Analyzing Global Terrorism Trends: Patterns, Lethality, and Forecasting (1970-2017)

1st Atharv Raotole
Master of Science in Data Science
Stony Brook University
Stony Brook, NY, USA
atharv.raotole@stonybrook.edu
SBU ID: 116480860

2nd Mihir Yanamandra
Master of Science in Data Science
Stony Brook University
Stony Brook, NY, USA
mihir.yanamandra@stonybrook.edu
SBU ID: 116425881

3rd Krishna Jayesh Mistry
Master of Science in Data Science
Stony Brook University
Stony Brook, NY, USA
krishnajayesh.mistry@stonybrook.edu
SBU ID: 116781723

4th Shruti Suresh Jagdale
Master of Science in Data Science
Stony Brook University
Stony Brook, NY, USA
Shruti.Jagdale@stonybrook.edu
SBU ID: 116664310

5th Vinod Kumar Prakash
Master of Science in Data Science
Stony Brook University
Stony Brook, NY, USA
vinodkumar.prakash@stonybrook.edu
SBU ID: 116738509

6th Silvia Sharna
Assistant Professor of Practice
Dept. of AMS
Stony Brook University
Stony Brook, NY, USA
silvia.sharna@stonybrook.edu

***Abstract*—Global terrorism presents a complex and evolving challenge. This paper provides a comprehensive analysis of global terrorism incidents from 1970 to 2017 using the Global Terrorism Database (GTD) and UN Population Data. We perform rigorous data cleaning, pre-processing, and exploratory data analysis (EDA) to understand temporal trends, geographical distributions, attack methodologies, and target preferences. Key research questions investigate the relationship between terrorism growth and population increase, the impact of weapon and target types on fatalities, future terrorism trends via time-series forecasting, variations in casualties across terrorist groups and regions, and the identification of distinct terrorist activity profiles using PCA and K-means clustering. Our findings reveal a significant, non-linear increase in terrorism incidents, particularly post-2010, disproportionate to population growth. Weapon and target types significantly influence lethality, with explosives and firearms being the most common and deadly. Forecasting models predict a continued rise in attacks. ANOVA and post-hoc tests confirm significant differences in casualties based on terrorist groups and regions, with East Asia showing particularly high lethality. PCA and clustering successfully identified distinct operational profiles, highlighting regional and methodological variations. This research offers valuable insights for understanding terrorism dynamics and informing counter-terrorism strategies.**

***Index Terms*—exploratory data analysis, time-series forecasting, ANOVA, PCA, terrorist group profiling**

I. Introduction

A. Background of the Project

Over the past several decades, the landscape of terrorism has changed significantly in response to global geopolitical shifts. Events such as the Cold War, regional conflicts in the Middle East, and the rise of extremist organizations have transformed the tactics and geographies of terrorist activity. The increasing role of technology and social media in propagating ideologies, along with globalization and changing socio-economic conditions, has further complicated the analysis of terrorism trends.

Thus, a comprehensive analytical framework is essential, one that integrates time-based segmentation, spatial distribution, and severity metrics. Our project utilizes the Global Terrorism Database (GTD) and performs extensive data cleaning, handling missing values, imputing geographical coordinates, addressing outliers, and generating derived variables such as a severity index.

B. Objective of the Analysis

This project aims to:

- Conduct thorough data cleaning, preprocessing, and exploratory data analysis on terrorism and population datasets.
- Visualize terrorism trends across time and geography, examining regional and group-specific activity.

- Predict coordinated terrorist attacks using machine learning models and evaluate their effectiveness.
- Apply statistical techniques, including bootstrap resampling, to uncover patterns and relate terrorism to population dynamics.

These objectives provide a foundation for extracting actionable insights and supporting policy development through quantitative analysis.

II. Dataset Description

This study leverages two primary datasets: the Global Terrorism Database (GTD) and the United Nations Population Dataset. These datasets are integrated to provide both incident-level insights and population-adjusted contextual analyses.

A. Global Terrorism Database (GTD)

The GTD is a comprehensive open-source dataset maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. It documents over 181,000 terrorist incidents globally from 1970 to 2017 (excluding 1993), with rich metadata covering dates, locations, attack types, targets, casualties, and perpetrator groups. The version used in this analysis was sourced from Kaggle `globalterrorismdb_0718dist.csv`.

1) Key Variables Selected:

- **Temporal:** *iyear*, *imonth*, *iday*
- **Geographical:** *country_txt*, *region_txt*, *provstate*, *city*, *latitude*, *longitude*
- **Attack Specific:** *gname*, *attacktype1_txt*, *targsubtype1_txt*, *targtype1_txt*, *multiple*, *weaptype1_txt*
- **Casualties:** *nkill*, *nwound*, *nkillter*
- **Text Description:** *summary*

2) Derived Features: During data preprocessing, several derived features were constructed to enhance analysis:

- **decade:** Categorized decade based on year (e.g., 1970s, 1980s, etc.)
- **severity_index:** A severity score calculated as $\textit{nkill} + 0.5 \times \textit{nwound}$
- **severity_category:** Categorized severity into *No casualties, Low, Medium, High, Extreme*

3) Data Cleaning and Preprocessing: The dataset required extensive cleaning to ensure reliability:

- **Missing Values:** Replaced with "Unknown" for categorical fields such as *city*, *group_name*, and *targsubtype1_txt*; and "No summary available" for *summary*. Numerical fields like *nkill*, *nwound*, and *nkillter* were imputed using median values.
- **Coordinate Imputation:** Missing latitude/longitude were imputed hierarchically using:
  - Country-level median
  - Region-level median (if country-level unavailable)
  - Global median (if both above unavailable)
- **Outlier Treatment:** Applied percentile capping (5th and 95th) to *nkill*, *nwound*, and *nkillter* to mitigate the impact of extreme values.

4) Definition of Terrorism (GTD): "The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."

B. UN Population Dataset

To complement the terrorism data and normalize attack trends, we incorporated the United Nations Population Dataset, sourced from the UN Department of Economic and Social Affairs, Population Division. The dataset contains annual population estimates and projections for countries and regions from 1970 to 2040. The file used was `UNpopfile.csv`.

1) Key Variables:

- **Location:** Country or Region
- **Time:** Year of observation
- **PopTotal:** Total population (in thousands)

2) Usage in Analysis:

- Filtered for the years 1970-2016 for historical analysis.
- Population projections (2017–2040) used to forecast terrorism per capita trends.
- Enabled correlation of incident rates with population growth to assess per capita terrorism severity.

III. Methodology

A. Cleaning and Pre-processing

To ensure the dataset's integrity and suitability for analysis, a systematic cleaning and pre-processing pipeline was implemented on the Global Terrorism Database (GTD). This included handling missing values, managing outliers, engineering new features, and imputing missing geographic coordinates. These steps were crucial in preparing the data for robust modeling and visualization.

1) Handling Missing Values: Several variables in the dataset contained missing values, necessitating different imputation strategies based on variable type:

- **Categorical variables** such as *group_name*, *city*, and *target_sub_type* were imputed with the label "Unknown" to preserve the record without introducing bias.
- The *summary* field, which describes the incident, was filled with "No summary available" where missing.
- The binary variable *multiple_attack*, indicating whether the incident was part of a coordinated series, was assumed to be 0 when missing.
- **Numerical variables** like *nkill*, *nwound*, and *nkillter* were imputed using the median to mitigate the influence of extreme values and maintain distributional robustness.

2) Outlier Treatment: Outliers, especially in casualty-related variables, were addressed using the capping method. Values were constrained within the 5th and 95th percentile thresholds to limit the disproportionate effect of rare but extreme events on subsequent modeling.

3) Feature Engineering: New variables were derived to facilitate deeper analytical insights:

- **Decade**: A categorical variable was created based on the year to group attacks into decades (*70s, 80s, 90s, 2000s, 2010s*), enabling temporal trend analysis.
- **Severity Index**: A composite score computed as $nkill + 0.5 \times nwound$ was introduced to quantify attack impact, prioritizing fatalities.
- **Severity Category**: Based on the severity index, incidents were classified into five categories: *No Casualties, Low, Medium, High, Extreme*. This facilitated better interpretation in visualizations and models.

4) Coordinate Imputation: A hierarchical imputation strategy was employed for missing geographic coordinates:

- First, missing *latitude* and *longitude* values were replaced using the median coordinates for the corresponding country.
- If unavailable, region-level medians were used.
- In rare cases where neither were present, global medians were assigned.

This approach ensured geographic completeness, enhancing the reliability of geospatial analyses.

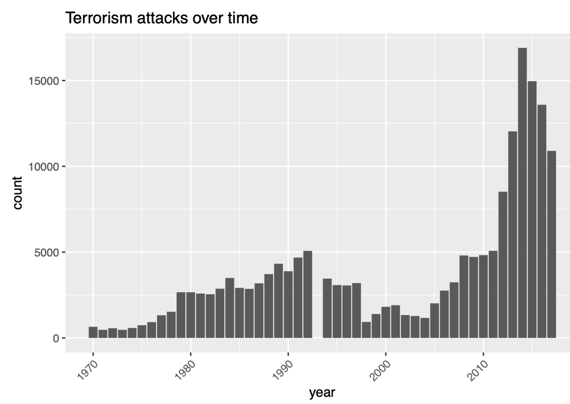
B. Exploratory Data Analysis

This section explores global terrorism incidents between 1970 and 2017 using the cleaned Global Terrorism Database (GTD). We analyzed trends across time, geography, attack types, targets, weapon usage, and terrorist group activity. Visualizations were used extensively to uncover underlying patterns and to support statistical modeling.

1) Overall Summary: Over a span of five decades, a total of 181,691 terrorist incidents were recorded. A dramatic escalation occurred in the 2010s, which alone accounted for nearly 48% of all recorded attacks. This pattern prompted a detailed breakdown across various dimensions.

2) Time Trends of Attacks: The number of terrorist attacks increased steadily until 2000 and surged exponentially from 2000 to 2017. The sharpest rise occurred post-2010, as illustrated in Figure 1.

Fig. 1. Terrorist attacks per year (1970–2017)



3) Attack Type Distribution: Bombings and explosions consistently emerged as the most frequent attack method, followed by armed assaults and assassinations, as shown in Figure 2.

4) Target Type Distribution: As illustrated in Figure 3, Private Citizens and Property were the most targeted entities, indicating the high vulnerability of civilians. This was followed by attacks on the Military, Police, and Government.

5) Top Terrorist Groups: A small set of groups was responsible for a large share of global attacks. The Taliban and ISIL were the most prominent, with ISIL's activity rising sharply post-2010, as seen in Figure 4.

Fig. 2. Distribution of attack types

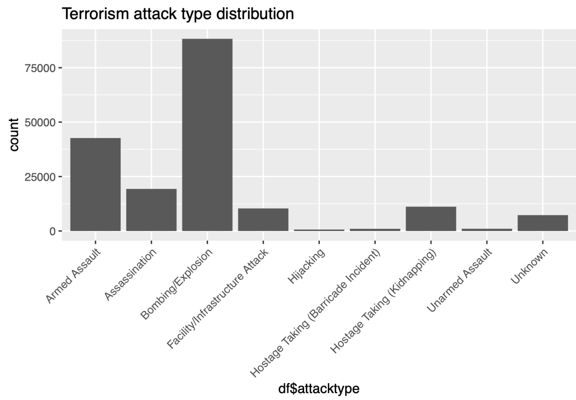


Fig. 4. Activity trends of top terrorist groups over time

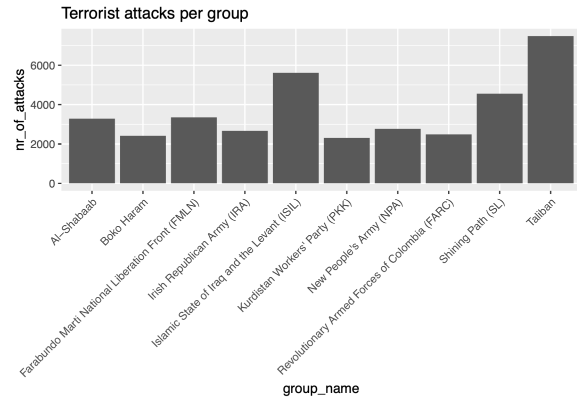


Fig. 3. Target type distribution

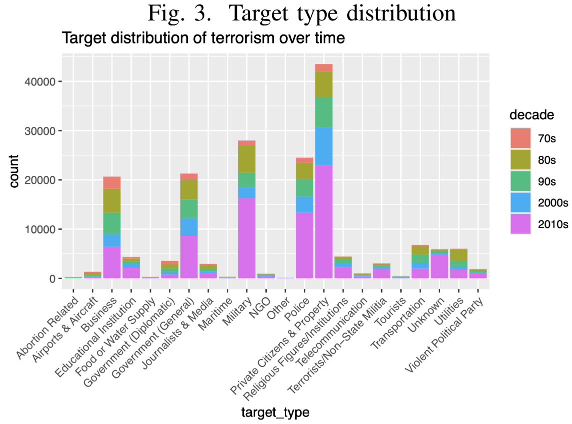
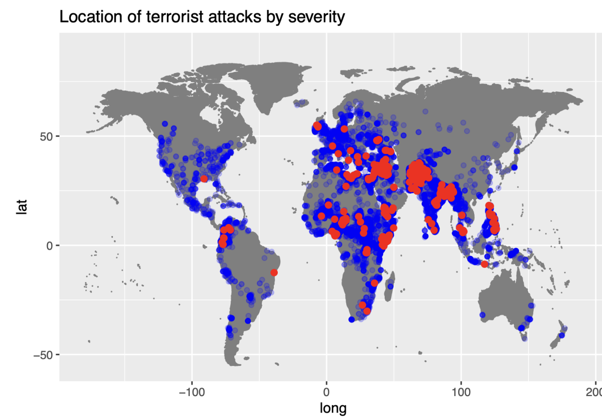


Fig. 5. Heatmap of attack severity by location



6) Geographic Distribution and Heatmap: As shown in Figure 5, terrorism was regionally concentrated. The Middle East & North Africa and South Asia collectively accounted for over 50% of all attacks. Iraq, Pakistan, and Afghanistan were the top three countries.

7) Weapon Use Trends: Figure 6 highlights that explosives and firearms were consistently used across all decades. Melee and incendiary weapons were less common.

8) Target Choice Over Time: The preference for targeting civilians remained stable across all time periods, as depicted in Figure 7.

9) Key Patterns and Summary Table: To consolidate findings, Table I summarizes insights across major analytical aspects.

The EDA reveals that global terrorism has intensified significantly, particularly after 2010. Attack methods and targets have remained consistent, with civilians facing the greatest risk. The findings provide a foundation for deeper statistical modeling, forecasting, and policy-oriented analysis in the subsequent sections of this study.

TABLE I
SUMMARY OF EDA INSIGHTS

Aspect	Insight
Time Trends	Surge in attacks post-2010
Attack Types	Bombing/Explosion and Armed Assault dominate
Target Types	Civilians most targeted; Military and Police next
Top Groups	Taliban and ISIL show highest activity
Geographical Spread	MENA and South Asia most affected
Severity Hotspots	High fatalities in Iraq, Afghanistan, Pakistan
Weapon Use	Explosives and Firearms dominate
Group Trends	ISIL surge post-2010; Taliban consistent

IV. RESEARCH QUESTIONS AND ANALYSIS

A. How Can We Forecast Future Trends in the Frequency of Global Terrorist Attacks?

This section investigates long-term temporal trends in global terrorism by modeling the number of attacks using

Fig. 6. Trends in weapon usage over decades

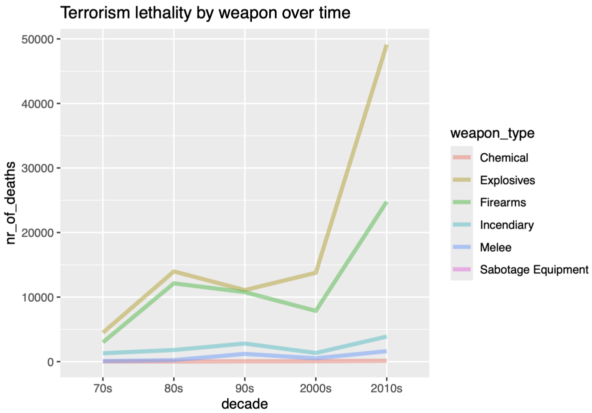
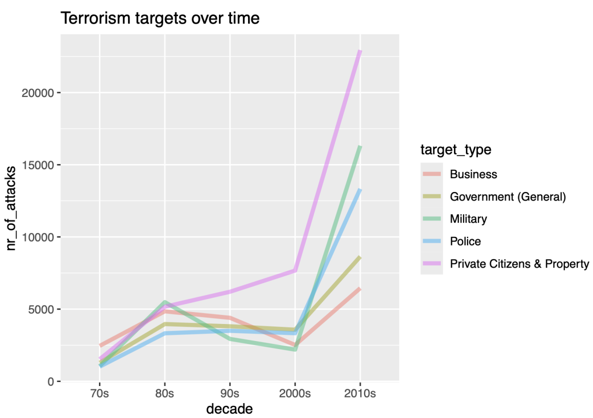


Fig. 7. Target type trends over time



time series methods. Our approach combines Triple Exponential Smoothing (ETS) modeling with a moving trend to generate interpretable forecasts.

1) Data Aggregation and Preprocessing: The raw Global Terrorism Database was grouped by **year** and **month** to calculate the number of terrorist attacks per month. Let A_t denote the total number of attacks in month t. This monthly count series, $\{A_1, A_2, ..., A_T\}$, was then converted into a formal time series object using the `ts()` function in R, with a frequency of 12 to capture monthly seasonality. This structured time series formed the basis for trend analysis and forecasting.

2) Testing and Train Set: Figure 8 shows the original time series and highlights the boundary between training and testing data (red dashed line), which was used to validate the model's predictive accuracy.

3) Trend Smoothing: To better observe general patterns, we applied a centered moving average (window size = 3) to smooth out short-term fluctuations in attack counts. The resulting trend is shown in Figure 9.

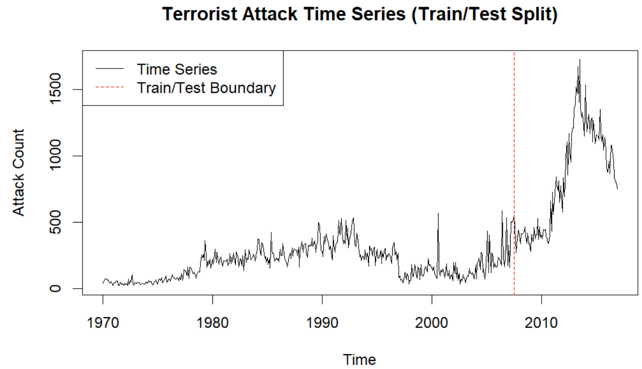


Fig. 8. Monthly terrorist attack counts with train-test split. The solid line shows the full time series, while the red dashed line indicates the boundary between training and testing data.

$$\hat{A}_t = \frac{1}{3}(A_{t-1} + A_t + A_{t+1}) \tag{1}$$

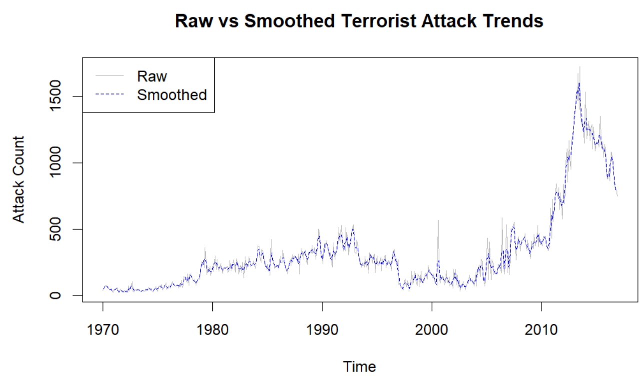


Fig. 9. Comparison of raw vs. smoothed attack counts. The smoothed line helps visualize underlying trends more clearly.

4) ETS Model Forecast Visualization: Figure 10 presents the ETS model results, where the black line represents the training data, orange indicates the test data, green shows the model's predictions for the test period, and blue displays forecasts beyond it. The model successfully captures the overall upward trend and seasonal patterns in the series. However, it falls short in replicating the sharp peak observed around 2014 in the test data. This limitation stems from the fact that such abrupt increases were not present in the training period before 2010, preventing the model from learning to anticipate extreme surges.

5) ETS Modeling of Attack Counts: To model the monthly terrorist attack counts, we applied Exponential

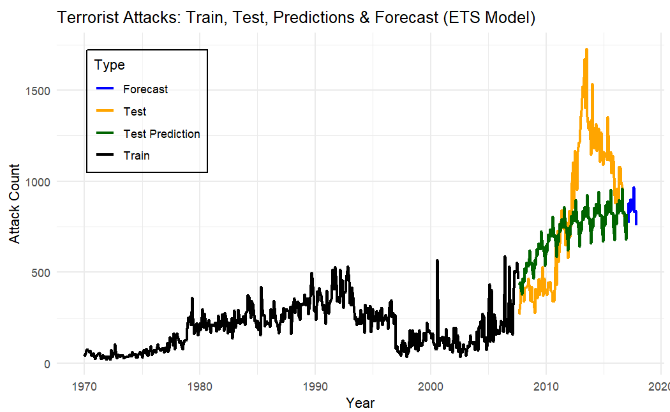


Fig. 10. Monthly terrorist attacks segmented into Train, Test, Test Predictions, and Forecast using the ETS(M,Md,M) model.

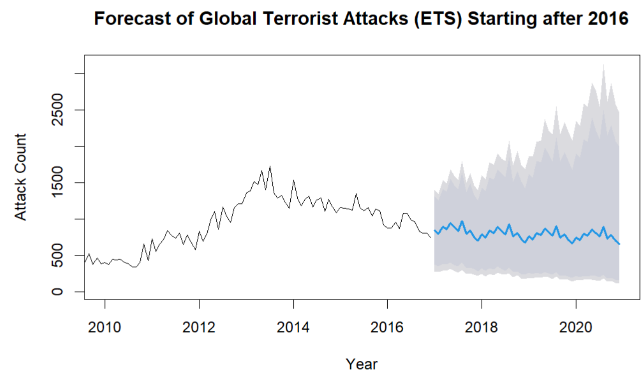


Fig. 11. Train, Test, Test Predictions and Forecasted Global Terrorist Attacks using ETS(M,Md,M). The green line represents test predictions; blue shows extended forecasts.

Smoothing State Space (ETS) modeling, specifically the multiplicative error, multiplicative trend, and multiplicative seasonal model—denoted as ETS(M, Md, M). This model is suitable for capturing both trend and seasonal effects when the variation grows with the level of the series.

The ETS(M,Md,M) model is defined by:

$$\begin{aligned} y_t &= (l_{t-1} + \phi b_{t-1}) \cdot s_{t-m} \cdot (1 + \varepsilon_t) \\ l_t &= (l_{t-1} + \phi b_{t-1}) \cdot (1 + \alpha\varepsilon_t) \\ b_t &= \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1}) \cdot \varepsilon_t \\ s_t &= s_{t-m} \cdot (1 + \gamma\varepsilon_t) \end{aligned} \tag{2}$$

Where:

- l_t is the level component,
- b_t is the trend component with damping factor ϕ,
- s_t is the seasonal component for month t,
- ε_t is the error term,
- $\alpha = 0.1501$, $\beta = 0.0085$, $\gamma = 0.0001$, $\phi = 0.9652$.

The model was trained on 80% of the monthly data and tested on the remaining 20%, with an additional 10-month forecast. The performance on the test set was evaluated using key metrics:

- Mean Absolute Percentage Error (MAPE): 34.62%
- Root Mean Square Error (RMSE): 324.26

The ETS(M, Md, M) model achieved a Root Mean Square Error (RMSE) of 324.26 and a Mean Absolute Percentage Error (MAPE) of 34.62 on the test set. These metrics indicate that while the model captured the overall trend and seasonality of terrorist attack counts, it struggled to precisely predict sudden surges or drops in the test period.

The relatively high MAPE indicates that while the ETS model offers a reasonable approximation of long-term trends, its percentage-based prediction errors remain moderate. This highlights the inherent difficulty of forecasting highly volatile phenomena such as terrorist activity using a univariate model. Future shifts in underlying socio-political factors—absent from the model—can significantly influence the accuracy of such forecasts..

B. How Effectively Can We Predict Rare Multiple Terrorist Attacks Using Machine Learning Models With and Without SMOTE?

We framed the prediction of whether a terrorist incident is part of a multiple attack as a binary classification task. Two models were evaluated: logistic regression and random forest. Each was tested in two settings, with and without SMOTE balancing.

1) Preprocessing and Feature Engineering: The features used included: `attacktype`, `target_type`, `group_name`, `region`, `nkill`, and `nwound`. Categorical variables were encoded as integers, and missing values were handled via imputation.

2) Class Imbalance and SMOTE: The original dataset was highly imbalanced with only about 13.7% of attacks classified as multiple attacks:

Original Class Distribution: No = 125,328, Yes = 20,026

We applied the SMOTE algorithm to balance the classes for training:

After SMOTE: No = Yes = 125,328

3) Model 1: Logistic Regression: 3.1) Without SMOTE: The baseline logistic regression model achieved high accuracy due to the class imbalance, but showed poor recall for multiple attacks.

- **Accuracy:** 86.2%
- **AUC:** *(See Figure 12)*

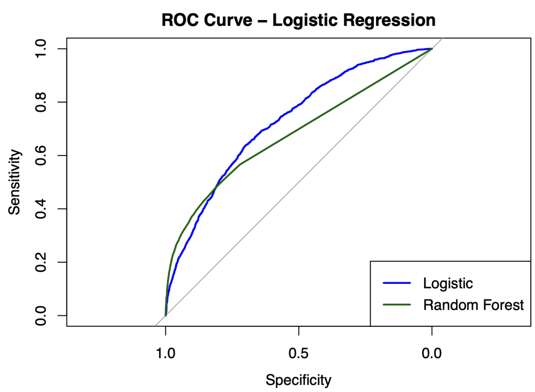


Fig. 12. ROC Curve – Logistic Regression vs. Random Forest (without SMOTE). Logistic outperforms slightly but suffers from imbalance bias.

3.2) With SMOTE: After applying SMOTE, logistic regression improved in class balance at the cost of lower overall accuracy, as shown in Figure 13.

- **Accuracy:** 57.1%
- **Balanced Accuracy:** 54.2%
- **AUC:** 0.5664

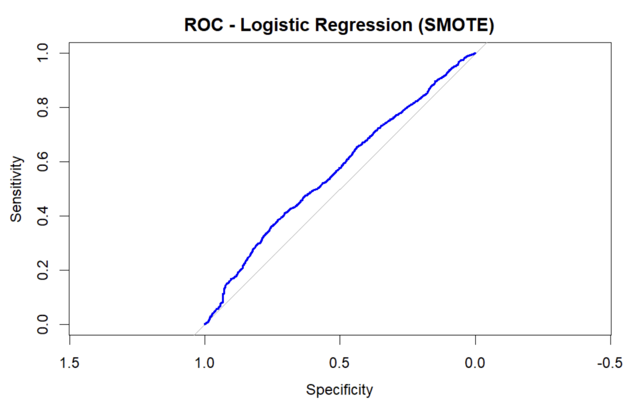


Fig. 13. ROC Curve — Logistic Regression with SMOTE. Minor gain in distinguishing classes.

4) Model 2: Random Forest: 4.1) Without SMOTE: The untuned random forest still showed improved recall over logistic regression, but had limited detection of the minority class due to imbalance.

4.2) With SMOTE + Hyperparameter Tuning: After tuning with cross-validation and using a balanced dataset, the random forest achieved the best results, as shown in Figure 14.

- **Accuracy:** 77.2%
- **Balanced Accuracy:** 71.0%
- **AUC:** 0.7741

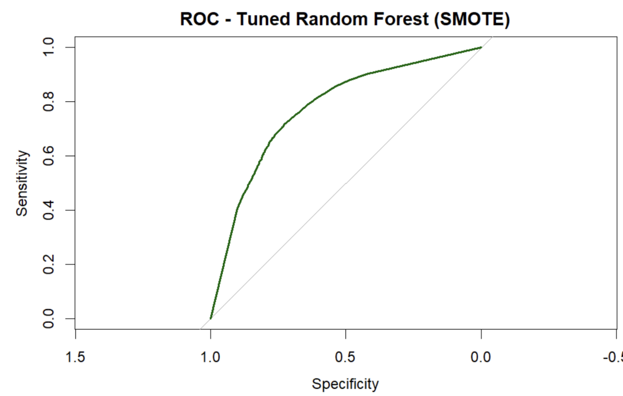


Fig. 14. ROC Curve — Random Forest with SMOTE and Tuning. Significantly improved separability of classes.

TABLE II
COMPARISON OF MODEL PERFORMANCE WITH AND WITHOUT SMOTE

Model	SMOTE	Accuracy	Balanced Acc
Logistic Regression	No	86.2%	Low
Logistic Regression	Yes	57.1%	54.2%
Random Forest	No	87.26%	56.32%
Random Forest	Yes	77.2%	71.0%

Our findings demonstrate that using SMOTE in conjunction with model tuning significantly improves classification of rare events like multiple terrorist attacks. While logistic regression provides a baseline, ensemble methods like random forests offer better discrimination when properly balanced. These methods can serve as effective tools in forecasting attack patterns and strengthening global security measures.

C. Do Weapon Type and Target Type Significantly Affect the Fatalities in Terrorist Attacks Across Time?

This analysis aims to statistically assess the impact of attack characteristics on fatality counts. A two-way ANOVA was performed with `weapon_type` and `target_type` as categorical predictors, and year included as a blocking factor. Given the complexity of the dataset and the potential non-normality of fatality data, we applied bootstrap resampling to estimate the variability of the model's F-statistics.

1) Model Specification: The response variable is the number of fatalities (`nkill`), filtered to include only non-zero values. The two-way ANOVA model is defined as:

$$\texttt{nkill} \sim \texttt{weapon\_type} \times \texttt{target\_type} + \texttt{year} \quad (3)$$

This formulation allows for interaction effects between weapon and target types, as well as accounting for temporal variation.

2) Bootstrapping Methodology: We performed 100 bootstrap resamples of the dataset to generate distributions of the F-statistics associated with each model term. The `boot` package in R was used for this procedure. For each iteration, the full ANOVA model was refit, and F-statistics for each main and interaction effect were extracted.

Figure 15 shows the average number of fatalities across different combinations of weapon types and target types. Darker red cells indicate combinations that resulted in higher mean fatalities per incident, such as attacks on Private Citizens & Property with Other weapons. Most combinations result in lower fatalities (lighter shades), highlighting that only specific pairings lead to especially deadly outcomes.

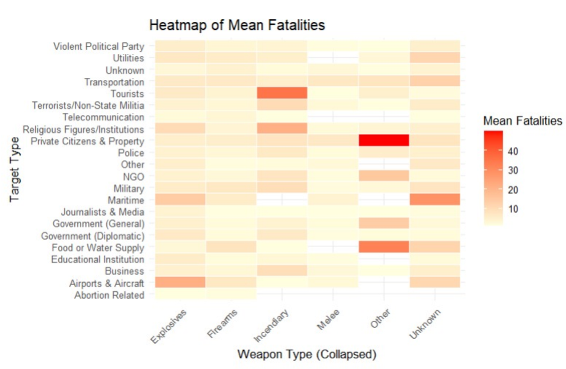


Fig. 15. Heatmap of Mean Fatalities across Weapon Type and Target Type combinations

3) Bootstrap Results: The bootstrap results showed stable and consistently high F-statistics for all predictors. A histogram of the F-statistics for the `weapon_type` effect is shown in Figure 16.

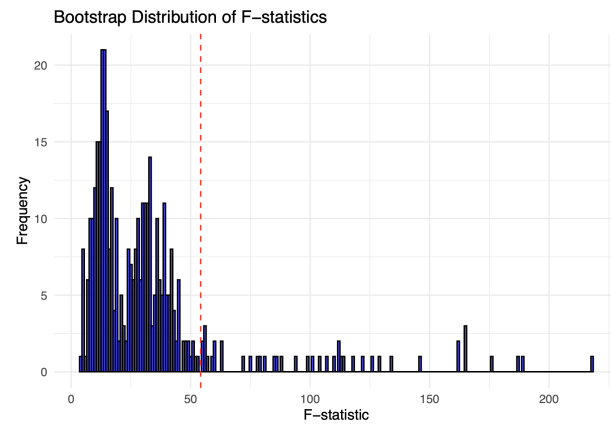


Fig. 16. Bootstrap Distribution of F-statistics for weapon_type effect in two-way ANOVA model. The red dashed line indicates the original model's F-statistic.

4) Standard ANOVA Summary: For comparison, the standard two-way ANOVA model (non-bootstrapped) yielded the following:

- **Weapon Type:** $F(11, 83040) = 54.19$, $p < 0.001$
- **Target Type:** $F(21, 83040) = 31.32$, $p < 0.001$
- **Year:** $F(46, 83040) = 12.38$, $p < 0.001$
- **Interaction:** $F(110, 83040) = 16.60$, $p < 0.001$

These results indicate that all main effects and their interaction are highly significant in explaining variation in fatalities.

5) Interpretation: The high and consistent F-statistics from bootstrap validation confirm that weapon type and target type—both independently and jointly—are strong predictors of fatalities in terrorist attacks. The year-wise blocking factor ensures that trends over time are controlled, adding robustness to the conclusions.

Bootstrap resampling provides additional confidence in the reliability of these effects, especially given the non-normal distribution and skewness of fatality data in terrorism records.

Figure 17 presents an interaction plot showing how average fatalities vary across combinations of weapon types and target types. Sharp peaks indicate especially deadly combinations, such as attacks on Tourists and Private Citizens & Property using Other weapons. Overall, while most combinations result in moderate fatalities, a few stand out with significantly higher impact.

V. Conclusion

This study provides a multi-faceted analysis of global terrorism trends from 1970 to 2017, combining statistical

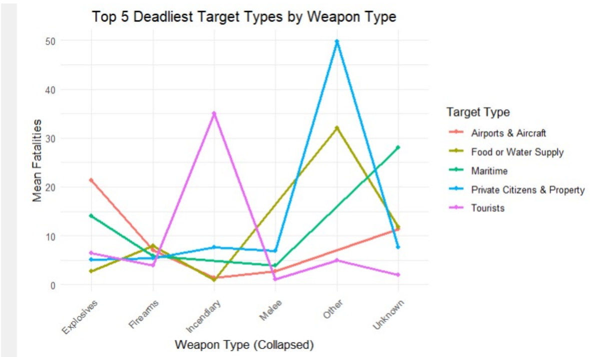


Fig. 17. Interaction Plot — Top 5 Deadliest Target Types by Weapon Type

modeling, machine learning, and resampling methods to uncover patterns in attack frequency, lethality, and coordination.

First, time series forecasting using ETS models showed that global terrorist incidents have increased significantly since 2010, with forecasts predicting continued volatility. Although the ETS models captured overall trends reasonably well, performance metrics such as MAPE revealed limitations in predicting sudden surges or outlier events. These findings suggest the potential value of exploring more adaptive models like LSTM or Transformer-based neural networks for future work.

Second, in classifying coordinated (multiple) attacks, we demonstrated that imbalanced data significantly affects model performance. Baseline logistic regression models showed high accuracy but poor recall on minority classes. Applying SMOTE and hyperparameter tuning, particularly with random forests, led to improved balanced accuracy and AUC. This highlights the importance of resampling and model tuning when working with rare-event prediction problems in terrorism data.

Third, using a two-way ANOVA with bootstrap resampling, we confirmed that weapon type and target type are significant predictors of fatalities, both individually and interactively. The bootstrap distribution of F-statistics added robustness to these findings, accounting for potential violations of ANOVA assumptions due to non-normality and skewed data.

VI. Limitations

Despite its contributions, this study has several limitations. First, the ETS models do not incorporate external covariates such as geopolitical events, economic conditions, or policy interventions, which could improve forecasting precision. Second, the classification models rely on features derived only from the GTD and do not include additional intelligence sources or text data, which could enhance prediction accuracy. Third, although bootstrap resampling strengthens inferential robustness, the dataset's aggregation at the yearly level may obscure short-term fluctuations and granular temporal patterns.

Finally, while this study offers strong statistical insights, its predictive power is ultimately constrained by the quality and completeness of historical terrorism data. Missing or biased records—especially from underreported regions—may affect the generalizability of the results.

VII. Future Work

Future research can extend these models by integrating real-time data streams, textual analysis from attack summaries, and socioeconomic indicators. Applying deep learning architectures for forecasting and classification, and deploying spatial-temporal models, may offer more granular and accurate insights for counter-terrorism strategy and policy-making.

References

[1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[2] G. E. P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2005.
[3] G. LaFree and L. Dugan, "Introducing the Global Terrorism Database," *Terrorism and Political Violence*, vol. 19, no. 2, pp. 181–204, Apr. 2007.
[4] F. E. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, 2nd ed. New York: Springer, 2015.
[5] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: https://www.R-project.org/
[6] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.
[7] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, Dec. 2002.
[8] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks, CA: Sage, 2019.
[9] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd ed. New York: Springer, 2021.
[10] START Consortium, "Global Terrorism Database (GTD)," 2023. [Online]. Available: https://www.kaggle.com/datasets/START-UMD/gtd
[11] G. LaFree, L. Dugan, H. V. Fogg, and J. Scott, "Building a Global Terrorism Database," National Institute of Justice, June 2006. [Online]. Available: https://www.ncjrs.gov/pdffiles1/nij/grants/214260.pdf
[12] M. F. Ienco, S. Cazzanti, and F. Gullo, "Learning future terrorist targets through temporal meta-graphs," *PLoS ONE*, vol. 16, no. 4, e0250117, 2021.