Urban Data Labs

Proposal for Real-Time Anomaly Detection with Smart Building Data

Master of Data Science Capstone

Nathan Smith, Mitch Harris, Ryan Koenig
5-9-2021

Abstract

Proposal to develop a real-time anomaly detection system with smart building sensor data for Urban Data Lab (UDL). Data is managed by UDL in both InfluxDB and Azure which gather sensor data from the University of British Columbia (UBC) Vancouver Campus Energy and Water Services (EWS) ION and SkySpark databases. The real-time anomaly detection system will be initially implemented with InfluxDB and Python and then potentially twinned to operate with the Azure system. A dashboard will allow users to visualize detected anomalous data. The detection system will provide campus and building managers with the data-driven insight to monitor their systems' operations and predict maintenance issues, thus, reducing operational costs, downtime, and unexpected maintenance.

Background

Urban Data Lab

The Urban Data Lab (UDL) was founded in 2019 to advance data access, data management and data analytics capabilities on the University of British Columbia (UBC) campus with the goal of addressing campus-wide sustainability challenges. It accomplishes this by providing open access of UBC sustainability data to researchers, policymakers and operational staff. It also supports the monitoring and measurement of sustainability performance for buildings, transportation, and specifically as it relates to the policy commitments of UBC Sustainability Initiative and Campus and Community Planning. (Wei J.)

UBC ENERGY AND WATER SERVICES

UBC Energy and Water Services (EWS) manages and operates UBC's utility assets, delivers utilities master plans and load forecasts, and associated capital upgrades. Established in 2014, the unit was created as a dedicated centre of energy and utility expertise at UBC. Its mission is to generate, distribute and conserve the Vancouver campus' energy and water resources in an open and transparent manner that maintains environmental comfort, ensures system reliability, minimizes life cycle cost, reduces greenhouse gas (GHG) emissions, and seeks out innovative solutions. (UBC Energy and Water Services) EWS has taken two approaches for dealing with UBC's heating needs through the hot water boiler facilities Campus Energy Centre (CEC) and Bioenergy Research Demonstration Facility (BRDF). This allowed EWS to track and store information about energy consumption into two databases, ION and SkySpark.

ANOMALY DETECTION

UDL gained access to EWS's ION and SkySpark databases in 2019 and recorded data to their InfluxDB instance. InfluxDB is a multi-tenanted time series database, UI and dashboarding tools, background processing and monitoring agent. UDL noticed some of the data streaming from the meters appears to be inconsistent with the surrounding data and there is no system currently in place to catch or flag these data. According to UDL, only some EWS sensors are monitored regularly for unusual behavior and being able to have autonomous monitoring on all sensors would be advantageous. UDL are interested in deployment of an anomaly detection system operating on multiple data streams and notifying users of unusual behavior. Unusual behaviour may include SkySpark server interruptions, sensors disconnecting, interruptions in the in the server-to-InfluxDB streaming process, and unusual sensor readings, potentially indicating a maintenance-related concern.

LITERATURE REVIEW

A high-level literature review was completed to identify initial methods that could be explored for streaming sensor data anomaly detection. Anomaly detection of streaming data appears to be an active area of research and a more detailed literature review will be required as part of the study to identify candidate approaches. Challenges in streaming anomaly detection include selecting the window (timeframe) of data used to analyze anomalies, use of labelled data for supervised methods, and changing trends in data with time. Current research addresses these challenges, with some methodologies attempting to solve multiple challenges at once. For example, Hierarchical Temporal Memory models, provides an unsupervised approach on streaming data that learns on a continuous basis (Subutai Ahmad, 2017). Another recent method is an Online Evolving Spiking Neural Network (Piotr S.Maciąg, 2021), which similarly provides unsupervised anomaly detection. Both of these approaches do not require labelled data which would be a challenge for the UBC SkySpark dataset. Accordingly, it is anticipated that an unsupervised approach capable of learning on a continual basis will be used. Semi-supervised approach requiring minimal labelled data will also be considered.

Dataset

EWS collects data from 370 energy meters, 5000 HVAC equipment and three Building Management Systems. The EWS ION server provides data on Energy, Power, Gas, and Water consumption of buildings on the UBC Campus. The EWS SkySpark database provides data recorded by smart meters and smart devices in UBC buildings every 15 minutes. UDL collects data from these servers and stores the data in InfluxDB. (Wei J. , 2020). The ION server provides updates every 5 seconds and is monitored by UDL through an OPC Router which updates with every 2kW change in values. UDL currently only has access to historical SkySpark data but anticipate access to real-time data shortly.

A subset of EWS data points available on SkySpark will be initially used to evaluate and select the streaming anomaly detection model. Sensors and meters in the Campus Energy Centre (CEC) and Bioenergy Research Demonstration Facility (BRDF) will be used for the subset. A streaming simulation using the historical data will be initially completed for the study as the UDL InfluxDB currently only has historical SkySpark data.

Objectives

Goal

The goal of the project is to develop a streaming anomaly detection system for the EWS smart sensor network. The system should allow users to understand when a data point reading is faulty such that the data can be safely ignored or removed from analyses and sensors can be fixed or replaced. The anomaly detection system will include a dashboard for visualization and will ideally provide automated notifications. The primary goal is to implement the streaming system with InfluxDB and a secondary goal is to twin the system in Azure.

An additional scope that can be explored depending on the project timeline includes linking the anomaly detection system with sensors directly, through Azure's digital twin (via IoT Hub) or Over-the-Air communications, depending on sensors' systems. This would be done via TinyML-edge AI (embedded

machine learning/ machine learning on embedded systems) and the models would need to be optimized/quantized to run on low-power/low-storage IoT devices.

Methodology

The proposed methodology for the project is listed below. Note that the approach provides several items that are dependent on the project progress/timeline including twinning the system in Azure and implementing the system on additional data sources.

- A subset of building sensor data from the historical SkySpark (not real-time) database in InfluxDB will be selected for initial study. The subset will be focused on the sensor data coming from both the CEC and BRDF buildings.
- An offline anomaly detection model will be created in Python based on the subset of data. Various machine learning approaches will be explored for the model.
- Streaming data will initially be simulated in Python to test the anomaly detection model using the subset of historical SkySpark data. It is understood that a real-time version of the SkySpark in InfluxDB should be in operation shortly. The streaming anomaly detection model can then be implemented with the real-time database.
- Twinning the system in Azure can be explored/implemented after the InfluxDB streaming anomaly detection model is functional, depending on the project timeline.
- The anomaly detection model can be tested and implemented on a larger set of streaming sources depending on the project timeline.
- Output from the anomaly detection model will be provided in a dashboard. This may include dashboards for both the InfluxDB and the Azure systems depending on the success of twinning the system. The InfluxDB dashboard will use either Grafana, Dash, or Power BI while the Azure dashboard would likely use Power BI.

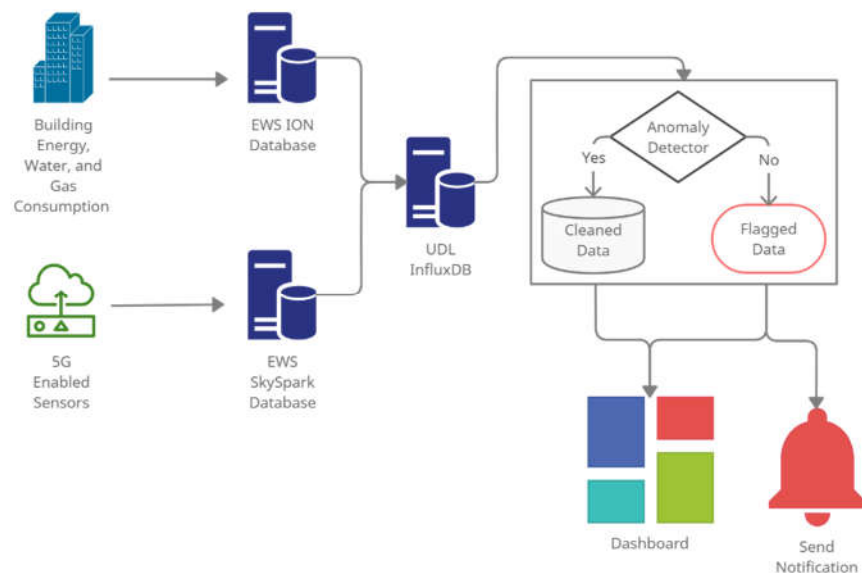A high-level schematic of the system is provided below.

FIGURE 1: SCHEMATIC OF DATA FLOW.

Statement of Work

The project scope is divided into weekly goals over the seven-week project schedule. A description of the proposed goals is provided in the following sections and a Gantt chart of the schedule is provided in Appendix A. It is expected that the timeline and goals will evolve throughout the project, but the initial scope provides a basis for tracking progress and understanding when adjustments are required.

A summary of project deliverables include:

- Proposal: May 9
- Proposal Presentation: May 11
- Status Presentation: June 1
- Final Report and Presentation: June 22

Week 1 – Project Definition and Proposal

The first week of the project includes meeting with UDL, understanding the project, and writing the proposal. A list of tasks included in this week (now complete) are provided below:

- **Select/Define Project**: Defining the project based on various options provided by UDL.
- **Create Project Management Tools:** Set up project templates, file sharing, communication, and project management tools.
- **Write Project Proposal:** Includes completing this proposal document due Sunday May 9 and a presentation for MDS course requirements due Sunday May 11.
- **Project Background/Data Review:** Review of preliminary project background information provided.

Week 2 – Data and Systems Understanding

Week 2 of the project includes understanding the UDL data and systems, learning required tools, and starting anomaly detection research. Tasks include:

- **Select Project Data and Exploratory Data Analysis**: A recommended subset of data will be discussed with UDL and an exploratory analysis will be completed.
- **Understand Data Systems:** A review of data systems will be completed including gaining a better understanding of the ION and SkySpark databases.
- **Learn InfluxDB:** Streaming anomaly detection will be initially completed using the InfluxDB framework as requested by UDL. The MDS team will need to learn this tool.
- **Simulate Streaming in InfluxDB:** The data within the UDL InfluxDB is currently not streaming and will need to be initially simulated for the study. This task also includes understanding how a streaming anomaly detection system could work with InfluxDB.
- **Anomaly Detection Research:** Literature review of anomaly detection methods in an online streaming approach. A short-list of methods will be compiled.

Week 3 – Anomaly Detection Model

Week 3 focuses on completing the anomaly detection model. Tasks include:

- **Meet with Domain Experts**: Anomalous data discussions with experts including Technical Safety BC, SkySpark specialists from EWS, and CEC/BRDF.

- **Data Cleaning/Feature Steps in Pipeline**: Required data cleaning and feature engineering to support the anomaly detection model within a pipeline.
- **Build Anomaly Detection Model:** The short-list of anomaly detection methods will be explored, and the best method will be selected for the project. It is not anticipated that the model will be implemented in the pipeline this week.
- **Review Azure:** Another time dependent goal of the project is implementing the streaming anomaly detection approach in Azure. An initial exploration of how this could be done will be discussed with UDL.

Week 4 – Implement Streaming Pipeline

The goal of this week is to complete the streaming anomaly detection implementation in InfluxDB.

- **Streaming Anomaly Detection Pipeline:** It is anticipated that the streaming pipeline without the anomaly detection model will be in place at the end of Week 3. This task includes implementing the anomaly detection model in the pipeline.
- **Model Tuning and Evaluate Performance:** This may be done outside of the streaming pipeline and will ideally include evaluation on datasets outside of the subset selected in Week 2.
- **Status Presentation:** A status presentation as part of the MDS course requirements is due June 1. The presentation will be completed this week.

Week 5 – Value Week

The purpose of this week is to complete the set of tasks that provide the most value given the current state of work. This will be discussed with UDL and may include one or more of:

- **Finish Outstanding Week 2-4 Tasks:** Depending on progress, additional time may be required to complete the streaming anomaly detection model in InfluxDB.
- **Implement in Azure:** Implement the streaming anomaly detection model in Azure.
- **Scale to Additional Data Systems:** Review the ability to scale the model to other data systems and implement as possible.
- **Scale to Additional Data Systems:** Review the ability to scale the model to other data systems and implement as possible.
- **Linking the Anomaly Detection directly with Sensors**: Explore/implement TinyML-edge AI.
- **Improve Anomaly Detection Model:** Time could be allocated to improve the model depending on the performance evaluation completed in Week 4.
- **Start Dashboard:** Week 6 is allocated for work on the anomaly detection dashboard. Starting the dashboard this week would provide additional time for a higher quality dashboard.

Week 6 – Dashboard

The goal of Week 6 is to complete the anomaly detection model dashboard. The InfluxDB dashboard will be prioritized. The dashboard in Azure can be considered depending on project progress.

Week 7 – Reporting

Reporting and project wrap-up will be completed in Week 7. Tasks and deliverables include:

- **Final Project Report:** The final report for the MDS program is due on Tuesday June 22. A draft report for review will be provided to UDL on Thursday June 17.
- **Final Presentation:** The final presentation for the MDS program is due on Tuesday June 22. The final presentation will be provided to UDL on Friday June 18.
- **Project Wrap-up**: Final project documentation and code will be wrapped up. Hand-off to UDL will also be completed.

Project Management and Roles

Project management will be completed using a variety of tools including Kanban boards on Clubhouse.io for task allocation, OneDrive for document sharing with UDL, Slack for communication, and GitHub for internal MDS Team documents and code organization. Weekly scrum meetings will be held with UDL to allocate and discuss weekly goals and tasks. Technical meetings will be held with UDL on an as needed basis and it is expected this will occur once per week.

Project management responsibilities will be shared by all MDS team members, but the following roles have been identified:

- **Nathan Smith:** Project management and communication
- **Mitch Harris:** Code management
- **Ryan Koenig:** Research material and reference management

Conclusion

The MDS Team looks forward to working with UDL on the streaming anomaly detection for smart building data. The goal of the project is to provide an implementation or approach that can be used for real-time notification of anomalous data. The system will allow users to visualize the anomalous data on a dashboard and assess if readings can be safely ignored or removed from analyses. It should also provide an indication of sensors that require more frequent monitoring, repair, or replacement.

References

Piotr S.Maciąg, M. K. (2021, July). Unsupervised Anomaly Detection in Stream Data with Online Evolving Spiking Neural Networks. *Neural Networks, 139*, pp. 118-139. Retrieved from https://www.sciencedirect.com/science/article/pii/S0893608021000599

Subutai Ahmad, A. L. (2017, November). Unsupervised real-time anomaly detection for streaming data. *Neurocomputing, 262*, pp. 134-147. Retrieved from https://www.sciencedirect.com/science/article/pii/S0925231217309864

UBC Energy and Water Services. (n.d.). About Us. Retrieved May 2021, from https://energy.ubc.ca/about-us/

UBC Energy and Water Services. (n.d.). Projects. Retrieved May 2021, from https://energy.ubc.ca/projects/

UBC Energy and Water Services. (n.d.). SkySpark. Retrieved May 2021, from
https://energy.ubc.ca/projects/skyspark/

Wei, J. (2020, January 9). UBC Building Data. Retrieved May 2021, from
https://urbandatalab.io/news/ubc-building-energy-data/

Wei, J. (n.d.). *About.* Urban Data Lab. Retrieved May 2021, from https://urbandatalab.io/about

Appendix A - Project Schedule

TASKS	WEEK	May 3, 2021 Project Definition and Proposal	May 10, 2021 Data and Systems Understanding	May 17, 2021 Anomaly Detection Model	May 24, 2021 Implement Streaming Pipeline	May 31, 2021 Value Week	June 7, 2021 Dashboard	June 14, 2021 Reporting
		3 4 5 6 7 / M T W R F	10 11 12 13 14 / M T W R F	17 18 19 20 21 / M T W R F	24 25 26 27 28 / M T W R F	31 1 2 3 4 / M T W R F	7 8 9 10 11 / M T W R F	14 15 16 17 18 / M T W R F / 21 22 / M T
Project Definition and Proposal								
Select/Define Project	1	███						
Create Project Management Tools	1	███						
Write Project Proposal	1	███						
Project Background/Data Review	1	███	███					
Data and Systems Understanding								
Select Project Data and Complete EDA	2		███					
Understand Data Systems	2		███					
Learn InfluxDB	2		███					
Simulate Streaming in InfluxDB	2		███					
Anomaly Detection Research	2		███	███				
Anomaly Detection Model								
Meet with Domain Experts	3			███				
Data Cleaning/Feature Steps in Pipeline	3			███				
Build Anomaly Detection Model	3			███				
Review Azure	3			███	███			
Implement Streaming Pipeline								
Streaming Anomaly Detection Pipeline	4				███			
Model Tuning and Evaluate Peformance	4				███			
Status Presentation	4				███			
Value Week								
One or More of the Items in Notes	5					███		
Dashboard								
Create and Complete Dashboard	6						███	
Reporting								
Final Project Report	7							███
Final Presentation	7							███
Project Wrap-up	7							███

Note: Value week will consist of one or more of the following based on priorities
1) Finish Outstanding Week 2-4 Tasks
2) Implement in Azure
3) Scale to Additional Data Systems
4) Explore/Implement TinyML-Edge AI
5) Improve Anomaly Detection Model
6) Start Dashboard