Radiant Earth Spot the Crop XL Challenge

1. Data Creation :

- 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 create a zeros vector instead
- I have 5 notebooks for creating S2 train , 5 notebooks for creating S2 test , 4 notebooks for creating S1 train , 4 notebooks for creating S1 test .
- I used the same code provided by radiant earth , If you faced a RAM problem , don't worry I'll provide the data generated from each of the 18 notebooks .

- Here I provide each notebook name , it's output and the link :
  - S2Train_Observation1.ipynb :
    - Output : TrainObs1.csv
    - Link : TrainObs1.csv
  - S2Train_Observation2.ipynb :
    - Output : TrainObs2.csv
    - Link : TrainObs2.csv
  - S2Train_Observation3.ipynb :
    - Output : TrainObs3.csv
    - Link : TrainObs3.csv
  - S2Train_Observation4.ipynb :
    - Output : TrainObs4.csv
    - Link : TrainObs4.csv
  - S2Train_Observation5.ipynb :
    - Output : TrainObs5.csv
    - Link : TrainObs5.csv

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

- S1Train_Observation1.ipynb :
  - Output : S1TrainObs1.csv
  - Link : S1TrainObs1.csv
- S1Train_Observation2.ipynb :
  - Output : S1TrainObs2.csv
  - Link : S1TrainObs2.csv
- S1Train_Observation3.ipynb :
  - Output : S1TrainObs3.csv
  - Link : S1TrainObs3.csv
- S1Train_Observation4.ipynb :
  - Output : S1 TrainObs4.csv
  - Link : S1TrainObs4.csv

- S1Test_Observation1.ipynb :
  - Output : S1TestObs1.csv
  - Link : S1TestObs1.csv
- S1Test_Observation2.ipynb :
  - Output : S1TestObs2.csv
  - Link : S1TestObs2.csv
- S1Test_Observation3.ipynb :
  - Output : S1TestObs3.csv
  - Link : S1TestObs3.csv
- S1Test_Observation4.ipynb :
  - Output : S1TestObs4.csv
  - Link : S1TestObs4.csv

2. Modeling :

- I trained for both S2 data and XL Data 5 different models , 3 gradient boosting models and I created 2 different DNN Architectures , 1st one is a combination between classic layers and residual layers . the 2nd one is a combination between Convolutions and attentions .
- All notebooks will use first 4 observation except LGBM ,in LGBM notebook we will the first 5 observations . please refer to the link I provided to accelerate the review process .

- Each notebook will generate a submission file and out of folds numpy array , those outputs will be used in the stackingbelnding notebook where we will first ensemble our submissions , then use stacking strategy using LGBM as a meta model and at final stage we will blend the ensemble and the stacking To generate out final submission file .

- For LightGBM , it takes 1 hour and 30 minutes to complete 1 fold , So it will take 13 hours to complete running 10 folds . for that I decided to train the notebook on 3 parts , each part will generate an out of fold and the test prediction relative to this fold .

3. Environment:

- I trained my models on Both Google colab and Kaggle environment , to reproduce the same score , I will specify each notebook environment .
- Changing the gradient boosting model version will change the score , Please be careful !
- Catboost :
    - **Google colab**
    - GPU : Enabled
    - GPU Machine : Tesla k80
- LightGBM :
    - **Google colab**
    - GPU : disabled
- NN :
    - **Kaggle**
    - GPU : disabled
- NN Attention :
    - **Kaggle**
    - GPU : Enabled
    - GPU Machine : Tesla P100
- XGBOOST:
    - **Kaggle**
    - GPU : Enabled
    - GPU Machine : Tesla P100