 You are accessing a free view of the Web of Science

[Learn More](#)



Results for CYBERSECURITY... >

Cybersecurity Intelligence Through Textual Data Analysis: A Framework Usi...








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

By	Atoum, MS (Atoum, Mohammed Salem) ; Alarood, AA (Alarood, Ala Abdulsalam) ; Alsolami, E (Alsolami, Eesa) ; Abubakar, A (Abubakar, Adamu) ; Al Hwaitat, AK (Al Hwaitat, Ahmad K.) ; Alsmadi, I (Alsmadi, Izzat)
	<a href="#">View Web of Science ResearcherID and ORCID</a> (provided by Clarivate)
Source	FUTURE INTERNET Volume: 17   Issue: 4 DOI: 10.3390/fi17040182
Article Number	182
Published	APR 21 2025
Indexed	2025-05-04
Document Type	Article
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****Author Keywords:** cyber intelligence; terrorism; machine learning**Addresses**

-  <sup>1</sup> Univ Jordan, Dept Comp Sci, Amman 11942, Jordan
-  <sup>2</sup> Univ Jeddah, Coll Comp Sci & Engn, Jeddah 21959, Saudi Arabia
-  <sup>3</sup> Int Islamic Univ Malaysia, Dept Comp Sci, Kuala Lumpur 53100, Malaysia
-  <sup>4</sup> Texas A&M Univ, Dept Comp Engn & Math Sci, San Antonio, TX 78224 USA
-  <sup>5</sup> Univ Jordan, Dept Comp Informat Syst, Aqaba 77110, Jordan

**Categories/  
Classification**

Research Areas: Computer Science

Citation Topics:	1 Clinical & Life Sciences	> 1.104 Virology - General	> 1.104.2810 AI in COVID-19
---------------------	-------------------------------	-------------------------------	--------------------------------

**Web of Science  
Categories**

Computer Science, Information Systems

[+ See more data fields](#)**Citation Network**

In Web of Science Core Collection

**1**

Citation

**45**

Cited References

**Use in Web of Science****0**

Last 180 Days

**0**

Since 2013

**This record is from:****Web of Science Core Collection**

- Emerging Sources Citation Index (ESCI)

**Suggest a correction**

If you would like to improve the quality of the data in this record, please [Suggest a correction](#)

