SOLUTION DOCUMENTATION FOR RANK 4 RADIANT EARTH SPOT THE CROP CHALLENGE

General Approach:

- Calculated various vegetation indexes and trained them on a LGBMClassifier
- Used raw image pixels to train a neural network.
- Final submission was a weighted average from LGBClassifier predictions and the neural network on a ratio of 50:50

Data Download

Leveraged the starter notebook provided by the organizers to download data and loading the asset file paths into pandas dataframes, i.e for training and testing.

Data Preprocessing

Data provided had lots nulls for some dates. To capture a representative frequency of days during the growing period, a 10 day gap was used. Only images within a 10 day frequency were downloaded.

LGBMClassier

Calculated vegetation indices were trained on a LGBMClassifier. Vegetation indices that captured the soil properties or background of crops proved to be really useful.

LGBM parameters- lgbm_params = {

```
'n_estimators': 50000,
'learning_rate': 0.05,
'colsample_bytree': 0.2,
'subsample': 0.2,
'reg_alpha': 10,
'reg_lambda': 5,
'num_leaves': 20,
'early_stopping_rounds': 300,
'n_jobs': -1,
'objective': 'multiclass',
'boosting': 'gbdt',
'feature_name': 'auto',
```

```
  'categorical_features': 'auto'}
```

Neural nets

The raw image pixels were trained on a pytorch neural net using the pytorch tabular library.

```
Parameters - batch_size = 2048

max_epochs = 300
lr = 1e-3
layers = '2048-1024-512'
target = ['crop_type']
continuous_cols = list(train.drop(['Field ID', 'crop_type'],1).columns)
categorical_cols = []
gpus = 1
activation="LeakyReLU"
```

Hardware Specs:

Data Download, Preprocessing and Experimentation:

Processor: Xeon Platinum 9282 56 cores 112 threads 3.8GHz

RAM: 1024GB

HDD: 5TB SSD

Model Training

LGBM Classifier – Colab PRO – TPU with 20 cores 40 threads

Neural Net – Colab PRO – GPU P100

STEPS TO REPRODUCE SUBMISSION

1. Download data - Plato_Radiant_Data_Download

2. Preprocess data to numpy arrays - Plato_Radiant_Data_Preprocessing

3. Train LGBM model - Plato_Lgbm

- - For faster training, upload the Plato_Lgbm notebook to colab
- - Upload the radiant pixels dataset and the sample submission file
- - Enable TPU which has 40 cores
- - Run all to get the LGBM_SUB file

4. Train Neural Network - Plato_Neural_Net

- - Upload Plato_Neural_Net notebook to colab
- -Upload the Plato_Neural_Net notebook_requirements.txt file
- - Upload the radiant pixels dataset and sample submission file
- - Enable GPU runtime
- - Run all to get the pytorch_tabular file

5. Blend predictions from the two models - Plato_Blend_Predictions

- - Upload the Plato_Blend_Predictions notebook to colab
- - Upload the LGBM_SUB file and the pytorch_tabular file
- - Run all to get the final submission file

Link to data
used: https://drive.google.com/file/d/1vVP0ekUBPXrqG6vaoYb5R6MyyPIJL0pL/view?usp=sharing

Note that the score will vary due to the nature of deep learning randomness irrespective of setting seed.