

Dataset Selection Form

Group Members:

Umer	Faisal	Umer.Faisal@city.ac.uk
Inthirajith	Sinniah	Inthirajith.Sinniah@city.ac.uk
Kaloyan	Dimitrov	Kaloyan.Dimitrov@city.ac.uk
William	Dibble	William.Dibble@city.ac.uk

Dataset source link:

<https://www.kaggle.com/datasets/brsdincer/all-natural-disasters-19002021-eosdis>

Dataset size (and number of features, if appropriate): **Size:**

16126 Observations

45 Variables

Dataset description: (One or two sentences):

All-Natural Disaster 1900-2021

As we know, the global climate disaster makes its impact on us more felt day by day. Understanding the parameters created by the climate crisis will be helpful in deciding the measures we will take against it.

In this dataset, you will see the natural disasters of all countries.

Dataset, does your source come with copyright information? (Yes/no/details if appropriate) – **Yes**

Database Contents License (DbCL) v1.0

<https://opendatacommons.org/licenses/dbcl/1-0/>

Problem to be solved: (One or two sentences. Including the kind of problem, e.g., classification or regression)

Predicting when and where a natural disaster/s is/are likely to occur using regression.

Github link:

<https://github.com/WDibble/IntroAI-Team14>

Feedback on your Dataset

Suitability score: +1

(+1 means suitable, 0 means adequate for task but might not provide a strong challenge, -1 means unsuitable for the task)

Brief comments: (One or two sentences)

Dataset looks suitable for the problem. You will need to carefully perform pre-processing of the data. Apparently it seems some columns might be needed to delete (there are too many null values, or they contain too less information). Exploratory data analysis will help here. Your objective seems not very clear, it has two parts “**predicting i) when and ii) where a natural disaster will occur**”. From “**where** a disaster occur” possibly you mean country(?) it will be a classification problem. And if you want to predict year (the **where** part), it is regression.