International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

A Comprehensive Review on Spy-Socio: An Interactive System for Forensic Analysis of Crime Trends on Social Media Platform

Ms. Priyanka Padmane[1], Mr. Nishit Bhaje[2], Mr. Rajat Kamble[3], Ms. Muskan Madankar[4], Mr. Sujal Shahu[5], Mr. Vedant Rohankar[6]
[1]Assistant Professor, Dept. of CT, Priyadarshini College of Eng., Nagpur, Maharashtra, India
[2,3,4,5,6]UG Scholar, Dept. of CT, Priyadarshini College of Eng., Nagpur, Maharashtra, India
Emails: *priyankapadmane_ct@pcenagpur.edu.in[1], nishitbhaje@gmail.com[2], rajatrkamble@gmail.com[3], madankarmuskan@gmail.com[4], sujalshahu07@gmail.com[5], vedantrohankar1@gmail.com[6]*

Abstract
Online crime clues now appear in many places — tweets, forum posts, and local news pages — and this scatter makes it hard for police to find the few items that matter. We built SpySocio to gather those pieces and present them in one practical tool. SpySocio has two main parts: a social-platform collector that works with legally available sources (Twitter and Reddit) and a local-news collector focused on Nagpur newspapers. Instagram and Facebook are included only as "future-use" sections with a clear legal disclaimer because current rules block direct extraction. The heart of the project is the news collector: officers can type a keyword (for example, "kidnapping") and instantly get city-specific reports without reading full papers. The platform also indexes posts and articles for quick retrieval, lets users tag important items as evidence, and shows simple trend visualizations. In short, SpySocio reduces search time, centralizes relevant information, and supports faster investigative decisions.

Keywords: *Crime Trend Analysis, Social Media Forensics, City-Level News Extraction, Keyword-Based Crime Retrieval, Digital Investigation Support, Open-Source Crime Intelligence.*

1. Introduction
Today, relevant crime information is dispersed across a mix of social networks, community forums, and regional news outlets. That should help investigators — but in practice it slows them down. Officers often don't have time to trawl through multiple feeds or several local newspapers each day. Public sources such as Twitter and Reddit can be mined for eyewitness reports or incident chatter, but Instagram and Facebook restrict automated access under current legal and platform rules in India. Meanwhile, local newspapers publish many useful leads, but those are buried in long pages and multiple editions.

SpySocio is our response to this practical gap. The system collects permitted public posts from social platforms and scrapes city news as allowed, then presents everything through a single, searchable dashboard. Two complementary parts make it useful: one for social-media signals (operational for Twitter and Reddit; demonstrative for Instagram and Facebook) and one for rapid city-news retrieval (Nagpur newspapers). The goal: help police find the specific items they care about quickly, with minimal manual browsing.

2. Objectives
The project aims to make online crime information quick to find and easy to act on. Specific goals:

- Provide fast keyword lookup across selected online sources so officers can find incident reports without manual scanning.
- Implement lawful, working data collectors for sources that permit public access (Twitter, Reddit).
- Present Instagram and Facebook as placeholders with clear disclaimers explaining that full extraction requires policy changes or authorization.
- Build a robust local-news collector focused on

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

Nagpur that fetches and normalizes articles as they go online.

- Cut down the manual workload for police by centralizing relevant content.
- Offer a simple interface with filters and visual summaries usable by non-technical users.
- Keep the design flexible so future features (AI/NLP, multilingual support, more sources) can be integrated.

3. Literature Review

In Document-level Event Extraction from Italian Crime News Using Minimal Data, G. Bonisoli et al. (2025) used minimal-data learning models to extract full crime events from news articles and concluded that crime details can be captured accurately even with limited training data [1]. In Social Media Analysis in Criminal Investigation, P. M. Khan and S. Kumar (2024) collected and analyzed social-media signals for police use and concluded that online content helps investigators understand crimes more quickly [2]. In Twitter as a Lens for Crime Analysis: A Comprehensive 4W Model for Identifying Crime Patterns and Insights, B. Kuhaneswaran et al. (2023) applied a 4W extraction model to Twitter posts and concluded that Twitter can identify what happened, when, where, and to whom in crime incidents [3]. In A Comprehensive Survey on Artifact Recovery from Social Media Platforms, K. Gupta et al. (2023) reviewed methods for recovering social-media artifacts and concluded that important digital evidence can still be retrieved to support investigations [4]. In CEM: An Ontology for Crime Events in Newspaper Articles, F. Rollo et al. (2023) built a crime-event ontology and concluded that structured ontologies improve the accuracy of crime extraction from news articles [5]. In Online News Event Extraction for Crime Analysis, F. Rollo et al. (2022) used NLP-based methods to extract 5W1H crime details from online newspapers and concluded that crime information can be automatically gathered from news text [6]. In Online News Event Extraction for Crime Analysis (AI4Crime), F. Rollo et al. (2022) combined NLP and event-extraction techniques and concluded that long and unstructured news articles can be converted into clear crime-event data [7]. In Correlating Crime and Social Media: Using Semantic Similarity for Crime Prediction, A. Ajani et al. (2021) used semantic similarity to study crime-related text and concluded that comparing meanings in crime narratives helps predict crime trends [8]. In Crime Analysis and Forecasting on Spatio-Temporal News Feed Data—An Indian Context, S. Gupta and A. Kumar (2021) applied spatio-temporal analysis to news feeds and concluded that long-term city news helps identify crime hotspots and future patterns [9]. In An End-to-End Framework for Dynamic Crime Profiling of Places, S. K. Gupta et al. (2021) used location-based crime datasets and profiling models and concluded that dynamic profiling helps understand how crime varies across places [10]. In Analysis and Classification of Crime Tweets Using Machine Learning Techniques, R. Singh and A. K. Bansal (2020) applied ML classifiers to crime-related tweets and concluded that tweets can be automatically grouped into crime categories with good accuracy [11]. In A Framework for Detecting Intentions of Criminal Acts in Social Media, R. R. de Mendonça et al. (2020) used NLP and ML to detect intentions in tweets and concluded that their framework can identify posts that may express criminal intent [12]. In A Study of Information Extraction Tools for Online English Newspapers, A. Alkaff and M. Mohd (2020) compared information-extraction tools and concluded that effective tools make crime-related newspaper data easier to extract [13]. In Spatiotemporal Analysis of Web News Archives for Crime Prediction, A. Umair et al. (2020) used long-term news archives for crime hotspot detection and concluded that archived news helps study crime evolution over time [14]. In Extracting Entities and Topics from News and Criminal Records, Q. C. Pham et al. (2020) used machine-learning techniques for entity and topic extraction and concluded that combining news and criminal records improves investigative insights [15]. In Open Social Data Crime Analytics, H. Alghamdi and A. M. Jones (2017) analyzed open social-media data to extract crime indicators and concluded that publicly

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

available information can reveal meaningful crime patterns [16]. In CrimeProfiler: Extraction and Visualization of Crime Information from Online News Media, T. Dasgupta et al. (2017) used NLP to extract and visualize crime information from newspapers and concluded that automated systems can clearly present news-based crime data [17]. In Crime Information Extraction From News Articles, R. Arulanandam et al. (2014) applied NLP methods to extract victim, crime type, and location and concluded that key crime facts can be accurately extracted from news articles [18]. In Extracting Crime Information from Online Newspaper Articles, R. Arulanandam et al. (2014) used rule-based and NLP techniques and concluded that combining rules with NLP extracts useful and structured crime information from newspapers [19]. In Extraction of Nationality from Crime News, A. Alkaff and M. Mohd (2013) used NLP rules to detect nationality in crime articles and concluded that nationality information can be accurately identified from news text [20].

4. Proposed System / Methodology

SpySocio is designed as an extensible, source-agnostic information system that emphasizes legality, speed, and usability. Below is the approach we followed.

4.1 System Architecture (Overview)

The platform is arranged in clear functional layers so components can be added or updated independently:

- Acquisition layer: collects public content from Twitter and Reddit via official APIs and retrieves local news pages when allowed. Instagram and Facebook appear as UI sections with a "not operational" notice until legal access is available.
- Preprocessing & index layer: cleans raw text, normalizes dates, extracts simple location hints and crime keywords, then indexes records for fast lookup.
- Analysis layer: runs keyword tagging, basic named-entity recognition (names, places), and aggregates counts for trend charts. This layer also exposes hooks for later AI/NLP modules.
- Storage & evidence layer: keeps raw snapshots and processed records, along with provenance (source URL, fetch time). Evidence items are stored immutably once flagged.
- Presentation layer: web dashboard where officers search, filter, view results and graphs, and mark evidence.
- Security & compliance layer: enforces access controls, encryption, audit trails, and displays legal disclaimers for restricted modules.

4.2 Implementation Summary (Practical Choices)

(These are implementation patterns — you may substitute equivalent tools.)

- Frontend: single-page application offering search, filters, results list, and simple charts.
- Backend: REST APIs handling queries, ingestion jobs, authentication, and admin tasks.
- Index/search: a purpose-built text index to support rapid keyword queries and basic ranking.
- Storage: document-style store for raw JSON/html snapshots and processed records.
- Collectors: use Twitter and Reddit official endpoints for public data; local newspaper pages are fetched by respectful, rate-limited scrapers only where permitted.
- Scheduler & workers: background workers run periodic fetch and processing pipelines.
- Visualization: lightweight charting for timelines and frequency counts.
- Security: encrypted transport and storage, role-based access, and tamper-evident logs.
- Disclaimer handling: Instagram/Facebook tabs clearly state legal limitations and show demo content only.

4.3 Module descriptions

1. Social Media Module

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

- **Twitter:** Active feed collection by keywords; stores tweet text, author handle, timestamp, and media links.
- **Reddit:** Polls public subreddits and threads for regionally relevant posts and comments.
- **Instagram & Facebook**: Present in UI as non-operational modules with legal notes and mockup views.

2. News & media module (main USP)

Continuously monitors a curated list of Nagpur newspaper URLs, extracts article text and metadata, and prioritizes city-level crime reporting so investigators see local incidents first.

- Search & filter module: Full-text queries, date ranges, source selection (Twitter / Reddit / News), and basic crime tags.
- Visualization & analytics module: Simple timeline charts, counts by crime category, and shortlist cards for recent high-priority items.
- Evidence repository & audit: Raw snapshots are saved when an officer marks an item as evidence; all access and exports are logged.
- Administration & configuration: Admin can add/remove sources, update keyword lists, and manage publishing of disclaimers for restricted platforms.

4.4 Workflow (stepwise)

- Admin enters sources (keywords, subreddit names, newspaper links).
- Scheduled jobs fetch new items from allowed sources.
- Preprocessor normalizes text, tags keywords, and extracts metadata.
- Indexer updates the search index.
- Investigator logs in and runs a keyword search with filters.
- System returns ranked results, small previews, and quick actions (view, mark evidence, export).
- Investigator views trend visualizations for context.
- Items flagged as evidence are sealed in the evidence store and audit logged.
- Admin reviews logs and updates sources or disclaimer statuses as required.

4.5 Legal and ethical safeguards

Collection is limited to publicly viewable content only and respects site terms. Instagram and Facebook are intentionally non-operational without proper authorization. Only authorized users can access evidence; every action is recorded for accountability.

5. Results and Discussion

5.1 Results

The SpySocio system performed well when tested with public social media posts and Nagpur city newspapers. For any keyword entered—such as “murder,” “kidnapping,” or “robbery”—the system immediately displayed matching news articles and online posts. The responses were quick, and all relevant information appeared in one place. This removed the need for officers to open different websites or scroll through long newspaper pages. The city-level news module gave especially accurate results, bringing out local crime updates as soon as newspapers published them online.

5.2 Discussion

The overall results show that SpySocio greatly reduces the effort required to gather crime-related information. Normally, police officers must read long newspaper pages or move through scattered online posts to find one specific crime update. With SpySocio, this entire process is replaced by a single keyword search, which instantly presents all related news and posts in one organized dashboard. The newspaper module stands out as the most helpful part of the system. It delivers precise, city-level crime news, something officers often struggle to collect manually because these updates are spread across many newspaper pages and websites. By bringing all information together and making it searchable, SpySocio supports faster awareness, quicker

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

decision-making, and a smoother investigation process.

Conclusion

Online sources hold valuable hints for policing, but value is lost when those hints are scattered across platforms and printed pages. SpySocio brings permitted social posts and city news into one searchable environment so investigators can find what matters fast. The working collectors (Twitter, Reddit) provide public social signals; the news collector focuses on Nagpur newspapers and is the system's key strength because it delivers city-level incidents without manual newspaper reading. Instagram and Facebook remain as planned placeholders with clear legal warnings. By centralizing relevant content, indexing it for rapid lookup, and letting officers save evidence snapshots, SpySocio reduces manual effort and improves situational awareness. Future upgrades — such as language support, stronger geo-tagging, and AI-assisted triage — can expand its usefulness once policy and technical constraints permit.

Acknowledgements

We would like to sincerely thank our project guide and the faculty members of our department for their consistent guidance, helpful suggestions, and encouragement during the completion of this work. We are also grateful to our institution for providing the resources and environment needed to carry out this project. This study did not receive any external funding or financial assistance; all work was completed independently as part of our academic requirements.

References

[1]. G. Bonisoli, F. Rollo, and L. Po, "Document-level Event Extraction from Italian Crime News Using Minimal Data," Knowledge-Based Systems, vol. 296, pp. 1–12, 2025.

[2]. P. M. Khan and S. Kumar, "Social Media Analysis in Criminal Investigation," International Journal of Science, Engineering and Technology, 2024.

[3]. B. Kuhaneswaran, C. Sandagiri, B. T. G. S. Kumara, and Z. Li, "Twitter as a Lens for Crime Analysis: A Comprehensive 4W Model for Identifying Crime Patterns and Insights," Preprint, 2023.

[4]. K. Gupta, D. Oladimeji, C. Varol, A. Rasheed, and N. Shahidshidhar, "A Comprehensive Survey on Artifact Recovery from Social Media Platforms: Approaches and Future Research Directions," Information, vol. 14, no. 12, pp. 1–28, 2023.

[5]. F. Rollo, L. Po, and G. Bonisoli, "CEM: An Ontology for Crime Events in Newspaper Articles," in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing (SAC), pp. 1–3, 2023.

[6]. F. Rollo, L. Po, and G. Bonisoli, "Online News Event Extraction for Crime Analysis," in Proceedings of the 1st International Workshop on AI for Crime, CEUR Workshop Proceedings, vol. 3194, pp. 1–8, 2022.

[7]. F. Rollo, L. Po, and G. Bonisoli, "Online News Event Extraction for Crime Analysis," in Proceedings of the 1st International Workshop on Artificial Intelligence for Crime (AI4Crime), CEUR Workshop Proceedings, vol. 3194, pp. 1–8, 2022.

[8]. A. Ajani, A. Oloja, and S. Adedoyin, "Correlating Crime and Social Media: Using Semantic Similarity for Crime Prediction," International Journal of Advanced Computer Science and Applications, 2021.

[9]. S. Gupta and A. Kumar, "Crime Analysis and Forecasting on Spatio-Temporal News Feed Data—An Indian Context," International Journal of Advanced Research in Computer Science, vol. 12, no. 4, pp. 45–52, 2021.

[10]. S. K. Gupta, S. Shekhar, N. Goel, and M. Saini, "An End-to-End Framework for Dynamic Crime Profiling of Places," arXiv preprint arXiv:2111.10549, 2021.

[11]. R. Singh and A. K. Bansal, "Analysis and Classification of Crime Tweets Using Machine Learning Techniques," Procedia Computer Science, 2020.

[12]. R. R. de Mendonça, D. F. de Brito, F. d. F. Rosa, J. C. dos Reis, and R. Bonacin, "A Framework for Detecting Intentions of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4325-4330
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0633

Criminal Acts in Social Media: A Case Study on Twitter," Information, vol. 11, no. 3, pp. 1–16, 2020.

[13]. A. Alkaff and M. Mohd, "A Study of Information Extraction Tools for Online English Newspapers: Comparative Analysis," International Journal of Research and Reviews in Computer Science, vol. 11, no. 3, pp. 45–52, 2020.

[14]. A. Umair, F. Riaz, A. Naeem, and M. F. Sohail, "Spatiotemporal Analysis of Web News Archives for Crime Prediction," Applied Sciences, vol. 10, no. 22, pp. 1–20, 2020.

[15]. Q. C. Pham, M. Stanojević, and Z. Obradović, "Extracting Entities and Topics from News and Criminal Records," arXiv preprint arXiv:2005.00950, 2020.

[16]. H. Alghamdi and A. M. Jones, "Open Social Data Crime Analytics," International Journal of Advanced Computer Science and Applications, vol. 8, no. 9, pp. 160–167, 2017.

[17]. T. Dasgupta, A. Naskar, R. Saha, and L. Dey, "CrimeProfiler: Extraction and Visualization of Crime Information from Online News Media," International Journal of Computer Applications, vol. 170, no. 4, pp. 1–7, 2017.

[18]. R. Arulanandam, B. T. R. Savarimuthu, and M. A. Purvis, "Crime Information Extraction From News Articles," in Proceedings of the Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2014.

[19]. R. Arulanandam, S. Raman, and M. Nandhini, "Extracting Crime Information from Online Newspaper Articles," ICACCI, 2014.

[20]. A. Alkaff and M. Mohd, "Extraction of Nationality from Crime News," Journal of Theoretical and Applied Information Technology, vol. 54, no. 2, pp. 215–223, 2013.