**Flood Prediction Dataset**

Regression with flood prediction on full dataset

## **About Dataset**

Flood detection refers to identifying, monitoring, and alerting authorities or individuals about the presence or likelihood of flooding in a particular area. It involves using various technologies and methods to detect, predict, and mitigate the impacts of floods. Flood prediction is a critical area of research due to its significant impact on human life, infrastructure, and the environment. Accurate flood prediction models can aid in disaster preparedness and risk management, reducing the adverse effects of floods.

* The dataset used in this study, sourced from flood.csv, contains multiple features relevant to flood prediction, including environmental factors and socio-economic indicators. The Dataset has 50000 rows and 21 columns.
* The dataset includes 21 numeric variables such as 'MonsoonIntensity,' 'TopographyDrainage,' 'RiverManagement,' 'Deforestation,' 'Urbanization,' 'ClimateChange,' 'DamsQuality,' 'Siltation,' 'AgriculturalPractices,' 'Encroachments,' 'IneffectiveDisasterPreparedness,' 'DrainageSystems,' 'CoastalVulnerability,' 'Landslides,' 'Watersheds,' 'DeterioratingInfrastructure,' 'PopulationScore,' 'WetlandLoss,' 'InadequatePlanning,' 'PoliticalFactors,' and 'FloodProbability.'
* There are no missing values in these numeric columns.
* The dataset includes no categorical variables. All columns are int64 type. This makes the dataset generally suitable for regression models without extensive pre-processing.

**IMPORTANT FEATURES OF DATASET:**  
The dataset used for this analysis comprises numerous features, providing essential insights into environmental and human factors that influence flood occurrence and severity. These features include:

1. **Monsoon Intensity:** Higher volumes of rain during monsoons increase the probability of floods.
2. **Topography Drainage:** The drainage capacity based on the region's topography. Efficient drainage can help drain rainwater and reduce the risk of floods.
3. **River Management:** The quality and effectiveness of river management practices. Proper river management, including dredging and bank maintenance, can improve water flow and reduce floods.
4. **Deforestation:** The extent of deforestation in the area.\* Deforestation reduces the soil's ability to absorb water, increasing surface runoff and the risk of floods.
5. **Urbanization**: The level of urbanization in the region. Urban areas have impermeable surfaces (asphalt, concrete), which reduce water infiltration, raising the risk of floods.
6. **Climate Change:** The impact of climate change on the region. Climate change can lead to more extreme precipitation patterns, including torrential rains that can cause floods.
7. **Dams Quality:** The quality and maintenance status of dams. Well-maintained dams can control floods, and dams with structural problems can break and cause catastrophic floods.
8. **Siltation**: The extent of siltation in rivers and reservoirs. The accumulation of sediments in rivers (siltation) reduces drainage capacity and increases the risk of floods.
9. **Agricultural Practices:** The types and sustainability of agricultural practices. The intensification of agriculture can lead to deforestation, excessive use of fertilizers and pesticides, and inappropriate irrigation practices, reducing soil biodiversity and increasing the risk of floods.
10. **Encroachments:** The degree of encroachment on flood plains and natural waterways. Construction in flood-prone areas impedes the natural flow of water and increases the risk of floods.
11. **Ineffective Disaster Preparedness:** The lack of emergency plans, warning systems, and simulations increases the negative impact of floods.
12. **Drainage Systems:** Well-maintained and adequately sized drainage systems help drain rainwater and reduce the risk of floods.
13. **Coastal Vulnerability:** Low-lying coastal areas are prone to flooding from storm surges and sea level rise.
14. **Landslides:** Steep slopes and unstable soils are more prone to landslides.
15. **Watersheds:** Regions with more watersheds may have a higher or lower risk of flooding, depending on various factors.
16. **Deteriorating Infrastructure:** Clogged culverts, damaged drainage channels, and other deficient infrastructure can increase the risk of floods.
17. **Population Score:** Densely populated areas can suffer more severe losses.
18. **Wetland Loss:** Wetlands act as natural sponges, absorbing excess water and helping to prevent floods.
19. **Inadequate Planning:** Urban planning that does not consider the risk of flooding increases the vulnerability of communities.
20. **Political Factors:** Factors such as corruption and a lack of political will to invest in drainage infrastructure can make it difficult to manage flood risk.
21. **Flood Probability:** The overall probability of flooding in the region. This is the target variable for predictive analysis. This is the target Variable.