DETECTION OF ONLINE SPREAD TERRORISM USING WEB DATA MINING

***Project Reference No**: 49S_BE_6471*

College : G. Madegowda Institute of Technology, Maddur
Branch : Department of Computer Science and Engineering
Guide(s) : Prof. Madhushree
Student(s) : Mr. Chidambara L D
Mr. Mariswamy U
Mr. Monish M K
Mr. Vikas S

Keywords:

Web Mining, Terrorism Detection, Natural Language Processing (NLP), K-Means Clustering, Machine Learning, Cybersecurity

Introduction:

The rapid growth of the internet has transformed communication and information sharing, but it has also enabled terrorist organizations to spread extremist ideologies, recruit members, and coordinate activities online. Traditional monitoring systems are insufficient due to the massive volume and dynamic nature of online content.

This project proposes an automated system to detect terrorism-related content using web data mining techniques. By integrating Natural Language Processing (NLP) and K-Means clustering, the system analyzes large-scale online data such as social media posts, forums, and news articles. The aim is to identify harmful patterns, keywords, and sentiments to support cybersecurity agencies in preventing digital extremism and ensuring a safer online environment.

Objectives:

1. To collect and analyze data from online platforms such as social media, forums, and news sites.
2. To preprocess and clean web data for efficient analysis.
3. To extract meaningful features like keywords, sentiment, and topics using NLP.
4. To apply K-Means clustering for identifying patterns related to terrorist activities.
5. To develop an automated system for monitoring and detecting online extremist content.

Methodology:

The system follows a structured approach:

- **Data Collection:** Gather data from various online sources using web scraping techniques.
- **Data Preprocessing:** Remove noise, tokenize text, and normalize data using stemming or lemmatization.
- **Feature Extraction:** Apply NLP techniques such as TF-IDF, sentiment analysis, and topic modeling.
- **Clustering:** Use K-Means algorithm to group similar data into clusters.
- **Analysis & Evaluation:** Identify patterns and evaluate performance using clustering metrics.
- **Real-Time Monitoring:** Continuously update and analyze incoming data for early detection.

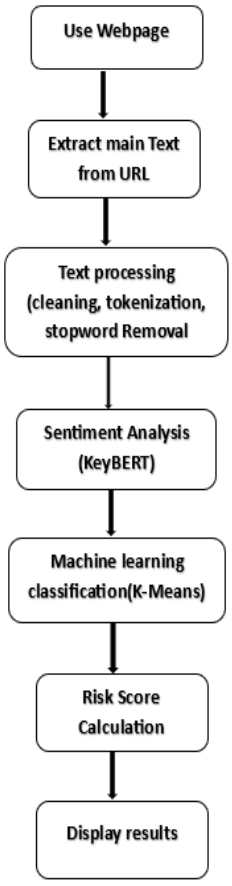

Fig: System Architect

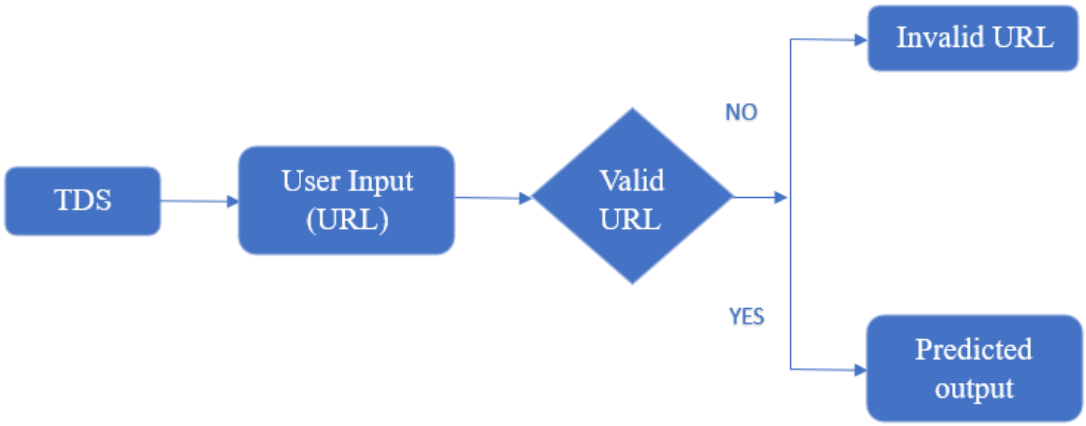

Fig: Proposed System

Result and Conclusion:

The proposed system successfully automates the detection of terrorism-related content from large volumes of online data. Compared to traditional keyword-based systems, it provides improved accuracy by understanding context and semantic meaning.

The use of NLP and clustering techniques reduces false positives and enables identification of hidden patterns in data. The system is scalable and can process real-time data, making it highly useful for cybersecurity agencies.

In conclusion, this project demonstrates how web data mining and AI can play a crucial role in combating the online spread of terrorism and enhancing global security.

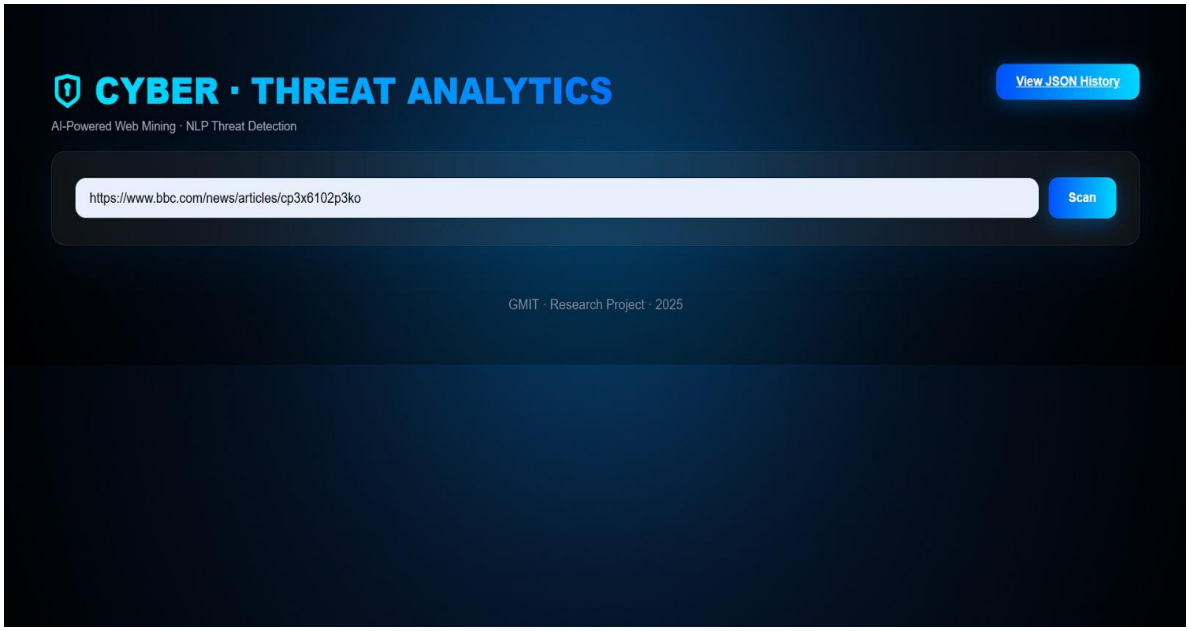

Fig: URL Input.

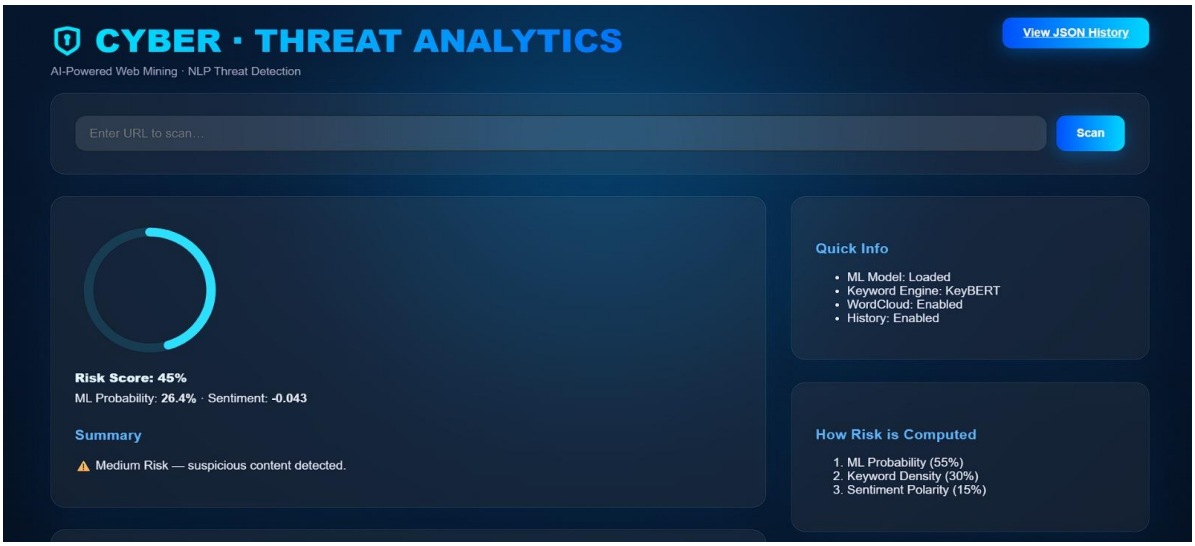

Fig: Prediction page

References:

1. Cyberbullying Detection using Machine Learning – IRJET, 2022
2. Detection of Cyberbullying using ML Techniques – 2021
3. Automated Detection of Cyberbullying – 2020
4. NLP and Web Mining Research Papers
5. Machine Learning and Data Mining Journals

Future Scope:

1. Implementation of advanced deep learning models like BERT and LLMs.
2. Integration of multimedia analysis (text, images, videos).
3. Real-time alert systems for detecting threats.
4. Cross-platform intelligence sharing among agencies.
5. Development of scalable and deployable security systems