STREAM: A Framework for Sequence Data Analysis, Modeling, and Anomaly Alerts

Wan D. Bae[1], Shayma Alkobaisi ✉[2], and Pavleen Kaur[3]

[1] Computer Science, Seattle University, Seattle, WA, USA,
baew@seattleu.edu
[2] College of Information Technology, United Arab Emirates University, Al Ain, UAE,
shayma.alkobaisi@uaeu.ac.ae
[3] Geospaces.org, Denver, CO, USA,
pave@geospaces.org

Abstract. The increasing prevalence of time-series data in various domains necessitates efficient tools for real-time analysis and anomaly detection. Existing solutions often face challenges with scalability, usability, or domain-specific integration. To address these issues, we introduce STREAM, a modular framework that integrates data ingestion, ML model training, anomaly detection, visualization, and alert management within a user-friendly browser interface. STREAM utilizes advanced AI, including transformer models and expert systems, to ensure robust and interpretable monitoring. Key features—such as multi-user support, Docker-based scalability, metadata enrichment, and a searchable model repository—enable seamless data analysis and automation without programming expertise. STREAM's real-time capabilities make it a valuable resource for both academia and industry, streamlining AI workflows and delivering actionable insights.

Keywords: time-series data, anomaly detection, AI workflows

1 Motivation

The growth of sequence data, also called time-series data, in fields like IoT, finance, healthcare, and cybersecurity has created a need for robust analytical frameworks to process real-time, time-indexed data from diverse sources such as satellite telemetry, stock markets, health monitors, and manufacturing sensors. Managing and analyzing this data for actionable insights is a complex task, often requiring manual or semi-automated workflows involving highly trained experts. Detecting anomalies in the resulting stream of time sequence data and issuing relevant alerts and responses are extremely complex tasks, often performed manually or semi-manually, relying on experts trained for monitoring a particular aspect of the process.

Traditional tools, while effective in specific cases, often fall short in scalability, flexibility, or user accessibility [2]. Moreover, challenges such as integrating domain-specific metadata, building explainable models, and enabling real-time workflows remain inadequately addressed [3]. Time-sequence data analysis continues to benefit from advances in machine learning (ML) algorithms. Such algorithms have recently been deployed in some specific cases, e.g., monitoring of NOAA GOES-R weather satellites [1]. As the market for time-series analysis software is projected to grow from 1.8 billions in 2024 to 4.7 billions by 2031 [5], there is an increasing demand for a general-purpose solution that integrates data exploration, ML model training, anomaly detection, visualization, and automated alert management.

This paper introduces "STREAM", a comprehensive and user-friendly framework designed to address these challenges. STREAM is a modular and extensible platform for sequence data analysis, which addresses the full spectrum of analytical challenges in sequence data workflows presented in [4] and empowers users to assemble and visualize datasets, fine-tune ML algorithms, and establish autonomous workflows for anomaly detection and response, all without the need for advanced programming expertise.

STREAM utilizes state-of-the-art AI techniques, such as transformer-based ML models, to enable robust anomaly monitoring and nonlinear regression. To ensure safety and interpretability, a rule-based expert system operates in conjunction with ML algorithms, merging the flexibility and power of ML with the reliability of formal logic. Key features include an intuitive graphical user interface (GUI) for data exploration, metadata enrichment, and visualization, as well as scalable deployment options like Docker environments and multi-user access.

2 System Framework

STREAM is a framework for anomaly monitoring and detection in time-series data, designed with a browser-based graphical user interface (UI) and a flexible, automated backend. Deployed via Docker, it supports Single Sign-On (SSO) and organizes data into missions, which group related datasets for streamlined management. Each dataset comprises time-series files with timestamped rows and labeled variables (e.g., sensor names). As STREAM facilitates data exploration through sorting, filtering, graphing, and statistical analysis, it offers users quick and actionable insights. The system consists of the following four main layers as depicted in Fig. 1:

1. User interface: Displays maps, data collections, 3D models, and geographic locations. It allows users to load, view, and analyze datasets, including specific data with meaningful patterns.
2. Data and model management: Handles data ingestion and integration, feeding data into the UI, backend for further processing, and model management.
3. Algorithms: Implements a range of machine learning algorithms for corelation analysis, prediction, and anomaly detection.
4. Data storage: Manages storage using local flat files, Parquet-formatted files, Elasticsearch, or links to external storage with appropriate access permissions.

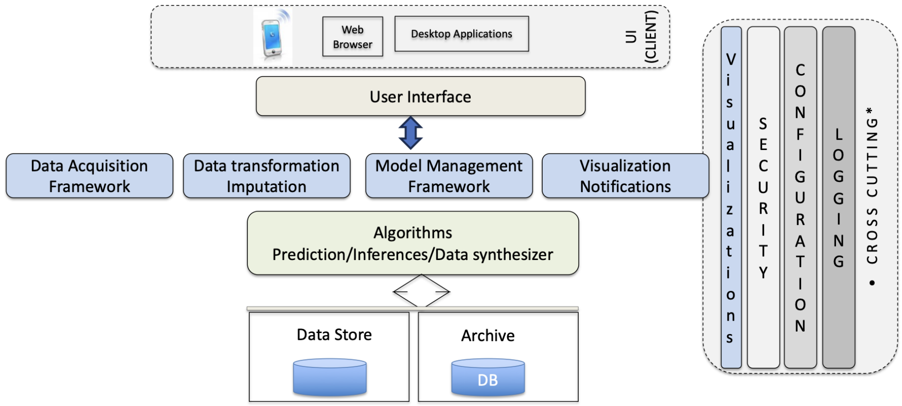


Fig. 1. An overview of STREAM

To enrich the context, STREAM allows annotation of variables with descriptive names, units, ranges, subsystem details, and 3D coordinates. These annotations are managed in editable dictionaries and are propagated system-wide with localized priorities.

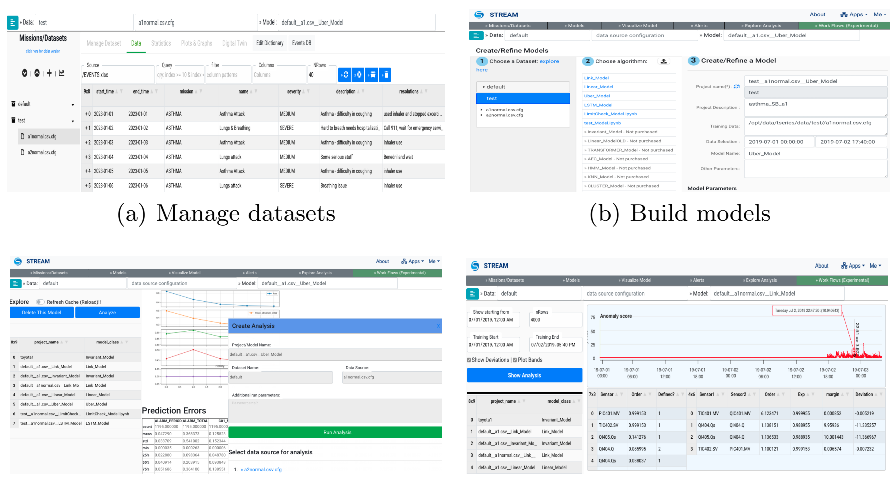
The framework includes robust statistical and machine learning algorithms for analyzing correlations, prediction models, and detecting anomalies. Trained models are systematically named, stored, and managed in a searchable repository. These models generate anomaly scores by comparing predicted and actual values, with scores normalized for consistency. Anomalies trigger alerts composed of three components: (1) Triggers: Conditions for activation (e.g., exceeding an 80% anomaly score within a 4-hour window), (2) Notifications: Communication actions, such as sending emails or messages, and (3) Responses: Prescribed actions, from manual recommendations to automated processes.

Critical alerts can be flagged as events, which capture snapshots of contributing variables and enabling user annotations for deeper analysis. Iterative model validation supports continuous refinement, allowing users to test models on new datasets, review anomaly scores, and adjust alert criteria. STREAM also visualizes correlations between variables and subsystems, with ongoing development of enhanced 3D spatial visualization tools.

In live monitoring scenarios, STREAM processes real-time data streams, displaying anomaly scores and managing alerts in a “situation room”. Users can deploy and monitor multiple models concurrently, making STREAM a versatile solution for dynamic anomaly detection and monitoring.

3 Demonstration

As STREAM is a critical tool for both technical and non-technical users across academia and industry, we demonstrate its main functionalities: (1) manage data, (2) build AI models, (3) explore models, (4) create analysis, and (5) monitor and alert. For the currently selected project, data file and model should be displayed under the main menu. The user can easily select a different project, data file and model by clicking on the displayed names. Prototypes of STREAM are shown in Fig. 2. and A demo video provides an overview of the system. Users can also access the framework development tools.



(a) Manage datasets

(b) Build models

(c) Create model analysis

(d) Explore models

Fig. 2. STREAM: data management, modeling development and analysis

References

1. Spacenews: How ai improves operations on noaa goes-r weather and environmental satellites. https://spacenews.com/ai-already-improving-operations-noaa-goesr-weather-environmental-satellites/, accessed: September 2024
2. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM computing surveys (CSUR) **41**(3), 1–58 (2009)
3. Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T.: Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 387–395 (2018)
4. Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Xing, T., Yang, M., Tong, J., Zhang, Q.: Time-series anomaly detection service at microsoft. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 3009–3017 (2019)
5. Research, V.M.: Global time series analysis software market size by product, geographic scope and forecast. https://www.verifiedmarketresearch.com/product/time-series-analysis-software-market/, accessed: September 2024