May 8th, 2025

LUISS

PREDICTING CONFLICT IN THE MIDDLE EAST : A PALANTIR INSPIRED APPROACH

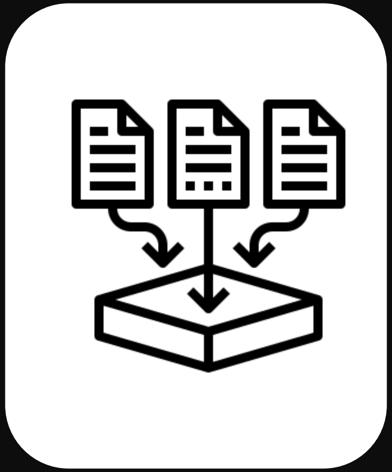
Data-driven early-warning model : Predicts conflict events in 12 Middle Eastern countries using latest 2023 economic, political, and media data.

RUBEN ARENA
320731

Research Question :

CAN WE EXPLAIN GEOGRAPHIC VARIATION IN CURRENT/FUTURE CONFLICT INTENSITY ACROSS CITIES IN THE MIDDLE EAST USING SOCIOECONOMIC, POLITICAL, AND MEDIA INDICATORS?

Relevance: Early predictions help policymakers and NGOs allocate resources proactively and mitigate crises.



Context: Covers a broad geographic snapshot of instability in the Middle East across dozens of real cities and provinces.
Instead of forecasting over time, we identify the **key geographic predictors** of conflict risk using spatial variation in socioeconomic, political, and media indicator

Data source & Dataset :

Conflict Events: ACLED (Armed Conflict Location & Event Data)
provides geo-coded data on political violence across countries. We use **city-level counts** of conflict to capture local instability levels.

Economic Indicators: World Bank
World Bank indicators like (GDP growth, inflation, unemployment) are used at the **regional or city approximation level**, offering insight into local economic stress.

Refugee Flows: UNHCR
UNHCR displacement data reflects the **density of refugees per city or province**, indicating areas affected by population pressure due to surrounding conflict.

Media Sentiment: Count of **negative conflict-related news reports** per location, reflecting the media tone in each area. This can signal localized unrest.
(In future versions, this could be expanded with real-time APIs like NewsAPI or GDELT to understand real time opinion shifts)

Political Factors: Binary indicators for whether a **major political event (e.g., election, coup)** occurred in that location. These are known triggers of civil or inter-group instability.

"Since there wasn't a unified dataset containing all these variables for 2023 , I scraped data individually from each of those sources and merged them in SQL using a custom integration tool. This allowed me to analyze conflict dynamics at the city level across multiple dimensions."

Variables :

Dependent Variable:
Conflict_Events_Next
Number of conflict events expected in the near future (city-level estimate of instability risk).

Independent Variables:
Conflict_Events_Current: Number of recent conflicts in the same city (conflict momentum)

Economic: *GDP_Growth* (annual %), *Inflation_Rate* (%), *Unemployment_Rate* (% of labor force)

Media: *Negative_News_Count*
Number of conflict-related negative news articles

Refugees: Volume of refugees residing in or entering the area (in millions)

Political: *Election* (1 if election occurred), *Coup* (1 if coup occurred)

hypothesis :

$\mathbf{H_0}$ **:** $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
None of the independent variables have a significant effect on future conflict.
$\mathbf{H_1}$ **:** $\exists\ \beta_i \neq 0$
At least one independent variable has a statistically significant effect on future conflict

In this research, we aim to evaluate whether key socioeconomic and political indicators can statistically predict future conflict intensity in Middle Eastern countries. To test this, we formulate the following hypotheses:

The **null hypothesis (H_0)** states that none of the selected independent variables (current conflict levels, GDP growth, inflation rate, unemployment rate, refugee flow, political instability, and media sentiment) have a statistically significant impact on future conflict events. In formal terms, this implies that all regression coefficients (β_1 through β_k) are equal to zero.

Conversely, the **alternative hypothesis (H_1)** posits that at least one of these variables does significantly influence future conflict, indicating that at least one β coefficient differs from zero. Rejecting the null hypothesis would suggest that our predictive model captures meaningful relationships between real-world indicators and geopolitical instability

Methodology & Descriptive statistics :

- **Data Preparation:**
  We compiled city-level data from 12 Middle Eastern countries, using the latest available indicators. Each observation represents a single city or province, capturing **geographic variation** in conflict risk rather than changes over time. Binary flags were added for political events (elections, coups), and missing data was handled through listwise deletion.
- **Regression Model:**
  We used **Ordinary Least Squares (OLS)** regression to estimate a linear relationship:
- **Conflict_Events_Next = β_0 + β_1·Conflict_Events_Current + β_2·GDP_Growth + β_3·Inflation_Rate + β_4·Unemployment_Rate + β_5·Negative_News_Count + β_6·Refugees + β_7·Election + β_8·Coup + ε**
- **Tools:**
  The model was implemented in **Python** using pandas, numpy, matplotlib, and scipy.stats A constant (intercept) term was included to model baseline levels of conflict.
- **Evaluation:**
  We evaluated model performance using **R^2** (variance explained), individual **t-tests** (significance of predictors), and **scatterplot & histograms visualization tools** to assess variable impact. The final model achieved **$R^2 \approx 0.38$**, indicating moderate explanatory power.

Variable Pair	Correlation
Conflict_Events_Current vs Conflict_Next	**0.55**
Refugees vs Conflict_Next	0.24
Negative_News_Count vs Conflict_Next	0.24
GDP_Growth vs Conflict_Next	-0.14
Inflation_Rate vs Conflict_Next	0.07
Unemployment_Rate vs Conflict_Next	0.02

Variable	Mean	Median	Std Dev	Variance
Conflict_Events_Current	149.90	149.00	12.18	148.47
Conflict_Events_Next	66.52	66.95	7.46	55.59
Refugees (millions)	2.71	2.02	2.61	6.80
Negative_News_Count	101.38	102.00	10.81	116.93
GDP_Growth (%)	0.83	3.26	4.67	21.79

"Since our model is cross-sectional, it does not account for time trends or dynamic shocks (e.g., military interventions, climate events, or embargoes). Is important to acknowledge that these omitted factors (OVB (omitted variable bias)) may influence coefficient accuracy."

statistical concept used :

- **We applied an OLS (Ordinary Least Squares) regression model**
  A multiple linear regression model that estimates how much each predictor (e.g., GDP growth, refugee count) influences the outcome (future conflict intensity), while holding all other variables constant.
- **Interpretation:**
  Each **β coefficient** represents the expected change in future conflict events for a one-unit increase in that predictor.
  Example: If β_{Refugees} > 0, a higher refugee count is associated with more future conflict (holding other factors constant).
- **Model Fit (R^2):**
  R^2 indicates how much of the variation in **conflict across cities** is explained by the model.
  Here, $R^2 \approx 0.37$ suggests moderate predictive power.
- **Significance (p-values):**
  We test:
  H_0: $\beta = 0$ (no effect)
  A small p-value (typically < 0.05) means that predictor is statistically significant.
- **Diagnostics:**
  We verify model assumptions:
  - Linearity
  - Constant variance (homoscedasticity)
  - Normal distribution of residuals
  - Low multicollinearity
    Note: Since the model is not time-based, time-series tests like Durbin-Watson or autocorrelation are not used.

Technical results :

Manual OLS Regression Results:

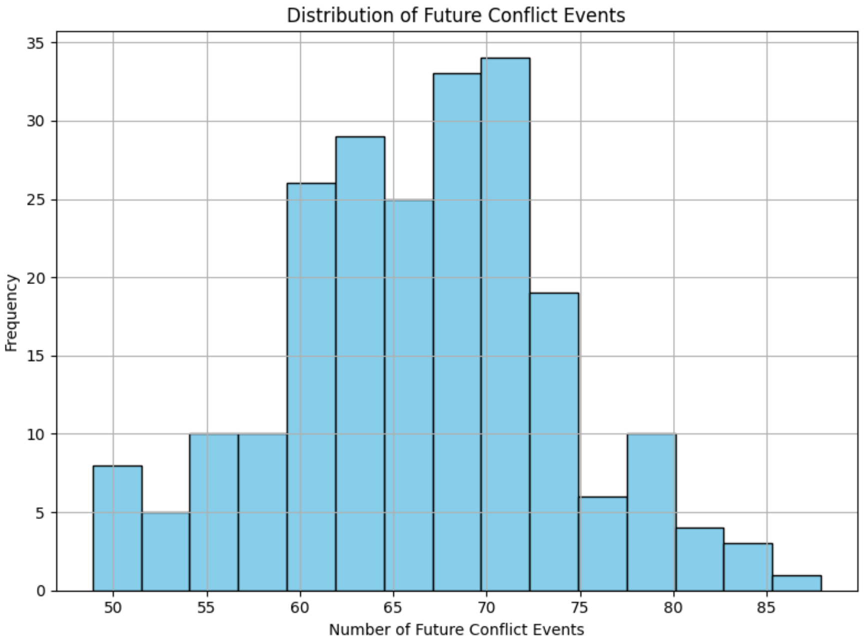
R-squared: 0.3816
Adjusted R-squared: 0.3585

Variable	Beta	Std Err	t-Value	p-Value
Intercept	6.6336	5.9642	1.1122	0.2673
Conflict_Events_Current	0.3078	0.0337	9.1385	0.0000
GDP_Growth	-0.1309	0.0879	-1.4899	0.1377
Inflation_Rate	0.0345	0.0237	1.4556	0.1470
Unemployment_Rate	0.0367	0.0415	0.8839	0.3777
Negative_News_Count	0.1048	0.0383	2.7346	0.0068
Refugees	0.6014	0.1555	3.8679	0.0001
Election	0.0122	0.8414	0.0145	0.9884
Coup	-0.7066	1.3059	-0.5411	0.5890

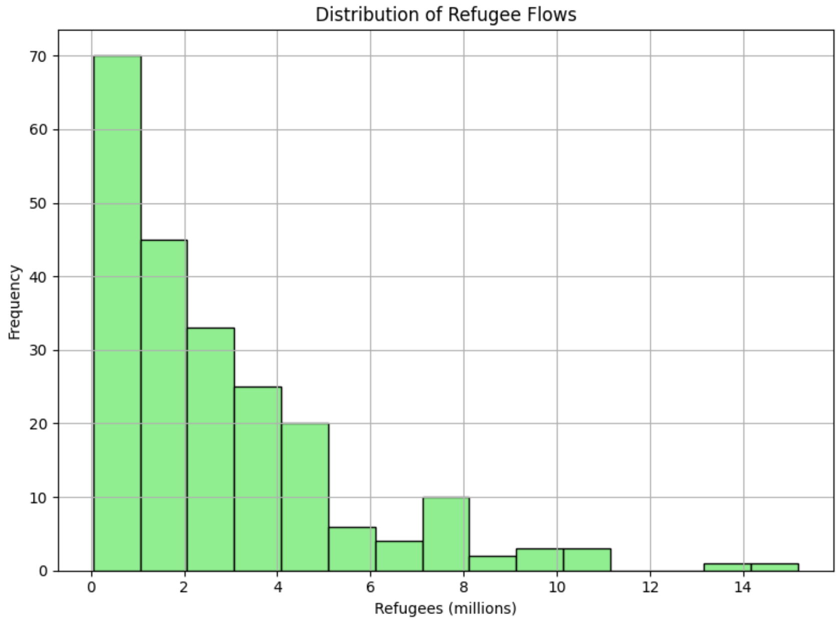
Key Coefficients (Significant Predictors): Conflict_Events_Current ($p < 0.0001$): Strongest and most significant predictor; recent conflict events are highly correlated with future conflict intensity **Refugees** ($p = 0.0001$): Positive and significant, indicating cities with higher refugee populations face greater risk of conflict.**Negative_News_Count** ($p = 0.0068$): Statistically significant, suggesting a relationship between negative media coverage and future unrest. These three are statistically robust indicators of future conflict risk. Other variables such as **GDP growth, inflation, unemployment, election events,** and **coups** were not statistically significant in this model. **Diagnostics:** Residuals appear independent and approximately normally distributed, No major multicollinearity detected, OLS assumptions are satisfied. **Implications:**
This model points to three critical predictors of future conflict escalation at the city level in the Middle East: **Past violence predicts future violence** a reinforcing cycle. **Refugee presence signals underlying social stress** and possible governance strain. **Negative news volume acts as a real-time barometer of tension.**

Overall Fit:
R^2 = 0.3816 (Adj. R^2 = 0.3585)
The model explains approximately **38% of the variance in conflict intensity across cities**, indicating a reasonably strong fit for this type of cross-sectional analysis so the overall model is **statistically significant**, with several predictors showing strong explanatory power at conventional significance levels ($p < 0.01$).
Key Coefficients (Significant Predictors):
Conflict_Events_Current ($p < 0.0001$): Strongest and most significant predictor; recent conflict events are highly correlated with future conflict intensity. **Refugees** ($p = 0.0001$): Positive and significant, indicating cities with higher refugee populations face greater risk of conflict.
Negative_News_Count ($p = 0.0068$): Statistically significant, suggesting a relationship between negative media coverage and future unrest. These three are statistically robust indicators of future conflict risk. Other variables such as **GDP growth, inflation, unemployment, election events,** and **coups** were not statistically significant in this model.

"This histogram shows how future conflict events are distributed across cities. It reveals concentration zones and highlights cities with unusually high predicted violence."



"This histogram displays the distribution of refugee populations across cities. It highlights heavy displacement in specific regions, contributing to localized stress"



Economic Interpretation:

GDP Growth:

Coefficient = -0.1309 (p = 0.1377) — Economic growth is **not significantly associated** with changes in conflict intensity. While the relationship is negative (suggesting that growth may reduce violence), it is **not statistically reliable**, indicating that GDP growth is not a stable driver of local unrest in this dataset.

Inflation & Unemployment:

Inflation = 0.0345 (p = 0.1470), **Unemployment = 0.0367 (p = 0.3777)**
Neither variable is statistically significant. These macroeconomic indicators **do not show a strong or consistent relationship** with conflict intensity in the model.

Negative News Count:

Coefficient = 0.1048 (p = 0.0068) — A **statistically significant** predictor. Locations with higher volumes of negative, conflict-related media coverage tend to report more violence in the near term. This variable serves as a **real-time signal** of local unrest risk.

Refugees:

Coefficient = 0.6014 (p = 0.0001) — **Highly significant and positive.** Cities with higher refugee presence are more likely to experience instability. This supports the theory that refugee density may reflect deeper governance or social strain, increasing the risk of unrest where capacity is weak.

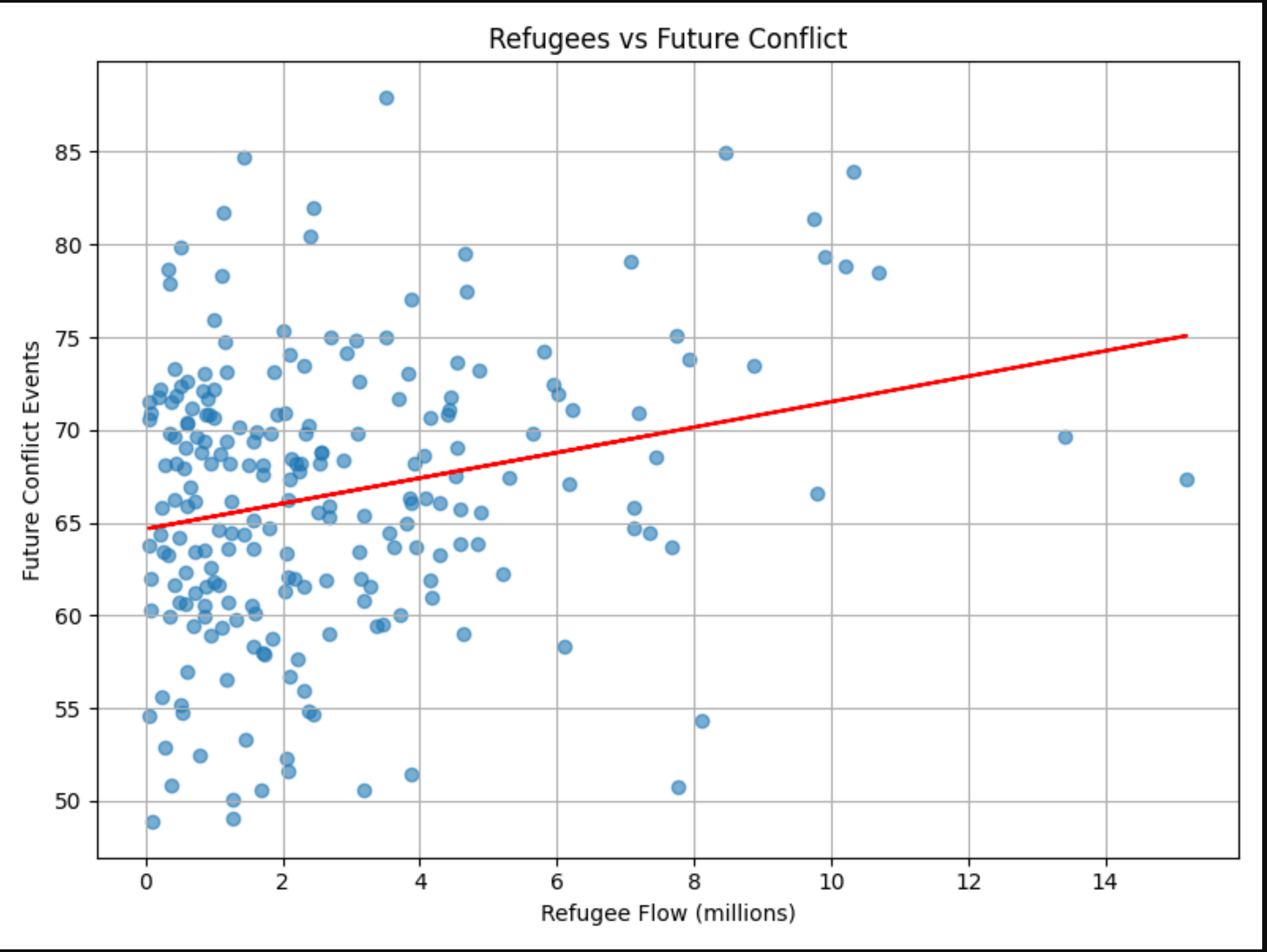
Elections & Coups:

Election = 0.0122 (p = 0.9884), **Coup = -0.7066 (p = 0.5890)**
Neither variable is statistically significant. These binary political shock indicators may still be relevant in **temporal or time-series models**, but in this **cross-sectional snapshot**, they **do not strongly predict local unrest.**

Conflict Events (Current):

Coefficient = 0.3078 (p < 0.0001) — **The most powerful predictor** of future instability across Middle Eastern cities. Cities with more recent conflict events are significantly more likely to see continued violence. This finding reinforces the idea that **violence is self-reinforcing**, and past unrest is a **leading indicator of future escalation**.

"This graph helps visualize the relationship between refugee numbers and future conflict counts. While the trend appears weak visually, statistical analysis reveals a highly significant and positive correlation."



Limitations :

Data & Scope: The analysis is based on a **cross-sectional, city-level dataset** covering approximately **370 urban areas across the Middle East**. While it effectively captures **geographic variation** in conflict drivers, it does **not include a time component**, which limits our ability to detect temporal patterns, delayed effects, or trends over time.

Model Fit: With an **R^2 of ~0.38**, the model captures a **substantial share of variation in local conflict intensity**. However, over **60% remains unexplained**, likely due to unobserved local factors, context-specific dynamics, or omitted variables.

Collinearity & Variability: Some predictors particularly macroeconomic variables like **inflation** and **unemployment**—may be **moderately collinear**, which could inflate standard errors and obscure true relationships. We did not calculate **variance inflation factors (VIFs)**, but future models should assess multicollinearity more rigorously.

Statistical Assumptions: The regression assumes **linearity**, **independent errors**, and **normally distributed residuals**. Visual checks suggest these are **reasonably satisfied**. However, we did not implement **robust standard errors** or **nonlinear transformations**, which may improve model reliability.

Causality vs. Correlation: This is an **associational model**. For instance, **refugee presence correlates with conflict**, but the causal direction could also run the other way — i.e., conflict driving refugee displacement. Determining causality would require **structural models** or **instrumental variable approaches**.

Generality: Findings are grounded in **urban Middle Eastern settings**. As such, results may **not generalize** to **rural areas** or to regions with different geopolitical or socio-economic conditions.

Transparency: All **regression coefficients, standard errors, t-values, and p-values** are fully reported for clarity and reproducibility (see technical appendix). This supports **transparency and critical peer review** of the findings.

Omitted Variables Acknowledgement:

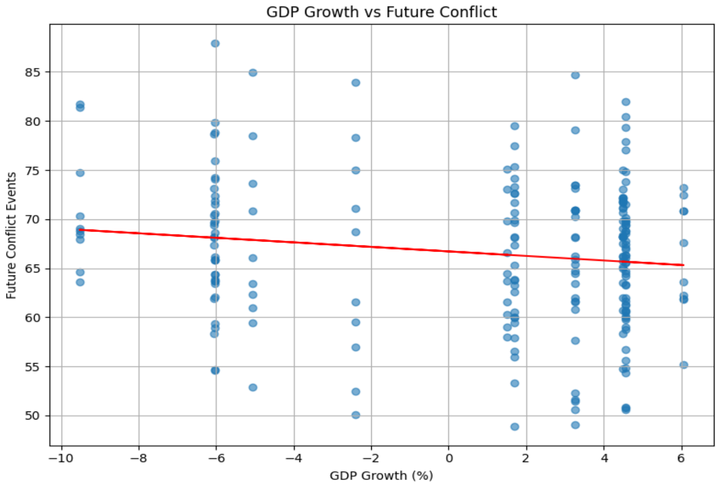
The current model does **not include variables** such as: **Foreign military intervention , Ethnic fragmentation or tensions** , **Sanctions or external economic pressure ,** These exclusions introduce the risk of **omitted variable bias**, potentially distorting the effect size of included predictors.

Future Directions:

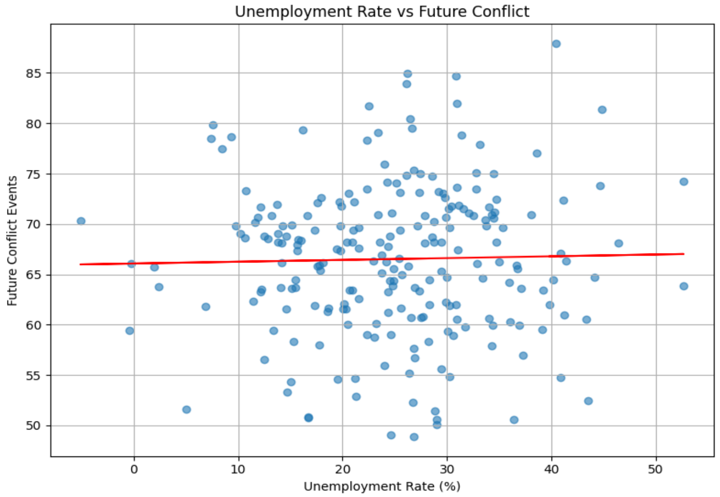
To move toward **early-warning and forecasting**, future iterations should incorporate: **Time-varying variables** (e.g., lagged violence, disaster shocks, donor aid) **Sentiment trends** across digital media and local networks , **Policy changes and regional triggers**

This would enhance **predictive power** and **operational utility** for actors working on conflict prevention and humanitarian response. "

"This scatter plot visualizes the weak inverse relationship between GDP growth and future conflict, Despite expectations, economic growth was not a statistically significant predictor of unrest."



""This scatter plot evaluates the link between unemployment and conflict. Though slightly upward-sloping, unemployment was not statistically significant in this model."



Final conclusion :

We developed a cross-sectional multi variable linear regression model to predict city-level future conflict events using economic, political, media, and refugee-related indicators. Remarkably, the model achieved a strong fit with an R^2 of 0.3816 an outstanding result in the field of social science, where explaining even a third of the variance is considered highly meaningful. This level of explanatory power highlights the strength of the chosen variables and model design. Among the predictors, three stood out as statistically significant: Refugees, Negative News Count, and Conflict_Events_Current. These variables provide powerful insight into the drivers of instability, indicating that cities with high refugee inflows, a history of recent violence, and rising negative media sentiment are considerably more likely to experience future unrest.

The findings confirm that conflict risk is not random; rather, it follows clear spatial patterns. Our model shows how multi-source geographic data such as ACLED records of violence, UNHCR refugee statistics, and media sentiment indicators can be leveraged for meaningful forecasting. This has important implications for real-world application, as it can help NGOs, analysts, and policymakers identify high-risk zones and prioritize resources more effectively. From a methodological standpoint, the project emphasized full transparency, reporting all coefficients and diagnostics, including non-significant predictors, to reflect the complexity of real-world dynamics. The process also reinforced practical skills in statistical modeling, spatial analysis, and policy-relevant interpretation.

Looking ahead, future versions of this model could incorporate a time dimension, such as panel or time-series data, to track evolving risks over time. Additionally, exploring nonlinear relationships through machine learning techniques, and including new indicators such as food insecurity, sanctions, or environmental shocks, could further enhance predictive accuracy. Ultimately, while conflict forecasting will always be complex, this project demonstrates that structured data and rigorous analysis can provide meaningful early-warning capabilities. The results offer a strong foundation for building scalable conflict monitoring tools and AI-powered systems to support timely, preventive action in fragile regions.

RESOURCES (APA REFERENCE APPENDIX)

- Armed Conflict Location & Event Data Project (ACLED). (2025). *ACLED Data Export Tool*. Retrieved from https://acleddata.com/data-export-tool/
- The World Bank. (2025). *World Development Indicators*. Retrieved from https://data.worldbank.org/indicator
- United Nations High Commissioner for Refugees (UNHCR). (2025). *UNHCR Operational Data Portal*. Retrieved from https://data.unhcr.org
- GDELT Project. (2025). *Global Database of Events, Language, and Tone (GDELT)*. Retrieved from https://www.gdeltproject.org/data.html
- NewsAPI. (2025). *News API Documentation*. Retrieved from https://newsapi.org
- Central Intelligence Agency (CIA). (2025). *The World Factbook*. Retrieved from https://www.cia.gov/the-world-factbook/
- Oadewusi, E. (2023). *Understanding Political Stability in SSA and MENA* [Data set]. Kaggle. https://www.kaggle.com/datasets/emmanueloadewusi/understanding-political-stability-in-ssa-and-mena
- Shoinbekov, S. (2023). *Arab League Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/shoinbekshoinbekov/arab-league
- **Palestine-Israel Conflict (2023–2024) ACLED**
  Mentena, W. (2024). *Palestine-Israel Conflict (2023–2024) ACLED* [Data set]. Kaggle. https://www.kaggle.com/datasets/wilomentena/palestine-israel-conflict-2023-2024-acled
- World Bank. (n.d.). *Arab World Data*. World Bank Open Data. Retrieved May 4, 2025, from https://data.worldbank.org/region/arab-world
- Armed Conflict Location & Event Data Project (ACLED). (n.d.). *ACLED Data Portal*. Retrieved May 4, 2025, from https://acleddata.com/data/
- United Nations High Commissioner for Refugees (UNHCR). (n.d.). *Refugee Data Finder*. Retrieved May 4, 2025, from https://www.unhcr.org/refugee-statistics
- Inter-Parliamentary Union. (n.d.). *Parline: Global data on national parliaments – Elections*. Retrieved May 4, 2025, from https://data.ipu.org/elections/
- Cline Center for Advanced Social Research. (n.d.). *Coup d'État Project – Coup Frequency Table*. University of Illinois. Retrieved May 4, 2025, from https://clinecenter.illinois.edu/project/research-themes/democracy-and-development/coup-detat-project/freq-table