
1. INTRODUÇÃO

O presente relatório atende à proposta da Global Solution do 1º Semestre de 2025, que desafia os estudantes a desenvolverem soluções tecnológicas para enfrentar os impactos de eventos climáticos extremos. O objetivo deste trabalho é realizar uma análise estatística sobre dados de risco de enchentes e desenvolver um modelo preditivo inicial para identificar fatores de risco.

Para esta análise, foi utilizada a base de dados pública "Flood Risk Dataset India", obtida na plataforma Kaggle. O processo envolveu a análise de frequência de variáveis, a criação de visualizações, o cálculo de estatísticas descritivas e a implementação de um modelo de Regressão Linear Simples para avaliar a relação entre a precipitação e a ocorrência de enchentes.

2. ANÁLISE DE FREQUÊNCIA

A análise de frequência permite entender a distribuição dos dados em variáveis chave.

Tabela 1: Frequência da Variável Discreta 'Historical Floods'

Frequência	
Historical	Floods
0	5013
1	4987

Interpretação: A Tabela 1 mostra a frequência de registros com e sem enchentes históricas. Observa-se que a base de dados é notavelmente balanceada, com um número quase idêntico de ocorrências para ambas as categorias (5013 para "não" e 4987 para "sim"). Isso é positivo, pois evita vieses em modelos preditivos que poderiam ser causados por uma representação desproporcional de uma das classes.

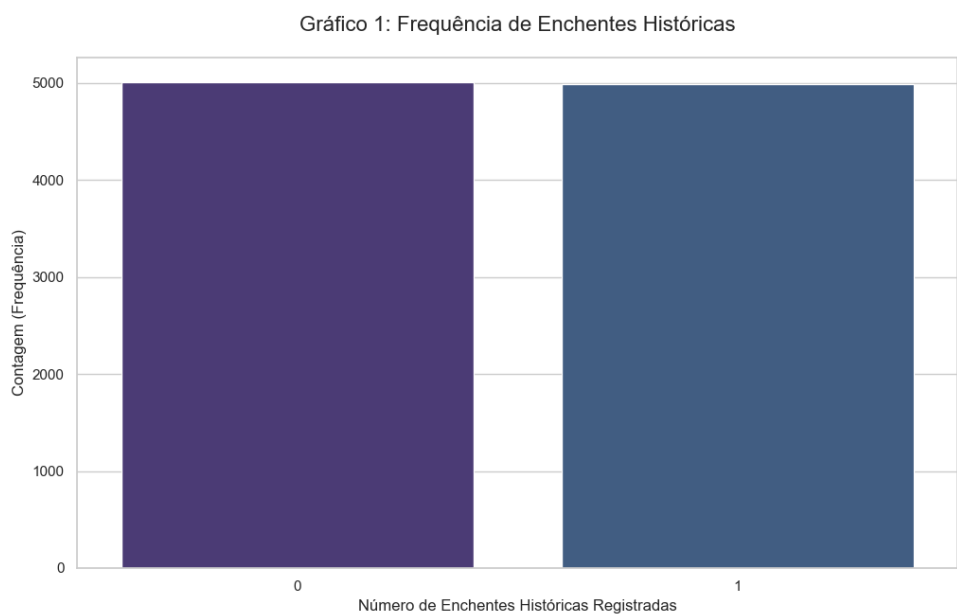
Tabela 2: Frequência da Variável Contínua 'Rainfall (mm)'

Frequência	
Rainfall (mm)	
[0.0144, 37.509)	1248
[37.509, 75.003)	1217
[75.003, 112.498)	1261
[112.498, 149.992)	1252
[149.992, 187.487)	1276
[187.487, 224.981)	1297
[224.981, 262.476)	1234
[262.476, 300.27)	1215

Interpretação: A Tabela 2 agrupa os dados de precipitação em 8 faixas (classes). A distribuição de frequência entre as classes é bastante uniforme, sem uma concentração excessiva em uma faixa específica. Isso sugere que a base de dados cobre uma ampla e bem distribuída gama de volumes de chuva.

3. ANÁLISE VISUAL DOS DADOS

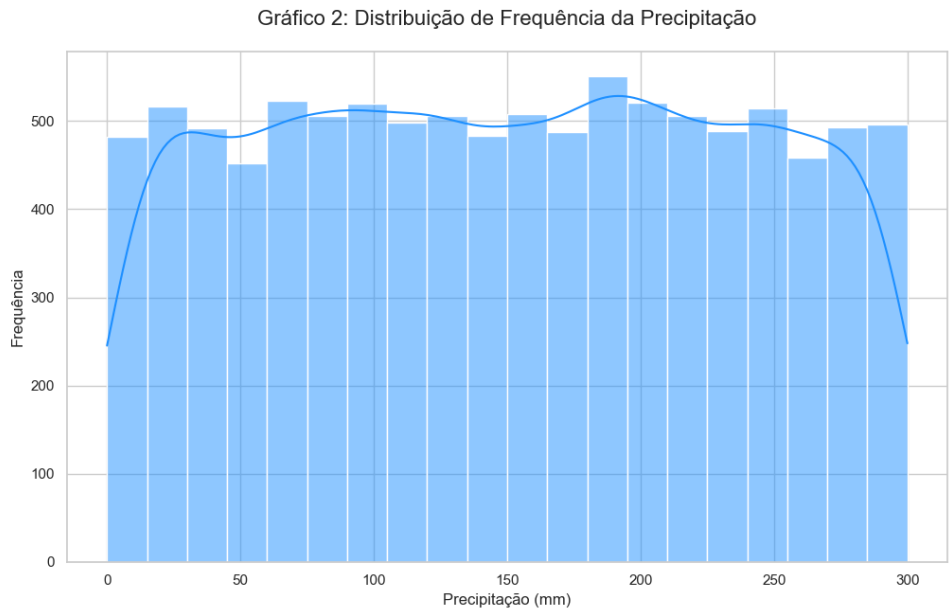
Gráfico 1: Frequência de Enchentes Históricas



Interpretação: O gráfico de barras ilustra visualmente o equilíbrio apresentado na Tabela 1. As duas barras têm alturas muito semelhantes, reforçando a conclusão

de que há uma paridade entre os registros com e sem histórico de enchentes na base de dados.

Gráfico 2: Distribuição de Frequência da Precipitação



Interpretação: O histograma mostra a distribuição dos dados de precipitação. A forma do gráfico se assemelha a uma distribuição uniforme, sem picos ou vales acentuados, o que confirma a análise da Tabela 2. A linha de densidade (KDE) sobreposta suaviza a distribuição e também evidencia essa uniformidade.

4. ANÁLISE ESTATÍSTICA DESCRITIVA

Foi realizada uma análise descritiva detalhada sobre a variável de precipitação para extrair medidas de tendência central, dispersão e separatrizes.

Tabela 3: Medidas Estatísticas da Variável 'Rainfall (mm)'

Medida Valor	
Média 150.02 Mediana 150.62 Moda 0.01	
Mínimo 0.01 Máximo 299.97 Amplitude Total 299.96 Variância 7,401.53	
Desvio Padrão 86.03 Coeficiente de Variação (%) 57.35%	
1º Quartil (Q1) 76.12 2º Quartil (Q2 - Mediana) 150.62 3º Quartil (Q3) 223.40	

Interpretação: A média (150.02 mm) e a mediana (150.62 mm) são muito próximas, o que sugere uma distribuição simétrica, sem valores extremos

puxando a média para um lado. O Desvio Padrão de 86.03 mm, combinado com um Coeficiente de Variação de 57.35%, é considerado alto e indica uma dispersão significativa dos dados em torno da média. Isso significa que, embora a média seja 150 mm, os valores de precipitação variam consideravelmente. Os quartis mostram que 50% dos dados centrais estão concentrados entre 76.12 mm e 223.40 mm.

5. ANÁLISE DO MODELO PREDITIVO (REGRESSÃO LINEAR)

Para atender aos requisitos da avaliação, foi construído um modelo de Regressão Linear Simples com o objetivo de prever a ocorrência de enchentes ('Flood Occurred') utilizando como variável preditora a precipitação ('Rainfall (mm)').

O modelo treinado gerou a seguinte equação: Probabilidade de Enchente = $-0.000057 * (\text{Rainfall (mm)}) + 0.5114$

Para avaliar a eficácia do modelo, foi calculado o coeficiente de determinação, ou R-quadrado (R^2).

Score R^2 (Teste): -0.0015 (-0.15%)

Interpretação do Resultado: O R^2 é uma métrica que indica o quanto o modelo é capaz de explicar a variação dos dados. Um R^2 de -0.0015 é um resultado estatisticamente muito pobre. Um valor negativo significa que o modelo linear é pior do que simplesmente usar a média como previsão para todos os casos. Esta é a descoberta mais crucial desta análise: não há uma relação linear simples entre o volume de precipitação e a ocorrência de uma enchente. Portanto, um modelo de Regressão Linear Simples é inadequado para este problema.

6. CONCLUSÃO GERAL

A análise estatística dos dados sobre risco de enchentes na Índia revelou um dataset balanceado e com ampla distribuição de valores de precipitação. As análises de frequência e descritiva forneceram um panorama claro das características dos dados.

O desenvolvimento do modelo de Regressão Linear Simples, embora tenha sido executado com sucesso, demonstrou de forma conclusiva sua ineficácia para prever um fenômeno complexo como uma enchente usando apenas uma variável. O R^2 negativo é a evidência estatística que comprova essa limitação.

Conclui-se que, para a criação de uma solução tecnológica eficaz, seria imprescindível o uso de modelos de classificação mais sofisticados (como

Regressão Logística, Árvores de Decisão ou Redes Neurais) e a inclusão de múltiplas variáveis, como tipo de solo, elevação e descarga de rios, para capturar a natureza multifatorial do risco de enchentes. O trabalho cumpriu seu objetivo ao não só aplicar as técnicas estatísticas, mas também ao extrair insights valiosos sobre a complexidade do problema e a limitação de modelos simplistas.