Machine Learning Team Flow

Before starting with step 1, please do apply for a google earth engine account. [It may take 1-2 days for verification]

Link :- https://developers.google.com/earth-engine

Step 1 : Finalise dataset [Sentinel 2A - to begin with]

After finalising the dataset, we can import it in python using geemap. Then convert it into an array using ee_to_numpy [package needed ee].

Reference : https://omdena.com/blog/yield-prediction/

Step 2 : Water Indices to identify water surface in the image.

We need to compare a few indices to see which all works well for our area. The formulas and bands needed for the same are as follows (note if you cannot see the image it's in the paper shared in this step):

Multiband indices used for water feature extraction.

Multiband Index	Equation	Water Value	
Normalized Difference Vegetation Index	$NDVI = (NIR - Red)/(NIR + Red)$	Negative	
Normalized Difference Water Index	$NDWI = (Green - NIR)/(Green + NIR)$	Positive	
Modified Normalized Difference Water Index	$MNDWI1 = (Green - SWIR1)/(Green + SWIR1)$ $MNDWI2 = (Green - SWIR2)/(Green + SWIR2)$	Positive	
Automated Water Extraction Index	$AWEIsh = Blue + 2.5 \times Green - 1.5 \times (NIR + SWIR1) - 0.25 \times SWIR2$ $AWEInsh = 4 \times (Green - SWIR1) - (0.25 \times NIR + 2.75 \times SWIR1)$	Positive	

The paper below shows how we can consider these indices, work around it. It has done the exact thing for nepal region, that we want to do for initially Kutch then generalise it.

Link :- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111878/

Step 3 - Water Quality Parameter Extraction

We need to derive the parameter using the water indices and bands.

Chl_a = https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2905558/ [reference on how you can search for the formula]

Temperature

Elevation

Turbidity

Salinity

Dissolved Oxygen

pH

and more parameters can be found in the sheet below (2nd Tab):

Sheet : https://docs.google.com/spreadsheets/d/1_uNbSpXKCI0oSsdHCsmXU45l9odqEWaCTs5SUu

Ps: As discussed in the meet we will be selecting one parameter voluntarily and put a message in the slack channel mentioning the parameter you have chosen. You may choose after getting a headstart or before totally upto you. You need to find a dataset [after checking resolution compatibility] or a formula to compute the parameters from the bands or water indices extracted from the Sentinel 2A dataset. Anyone finding a better formula can send it too. You also need to code the formula in the chosen language.

Step 4 - Multiclass Classification

We can look into the various machine learning and deep learning models for this task and see which performs the best. Before starting with this we need to finalise the classes.

```
In [ ]:
```