frontier tech leaders

A new generation of tech specialists

Offline AI-Powered Community Support & Vulnerability Prediction System

Bridging the Information Gap in Conflict-Affected Myanmar

Team Name: **Team 2**

frontier tech leaders

Outline

- Concept note and implementation plan:
  - Background
  - Objectives
  - SDG Relation
- Data
  - Data Collection
  - Exploratory Data Analysis (EDA) and Feature Engineering
- Model Selection and Training
  - Model Evaluation and Hyperparameter Tuning
  - Model Refinement and Testing
- Results
- Deployment
- Future Work

Background

- **Brief project overview**: We developed an offline-first system that combines Machine Learning (to predict conflict vulnerability) with a Mobile Chatbot (to deliver life-saving aid information) for disconnected communities in Myanmar.

- **Provide brief background**: Myanmar faces severe internet shutdowns and escalating conflict. Displaced populations lack access to safety information, and aid organizations lack data on where to allocate resources safely.

- **Importance of the problem being solved**: Without internet, standard API-based apps fail. Our solution provides predictive safety intelligence and essential services without requiring connectivity.

Objectives

- Objective 1: To develop a Random Forest Regression Model that predicts monthly conflict fatalities (vulnerability) using multi-source data (ACLED, World Bank, NASA).
- Objective 2: To engineer an Offline Intelligence Pipeline that converts complex model predictions into a lightweight format (JSON) for mobile devices.
- Objective 3: To deploy a Flutter-based Chatbot Application that provides region-specific risk alerts and aid information (Health, Food, Education) entirely offline.

SDG Relation

- SDG 16 (Peace, Justice, and Strong Institutions): By predicting conflict hotspots, we enable proactive safety measures and violence reduction.
- SDG 3 (Good Health and Well-being): The chatbot provides offline medical triage and health guidance to areas without clinics.
- SDG 9 (Industry, Innovation, and Infrastructure): We bridge the "digital divide" by creating resilient infrastructure that functions without the internet.

Data

Data Collection and Preprocessing

Sources

- ACLED: Conflict event logs (2018–2024).
- World Bank/UNDP: Socio-economic indicators (GDP, HDI).
- NASA: Satellite Nighttime Lights (Economic activity proxy).

Preprocessing

- Aggregated granular event data into Monthly Time-Steps per Region.
- Merged geospatial datasets using "Admin1" (State/Region) codes.

Handling Missing Values

- Used Forward Filling for temporal socio-economic data (assuming stability over short periods).
- Normalized numerical features to a 0-1 scale for model stability.

Exploratory Data Analysis (EDA) and Feature Engineering

<u>EDA Insights</u>:

- Seasonality: Visualized conflict trends showing distinct spikes during dry seasons (operational windows).
- Hotspots: Identified Sagaing and Magway as the regions with the highest sustained conflict intensity (fatalities).
- Correlation: Found a strong inverse relationship between Nighttime Lights (economic activity) and conflict intensity in rural areas.

<u>Feature Engineering</u>:

- Lag Features: Created "lagged" variables (e.g., fatalities_last_month) because past violence is the strongest predictor of future violence.
- Harmonization: Aggregated weekly ACLED data and annual World Bank data into a unified Monthly-State timeframe to allow for regression analysis.

Model

Model Selection and Training

Model Selection:

- We compared Random Forest Regressor vs. 1D-CNN (Deep Learning).
- Choice: Random Forest was selected for its superior interpretability and ability to handle tabular socio-economic data without overfitting.

Training Strategy:

- Train Set: 2018–2023 (Historical context).
- Test Set: 2024–2025 (Unseen future data).
- Cross-Validation: TimeSeriesSplit (3 folds) to respect temporal order.

Model Refinement and Testing

Refinement Strategy:

- Hyperparameter Tuning: Used RandomizedSearchCV to optimize the Random Forest model. We tuned n_estimators (number of trees) and max_depth to prevent overfitting.
- Feature Selection: We tested "Enriched" datasets (adding detailed weather data) but found they added noise. We refined the model by sticking to the core features (Previous Fatalities + Economic Baseline) for higher stability..

Testing Protocol:

- Temporal Split: We strictly separated data by time (Train: 2018–2023, Test: 2024–2025) to simulate real-world forecasting.
- Metric: We focused on minimizing RMSE (Root Mean Squared Error) to penalize large, dangerous prediction errors.

Result

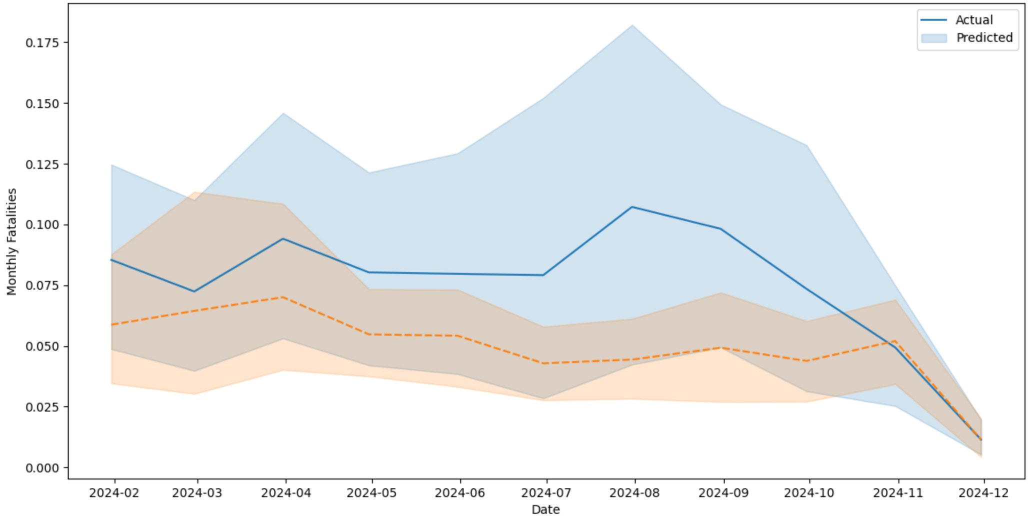
Evaluation Results

Quantitative Results:

- RMSE (Root Mean Squared Error): 0.0870
- MAE (Mean Absolute Error): 0.0398
- Key Insight: The Test Score (0.0870) outperformed the Validation Score (0.0884), proving the model generalized well and did not overfit.

Visualization:

Figure 1: Model Evaluation on Test Data (Actual vs. Predicted Fatalities)



S D G
A I
L A B

UN TECHNOLOGY
UNITED NATIONS TECHNOLOGY BANK FOR LEAST DEVELOPED COUNTRIES

icpsd
ISTANBUL INTERNATIONAL CENTER FOR PRIVATE SECTOR IN DEVELOPMENT

UN
DP

frontier
tech
leaders

Deployment

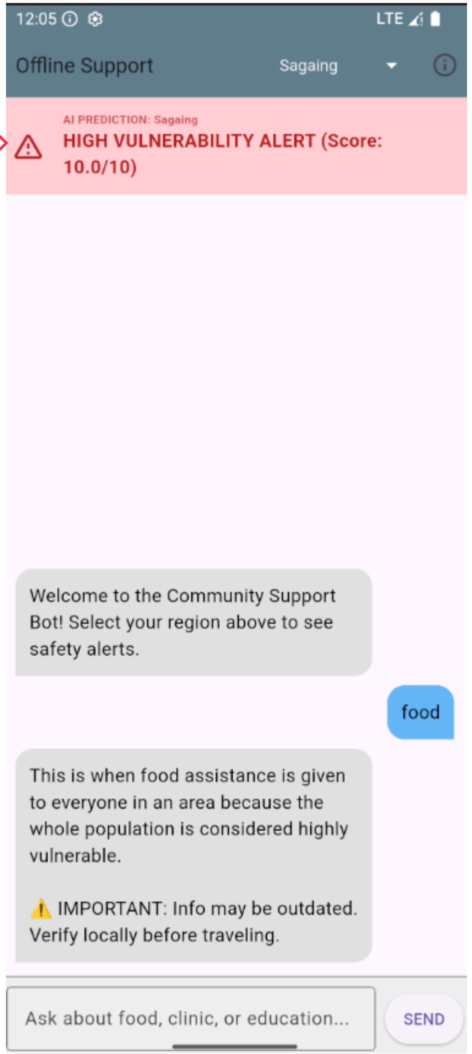
frontier tech leaders

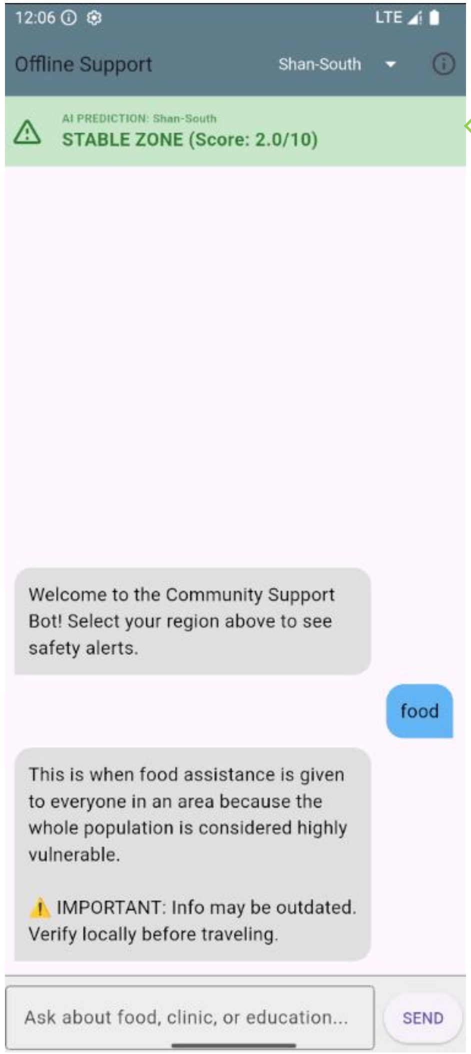
Deployment

Overview: The "Offline Intelligence Pipeline."

Process:

- Python Model: Generates predictions for all 15 regions.
- Serialization: predictions are converted to a lightweight regional_risk_scores.json.
- Flutter App: Loads the JSON asset on startup to drive the UI.





Conclusion and Futurework

We successfully built an end-to-end system that translates complex data science into actionable, offline-accessible safety tools for vulnerable populations.

Future Work:

- GPS Integration: Automatic region detection.
- Peer-to-Peer Sharing: Allow users to share the updated JSON risk files via Bluetooth/Zapya.
- Voice Support: Adding Speech-to-Text for users with low literacy.

Thank you!

frontier tech leaders