Scopus

Back

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

Future Internet • Article • *Open Access* • 2025 • DOI: 10.3390/fi17040182

Atoum, Mohammed Salem [a] ✉ ; Alarood, Ala Abdulsalam [b] ✉ ; Alsolami, Eesa [b] ✉ ; Abubakar, Adamu [c] ✉ ; Hwaitat, Ahmad K. Al [a] ✉ ; +1 author

[a] Department of Computer Science, The University of Jordan, Amman, 11942, Jordan

Show all information

1 78th percentile
Citation

1.18
FWCI

View PDF Full text Export Save to list

Document Impact Cited by (1) References (45) Similar documents

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. © 2025 by the authors.

Author keywords

cyber intelligence; machine learning; terrorism

Indexed keywords

Engineering controlled terms

Cybersecurity

Engineering uncontrolled terms

Cybe intelligence; Cyber security; Data-source; F1 scores; Global terrorism; Lexical patterns; Machine-learning; Open-source; Textual contexts; Textual data

Engineering main heading

Adversarial machine learning

Corresponding authors

Corresponding author M.S. Atoum