

FloodGuard: Sistema Inteligente de Monitoramento e Alerta de Enchentes

FIAP - Faculdade de Informática e Administração Paulista

Global Solution 2025.1 - Mitigação de Eventos Naturais Extremos

Integrantes:

- Bruno Castro - RM558359
- Hugo Mariano - RM560688
- Matheus Castro - RM559293

GitHub: https://github.com/Hugo345-l/global_solution_2_rm560688

Introdução e Justificativa

As enchentes representam o tipo de desastre natural que mais afeta pessoas globalmente, sendo responsáveis por aproximadamente 40% de todos os desastres registrados no Brasil. No contexto urbano brasileiro, especialmente em regiões metropolitanas como São Paulo, a combinação de eventos extremos de precipitação com urbanização inadequada e mudanças climáticas tem resultado em perdas humanas e econômicas significativas que afetam desproporcionalmente comunidades vulneráveis.

Atualmente, o Brasil conta com infraestrutura oficial robusta para monitoramento de desastres através do CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais), que mantém uma rede nacional de pluviômetros automáticos em 959 municípios. Contudo, identificamos uma lacuna crítica: os sistemas centralizados, embora eficazes para planejamento governamental e coordenação da Defesa Civil, não chegam diretamente às comunidades que mais necessitam de alertas imediatos e capacidade de ação autônoma.

O **FloodGuard** surge como proposta de **democratização** do monitoramento de enchentes, desenvolvendo um sistema distribuído que combina análise preditiva baseada em machine learning com sensoriamento IoT local. O diferencial conceitual está na **complementaridade** com a infraestrutura oficial existente, não na competição. Enquanto o CEMADEN oferece monitoramento nacional centralizado, o FloodGuard foca na proteção comunitária descentralizada, criando uma rede colaborativa de sensores que implementa lógica de cascata hidrológica para alertas preventivos.

A escolha por enchentes como foco se justifica pela **previsibilidade** destes eventos baseada em padrões pluviométricos identificáveis, pela **viabilidade técnica** de implementação com tecnologias IoT acessíveis, e pelo **impacto social** direto que soluções de monitoramento podem ter em comunidades vulneráveis urbanas.

Metodologia e Decisões Técnicas

Análise Estratégica: Complementaridade vs Competição

Aspecto	CEMADEN (Sistema Oficial)	FloodGuard (Nossa Proposta)
Escopo de Atuação	Nacional, centralizado	Comunitário, descentralizado
Investimento Inicial	Alto (infraestrutura gov.)	Baixo (~R\$ 50 por sensor)
Complexidade de Instalação	Técnica especializada	DIY (Do It Yourself)
Público-Alvo	Defesa Civil e órgãos oficiais	Comunidades vulneráveis
Processamento de Dados	Centralizado em data centers	Edge computing + servidor central
Comunicação	Hub-central apenas	Rede colaborativa sensor-to-sensor
Alertas	Via órgãos intermediários	Direto à população local

Base de Dados e Fundamentação Científica

Para desenvolvimento e treinamento do modelo preditivo, utilizamos dados oficiais da **rede de monitoramento do CEMADEN**, especificamente focando nos pluviômetros automáticos da região metropolitana de São Paulo. Esta escolha se justifica por múltiplos critérios técnicos:

- **Qualidade dos dados:** Pluviômetros automáticos com transmissão a cada 10 minutos
- **Densidade de instrumentação:** São Paulo possui alta concentração de equipamentos CEMADEN
- **Relevância histórica:** Município prioritário para gestão de riscos de desastres naturais
- **Disponibilidade temporal:** Séries históricas robustas (utilizamos dados de Janeiro a Maio 2025)

Decisões de Arquitetura e Modelagem

Agregação temporal: Convertemos dados de 10 minutos para janelas horárias, alinhando com padrões hidrológicos de formação de enchentes urbanas.

Classificação de risco: Implementamos limiares calibrados baseados na distribuição dos dados:

- **Baixo risco (0):** $< 5.5\text{mm/h}$
- **Moderado risco (1):** $5.5\text{mm} \leq \text{chuva} < 18\text{mm/h}$
- **Alto risco (2):** $\geq 18\text{mm/h}$

Seleção de features: Optamos por abordagem híbrida utilizando `acumulado_chuva_1_h_mm` (variável contínua de intensidade) e `cod_estacao` (variável categórica de localidade), permitindo ao modelo aprender padrões geográficos específicos.

Pipeline de pré-processamento:

python

```
preprocessor = ColumnTransformer(  
    transformers=[  
        ('onehot', OneHotEncoder(handle_unknown='ignore', sparse_output=False),  
         ['cod_estacao'])  
    ],  
    remainder='passthrough'  
)
```

Comparação de algoritmos: Testamos Logistic Regression versus Random Forest, com Random Forest demonstrando melhor capacidade de capturar interações não-lineares entre localidade e intensidade pluviométrica.

Stack Tecnológica Implementada

Servidor e ML: Python com scikit-learn, pandas para manipulação de dados, e joblib para persistência de modelos. Pipeline completo automatizado desde ingestão até predição.

Simulação IoT: ESP32 emulado em Python com sensores DHT22 para temperatura/umidade, sistema de alertas RGB (Verde/Amarelo/Vermelho) e feedback sonoro graduado por buzzer piezoelétrico.

Comunicação: Protocolo HTTP/JSON para transmissão entre sensores simulados e servidor central, com capacidade de comunicação inter-sensores para implementação futura de lógica de cascata hidrológica.

Desenvolvimento do MVP

Pipeline de Processamento de Dados

O script `1_process_official_data.py` implementa o processamento automatizado dos dados mensais do CEMADEN:

python

```
def processar_dados_cemaden_oficiais():  
    # Unificação de múltiplos CSVs mensais  
    df_completo = pd.concat(lista_dfs_mensais, ignore_index=True)  
  
    # Conversão temporal UTC → Brasília  
    df_completo['datahora_brasilia'] = df_completo['datahora_utc'] - pd.Timedelta(hours=3)  
  
    # Agregação horária  
    df_horario = df_completo.groupby(['cod_estacao', 'hora_utc_agrupada']).agg(  
        acumulado_chuva_1_h_mm=('chuva_10min_mm', 'sum')  
    ).reset_index()  
  
    # Classificação de risco baseada em limiares calibrados  
    def classificar_risco(chuva_h):  
        if chuva_h < 5.5: return 0 # Baixo  
        elif chuva_h < 18: return 1 # Moderado  
        else: return 2 # Alto
```

Modelagem e Treinamento

O script `2_train_model.py` implementa o pipeline completo de machine learning:

python

```
# Pipeline integrado com pré-processamento  
pipeline_rf_clf = Pipeline(steps=[  
    ('preprocessor', preprocessor),  
    ('classifier', RandomForestClassifier(random_state=42, n_estimators=100))  
)  
  
# Treinamento com todos os dados disponíveis  
pipeline_rf_clf.fit(X, y)  
  
# Persistência do modelo treinado  
joblib.dump(pipeline_rf_clf, 'ml_model/cemaden_flood_risk_model_pipeline.joblib')
```

Simulação IoT e Integração Híbrida

O script `3_run_simulation_with_local_sensor.py` demonstra a arquitetura híbrida proposta:

python

```
def determinar_risco_final(risco_ml, intensidade_local, mm_h_local):  
    """Combina predição regional (ML) com dados locais (sensores)"""  
    ajuste = 0  
    if intensidade_local == "Forte": ajuste = 1  
    elif intensidade_local == "Extrema": ajuste = 2  
    return min(risco_ml + ajuste, 2)  # Máximo nível 2  
  
def main():  
    modelo = carregar_modelo("ml_model/cemaden_flood_risk_model_pipeline.joblib")  
    for ciclo in range(SIM_DURATION):  
        # Simula dados locais do sensor  
        intensidade_local, mm_h_local = simular_evento_chuva()  
        # Predição regional via ML  
        risco_ml = prever_risco_ml(modelo, mm_h_local, SENSOR_COD_ESTACAO)  
        # Decisão final híbrida  
        risco_final = determinar_risco_final(risco_ml, intensidade_local, mm_h_local)  
        exibir_alerta(risco_final)
```

Funcionalidades Implementadas no MVP

- ✅ **Processamento automatizado:** Ingestão de 5 meses de dados CEMADEN com tratamento de múltiplos formatos CSV e padronização temporal
 - ✅ **Modelo ML treinado:** Random Forest com validação estratificada, capaz de processar dados de localidade via OneHotEncoder
 - ✅ **Simulação ESP32 completa:** Sistema de alertas visuais (RGB) e sonoros (buzzer) com graduação proporcional ao risco
 - ✅ **Arquitetura híbrida validada:** Integração funcional entre predições regionais (modelo ML) e refinamento local (sensores simulados)
 - ✅ **Pipeline end-to-end:** Fluxo completo desde dados brutos até alertas finais, com persistência de modelo e capacidade de re-treinamento
 - ✅ **Testes automatizados:** Validação de cenários extremos e verificação de consistência das predições
-

Resultados e Conclusões

Eficácia Demonstrada

O MVP do FloodGuard demonstra **viabilidade técnica completa** da arquitetura proposta. O modelo Random Forest treinado com dados do CEMADEN apresenta capacidade consistente de distinguir níveis

de risco baseado tanto em intensidade pluviométrica quanto em características geográficas das estações. A simulação IoT valida a lógica de integração híbrida, onde previsões regionais são contextualizadas por dados locais para gerar alertas mais precisos.

Validação conceitual: A abordagem de complementaridade com sistemas oficiais se mostra tecnicamente sólida. O FloodGuard não substitui a infraestrutura CEMADEN, mas democratiza o acesso a alertas preventivos, criando uma camada de proteção comunitária que reduz o tempo entre detecção de risco e ação preventiva local.

Limitações Reconhecidas

Hardware: Implementação em simulação Python, não em ESP32 físico com sensores reais **Escopo**

geográfico: Dados limitados à região metropolitana de São Paulo

Período temporal: Treinamento baseado em 5 meses (Janeiro-Maio 2025) **Comunicação inter-sensores:**

Lógica conceitual não implementada fisicamente **Interface de usuário:** Sistema básico via terminal, sem dashboard web/mobile

Impacto e Contribuição

O FloodGuard representa uma **proposta inovadora de democratização** de tecnologias de monitoramento de desastres. A combinação de dados científicos oficiais (CEMADEN) com sensoriamento comunitário acessível (ESP32) cria um modelo híbrido que:

1. **Complementa** a infraestrutura governamental sem duplicação de esforços
2. **Empodera** comunidades vulneráveis com capacidade autônoma de detecção e resposta
3. **Reduz significativamente** o tempo entre identificação de risco e ação preventiva
4. **Escalabiliza** para múltiplas regiões com investimento baixo (~R\$ 50 por sensor)
5. **Democratiza** o acesso a tecnologias até então restritas a órgãos governamentais

Trabalhos Futuros e Escalabilidade

Implementação física: Desenvolvimento com ESP32 real, sensores DHT22, sistema de comunicação WiFi

e alertas RGB/sonoros **Expansão geográfica:** Treinamento com dados CEMADEN de outras regiões

brasileiras prioritárias **Comunicação mesh:** Rede inter-sensores para propagação de alertas em cascata

hidrológica **Interface avançada:** Aplicativo mobile e dashboard web para gestão comunitária **Validação**

em campo: Testes piloto em comunidades vulneráveis com métricas de eficácia

Relevância para Global Solution

O FloodGuard cumpre integralmente os objetivos da Global Solution 2025.1, demonstrando aplicação prática de inteligência artificial para mitigação de eventos naturais extremos. A solução integra de forma coerente conhecimentos em lógica computacional, programação Python, estruturação de dados e machine learning, criando um impacto social direto em comunidades que mais necessitam de proteção contra enchentes.

A proposta de **complementaridade** com sistemas oficiais, ao invés de competição, representa uma abordagem madura e viável que reconhece tanto as limitações quanto as oportunidades do cenário tecnológico atual. O MVP desenvolvido não é apenas um protótipo conceitual, mas um **sistema funcional** que valida a arquitetura proposta e estabelece base sólida para implementação real.

O FloodGuard demonstra como a democratização de tecnologias pode gerar impacto social significativo, criando soluções que empoderam comunidades vulneráveis enquanto respeitam e complementam a infraestrutura institucional existente.

Referências

- Dados da Rede de Monitoramento de Desastres Naturais do CEMADEN/MCTIC. Disponível em: <https://mapainterativo.cemaden.gov.br/>
- Agência Gov. "Cemaden registra recorde de alertas e mais de 1,6 mil ocorrências de desastre em 2024". Disponível em: <https://agenciagov.ebc.com.br/noticias/202501/cemaden-registra-recorde-de-alertas-e-mais-de-1-6-mil-ocorrencias-de-desastre-no-brasil-em-2024>