

## Dataset Introduction

The ML Coursework Dataset was created for the 2024 Machine Learning module assessment. It is made up of data extracted from the FAOSTAT database (<https://www.fao.org/faostat/en/#home>), which gives open access to food and agricultural data for over 245 countries and covers years from mid 1990s to present day.

## Dataset Content

The ML Coursework Dataset contains 13 csv files, each covering a category of variables relevant to food and agriculture. See below for the 13 categories covered (with the corresponding FAOSTAT source for each category).

1. Consumer prices indicators (<https://www.fao.org/faostat/en/#data/CP>)
2. Crops production indicators (<https://www.fao.org/faostat/en/#data/QCL>)
3. Emissions (<https://www.fao.org/faostat/en/#data/GCE> & <https://www.fao.org/faostat/en/#data/GV>)
4. Employment (<https://www.fao.org/faostat/en/#data/OEA>)
5. Exchange rate (<https://www.fao.org/faostat/en/#data/PE>)
6. Fertilizers use (<https://www.fao.org/faostat/en/#data/RFB>)
7. Food balances indicators (<https://www.fao.org/faostat/en/#data/FBS>)
8. Food security indicators (<https://www.fao.org/faostat/en/#data/FS>)
9. Food trade indicators (<https://www.fao.org/faostat/en/#data/TCL>)
10. Foreign direct investment (<https://www.fao.org/faostat/en/#data/FDI>)
11. Land temperature change (<https://www.fao.org/faostat/en/#data/ET>)
12. Land use (<https://www.fao.org/faostat/en/#data/RL>)
13. Pesticides use (<https://www.fao.org/faostat/en/#data/RP>)

See <https://www.fao.org/faostat/en/#definitions> for details about the headers in the data files (Note - 'Area' in the headers refers to 'Country' in this FAOSTAT glossary). Note that the data files do not have the same number of rows and columns due to differences in the year period and the number of variables (and number of variable sub-categories) covered. However, year and country variables are common to all files. The table below provides a summary of the variables included in each file.

<b>Data file/category</b>	<b>Summary of variables (variable units are indicated in the file, where applicable)</b>
Consumer price indicators	<ul style="list-style-type: none"> <li>• Consumer price, food index</li> <li>• Country</li> <li>• Food price inflation</li> <li>• Month</li> <li>• Year</li> </ul>
Crops production indicators	<ul style="list-style-type: none"> <li>• Country</li> <li>• Year</li> <li>• Yield for different crop products</li> </ul>
Emissions	<ul style="list-style-type: none"> <li>• Country</li> <li>• Crops CH4 emissions</li> <li>• Crops N2O emissions</li> <li>• Drained soil CO2 emissions</li> <li>• Drained soil N2O emissions</li> <li>• Year</li> </ul>
Employment	<ul style="list-style-type: none"> <li>• Country</li> <li>• Employment (male and female total) in agriculture, forestry, and fishing</li> <li>• Mean weekly hours worked per person (no distinction between male and female) in agriculture, forestry, and fishing</li> <li>• Year</li> </ul>
Exchange rate	<ul style="list-style-type: none"> <li>• Country</li> <li>• Currency</li> <li>• Local currency units per USD</li> <li>• Months</li> <li>• Year</li> </ul>
Fertilizers use	<ul style="list-style-type: none"> <li>• Agricultural use of different categories of fertilizers</li> <li>• Country</li> <li>• Year</li> </ul>
Food balances	<ul style="list-style-type: none"> <li>• Country</li> <li>• Export quantity for different crop and livestock products</li> <li>• Food uses for different crop and livestock products</li> <li>• Import quantity for different crop and livestock products</li> <li>• Losses for crop and livestock products</li> <li>• Other uses for crop and livestock products</li> <li>• Year</li> </ul>
Food security	<ul style="list-style-type: none"> <li>• Cereal import dependency ratio</li> <li>• Country</li> <li>• Dietary energy supply adequacy</li> <li>• Per capita food production variability</li> <li>• Per capita food supply variability</li> </ul>

Data file/category	Summary of variables (variable units are indicated in the file, where applicable)
	<ul style="list-style-type: none"> <li>• Percentage of arable land equipped for irrigation</li> <li>• Political stability and absence of violence/terrorism index</li> <li>• Prevalence of anaemia in women of reproductive age</li> <li>• Prevalence of low birthweight</li> <li>• Protein energy supply</li> <li>• Value of food imports in total merchandise exports</li> <li>• Year</li> </ul>
Food trade	<ul style="list-style-type: none"> <li>• Country</li> <li>• Export value</li> <li>• Import value</li> <li>• Year</li> </ul>
Foreign direct investment (FDI)	<ul style="list-style-type: none"> <li>• Country</li> <li>• FDI inflows to agriculture, forestry, and fishing</li> <li>• FDI inflows to food, beverages, and tobacco</li> <li>• FDI outflows to agriculture, forestry, and fishing</li> <li>• FDI outflows to food, beverages, and tobacco</li> <li>• Total FDI inflows</li> <li>• Total FDI outflows</li> <li>• Year</li> </ul>
Land temperature change	<ul style="list-style-type: none"> <li>• Country</li> <li>• Months</li> <li>• Temperature change</li> <li>• Standard deviation</li> <li>• Year</li> </ul>
Land use	<ul style="list-style-type: none"> <li>• Area for different categories of land use</li> <li>• Country</li> <li>• Year</li> </ul>
Pesticides use	<ul style="list-style-type: none"> <li>• Agricultural use for each of fungicides (and bactericides), herbicides, insecticides, pesticides, rodenticides</li> <li>• Country</li> <li>• Use per area of cropland for each of fungicides (and bactericides), herbicides, insecticides, pesticides, rodenticides</li> <li>• Use per value of agricultural production for each of fungicides (and bactericides), herbicides, insecticides, pesticides, rodenticides</li> <li>• Year</li> </ul>