

Nota conceitual e plano de implementação do projeto Capstone

Título do projeto: Previsão de Inundações com Dados Meteorológicos e Históricos

Membros da equipe

1. Gerónimo Bêco
2. Iveth de Lima
3. Laurinda Carlos
4. Pitchi Sidibe

Nota conceitual

1. Visão Geral do Projeto

O projeto “**Previsão de Inundações com Dados Meteorológicos e Históricos**” visa desenvolver um modelo preditivo baseado em *Machine Learning* capaz de antecipar a ocorrência de inundações em regiões vulneráveis de Angola. Esta iniciativa está alinhada com os **Objetivos de Desenvolvimento Sustentável (ODS)**, nomeadamente o **ODS 11 – Cidades e Comunidades Sustentáveis** e o **ODS 13 – Ação Climática**, ao promover a resiliência urbana e a mitigação de desastres naturais. O problema central a ser abordado é a ausência de sistemas automatizados e confiáveis de previsão de inundações no país, o que leva a perdas humanas e materiais significativas. A solução proposta permitirá melhorar o planeamento urbano, reduzir danos ambientais e proteger comunidades vulneráveis.

2. Objetivos

O objetivo principal é criar um sistema inteligente que utilize dados meteorológicos e históricos para prever a probabilidade de inundações.

Os objetivos específicos incluem:

- Coletar e integrar dados climáticos (chuva, temperatura, umidade, pressão atmosférica) e geográficos (altitude, proximidade de rios e zonas costeiras).
- Treinar e avaliar modelos de *Machine Learning* capazes de identificar padrões associados à ocorrência de inundações.
- Desenvolver uma interface simples que apresente alertas e previsões em tempo real.

O projeto visa contribuir para a redução de riscos ambientais, o fortalecimento da infraestrutura resiliente e a proteção de vidas humanas.

3. Histórico

Nos últimos anos, várias províncias de Angola têm enfrentado episódios de inundações, sobretudo durante a época chuvosa, com destaque para Luanda, Benguela e Huambo. Esses eventos causam desalojamentos, prejuízos económicos e danos às infraestruturas públicas. Embora existam iniciativas globais de monitoramento climático — como o **Copernicus Climate Change Service** e o **NASA POWER Data Access Viewer** —, a aplicação dessas tecnologias em contexto angolano é limitada pela escassez de dados locais. Uma abordagem baseada em *Machine Learning* permite superar essas lacunas ao identificar relações complexas entre variáveis ambientais e eventos de inundação, fornecendo previsões mais precisas e contextualizadas para o território nacional.

4. Metodologia

A metodologia adotada baseia-se em técnicas de *Machine Learning supervisionado*. Serão utilizados algoritmos como **Random Forest**, **XGBoost** e **Regressão Logística**, conhecidos pela sua capacidade de lidar com dados multivariados e não lineares. As etapas incluem a coleta e limpeza dos dados, engenharia de atributos (extração de variáveis relevantes), divisão em conjuntos de treino e teste, treino do modelo, validação cruzada e avaliação do desempenho com métricas como *Acurácia*, *Precisão*, *Recall* e *F1-Score*. A implementação será feita em **Python**, utilizando bibliotecas como *Scikit-learn*, *Pandas*, *Matplotlib* e *NumPy*.

5. Diagrama de Projeto de Arquitetura

Descrição geral da arquitetura:

1. **Coleta de Dados:** Reúne dados meteorológicos e geográficos de fontes públicas (ex.: NASA, Copernicus, WorldClim).
2. **Pré-processamento:** Limpeza, normalização e tratamento de dados ausentes.
3. **Treino do Modelo:** Aplicação dos algoritmos de *Machine Learning* para aprendizagem e otimização de parâmetros.
4. **Avaliação:** Validação dos modelos com base em dados históricos de inundações.
5. **Interface e Visualização:** Exibição dos resultados e previsões em painéis interativos.

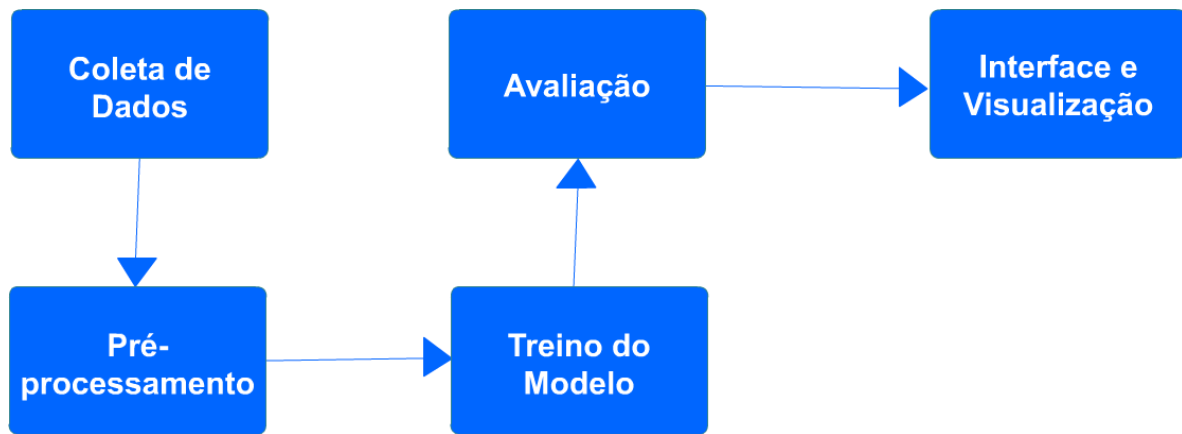


Fig1. O diagrama de fluxo entre esses cinco módulos com setas indicando o processamento sequencial.

6. Fontes de Dados

Os dados serão obtidos de bases meteorológicas e geográficas de acesso livre, incluindo o **NASA POWER**, o **WorldClim**, o **Copernicus Climate Data Store** e o **OpenStreetMap**. Esses conjuntos de dados fornecem variáveis climáticas históricas (temperatura, precipitação, umidade, pressão atmosférica) e informações geoespaciais (altitude, uso do solo, proximidade de rios). O pré-processamento envolverá a padronização dos formatos (CSV, GeoTIFF), interpolação de valores ausentes e normalização das variáveis numéricas.

7. Revisão de Literatura

Pesquisas recentes, como **Rahman et al. (2021)** – “*Flood Prediction Using Machine Learning Models in Bangladesh*” e **Ali et al. (2022)** – “*Machine Learning-Based Flood Risk Assessment in Data-Scarce Regions*”, demonstram que técnicas de *Machine Learning* podem alcançar alta precisão na previsão de inundações mesmo em contextos com dados limitados. O presente projeto baseia-se nessas abordagens, adaptando-as à realidade angolana, de modo a criar um sistema localmente relevante e tecnicamente robusto para mitigação de desastres naturais.

Plano de Implementação

1. Pilha de Tecnologia

O projeto utilizará uma combinação de ferramentas e linguagens voltadas para ciência de dados e machine learning:

- **Linguagem de Programação:** Python
- **Bibliotecas:**
 - *Pandas, NumPy* → tratamento e análise de dados
 - *Scikit-learn, XGBoost* → modelagem e previsão

- *Matplotlib, Seaborn* → visualização e análise exploratória
- *Flask* ou *Streamlit* → construção da interface de visualização
- **Ambiente de Desenvolvimento:** Google Colab / Jupyter Notebook e vscode
- **Armazenamento e Dados:** CSVs locais e APIs meteorológicas (NASA POWER, WorldClim)
- **Controle de Versão:** GitHub
- **Infraestrutura:** Computador pessoal / Google Cloud (para eventuais testes de implantação)

2. Linha do Tempo

Etapa	Descrição	Duração Estimada	Período
1. Coleta e Pré-processamento de Dados	Recolha, limpeza e normalização dos dados	2 dia	Dia 1-2
2. Análise Exploratória	Identificação de padrões e correlações	1 dia	Dia 3
3. Desenvolvimento do Modelo	Implementação de algoritmos ML (Random Forest, XGBoost)	1 dia	Dia 4
4. Treinamento e Avaliação	Treinar e validar os modelos com métricas adequadas	1 dia	Dia 5
5. Interface e Visualização	Criação de dashboard para previsão e alertas	3 dias	Dia 6-8
6. Implantação e Testes	Apresentação do protótipo funcional	1 dias	Dia 9

Distribuição de Tarefas

Membro	Responsabilidade Principal
Gerónimo Bêco	Desenvolvimento do modelo de Machine Learning e integração de dados
Iveth de Lima	Análise de dados
Laurinda Carlos	Pesquisa e documentação técnica
Pitchi Sidibe	Interface e visualização interativa

3. Marcos

1. ✓ Conclusão da proposta e plano de implementação do projecto.

4. Desafios e Mitigações

Desafio Potencial	Impacto	Estratégia de Mitigação
Escassez de dados locais precisos	Pode afetar a precisão do modelo	Utilizar fontes globais (NASA, Copernicus) e gerar dados sintéticos
Desequilíbrio de classes (poucos eventos de inundação)	Dificulta o treino do modelo	Aplicar técnicas de <i>oversampling</i> como SMOTE
Limitações computacionais	Pode atrasar o treino do modelo	Usar Google Colab (GPU gratuita)
Integração de APIs meteorológicas	Possível incompatibilidade de formato	Criar scripts de conversão automática e padronização

5. Considerações Éticas

O projeto respeita a **privacidade e segurança dos dados**, utilizando apenas informações públicas e agregadas, sem identificar indivíduos. Todas as previsões terão caráter informativo, não substituindo decisões oficiais de defesa civil.

Para evitar **viés algorítmico**, serão aplicadas técnicas de validação cruzada e monitoramento de métricas equilibradas entre regiões. O impacto esperado é positivo, promovendo a segurança de comunidades vulneráveis e fortalecendo a resiliência urbana face às mudanças climáticas.

6. Referências

- Rahman, M. et al. (2021). *Flood Prediction Using Machine Learning Models in Bangladesh*.
- Ali, S. et al. (2022). *Machine Learning-Based Flood Risk Assessment in Data-Scarce Regions*.
- NASA POWER Data Access Viewer – <https://power.larc.nasa.gov/>
- Copernicus Climate Data Store – <https://cds.climate.copernicus.eu/>
- WorldClim Global Climate Data – <https://www.worldclim.org/>