

Rising Waters: A Machine Learning Approach to Flood Prediction

DATE	28-02-2026
TEAM ID	LTVIP2026TMIDS89043
PROJECT NAME	Rising Waters: A Machine Learning Approach to Flood Prediction
MAXIMUM MARKS	2 MARKS

6.2 DATA COLLECTION:

- Download the dataset from the below link
- <https://www.kaggle.com/datasets/arbethi/rainfall-dataset>
- Login to your Kaggle account.
- Open your dataset page (using your link or search it).
- Click the Download button (top right side).
- The file will download as a ZIP file.
- Go to your Downloads folder.
- Right-click the ZIP file → Click Extract All.
- Open the extracted folder — your dataset (CSV/Excel file) is ready

The screenshot shows the Kaggle website interface. On the left, there's a sidebar with navigation links like Home, Competitions, Datasets, Models, Benchmarks, Game Arena, Code, Discussions, Learn, and More. The main content area is titled 'Rainfall dataset'. It features a 'Data Card' tab, a 'Code (0)' link, a 'Discussion (0)' link, and a 'Suggestions (0)' link. Below the tabs, there's a file named 'flood dataset.xlsx' (16.03 kB). To its right is a 'Download' button and a 'Suggest Edits' link. Further down, there's an 'About this file' section stating 'This file does not have a description yet.' On the far right, there's a 'Data Explorer' section showing the file structure: 'flood dataset.xlsx' (version 2, 531.02 kB), containing 'Sheet5' and 'rainfall in india 1901-2015'. Below that is a 'Summary' section indicating 2 files and 31 columns.

Create the folder called data and put the dataset 'flood dataset.xlsx' in it.

IMPORT THE DATASET:

Import the dataset to work by given code

Rising Waters: A Machine Learning Approach to Flood Prediction

```
#read the dataset  
dataset=pd.read_excel('flood dataset.xlsx')
```

The screenshot shows a Python development environment with a project structure on the left and a code editor on the right. The code editor contains a file named 'sample.py' with the following content:

```
1 import numpy as np  
2 import pandas as pd  
3 import matplotlib.pyplot as plt  
4 import seaborn as sea  
5  
6 dataset = pd.read_excel('data/flood dataset.xlsx')
```

The output window below the code editor shows the execution results:

```
C:\Users\CJohn\AppData\Local\Programs\Python\Python314\python.exe "E:\AI&ML PROJECT\sample.py"  
   Temp Humidity Cloud Cover ANNUAL ... Oct-Dec avgjune sub flood  
0     29       78      30  3248.6 ...  666.1  274.866667  649.9    0  
1     28       75      40  3326.6 ...  658.2  130.300000  256.4    1  
2     28       75      42  3271.2 ...  570.1  186.200000  308.9    0  
3     29       71      44  3129.7 ...  365.3  366.066667  862.5    0  
4     31       74      40  2741.6 ...  458.1  283.400000  586.9    0  
..   ...       ...      ...  ...  ...  ...  ...  ...  
110   28       71      30  3035.1 ...  446.3  262.833333  664.3    0  
111   29       71      37  2151.1 ...  309.8  143.433333  335.0    0  
112   30       74      42  3255.4 ...  431.8  347.566667  923.4    1  
113   31       71      31  3046.4 ...  502.1  151.466667  203.4    0  
114   28       71      34  2600.6 ...  611.1  187.866667  361.8    0  
  
[115 rows x 11 columns]  
<class 'pandas.DataFrame'>  
RangeIndex: 0 to 114  
Data columns (total 11 columns):
```

The status bar at the bottom indicates the file is 196:3, uses CRLF line endings, is in UTF-8 encoding, has 4 spaces indentation, and is using Python 3.14.