Sarah D

Contact: beyond-features.com/contact

SUMMARY

Experienced Data Scientist with a strong background in security data, machine learning, and automation. Proven track record of developing and refining algorithms to detect and mitigate cyber threats. Skilled in developing data pipelines, automation workflows, and statistical models to drive security-focused insights and decision-making. Adept at working with large-scale cybersecurity data and integrating AI-driven solutions to enhance security operations.

EXPERIENCE

Senior Data Scientist | ThreatSTOP, Inc. | Dec 2024 - Present

- Leading the development and deployment of machine learning pipelines for phishing detection in DNS traffic on AWS.
- Architected a multi-stage phishing domain detection pipeline using an ensemble of Random Forest, Gradient Boosting, Logistic Regression, and Linear SVC models with character-level TF-IDF 3-gram features.
- Engineered a layered post-ML analysis framework combining WHOIS/DNS risk scoring, VirusTotal enrichment, DGA signals, and WHOIS anomaly detection to improve detection quality and reduce false positives.
- Designed solutions around real-world infrastructure constraints, optimizing shared VirusTotal API quotas while leveraging OpenSearch Serverless for deduplication and Amazon S3 for model and artifact storage.
- Built an Atlassian MCP connector enabling AI-powered access to Confluence documentation for internal knowledge retrieval.
- Enhanced threat intelligence by integrating new data sources into detection pipelines.
- Collaborated with security teams to operationalize AI-driven insights for threat mitigation.
- Managed scalable data pipelines supporting cybersecurity analytics and machine learning workloads.

Data Scientist / AI Engineer | Cisco Systems, Learning & Development | Nov 2023 - Nov 2025

- Led the Cisco U. Generative AI services team, driving the design and delivery of AI-powered capabilities that enhanced the learning experience.
- Designed, deployed, and maintained a production Retrieval-Augmented Generation (RAG) chatbot on AWS using enterprise knowledge sources.
- Built AI applications supporting semantic search, knowledge retrieval, and LLM-powered user experiences.
- Analyzed user behavior, application telemetry, and feedback to continuously improve retrieval quality and user experience.
- Contributed to a graph-based data catalog initiative supporting content recommendation systems through graph data models and metadata management.
- Led product quality initiatives by developing automated reporting, quality metrics, and data-driven processes across AI services.

Security Engineer | Cisco Systems, Talos | Mar 2016 - Oct 2023

- Developed machine learning models to improve phishing URL detection accuracy.
- Automated data analysis and threat intelligence reporting for large-scale cybersecurity datasets.
- Conducted statistical analysis on threat intelligence data to support cybersecurity operations.
- Created and maintained cyber threat intelligence databases.

EDUCATION

MS in Environmental Science & Policy, George Mason University (2013-2015)

BA in English & Business Economics, University of Kassel, Germany (2008-2012)

SKILLS

Programming: Python, SQL, Bash

Data Platforms: MySQL, Elastic Stack, Excel, Tableau, REST APIs

Cloud: AWS (S3, Redis, Pinecone, DocumentDB, PostgreSQL), CircleCI

Software: Docker, Kubernetes, Git

Operating Systems: Linux, Windows, macOS, BSD

Data Science: Machine Learning, Data Cleaning, APIs, RegEx, Statistics, Data Visualization

Generative AI: AI Agents, RAG, Chatbot Development