GLOBAL TERRORISM ANALYSIS AND PREDICTION

S. Sai praneeth[1], S. Uday[2]

ABSTRACT

Crime prediction is a challenging task that requires advanced analytical tools to complement traditional detection methods. Machine learning and deep learning algorithms have shown promising potential in this domain, enabling researchers to develop innovative solutions for uncovering criminal activities, forecasting crime patterns, and preventing offenses. This project aims to explore the Global Terrorism Database (GTD) to gain a deeper understanding of terrorism patterns over the past decades. Through data manipulation, numerical analysis, and graphical visualization, the project will generate a robust and thorough examination of the dataset, leading to substantial key findings related to global terrorism.

The project will focus on the following aspects:

1.Revealing potential trends in attack frequency, the most targeted countries, preferred methods of attack, types of weapons used, casualties, and the evolution of terrorist organizations. Identifying correlations between terrorist attacks and environmental or demographic factors. Providing a comprehensive overview of global terrorism trends to inform counter-terrorism strategies and policies. Understanding the characteristics of regions prone to attacks and the underlying reasons for their vulnerability

The project's findings will have a significant impact on counter-terrorism efforts, enabling policymakers and law enforcement agencies to develop more effective prevention and response strategies. Additionally, the research will contribute to a better understanding of the complex phenomenon of terrorism, paving the way for future research in this field.

1.INTRODUCTION

Crime prediction presents a multifaceted challenge, necessitating advanced analytical tools to complement existing detection methods. The increased availability of crime data and technological advancements have provided an opportunity for researchers to explore crime detection using machine learning and deep learning techniques. This article examines recent breakthroughs to highlight the current trends in machine learning and deep learning for crime prediction. It emphasizes their roles in uncovering criminal activities, forecasting crime patterns, and preventing offenses, with the goal of providing a holistic perspective on this evolving field and inspiring future research [1].

According to a survey, approximately 218 million individuals are affected by both natural and human-induced disasters each year, resulting in around 68,000 annual casualties. While the occurrence of natural disasters like earthquakes and volcanic eruptions has exhibited relatively stable patterns, there has been a noticeable increase in crime-related incidents during the same period [4].

Machine learning and deep learning algorithms have played a pivotal role in predicting crime patterns by analyzing crime data. Algorithms such as decision trees, random forests, and support vector machines have been trained on city-specific crime data to accurately discern crime trends. These algorithms not only predict crime patterns but also provide insights into crime trends, facilitating efficient resource allocation for crime control. Moreover, they can uncover correlations between crime incidents and environmental or demographic factors, enabling tailored crime prediction strategies. Predictive policing is another valuable application of machine learning. By analyzing crime data from specific areas, machine learning can identify crime hotspots and predict future incidents. This knowledge guides law enforcement agencies in optimizing resource deployment, thereby enhancing their effectiveness [5].

This project aims to explore the Global Terrorism Database (GTD) to gain a deeper understanding of terrorism patterns over the past decades. The objective is to reveal potential trends in attack frequency, the most targeted countries, preferred methods of attack, types of weapons used, casualties, and the evolution of terrorist organizations, among other relevant dimensions [7].

Through this analysis, the project intends to provide a comprehensive overview of global terrorism trends, which can inform counter-terrorism strategies and policies. Additionally, the findings may contribute to understanding the characteristics of regions prone to attacks and the underlying reasons for their vulnerability [10].

In summary, this project offers a data-driven exploration of the intricate realm of terrorism, with the goal of shedding light on the complex patterns hidden within the vast GTD. The final outcome of this project will consist of a collection of valuable insights that have the potential to significantly support ongoing counter-terrorism efforts and guide future research in this field. By combining data manipulation, numerical analysis, and graphical visualization, we anticipate generating a robust and thorough examination of the dataset, leading to substantial key findings related to global terrorism [13]

2.SUMMARY:

K. Venkata Naga Sai and K. Sai Tarun Kumar developed a system to predict high-crime areas, addressing the global issue of crime's impact on quality of life and economic growth. Their approach emphasizes enhancing crime analytics for safer communities.[1]

In this study, Ginger Saltos and Mihaela Cocea investigate predictive models for various crimes in the UK, such as anti-social behavior, utilizing LSOA codes. They employ three distinct algorithms—instance-based learning, regression, and decision trees—to analyze and predict crime patterns effectively.[2]

Shiju Sathyadevan, M.S. Devan, and S. Surya Gangadharan have created a system that predicts high-crime regions and visualizes criminal hotspots. Their research emphasizes the importance of computerized systems and data mining techniques in expediting crime-solving processes for law enforcement.[3]

Tushar Sonawane and Shirin Shaikh emphasize the value of extensive crime data, highlighting its crucial role in analysis and prediction. Their research underscores the significance of crime analysis for law enforcement, revealing hidden patterns and relationships within crimes, facilitated by powerful data mining techniques.[4]

Khushab U and A. Bokde classify crimes based on occurrence frequency across years. They enhance accuracy by assigning feature weights and use Genetic Algorithm to optimize Outlier Detection in RapidMiner.[5]

Benjamin Fredrick David H surveys the use of police and criminal databases for predicting criminal movements. He explores supervised and unsupervised learning techniques, aiming to enhance crime analysis and prediction methods.[6]

Suhong Kim and Param Joshi investigate leveraging police databases for predicting criminal movements, utilizing supervised and unsupervised learning techniques. Their study aims to survey the effective application of these methods in crime analysis and prediction.[7]

Miquel Vaquero Barnadas conducts a study on machine learning-based crime prediction using 15 years of Vancouver's crime data. Employing two unique data-processing approaches, the research achieves a crime prediction accuracy between 39% to 44% for forecasting crime occurrences in Vancouver.[8]

Ricardo Francisco and Sergio Luis utilized the SEMMA (Sample, Explore, Modify, Model, and Assess) model for their analysis. They prepared the dataset through data manipulation, standardization, and cleaning techniques.[9]

Jingyi He and Hao Zheng develop a crime distribution predictive model in urban areas using GAN neural networks. Their research emphasizes how advancements in predictive urban planning can improve security measures and urban design by offering a data-driven approach to optimize city layouts for crime prevention.[10]

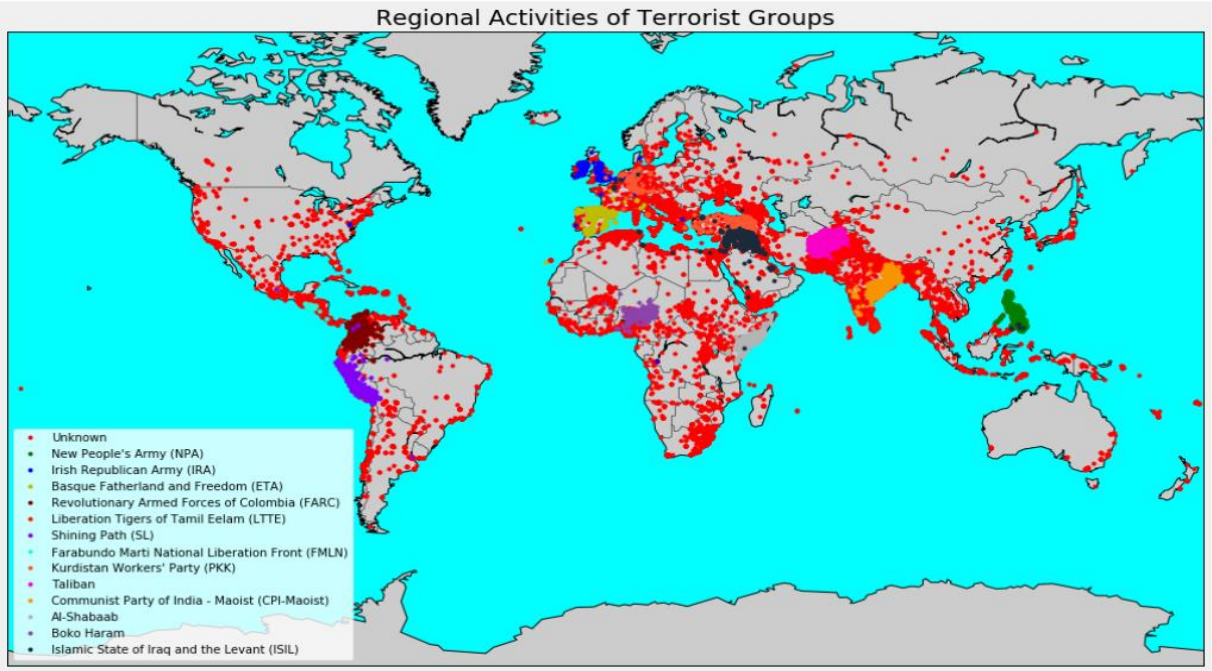
Rajkumar S and Sakkarai Pandi M develop a system predicting high-crime regions and visualizing vulnerable areas. Their research underscores the significance of data mining in extracting insights from unstructured data, aiding in predictions based on existing datasets.[11]

Sapreet Kaur and Dr. William Jeet Singh emphasize the development of a system predicting high-crime regions and visualizing vulnerable areas. Their work highlights the crucial role of data mining in extracting insights from unstructured data, enabling predictions based on existing datasets.[12]

Almaw and Kal Yani Kadam aim to reduce future crimes using historical data, predicting upcoming incidents based on various factors. They address rising crime rates as motivation, offering a detailed study on data mining methods and ensemble classification techniques for precise crime analysis and prediction.[13]

Ruaa Mohammed Saeed offers a concise review of crime prediction algorithms, aiding crime researchers in understanding these techniques for future analyses. Logistic Regression emerges as the most powerful method for crime prediction based on the literature examined.[14]

Xu Zhang and Lin Liu utilized the XGBoost algorithm to train a crime prediction model. They employed the Shapley additive explanation (SHAP) method to assess individual variable contributions, mapping local SHAP values to assist police in prioritizing factors for specific locations.[15]



2.1. LITERATURE SURVEY

AUTHOR	MODEL USED	PARA METERS	MERITS	DEMERITS	FUTURESCOPE
k.ventaka naga sai, k.sai tarun kumar (2021) [1]	KNN Decision tree Random forests	Data-Related Parameters. -Feature Engineering Parameters	Improved Crime Analysis. Quick Identification of Suspects. Evidence collection.	Technologies and data mining methods can raise significant privacy concerns. Between ensuring public safety and infringing on individuals' right to privacy.	Improved Predictive Accuracy. -Real-time Crime Prediction. Integration of Multiple Data Sources.
Ginger Saltos, Mihaela Cocea. (2017) [2]	regression, decision tree. open data.	Model-Specific Parameters. Evaluations	Evidence based decision making Crime Hot Spot Identification Proactive Crime prevention	-Data bias and Accuracy. -Risk of over policing. -Privacy concerns	Enhanced Data Sources. Geo-Spatial Analysis. Temporal Analysis. Network Analysis.
Shiju Sathyade van; M.S. Devan; S. Surya Gangadh aran (2012) [3]	Navie bayes, Apriori algorithm	Crime type, geograp hical location	This proactive approach enables law enforcement to allocate resources strategically and prevent crimes before they happen.	-Misinterpretation of patterns. -Over-reliance on technology. -Ethical consideration.	Analysis and prevention has the potential to greatly enhance law enforcement efforts. Challenges related to privacy, bias, ethics, and technological implementation.
Tushar Sonawanev ,2 Shirin Shaikh (2013) [4]	k-means, cluster, correlation	Visualization sParameters. -Patterns Detections and Analysis	The use of data - mining techniques to analyze and predict crime against women can aid in developing effective strategies for prevention and reduction of such crimes.	Data Quality and Availability. Bias and Fairness. Privacy Concerns Overfitting and Generalization.	Analyze correlate and predict the crimes from huge data set. Results will be in the form of correlation.
Khushabu,A.bokde (2017) [5]	Clustering, k-means	Victim characteristic. Number of clusters. Distance Metric	Enhanced data quality Outlier detection. Optimization through genetic algorithm. Feature weighting.	Model overfitting Algorithm selection Data interpretation.	Multi-model Data Fusion. Dynamic Clustering. Hybrid Models Real-time Crime

Benjamin Fredrick David .H (2017) [6]	Data mining	Data Collections and Preprocessing Parameters. Model Selection and Evaluations	Enhanced investigative techniques. Crime tracking and prevention	Over-relying on data-driven Approaches might lead to overlooking. Importance of human intuition and qualitative insights in crime solving.	Technological Advancements. Ethical Considerations. Collaboration and Data Sharing.
Suhong kim, Param joshi (2018) [7]	KNN Decision tree	Distribution of crimes per-year, month, day.	Efficient crime prediction supports optimized allocation of police resources. Predictive models aid in resource allocation for crime prevention.	Over-reliance on predictive models might overlook contextual nuances.	Urban Planning Integration. Long-Term Impact Evaluation.
Miquel Vaquero Barnadas. (2016)[8]	KNN Neural Network.	Address latitude& Longitude. Resolution	Simplifying Complexity. Problem-Centric Approach. Diverse Algorithm Exploration.	Real-World Variability. No Interdisciplinary Focus. Algorithm Limitation.	Real-world Data Integration. Social Context Exploration
Ricardo Francisco, Sergio Luis, (2021) [9]	KNN Neural Network.	Geographic area. Timing	Comprehensive Data Selection. Following the SEMMA framework, the study employs a systematic approach from data preparation to model assessment.	The model's effectiveness might be limited to Buenos Aires and not universally applicable to other cities. Simplification of Urban Factors.	cross-city adaptability. multivariate analysis.
Jingyi He, Hao Zheng (2021) [10]	-GAN Neural network. SVM	- crime concentration areas	- Visualization Enhancement (Crime hotspot maps.) - Holistic Crime Prediction.	The utilization of GAN neural networks might hinder full transparency and clear interpretation of prediction results.	Multi-City Validation. Predictive Policing Strategies.

Rajkumar.s, sakkarai pandi.M (2018) [11]	SVM	Data Parameters Feature Engineering Parameters Model Selection Parameters. Evaluation Metrics	By,analyzing patterns and trends, data mining can help in predicting potential crime occurrences. Utilizing visualization techniques helps in presenting crime trends and Patterns	Analyzing data from diverse sources might Introduce bias and privacy concerns. Criminal behavior and tactics can evolve over time,	Analyzed and compared different algorithm on crime data determine.
Sapreet kaur, Dr.William jeet singh (2020) [12]	Crime Data Mining Techniques. Systematic Review.	Research Questions and Objectives. Data Sources. Data Mining Techniques Feature Engineering and Selection.	. Early Warning systems Analyzing crime data can involve communities in crime. Prevention efforts, promoting collaboration.	Some crime patterns might be influenced by unpredictable events External factors that are not captured in historical data	The future scope of crime data mining techniques is promising and opens up numerous Integration with emerging technology.
Ayishesh im Almaw,kal yani kadam (2022) [13]	Naïve bayes. J48. Aritifical network.	Environment al Factors. Economic Indicators	Crime data mining can lead to tailored interventions and strategies based on specific crime patterns. The use of various data mining techniques and ensemble learning provides a comprehensive and multi-faceted	As criminal behavior evolves, historical data might not accurately predict new and emerging crime patterns. The use of predictive models might be perceived as invasive, dystopian, or discriminatory by the public	Contextual analysis Interdisciplinary collaboration Ethical frameworks.
Ruaa Mohamme d Saeed (2020) [14]	Regression Model	-Data Collections and Preprocessing Parameters. -Model Selection and Hyperparamet ers.	-Timely Interventions -Holistic approach -Real-time prediction	-Generalization Limitations. -Changing Criminal Tactics	-Hybrid Approaches -Temporal Dynamics -Bias Mitigation
Xu Zhang, Lin Liu (2021) [15]	XG Boost algorithm,	Accuracy Scalability efficiency	Post-hoc Interpretation. Spatial Insight Generation.	Data Quality Dependence. SHAP's post-hoc interpretation might struggle with intricacies in complex models.	Temporal Interpretability. Longitudinal Impact Assessment

3.PROBLEM STATEMENT:

Despite the advancements in machine learning and deep learning techniques for crime prediction and counter-terrorism efforts, there remains a critical need to address the evolving patterns and intricacies within global terrorism incidents. The existing research primarily focuses on crime prediction in specific cities or regions, leaving a gap in understanding the broader global terrorism landscape.

3.1.OBJECTIVE

The project aims to perform an exploratory data analysis (EDA) of the GTD to uncover key trends and patterns in terrorism-related activities. The objective is to visually present these findings, enhancing the understanding of terrorism dynamics.

- Enhanced Crime Prediction Accuracy.
- Optimized Resource Allocation.
- Customized Crime Solutions.
- Reducing Crime Impact.
- Ethical Use of Technology.
- Public Awareness and Collaboration.
- Improving Predictive Policing.

The project will explore the following aspects of the GTD:

- Attack frequency
- Most targeted countries
- Preferred methods of attack
- Types of weapons used
- Casualties
- Evolution of terrorist organizations
- Data Visualization

4.METHODOLOGY:

The GTD dataset is intricate, making comprehension and analysis challenging without a defined methodology. We propose a systematic approach, involving four key steps, to classify and predict terrorist organizations using ensemble learning.

Existing methodologies: There are a number of existing methodologies for analyzing terrorism data. These methodologies vary in their focus, scope, and complexity.

STEPS

1)Data Preprocessing:

Raw data undergoes cleaning, involving outlier removal, missing value imputation, data type conversion, and scaling to prepare it for modeling.

2)Data Splitting:

The preprocessed data is divided into training and testing sets. The training set educates the machine learning model, while the testing set evaluates its performance on unseen data.

3)Model Training:

Various machine learning models, including decision trees, random forests, XG Boost, bagging, and extra trees, are trained on the dataset.

4)Prediction and Evaluation:

The trained model predicts outcomes on the testing set, evaluated using metrics like accuracy, precision, recall, and F1 score to measure its performance.

5)Final Prediction:

The most effective model is chosen and applied to new data for making accurate predictions.

FORMULAS:

Accuracy:

Accuracy = (True Positives + True Negatives) / (Total Predictions)

Precision:

Precision = True Positives / (True Positives + False Positives)

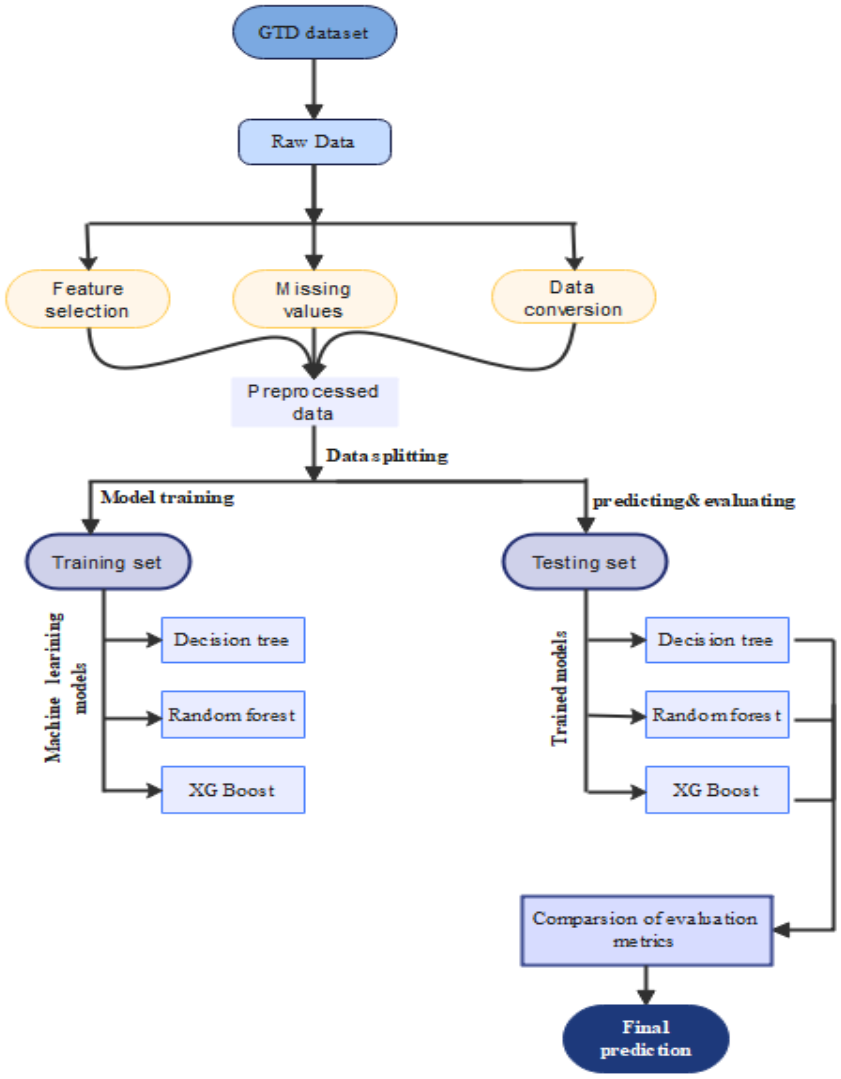
Recall:

Recall = True Positives / (True Positives + False Negatives)

F1 Score:

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

4.1ARCHITECTURE:



4.2.Algorithm:

XG Boost Algorithm

XG Boost (Extreme Gradient Boosting) is a machine learning algorithm that is commonly used for classification and regression tasks. It is a type of ensemble learning algorithm, which means that it combines the predictions of multiple weak learners to produce a more accurate prediction.

XGBoost is particularly well-suited for crime prediction because it can handle large and complex datasets. It is also able to learn from both numerical and categorical features.

F(x) = \sum_{i=1}^n fi(x)

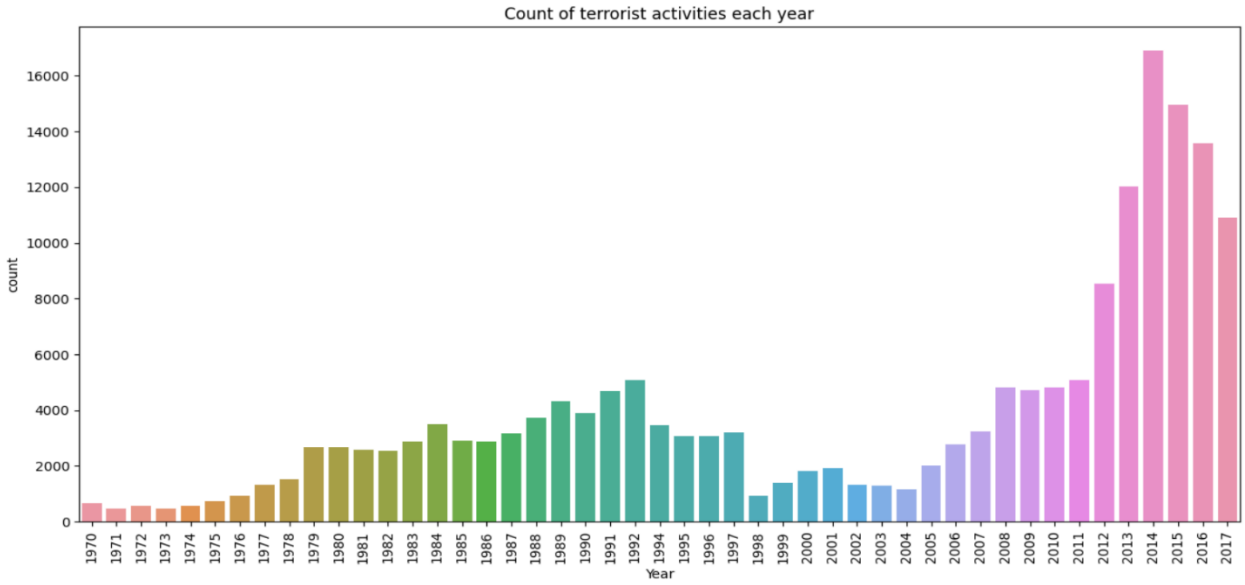
Where,

F(x) is the final prediction of the model.

fi(x) is the prediction of the ith weak learner.

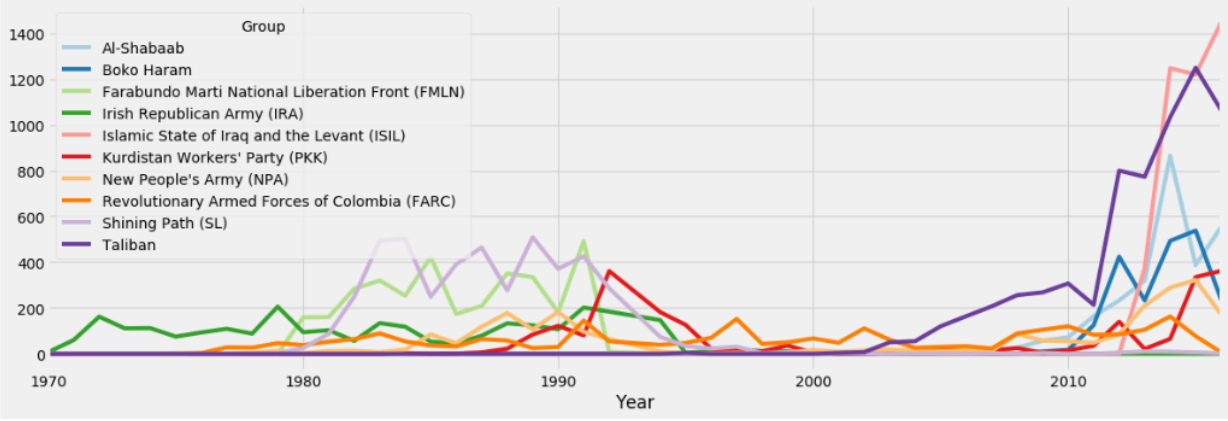
n is the number of weak learners.

5.CODING

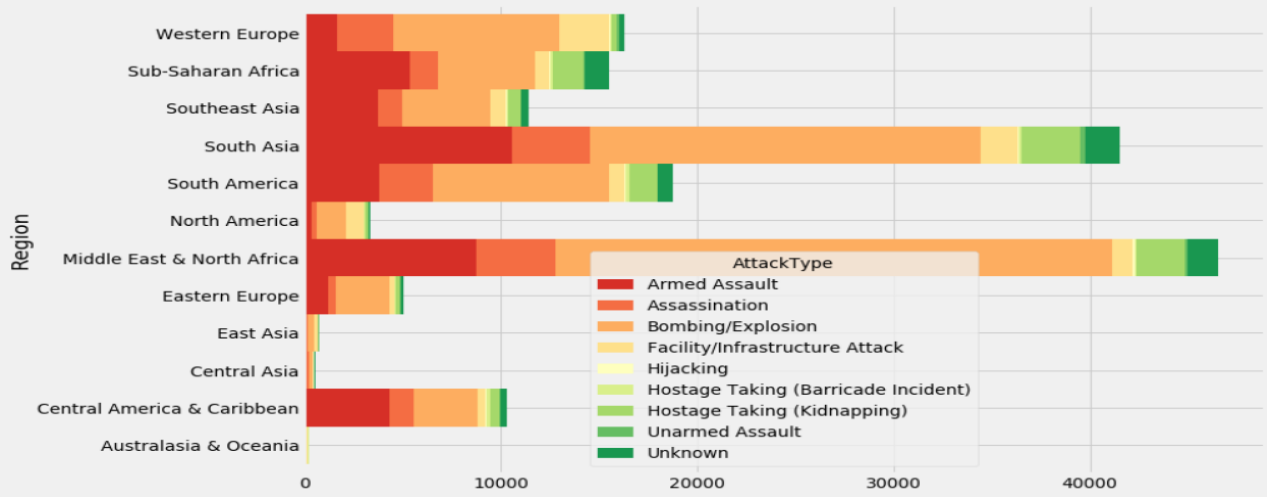


The insights obtained from analyzing the trend of terrorist activities over the years are invaluable for assessing their frequency and patterns. This information is vital for predicting future trends and assisting law enforcement and security agencies in resource allocation and strategic planning.

However, if the analysis reveals a consistent increase in terrorist activities, it could have adverse effects on various sectors. A growing trend in attacks signifies a worsening problem, potentially leading to negative consequences for businesses, public safety, and overall societal well-being. This negative impact may arise due to increased security concerns, higher insurance costs, and decreased consumer confidence, all of which can hamper economic growth and stability.



The Irish Republican Army(IRA), is the oldest terrorist group started back in the 1960-1970, maybe after the World War 2 due to the mass killing. However, it has probably stopped its activities in the late 90's. Some of the groups that have started lately in 2000's like the ISIL and Taliban, have shown a shoot in the number of attacks in the past years.



5.1.DATASETS DESCRIPTION:

The Global Terrorism Database (GTD) is a freely accessible database managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. Covering the period from 1970 to 2017, this open-source database offers extensive information on both domestic and international terrorist incidents worldwide, totaling over 180,000 recorded attacks.

The GTD includes over 200 data fields, including the following:

- Date of attack
- Location of attack
- Target of attack
- Perpetrator of attack
- Weapons used
- Casualties (deaths, injuries, kidnappings)
- Monetary damage
- Political, economic, religious, or social motivation Context of attack (e.g., conflict, protest, crime)

Data for the GTD is gathered from diverse sources like news reports, academic studies, and government records. Experts at START meticulously code the information to ensure accuracy and consistency. However, it's important to note that the GTD has its limitations. Since it relies on news reports, some attacks might not be included, and data completeness and accuracy can vary across different countries.

Dataset

Global Terrorism Database (GTD). (n.d.). University of Maryland.[16]

5.2.RESULT

XGBoost has been shown to be effective for crime prediction in a number of studies.
In general, XGBoost has been shown to be a promising algorithm for crime prediction. It is able to handle large and complex datasets, and it can learn from both numerical and categorical features. However, it is important to note that the performance of XGBoost will depend on the quality of the data that is used to train the model.

ACCURACY: 0.898456204078263

	Precision	**recall**	**F1-score**	**support**
0	0.64	0.17	0.26	3978
1	0.91	0.99	0.95	32361
Accuracy			**0.90**	**36339**
Macro avg	**0.77**	**0.58**	**0.61**	**36339**
Weighted avg	**0.88**	**0.90**	**0.87**	**36339**

6.CONCLUSION

The Exploratory Data Analysis (EDA) performed on the Global Terrorism Dataset offered valuable insights into global terrorism trends from 1970 to 2017. Leveraging Python libraries like Pandas, Matplotlib, Seaborn, and NumPy, we efficiently processed and visualized complex data related to terrorist activities.
The process highlighted the importance of data-driven decision-making. EDA allowed us to translate raw data into meaningful insights. For example, understanding that certain regions face a higher risk of terrorist attacks and recognizing active terrorist groups empowers security agencies and policymakers to allocate resources more effectively. This targeted approach has the potential to save lives and protect property.
However, it's crucial to acknowledge that combating terrorism involves more than just historical data analysis. It demands a holistic strategy that incorporates real-time intelligence, geopolitical factors, and ground-level realities.
In conclusion, this project showcases the power of data analysis in informing and guiding counter-terrorism initiatives. It serves as a valuable starting point for further research and action, emphasizing the continuous need for data collection, analysis, and interpretation in addressing global security challenges such as terrorism.

6.1.FUTURE SCOPE

Analyzing the Global Terrorism Dataset opens doors to future advancements in counter-terrorism. Predictive analytics, social network analysis, and integrating sentiment with geospatial data offer avenues for targeted interventions and understanding public sentiment. Exploring the intersection of cybersecurity threats and terrorism is crucial in the digital age. Additionally, assessing real-time policy impact and fostering international collaborations are vital. Upholding ethical standards in data-driven approaches is paramount. Embracing these interdisciplinary strategies is key to effective, evidence-based counter-terrorism efforts.

7.REFERENCE

[1] Sathyabama Institute of Science and Technology (Deemed to be University). (2022, March). k.ventaka naga sai. (2021). Crime prediction and analysis Using machine learning.

[2] Saltos, G., & Cocea, M. (2017). An exploration of crime prediction using data mining on open data. International Journal of Information Technology and Decision Making, 16(05), 1155–1181.

[3] First International Conference on Networks & Soft Computing (ICNSC), 2014. (2012). Shiju SAthyadevan, Devan M.S, 2014. Crime Analysi s and predicti on using data minning.

[4] Tushar Sonawanev. (2014). Crime Pattern Analysis, Visualization And Prediction Using Data Mining.

[5] Wadhai, C. G., Kakade, D. S. T. T. P., Bokde, K. A., & Bhattacharya, D. (2018). Crime detection technique using data mining and K-Means. International Journal of Engineering Research and Technology, V7(02)

[6] David, H. B. F., & Suruliandi, A. (2017). SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES. ICTACT Journal on Soft Computing, 7(3), 1459–1466.

[7] Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018). Crime Analysis Through Machine Learning. IEEE.

[8] Miquel Vaquero Barnadas. (2016). MACHINE LEARNING APPLIED TO CRIME PREDICTION.

[9] Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vazquez, J., & Rodriguez, M. L. (2021). Applied Machine Learning in Social Sciences: Neural Networks and Crime Prediction. Social Sciences, 10(1), 4.

[10] Sapreet kaur, Dr.Williamje et singh. (2015). Systematic Review of crime data mining. http://www.ijarcs.info/index.php/Ijarcs/article/view/3607.

[11] Ayishesh im Almaw,kalya ni kadam 2018. (2018). Survey paper on crime presicition using ensemble approach.

[12] Ruaa Mohammed Saeed. (2020). A study on predicting crime rates through machine learning and data mining.

[13] Xu Zhang,Lin Liu. (2022). Interpretable machine learning models for crime prediction.

[14] Forradellas, R. F. R., Alonso, S. L. N., Jorge-Vázquez, J., & Rodriguez, M. L. (2020c). Applied Machine Learning in Social Sciences: Neural networks and crime prediction. Social Sciences, 10(1), 4.

[15] Jingyi He, Hao Zheng 2021. (2021). Prediction of crime rate in urban neighborhood s based on machine learning.

[16] Dataset https://www.start.umd.edu/gtd/.