to extract the first 5 observations for each month , if a month has less than 5 observations , we will replace the row observation by a zeros vector .
- I have 5 notebooks for creating S2 train , 5 notebooks for creating S2 test .
- I used the same code structure provided by radiant earth tutorial , But **I added some tricks that speed up the data generation with less memory consumption** 😊 .
- If you faced a RAM problem , please generate the new data by tiles , I mean run the notebook from 0 : 500 tiles then save output , then from 500 : 1000 tiles then save output … until 2500 : 2650 tiles then save the output .

you'll have 6 outputs saved in the drive or in your local , then with simple code read all 6 csv files and concatenate them . I know it's hard , but hopefully you'll have a good ram for review so

you'll generate the data from single run for each notebook !

- here I provide each notebook name and it's output :

`S2Train_Observation1.ipynb` : Output : TrainObs1.csv
`S2Train_Observation2.ipynb` : Output : TrainObs2.csv
`S2Train_Observation3.ipynb` : Output : TrainObs3.csv
`S2Train_Observation4.ipynb` : Output : TrainObs4.csv
`S2Train_Observation5.ipynb` : Output : TrainObs5.csv

`S2Test_Observation1.ipynb` : Output : TestObs1.csv
`S2Test_Observation2.ipynb` : Output : TestObs2.csv
`S2Test_Observation3.ipynb` : Output : TestObs3.csv
`S2Test_Observation4.ipynb` : Output : TestObs4.csv
`S2Test_Observation5.ipynb` : Output : TestObs5.csv

b. Modelling Folder : contains notebooks that generate the final winning submission

More details about modelling folder :

- **`LGBM folder :`** my lgbm model will run for 13hours on Colab ,

the Colab session timeout is 12hours , so I run LGBM model in 3 separated notebooks .
the 1st & 2nd notebooks will output oof and test_preds .npy files ,
those files will be uploaded from drive in the 3rd notebook to generated the lgbm OOF and TEST_PREDS files .
** 1st and 2nd notebooks are : -- Sentinel_2_LGBM_Folds_0_1_2.ipynb
-- Sentinel_2_LGBM_Folds_3_4_5_6.ipynb
** 3rd notebook is : -- Sentinel_2_LGBM_Folds_7_8_9.ipynb

To resume , for LGBM model the **input** will be **:**
TrainObs1.csv , TrainObs2.csv ,TrainObs3.csv ,TrainObs4.csv ,TrainObs5.csv
TestObs1.csv ,TestObs2.csv ,TestObs3.csv ,TestObs4.csv ,TestObs5.csv

and the **output** will be :
-- S2_LightGBM.csv
-- S2_oof_lgbm.npy

- **`S2_Catboost.ipynb :`**
  - **N**otebook input wil be :
    - TrainObs1.csv , TrainObs2.csv ,TrainObs3.csv ,TrainObs4.csv
    - TestObs1.csv ,TestObs2.csv ,TestObs3.csv ,TestObs4.csv
  - Outputs will be :
    - S2_Catboost.csv
    - S2_oof_cat.npy
- **`xgboost-s2.ipynb -- NNA_S2.ipynb -- NN_s2.ipynb :`**
  - Same as S2_Catboost.ipynb spirit !

- **`StackingBlending_S2.ipynb`** :
  - **N**ow we will use the outputs of the previous notebooks
  - The output will be my S2 competition 5th place submission file .

3. XL Folder :

i. It has same spirit as S2 folder !

ii. **NOTE :** We will need the outputs of S2 notebooks , so Kindly start with S2 Folder .

b. **Data Creation :**

i. It's the same as S2 data creation , but we added here the S1 data creation notebooks .