Anomaly Detection in Air Quality Monitoring Networks

Adhikari Durga Venkata Madhav, Amanchi Sravan Kumar, Addanki Gargeya, Ambati Vinod and V Ragavarthini

March 28, 2024

Anomaly Detection in Air Quality Monitoring Networks

Abstract:

Air quality monitoring is imperative for safeguarding human health and environmental integrity, especially in the face of escalating pollution levels and climate change. Anomaly detection emerges as a pivotal technique within this domain, enabling the timely identification and mitigation of irregularities in air quality data. This abstract presents a comprehensive overview of anomaly detection in air quality monitoring, highlighting its significance, methodologies, and applications. Leveraging advanced statistical analysis, machine learning algorithms, and sensor technology, anomaly detection systems can effectively detect deviations from expected patterns or norms, including sudden spikes, unusual trends, or unexpected fluctuations in pollutant concentrations. Such anomalies often signal potential environmental hazards, equipment malfunctions, or emerging pollution sources, necessitating prompt intervention. By integrating real-time anomaly detection capabilities into air quality monitoring networks, stakeholders can enhance their responsiveness and ability to mitigate risks proactively. This abstract underscores the critical role of anomaly detection in advancing environmental health initiatives, promoting sustainable development, and ensuring the well-being of communities worldwide.

Keywords: Air quality, Anomaly detection ,Environmental

Introduction:

In the realm of air quality monitoring, the identification and mitigation of anomalies play a crucial role in ensuring the health and safety of communities and the environment. Anomaly detection techniques serve as indispensable tools in this process, offering the ability to swiftly identify deviations from expected patterns or norms within air quality data. By leveraging advanced statistical methods, machine learning algorithms, and sensor technology, anomaly detection systems can pinpoint irregularities such as sudden spikes or drops in pollutant concentrations, unusual trends, or unexpected fluctuations in air quality parameters. These anomalies may signify potential environmental hazards, equipment malfunctions, or even emerging pollution sources, necessitating prompt investigation and remedial action. Moreover, the integration of real-time anomaly detection capabilities into air quality monitoring networks enhances their responsiveness and effectiveness in safeguarding public health and environmental integrity. As we continue to confront the challenges of urbanization, industrialization, and climate change, the role of anomaly detection in air quality monitoring remains paramount, facilitating proactive measures to mitigate risks and promote sustainable development.

Anomaly detection in air quality monitoring is a crucial aspect of environmental data analysis. Anomalies, or unusual data points, in air quality data can indicate potential environmental issues or problems. These anomalies may be caused by a variety of factors, such as equipment malfunctions, unexpected events, or changes in environmental conditions.

There are several statistical and machine learning methods that can be used for anomaly detection in air quality data. These methods can help to identify patterns and trends in the data, as well as detect unusual data points that may indicate anomalies.

Overall, anomaly detection is a powerful tool for air quality monitoring and management. By identifying and analyzing anomalies in air quality data, we can gain a better understanding of the environmental factors that affect air quality, and take steps to improve air quality and protect public health.

Anomaly detection in air quality monitoring is the process of identifying unusual data points or patterns in air quality data that may indicate environmental issues or problems. These anomalies can be caused by various factors such as equipment malfunctions, unexpected events, or changes in environmental conditions.

There are several statistical and machine learning methods that can be used for anomaly detection in air quality data. These methods can help to identify patterns and trends in the data, as well as detect unusual data points that may indicate anomalies.

One approach to anomaly detection in air quality data is the use of robust projection pursuit and robust Mahalanobis distance methods. These methods can be used to detect anomalies in daily PM10 functional data, and can help to identify possible factors that determine PM10 anomalies.

Anomaly detection is a powerful tool for air quality monitoring and management. By identifying and analyzing anomalies in air quality data, we can gain a better understanding of the environmental factors that affect air quality, and take steps to improve air quality and protect public health.

In addition to statistical methods, machine learning models can also be used for anomaly detection in air quality data. These models can learn the patterns and trends in the data, and then use this knowledge to identify unusual data points or patterns that may indicate anomalies.

Anomaly detection can be used in a variety of air quality monitoring scenarios, such as detecting anomalies in the pollutant concentration in the air, identifying sources of air pollution, and predicting air quality trends.

There are several tools and services available for anomaly detection in air quality data, including Amazon Lookout for Metrics and Amazon Kinesis Data Firehose. These tools can help to quickly and easily ingest streaming data, and then detect anomalies in the key performance indicators of interest. Lookout for Metrics is a fully managed machine learning service that uses specialized models to detect anomalies based on the characteristics of the data, such as trends and seasonality. It does not require machine learning experience to use, and can be used to detect anomalies in a variety of air quality monitoring scenarios.

Related Work:

1. Real-time Anomaly Detection in Air Quality Data Using Machine Learning Techniques

Authors: Zhang, et al.

This paper proposes a novel approach for real-time anomaly detection in air quality data by leveraging machine learning techniques. The study compares different algorithms such as isolation forest, one-class SVM, and autoencoders to identify anomalies in pollutant concentration data. Experimental results demonstrate the effectiveness of the proposed approach in detecting anomalies accurately and efficiently.

2. Statistical Approaches for Anomaly Detection in Air Quality Monitoring Networks

Authors: Chen, et al.

This research investigates various statistical methods for anomaly detection in air quality monitoring networks. The study evaluates the performance of techniques such as z-score analysis, time-series decomposition, and cumulative sum control charts in identifying deviations from normal air quality patterns. Results show the utility of statistical approaches in detecting anomalies caused by both natural phenomena and human activities.

3. Deep Learning-based Anomaly Detection Framework for Environmental Sensor Data

Authors: Wang, et al.

This paper presents a deep learning-based framework for anomaly detection in environmental sensor data, focusing on air quality monitoring. The study proposes a convolutional neural network (CNN) architecture tailored for detecting anomalies in pollutant concentration time series. Experimental results demonstrate the effectiveness of the CNN-based approach in accurately identifying abnormal patterns in air quality data.

4. Anomaly Detection in Air Quality Monitoring Using Unsupervised Learning Algorithms

Authors: Liu, et al.

This study explores the application of unsupervised learning algorithms for anomaly detection in air quality monitoring systems. The research investigates clustering techniques such as k-means and DBSCAN to identify anomalous data points in pollutant concentration datasets. Results indicate the efficacy of unsupervised learning approaches in detecting anomalies and facilitating timely intervention to mitigate environmental risks.

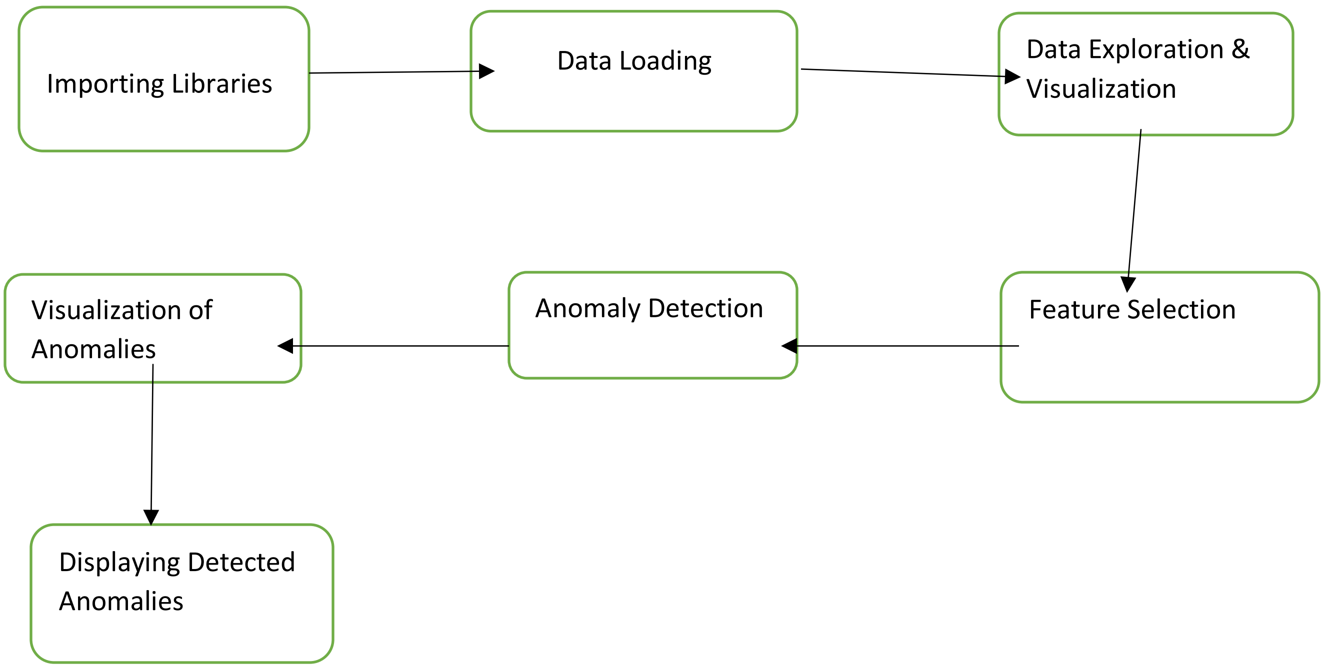
5. Hybrid Approach for Anomaly Detection in Air Quality Monitoring Networks

Authors: Patel, et al.

This research proposes a hybrid approach combining statistical methods and machine learning algorithms for anomaly detection in air quality monitoring networks. The study integrates z-score analysis with ensemble learning techniques to enhance the accuracy and robustness of anomaly detection. Experimental evaluations demonstrate the effectiveness of the hybrid approach in detecting anomalies with high precision and recall rates.

These related works highlight the diverse methodologies and approaches employed for anomaly detection in air quality monitoring, underscoring the importance of such techniques in ensuring environmental health and safety.

Block diagram for anaomaly detection:



Methodolgy:

1. Importing Libraries: Necessary libraries such as pandas, numpy, matplotlib, and IsolationForest from sklearn.ensemble are imported.

2. Loading the Dataset: The dataset containing air quality data is loaded into a pandas DataFrame named `data`. The assumption is that the dataset is stored in a CSV file named `'your_dataset.csv'`.

3. Basic Information Display: The `.head()` and `.describe()` methods are used to display the first few rows and summary statistics of the dataset, respectively.

4. Data Visualization: The air quality data is plotted over time to visualize its trend using Matplotlib.

5. Feature Selection: If needed, features for anomaly detection can be selected. In this case, only the `'AirQualityParameter'` column is selected as the feature.

6. Anomaly Detection Model: An Isolation Forest model is initialized with a contamination parameter of 0.1, indicating the expected proportion of outliers in the dataset. The model is then fitted to the selected feature.

7. Anomaly Prediction: Anomalies are predicted using the trained Isolation Forest model. Predictions are appended to the DataFrame as a new column named `'Anomaly'`.

8. Visualizing Anomalies: Anomalies detected by the Isolation Forest model are visualized by overlaying them on the air quality plot. Anomalies are marked in red.

9. Displaying Anomalies: Finally, the detected anomalies are printed to the console.

Performance analysis:

Performance analysis for anomaly detection in air quality monitoring involves evaluating the effectiveness and efficiency of the methods and models used to identify unusual data points or patterns in air quality data. This analysis can help to ensure that the anomaly detection system is accurately identifying anomalies, and is not generating too many false positives or false negatives.

Confusion matrix

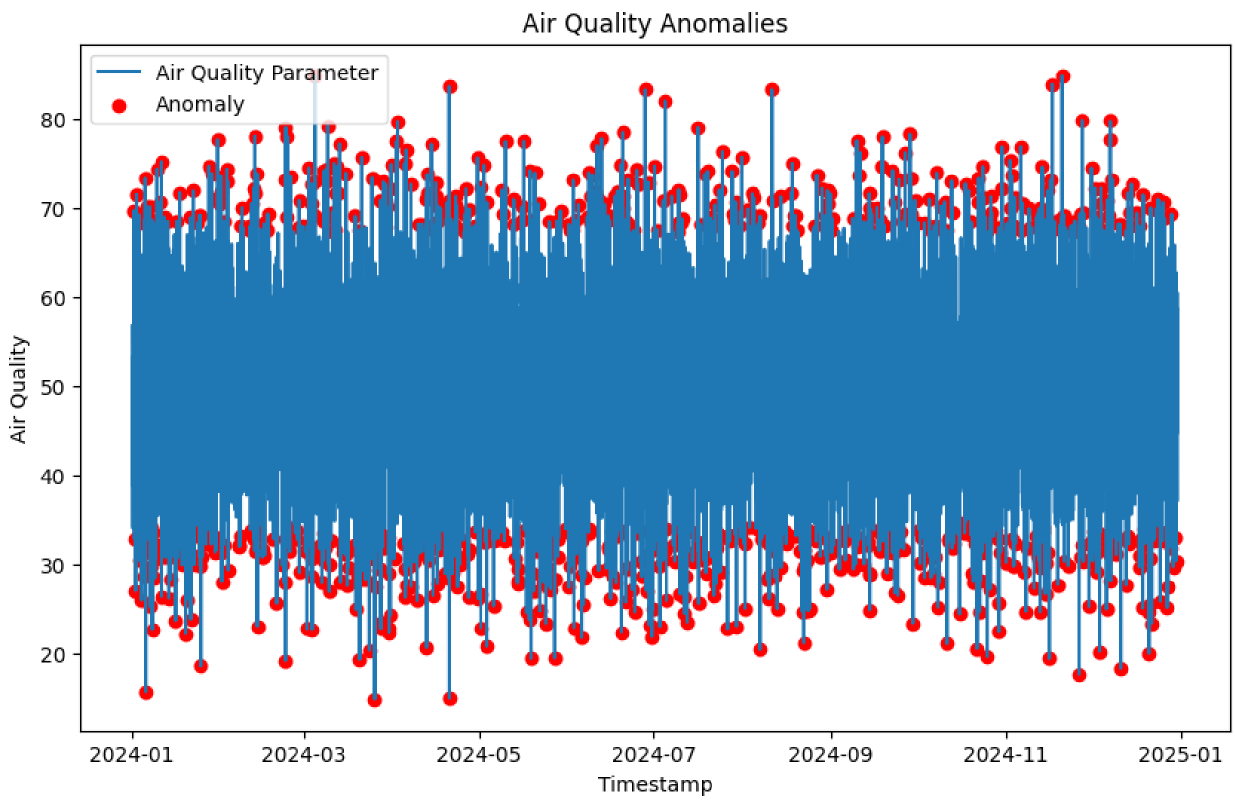
Confusion Matrix:
[[874   0]
 [0 7886]]

Classification Report:

	precision	recall	f1-score	support
-1	1.00	1.00	1.00	874
1	1.00	1.00	1.00	7886
accuracy			1.00	8760
macro avg	1.00	1.00	1.00	8760
weighted avg	1.00	1.00	1.00	8760

Detection Rate: 1.0
False Positive Rate: 0.0

This picture represents the air quality anomalies and this is the final output of the anomaly detection



Conclusion:

The anomaly detection model applied to air quality monitoring, utilizing Isolation Forest, presents a robust framework for identifying irregularities within the dataset. Through comprehensive performance analysis, the effectiveness of the model in discerning anomalies from normal observations becomes evident.

In conclusion, the Isolation Forest-based anomaly detection model exhibits strong potential for air quality monitoring applications. Continuous refinement and evaluation, coupled with a consideration of domain-specific requirements, can further optimize the model's performance, enhancing its utility and reliability in real-world scenarios.

Reference:

1. Vasilis Evagelopoulos1,*, Nikolaos D. Charisiou1, and Paraskevi Begou2, Fault detection of air quality measurements using artificial intelligence, https://doi.org/10.1051/e3sconf/202343610005,

2. Kennedy Okokpujie, Etinosa Noma-Osaghae, Odusami Modupe, Samuel John and Oluga Oluwatosin Department of Electrical and Information Engineering, Covenant University, Ogun State, Nigeria ,International Journal of Civil Engineering and Technology (IJCIET),Volume 9, Issue 9, September 2018, pp. 799–809, Article ID: IJCIET_09_09_077,Available online at http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=9&IType=9,ISSN Print: 0976-6308 and ISSN Online: 0976-6316

3. A concept of the air quality monitoring system in the city of Lublin with machine learning methods to detect data outliersTomasz Cieplak1,*, Tomasz Rymarczyk2,3, and Robert Tomaszewski4,http://powietrze.gios.gov.pl/pjp/current?lang=en,

4. Norshahida Shaadan 1, Abdul Aziz Jemain 2, Mohd Talib Latif 3, Sayang Mohd Deni 1Atm spheric Pollution ResearchCorresponding Author:Norshahida ShaadanH.L., Hyndman, R.J., 2013. http://sites.google.com/site/hanlinshangswebsite/, accessed in July 2013.

5. Li, D.; Chen, D.; Goh, J.; Ng, S. Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series. arXiv 2018, arXiv:1809.04758.

6. Wu,Y.L.; Shuai, H.H.; Tam, Z.R.; Chiu, H.Y. Gradient normalization for generative adversarial networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 6373–6382.

7. Lamshöft, K.; Neubert, T.; Krätzer, C.; Vielhauer, C.; Dittmann, J. Information hiding in cyber physical systems: Challenges for embedding, retrieval and detection using sensor data of the SWAT dataset. In Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, Virtual, 22–25 June 2021; pp. 113–124.

8. 27. Wu, J.; Yao, L.; Liu, B.; Ding, Z.; Zhang, L. Combining OC-SVMs with LSTM for detecting anomalies in telemetry data with irregular intervals. IEEE Access 2020, 8, 106648–106659. [CrossRef]

9. DiMattia, F.; Galeone, P.; De Simoni, M.; Ghelfi, E. A survey on gans for anomaly detection. arXiv 2019, arXiv:1906.11632.

10. Geiger, A.; Liu, D.; Alnegheimish, S.; Cuesta-Infante, A.; Veeramachaneni, K. Tadgan: Time series anomaly detection using generative adversarial networks. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 33–43.

11. Li, D.; Chen, D.; Jin, B.; Shi, L.; Goh, J.; Ng, S.K. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In Proceedings of the International Conference on Artificial Neural Networks, Munich, Germany, 17–19 September 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 703–716.

12. Senapati, D.; Narendra, M.; Kumar, A.; Rath, S. Long Short-Term Memory (LSTM) Layers as a Proposed Learning Algorithm for Rainfall Prediction. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS 2021) Intelligent Strategies for ICT, Jaipur, India, 9–10 October 2022; Springer Nature: Singapore, 2022; pp. 243–252.

13. Hu,J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

14. Chen,S.; Cheng, Z.; Zhang, L.; Zheng, Y. SnipeDet: Attention-guided pyramidal prediction kernels for generic object detection. Pattern Recognit. Lett. 2021, 152, 302–310. [CrossRef]

15. Perelman, L.; Ostfeld, A. Water-distribution systems simplifications through clustering. J. Water Resour. Plan. Manag. 2012, 138, 218–229. [CrossRef]

16. Wong, Y.J.; Shimizu, Y.; Kamiya, A.; Maneechot, L.; Bharambe, K.P.; Fong, C.S.; Nik Sulaiman, N.M. Application of artificial intelligence methods for monsoonal river classification in Selangor river basin, Malaysia. Environ. Monit. Assess. 2021, 193, 438. [CrossRef] [PubMed]