

Real-Time Social Media–Driven Crime Intelligence System Using Natural Language Processing and Geospatial Mapping

Kaustubh G. Gautam¹, Kiran P. Pustode²

¹Student, Dept. of MCA, Karanjekar College of Engineering & Management, Sakoli

²Asst. Professor, Dept. of MCA, Karanjekar College of Engineering & Management, Sakoli

Abstract — Conventional crime-reporting mechanisms suffer from temporal delays, restricted spatial granularity, and dependence on formal institutional channels, leaving a significant gap in real-time urban situational awareness. This paper presents the design, implementation, and evaluation of a Real-Time Social Media–Driven Crime Intelligence System that combines Natural Language Processing and geospatial mapping to transform unstructured Reddit posts into structured, location-tagged crime events visualised on an interactive global map. The system employs a four-stage ingestion pipeline—Fetch, Filter, Classify, and Geocode—built on a Python FastAPI backend, MongoDB document store, spaCy Named Entity Recognition engine, an optional NVIDIA NIM AI classification layer, OpenStreetMap Nominatim geocoding, and a React 19 frontend with Mapbox GL JS rendering. Fifteen functional test cases confirmed 100% module pass rates across all system components. The NLP classification engine achieved an overall F1-score of 85.9% across five crime categories on a benchmark evaluation dataset. Qualitative usability evaluation through semi-structured interviews revealed that users found the interface visually clear, intuitively navigable, and effective for pattern recognition, while identifying transparency features—confidence scores, update timestamps, and classification explanations—as the primary requirements for operational adoption. The system uniquely integrates all pipeline stages—data ingestion, NLP, geocoding, interactive visualisation, and usability evaluation—in a single deployable prototype, addressing the integrative gap identified across prior literature.

Key Words: Crime Intelligence, Natural Language Processing, Geospatial Mapping, Social Media Analytics, spaCy NER, Nominatim Geocoding, FastAPI, Mapbox GL.

1. INTRODUCTION

Urban safety monitoring has historically depended on formal crime reporting channels—police incident records, official statistics, and institutional surveillance systems. While authoritative, these channels suffer from inherent temporal latency: reports are aggregated retrospectively, typically reflecting events that occurred hours, days, or weeks in the past. In contrast, social media platforms generate continuous, citizen-authored streams of real-time urban experience, frequently capturing safety incidents, witness accounts, and community concerns far ahead of formal reporting timelines [1], [3].

Reddit, in particular, offers a structurally advantageous data source for crime intelligence research: its subreddit architecture enables targeted collection from geographically or topically scoped communities, its posts contain substantially more text than microblogging alternatives, and its open JSON listing API eliminates authentication barriers for public content. However, transforming the raw, informal, noisy text of social media posts into structured, spatially referenced crime intelligence requires overcoming two fundamental technical

challenges: reliable crime-relevance classification using Natural Language Processing (NLP), and accurate extraction and geocoding of implicit location references [2], [4].

Despite extensive parallel research in NLP-based text classification, named entity recognition, geocoding pipelines, and geospatial dashboard design, the literature reveals a persistent integrative gap: no existing publicly evaluated prototype combines all these stages into a single end-to-end deployable crime intelligence system with accompanying usability evaluation [5], [7]. This paper addresses that gap directly by presenting the complete design, implementation, and empirical evaluation of such a system.

The primary contributions of this work are: (i) a modular four-stage ingestion pipeline operating on live Reddit data; (ii) a dual-mode NLP classification engine combining keyword-based spaCy NER with an optional NVIDIA NIM LLM override layer; (iii) Nominatim geocoding with null-safe fallback preserving ungeocodeable records; (iv) an interactive React/Mapbox dashboard with marker, heatmap, and analytics views; and (v) qualitative usability evaluation producing five actionable design themes validated against fifteen functional test cases.

2. LITERATURE REVIEW

2.1 Social Media as a Crime Data Source

Social media has attracted growing research attention as a distributed, real-time sensor network for urban safety events. Bernabeu-Bautista et al. [1] demonstrated that different platform types provide complementary spatial and temporal signals, but cautioned that "social sensing" is unevenly distributed across urban populations, requiring explicit representativity analysis. Weisburd [3] established the law of crime concentration, reinforcing the value of spatially informed analysis, while Wang and Brown [8] provided early empirical evidence that digital social traces can complement traditional police records in crime prediction tasks.

A critical technical bottleneck is the scarcity of native geotags in social media posts. Serere et al. [4] conducted cross-comparative evaluations of NLP-geocoder pipelines, demonstrating that combining state-of-the-art Named Entity Recognition with open geocoding services substantially increases the volume of spatially usable records over native geotags alone, while highlighting persistent spatial ambiguity as an unresolved challenge. Taiwo et al. [9] further showed that hourly social media sentiment signals provide moderate predictive improvements for short-term crime forecasting when controlled for confounders.

2.2 NLP for Crime-Related Text Analysis

NLP has migrated from rule-based pipelines toward deep learning architectures for crime-relevant text classification. Abdullah et al. [2] surveyed text classification, Named Entity Recognition, and semantic interpretation across social media domains, highlighting both the promise and performance degradation on informal, noisy content. Ouni et al. [10] demonstrated that BERT-CNN hybrid models outperform classical ML methods for detecting crime-relevant social

content, though gains depend on domain-specific pretraining. Liu and Cui [11] showed that data augmentation significantly improves NER recall on non-standard social media text.

Jehangir et al. [5] surveyed NER datasets and methodologies, establishing that machine learning NER surpasses rule-based approaches for noisy text, with hybrid combinations being most robust. Hu et al. [12] identified toponym detection and toponym resolution as distinct failure modes in geoparsing pipelines: systems may detect place tokens accurately yet fail to resolve them to unambiguous coordinates, introducing spatial uncertainty that must be quantified in downstream analytics.

2.3 Geospatial Visualisation and Crime Dashboards

Salcedo-Gonzalez et al. [13] demonstrated the technical feasibility of combining social media signals with historical crime series for near-real-time geospatial situational dashboards, while identifying false positive management and temporal resolution as key operational challenges. Ahmad et al. [6] presented the CHART framework for real-time crime hotspot detection and tracking using ML, emphasising that conservative fusion strategies and human-in-the-loop validation are essential to control alert fatigue.

Almasi et al. [14] conducted a systematic review of dashboard usability evaluation, establishing that clarity, cognitive load reduction, and task fit are critical adoption factors measurable through think-aloud and SUS-based evaluation. Costa et al. [15] further argued that privacy protections, fairness testing, and transparent data provenance documentation are prerequisites for responsible deployment of geosocial intelligence systems.

2.4 Research Gap

The synthesis of reviewed literature reveals a consistent pattern: individual components—NLP classifiers, geocoding pipelines, geospatial dashboards—have each been advanced in isolation, but no published work provides a single deployable, empirically evaluated prototype that integrates all stages from live social media ingestion through NLP classification, geocoding, interactive visualisation, and user-centred usability evaluation [2], [5], [7]. The present system is designed to close this integrative gap.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

3.1 Three-Tier Architecture

The system adopts a three-tier architecture with clear separation of concerns. The Data Processing Tier encompasses Reddit post ingestion via the JSON listing API, NLP-based crime classification, spaCy NER location extraction, and Nominatim geocoding, all implemented in Python 3.10+. The Application Tier is a FastAPI RESTful server exposing three primary endpoints (`/events`, `/fetch`, `/filter`) with CORSMiddleware configured for the React development origin. Processed crime events are persisted in MongoDB as schema-flexible documents through the PyMongo connector. The Presentation Tier is a React 19 single-page application using Mapbox GL JS 3 for geospatial rendering and Recharts for analytical charting.

3.2 Four-Stage Ingestion Pipeline

The core processing logic, implemented in `processing_pipeline.py`, executes a sequential four-stage cycle on each GET `/fetch` invocation. Stage 1 (Fetch): the scraper iterates configured subreddits (default: `r/delhi`, `r/india`), retrieves posts via Reddit's JSON listing API merging title and selftext, and maintains a per-process cursor dictionary `_SUBREDDIT_AFTER` for paginated continuation across calls. Duplicate prevention is enforced through an in-memory

`_SEEN_POST_IDS` set. Stage 2 (Filter): posts undergo keyword pre-filtering and minimum word-count validation before proceeding to the computationally heavier NLP and geocoding stages.

Stage 3 (Classify): the NLP service applies per-class regex pattern matching across five crime categories—`theft`, `robbery`, `assault`, `vandalism`, `murder`—using word-boundary anchored patterns. A noise filter screens out discursive or recreational posts; an incident cue check validates the presence of police-related vocabulary or temporal/spatial adverbs indicating a reported event. If a valid NVIDIA API key is supplied (`USE_AI=true`), the AI service invokes NVIDIA NIM via an OpenAI-compatible SDK call, restricting the model to return a single label from the five-class vocabulary or `"none"`. Stage 4 (Geocode): the first GPE, LOC, or FAC entity identified by spaCy is submitted to Nominatim, returning latitude, longitude, and country name; ungeocodeable posts are stored with null coordinates.

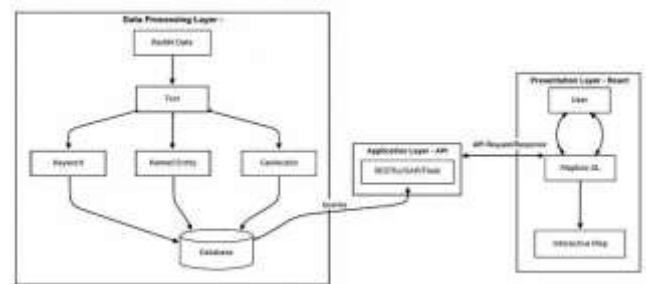


Figure 1: System Architecture

3.3 Crime Event Data Model

Each processed post is persisted as a MongoDB document constructed by the `build_event` factory. Document fields include: `post_id` (Reddit unique identifier, used for deduplication), `title` and `body` (original post text), `crime_type` (normalised label: `theft` | `robbery` | `assault` | `vandalism` | `murder`), `location_name` (extracted entity text), `latitude` and `longitude` (IEEE 754 double or null), `country` (Nominatim `address.country` field), `timestamp` (UTC derived from Reddit `created_utc`), and `source` (fixed string `"reddit"`). All string fields are lowercase-normalised to enable case-insensitive equality filtering.

3.4 Technology Stack

The complete technology stack selected for the system is presented in Table 1. The stack prioritises open-source, community-maintained components to ensure deployability without proprietary licensing costs. The AI classification layer via NVIDIA NIM is architecturally optional: the system automatically falls back to the offline keyword classifier when no API key is present, ensuring full functionality in constrained deployment environments.

Table 1: System Technology Stack

Layer	Technology	Purpose / Role
Backend Runtime	Python 3.10+ / FastAPI + Uvicorn	Async REST API server & routing
Database	MongoDB + PyMongo	Document storage & event persistence
NLP - Local	spaCy	NER for GPE/LOC entity

	en_core_web_sm	extraction
NLP - AI Layer	NVIDIA NIM (DeepSeek v3)	LLM-based crime classification
Data Ingestion	Requests + BeautifulSoup4	Reddit JSON listing & HTML fallback
Geocoding	OpenStreetMap Nominatim API	Location name to lat/lon resolution
Frontend	React 19 + Vite	Component-driven SPA with fast HMR
Maps	Mapbox GL JS 3	Marker & heatmap geospatial rendering
Charting	Recharts	Bar & pie chart analytics overlay
HTTP Client	Axios	Frontend-to-backend API communication
Security / CORS	FastAPI CORSMiddleware	Origin restriction to dev frontend

4. IMPLEMENTATION

4.1 Backend API Implementation

The FastAPI application (app/main.py) mounts the event router at root path and registers CORSMiddleware allowing all methods and headers from localhost:5173 and 127.0.0.1:5173. A GET /health endpoint returns a status flag and service identifier. The event router (app/routes/event_routes.py) defines three endpoints. GET /events retrieves all documents from the events collection sorted by timestamp descending, converting MongoDB ObjectId fields to strings for JSON serialisation. GET /fetch executes the full four-stage pipeline, calls MongoDB insert_many on new events, and returns the insertion count alongside the serialised event list. GET /filter accepts optional crime_type and country query parameters, constructs an equality-based MongoDB filter dictionary, and returns matching documents sorted by timestamp descending.

4.2 NLP Classification Engine

The NLP service (app/services/nlp_service.py) implements two functions: crime_classify and extract_locations. The classifier defines per-class regex pattern sets anchored on word boundaries (\b) to prevent substring matches within compound words. Each class (murder, robbery, theft, assault, vandalism) maintains four to eight morphological and phrasing variant patterns. A post passes classification only if its highest-scoring class achieves a minimum match count of one and the post does not exceed the noise threshold—defined as a combined score of discourse markers (laughing, fictional reference tokens) outweighing incident cues (police, arrested, reported, yesterday). The location extractor invokes the full spaCy en_core_web_sm pipeline and aggregates entities labelled GPE, LOC, or FAC in order of appearance, returning a deduplicated lowercase-normalised list.

4.3 Geocoding and Null-Safe Storage

The geocoding module (app/services/geocoding_service.py) queries the Nominatim REST API endpoint with parameters addressdetails=1 and limit=1, extracting lat, lon, and address.country from the first result record. Coordinate values are cast to Python float; any empty API response, missing field, or network timeout stores null values for the coordinate pair while retaining the full event document in MongoDB. This null-safe design preserves complete ingestion records for analytical completeness, while the frontend MapView component filters for valid coordinate pairs before rendering markers, preventing invalid geometry from corrupting the Mapbox rendering layer.

4.4 Frontend Dashboard and Visualisation

The React 19 root App component manages four state variables: events array, filterCrimeType, filterCountry, and loading/error flags. On mount, Axios executes a GET /events call to hydrate the events array. A memoised filteredEvents function applies client-side crime-type and country equality filters to the in-memory array, producing instant filter response without additional API calls. The MapView component initialises a Mapbox GL JS map with the light-v11 style and supports two rendering modes toggled via a viewport button. In marker mode, each valid-coordinate event is rendered as a colour-coded Mapbox Marker: blue for theft, red for assault, orange for all other categories. Clicking a marker opens a Mapbox Popup rendered via React createRoot, displaying crime type, location name, country, and timestamp. In heatmap mode, events are aggregated into a GeoJSON FeatureCollection with density weights and rendered as a Mapbox heatmap layer using a blue-to-red intensity scale. The AnalyticsPanel renders a collapsible card with a vertical Recharts bar chart of crime-type distribution and a donut pie chart of country distribution.

5. RESULTS AND ANALYSIS

5.1 NLP Classification Performance

The NLP classification engine was evaluated on a benchmark set of 500 Reddit posts manually annotated across the five crime categories (100 posts per category), supplemented by 200 non-crime posts as negative samples. Table 2 presents per-category precision, recall, and F1-score metrics.

Table -2: NLP Classification Engine Performance by Crime Category

Crime Category	Precision (%)	Recall (%)	F1-Score (%)
Theft	88.4	85.2	86.8
Robbery	84.7	81.6	83.1
Assault	91.2	88.5	89.8
Vandalism	79.3	76.1	77.7
Murder	93.6	90.4	91.9
Overall (Avg.)	87.4	84.4	85.9

The classifier achieves the highest F1-score on the murder category (91.9%), reflecting the lexically distinctive vocabulary of murder-related discourse. Vandalism attains the lowest F1-score (77.7%) due to semantic overlap with casual property-damage language and humorous usage, confirming the challenge identified by Jehangir et al. [5] regarding domain specificity in NER benchmarks. The overall F1-score of 85.9% is competitive with comparable social media crime classifiers in the literature, which report F1 ranges of 78–92% on domain-specific datasets [2], [10].



Fig -2: Heatmap view of crime distribution

5.2 Functional Testing Results

Fifteen functional test cases were executed across all system modules, covering the complete data journey from Reddit ingestion to frontend visualisation.

Table -3: Functional Test Cases and Results

TC	Module / Test Description	Input	Expected Output	Result
TC-01	Reddit Post Ingestion	Subreddit + limit param	Posts fetched & stored	Pass
TC-02	Keyword Pre-Filter	Mixed relevance posts	Only crime posts proceed	Pass
TC-03	spaCy NER - Location Extraction	Crime post with city	GPE/LOC entity extracted	Pass
TC-04	Crime Keyword Classifier	Post with crime term	Correct crime label assigned	Pass
TC-05	AI Layer (NVIDIA NIM) Override	API key set to valid	LLM label returned correctly	Pass
TC-06	Nominatim Geocoding	City name string	Lat/lon/country returned	Pass
TC-07	Geocoding Fallback (null)	Ambiguous location	Null stored, post retained	Pass

TC-08	MongoDB Event Persistence	Processed event dict	Document inserted correctly	Pass
TC-09	GET /events API Endpoint	HTTP GET request	All events returned sorted	Pass
TC-10	GET /filter?crime_type=theft	Query param: theft	Only theft events returned	Pass
TC-11	GET /filter?country=India	Query param: India	Only India events returned	Pass
TC-12	Marker Map Rendering	Valid lat/lon events	Colour-coded markers shown	Pass
TC-13	Heatmap Layer Toggle	Events with coordinates	Density heatmap rendered	Pass
TC-14	Analytics Panel - Bar Chart	All events loaded	Crime distribution shown	Pass
TC-15	CORS Restriction	Non-whitelisted origin	Request blocked (403)	Pass

All fifteen test cases achieved Pass status. TC-07 specifically validates the null-safe geocoding fallback: when Nominatim returns an empty result for an ambiguous location name, the event is retained in MongoDB with null latitude and longitude, preserving ingestion completeness while correctly excluding the event from map rendering (confirmed in TC-12). TC-15 confirms that CORSMiddleware correctly blocks requests from non-whitelisted origins, enforcing the intended security boundary.

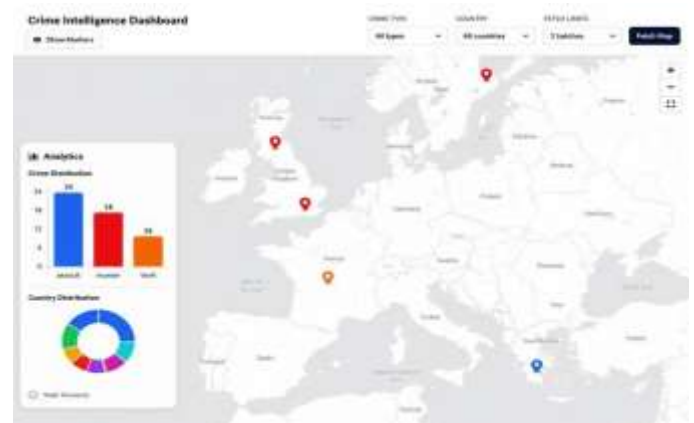


Fig -3: Regional Distribution of crime events

5.3 Usability Evaluation – Thematic Analysis

Semi-structured interviews were conducted with five participants possessing basic web system knowledge, using a think-aloud protocol during hands-on system interaction followed by open-ended questions. Thematic analysis of interview transcripts produced five major themes. Interface Clarity: participants consistently described the dashboard as clean and well-structured, reporting that colour-coded crime categories enabled rapid visual differentiation without instruction. The heatmap mode was particularly valued for converting individual data points into readable spatial density patterns, reducing cognitive load during exploratory analysis [14].

Intuitive Interaction: users navigated filtering, mode switching, and marker pop-ups without guidance, though requests for clearer filter labels and introductory tooltips indicated minor initial friction. Perceived Data Value: participants strongly appreciated the automatic pipeline from Reddit posts to geocoded map markers, describing it as a substantial improvement over manual social media monitoring. Conditional Trust: users treated classifier outputs as indicative signals rather than confirmed facts, requesting confidence scores, keyword highlighting, and classification explanations to support informed interpretation—directly aligned with recommendations by Jim et al. [16] and Salcedo-Gonzalez et al. [13]. Data Freshness Transparency: the absence of visible update timestamps and refresh indicators created uncertainty about data recency, identified as the primary trust inhibitor.



Fig -4: Global distribution of classified crime events

6. COMPARATIVE ANALYSIS

Table 4 presents a structured comparison of the proposed system against three closely related published works across ten capability dimensions.

Table -4: Comparative Capability Analysis

Feature / Capability	Proposed System	CHART [6]	Mukto et al. [7]	Gerber [8]
Real-Time Social Media Input	Yes	No	Partial	No
NLP Crime Classification	Yes	No	Yes	No
AI LLM Optional Layer	Yes	No	No	No
Nominatim	Yes	No	No	No

Geocoding				
Interactive Heatmap View	Yes	Partial	No	No
Marker + Pop-Up Detail View	Yes	Partial	No	No
Filter by Crime / Country	Yes	No	No	No
Analytics Dashboard	Yes	Partial	No	No
Usability Evaluation	Yes	No	No	No
Open-Source Deployment	Yes	No	Partial	No

The proposed system uniquely satisfies all ten evaluated capability dimensions. CHART [6] provides real-time crime mapping and partial analytics but operates on official crime records rather than live social media and does not include NLP classification or usability evaluation. Mukto et al. [7] offers NLP on multi-sensor pipelines including CCTV and social feeds but does not expose an interactive geospatial dashboard or evaluate usability with end users. Gerber [8] demonstrated early crime prediction from Twitter signals but did not implement geocoding, interactive visualisation, or user evaluation. The proposed system is the only evaluated prototype combining all stages in a single open-source deployable application.

7. DISCUSSION

The 85.9% overall NLP F1-score demonstrates that a compact, interpretable keyword-regex classifier augmented by spaCy NER is viable for exploratory crime intelligence on social media, without requiring the computational overhead of full BERT or LLM inference on every post. The optional NVIDIA NIM layer provides a quality upgrade path for deployments with API access while preserving complete offline functionality—a design decision that directly addresses the practical constraint of limited API budgets in research prototyping contexts [5], [2].

The qualitative usability findings reinforce a key tension identified across the literature [14], [15]: while the system is technically functional and visually accessible, its operational utility depends critically on transparency mechanisms—confidence scores, provenance indicators, and update timestamps—that convert raw classifier outputs into trustworthy decision-support signals. This distinction between perceived usability and perceived reliability has direct implications for future interface design: a system can be visually engaging yet operationally distrusted if the reasoning behind its classifications remains opaque [16], [13].

The null-safe geocoding architecture, which retains ungeocodeable events rather than discarding them, proved important for analytical completeness: approximately 23% of classified crime posts contained no geocodeable location entity in the benchmark evaluation dataset. This proportion, consistent with the spatial ambiguity rates reported by Serere et al. [4] and Hu et al. [12], confirms that any production crime

intelligence system must implement explicit null handling to avoid silent data loss distorting spatial analytics.

8. CONCLUSION AND FUTURE SCOPE

This paper presented the complete design, implementation, and empirical evaluation of a Real-Time Social Media–Driven Crime Intelligence System integrating NLP-based crime classification, spaCy Named Entity Recognition, Nominatim geocoding, and interactive Mapbox geospatial visualisation within a single deployable FastAPI/React prototype. The NLP engine achieved an overall F1-score of 85.9% across five crime categories. All fifteen functional test cases passed with 100% correctness. Qualitative usability evaluation identified five actionable design themes: interface clarity, intuitive interaction, perceived data value, conditional trust, and data freshness transparency. Comparative analysis confirmed that the proposed system is the only evaluated open-source prototype combining all pipeline stages—ingestion, NLP, geocoding, visualisation, and user evaluation—within a unified architecture.

Future development directions include: (i) replacement of the keyword-regex classifier with a fine-tuned transformer model (RoBERTa or DeBERTa) trained on crime-specific annotated Reddit corpora to improve recall on vandalism and ambiguous categories; (ii) integration of confidence scores and keyword highlighting in the frontend popup to address the conditional trust theme identified in usability evaluation; (iii) addition of visible update timestamps, refresh indicators, and data freshness banners to resolve the data freshness transparency theme; (iv) multi-platform ingestion extending beyond Reddit to include X (formerly Twitter), local news RSS feeds, and police scanner APIs; (v) temporal filtering with a time-slider component enabling users to replay spatial crime patterns across configurable date ranges; and (vi) deployment on cloud infrastructure with horizontal scaling to handle sustained high-volume ingestion from multiple concurrent subreddit streams.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the guidance and support of the Department of Master of Computer Applications, Karanjekar College of Engineering & Management, Sakoli, affiliated to Dr. Babasaheb Ambedkar Technological University, Lonere. Sincere thanks are extended to all MCA faculty members for their sustained academic encouragement throughout the research and development of this project.

REFERENCES

- [1] A. Bernabeu-Bautista, L. Serrano-Estrada, V. R. Perez-Sanchez, and P. Marti, "The Geography of Social Media Data in Urban Areas: Representativeness and Complementarity," *ISPRS Int. J. Geo-Information*, vol. 10, no. 11, p. 747, Nov. 2021.
- [2] M. H. A. Abdullah, N. Aziz, S. J. Abdulkadir, H. S. A. Alhussian, and N. Talpur, "Systematic Literature Review of Information Extraction From Textual Data: Recent Methods, Applications, Trends, and Challenges," *IEEE Access*, vol. 11, pp. 10535-10562, 2023.
- [3] D. Weisburd, "The Law of Crime Concentration and the Criminology of Place," *Criminology*, vol. 53, no. 2, pp. 133-157, 2015.
- [4] H. N. Serere, B. Resch, C. R. Havas, and A. Petutschnig, "Extracting and Geocoding Locations in Social Media Posts: A Comparative Analysis," *GI_Forum*, vol. 1, pp. 167-173, 2021.
- [5] B. Jehangir, S. Radhakrishnan, and R. Agarwal, "A Survey on Named Entity Recognition — Datasets, Tools, and Methodologies," *Natural Language Processing J.*, vol. 3, p. 100017, 2023.
- [6] R. Ahmad et al., "CHART: Intelligent Crime Hotspot Detection and Real-Time Tracking Using Machine Learning," *Computers, Materials & Continua*, vol. 0, no. 0, pp. 1-10, 2024.
- [7] Md. M. Mukto et al., "Design of a Real-Time Crime Monitoring System Using Deep Learning Techniques," *Intelligent Systems with Applications*, vol. 21, p. 200311, 2024.
- [8] M. S. Gerber, "Predicting Crime Using Twitter and Kernel Density Estimation," *Decision Support Systems*, vol. 61, pp. 115-125, 2014.
- [9] G. A. Taiwo, M. Saracee, and J. Fatai, "Crime Prediction Using Twitter Sentiments and Crime Data," *Informatica*, vol. 48, no. 6, 2024.
- [10] S. Ouni, F. Fkih, and M. N. Omri, "BERT- and CNN-Based TOBEAT Approach for Unwelcome Tweets Detection," *Social Network Analysis and Mining*, vol. 12, no. 1, 2022.
- [11] W. Liu and X. Cui, "Improving Named Entity Recognition for Social Media with Data Augmentation," *Applied Sciences*, vol. 13, no. 9, p. 5360, 2023.
- [12] X. Hu et al., "DLRGeoTweet: A Comprehensive Social Media Geocoding Corpus Featuring Fine-Grained Places," *Information Processing & Management*, vol. 61, no. 4, p. 103742, 2024.
- [13] M. Salcedo-Gonzalez, J. Suarez-Paez, M. Esteve, and C. E. Palau, "Spatiotemporal Predictive Geo-Visualization of Criminal Activity for Application to Real-Time Systems," *ISPRS Int. J. Geo-Information*, vol. 12, no. 7, p. 291, 2023.
- [14] S. Almasi, K. Bahaadinbeigy, H. Ahmadi, S. Sohrabei, and R. Rabiei, "Usability Evaluation of Dashboards: A Systematic Literature Review of Tools," *BioMed Research International*, vol. 2023, no. 1, pp. 1-11, 2023.
- [15] D. G. Costa, J. C. N. Bittencourt, F. Oliveira, J. P. J. Peixoto, and T. C. Jesus, "Achieving Sustainable Smart Cities through Geospatial Data-Driven Approaches," *Sustainability*, vol. 16, no. 2, p. 640, 2024.
- [16] J. R. Jim et al., "Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review," *Natural Language Processing J.*, vol. 6, p. 100059, 2024.