AI-Augmented Digital Forensics Assistant

Project Title : AI-Augmented Digital Forensics Assistant

Submitted By : Bhakti Janardan Indulkar

Institution : MCC, Mulund West, Mumbai

Organized by : Digisuraksha Parhari Foundation

Powered by : Infinisec Technologies Pvt. Ltd.

Date : 10 may 2025

Abstract

In the evolving field of cybersecurity and digital forensics, analysts are often challenged with identifying anomalous patterns, hidden threats, and extracting crucial metadata from files and images. The "AI-Augmented Digital Forensics Assistant" is an integrated tool that assists forensic analysts by automating these tasks using AI and open-source technologies. The system analyzes file metadata, detects log anomalies using machine learning (Isolation Forest), extracts EXIF data from images (including GPS coordinates), and integrates VirusTotal API for file threat intelligence. This tool empowers investigators with a single interface for efficient, intelligent evidence gathering and analysis.

Problem Statement & Objective

Manual digital forensic investigations are time-consuming and prone to oversight due to the vast number of logs, files, and multimedia content that must be reviewed. The objective of this project is to develop a unified AI-driven toolkit that assists digital forensics professionals by:

- Detecting anomalies in large system logs.
- Extracting file and image metadata for evidence.
- Mapping GPS data from image EXIF.
- Checking files against VirusTotal for malware indicators.

Literature Review

1. **Digital Forensics & Cybercrime** – Casey, E. (2011)
2. **AI for Log Analysis** – IBM Security Reports, 2022
3. **Machine Learning in Anomaly Detection** – Chandola et al., ACM Computing Surveys
4. **Forensic Use of Metadata** – NIST Special Publication 800-86
5. **EXIF GPS Location in Forensics** – Journal of Digital Forensics, Security and Law

6. **VirusTotal API for Malware Detection** – VirusTotal Documentation
7. **Metadata Spoofing in Cybercrime** – IEEE Transactions on Information Forensics
8. **Streamlit for Forensic Visualization** – Open Source Project Reports
9. **Image Forensics and EXIF Tampering** – Digital Investigation Journal
10. **AI Integration in Cybersecurity** – MIT CSAIL Technical Paper

---

Research Methodology

This project employs a modular approach:

- **File Analysis**: Python's os, hashlib, and magic libraries gather critical file properties like timestamps and hashes.
- **AI Log Anomaly Detection**: Machine learning with Isolation Forest identifies suspicious log entries.
- **Image Metadata Extraction**: Pillow is used to extract EXIF including GPS, and folium for geo-visualization.
- **Threat Intelligence**: VirusTotal API scans uploaded files' SHA256 for known threats. All modules are unified through a Streamlit GUI for interactivity and simplicity.

Tool Implementation

- **Language**: Python
- **Framework**: Streamlit for GUI
- **Modules**:
    - file_analyzer.py: Analyzes file size, type, hashes, and timestamps.
    - ai_log_analyzer.py: Reads CSV log files, detects anomalies using scikit-learn.
    - image_metadata.py: Extracts and maps EXIF data.
    - virustotal_checker.py: Queries VirusTotal with file hashes.
- **Dependencies**:

```
plaintext
CopyEdit
streamlit==1.35.0
pandas==2.2.2
scikit-learn==1.4.2
Pillow==10.3.0
python-magic-bin==0.4.14
folium==0.16.0
streamlit-folium==0.18.0
requests==2.31.0
```

Results & Observations

- Uploaded CSV logs flagged multiple anomalous IPs during simulation.
- File metadata matched OS-generated timestamps and calculated correct MD5/SHA256 hashes.
- GPS data was successfully extracted and visualized from sample images containing EXIF data.
- VirusTotal scans returned accurate file threat statistics (e.g., 2 malicious, 0 suspicious).

Ethical Impact & Market Relevance

Ethical Considerations:

- Avoid misuse of GPS metadata extraction and image tracing.
- Ensure VirusTotal API access complies with privacy guidelines.

Market Use Cases:

- Law enforcement digital evidence analysis
- Incident response teams
- Academic cybersecurity labs
- Custom SOC (Security Operations Center) tool integrations.

Future Scope

- Include PDF/DOC metadata and text extraction.

- Integrate facial recognition to correlate images with known suspects.
- Add real-time alerting features for log anomalies.
- Expand to mobile forensics capabilities (Android/iOS file systems).
- Offer cloud version for scalable forensic analytics.

References

1. Casey, E. *Digital Evidence and Computer Crime*
2. Chandola et al., *Anomaly Detection: A Survey* – ACM
3. NIST SP 800-86 – *Guide to Integrating Forensic Techniques*
4. VirusTotal – https://www.virustotal.com
5. Pillow Documentation – https://pillow.readthedocs.io
6. scikit-learn Docs – https://scikit-learn.org
7. Streamlit Docs – https://docs.streamlit.io
8. GPS Forensics – JDFSL, 2019
9. IEEE: *Metadata Exploitation in Malware*
10. MIT CSAIL – *AI in Cybersecurity Research*