Classification and Predictions of the Palestine – Israel Conflict: An ML and Econometrics Approach

Arya Amarnath

Motivation and Background

History On this Ongoing Conflict Since 1948

- Rise of Zionism + Jewish Immigration to Palestine after WW2
- Israel Formed as a State in 1948 Sparking First Israeli-Arab War
- Six-Day War, First and Second Intifada
- Hundreds of Failed Peace Negotiations
- Over 6 Million Palestinians Displaced

Since October 7 2023, Israel – Hamas 'War':

- 35,000 Palestinian and 1400 Israeli Dead
- 2 Million Palestinians Trapped

How can Data Science Be Used For Peacemaking and Conflict Resolution?
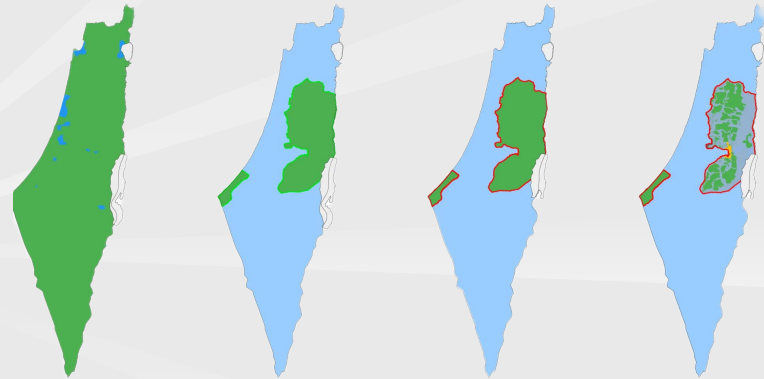
- Forecasting and Early Warning Systems

Figure 2, Al-Jazeera

Goals

Main Objective: Classify and Predict the 'Sub-Event Type' Variable

Sub Objective: Utilizing Predictions, Create a Time-Series Forecast of Sub-Event Type

Data

Sourced From: Armed Conflict Location & Event Data Project (ACLED)

Utilizing Data Export Tool:

- **April 20 2021 -> April 19, 2024**
    - **Israel Conflict Data**
    - **Palestine Conflict Data**

38,130 Rows + 32 Columns

Table 2: ACLED Event Types

Event type	Sub-event type	Disorder type
Battles	Government regains territory	Political violence
	Non-state actor overtakes territory	
	Armed clash	
Protests	Excessive force against protesters	Political violence; Demonstrations
	Protest with intervention	Demonstrations
	Peaceful protest	
Riots	Violent demonstration	
	Mob violence	
Explosions/ Remote violence	Chemical weapon	Political violence
	Air/drone strike	
	Suicide bomb	
	Shelling/artillery/missile attack	
	Remote explosive/landmine/IED	
	Grenade	
Violence against civilians	Sexual violence	
	Attack	
	Abduction/forced disappearance	
Strategic developments	Agreement	Strategic developments
	Arrests	
	Change to group/activity	
	Disrupted weapons use	
	Headquarters or base established	
	Looting/property destruction	
	Non-violent transfer of territory	
	Other	

Figure 3, ACLED Codebook

Data

Event Date	Year	Disorder Type	Event Type	Sub Event Type	Actor 1	Inter 1	Actor 2
19-Apr-24	2024	Political violence	Explosions/Remote violence	Shelling/artillery/missile attack	Hamas Movement	3	Military Forces of Israel
19-Apr-24	2024	Demonstrations	Protests	Peaceful protest	Protestors (Israel)	6	
19-Apr-24	2024	Demonstrations	Riots	Violent demonstration	Rioters (Israel)	5	Rioters (Israel)

Data

Inter 2	Interaction	Civilian Targeting	Region	Country	Admin 1	Latitude	Fatalities
0	30	0	Middle East	Israel	HaDarom	31.525	0
0	60	0	Middle East	Israel	Haifa	32.6363	0
5	55	0	Middle East	Israel	HaMerkaz	31.8766	0

Exploratory Data Analysis

Monthly Count of Events by Event Type
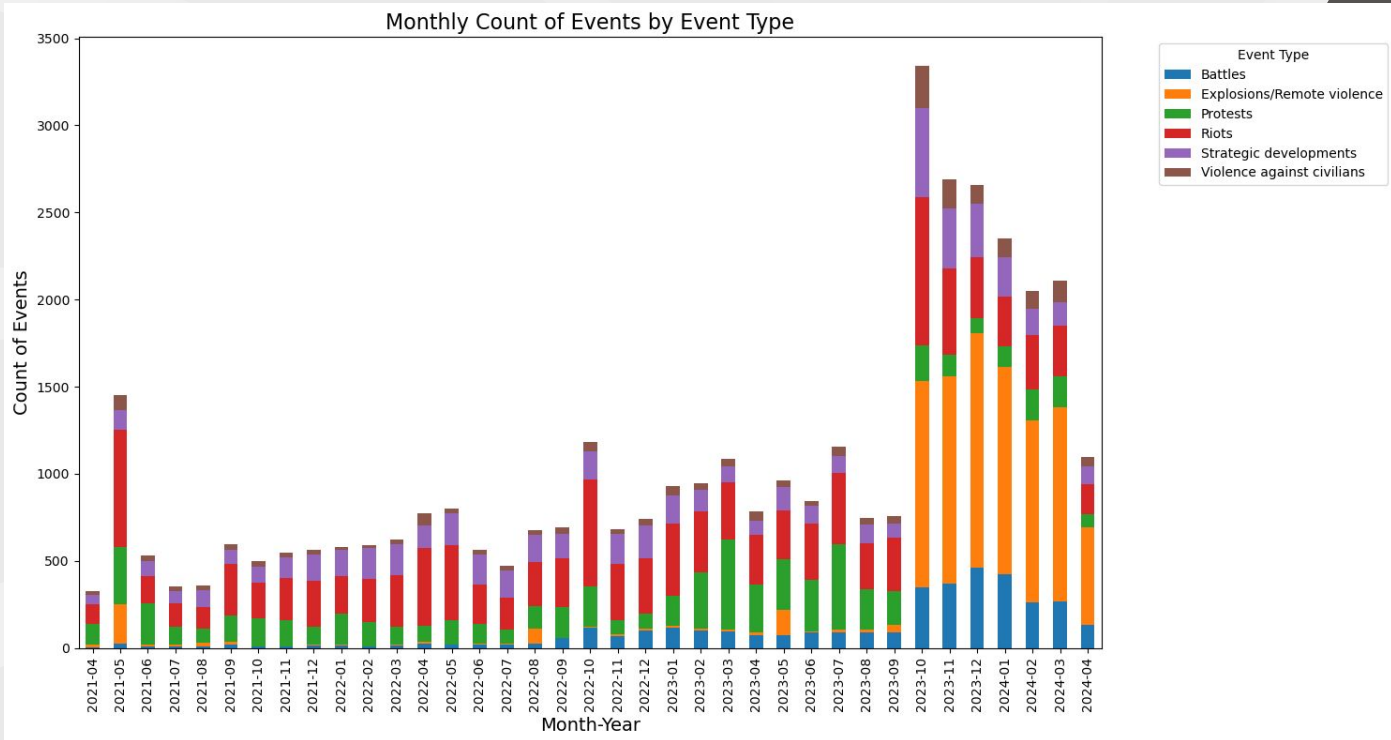
Figure 4, Amarnath
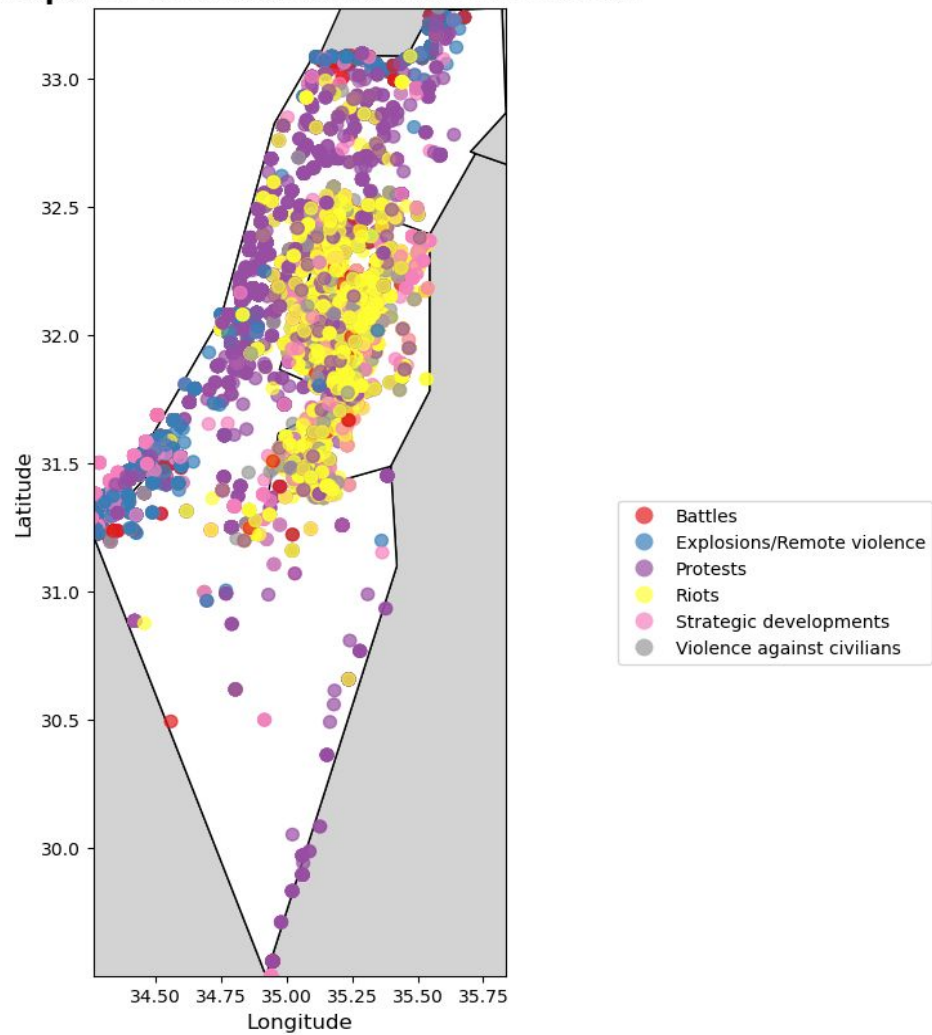
Upsurge in Conflict Events since October 7th, 2023 -> Hamas-led Attack on Israel in response to Illegal Israeli Occupation of Palestinian Territories

Geospatial Distribution of Conflict Events

EDA Continued

Data Wrangling

38,130 Rows + **32** Features

28,252 Rows + **14** Features

28,252 Rows + **36** Features

Renaming NA's, Dropping Duplicates and Irrelevant Features

Creating Temporal Features, Lagged Variables and Log Transform

9878 Duplicates

'event_type', 'assoc_actor_1', 'latitude/longitude', 'source', 'notes'...

All Removed

Temporal Features -> Date-Related, Time Since Last 'X', Rolling Average, Cumulative Count of 'X'

Lag Factor of 3 -> Value of a Predictor Variable 3 Days Prior

Log Transformations -> Stabilize Variance and Reduce Extreme Values

Encoding

1 Label Encoding

Converts each value within a categorical column into a unique integer

Used On -> Actor 1 and Actor 2

Actor Before	Actor After
Military Forces of Israel	1
Civilians (Pal.)	2
Civilians (Isr.)	3

2 One-Hot Encoding

Transforms each categorical variable into a set of binary variables, shown through 1's and 0's

Used On -> All Other Features

Disorder Type Before	Disorder Type Demonstrations	Disorder Type Political Violence	Disorder Type Strategic Developments
Demonstrations	0	1	0
Political Violence	0	0	1
Strategic Developments	1	0	0

Feature Selection

Combined Top Features ->

Top 20% of Selected Features From **Mutual Information** and **Random Forest** Importance
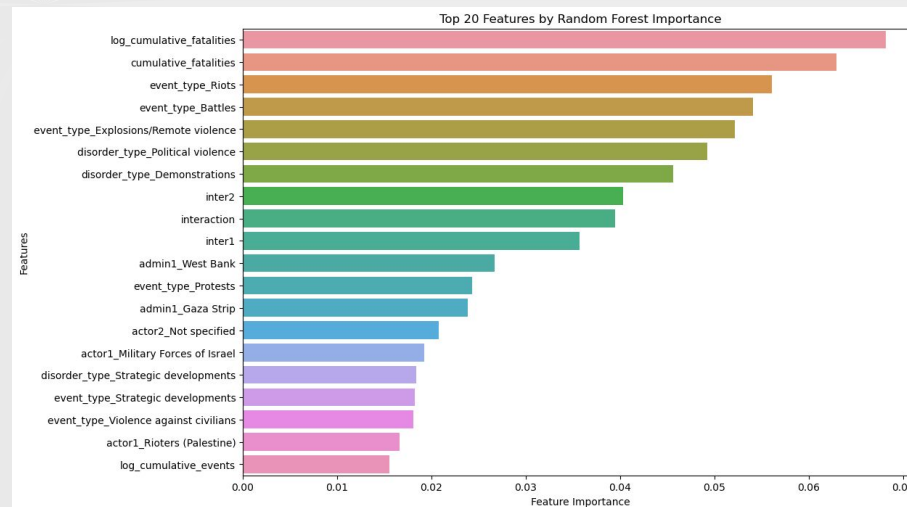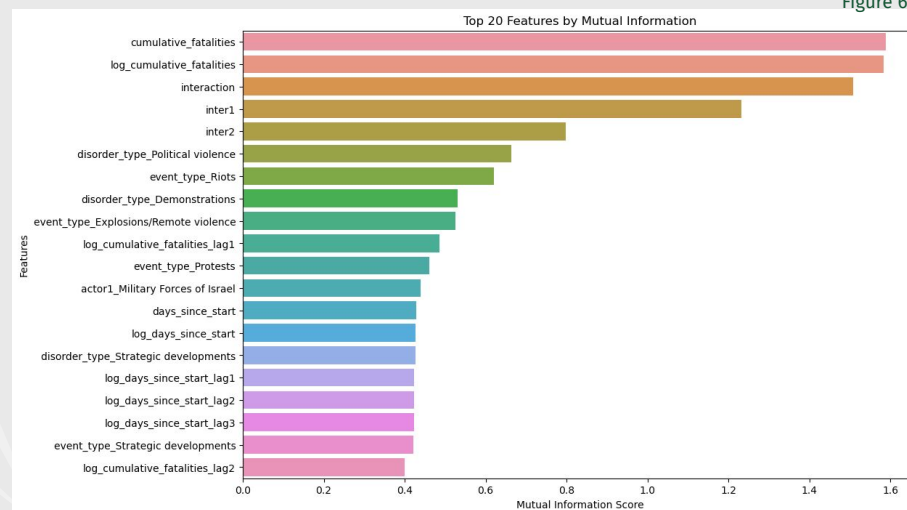
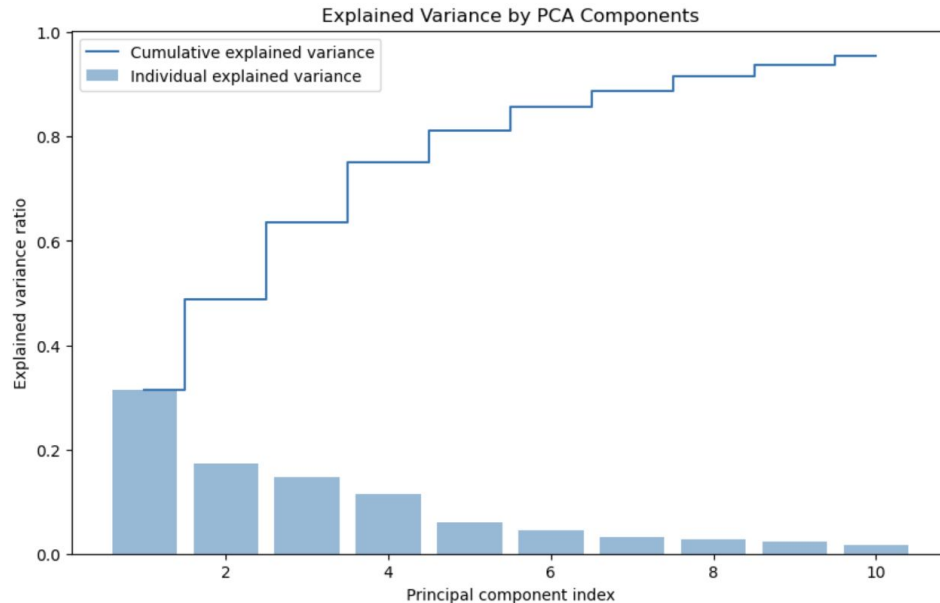**155** Features down to **19** Features

Top 20 Features by Mutual Information

Top 20 Features by Random Forest Importance

Principal Component Analysis

Explained Variance by PCA Components

Reducing the **Dimensional Size** **While** **Retaining 95%** of **Variance** **Explained**

19 Features to **10** Principal **Components**

Problems Within Data

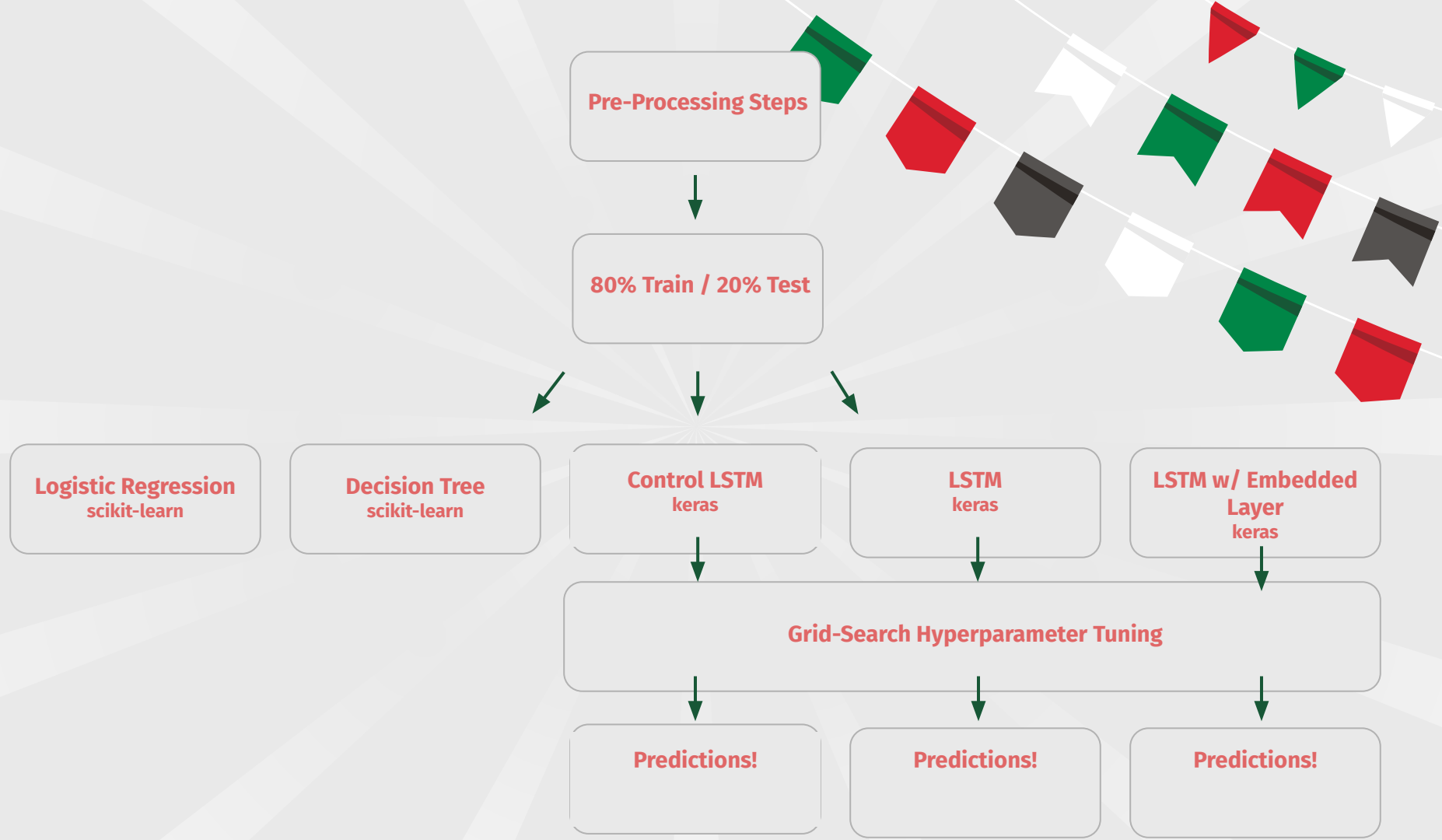**Heteroskedasticity -> Consistent Variance of Residuals**

- Breusch-Pagan Test
- Not Solved

Multicollinearity -> Highly Correlated Features

- Variance Inflation Factor
- Somewhat Addressed Through Removal of Collinear Values > 0.75

Ensuring Data is Stationary -> Constant mean, variance, autocorrelation

- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test
- Solved with Majority P-Values > 0.05

```
Pre-Processing Steps
        │
        ▼
80% Train / 20% Test
```

| Logistic Regression | Decision Tree | Control LSTM | LSTM | LSTM w/ Embedded Layer |
| scikit-learn | scikit-learn | keras | keras | keras |

```
        Control LSTM          LSTM          LSTM w/ Embedded Layer
             │                 │                      │
             ▼                 ▼                      ▼
        Grid-Search Hyperparameter Tuning
             │                 │                      │
             ▼                 ▼                      ▼
        Predictions!      Predictions!           Predictions!
```

LSTM Parameters

Training Data

Forget Gate

Input Gate

Output Gate

Constants

Epochs -> 50

Batch Size -> 64

Validation Split -> 0.1

Early Stopping -> 10 or 5

# of Features per time step -> 24

Output Class -> 20

Grid Search

Time steps -> 7, 14, 24, 30

Neurons -> 50, 100, 200

Dropout Rate -> 0.1 - 0.5

Model Results

	Logistic Regression	Decision Tree	Control LSTM	LSTM w/ No Embedded	LSTM w/ Embedded Layer
Test Accuracy	86.0%	77.0%	65.0%	81.9%	85.0%
Test MAE	-	-	3.58	1.93	1.93
Test MSE	-	-	48.43	26.53	25.42
Test RMSE	-	-	6.96	4.96	4.96

Classification Predictions

Actual: Looting/property destruction, Predicted: Looting/property destruction
Actual: Disrupted weapons use, Predicted: Disrupted weapons use
Actual: Shelling/artillery/missile attack, Predicted: Armed clash
Actual: Grenade, Predicted: Grenade
Actual: Armed clash, Predicted: Armed clash
Actual: Mob violence, Predicted: Mob violence
Actual: Mob violence, Predicted: Mob violence
Actual: Mob violence, Predicted: Mob violence
Actual: Armed clash, Predicted: Armed clash
Actual: Armed clash, Predicted: Armed clash
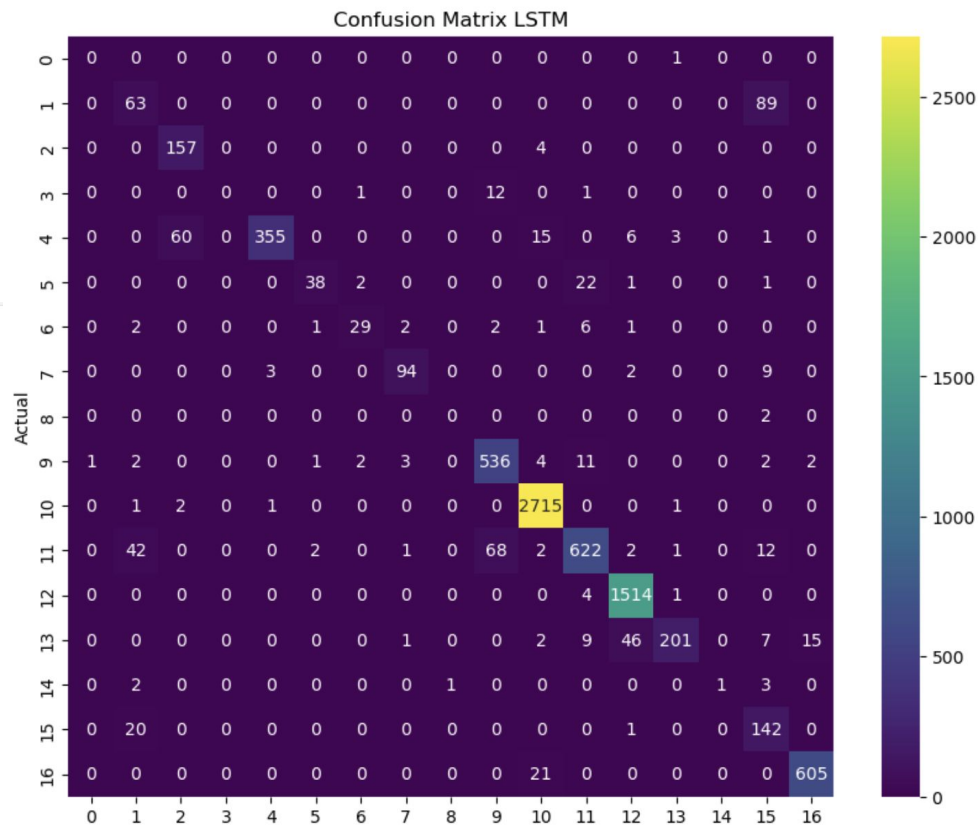
Figure 9, Amarnath

Figure 10, Amarnath

Time-Series Prediction

Multicollinearity, Heteroscedastic data and **Class Imbalance** ->

Inaccurate Predictions

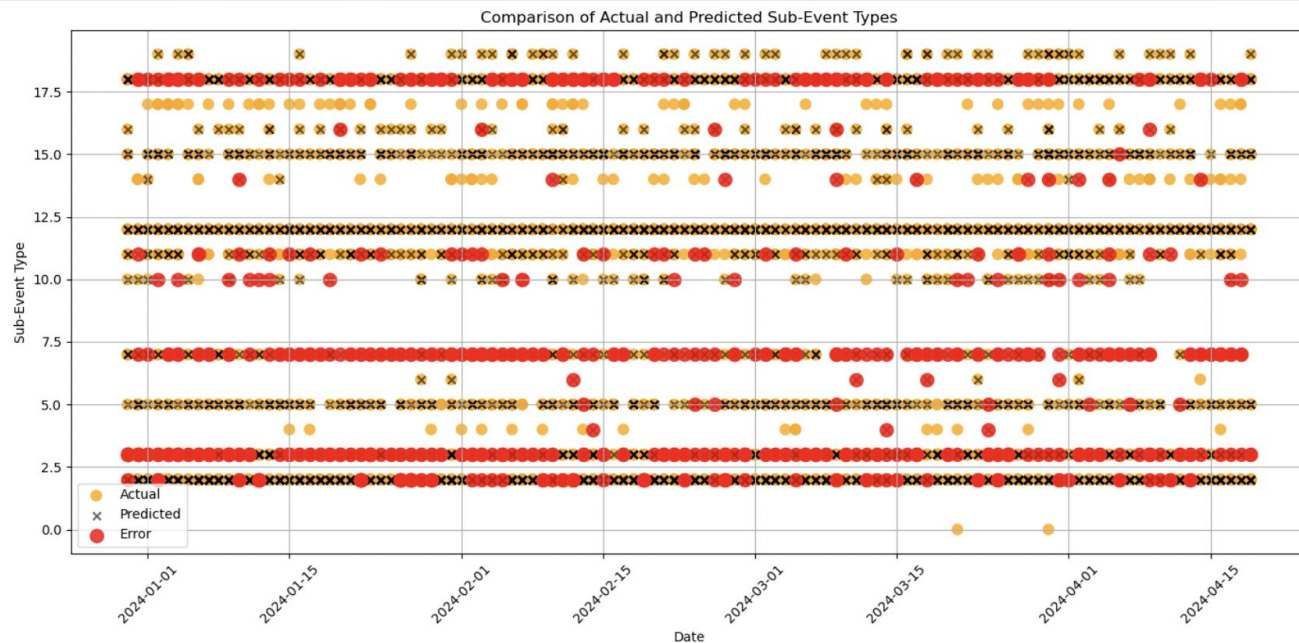**Unsuccessful** in **Generating Accurate Predictions**

Figure 11, Amarnath

Inconsistency

```
Number of errors With Embed: 764 / 5628
Index: 2, Predicted: 3, Actual: 18
Index: 18, Predicted: 3, Actual: 18
Index: 19, Predicted: 3, Actual: 18
Index: 20, Predicted: 3, Actual: 18
Index: 35, Predicted: 2, Actual: 18
Index: 39, Predicted: 3, Actual: 18
Index: 40, Predicted: 3, Actual: 18
Index: 45, Predicted: 2, Actual: 18
Index: 57, Predicted: 3, Actual: 18
Index: 70, Predicted: 18, Actual: 5
```

Future Work:

**Properly Addressing
Time-Series Data Issues**

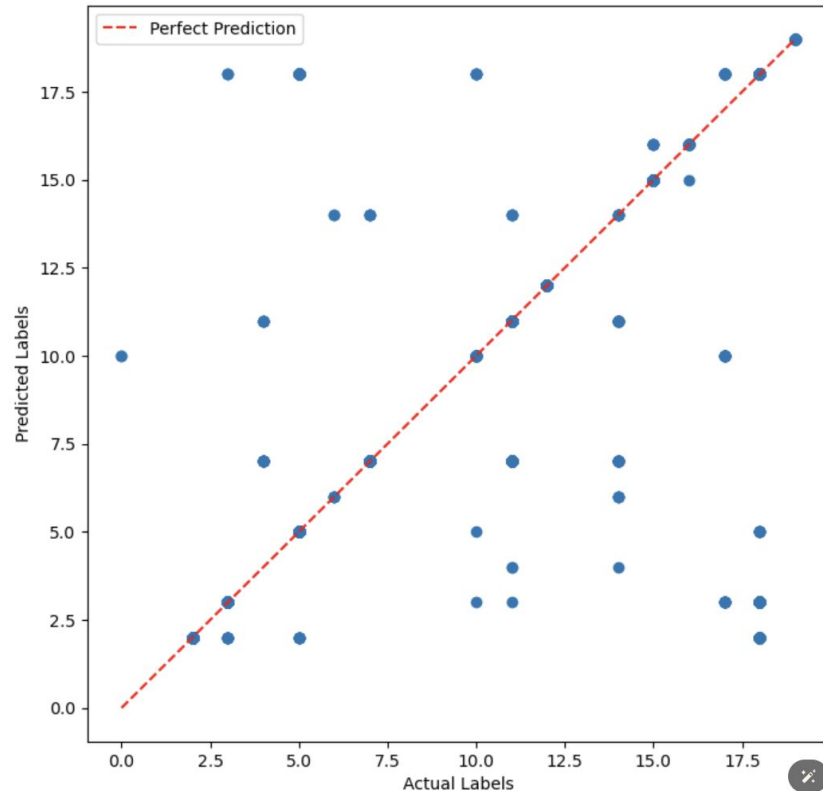**Swapping Datasets or Adding
in External Temporal
Variables**

Figure 12, Amarnath

Thank you and Free Palestine!

References:

Al Jazeera. (2023, November 27). Palestine and Israel: A brief history with maps and charts. Retrieved from https://www.aljazeera.com/news/2023/11/27/palestine-and-israel-brief-history-maps-and-charts

Ananth, Anagha D, and Sujatha Arun Kokatnoor. "An enhanced ensemble of long short-term memory and vector autoregression for energy consumption forecasting." *Indian Journal of Science and Technology*, vol. 14, no. 43, 12 Nov. 2021, pp. 3227–3236, https://doi.org/10.17485/ijst/v14i43.655.

Armed Conflict Location & Event Data Project (ACLED). (2023). ACLED Codebook 2023. Retrieved from https://acleddata.com/acleddatanew/wp-content/uploads/dlm_uploads/2023/06/ACLED_Codebook_2023.pdf

Brownlee, Jason. "How to Grid Search Deep Learning Models for Time Series Forecasting." *MachineLearningMastery.Com*, 27 Aug. 2020, machinelearningmastery.com/how-to-grid-search-deep-learning-models-for-time-series-forecasting/.

CPA. "Israeli-Palestinian Conflict | Global Conflict Tracker." Council on Foreign Relations, Center for Preventive Action, www.cfr.org/global-conflict-tracker/conflict/israeli-palestinian-conflict#:~:text=The%20Israeli%2DPalestinian%20conflict%20dates,into%20Arab%20and%20Jewish%20states. Accessed 1 May 2024.

Grace-Martin, Karen. "Eight Ways to Detect Multicollinearity." The Analysis Factor, 31 Dec. 2023, www.theanalysisfactor.com/eight-ways-to-detect-multicollinearity/.

SlidesGo. (n.d.). Palestine flag minitheme. Retrieved from https://slidesgo.com/theme/palestine-flag-minitheme