# Predictive Modeling of Conflict Events in Africa using ACLED Data

Ian Mbugua

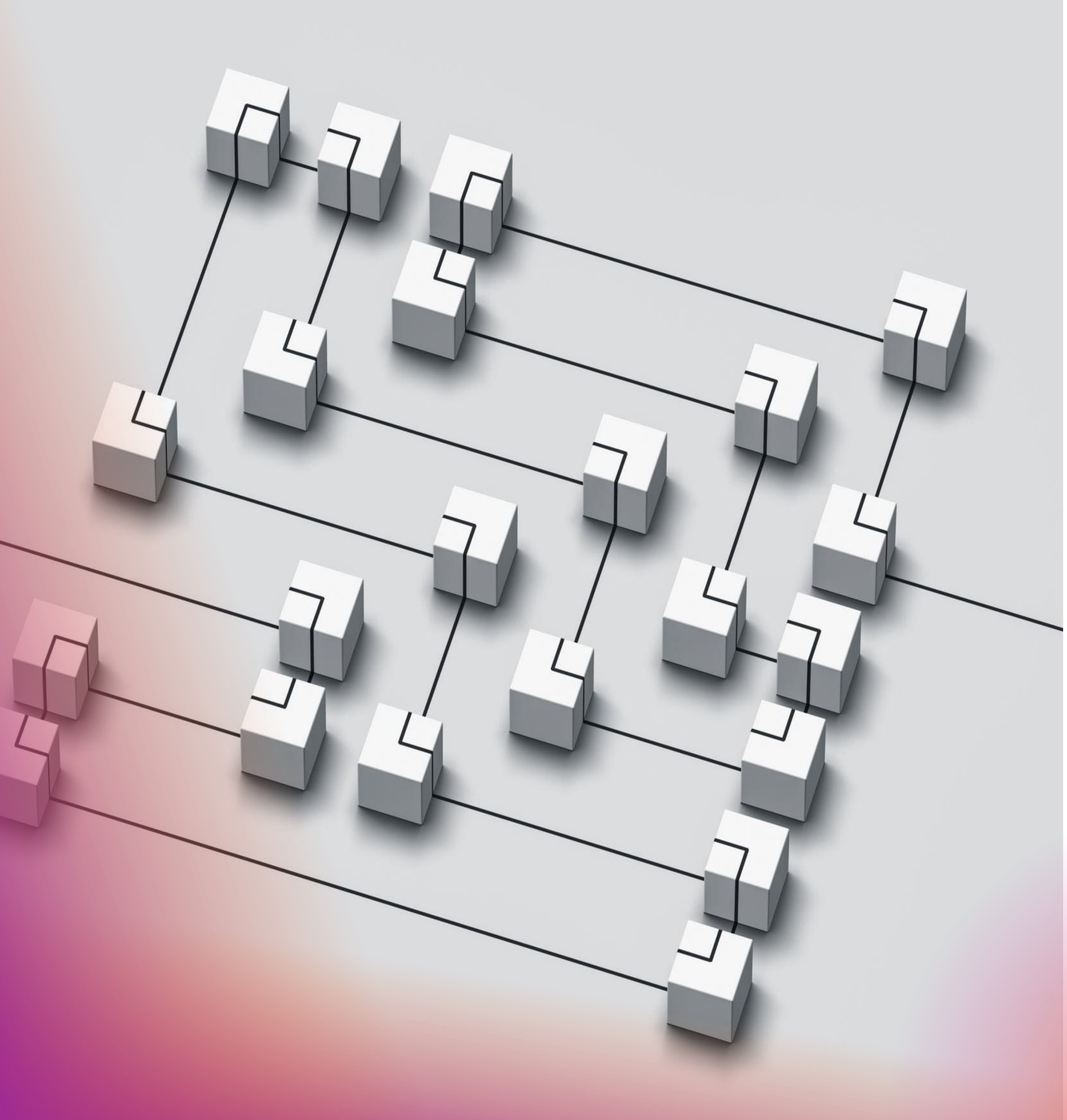
## • Introduction & Background


### Why Predict Conflict?

•Africa's complex conflict landscape.
•Importance of early warning for:

- Humanitarian aid planning
- Peacekeeping operations
- Policy intervention
- Resource allocation

•Goal: To leverage machine learning and ACLED data to develop a monthly conflict prediction model at the Admin1 level for Africa.



# Project Objectives

- Key Objectives
  - Analyze historical conflict trends in Africa (2012-2023) using ACLED data.
  - Engineer a comprehensive set of temporal and spatial features.
  - Develop and train machine learning models (XGBoost primary, baselines for comparison) to predict conflict likelihood one to three months ahead.
  - Evaluate model performance using appropriate metrics.
  - Identify key drivers and predictors of conflict.

# The ACLED Dataset

- **A**rmed **C**onflict **L**ocation & **E**vent **D**ata Project.
- Disaggregated event data: dates, actors, locations, fatalities, event types.
- Global coverage, real-time updates.
- **Scope for this project:** Africa, Jan 2012 - Dec 2023.
- Unit of Analysis: Country-Admin1-Month.

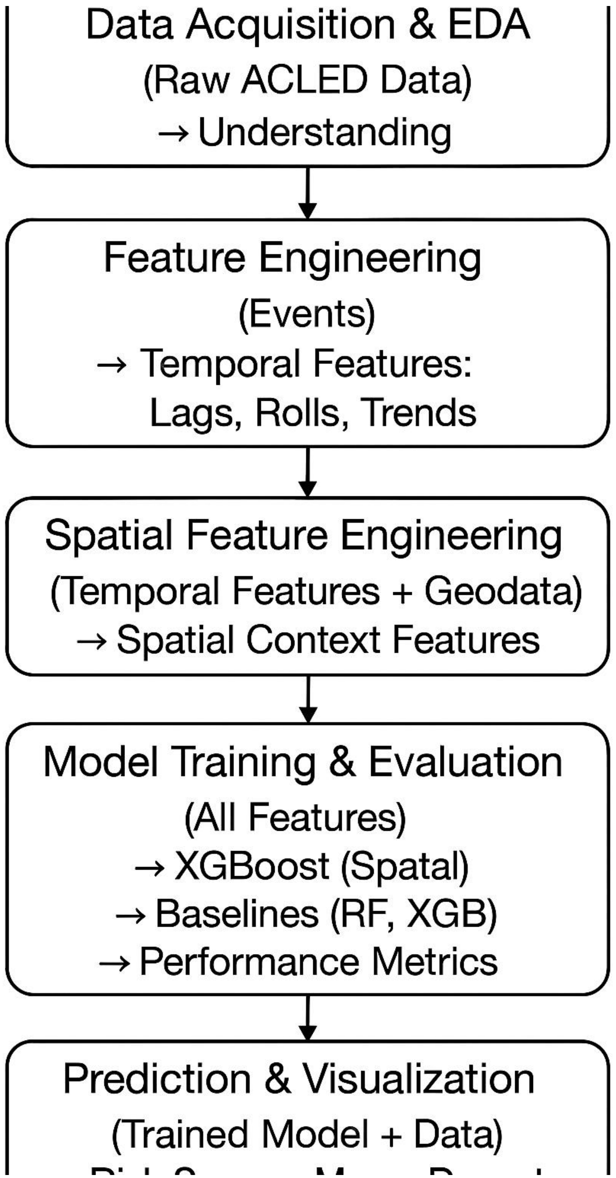
AF10

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | event_id_c | event_date | year | time_preci | disorder_type | event_type | sub_event | actor1 | assoc_act | inter1 | actor2 | assoc_act | inter2 | interactio | civilian_ta |
| 2 | SUD3517 | 16-May-25 | 2025 | 2 | Strategic deve | Strategic c | Arrests | Military Fo | Military Fo | State force | Civilians ( | Civilians (I | Civilians | State forces-Civilians | |
| 3 | ALG1305 | 16-May-25 | 2025 | 1 | Political violer | Battles | Armed cla | Military Forces of Alge | | State force | Unidentified Armed G | | Political m | State forces-Political | |
| 4 | BFO1381 | 16-May-25 | 2025 | 1 | Political violer | Violence a | Attack | JNIM: Group for Supp | | Rebel grou | Civilians (Burkina Fas | | Civilians | Rebel grou | Civilian tar |
| 5 | BFO1382 | 16-May-25 | 2025 | 1 | Strategic deve | Strategic c | Looting/p | JNIM: Group for Supp | | Rebel grou | Civilians (E | Labor Gro | Civilians | Rebel group-Civilians | |
| 6 | BFO1382 | 16-May-25 | 2025 | 1 | Political violer | Violence a | Attack | JNIM: Group for Supp | | Rebel grou | Civilians (Burkina Fas | | Civilians | Rebel grou | Civilian tar |
| 7 | BFO1382 | 16-May-25 | 2025 | 1 | Political violer | Violence a | Attack | JNIM: Group for Supp | | Rebel grou | Civilians (Burkina Fas | | Civilians | Rebel grou | Civilian tar |
| 8 | CAO1688 | 16-May-25 | 2025 | 1 | Political violer | Violence a | Attack | Boko Haram - Jamaat | | Rebel grou | Civilians ( | Kourgui C | Civilians | Rebel grou | Civilian tar |
| 9 | CDI3264 | 16-May-25 | 2025 | 1 | Demonstratio | Protests | Peaceful p | Protesters | PDCI: Dem | Protesters | | | | Protesters only | |
| 10 | CHA2077 | 16-May-25 | 2025 | 1 | Strategic deve | Strategic c | Arrests | Military Forces of Cha | | State force | Civilians ( | Transform | Civilians | State forces-Civilians | |
| 11 | CHA2079 | 16-May-25 | 2025 | 1 | Demonstratio | Protests | Peaceful p | Protesters | Muslim Gr | Protesters | | | | Protesters only | |
| 12 | CHA2080 | 16-May-25 | 2025 | 2 | Strategic deve | Strategic c | Arrests | Military Forces of Cha | | State force | Civilians (Chad) | | Civilians | State forces-Civilians | |

| | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | iso | region | country | admin1 | admin2 | admin3 | location | latitude | longitude | geo_precis | source | source_sc | notes | fatalities | tags | timestam |
| 2 | 729 | Northern A | Sudan | Red Sea | Al Ganab | | Arb'at | 19.7494 | 37.0956 | 1 | Al Rakoba | National | Around 16 | 0 | | 1.7E+09 |
| 3 | 12 | Northern A | Algeria | Khenchela | Chechar | | Chechar | 35.0381 | 7.0042 | 2 | Algeria Pre | National | On 16 May | 1 | | 1.7E+09 |
| 4 | 854 | Western A | Burkina Fa | Est | Tapoa | Diapaga | Diapaga | 12.0731 | 1.7884 | 1 | Al Zallaqa; | New medi | On 16 May | 0 | | 1.7E+09 |
| 5 | 854 | Western A | Burkina Fa | Centre-Es | Boulgou | Bane | Ouada-V1 | 11.4974 | -0.3536 | 1 | Whatsapp | New medi | Property d | 0 | | 1.7E+09 |
| 6 | 854 | Western A | Burkina Fa | Boucle du | Banwa | Solenzo | Kouna | 12.2017 | -4.1727 | 1 | Undisclos | Local parti | On 16 May | 0 | | 1.7E+09 |
| 7 | 854 | Western A | Burkina Fa | Boucle du | Banwa | Solenzo | Lekoro | 12.1943 | -4.2105 | 1 | Undisclos | Local parti | On 16 May | 0 | | 1.7E+09 |
| 8 | 120 | Middle Afr | Cameroon | Extreme-N | Mayo-Sava | Mora | Kourgui | 11.0881 | 14.1111 | 1 | Humanity | New medi | On 16 May | 2 | | 1.7E+09 |
| 9 | 384 | Western A | Ivory Coas | Lacs | Moronou | Bongouan | Bongouan | 6.6517 | -4.2039 | 1 | AIP (Ivory ( | National | On 16 May | 0 | crowd size | 1.7E+09 |
| 10 | 148 | Middle Afr | Chad | Ndjamena | Ndjamena | Ndjamena | Ndjamena | 12.1085 | 15.0482 | 1 | Alwihda (C | National-F | On 16 May | 0 | | 1.7E+09 |
| 11 | 148 | Middle Afr | Chad | Ouaddai | Ouara | Kachimel- | Abeche | 13.829 | 20.832 | 1 | Alwihda (C | National | On 16 May | 0 | crowd size | 1.7E+09 |
| 12 | 148 | Middle Afr | Chad | Logone Oc | Dodje | | Mandakao | 8.7595 | 15.3828 | 1 | Alwihda (C | National-I | Around 16 | 0 | | 1.7E+09 |

# The ACLED Dataset

# Project Pipeline



# Engineering Predictive Features

- **Goal:** Transform raw events into meaningful monthly signals for each Admin1 region.
- **Base Metrics Calculated Monthly:**
  - Event Counts (total, violent, specific types like battles, protests)
  - Fatalities (sum)
  - Actor Dynamics (distinct actor counts, distinct actor type counts)
  - Event Diversity (variety of event types, sub-types)
- **Temporal Features Derived from Base Metrics:**
  - **Lags:** (1, 2, 3, 6, 12 months) - Recent history.
  - **Rolling Statistics:** (3, 6, 12-month means & sums) - Sustained trends.
  - **Trend Features:** (Change over 1 & 3 months) - Rate of change.
  - days_since_last_violent_event.

# Target Variable & Dataset Balance

- Defining What We Predict
  - **Target (conflict_occurs):** Binary (1 = Conflict, 0 = No Conflict).
  - **Definition:** Conflict occurs if violent_events_count >= 1 in an Admin1 region, X months in the future (Project used X=3 months).
  - **Dataset Balance (Example from 3-month window run):**
    - No Conflict (0): ~73.3%
    - Conflict Occurs (1): ~26.7%

# Spatial Feature Engineering - The Geographic Context

- Rationale: Conflict in one region can influence its surroundings.
- **Primary Method (Used in final run with Africa_Countries.shp):**
  - Shapefile used to identify contiguous Admin1 neighbors.
  - Features: Avg. violent events, avg. fatalities, conflict density in neighboring regions (from t-1).
- **Fallback (If no shapefile):** Country-based features (activity in other Admin1s within the same country).

# Predictive Models & Validation

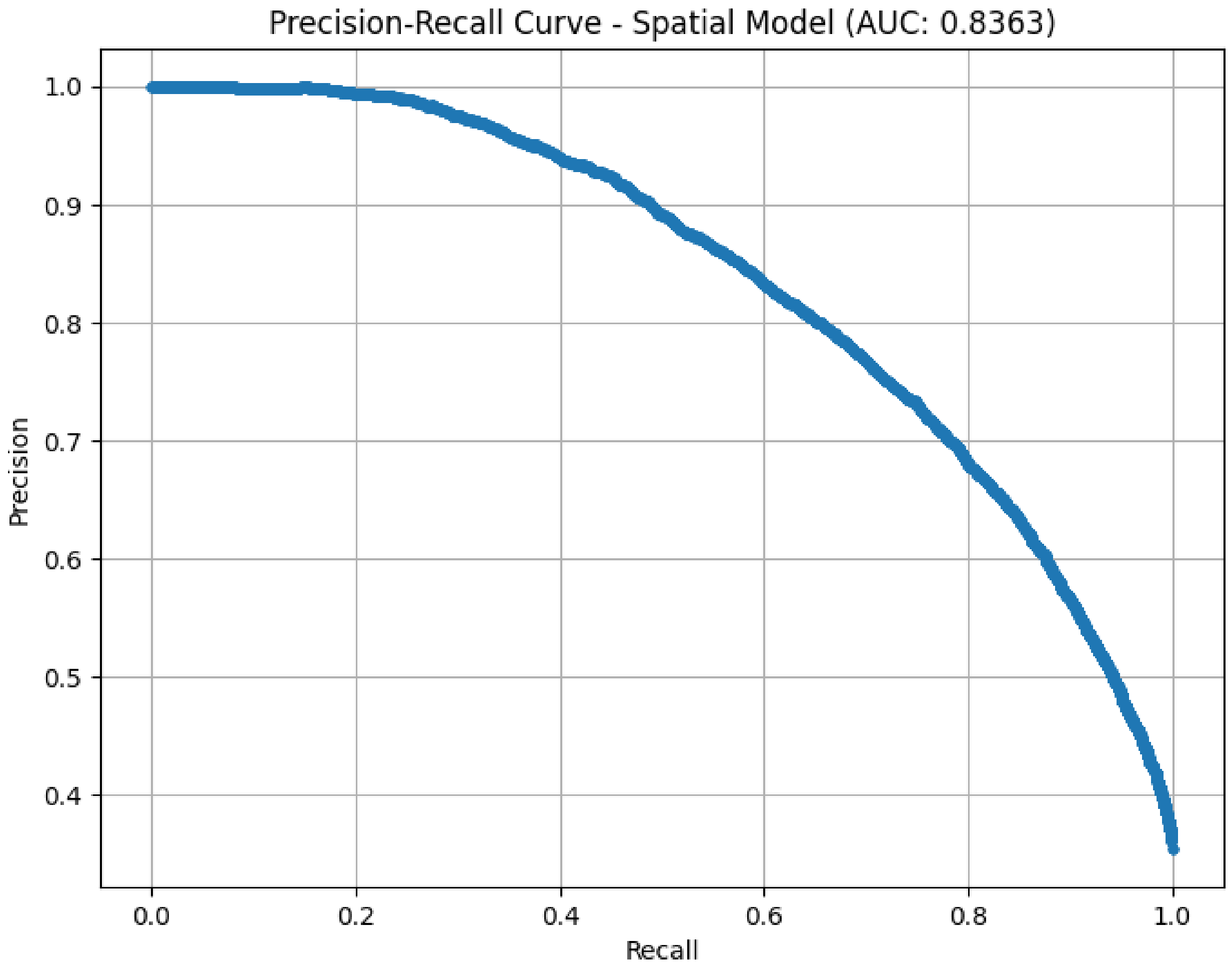
- **Primary Model:** XGBoost Classifier (with temporal + spatial features).
    - Why XGBoost? Performance, handles non-linearity, regularization.

- **Baseline Models:** (For comparison, using temporal features)
    - RandomForest Classifier
    - XGBoost Classifier (without explicit spatial spillover features)

- **Validation Strategy:** Temporal Train-Test Split (80% train, 20% test).
    - Ensures model is tested on data "future" to its training data.
    - Cutoff for main model: ~August 2020 (for 2012-2023 data, 3-month window).

# Evaluation Metrics

- **Accuracy:** Overall correctness.
- **Precision, Recall, F1-Score:** Focus on the "Conflict" class.
    - Precision: Of predicted conflicts, how many are real? (Minimizes false alarms)
    - Recall: Of actual conflicts, how many are caught? (Minimizes missed events)
- **Confusion Matrix:** Visualizes TP, TN, FP, FN.
- **ROC AUC:** Overall discrimination ability.
- **PR AUC:** Key for imbalanced data, focuses on positive class performance.

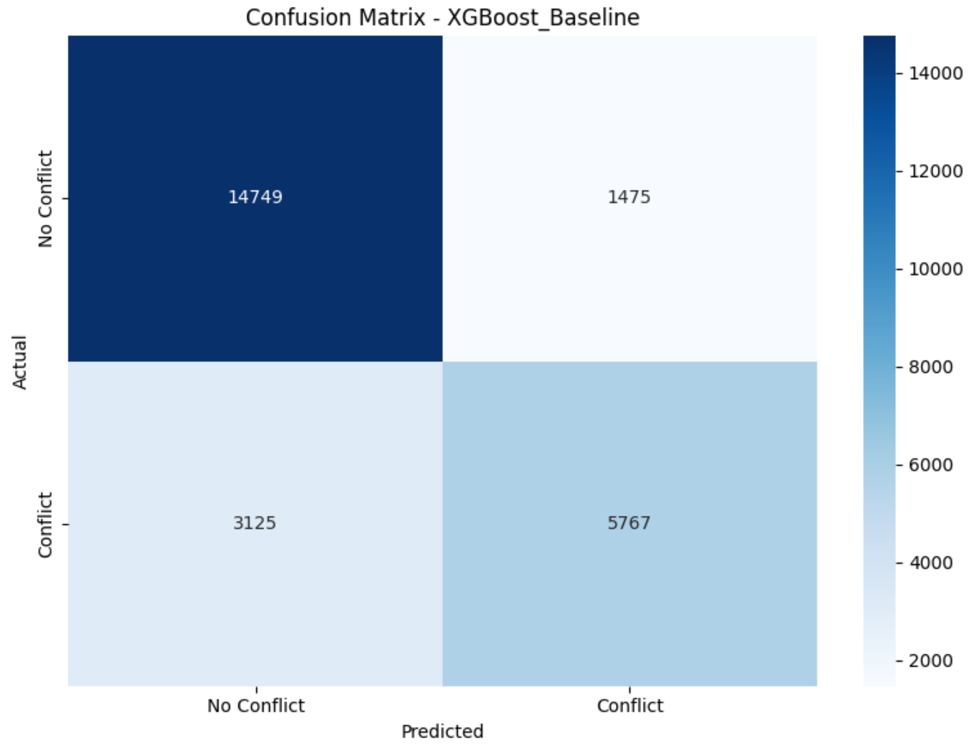


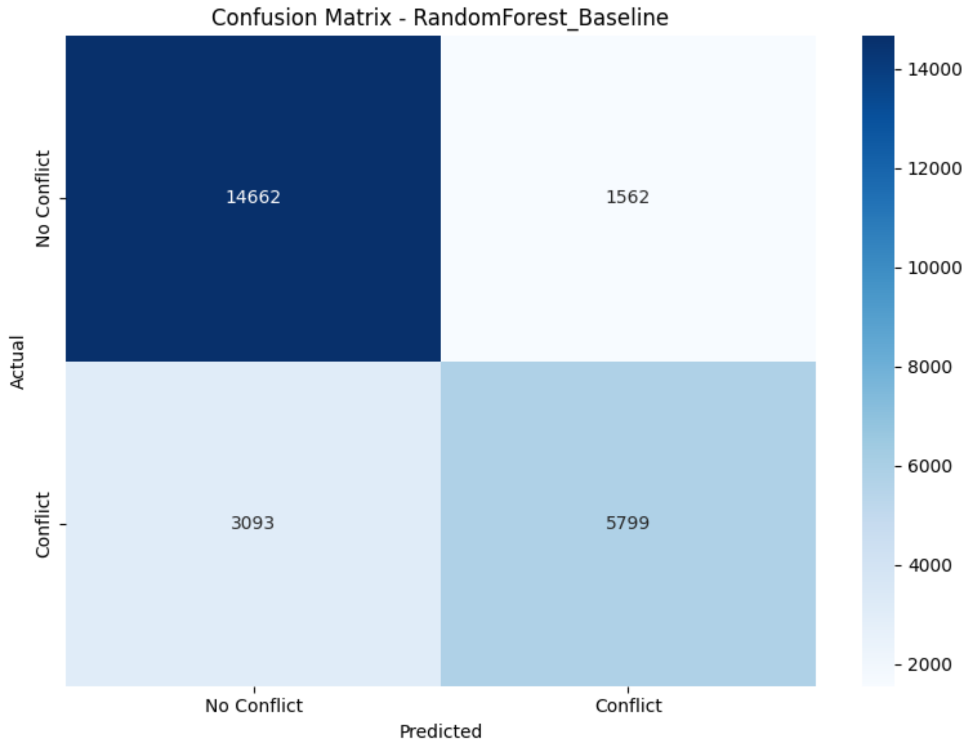
# Baseline Performance & Comparison

| Model | PR AUC | ROC AUC | Conflict Precision | Conflict Recall | Conflict F1-Score | Accuracy |
|---|---|---|---|---|---|---|
| XGBoost(Spatial Features) | 0.8363 | 0.8809 | ~0.79 | ~0.62 | ~0.70 | ~0.81 |
| XGBoost_Baseline | 0.8278 | 0.8722 | 0.80 | 0.65 | 0.71 | 0.82 |
| RandomForest_Baseline | 0.8265 | 0.8678 | 0.79 | 0.65 | 0.71 | 0.81 |

# Results - Spatial Model Performance

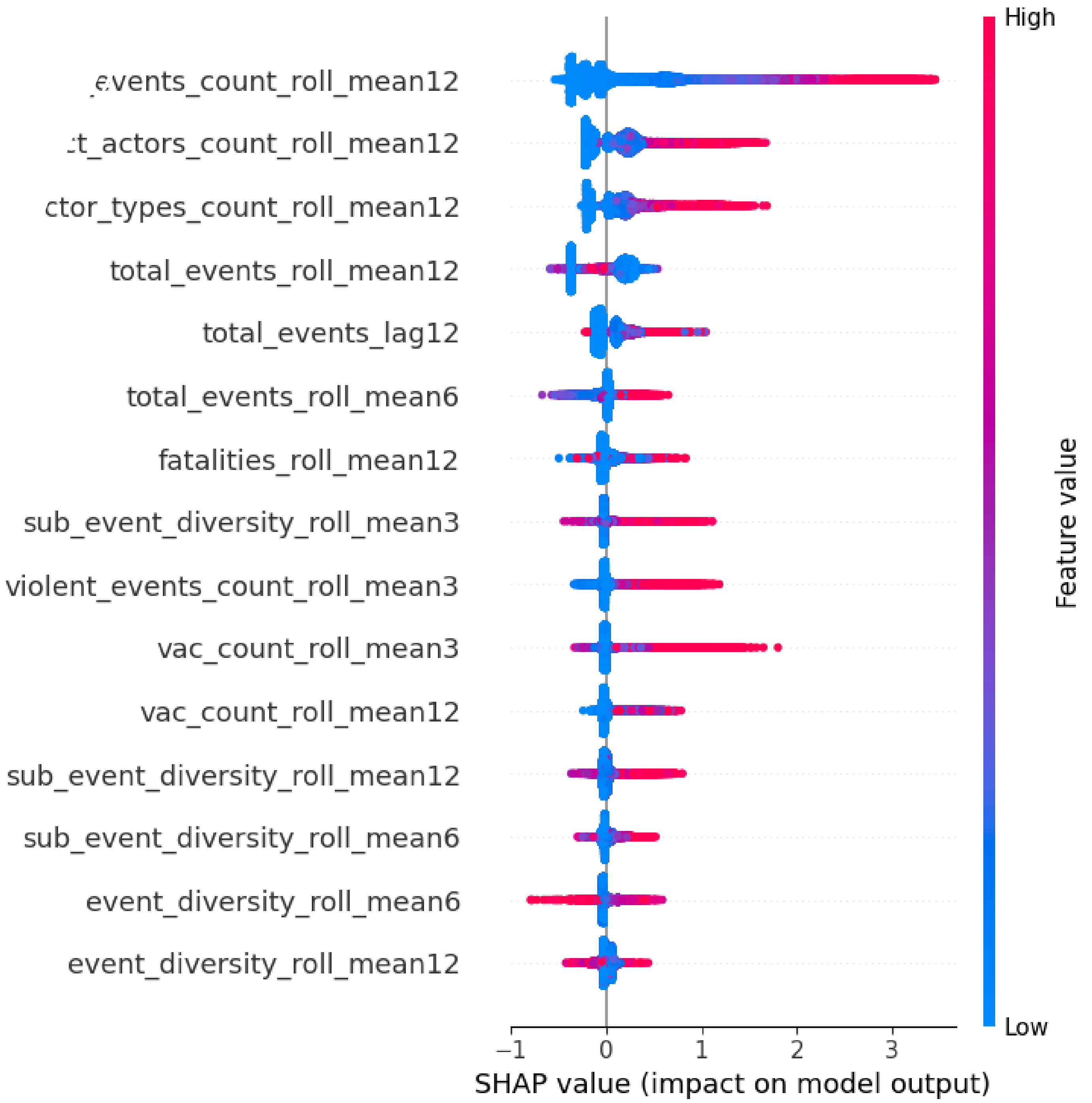
- **ROC AUC: 0.8809**
- **PR AUC: 0.8363**
- Brief interpretation: "Strong discriminative power, significantly better than random."



# Feature Importance - What Drives Predictions? (Baseline)

- Long-term rolling averages (especially 12-month) are dominant.
- Key predictors:
  - violent_events_count_roll_mean12 (High value -> High conflict risk)
  - distinct_actors_count_roll_mean12 (High value -> High conflict risk)
  - distinct_actor_types_count_roll_mean12 (High value -> High conflict risk)
- Shorter-term trends (e.g., 3-month violent events) also contribute.
- Higher historical violence, actor complexity, and event diversity consistently increase predicted risk.



# Predicted High-Risk Regions

- Middle Shabelle, Somalia (Prob: 0.9999, Very High)
- Est, Burkina Faso (Prob: 0.9999, Very High)
- Analamanga, Madagascar (Prob: 0.9998, Very High)
- Gao, Mali (Prob: 0.9998, Very High)
- Nord-Ouest, Cameroon (Prob: 0.9998, Very High)

# Discussion & Interpretation

- Machine learning models can effectively predict conflict likelihood with good accuracy using historical ACLED data.
- Sustained historical patterns (captured by long rolling windows) are more influential than isolated recent events for the baseline models.
- Incorporating spatial context (even simplified country-level effects or neighbor effects) provides a measurable uplift in predictive performance.
- The model identifies plausible high-risk areas.

# Challenges & Future Work

- **Challenges:**
  - ACLED data biases (reporting variations).
  - Defining "conflict" (sensitivity to event threshold).
  - Computational resources for SHAP / complex spatial features.
- **Future Work:**
  - Systematic hyperparameter tuning.
  - Refine spatial features (more advanced contiguity, distance decay).
  - Incorporate external data (socio-economic, governance, climate).
  - Focus on improving recall for the "conflict" class (e.g., scale_pos_weight, threshold tuning).
  - Deeper error analysis.

# Conclusion

- Successfully developed an end-to-end pipeline for conflict prediction.
- XGBoost model with spatial features achieved strong performance (PR AUC 0.8363, ROC AUC 0.8809).
- Demonstrated the value of both temporal feature engineering and spatial context.
- Provides a solid foundation for further research and potential application in early warning systems.