

# HVAC Anomaly System

Ganesh Nehru  
*Dept. of Computer Engineering*  
*San Jose State University*  
San Jose, CA  
ganesh.nehru@sjsu.edu

Pawan Kalyan Jonnalagadda  
*Dept. of Computer Engineering*  
*San Jose State University*  
San Jose, CA  
pawankalyan.jonnalagadda@sjsu.edu

Himaja Narina  
*Dept. of Computer Engineering*  
*San Jose State University*  
San Jose, CA  
himaja.narina@sjsu.edu

Mohamed Shafeeq Usman  
*Dept. of Computer Engineering*  
*San Jose State University*  
San Jose, CA  
mohamedshafeeq.usman@sjsu.edu

Bhavana Gangula  
*Dept. of Computer Engineering*  
*San Jose State University*  
San Jose, CA  
bhavana.gangula@sjsu.edu

**Abstract**—The use and number of IoT-based HVAC systems are rapidly increasing and will account for a significant portion of IoT-based household devices in the near future. These devices' operational and usage logs contain various sensor values logged over time, including normal data points and long-term anomalies. Modern anomaly detection methods are incapable of detecting these long-term anomalies which represents the deterioration of a sensors. This paper presents an scalable anomaly detection system processing time-series data from the Internet of Things (IoT)-based household devices in general and HVAC systems. We also present a pattern-based approach for anomaly detection in HVAC time-series data due to the number of issues observed during the evaluation of widely used distance-based, statistical-based, and cluster-based anomaly detection techniques. The presented approach solves this problem by constructing a knowledge base of long/short-term patterns based on regular data points that grow over time.

**Index Terms**—scalable, anomaly detection, HVAC, IoT

## I. INTRODUCTION

HVAC(Heating, ventilation, and air conditioning) systems are crucial in providing appropriate environmental conditions in terms of temperature, humidity, pressure, and flow rate while being energy efficient. With the rapid growth of IoT-based HVAC systems, it becomes crucial to detect short-term anomalies and long-term anomalies in heating and cooling, and air quality. To build a HVAC solutions for diverse use cases - ranging from monitoring HVAC systems in mission critical labs to residential, it becomes important to make it accessible to non-experts too. This becomes more challenging with large amounts of data. To solve this, we are building a system to scalability ingest HVAC IoT data and perform anomaly detection on it. This will cover

- 1) Ingestion of large volumes of data and it's efficient storage
- 2) Analysis of historical data and performing predictive maintenance on HVAC device
- 3) Real-time analysis on the incoming streaming data for anomalies detection and alerting

## II. PERSONAS

### A. Technician

As an HVAC technician it is his responsibility to keep things up and running and sometimes it is difficult to predict the downtime of the HVAC system so this tool can notify of any deviations out of ordinary by processing the sensor data generated by the HVAC sensors using advanced machine learning techniques.

## III. APPLICATION

The application is hosted on:

<https://main-majestic-syrniki-4d8693.netlify.app>

Link to GitHub repo:

<https://github.com/sjsucmpe272SP22/HVACAnomaly>

### A. Owner

As a business owner you want to ensure HVAC machines are running at their peak performance and don't want any downtimes that may affect the business. So, this tool can help to achieve that goal the smart solutions don't require any technical expertise in that area using the latest techniques one can monitor his HVAC systems on an easily readable dashboard.

## IV. METHODS

### A. Data Gathering

In order to produce a model with usable and valuable insight, the initial data must be of a high quality and actually substantial. In our case, HVAC IoT Anomaly detection is not that common of an issue so a decent data set was not easily accessible or even present on most open source data set providing websites. One option we found was an open source training data set used by some HVAC related study found on GitHub, but there was no context to the data and it also only dated back to three or four months when the ideal would be about a year's worth of time. After some more careful research, we were able to find a promising

data set on IEEE DataPort, a data platform developed and maintained by IEEE organization. This data set was collected by Grigore Stamatescu, a professor from University "Politehnica" of Bucharest, an educational facility based in Bucharest, Romania. The data set itself contains measurements taken from four air handling units installed in a medium-to-large size academic building. The building, according to the information provided, contains various conference rooms, research labs, administrative offices, and an auditorium. On-site cooling is handled by electric chillers while heating is provided by district heating. The sensors within the building had time series data recorded into the data set and were themselves placed into various locations within and outside the building. Sensors 1 and 2 were located on top of the building itself and were responsible for monitoring some of the higher level floors which mainly consisted of research labs. Meanwhile, the other sensors covered the bottom floors of the building which contained administrative and multi functional spaces. The time series data is from about a year of recording from these various sensors.

The only issue with the data was accessibility since it was locked behind an IEEE Dataport membership which was an obstacle for us. We managed to overcome it by just going straight to the source and contacting Grigore who was then considerate enough to provide the data himself to us.

### B. Data Analysis

Analyzing the sensor data, we found some daily, monthly and weekly patterns in it. The usage pattern from 10:00 to 20:00 very higher than the rest of the time across all sensors. Wednesday consistently saw lesser utilization and there were peaks in temperature in Summer.

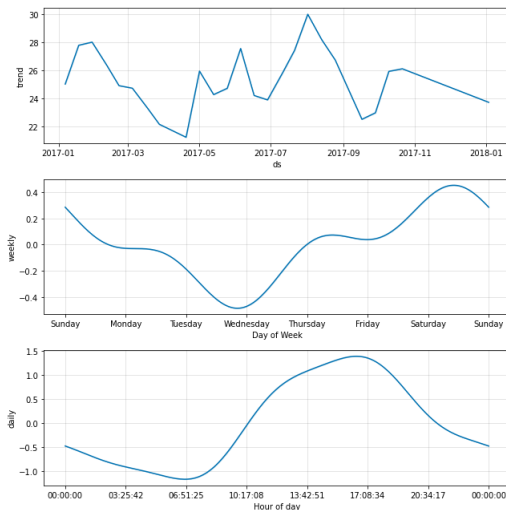


Fig. 1. Temperature Component Analysis

### C. Prediction

Post analyzing the data, multiple prediction techniques like ARIMA, SARIMAX were tried. The predictions from RNN

based model using Prophet gave most accurate fit and forecast and was able non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

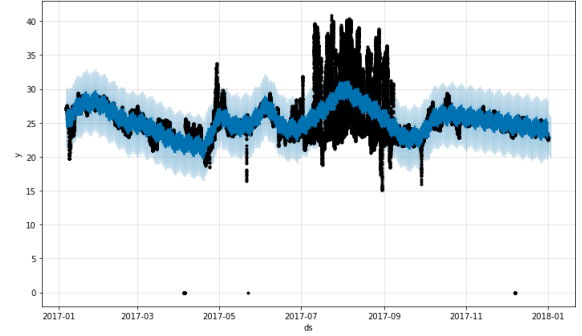


Fig. 2. Time-series analysis(Prophet)

### D. Key Metrics not sure what to title this

We developed KPIs after analyzing the data, and researching on important HVAC metrics.

1) *Real-time*: One advantage of using the data visualization platform, Tableau, in terms of real time data analysis is that it allows connections to various data sources and in our case we were able to connect to our PostgreSQL db known as TimescaleDB. This meant Tableau would be able to continuously, on some time based interval, update visualized data analysis based on new data being produced to the data source. Once the time series data was processed and classified in the data source, it was ready to be analyzed through the tools provided by Tableau.

One of the primary metrics used for measuring sensor performance in terms of anomalous values was anomaly percentage. Anomaly percentage was calculated by counting the total number of temperature measurements categorized as anomalies divided by the total number of temperature values recorded, all per sensor. The radial gauge meters within our dashboard displayed the calculated anomaly percentage along with the ID of the sensor and the location of the actual physical sensor for quick reference. The meter is designed to give the application user a quick and easily digestible insight into how each sensor is performing at the time they monitor the information. Once the meter reaches a certain threshold value it also changes visually to represent that by having the meter turn red which is supposed to indicate that the given sensor has exceeded the expected anomaly percentage and is now a cause for concern. This is done for each sensor on the same summary page so that someone that isn't as proficient in technology or statistical analysis like a property owner of a building with an HVAC system could use this application easily without having to do much research. A more in depth view of each sensor's information is also available just by clicking on the meter visual for the respective sensor.

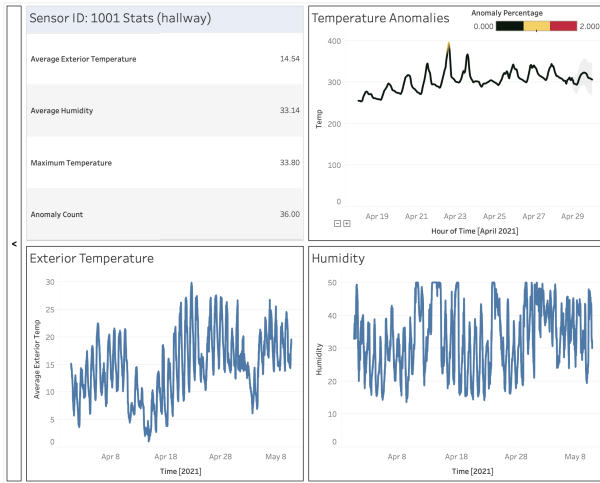


Fig. 3. In depth sensor analysis

The in depth sensor analysis for each sensor is split into four different sections on the respective pages. The top left section has statistics that are of potential interest to the user of the application: average exterior temperature, average humidity, maximum recorded temperature, and total anomaly count. The top right section is a particularly detailed graph of temperature versus time where temperature is colored based on its anomaly categorization. A specific temperature value is designated as an anomaly on a scale of 0 to 2 based on the severity of the anomaly in terms of variation from the norm. On the graph, highly anomalous temperatures are colored red while somewhat anomalous temperature values are yellow. There is also the option within the embedded Tableau to setup alarms that alert the application user whenever anomalies reach over a specific temperature threshold via email warnings. An application user is also able to hover over individual values on the graph to be able to see what those exact anomalies temperature values are along with accurate timestamps. This allows technically knowledgeable users to pinpoint the exact time and place anomaly events are occurring. Another feature of the temperature anomaly graph is a prediction of the temperature values for the next 25 hours after the end of the current scope of time data available. This is pictured on the graph by a highlighted section of the graph on the right side where the size of the highlight around the actual predicted line of values represents the confidence interval of potential values.

The bottom two graphs represent exterior temperature versus time and humidity versus time which could be useful in determining if a sensor is functional.

There is also an executive summary portion to our Tableau visualization. This section is designed with a number of sensor potential scalability in mind. It presents similar information to the main page but in a more condensed and tweak-able view. The anomaly percentage on this page also has a more diverse range by including a yellow coloring which indicates the level of anomalies within a particular sensor are not at a hazardous

level but they are something to be wary of. This view is also equipped with various filtering and sorting drop-down options which would be very useful on large scale sensor networks that each need to be actively monitored for anomalous behavior.

### E. Ingestion Pipeline

The data from the sensors gets ingested into the cloud via GCP Functions. Streaming the data using a serverless architecture helped reap benefits of elasticity with uniform data flow pattern and high-velocity bursty data pattern. The system was tested with bursty high-velocity data like seen on some sensors - the serverless functions were able to autoscaling upto 100 to support high velocity ingestion and scaling down. There are two parts of the system

1. Detection of real-time anomalies The algorithm learns the pattern from the data and detects real-time anomaly using Z-Score (with millisecond latency), The current alert is sends through an email notification.



Fig. 4. Long term pattern analysis from historical data

2. Prediction of future breakdowns

Every night, a cron job analyzes all the past data upto then and makes prediction for the next day using the RNN based Prophet model. We have the ability to convert this numerical prediction into meaningful messages that a non-expert can understand to send meaningful alerts.

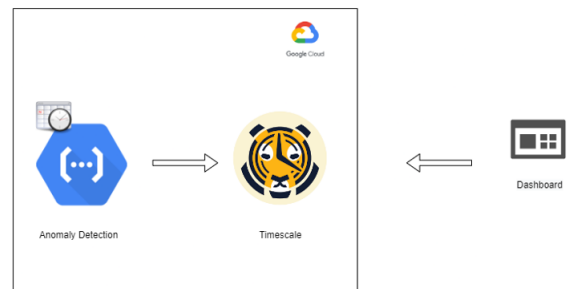


Fig. 5. Real-time anomaly detection

## F. Reporting and Frontend

The front end was implemented in ReactJS. The reporting consists of a landing page accessible to everyone who wishes to learn more about HVAC AnoML. The architecture diagrams have also been embedded for all to see. For registered users, there exists a login page. After login, one can access the dashboard.

## FUTURE SCOPE

One aspect of our project we would like to improve in the future is to provide more feedback and analysis based on the data collected. Ideally we would be able to detect anomalies in measurements like humidity and external temperature which would further assist in diagnosing issues in potentially faulty sensors. Another feature we would like to expand on further is our anomaly prediction since at the moment it has room for improvement in terms of further future predictions and also integrating it with our real time analysis dashboard. An additional component that could be expanded in our project would be our modeling algorithm. Instead of using z-score we could potentially use unsupervised ML learning models like Random Cut Forest that could be trained and then tested. This would provide us a more robust method of categorizing measurements as anomalies and it would lessen the amount of false positives we potentially have at the moment.

## CONCLUSIONS

This paper presents a thorough evaluation of anomaly detection techniques on a real IoT-based HVAC dataset as well as a visual evaluation on an data set. To address various issues that arise in traditional anomaly detection techniques, such as inability to detect changes in pattern and finding the precise number of clusters, a new technique for detecting anomalies based on context and history is also presented. HVAC system operational logs frequently contain long-term anomalies. It is critical to consider contextual information when detecting these long-term anomalies. Due to a lack of contextual knowledge, state-of-the-art anomaly detection methods are unable to detect these anomalies. The presented approach solves this problem by constructing a knowledge base of long/short-term patterns based on normal data points that grows over time. Furthermore, it does not analyze a single data point but considers its context to determine whether the current point is normal or abnormal. The detection of these trending long-term anomalous points in real time can be used to provide early warnings about the faulty sensor/component before the entire system fails. Aside from detecting anomalies, the presented method also generates a meaningful anomaly score. This means that it assigns a higher anomaly score to more significant anomalies and a lower score to less significant anomalies. The anomaly score generated by the existing methods for the HVAC dataset is meaningless and has no correlation with the significance of the anomalies.

## ACKNOWLEDGMENT

We would like to thank professor Rakesh Ranjan for the inspiration to build our project around the idea of HVAC IoT anomaly detection. We would also like to thank professor Grigore Stamatescu from Romania for generously providing an extensive data set for our purposes.

## REFERENCES

- [1] Building intelligence into buildings  
<https://www.ibm.com/downloads/cas/2GYNP5R9>
- [2] Grigore Stamatescu, Iulia Stamatescu, Nicoleta Arghira, and Ioana Fagarasan, "Data-Driven Modelling of Smart Building Ventilation Sub-system," *Journal of Sensors*, vol. 2019, Article ID 3572019, 14 pages, 2019. <https://doi.org/10.1155/2019/3572019>.
- [3] Pattern-Based Contextual Anomaly Detection in HVAC Systems  
<https://ieeexplore.ieee.org/document/8215784>
- [4] Anomaly Detection in the HVAC System Operation by a RadViz Based Visualization-Driven Approach  
[https://link.springer.com/chapter/10.1007/978-3-030-42048-2\\_6](https://link.springer.com/chapter/10.1007/978-3-030-42048-2_6)