Datasets
========

Ktk folder contains 3 datasets for the northern karnataka region-

1. test.txt & label.txt are preprocessed ndvi data and its label respectively for training (for 2016-2019) while test_img.txt & test_label.txt are for testing (for 2020) the model
2. SPEI_ 75.750000_ 14.750000.csv is a SPEI dataset for the years 1901-2018.
3. The SMI dataset we have is for 2 seasons i.e, monsoon and post-monsoon for the years 2018-2020.

Notebooks
=========

In the notebooks folder, we have two python notebooks namely 'Geospatial Data Analysis.ipynb' and 'Drought_Assessment KTK.ipynb'. Geospatial Data Analysis.ipynb mainly contains preprocessing steps for raw satellite data (in our case Landsat-8 data). After preprocessing the raw data(which was satellite images), we got a list of arrays which represents ndvi values corresponding to the image taken for preprocessing. We stored this ndvi data in the text files which later was used by another notebook to train the deep learning model. To get labels corresponding to each ndvi data, we first calculated ndvi mean value of the ndvi array and then again a mean of these ndvi mean values. Ndvi deviation was calculated for each data, based on the difference between ndvi mean and mean of ndvi mean. These ndvi deviation value was used to assign labels to the data as follows-

1) NDVI Deviation <= -0.2
   label= Severe Drought (extremely dry)
2) -0.2 < NDVI Deviation <= -0.05
   label= Drought (moderately dry)
3) -0.05 < NDVI Deviation <= 0.1
   label= Near normal
4) 0.1 < NDVI Deviation
   label= above optimum (extremely wet)

Similarly, SPEI data was labeled as follows-

1) SPEI <= -1.65
   label= Very Extreme Drought
2) -1.65 < SPEI <= -1.28
   label= Extreme Drought
3) -1.28 < SPEI <= -0.84
   label= Severe Drought
4) -0.84 < SPEI <= -0.5
   label= Moderate Drought
5) -0.5 < SPEI
   label= No Drought

Note:- these thresholds are taken from some research paper to get the labels

Models

=======

We have implemented two deep learning models to train on preprocessed NDVI data and SPEI data. For NDVI data we used a time distributed cnn network, where we pass 2 past samples as input and predict the label of the next sequential sample. We achieved 96% accuracy in this case. For SPEI data, we used the lstm network where we pass past 4 spei samples as input and predict the next spei value in the sequence. We achieved 74% of cosine similarity score in this case.