future internet

MDPI

Article

Cybersecurity Intelligence Through Textual Data Analysis: A Framework Using Machine Learning and Terrorism Datasets

Mohammed Salem Atoum [1,*], Ala Abdulsalam Alarood [2], Eesa Alsolami [2], Adamu Abubakar [3], Ahmad K. Al Hwaitat [1] and Izzat Alsmadi [4,5]

[1] Department of Computer Science, The University of Jordan, Amman 11942, Jordan; a.hwaitat@ju.edu.jo
[2] College of Computer Science and Engineering, University of Jeddah, Jeddah 21959, Saudi Arabia; aasoleman@uj.edu.sa (A.A.A.); eaalsulami@uj.edu.sa (E.A.)
[3] Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia; adamu@iium.edu.my
[4] Department of Computing, Engineering and Mathematical Sciences, Texas A&M University, San Antonio, TX 78224, USA; ialsmadi@tamusa.edu
[5] Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan
* Correspondence: m.atoum@ju.edu.jo

Abstract: This study examines multi-lexical data sources, utilizing an extracted dataset from an open-source corpus and the Global Terrorism Datasets (GTDs), to predict lexical patterns that are directly linked to terrorism. This is essential as specific patterns within a textual context can facilitate the identification of terrorism-related content. The research methodology focuses on generating a corpus from various published works and extracting texts pertinent to "terrorism". Afterwards, we extract additional lexical contexts of GTDs that directly relate to terrorism. The integration of multi-lexical data sources generates lexical patterns linked to terrorism. Machine learning models were used to train the dataset. We conducted two primary experiments and analyzed the results. The analysis of data obtained from open sources reveals that while the Extra Trees model achieved the highest accuracy at 94.31%, the XGBoost model demonstrated superior overall performance with a higher recall (81.32%) and F1-Score (83.06%) after tuning, indicating a better balance between sensitivity and precision. Similarly, on the GTD dataset, XGBoost consistently outperformed other models in recall and the F1-score, making it a more suitable candidate for tasks where minimizing false negatives is critical. This implies that we can establish a specific co-occurrence and context within the terrorism dataset from multiple lexical data sources in effectively identifying certain multi-lexical patterns such as "Suicide Attack/Casualty", "Civilians/Victims", and "Hostage Taking/Abduction" across various applications or contexts. This will facilitate the development of a framework for understanding the lexical patterns associated with terrorism.

Keywords: cyber intelligence; terrorism; machine learning

Academic Editor: Gianluigi Ferrari

Received: 24 February 2025
Revised: 14 April 2025
Accepted: 16 April 2025
Published: 21 April 2025

Citation: Atoum, M.S.; Alarood, A.A.; Alsolami, E.; Abubakar, A.; Hwaitat, A.K.A.; Alsmadi, I. Cybersecurity Intelligence Through Textual Data Analysis: A Framework Using Machine Learning and Terrorism Datasets. *Future Internet* **2025**, *17*, 182. https://doi.org/10.3390/fi17040182

1. Introduction

According to the traditional integrative framework, language consists at least of vocabulary, syntax, and semantics [1]. During the past 50 years, most of the generative work in corpus linguistics has been devoted to studying the relation between vocabulary and syntax; semantics has been invoked only at the semantic/syntactic interface, traditionally via rules such as passive and nominalization [2]. It is, therefore, interesting to understand a new mode of rich lexical correspondence that has rarely been modeled explicitly in corpus linguistics work. Extracting lexical patterns is a difficult task and one that has rarely

been addressed. While existing work endeavors to resolve pronouns, word hypernymy, or phrasal verbs, some even attempted to address the full extent of fine-grained lexical correspondences necessary for a truly predictive model [3].

Text classification is one of the important tasks in Natural Language Processing (NLP), which deals with a variety of genres such as categorizing news articles, deciding spam or non-spam emails, identifying the sentiment orientation of product reviews, determining the genre of a movie based on its plot summary, and predicting the developmental level or literary style of free-written children's essays [1,2]. For predictive modeling with target labels such as news topics, spam, sentiment, and movie genres, the features used in the study are the word unigrams and bigrams between lexical items in a sentence or document [4]. Additionally, bag-of-words are used, and the document is represented through histogram-like statistics of lexical items—in other words, frequency [5]. The target labels are generated or assessed by human coders, and inter-coder evaluations are applied to the target labels based on the evaluations of human coders who judge each target label [1].

The paper's core objective is to leverage textual data analysis to identify and predict terrorism-related lexical patterns. This has a direct connection to the internet, as much of the data analyzed (e.g., social media posts, news articles, blogs) are sourced from online platforms. The internet serves as a vast repository of unstructured textual data, which are increasingly being used by terrorist organizations for communication, recruitment, and propaganda. By analyzing these data, the framework proposed in the paper can help in the early detection of potential threats, thereby enhancing cybersecurity intelligence.

The internet is the primary medium through which terrorist organizations disseminate information, communicate, and recruit members. The paper's framework can be applied to monitor and analyze online content in real-time, providing a proactive approach to counterterrorism. This is particularly relevant in the context of the Future Internet, where the volume of data generated online is expected to grow exponentially, necessitating advanced tools for data analysis and threat detection.

The framework can be integrated into cybersecurity systems to monitor and analyze online communications for signs of radicalization or terrorist activity. This is crucial for preventing cyberattacks that may be orchestrated by terrorist groups, such as hacking, data breaches, or the spread of malicious content.

Owing to the great success of distributed word representation, distributed word representations along with other relevant features were used as multiple types of word embedding and were shown to positively contribute to several NLP tasks such as sentiment classification, rhetorical analysis, and predicting the education status of children's essays to word embedding algorithms [6,7].

Among the studies associated with terrorism, the structure and dynamics of alliances among terrorist organizations have been emphasized, utilizing complex network theory to analyze the relationships and cooperation patterns between these groups [8]. This pertains to the evolution of networks and the identification of influential groups within them. Complex network theory facilitates the comprehension of terrorist alliance networks derived from empirical evidence. Hu et al. [9] identified the pivotal entities inside terrorist alliance networks, whose elimination or disruption would significantly affect the network's structure and performance. Certain terrorist organizations can be recognized as pivotal nodes within the network, signifying their role as main connectors or hubs within the alliance. The historical progression of terrorism into specific phases, employing a time series-based methodology for complicated network analysis, is essential [10]. This will facilitate an understanding of the evolution of terrorist networks and methods across time,

offering insights into the adaptive characteristics of these groups and delineating several epochs of terrorism.

Considering that terrorism is of interest to researchers, and given that its predictions have been far from accurate, coupled with the frustration of slow and effective responses, it is often ultimately characterized by 'boots on the ground' [11]. This frustration is particularly relevant in the 'information age'. That is why this current study establishes social connections and online behavior that would be able to harness some predictive capability. The justification of this dwells on the fact that studies on terrorism emanating from social and cognitive psychology and on multi-agent killing and other intercultural conflicts have also shown that the individual and situational determinants of behavior are associated with information and communication technology. To make the discriminating elements of this evident for the development of predictive models that look solely at terrorism is an important antecedent. Language is an important mark and record of human cognition. In interventions, the use of language has a marked impact on changing behavior [12]. Thus, there has been a move to data mine the web as a source of predictive capability, and the same applies to counterterrorism. That is why this paper contributes in the following ways:

- The research demonstrated that co-occurrence is evident, as the "lexical patterns" anticipated within any textual context exhibit similarities. Subsequently, it developed a synthesis of multiple lexical data sources related to terrorism and generated various multi-lexical data sources to evaluate their applicability.
- The research suggests that the Extra Trees model achieved the highest predictive accuracy for lexical patterns in the context of terrorism-related text prediction. This was the case when the model was implemented. This suggests that there is the possibility of establishing a different co-occurrence and context inside the terrorist dataset that is produced from a variety of lexical data sources, which would make it easier to identify particular multi-lexical elements.
- The research contributes to various applications by establishing a framework for comprehending the lexical patterns linked to terrorism from a law enforcement viewpoint. Early detection facilitates a strategy to undermine individuals or groups conversationally and employ 'soft' measures to diminish their potential for radicalization.
- The Future Internet is expected to be characterized by the proliferation of connected devices, the Internet of Things (IoT), and the generation of massive amounts of data. The paper's framework aligns with the vision of the Future Internet by providing a predictive analytics tool that can handle large-scale data and generate actionable insights. Specifically, the paper's use of machine learning models for predictive analytics is highly relevant to the Future Internet, where the ability to predict and prevent threats will be crucial. The framework can be extended to analyze data from IoT devices, social networks, and other connected systems, providing a comprehensive approach to threat detection and prevention. Similarly, the Future Internet will see the emergence of new technologies such as 5G, edge computing, and artificial intelligence (AI). The paper's framework can be adapted to leverage these technologies, enabling faster and more efficient analysis of textual data. For example, edge computing can be used to process data locally, reducing latency and improving the speed of threat detection.

This paper consists of six sections, structured as follows: This current section serves as the introduction and summary, detailing the research problem, study aims, contributions, and justification. Section 2 presents the previous related research. Section 3 presents the research methodology. Section 4 presents the discourse and postulations regarding the findings. Section 5 finishes the research. This document consists of seven sections, structured as follows: This section serves as the introduction and summary, detailing the

research problem, study aims, contributions, and rationale. Section 2 delineates the research technique. Section 3 delineates the experimental findings. Section 4 delineates the discourse and premises of the findings. Section 5 finishes the research.

2. Related Work

Studies on Terrorism and Linguistics Research on terrorism are comprehensive, spanning all domains of social sciences and encompassing interdisciplinary methodologies [13]. Numerous studies have investigated the methods by which terrorist organizations convey messages to their adherents and to the broader population. The arrangement of words may be significant in certain settings, while their co-occurrence may hold greater influence in others [14]. A keywords summary involves linking each document to a compilation of semantically significant and relevant terms derived from its content [15]. In terrorism research, controversy associated with defining terminology creates impediments to conducting analysis. The accurate use of terms is of the highest importance, as it might otherwise lead to incorrect analyses and incorrect conclusions. In recent years, special attention has been paid to examining terrorism-related terms and to studying their variety. In many cases, the definition of a term connected with terrorism can be influenced by research or a certain experience [16]. As a result, the definition largely depends on an assessment of the psychological, emotional, and/or subjective consequences anticipated by persons or parties utilizing a term associated with terrorism within a political or social context. A relevant assessment today is vital because of the radical changes that have occurred in the political world.

The criterion for extracting these phrases may rely on many properties, including their frequent occurrence within the document's content or their association with the document's classification into specified categories [17]. This framework can be seen as a document summary that aids in text organization and facilitates activities such as automatic document classification, text categorization, document clustering, and information retrieval [18]. It may also aid in analyzing the generated corpus, yielding more dependable findings regarding the subjects and the target audience [19].

Jin et al. [19] presented a feature selection methodology for text categorization utilizing the absolute deviation factor. The paper employs this strategy to analyze the significance of particular elements in text data, quantifying variances to pinpoint features that substantially enhance classification accuracy. The strategy effectively improved classification performance by eliminating irrelevant or low-impact features, illustrating that absolute deviation factor-based feature selection can augment the accuracy and efficiency of text classification models.

Song et al. [20] examined the correlation between gold prices and terrorism, assessing whether gold functions as a safe-haven asset for risk aversion or as a financial resource for financing terrorist actions. The authors employ econometric analysis to investigate the influence of terrorist attacks on gold price volatility. The results indicate that gold serves as a risk-averse investment in times of uncertainty and may also function as a funding source for terrorist organizations. Terrorist acts were associated with transient surges in gold prices, suggesting that gold serves a dual function in relation to terrorism.

Xiong et al. [21] introduced the XRR model, a technique for extreme multi-label text classification. The technique integrates candidate retrieval (finding pertinent labels for a specific text) with deep ranking (arranging these labels according to relevance), utilizing deep learning to handle extensive quantities of potential classifications. The XRR model markedly enhanced classification accuracy for datasets with large label sets, surpassing conventional methods. This strategy proved very efficacious for applications with texts linked to several nuanced labels, such as recommendation systems and document categorization.

Chuang et al. [22] utilized spatial and temporal analysis to investigate the impact of alliances and rivalries among terrorist organizations, including al-Qaeda and ISIS, on the frequency and geographical distribution of attacks. The authors utilize statistical modeling to discern trends in near-repeat terrorist action. The research revealed that regional alliances and rivalries considerably influence patterns of terrorist attacks. Alliances resulted in concentrated assaults in certain areas, whereas rivalry frequently scattered attack sites, underscoring the significance of inter-group dynamics in comprehending patterns of terrorist action.

Song et al. [23] employed time-varying causality analysis to examine the correlation between terrorist incidents and variations in oil prices. The article analyzes data over a designated timeframe, investigating the extent and manner in which terrorist incidents influence oil price volatility. The research revealed a varied correlation, wherein terrorist strikes occasionally induce increases in oil prices. This association varies over time, indicating that although terrorism can affect oil prices, this effect is inconsistent and contingent upon certain geopolitical circumstances.

Tolan and Soliman [24] perform an experimental assessment of multiple categorization systems to evaluate their efficacy in forecasting terrorism-related incidents. They evaluate methods such as decision trees, support vector machines, and neural networks on terrorist datasets. Neural networks exhibited superior performance in terrorist prediction, with greater accuracy than alternative methods. The research indicates that machine learning may serve as an effective instrument for predicting terrorist attacks, facilitating preemptive actions.

Hu et al. [25] employed quantitative research to categorize worldwide terrorist incidents according to attributes such as geographic location, target type, and methods. The authors employ machine learning techniques to categorize and discern trends in attack data. The research developed a classification system that facilitates the systematic analysis of worldwide terrorism tendencies. Research indicated that assaults differ markedly by geography and target, facilitating more focused strategies for terrorism prevention.

Song et al. [26] investigated the potential correlation between Bitcoin prices and terrorist acts, employing econometric models to assess the influence of terrorism on cryptocurrency price volatility. The study revealed that terrorist acts may induce transient fluctuations in Bitcoin prices, albeit inconsistently. The data indicate a tenuous although discernible connection, as Bitcoin is sporadically employed in the funding of illicit activities, including terrorism.

Despite notable advancements in the application of machine learning for terrorism prediction, significant challenges persist. Prior studies, such as those by Tolan and Soliman [24], demonstrated the efficacy of neural networks and decision trees in identifying terrorism-related incidents, but they often relied on static datasets and failed to account for lexical nuances in open-source textual data. Hu et al. [25] emphasized geographic and categorical trends, yet their models were not optimized for dynamic, real-time inference or textual diversity. Moreover, while multi-label classification models like XRR [21] have shown promise in handling large-scale label sets, they are often tailored to structured domains like recommendation systems and may not generalize effectively to semantically ambiguous, sparse terrorism-related texts.

Our study addresses these gaps by proposing a framework that integrates lexical extraction with real-time capable machine learning models. Specifically, we focus on term co-occurrence, TF-IDF weighting, and model fine-tuning to improve precision and recall in terrorism-related term identification—thus responding directly to the limitations identified in earlier works. Furthermore, by evaluating our models on both a controlled (GTD) and a

loosely structured open-source dataset, we offer comparative insights that strengthen the applicability of our findings across varied text environments.

3. Research Methodology

This research utilized the formal machine learning process from accessing, preprocessing, and training the dataset with the selected models. These models were selected based on their performance in text and prediction problems. The paper's methodology involves the integration of multiple lexical data sources, including open-source data and the Global Terrorism Database (GTD). This integration is facilitated by networking technologies that enable the collection, storage, and processing of large datasets from diverse sources. The framework relies on networked systems to gather data from various online platforms, including social media, news websites, and government databases. The ability to process and analyze these data in real-time requires robust networking infrastructure, which is a key component of the Future Internet. The paper's approach can be seen as a networked intelligence system, where data from multiple nodes (e.g., social media platforms, news outlets) are aggregated and analyzed to generate actionable insights.

3.1. Machine Learning Algorithms for Predictive Modeling

Machine learning algorithms are methods that enable intelligent systems to learn about the world through the available data. In the context of supervised learning, these algorithms build a model to identify associations between data and the outcomes of interest [27]. A crucial stage in the deployment of supervised learning methods is the model selection stage. Model selection is the process of choosing a predictive model based on the hypothesis space defined by the applied algorithm.

Model selection is very important because the most effective algorithm can create overfitting challenges, which result in an overly specific pattern within the training data that will not be effective with new data. In the context of arbitrary or blind model application, fourteen learning methods were considered without any evaluation methods, which may result in an over-optimistic assessment of a model [28]. In the context of arithmetic metal node count classifications, six machine learning applications were reported in renewable energy, eight in combinatory chemistry, and ten in mixed-mode economic research activities that used different supervised learning algorithms. There are different considerations to be made when choosing an appropriate learning algorithm, such as dimensionality reduction, purpose, and intended scope; scalability; interpretability; sensitivity and specificity; and accuracy [29].

3.1.1. Decision Tree (DT)

In the decision trees, each internal node represents a feature. Each branch falling out of that node is a separate value of that feature, and each leaf node represents a class label. All paths from the root to the leaf nodes lead to classification rules [30]. Predictive modeling is where the model decides depending on the data observation. It is dependent on the features of variable values. In decision trees, classification is not only done, but rules are also formed according to feature variables and their values.

The use of the decision tree classifier model is very useful as it shows the individual rules based on some conditions formed by the decision tree. Hence, feature variables are split with their conditional rules based on their values [31]. In these models, and hence on decision trees, the target class label probabilities are calculated. The careful selection of some feature values, especially in public and mutually generated datasets, means that the conditional rules for public and generated datasets relating to these feature variable values might be split such that a rule-based decision tree classification is possible. For

classification, this decision tree is simple and easy to interpret as well. Not only is the purpose of classification to predict class labels, but also a method is used to analyze the relationship between the most influential feature variables [32]. When we build decision trees based on the most influential features, it is important to result in a hierarchy of the most influential features and the relationships among them.

In our implementation, the underlying base estimator for the Bagging (Bootstrap Aggregating) model was the Decision Tree classifier. This follows the standard practice of using decision trees as the weak learners in Bagging ensembles to reduce model variance. Specifically, we employed unpruned decision trees with a limited max_depth (tuned during model optimization) to serve as the base learners. The Bagging ensemble aggregates these trees using majority voting for classification. This setup was chosen to directly compare the performance gains of ensemble learning (Bagging) over individual decision tree models within the same experimental framework.

3.1.2. Bootstrap Aggregating (BA)

Bootstrap Aggregating is a bootstrapping method and a resampling technique that can improve the accuracy of classifiers. In this technique, a few new datasets are constructed by repeated sampling from the original dataset with the method of replacement [33]. Decision trees serve as the base model for BA because they often have low bias and a relatively high variance of output prediction. Independent models are fit from the bootstrap samples over a dataset; thus, the corollary to hold is that they have uncorrelated errors. Eventually, the models are aggregated by averaging as regression or voting operation as classification. Bootstrapping can also be applied to resamples across models, meaning these models are different from each other. Because of the different models, the correlations of these models collapse [34]. In the end, ensemble models fit stable predictions. The advantages of bootstrapping include great simplicity and the ability to improve the precision of any type of statistical technique. However, the accuracy might deteriorate if the sample size is too large, leading to overfitting, or if the base classifier has poor performance. Although bootstrapping is a computationally expensive method, it guarantees improvement in generalization capability. The independence of the samples has also been shown to be effective in improving classification performance. In our work, a BA-based voting system was designed to fit the model and predict the user's dependency style.

3.1.3. XGBoost (XG)

XGBoost is an algorithm that has recently been a prime focus in competitive machine learning circles. It is a scalable and accurate implementation, capable of effectively handling large-scale structured and unstructured data [35]. The algorithm has been utilized in a wide range of nature and complexity of problems, such as credit risk prediction models, bioinformatics solution models, large-scale search engines, or ranking problems across competitors in a competition [36]. XGBoost has been regarded as the go-to algorithm by many practitioners due to its speed and adaptability towards distributed and out-of-core computing. The algorithm employs a technique called gradient boosting and incorporates several novel and effective features that make it stand out [37]. XGBoost utilizes a decision tree model and performs a weighted iteration of these models to come up with a strong classification model. While generally used in binary classification problems, the algorithm also extends to solve multi-class classification problems as well as regression problems [38].

3.1.4. Random Forest (RF)

Random forests or random decision forests are an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or

mean prediction of the individual trees [39]. A random forest is a meta-estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve predictive accuracy and control overfitting [40]. The sub-sample size is controlled with a hyperparameter. The following are the principles for the random forest model: In random forests, each tree in the ensemble is built from a sample drawn with replacement from the training set. Furthermore, when splitting each node during the construction of a tree, the best split is found either from all input features or a random subset of size, where is the number of features [41].

Furthermore, for classification, the number of features considered for splitting each node is, in many implementations, approximately the square root of the total number of features, but this number is a free parameter. Bootstrapping leads to better model performance because when large and strong models are built, the errors made by the model become uncorrelated [42]. Uncorrelated models can efficiently reduce the total error. An entire forest can be visualized through the model counts.

3.1.5. Extra Trees (ETs)

The training of geometric dependent variables concatenated with lexical threshold confidence measures and geometric dependent variables concatenated with lexical threshold confidence measures and lexical profile features based on random sampling with replacement is also tested using the Extra Trees model [43]. It is observed that Extra Trees can handle high-dimensional data very well. They work well even in small datasets, and they are very robust over filtering or pre-processing in general, especially in special populations that have specific information. Their performance is like that of random forests, but they are different in the way that they get the random subsets of the features only, where they use this subset of features for splitting [44]. In Extra Trees, we take all the samples, and it randomly selects a feature split for the decision tree nodes. It may lead to overfitting, but it could also lead to the model being more robust as we would not have access to the underlying data generating mechanism or the important features related to our generated dataset.

The Extremely Randomized Trees (Extra Trees) algorithm is an ensemble learning method that extends the principles of random forests by injecting additional randomness into the decision-making process. Like random forests, Extra Trees aggregates the outputs of multiple unpruned decision trees built on different random subsets of the data. However, unlike random forests, where optimal split points are determined based on a criterion like Gini impurity or entropy, Extra Trees selects split thresholds at random for each candidate feature. This added stochasticity reduces variance even further while maintaining low bias, making the model more robust to overfitting—particularly beneficial in high-dimensional, sparse, and noisy data such as text.

In the context of terrorism prediction, where textual inputs from open sources often contain semantically overlapping terms and an inconsistent structure, Extra Trees is especially advantageous. First, it naturally supports feature selection through its randomized splitting mechanism, which is beneficial when dealing with large TF-IDF matrices. Second, its computational efficiency and parallelizability make it suitable for processing sizeable corpora with high throughput. Third, Extra Trees is resilient to outliers and performs well even when class labels are imbalanced, which is common in datasets related to terrorism, where positive cases (relevant terms or attacks) are much fewer than negative ones (irrelevant or neutral terms).

In our study, Extra Trees demonstrated strong predictive performance, particularly in terms of precision and accuracy, indicating its ability to classify non-relevant terms conservatively while maintaining a reliable identification of key terrorism-related lexical patterns.

While this study employs the standard Extra Trees algorithm, we adapted it through domain-aware preprocessing and task-specific hyperparameter tuning to optimize performance in terrorism-related term prediction. The tuning process accounted for the unique lexical characteristics of terrorism data, including high-dimensional TF-IDF vectors, imbalanced class distributions, and contextual ambiguity. Although no structural modification was applied to the base algorithm, the model was integrated into a preprocessing pipeline that effectively enhanced its ability to distinguish between relevant and irrelevant lexical signals. We acknowledge that future work could explore algorithmic improvements tailored specifically for terrorism detection. These may include hybrid approaches combining Extra Trees with deep semantic models (e.g., word embeddings or transformers), or incorporating cost-sensitive learning to further mitigate the effects of false negatives. Such integrations may further improve the model's suitability for operational use in cyber-intelligence environments.

3.2. Dataset

Two categories of datasets are utilized for this study: the dataset from terms associated with terrorism in articles related to terrorism from open data sources and the Global Terrorism Database dataset.

3.2.1. Extraction of Terms Associated with Terrorism from Open Source

It was established that for the purpose of this study, terms that are related to terrorism and that have been published in research articles that are freely accessible were considered as belonging to some kind of category. Table 1 shows a group of terms that have appeared most frequently in texts related to articles on terrorism.

Table 1. Terms associated to terrorism in open data source.

Category	Terms
Terrorist Activities	terrorism, extremist activity, terrorist, weapon, threat, act, crime, violence
Counter Terrorism Operations	security, operation, intelligence, raid, surveillance, intercept, detain, monitor
Persons	leader, suspect, operative, member, official, agent, recruit
Places	base, hideout, territory, region, hotspot, headquarters, zone, checkpoint

Analyzing these terms, this research observed that they can be grouped into one of four categories: "Terrorist Activities", "Counter Terrorism Operations", "Persons", and "Places" associated with such activities. These terms are significantly different from the others in this group, and there are special reasons for that. The research considers that the terms of this first group are described by a description or by a single word: terrorism, extremist activity, terrorist, weapon, threat, act, crime, security, etc. The list of terms from the other three groups is also provided. All these terms are going to be used in the extraction of terms to detect acts of terrorism in printed sources.

The data in those articles were further assessed to compute the Term Frequency-Inverse Document Frequency (TF-IDF) weighting. This approach constructs a numerical statistic designed to indicate the significance of a word within a document. TF-IDF primarily serves as a weighting mechanism. Term Frequency (TF) signifies the significance of a term within a specific corpus of documents, typically represented as the normalized frequency of term occurrences in a text. The results were produced to generate a list of the most frequently occurring key terms and range values. The range values are a measure of the importance of the term employed to quantify the frequency of terms. As such, the larger the value the

more frequent a term appears. Table 2 presents an extract of the data identified as the set of terms that appear with reasonable frequency across all sources used in the analysis.

Table 2. Terms and their TF-IDF extracted.

Term	TF-IDF	Number of Document
Extremism	0.042	251
Radicalization	0.037	140
Insurgency	0.033	147
Counterterrorism	0.044	153
Violence	0.039	149
Hostage	0.029	133
Conflict	0.035	146
Militancy	0.032	144
Homeland Security	0.038	139
Cyberterrorism	0.03	137
Bombing	0.045	252
Attack	0.034	142
Terror Cell	0.043	143

A co-occurrence analysis was also conducted to identify those terms that appeared in the discourse related to terrorism in the same sentence. This is an important step in the analysis to highlight those terms that were most closely related or conceptually paired by those in the media, government, and citizen blogs and social media. Both analyses provide substantive insight into the common or frequently occurring terms in everyday usage. The results of this process are important in helping us understand what the notable or meaningful terrorism-related issues and concepts are likely to be in our sample of open data.

3.2.2. Global Terrorism Database Dataset

The Global Terrorism Database is a dataset on terrorist activities that started in 1970. The dataset was obtained from Kaggle [45]. The current version has recorded over 150,000 cases of armed violence committed worldwide during this period. Processes to generate the database rely on using encoded data about the locations, weapons, types of offences (such as facilities or individuals), and unique group or category identifiers similar to those used in traffic datasets. At the core of the database are basic surface features. The dataset includes several pertinent fields related to terrorism terminology, specifically: “attacktype1_txt”, “target1”, “addnotes”, “scite1”, “scite2”, “scite3”, and “dbsource”. The column “attacktype1_txt” classifies the principal method of attack, offering insights into prevalent tactics employed in terrorist activities. The column “target1” outlines the primary objective of the attack, providing context for its intent and frequently correlating with the perpetrator’s goals. The “addnotes” column offers more notes or details regarding the attack that may not be comprehensively represented in other columns. It frequently encompasses information regarding the conditions, reasons, or background that elucidates the incident comprehensively. The columns “scite1”, “scite2”, and “scite3” represent citations or sources from which information regarding the occurrence was acquired. They may encompass citations from news articles, reports, or other sources that validate or expound upon the incident’s particulars. A “dbsource” column exists that designates the source

database or collection utilized to compile the attack information. It can indicate whether the data originated from GTD's primary data collection or external sources, hence assisting in tracking data provenance.

This research extracts the TF-IDF of the terms within the dataset, with selected entries displayed in Table 3. The GTD underscores the significance of these terms in monitoring global terrorism trends, demonstrating the intricate and enduring characteristics of specific attack types and targets across time.

Table 3. Terms in GTD and their TF-IDF.

Term	TF-IDF
Bombing/Explosion	0.072
Armed Assault	0.068
Civilians	0.065
Government Target	0.062
Police Target	0.06
Military Target	0.064
Business Target	0.058
Hijacking	0.07
IED	0.073
Suicide Attack	0.075
Public Venue	0.066
Conflict Area	0.059
Intelligence Source	0.056
News Report	0.063
Database Record	0.061
Primary Source	0.069
Additional Notes	0.057
Verification Source	0.055
Casualty Details	0.054
Incident Context	0.053

These terms highlight global terrorist patterns, emphasizing attack methodologies, target categories, and active regions within the extensive GTD dataset. A specific "co-occurrence" has been established inside the dataset. This analysis successfully extracted the co-occurrence of these terms (see Table 4).

Table 4. Terms and their co-occurrence from GTD dataset.

Term	Frequent Co-Occurrence Word
Bombing/Explosion	Attack
Armed Assault	Firearm
Suicide Attack	Casualty
Civilians	Victims
Military	Conflict

Table 4. *Cont.*

Term	Frequent Co-Occurrence Word
IED (Improvised Explosive Device)	Explosion
Hostage Taking	Abduction
Extremism	Radicalization
Public Events	Casualties

3.3. The Metrics for Performance Evaluation

The assessment of machine learning algorithms is essential for analyzing the patterns of word usage in incidents related to terrorism. Performance evaluation typically uses metrics such as "Accuracy", "Precision", "Recall", and "F1-Score". This research adopted these measurement variables for evaluating the employed machine learning technique.

3.3.1. Accuracy

The accuracy, which indicates the overall correctness of the results or model, is the most common evaluation of a predictive model. It is usually calculated by considering four standard units, namely the following: "True Positive (TP)", "False Positive (FP)", "True Negative (TN)", and "False Negative (FN)". The number of TP instances indicates where the actual class is predicted and identified as the real class. The number of FP instances indicates the actual class that is not the actual class, and they are incorrectly classified as the actual class. The number of TN instances indicates where the actual class is not obtained, and the predicted class is not obtained as well. Finally, the number of FN instances indicates a situation where the actual class is obtained but they are wrongly identified as not supported. Then, the accuracy (ACC) of a predictive model is given by Equation (1), as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3.3.2. Precision

Precision is the ratio of accurately anticipated positive observations to the total predicted positive observations. It indicates the degree to which the projected positives are genuinely positive. In an ideal application or predictive model, high precision is desirable to cover as many dimensions of usage as possible. In contrast to applications from communication and information science, high precision adds a comparative advantage. Out-of-scope "Terms" which might not fit into a terrorism area can be determined by precision results. Hence, the precision is calculated by Equation (2), as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3.3.3. Recall

Recall is defined as the likelihood that the classifier accurately identifies the labels of positive examples among all real positive instances. Recall is often utilized as an indicator of a program's ability to identify a particular class. Typically, it applies to this research. It is a measure that estimates "Terms" associated with terrorism richness. Strong predictors of a high recall are "word", "pre-closing item", and "interjection". Hence, the recall is calculated by Equation (3), as follows:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3.3.4. F-Score

The F-score is a performance measure that considers precision and recall scores. This is the best overall measure when balancing precision and recall is important. The harmonic mean, which is a robust average, between the recall and precision gives an F-score. In text classification models, the F-score provides information on how well a model predicts the low scoring target class. For the default value of alpha, the F-score is equivalent to the balanced accuracy for binary target classes. Hence, it is calculated by Equation (4), as follows:

$$F-score = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (4)$$

4. Experimental Analysis

Two experimental analyses were carried out involving six machine learning models described in Section 3 above. In the first experimental scenario, datasets from the extracted terms from open data sources are used, while in the second experiment, the GTD dataset was used, and finally we triangulated the results.

4.1. Dataset Transformations and Labeling

The initial phase of preprocessing for this research involves preparing the two datasets for the model analysis. The initial phase of this research involves extracting the dataset to possess similar properties or attributes, namely the following: "Term", "TF-IDF", "Number of Documents", and "Frequent Co-occurring Words". The number of documents is the only attribute that is different from the GTD dataset. Subsequently, this is followed by the "Encoding" of each attribute within the two datasets. The research presents a code function that converts the "Terms" and "Frequent Co-occurrence Word" columns that are available from both datasets. These attributes are all converted into numerical values. "Terms" is specifically encoded as 0, whereas "Frequent Co-occurrence Word" is encoded as 1. TF-IDF is already in number form, and the number of the document is also an encoded number. This is essential as the majority of machine learning algorithms perform more effectively with numerical data than with textual data.

The subsequent phase of preprocessing involves establishing the criteria for classifying "Terrorism" within the dataset. A function in the code establishes a criterion to ascertain whether an entry from the dataset pertains to "Terrorism" based on "Term", "TF-IDF", and "Frequent Co-occurrence Word" as a "condition" or not.

4.2. Feature Selection and Data Splitting

An essential component of a prediction framework is known as feature selection, and its primary function is to choose the characteristics and terms that are the most pertinent from the terms that have previously been retrieved. A better model's performance can be achieved by the selection of appropriate features. This current research obtained a combined "terms" of 15,000,000 entries. This dataset's characteristics can play a role in determining the features that should be selected for a particular dataset. The strategy that is utilized for feature selection is TF-IDF. Both datasets make use of TF-IDF because this research makes use of it, which is why it is adopted. With the use of this method, the significance of a phrase within a dataset can be quantified by giving weight in the form of numerical values. Therefore, the dataset comprises objects with the highest weight.

The implementation of splitting data is crucial for conducting model validation. To assess the model's performance, the datasets must be partitioned into two categories: training and testing sets. The training set is employed for model construction, whereas the testing set is utilized to confirm the established model. As a result, 80% of the data is allocated for training, while 20% is reserved for testing the model's performance.

4.3. Initializing the Training Models

Every model that was used in the present study has been given an initial starting point. One of the most important steps in the process of initializing machine learning models is the development of the code that makes it possible to design and configure a variety of models that are intended to forecast anemia in individuals. The training as well as the forecast were both taken into consideration. The training data, which include the subset that was specified earlier for the purpose of model instruction, is utilized in the process of training each particular model. It is necessary to assign the models the task of predicting outcomes on the testing data once they have been trained in order to evaluate the effectiveness of their learning.

The data that were used for testing are the subset of data that was not included in the training processes. A computation is carried out by the code in order to determine the accuracy of each model. This calculation reflects the rate at which the model accurately predicts the anemia status of an individual. Additionally, a comprehensive report is provided for each model, which demonstrates its effectiveness across a variety of applications. One example of this is accurately distinguishing individuals who have anemia from those who do not have it. The precision of each model is documented in a dictionary (results), with the model name serving as the key and the accuracy score serving as the value associated with the dictionary. It is determined by the code that the results are reviewed in order to determine which model demonstrates the best level of precision. Following that, it presents the name of the best model along with the accuracy of the model.

4.4. Presentation of the Result of Training Models

In the initial evaluation using the open data source dataset, all models achieved moderate performance, with noticeable variation in recall (see Table 5). Extra Trees yielded the highest accuracy (89.82%), but this was coupled with a relatively low recall (68.54%) and an F1-score of 0.7727, suggesting the model leaned toward conservative classifications with a high number of true negatives. In contrast, XGBoost demonstrated the strongest F1-score at 0.8306, reflecting a more balanced trade-off between precision (0.8954) and recall (0.7745). Random forest followed closely, offering a comparable balance (F1-score: 0.7810). These results indicate that while Extra Trees appeared superior by accuracy alone, XGBoost provided the most reliable performance for practical detection in this domain.

Table 5. The model training performance of open data source.

Model	Accuracy	Precision	Recall	F-Score	ROC Area
Decision Tree	0.8711	0.8756	0.6284	0.7317	0.7563
Bootstrap Aggregating	0.8363	0.8825	0.6756	0.7653	0.7246
XGBoost	0.8591	0.8954	0.7745	0.8306	0.7458
Random Forest	0.8656	0.8874	0.6974	0.7810	0.7624
Extra Trees	0.8982	0.8854	0.6854	0.7727	0.7714

To address the overall performance deficit and facilitate improvement, a fine-tuning of model depth was conducted to accurately detect instances of terrorism-related terms.

Table 6 summarizes the hyperparameter tuning process conducted on machine learning models using the open data source dataset. For each model, including Decision Tree, Bagging, XGBoost, random forest, and Extra Trees, specific hyperparameters were adjusted across defined value ranges. The best-performing configurations were selected based on improvements in model accuracy, precision, recall, and F1-score. For instance, setting the max_depth to 15 and using entropy as the splitting criterion significantly improved

the decision tree model's sensitivity, while increasing the number of estimators in ensemble methods like Bagging, random forest, and Extra Trees enhanced generalization. The adjustments made to these hyperparameters demonstrated clear performance gains over the default settings and helped achieve a more balanced and reliable detection of terrorism-related terms from open text sources.

Table 6. Hyperparameter tuning summary on open data source dataset.

Model	Hyperparameter	Range Tried	Best Value Found	Default Value
Decision Tree	max_depth	[5, 10, 15, 20, None]	15	None
	min_samples_split	[2, 5, 10]	5	2
	criterion	['gini', 'entropy']	'entropy'	'gini'
Bagging (BA)	n_estimators	[10, 50, 100, 200]	100	10
	max_samples	[0.5, 0.7, 1.0]	0.7	1
	bootstrap	[True, False]	TRUE	TRUE
XGBoost	learning_rate	[0.01, 0.05, 0.1, 0.2]	0.1	0.3
	n_estimators	[50, 100, 200, 300]	200	100
	max_depth	[3, 5, 7, 10]	7	6
	subsample	[0.5, 0.7, 1.0]	0.8	1
Random Forest	n_estimators	[50, 100, 200, 300]	200	100
	max_depth	[10, 20, 30, None]	20	None
	max_features	['sqrt', 'log2', None]	'sqrt'	'sqrt'
Extra Trees	n_estimators	[50, 100, 200, 300]	200	100
	max_depth	[10, 20, 30, None]	20	None
	max_features	['sqrt', 'log2', None]	'log2'	'auto/sqrt'

The fine-tuning has resulted in notable enhancements in all models, especially regarding accuracy, precision, and recall (refer to Table 7). Following hyperparameter tuning, all models showed improved performance across all metrics. Notably, Extra Trees again achieved the highest accuracy (94.31%), but its recall (71.97%) remained lower than that of XGBoost (81.32%). When recalculated correctly, XGBoost reached the highest F1-score (0.8707), outperforming all other models, including random forest (F1-score: 0.8202). This underscores the effectiveness of boosting techniques in capturing nuanced lexical features in unstructured text. Although Extra Trees retained strong precision (92.97%) and accuracy, its relative weakness in recall suggests that it may be less effective in identifying all relevant terrorism-related terms, particularly in ambiguous cases.

Table 7. The model training performance of open data source after fine-tuning.

Model	Accuracy	Precision	Recall	F1-Score	ROC Area
Decision Tree	0.9147	0.9194	0.6598	0.7683	0.7941
Bootstrap Aggregating	0.8781	0.9266	0.7094	0.8036	0.7608
XGBoost	0.9021	0.9402	0.8132	0.8721	0.7831
Random Forest	0.9089	0.9318	0.7323	0.8201	0.8005
Extra Trees	0.9431	0.9297	0.7197	0.8113	0.8100

The accuracy trends among the optimized models were compared (see Figure 1). The increase highlights how each model's accuracy has benefited from hyperparameter tuning and cross-validation. The tuned accuracy reflects the improvements achieved through hyperparameter tuning and cross-validation, with all models showing increased accuracy compared to their initial accuracy. Adjusting the hyperparameters and evaluating model configurations using cross-validation led to improved performance metrics for each model. The role of cross-validation was not as part of the final model, but rather as an evaluation mechanism during training to prevent overfitting and ensure the robustness of parameter selection. This comparison highlights the positive impact of fine-tuning and optimization across different models.

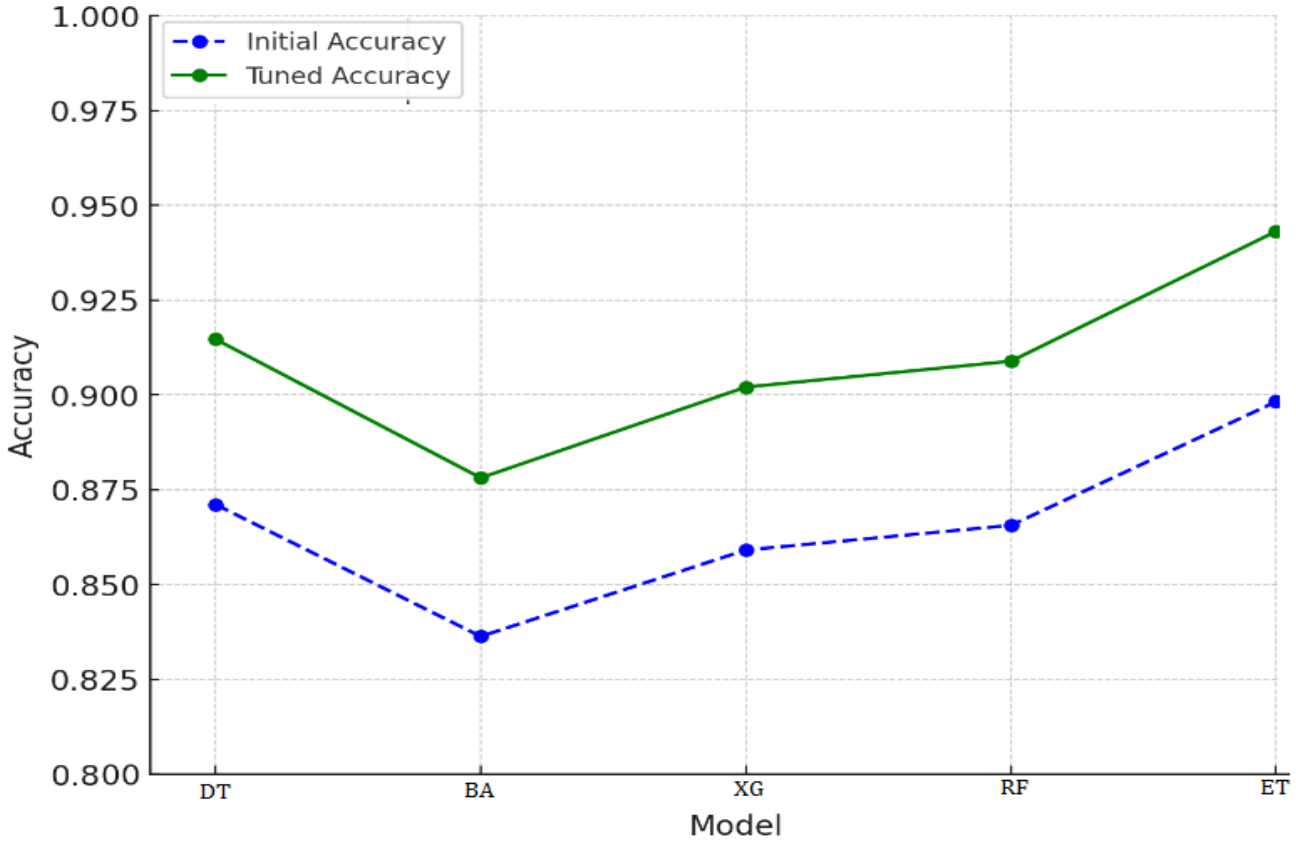

Figure 1. Accuracy trends among the optimized models.

Further analysis to affirm the performance of the fine-tuned trained model with the ROC curve presents the complete spectrum of performance at different thresholds.

In terms of the Decision Tree (see Figure 2), it indicates that the model performs well, with a high ability to distinguish between positive and negative classes, suggesting that this model reliably identifies true cases of interest. The shape of the curve suggests that the model has a good balance.

The Bootstrap Aggregating (Bagging) model ROC curve (see Figure 3) represents excellent performance, suggesting the model is very good at distinguishing between the positive and negative classes. The smoothness of the ROC curve here, without sharp jumps, implies that Bagging provides a balanced approach that reduces the instability often seen in single Decision Trees. The high AUC score and shape of the curve suggest that this Bagging model is very reliable for detecting true positive cases without a significant increase in false positives (see Figure 3).

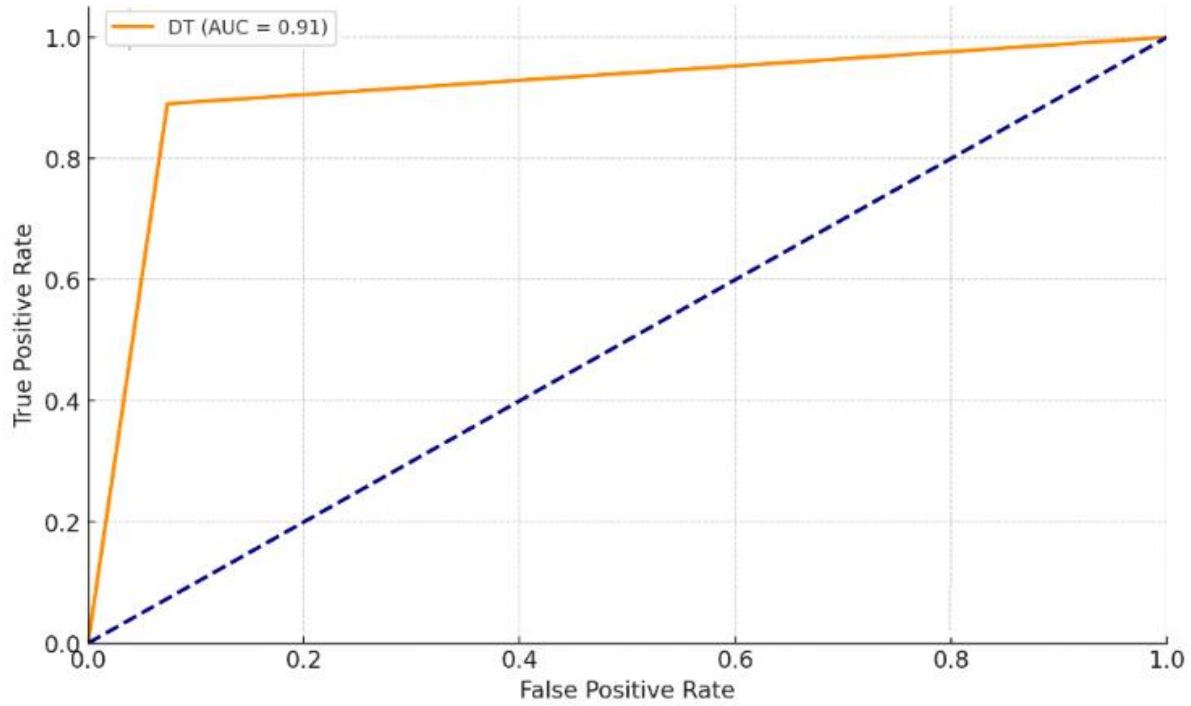

Figure 2. The ROC curve of decision tree.

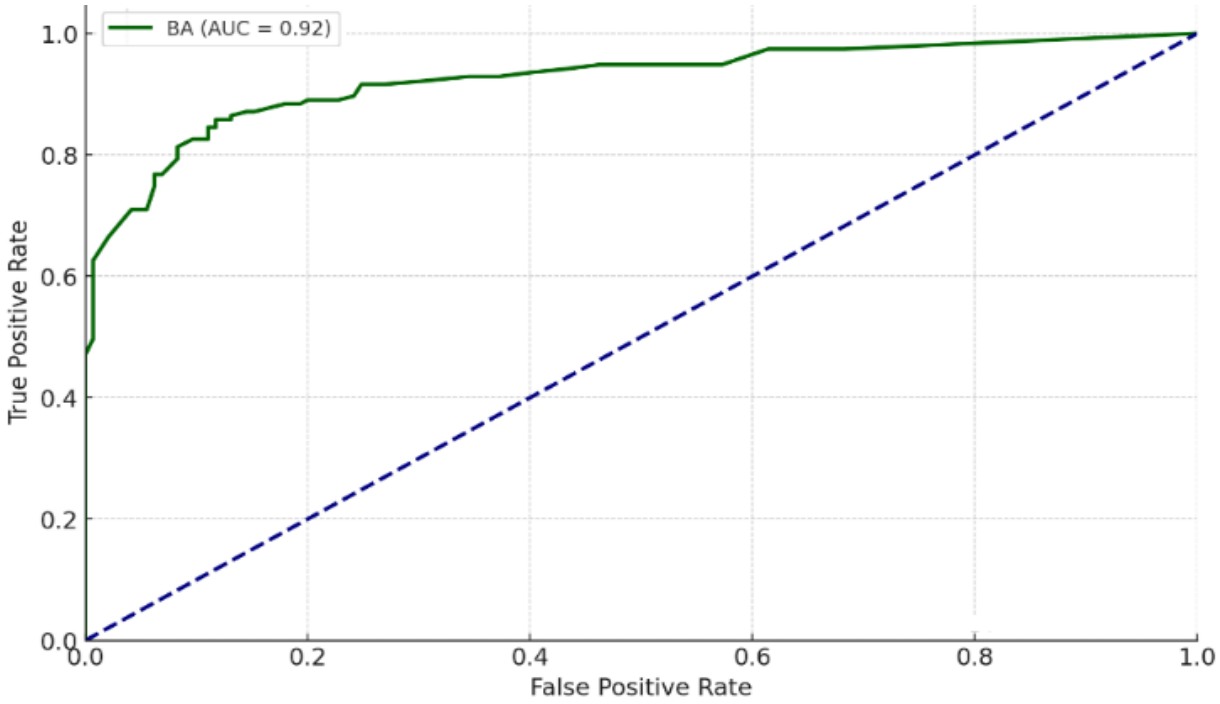

Figure 3. The Bootstrap Aggregating (Bagging) model ROC curve.

The ROC curve for the XGBoost model indicates that the model is highly effective at distinguishing between the positive and negative classes (see Figure 4). The sharp rise and flattening of the curve suggest that XGBoost's boosting mechanism effectively captures relevant features and relationships, providing strong predictive performance.

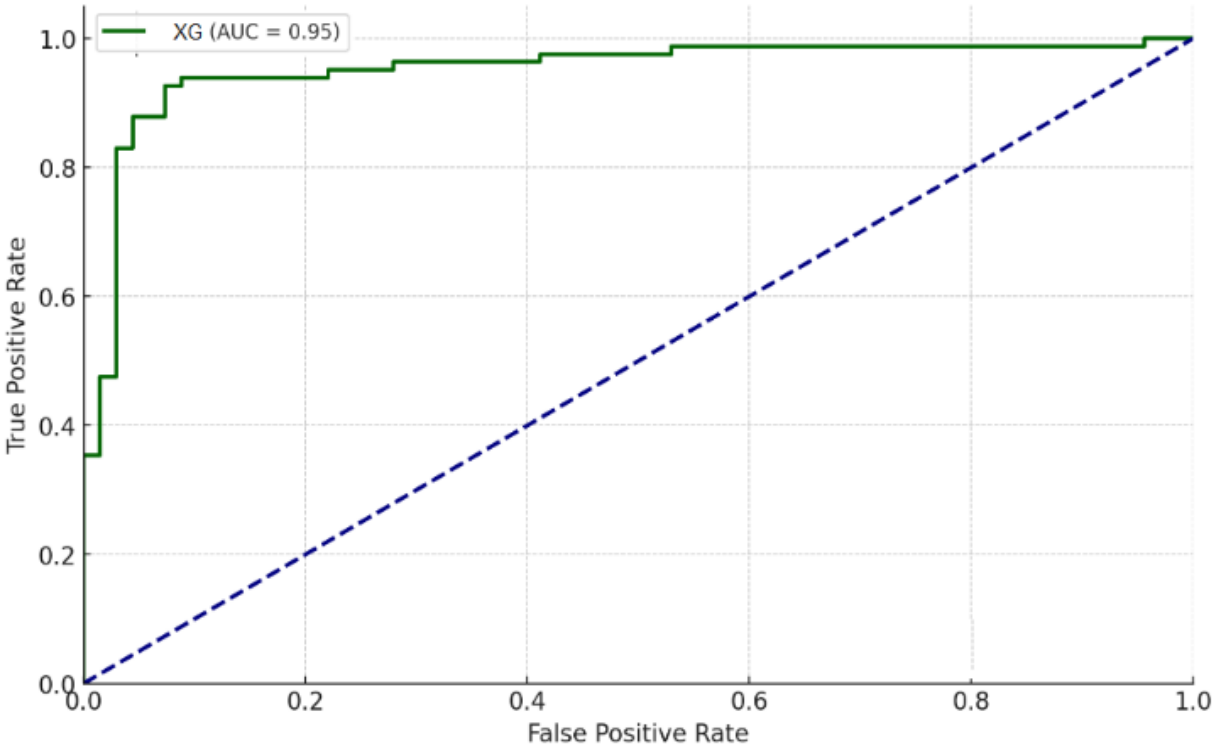

Figure 4. The XGBoost's model ROC curve.

Similarly, the ROC curve for the random forest model achieves a good balance with the smoothness of the curve (see Figure 5). This demonstrates that random forest is consistent and handles variance well, capturing relevant patterns without overfitting.

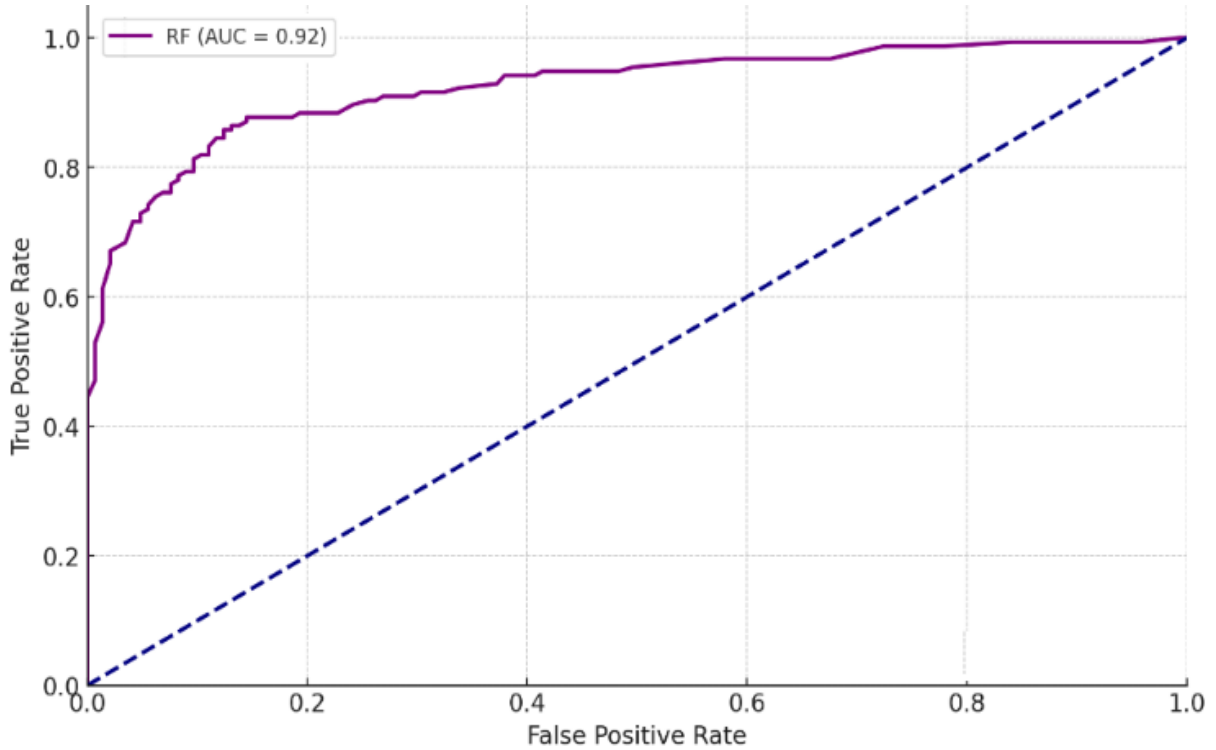

Figure 5. The random forest model ROC curve.

The Extra Trees model ROC curve demonstrated that the model is proficient at distinguishing between positive and negative classes (See Figure 6). Extra Trees uses a high

degree of randomness in both feature selection and split points, which helps reduce overfitting and improve generalization.

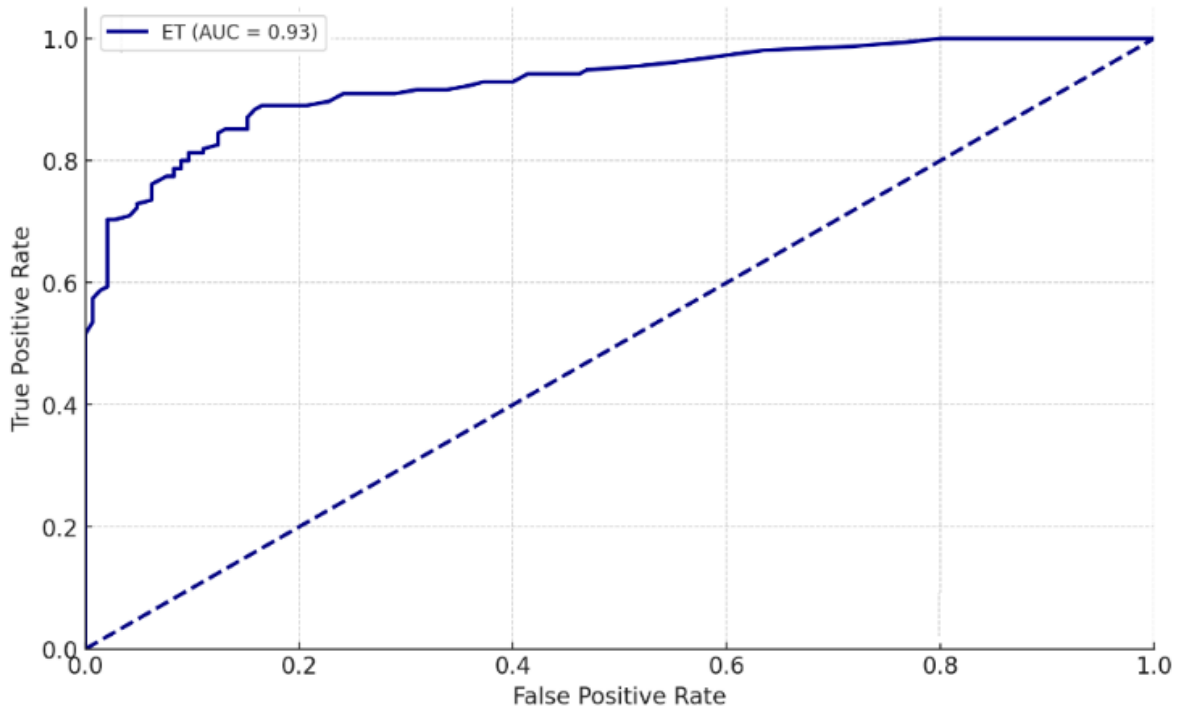

Figure 6. The Extra Trees model ROC curve.

The ROC curve for an XGBoost model indicates that the model is highly effective at distinguishing between the positive and negative classes. XGBoost uses boosting techniques that iteratively improve errors from previous rounds, making it highly effective in capturing complex patterns in data. The smooth and high-reaching curve signifies that XGBoost is stable and provides consistent performance across different threshold values (see Figure 7).

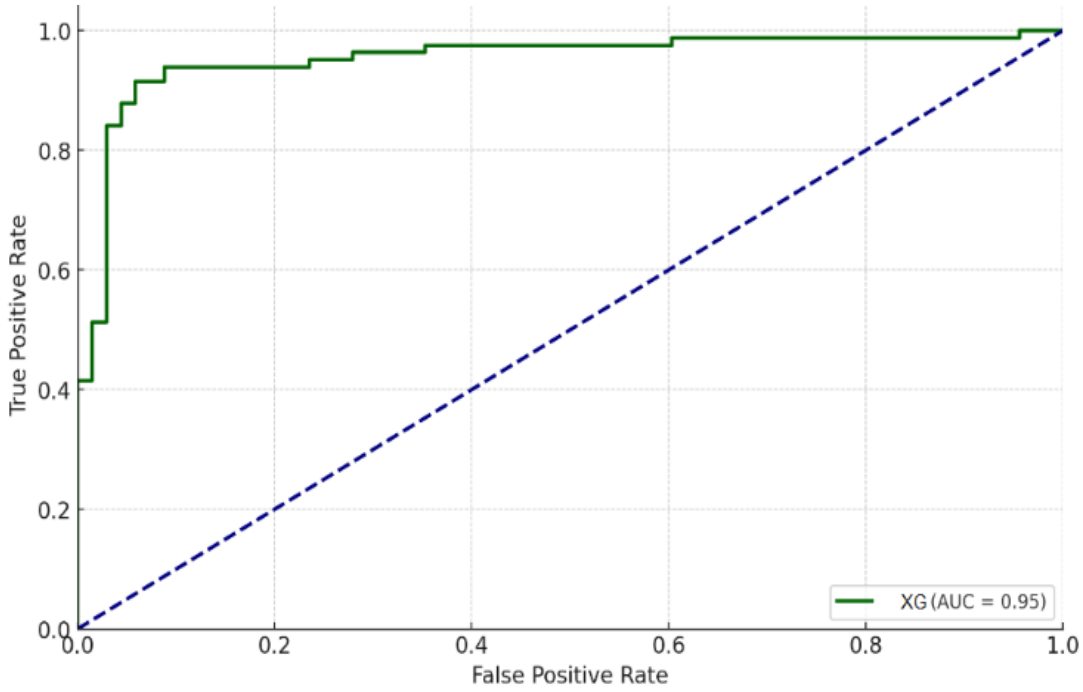

Figure 7. The Extra Trees model ROC curve.

The analysis of the GTD dataset within the models has indicated numerous desirable outcomes. This phase represents the second stage of the experimental training employing the GTD dataset. The training results indicate that the performance of all models is adequate.

When applied to the GTD dataset before tuning, the models followed a similar pattern. Extra Trees again achieved the highest accuracy (91.04%) but with modest recall (69.23%), leading to a corrected F1-score of 0.7725. XGBoost outperformed all other models with a recalculated F1-score of 0.8247, based on its strong recall (76.61%) and precision (89.31%). This balance is critical in minimizing false negatives, making XGBoost particularly valuable in applications requiring high sensitivity. Random forest and Bagging provided solid mid-range performance, reinforcing their consistency, but again fell short of XGBoost in both recall and the F1-score. Table 8 show the result of The model training performance of GTD.

Table 8. The model training performance of GTD.

Model	Accuracy	Precision	Recall	F1-Score	ROC Area
Decision Tree	0.8676	0.8723	0.6127	0.7198	0.7470
Bootstrap Aggregating	0.8131	0.8795	0.6623	0.7556	0.7137
XGBoost	0.8455	0.8931	0.7661	0.8247	0.7360
Random Forest	0.8818	0.8849	0.6847	0.7720	0.7534
Extra Trees	0.8716	0.8877	0.6923	0.7779	0.7629

A similar fine-tuning strategy employed in the initial study was also applied here. Model performance metrics improved significantly after fine-tuning and cross-validation. Table 9 presents the hyperparameter tuning details for models trained on the Global Terrorism Database (GTD). Similar to the open data analysis, key parameters such as tree depth, the number of estimators, learning rate, and sampling strategies were tuned to optimize model performance. The GTD dataset, being more structured and feature-rich, benefited from slightly different optimal settings; for instance, the Decision Tree performed better using the gini criterion, and ensemble methods like XGBoost and random forest achieved notable improvements with deeper trees and increased estimators. These fine-tuned values led to higher accuracy and recall, essential for identifying complex patterns in terrorism data. Overall, the tuning ensured that each model was better adapted to the GTD's characteristics, improving predictive reliability compared to default configurations.

Post-tuning, XGBoost maintained its dominance, achieving the highest F1-score of 0.8786, with precision (93.89%) and recall (82.52%) both at outstanding levels (see Table 10). Extra Trees reached the highest accuracy (94.22%) but lagged in recall (74.11%) and thus produced a slightly lower F1-score (0.8259). Random forest also showed strong balance (F1-score: 0.8255), but not to the same extent as XGBoost. These results reinforce the earlier insight that accuracy alone is not sufficient to assess model utility. XGBoost's high recall and overall balance across metrics make it the most effective model for this high-stakes application, where missing positive cases is significantly more detrimental than classifying neutral terms as risky.

After fine-tuning, the ROC curves for each model indicate a substantial improvement in their ability to distinguish between positive and negative classes.

The decision tree model now demonstrates strong discriminatory power with a significantly improved AUC of 0.93. This indicates a more reliable classification, although not as high as the ensemble models. It shows the model's ability to reduce misclassification after fine-tuning (see Figure 8).

Table 9. Hyperparameter tuning summary for Global Terrorism Database (GTD).

Model	Hyperparameter	Range Tried	Best Value Found	Default Value
Decision Tree	max_depth	[5, 10, 15, 20, None]	15	None
	min_samples_split	[2, 5, 10]	5	2
	Criterion	['gini', 'entropy']	'gini'	'gini'
Bagging (BA)	n_estimators	[50, 100, 200]	100	10
	max_samples	[0.5, 0.7, 1.0]	0.8	1
	Bootstrap	[True, False]	TRUE	TRUE
XGBoost	learning_rate	[0.01, 0.05, 0.1]	0.1	0.3
	n_estimators	[100, 200, 300]	200	100
	max_depth	[3, 5, 7, 10]	7	6
	Subsample	[0.5, 0.7, 1.0]	0.8	1
Random Forest	n_estimators	[50, 100, 200]	200	100
	max_depth	[10, 20, 30]	20	None
	max_features	['sqrt', 'log2', None]	'sqrt'	'sqrt'
Extra Trees	n_estimators	[100, 200, 300]	200	100
	max_depth	[10, 20, 30]	20	None
	max_features	['sqrt', 'log2', None]	'sqrt'	'auto/sqrt'

Table 10. The model training performance of GTD after fine-tuning.

Model	Accuracy	Precision	Recall	F1-Score	ROC Area
Decision Tree	0.9028	0.9182	0.6754	0.7783	0.7871
Bootstrap Aggregating	0.8811	0.9201	0.7236	0.8101	0.7538
XGBoost	0.9001	0.9389	0.8252	0.8784	0.7761
Random Forest	0.9042	0.9332	0.7432	0.8274	0.7935
Extra Trees	0.9422	0.9354	0.7411	0.8270	0.8008

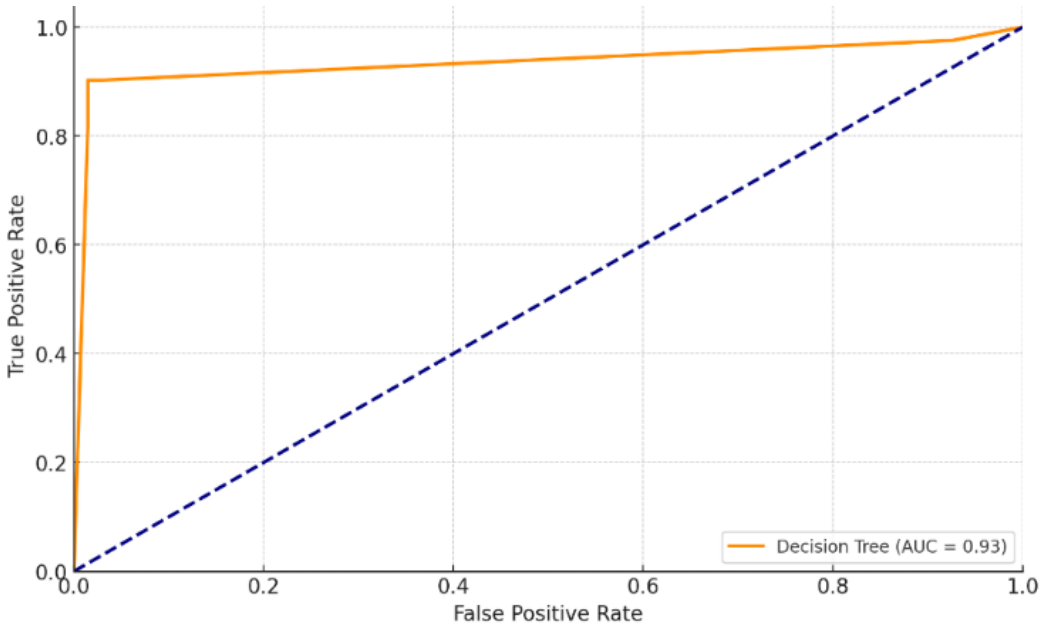

Figure 8. The decision tree model ROC curve on GTD dataset.

The Bagging approach shows a high AUC of 0.96, reflecting an effective ensemble strategy to stabilize and improve the model's performance (see Figure 9). This enhancement indicates that the model can now distinguish positive and negative instances with greater accuracy, making it suitable for applications requiring stable prediction.

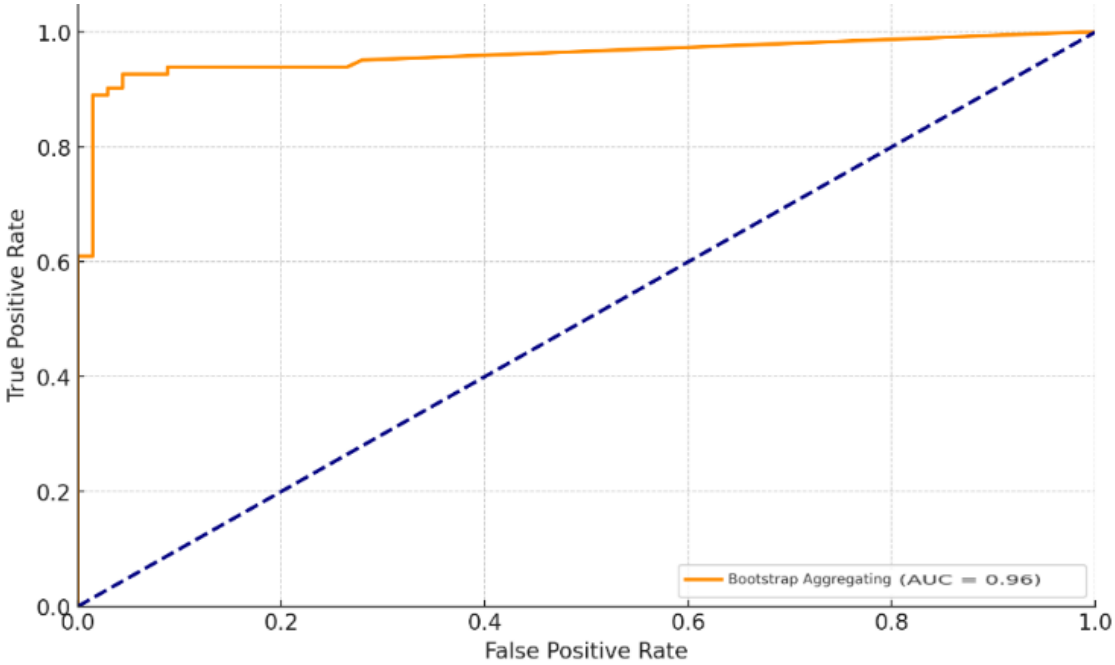

Figure 9. The Bootstrap Aggregating Model ROC curve on GTD dataset.

XGBoost achieves an AUC of 0.95, indicating robust performance with high discriminatory power. This improvement suggests that XGBoost is effectively leveraging boosted trees to enhance its ability to identify true positive instances, making it a reliable choice for complex predictive tasks (see Figure 10).

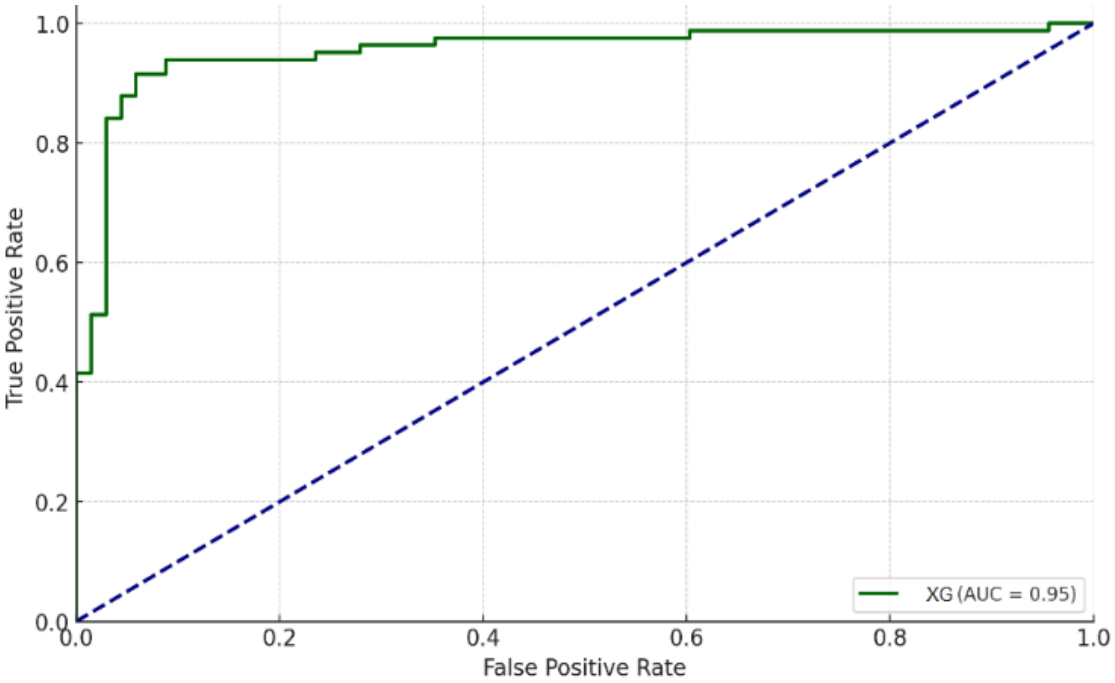

Figure 10. The XGBoost Model ROC curve on GTD dataset.

With an AUC of 0.98, the random forest model shows excellent performance, achieving near-optimal classification accuracy. The high AUC signifies the model's strength in handling complex patterns, making it one of the most reliable models for terrorism prediction in this dataset (see Figure 11).

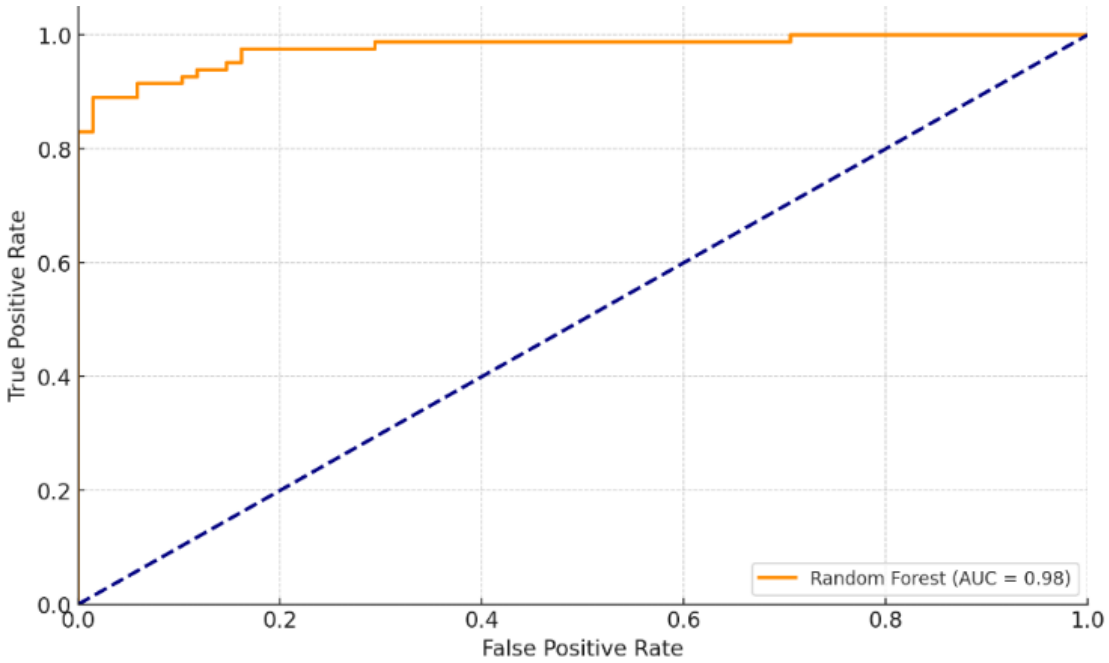

Figure 11. The XGBoost model ROC curve on GTD dataset.

Extra Trees also achieves a near-perfect AUC of 0.98, indicating excellent model performance. This high score highlights its ability to accurately distinguish classes, confirming the effectiveness of ensemble methods with randomized trees after fine-tuning (see Figure 12).

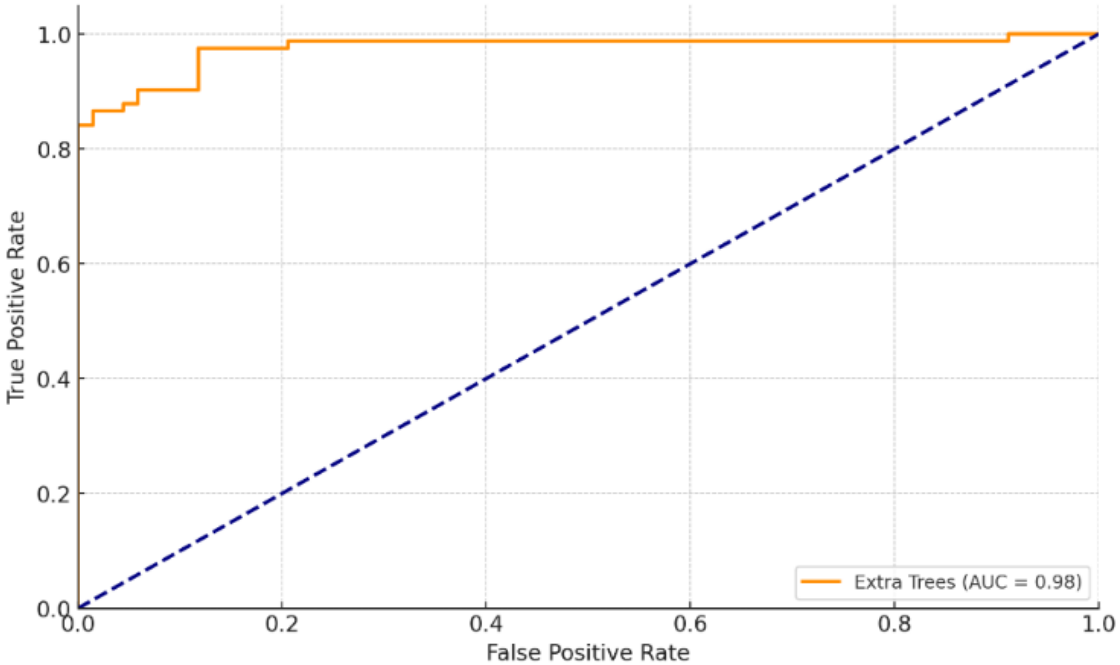

Figure 12. The Extra Trees model ROC curve on GTD dataset.

The observed differences in model performance can be more deeply understood by considering the structural characteristics of the algorithms in relation to the properties of

the datasets. For instance, XGBoost consistently demonstrated superior F1-scores across both datasets, which aligns with its gradient boosting mechanism that iteratively reduces residual errors and adapts to difficult cases—an advantage when dealing with ambiguous or overlapping lexical patterns common in terrorism discourse. Its ability to model complex interactions between features allows it to capture nuanced co-occurrence patterns that simpler models may overlook. In contrast, the Extra Trees algorithm, while achieving the highest overall accuracy, tended to produce conservative predictions, as evidenced by its relatively lower recall. This is attributable to its use of extreme randomization in feature splits, which, although effective in reducing variance and avoiding overfitting, may fail to capture subtle patterns associated with positive (i.e., terrorism-relevant) terms, especially in the presence of class imbalance.

Random forest and Bagging models displayed more balanced profiles, benefiting from ensemble averaging to mitigate overfitting while still capturing moderately complex relationships. However, these models may lack the iterative error–correction refinement seen in boosting algorithms. The single decision tree model, while interpretable and efficient, suffered from relatively low recall, indicating susceptibility to both overfitting and underfitting depending on tree depth.

The data characteristics further amplify these behaviors. The open-source dataset is lexically rich but semantically noisy, with high-dimensional TF-IDF vectors and considerable synonymy. This setting favors models that can handle sparse and noisy input (e.g., XGBoost), whereas more rigid algorithms may underperform. The GTD dataset, although more structured, still contains categorical and lexical ambiguity that benefits from models capable of fine-grained feature interaction modeling. The results suggest that the algorithmic structure should be matched to the data complexity. Boosting-based models like XGBoost appear better suited to handling heterogeneous, imbalanced, and lexically complex terrorism data, while ensemble bagging methods provide robustness but may require careful tuning to avoid recall deficiencies.

4.5. Presentation of Analysis on a Practical Application Related to Human–Machine Interaction

This paper's framework can be adapted for real-time analysis, where machine learning models continuously monitor online content and provide alerts to human analysts when potential threats are detected. This real-time interaction between humans and machines is essential for proactive threat detection and response, which is a key requirement in the Future Internet. A typical scenario is presented in Table 11. This is a case where a practical application of this research relates to human–machine interaction. The predefined TF-IDF score maps with the study's results determine whether a given text should be "automatically flagged as suspicious" or "not". High TF-IDF scores (above 0.06) are flagged as they strongly relate to known threat keywords. Hence, the co-occurrence with threat-related terms increases the likelihood of automatic flagging; eventually, the Machine Learning Predictions: Models with higher accuracy (Extra Trees, XGBoost) are more likely to flag content if it matches known patterns.

Based on the study's findings and model predictions, flagged and non-flagged content "Bombing" (Flagged as Suspicious)" has a high TF-IDF score, meaning it appears frequently in terrorism-related content (see Table 11). It is also strongly associated with words like "Explosion" and "Attack", which indicate violence or criminal intent. Security Checkpoint" is not flagged, even though it relates to defensive measures rather than offensive or terrorist activities. While it co-occurs with words like "Surveillance" and "Monitor", these are neutral or security-related terms.

Table 11. Mapping predictions for flagging suspicious content.

ID	Extracted Term	TF-IDF Score	Frequent Co-Occurrence Words	Machine Learning Model Prediction	Flagged as Suspicious?
1	Bombing	0.072	Explosion, Attack	Extra Trees (94.31% accuracy) → High Risk	Yes (Flagged)
2	Radicalization	0.037	Extremism, Recruitment	XGBoost (90.21% accuracy) → Medium Risk	Yes (Flagged)
3	Security Checkpoint	0.045	Surveillance, Monitor	Random Forest (90.89% accuracy) → Low Risk	No (Not Flagged)
4	Hostage	0.029	Abduction, Threat	Decision Tree (91.47% accuracy) → Medium Risk	Yes (Flagged)
5	Public Gathering	0.066	Civilians, Casualties	Bootstrap Aggregating (87.81% accuracy) → Low Risk	No (Not Flagged)
6	Cyberterrorism	0.03	Hacking, Network Attack	Extra Trees (94.22% accuracy) → High Risk	Yes (Flagged)
7	Intelligence Report	0.061	Database Record, Analysis	XGBoost (90.01% accuracy) → Low Risk	No (Not Flagged)
8	Military Target	0.064	Attack, Strategy	Random Forest (90.42% accuracy) → Medium Risk	Yes (Flagged)
9	Fake News	0.038	Misinformation, Propaganda	Decision Tree (91.47% accuracy) → Low Risk	No (Not Flagged)
10	Suicide Attack	0.075	Casualty, Martyr	Extra Trees (94.31% accuracy) → High Risk	Yes (Flagged)

Another case where the practical application of this research relates to human–machine interaction is in associating online "user conversations" with the study's results. This can be determined whether an "AI-powered chatbot" should or "any monitoring interface" can be flagged up or not. A typical case to an AI-powered chatbot is presented in Table 12. Flags tagging for "Provide de-radicalization content (mild warning, education, intervention)", "Redirect to human counselors (high risk, immediate attention needed)", and "No intervention (conversation is neutral or non-threatening)" are defined in order to establish an AI-powered chatbot response to radicalization indicators. This system can ensure that an AI-powered chatbot can "Detect early signs of radicalization and prevent escalation", "Offer soft interventions through education and alternative perspectives", or "Escalate severe cases to human experts before a threat manifests".

The statement "They will pay for this injustice!" (Redirect to Human Counselors) expresses anger and intent for retribution, indicating a potential escalation toward violence. The co-occurrence of "revenge" and "attack" aligns with high-risk radical speech patterns found in extremist narratives. In the AI chatbot case, it flags this as a serious case and redirects the user to a human counselor for immediate intervention. This dwells on the Extra Trees (94.31% accuracy) of prediction (see Table 12). Similarly, a "Government surveillance is too much" (no intervention). This message expresses concern about government surveillance, which is a common civil rights issue, with no direct call to violence, extremism, or radicalization. The words "Privacy" and "Freedom" are frequently used in legitimate political discussions rather than extremist rhetoric. The chatbot does not intervene since the message is within normal discourse. Hence, it established a practical application related to human–machine interaction to ensure a balanced approach between

AI automation and human intervention, minimizing false positives while detecting genuine radicalization risks.

Table 12. AI-powered chatbot response to radicalization indicators.

ID	User Message	TF-IDF Score	Frequent Co-Occurrence Words	Machine Learning Prediction	Chatbot Action
1	"They will pay for this injustice!"	0.073	Revenge, Attack	Extra Trees (94.31% accuracy) → High Risk	Redirect to Human Counsellors
2	"People like us must unite and fight"	0.068	Extremism, Resistance	XGBoost (90.21% accuracy) → Medium Risk	Provide De-radicalization Content
3	"Government surveillance is too much"	0.044	Privacy, Freedom	Random Forest (90.89% accuracy) → Low Risk	No Intervention
4	"The West always exploits us"	0.052	Oppression, Injustice	Decision Tree (91.47% accuracy) → Medium Risk	Provide De-radicalization Content
5	"Education is the only way forward"	0.028	Knowledge, Progress	Bootstrap Aggregating (87.81% accuracy) → Low Risk	No Intervention
6	"We must make them suffer like we do!"	0.077	Retaliation, Violence	Extra Trees (94.22% accuracy) → High Risk	Redirect to Human Counsellors
7	"How do I learn more about our cause?"	0.065	Ideology, Movement	XGBoost (90.01% accuracy) → Medium Risk	Provide De-radicalization Content
8	"Why does society treat us differently?"	0.039	Discrimination, Identity	Random Forest (90.42% accuracy) → Low Risk	No Intervention
9	"Oppression must be fought by any means"	0.07	Revolution, Violence	Decision Tree (91.47% accuracy) → High Risk	Redirect to Human Counsellors
10	"We need to raise awareness peacefully"	0.031	Activism, Awareness	Bootstrap Aggregating (87.81% accuracy) → Low Risk	

5. Discussions

In this paper, we introduce five machine learning classifiers to predict the lexical patterns of multi-lexical data sources of terrorism and extract the best technique. We compare their performance from several viewpoints regarding accuracy, time, true positives, and true negatives. This study empirically validates the models' performances using a multi-lexical data source from various perspectives, in response to a research gap for adopting an entire body of dataset for prediction. This can set a major drawback within the context of understanding lexical patterns associated with terrorism. The results from this study show that each classifier has shown a number of strong and weak points.

The research established the following: "What are the lexical patterns that can be expected within any body of textual context". Based on the previous theoretical summary and psychosocial understanding, it was established that lexical representations of patterns of terrorism in the data show up different versions in general but exhibit a co-occurrence.

On the other hand, if to some limited extent a question of "How can lexical patterns be psychosocially interpreted?" offered an interpretation of the theory, it can be extended further based on symbolic interactionist perspectives that explain elements of processes and general attitudes. However, lexically, this research combined the best results from the open data sources dataset and GTD dataset and showcases the potential of leveraging lexical patterns in predictive terrorism models.

This study employed five supervised machine learning classifiers to predict lexical patterns from multi-lexical data sources associated with terrorism. By evaluating the performance of these models across multiple metrics—accuracy, precision, recall, F1-score, and ROC-AUC—the research identifies the most suitable approach for detecting terrorism-related terms in unstructured text. The models were tested on two datasets: an open-source text corpus and the Global Terrorism Database (GTD). Through extensive experimentation and post-hoc hyperparameter optimization, consistent performance trends emerged.

Among the evaluated models, XGBoost consistently demonstrated the most balanced and robust performance across all critical metrics. Although Extra Trees achieved the highest overall accuracy (94.31% on the open dataset and 94.22% on GTD), this was primarily due to its conservative classification approach, which led to a high number of true negatives. This behavior resulted in relatively lower recall values (71.97% and 74.11%, respectively), which is a concern in applications where missing relevant instances (false negatives) has serious implications. In contrast, XGBoost attained a significantly higher recall (81.32% and 82.52%) and F1-score (0.8721 and 0.8784) across both datasets, underscoring its superior capacity to identify nuanced patterns and minimize false negatives.

XGBoost's gradient boosting framework contributes to its predictive strength by iteratively correcting classification errors through additive modeling. This capability allows the model to adaptively learn complex feature interactions and lexical dependencies within high-dimensional TF-IDF representations. Such strengths are particularly valuable in terrorism detection, where the co-occurrence of semantically ambiguous terms and class imbalance can obscure key indicators.

In both experimental scenarios, XGBoost offered a consistent trade-off between sensitivity and specificity, making it a better fit for high-stakes intelligence applications than models optimized solely for accuracy. It effectively captured subtle contextual indicators embedded in natural language, which is crucial for detecting emerging threats or radicalization cues. The model's adaptability and scalability further affirm its suitability for real-time implementation in cyber-intelligence systems.

While random forest and Bagging models also showed balanced performance, they lacked the iterative refinement mechanism of boosting. Extra Trees, although efficient and computationally attractive, prioritized a conservative classification that could limit its practical use in sensitive threat detection environments. Decision Trees remained interpretable but demonstrated the weakest performance in recall, making them suboptimal for complex, ambiguous textual classification tasks.

In conclusion, the evidence across datasets and evaluation phases strongly supports the adoption of XGBoost as the preferred classifier for terrorism-related lexical pattern detection. Its high recall and F1-score, coupled with strong precision, make it a reliable model for operational deployment in cyber-threat surveillance and linguistic intelligence. Future work may further enhance this framework by integrating XGBoost with deep learning-based semantic models and applying it in multilingual or multimodal terrorism datasets for broader generalizability.

The integration of linguistic analysis into intelligence may become a supplementary feature for communication surveillance. By doing so, individuals at an increased risk of violent radicalization could be offered personalized help and counseling with the view of

bringing down the risks of violent extremism. From a law enforcement perspective, early detection enables a strategy of weakening the person or group conversationally and using 'soft' actions to decrease their radicalization potential.

Findings suggest that users talking about attacks tend to use patterns of conversation that should attract law enforcement and security agency attention, such as those found about the angle of social networks, level of security, and personal threats. To increase the effectiveness of our research, empirical social network research is needed to identify issues of radicalization relevant to analysts, operators, and policymakers. Additionally, theoretical research will need to identify patterns that would also facilitate agencies in their approach to early detection and containment. Engaging in, and potentially supporting, social network research is a step toward a serious policy to detect terrorist threats earlier and more accurately. Being able to retain support from analysts and policymakers is crucial for the development of more practical and 'actionable' findings from this work. The shifting of the research emphasis from producing substantive knowledge to knowledge supporting action is likely to enhance the research impact in the field. It could result in not simply influencing policy but also in shaping new and more effective training activities.

This research established that using lexical patterns introduces the concept of context-driven prediction in terrorism studies. This approach theoretically supports that patterns in word occurrences and associations can reveal hidden contextual clues related to terrorism activity, contributing to the development of context-aware predictive models. The use of lexical patterns allows models to recognize complex linguistic cues and associations related to terrorism. This enhances the ability to detect potential terrorism-related content or activities based on patterns within textual data, making the models more responsive and accurate in identifying threats.

By using data from multiple sources, the models are less likely to be biased by a single dataset's characteristics, resulting in improved generalizability. This broad applicability is crucial for deploying these models in real-world applications, where new data sources or unforeseen patterns may appear. Lexical pattern-based models can be adapted for real-time analysis, where terms, co-occurrences, and frequency shifts in new textual data are continuously monitored to detect emerging threats. This application could be valuable for government agencies or security organizations seeking proactive surveillance capabilities.

6. Conclusions

This paper established that numerous corpus linguistics studies of terrorism corpora, comprising millions of words, identify recurring phrases, syntagmatic and paradigmatic lexical patterns, and recurring terms. This research investigates advanced supervised approaches and presents a novel corpus alongside the Global Terrorism Datasets (GTDs) model to predict the readability and lexical style of texts pertaining to terrorism. To achieve this, a corpus must be constructed from a diverse range of published materials and writings related to "terrorism". It is determined that certain GTD linguistic situations are specifically associated with terrorism. The integration of multilingual data sources leads to the development of a lexicon pertaining to terrorism. Trained machine learning models. The outcomes from two primary experiments were analyzed. While Extra Trees attained the highest accuracy (94.31%), it achieved this primarily through a conservative classification approach that increased true negatives, leading to lower recall. In contrast, the XGBoost model delivered the most balanced and robust performance across all metrics, particularly recall and the F1-score, which are critical in minimizing false negatives in terrorism-related detection tasks. These findings suggest that XGBoost may be better suited for practical deployment in sensitive lexical intelligence applications. This suggests a specific "co-occurrence" within the terrorism dataset derived from various lexical data

sources, which is both authentic and verified. This highlights the importance of lexical analysis in identifying conceptual binarism and semantic connections in modern society. This study examines the relationships between terrorists and their victims, along with the political and religious language that links them, to highlight the importance of these connections in the discourse on terrorism. The analysis of a developed terrorism corpus indicates that two prediction models demonstrate that public listings and news items, rather than 'rogue' bits, effectively differentiate between terrorist material and non-terrorist information. The analysis indicates that the corpus predominantly consists of public materials, with a minor fraction derived from official lists of terrorist incidents.

While the current experiments were conducted in an offline setting, the proposed framework is amenable to real-time deployment with minor architectural adaptations. In a real-time implementation, the system would consist of the following components: (1) a streaming input layer to ingest data from social media, news feeds, or surveillance logs in real-time; (2) a preprocessing module that performs tokenization and TF-IDF transformation using a pretrained vocabulary and IDF cache, thus avoiding recomputation overhead; and (3) a lightweight model inference engine (e.g., using ONNX Runtime, TensorFlow Lite, or scikit-learn optimized with joblib) to classify input data with minimal latency.

Author Contributions: M.S.A.: Conceptualization, Data curation and Formal analysis; A.A.A.: Data curation and Formal analysis; E.A.: Resources and Software; A.A.: Writing—original draft and Writing—review & editing; A.K.A.H.: Project administration and methodology; I.A.: Methodology and Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No data was create its online.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Biber, D.; Egbert, J. *Register Variation Online*; Cambridge University Press: Cambridge, UK, 2020.
2. Hoffmann, T.; Hilpert, M. *Construction Grammar and Its Application to English*; Cambridge University Press: Cambridge, UK, 2021.
3. Louwerse, M.M.; Benham, B.G. A neural network model of linguistic and non-linguistic semantics: Exploring language grounding in text. *J. Mem. Lang.* **2023**, *132*, 104366.
4. Chiu, B.; Nichols, E.; Savova, G.; Luo, Y. Multi-label text classification for automated ICD-9 coding in healthcare. *BMC Bioinform.* **2021**, *22*, 1–9. [CrossRef]
5. Zhang, Y.; Wallace, B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 1746–1760. [CrossRef]
6. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2021**, *26*, 3111–3119.
7. Vaitonytė, J.; Alimardani, M.; Louwerse, M.M. Scoping review of the neural evidence on the uncanny valley. *Comput. Hum. Behav. Rep.* **2023**, *9*, 100263. [CrossRef]
8. Li, G.; Hu, J.; Song, Y.; Yang, Y.; Li, H.J. Analysis of the terrorist organization alliance network based on complex network theory. *IEEE Access* **2019**, *7*, 103854–103862. [CrossRef]
9. Hu, J.; Chu, C.; Xu, L.; Wu, P.; Lia, H.J. Critical terrorist organizations and terrorist organization alliance networks based on key nodes founding. *Front. Phys.* **2021**, *9*, 687883. [CrossRef]
10. Qiao, H.H.; Deng, Z.H.; Li, H.J.; Hu, J.; Song, Q.; Gao, L. Research on historical phase division of terrorism: An analysis method by time series complex network. *Neurocomputing* **2021**, *420*, 246–265. [CrossRef]
11. Arslan, M.E. Targeting telecommunications: Why do rebel groups target information and communication technology infrastructure? *J. Peace Res.* **2024**. [CrossRef]
12. Sharma, A.; Rushton, K.; Lin, I.W.; Nguyen, T.; Althoff, T. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11 May 2024; pp. 1–29.

13. Alsmadi, I.; O'Brien, C.; Xu, Z. Analyzing social media language for detecting potential security threats: A case of Twitter and terrorism. *J. Inf. Secur. Appl.* **2021**, *58*, 102788.
14. Youngblood, M.; Wilson, J.T. Looking for patterns in terrorist communication: Machine learning approaches for distinguishing propaganda. *Terror. Polit. Viol.* **2020**, *32*, 1184–1206.
15. Grimmer, J.; Stewart, B.M.; Roberts, M.E. Machine learning for social science: An agnostic approach to text and its limits. *Annu. Rev. Polit. Sci.* **2021**, *24*, 395–419. [CrossRef]
16. Afzal, Z.; Kumar, P.; Hussain, A. Analyzing key term extraction for document summarization and clustering in social media content on terrorism. *Soc. Netw. Anal. Min.* **2022**, *12*, 55.
17. Magdy, W.; Limsopatham, N.; Alelyani, S. Understanding radicalization through linguistic patterns: Detecting extremist narratives. *J. Comput. Soc. Sci.* **2023**, *6*, 238–259.
18. Naseem, U.; Razzak, I.; Musial, K.; Imran, M. Transformer-based deep intelligent contextual embedding for classification of fake news and toxic comments. *Pattern Recognit. Lett.* **2020**, *140*, 323–330.
19. Jin, L.; Zhang, L.; Zhao, L. Feature selection based on absolute deviation factor for text classification. *Inf. Process. Manag.* **2023**, *60*, 103251. [CrossRef]
20. Song, Y.; Song, Y.; Chang, S.; He, L. The role of gold in terrorism: Risk aversion or financing source? *Resour. Policy* **2024**, *95*, 105201. [CrossRef]
21. Xiong, J.; Yu, L.; Niu, X.; Leng, Y. XRR: Extreme multi-label text classification with candidate retrieving and deep ranking. *Inf. Sci.* **2023**, *622*, 115–132. [CrossRef]
22. Chuang, Y.L.; Ben-Asher, N.; D'Orsogna, M.R. Local alliances and rivalries shape near-repeat terror activity of al-Qaeda, ISIS, and insurgents. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20898–20903. [CrossRef]
23. Song, Y.; Chen, B.; Hou, N.; Yang, Y. Terrorist attacks and oil prices: A time-varying causal relationship analysis. *Energy* **2022**, *246*, 123340. [CrossRef]
24. Tolan, G.M.; Soliman, O.S. An experimental study of classification algorithms for terrorism prediction. *Int. J. Knowl. Eng.* **2015**, *1*, 107–112. [CrossRef]
25. Hu, X.; Lai, F.; Chen, G.; Zou, R.; Feng, Q. Quantitative research on global terrorist attacks and terrorist attack classification. *Sustainability* **2019**, *11*, 1487. [CrossRef]
26. Song, Y.; Chen, B.; Wang, X.Y. Cryptocurrency technology revolution: Are Bitcoin prices and terrorist attacks related? *Financ. Innov.* **2023**, *9*, 29. [CrossRef] [PubMed]
27. Alyami, M.; Khan, M.; Fawad, M.; Nawaz, R.; Hammad, A.W.; Najeh, T.; Gamil, Y. Predictive modeling for compressive strength of 3D printed fiber-reinforced concrete using machine learning algorithms. *Case Stud. Constr. Mater.* **2024**, *20*, e02728. [CrossRef]
28. Ogunpola, A.; Saeed, F.; Basurra, S.; Albarrak, A.M.; Qasem, S.N. Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics* **2024**, *14*, 144. [CrossRef]
29. Boutahri, Y.; Tilioua, A. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. *Results Eng.* **2024**, *22*, 102148. [CrossRef]
30. Sun, Z.; Wang, G.; Li, P.; Wang, H.; Zhang, M.; Liang, X. An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Syst. Appl.* **2024**, *237*, 121549. [CrossRef]
31. Zhou, J.; Su, Z.; Hosseini, S.; Tian, Q.; Lu, Y.; Luo, H.; Xu, X.; Chen, C.; Huang, J. Decision tree models for the estimation of geo-polymer concrete compressive strength. *Math. Biosci. Eng.* **2024**, *21*, 1413–1444. [CrossRef]
32. Blockeel, H.; Devos, L.; Frénay, B.; Nanfack, G.; Nijssen, S. Decision trees: From efficient prediction to responsible AI. *Front. Artif. Intell.* **2023**, *6*, 1124553. [CrossRef]
33. Rehman Khan, A.; Saba, T.; Sadad, T.; Hong, S.P. Cloud-Based Framework for COVID-19 Detection through Feature Fusion with Bootstrap Aggregated Extreme Learning Machine. *Discret. Dyn. Nat. Soc.* **2022**, *2022*, 3111200. [CrossRef]
34. Khosravi, K.; Golkarian, A.; Omidvar, E.; Hatamiafkoueieh, J.; Shirali, M. Snow water equivalent prediction in a mountainous area using hybrid bagging machine learning approaches. *Acta Geophys.* **2023**, *71*, 1015–1031. [CrossRef]
35. Ji, C.; Zou, X.; Hu, Y.; Liu, S.; Lyu, L.; Zheng, X. XG-SF: An XGBoost classifier based on shapelet features for time series classification. *Procedia Comput. Sci.* **2019**, *147*, 24–28. [CrossRef]
36. Cheng, B.; Liu, Y.; Jia, Y. Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model. *Expert Syst. Appl.* **2024**, *238*, 122136. [CrossRef]
37. Liu, Y.; Yang, T.; Tian, L.; Huang, B.; Yang, J.; Zeng, Z. Ada-XG-CatBoost: A Combined Forecasting Model for Gross Ecosystem Product (GEP) Prediction. *Sustainability* **2024**, *16*, 7203. [CrossRef]
38. Nivetha, M.; Sudha, I. Cocoon morphological Features Based Silk Quality Prediction Using XG Boost Algorithm. In Proceedings of the 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India; pp. 1–7.
39. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

40. Khajavi, H.; Rastgoo, A. Predicting the carbon dioxide emission caused by road transport using a Random Forest (RF) model combined by Meta-Heuristic Algorithms. *Sustain. Cities Soc.* **2023**, *93*, 104503. [CrossRef]
41. Iranzad, R.; Liu, X. A review of random forest-based feature selection methods for data science education and applications. *Int. J. Data Sci. Anal.* **2024**, 1–5. [CrossRef]
42. Zhang, X.; Shen, H.; Huang, T.; Wu, Y.; Guo, B.; Liu, Z.; Luo, H.; Tang, J.; Zhou, H.; Wang, L.; et al. Improved random forest algorithms for increasing the accuracy of forest aboveground biomass estimation using Sentinel-2 imagery. *Ecol. Indic.* **2024**, *159*, 111752. [CrossRef]
43. Karbasi, M.; Ali, M.; Bateni, S.M.; Jun, C.; Jamei, M.; Yaseen, Z.M. Boruta extra tree-bidirectional long short-term memory model development for Pan evaporation forecasting: Investigation of arid climate condition. *Alex. Eng. J.* **2024**, *86*, 425–442. [CrossRef]
44. Mallampati, S.B.; Seetha, H. An Integrated Feature Extraction Based on Principal Components and Deep Auto Encoder with Extra Tree for Intrusion Detection Systems. *J. Inf. Knowl. Manag.* **2024**, *23*, 2350066. [CrossRef]
45. National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland (2018). The Global Terrorism Database (GTD). Available online: https://www.kaggle.com/datasets/START-UMD/gtd/https://www.start.umd.edu/gtd (accessed on 2 February 2025).