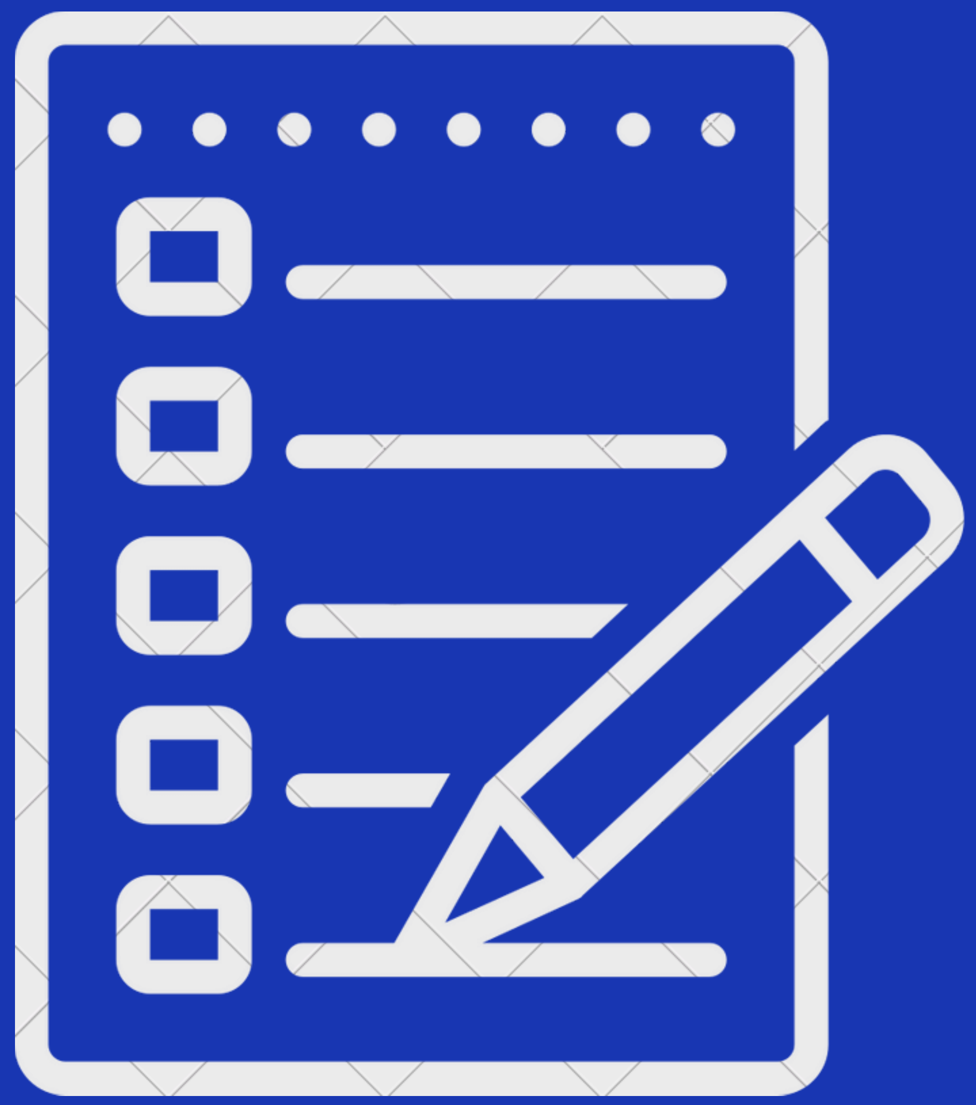
Infosys Intern Program

IMPACTSENSE - EARTHQUAKE IMPACT PREDICTION AND RISK VISUALIZATION

- Data Science Intern Team

Agenda



- Problem Statement
- Technical Architecture and System Design
- Data Documentation
- Model Development and Explainability
- Deployment & Maintenance
- Result Analysis
- Future Scope
- Conclusion
- Thank You

Problem Satement

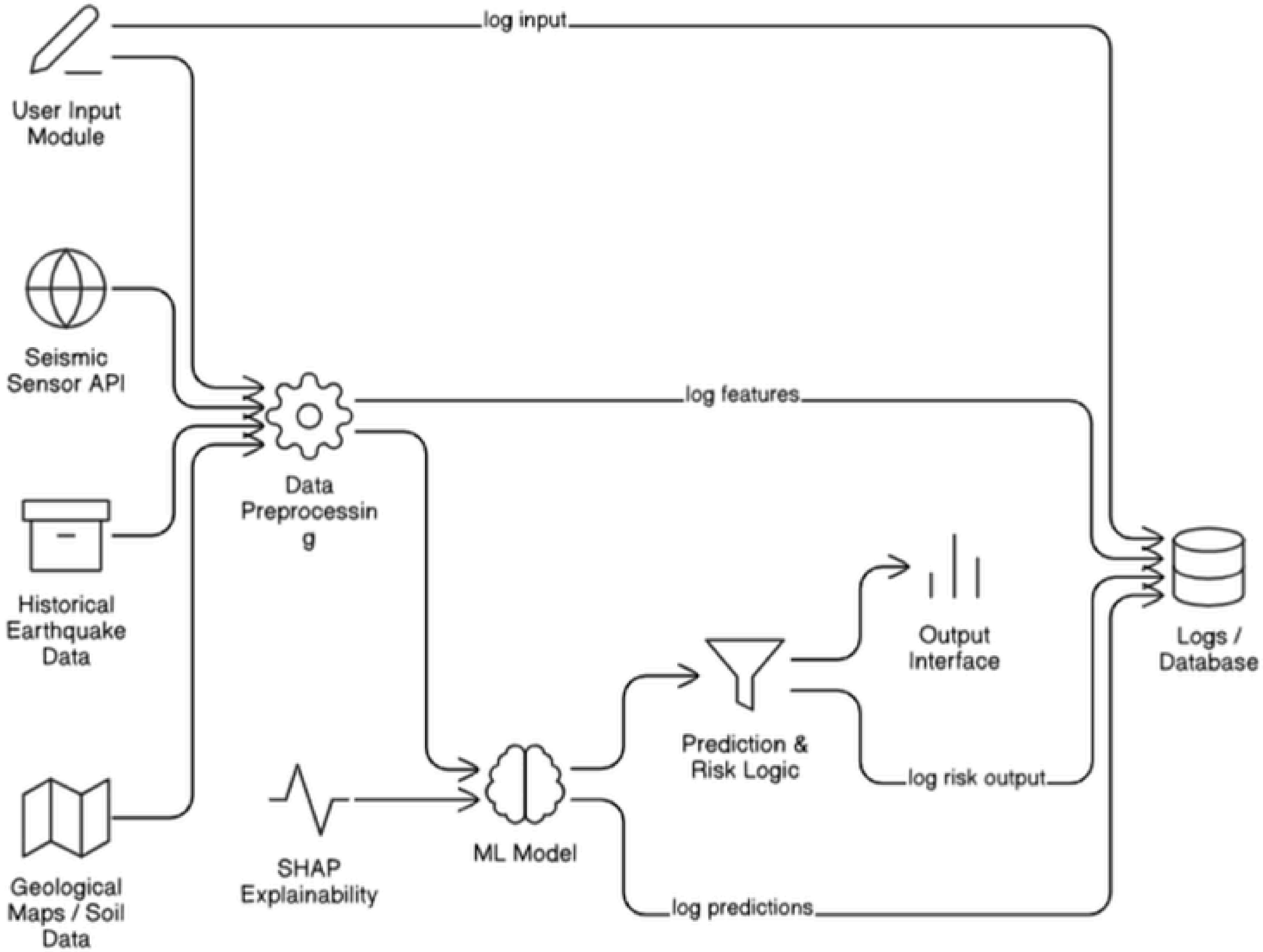
Problem Definition

ImpactSense leverages machine learning and geospatial visualization to enhance earthquake risk management. Its accurate LightGBM model (R^2 = 0.999) and intuitive Streamlit interface provide real-time, interpretable insights for informed decisions. Planned upgrades like real-time API integration, population calibration, and mobile support will further expand its scalability and effectiveness.

Business Context

- Stakeholders: Disaster agencies, city planners, utilities, and insurers.
- Use Cases: Quick risk assessment, resource planning, urban resilience.
- ROI: Better situational awareness and preparedness.
- Model: High accuracy (RMSE, R^2) with sub-second response time.
- Usability: Clear insights and an intuitive interface for all users.

Technical Architecture and System Design



Technical Architecture and System Design

System Overview

Data Ingestion

Upload CSV with earthquake events

Preprocessing

Pipeline for missing value imputation, feature engineering, encoding

Model Inference

LightGBM for Damage Potential prediction

Visualization

Statistical views, SHAP explainability, risk map

Tech Stack

FrontEnd

BackEnd

Explainability

Data Processing

Data Visualization

Data Documentation

Dataset Overview

- Source: Global seismic catalogs (USGS / ISC-GEM)
- Records: 23,412 total → 23,229 earthquakes → 18,583 for modeling
- Core Features: Latitude, Longitude, Depth, Magnitude, RMS, Type, Status

Preprocessing Pipeline

- .Selected 8 key features including RMS
- Imputed RMS using Random Forest (Lat, Lon, Depth, Magnitude)
- Filtered non-earthquake events
- Engineered Feature:
- Damage_Potential = 0.6×Magnitude + 0.2×((700−Depth)/700)×10
- Encoded categorical fields (Magnitude Type, Status)

Model Development & Explainability

Problem Formulation & Model Training

Problem Formulation

- Task: A regression model predicts a continuous score indicating earthquake severity for better decision-making.
- Target: Damage_Potential — a composite score based on magnitude and depth, estimating risk to people, infrastructure, and the environment.
- Features: 16 inputs — 5 numeric (e.g., Latitude, Longitude, Depth, Magnitude, RMS) and 11 categorical (e.g., Magnitude Type, Status) encoded to represent quantitative and qualitative factors.

Model Training

- Algorithm: The model uses LightGBM Regressor, a gradient boosting framework, with 600 trees and a learning rate of 0.05. LightGBM is chosen for its fast training, high accuracy, and ability to handle structured tabular data efficiently.
- Data Split: The dataset is divided into 80% training and 20% testing to ensure the model is trained effectively while preserving a set of unseen data for evaluation.

Deployment & Maintenance

Local Setup

The application can be deployed locally by installing dependencies with pip install -r requirements.txt and running the Streamlit app using streamlit run app.py for instant access.

CI/CD & Docker

An optional Dockerfile enables containerized deployment for consistent environments, while GitHub Actions automate CI/CD to validate model and pipeline updates.

Monitoring & Retraining

ImpactSense logs predictions for performance tracking, and periodic retraining keeps the model updated with new earthquake data for improved accuracy.

Result Analysis

Deploy

Upload earthquake CSV file

Drag and drop file here

Limit 200MB per file • CSV

Browse files

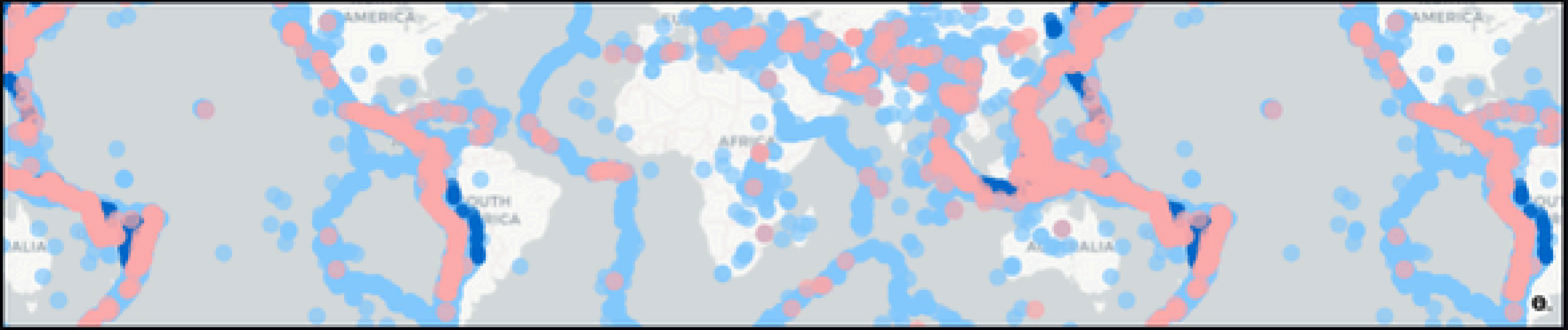
database.csv

2.3MB

Page

- Data
- Predict
- Map
- About

Risk Map



Risk_Label

- Moderate
- Low
- High

Sample Data

	Latitude	Longitude	Depth	Magnitude	Root Mean Square	Magnitude Type_MD	Magnitude Type_MH	Magnitude Type_ML	Magnitude Type_MS	Magnitude Type_MW	Magnitude Type_MWB	Magnitude Type_MWC	Magnitude Type_MWR	Magr
0	19.246	145.616	131.6	6	1.0284	0	0	0	0	1	0	0	0	
1	1.863	127.352	80	5.8	1.3056	0	0	0	0	1	0	0	0	
2	-20.579	-173.972	20	6.2	1.035	0	0	0	0	1	0	0	0	
3	-59.076	-23.557	15	5.8	1.0655	0	0	0	0	1	0	0	0	
4	11.938	126.427	15	5.8	1.0001	0	0	0	0	1	0	0	0	
5	-13.405	166.629	35	6.7	1.1787	0	0	0	0	1	0	0	0	

Future Scope

Realtime Data Integration

Incorporate USGS API to automatically ingest live earthquake data for up-to-date predictions.

Urban Risk Calibration

Enhance the model by including population-weighted metrics to better estimate human impact in urban areas.

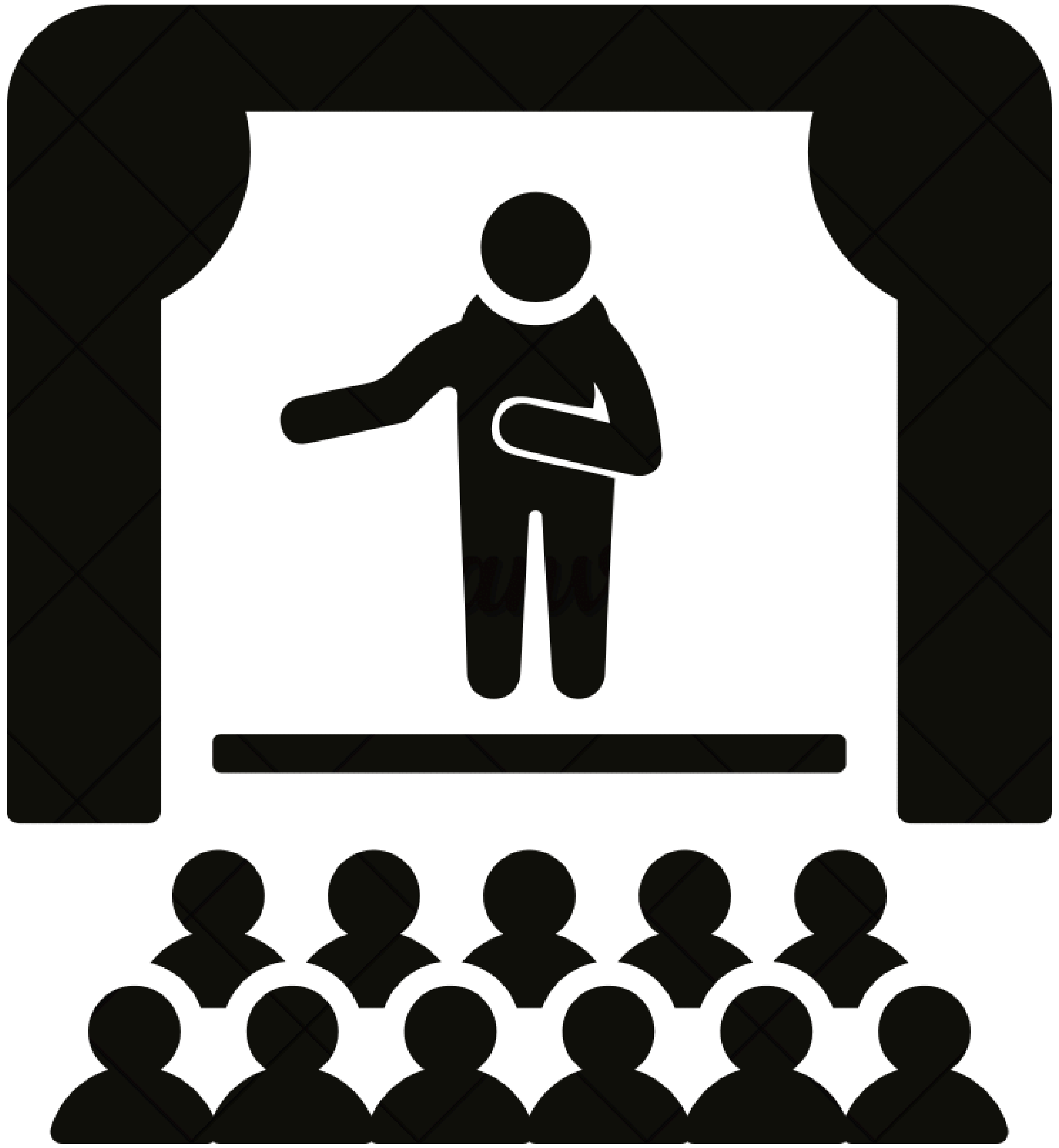
Model Optimization

Apply hyperparameter tuning with Optuna to further improve prediction accuracy and efficiency.

Mobile-Friendly Interface

Extend the platform with a responsive UI for easy access on smartphones and tablets.

Conclusion



- ImpactSense demonstrates how machine learning and geospatial visualization can revolutionize earthquake risk management. With its highly accurate LightGBM model ($R^2 = 0.999$) and user-friendly Streamlit interface, it offers real-time, interpretable insights for better decision-making. Future upgrades like real-time API integration, population calibration, and mobile support will enhance its effectiveness and scalability.

Thank You