The ML Coursework 2025A1 dataset was created for the 2025 Machine Learning module assessment. It is made up of data extracted from FAO and NASA databases.

The dataset contains 17 csv files, each covering a category of variables. See each csv file for the variables it contains (you can open each file using Microsoft Excel). The table below provides descriptions of the variables. Note that the files do not have the same number of rows and columns due to differences in resolution (see Resolution column in table below) and organization. However, longitude and latitude variables (and year also when relevant) are common to all files.

| **Variable**  **(\*file not variable)** | **Full name** | **Additional info** | **Unit** | **Resolution** | **Source info summary** |
| --- | --- | --- | --- | --- | --- |
| longitude | Longitude | where "centroid" is included in the variable name, the value refers to the midpoint for the given country | Degrees east | 1.0 degree (except for *country\_longitu de\_latitude\_area\_lookup* file) |  |
| latitude | Latitude | where "centroid" is included in the variable name, the value refers to the midpoint for the given country | Degrees north | 1.0 degree (except for *country\_longitu de\_latitude\_area\_lookup* file) |  |
| country | Country |  |  |  |  |
| *country\_longitude\_latitude\_area\_lookup*\* | Lookup table to match country by name to centroid longitude and latitude for the country |  |  |  | https://developers.google.com/public-data/docs/canonical/countries\_csv |
| centroid radius |  | You can use this to match country centroid longitude and latitude to other longitude and latitude data.  You can ignore the area variable. It was only include for computing the radius. The area data was obtained from https://unstats.un.org/unsd/environment/totalarea.htm | Degrees |  | Computed from the area assuming circular region (i.e. area pi\*radius2 ) and assuming that 1 degree (latitude/longitude) is equivalent to 100km |
| year | Year | Months 1, 2, 3, ... in a year always correspond to Jan, Feb, March, ... |  |  |  |
| Snowf\_tavg | Snow precipitation rate | Rate of snow fall | Kilogram per squared metre per second | per longitude and latitude pair per month per year | https://disc.gsfc.nasa.gov/datasets/GLDAS\_CLSM10\_M\_2.1/summary |
| Rainf\_tavg | Rain precipitation rate | Rate of rainfall | Kilogram per squared metre per second |
| TVeg\_tavg | Transpiration | Evaporation of water from plant | Watts per squared metre |
| ESoil\_tavg | Direct Evaporation from Bare Soil |  | Watts per squared metre |
| CanopInt\_inst | Plant canopy surface water | Water on plant surfaces | Kilogram per squared metre |
| TWS\_inst | Terrestrial water storage | Typical indicator of drought | millimetre |
| SoilMoi0\_10cm\_inst | Soil moisture at 0-10cm |  | Kilogram per squared metre | per longitude and latitude pair per month per year | https://disc.gsfc.nasa.gov/datasets/GLDAS\_NOAH10\_M\_2.1/summary |
| SoilMoi10\_40cm\_inst | Soil moisture at 10-40cm |  | Kilogram per squared metre |
| SoilMoi40\_100cm\_inst | Soil moisture at 40-100cm |  | Kilogram per squared metre |
| SoilMoi100\_200cm\_inst | Soil moisture at 100-200cm |  | Kilogram per squared metre |
| SoilTMP0\_10cm\_inst | Soil temperature at 0-10cm |  | Kelvin |
| SoilTMP10\_40cm\_inst | Soil temperature at 10-40cm |  | Kelvin |
| SoilTMP40\_100cm\_inst | Soil temperature at 10-400cm |  | Kelvin |
| SoilTMP100\_200cm\_inst | Soil temperature at 100-200cm |  | Kelvin |
| Land\_cover\_percent | Percentage of each of 17 land cover classes | Land cover classes code & description  1 – Evergreen needleleaf forests  2 – Evergreen broadleaf forests  3 – Deciduous needleleaf forests  4 – Deciduous broadleaf forests  5 – Mixed forests  6 – Closed shrublands  7 – Open shrublands  8 – Woody savannas  9 – Savannas  10 – Grasslands  11 – Permanent wetlands  12 – Croplands  13 – Urban and built-up lands  14 – Cropland / Natural vegetation mosaics  15 – Permanent snow and ice  16 – Barren  17 – Water bodies | % | per year per longitude and latitude pair | https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD12C1#overview  [Land cover classes - https://modis-land.gsfc.nasa.gov/pdf/MCD12Q1\_C6\_Userguide04042018.pdf] |
| Yield | Yield |  | Kilogram per hectare | per year per country per crop category | https://www.fao.org/faostat/en/#data/QCL |
| Production | Production quantity |  | tonnes |