JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TIME SERIES ANOMALY DETECTION

Shital Pazare[1], Aakshun Chatla[2], Sagar DhudhBhate[3], Harsh Kamani[4], Dharmendra Mishra[5]

1. *Artificial Intelligence and Data Science Department, Shah & Anchor Kutchhi Engineering College*
2. *Artificial Intelligence and Data Science Department, Shah & Anchor Kutchhi Engineering College*
3. *Artificial Intelligence and Data Science Department, Shah & Anchor Kutchhi Engineering College*
4. *Artificial Intelligence and Data Science Department, Shah & Anchor Kutchhi Engineering College*
5. *Artificial Intelligence and Data Science Department, Shah & Anchor Kutchhi Engineering College*

Abstract - The Time Series Anomaly Detector is a powerful tool designed to identify unusual patterns or events within a sequence of data over time. Whether monitoring financial transactions, temperature fluctuations, or any time-dependent dataset, this app helps user spin point anomalies that may indicate irregularities or unexpected occurrences.The app employs advanced algorithms to analyze historical data and normal patterns. Users can set thresholds to define what is considered typical behavior, and the app automatically flags deviations beyond these thresholds as anomalies.One key feature of the app is its user-friendly interface, making it accessible to a broad audience, including those without extensive data science expertise. The visual representation of time series data and highlighted anomalies simplify the interpretation of results. Additionally, the app supports customizable alerts, notifying users when anomalies are detected, facilitating aprompt response.In practical terms, the Time Series Anomaly Detector finds applications in diverse fields. For instance, in finance, it can identify unusual transaction patterns that may indicate fraudulent activity. In environmental monitoring, it could detect abnormal weather patterns or sensor malfunctions. The versatility of the app makes it a valuable asset across industries were understanding and responding to anomalies in time series data are critical.In summary, the Time Series Anomaly Detector is an intuitive and efficient solution for anyone seeking to gain insights from time-dependent data by swiftly identifying and addressing anomalies.

***Key Words*:** Anomaly, Detection, Outlier, Deviation, Fraud, Unusual patterns, Fault detection,

1. INTRODUCTION

In many industrial, medical, and scientific applications, anomaly detection (AD) in time series is important. For example, millions of people could be harmed by undiscovered anomalies in chemical or water treatment plants.

Finding unusual things, occurrences, or observations that stick out because they significantly deviate from expected patterns or behaviors is the essence of anomaly detection. These anomalies, which are also known as noise, standard deviations, outliers, novelties, or exceptions, are essential for anticipating and mitigating possible disruptions or dangers in a dataset.

Time series analysis can be useful in figuring out how a particular asset, security, or economic attribute changes over time. Comparing changes in the chosen data point to changes in other variables during the same time period is another application for it. Anomaly detection is the process of identifying odd objects, events, or observations that raise red flags because they significantly diverge from accepted patterns or behaviors. Additional words for anomalies in data include standard deviations, outliers, noise, novelty, and exceptions. In this context, our Time Series Anomaly Detector emerges as a crucial ally in fortifying data security. By detecting sudden spikes or anomalies in both manually input and uploaded data, spanning diverse types such as bank statements, company expenses, or report values, the application ensures a vigilant watch over numeric values. The app's API is designed to scrutinize data for irregular patterns or spikes, offering a proactive defense mechanism against potential risks or abnormalities, regardless of the data type. As we delve deeper into the report, we will unravel the technical intricacies and practical applications that make our Time Series Anomaly Detector a revolutionary force in the domain of data security.

2. METHODOLOGY

Problem Definition:

The objective of the project is to develop a comprehensive web application aimed at evaluating and comparing various model selection methods for anomaly detection in time series data. By clearly defining this objective, the project aims to address the growing need for robust anomaly detection solutions across diverse domains.

Data Collection and Preparation:

In this phase, the project will focus on gathering a wide array of time series datasets that accurately represent the characteristics and challenges present in real-world scenarios. These datasets will undergo rigorous preprocessing to handle missing values, normalize the data, and rectify any other quality issues that could impact the effectiveness of anomaly detection algorithms.

Algorithm Selection:

The project will meticulously identify and curate a selection of anomaly detection algorithms to be integrated into the web application. This selection process will prioritize algorithms spanning various methodologies, including statistical, machine learning, and deep learning approaches, ensuring a comprehensive evaluation of diverse techniques.

Web Application Development:

The development phase will concentrate on creating a user-friendly web interface that facilitates seamless uploading of datasets and selection of algorithms for evaluation. The backend functionality will be robustly engineered to efficiently process uploaded data, train models, and conduct evaluation metric calculations. Additionally, the incorporation of visualization tools will enable users to intuitively explore evaluation results and visualize detected anomalies.

Model Training and Evaluation:

Each chosen anomaly detection model will undergo rigorous training using the uploaded datasets, followed by comprehensive evaluation using predefined metrics such as precision, recall, F1-score, AUC-ROC, and AUC-PR. This systematic evaluation approach will enable the identification of each model's strengths and weaknesses, facilitating informed decision-making during the selection process.

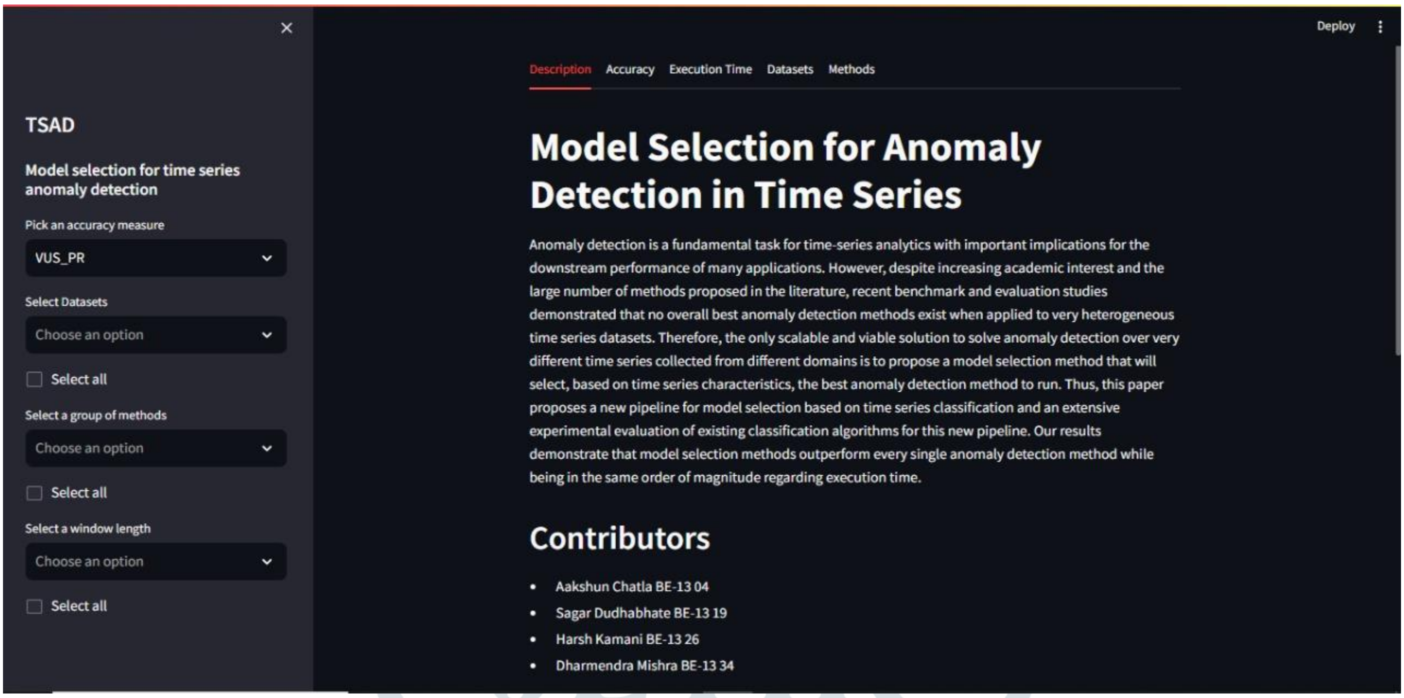
Documentation:

Throughout the project lifecycle, meticulous documentation will be maintained, detailing every step of the methodology, including data collection, preprocessing, algorithm selection, and evaluation procedures. Comprehensive reports summarizing the project's findings will be prepared, incorporating tables, charts, and other visual aids to effectively communicate the outcomes of the evaluation process.

3. IMPLEMENTATION

The implementation of the performance evaluation parameters outlined in the above project entails a systematic and rigorous approach to assessing the effectiveness of anomaly detection models for time series data. It begins with meticulous data preparation, involving the collection and preprocessing of diverse time series datasets representative of various domains and characteristics. These datasets are then partitioned into training, validation, and test sets to facilitate model training and evaluation.

The next step involves the selection and implementation of a range of anomaly detection models, including statistical methods, machine learning algorithms, and deep learning architectures. Each model undergoes rigorous training using the training dataset, with hyper parameters optimized through techniques such as grid search or Bayesian optimization. Validation using the validation dataset allows for the assessment of model performance and fine-tuning of parameters if necessary.



TSAD

Model selection for time series anomaly detection

Pick an accuracy measure: VUS_PR

Select Datasets: Choose an option

□ Select all

Select a group of methods: Choose an option

□ Select all

Select a window length: Choose an option

□ Select all

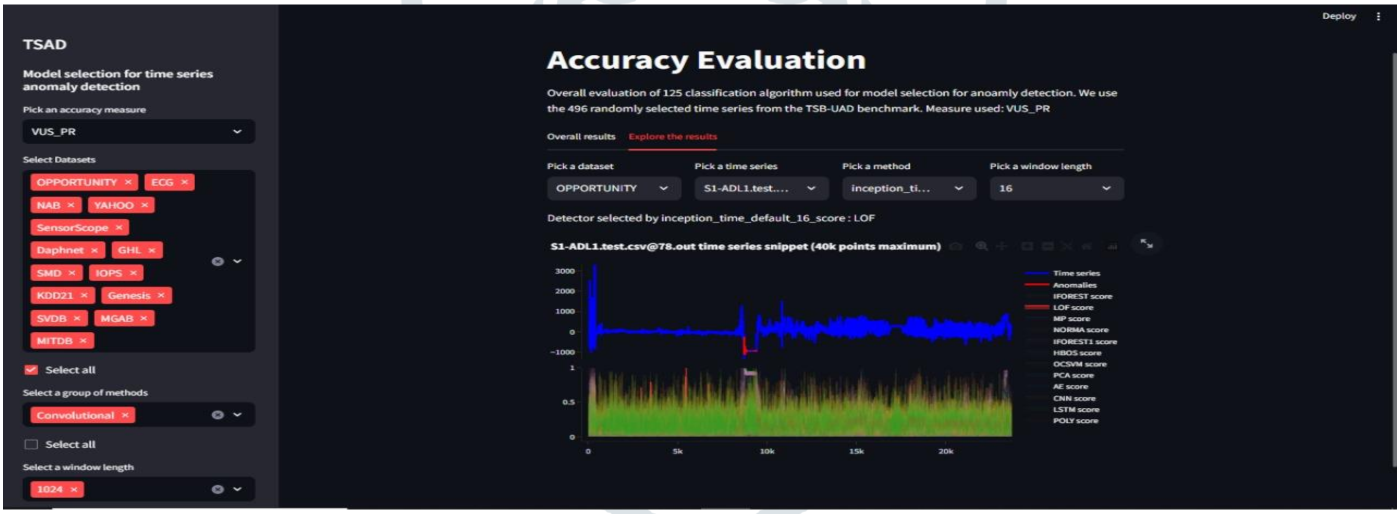
Model Selection for Anomaly Detection in Time Series

Anomaly detection is a fundamental task for time-series analytics with important implications for the downstream performance of many applications. However, despite increasing academic interest and the large number of methods proposed in the literature, recent benchmark and evaluation studies demonstrated that no overall best anomaly detection methods exist when applied to very heterogeneous time series datasets. Therefore, the only scalable and viable solution to solve anomaly detection over very different time series collected from different domains is to propose a model selection method that will select, based on time series characteristics, the best anomaly detection method to run. Thus, this paper proposes a new pipeline for model selection based on time series classification and an extensive experimental evaluation of existing classification algorithms for this new pipeline. Our results demonstrate that model selection methods outperform every single anomaly detection method while being in the same order of magnitude regarding execution time.

Contributors

- Aakshun Chatla BE-13 04
- Sagar Dudhabhate BE-13 19
- Harsh Kamani BE-13 26
- Dharmendra Mishra BE-13 34

Figure : Base page of website



TSAD

Model selection for time series anomaly detection

Pick an accuracy measure: VUS_PR

Select Datasets: OPPORTUNITY × ECG × NAB × YAHOO × SensorScope × Daphnet × GHL × SMD × IOPS × KDD21 × Genesis × SVDB × MGAB × MITDB ×

☑ Select all

Select a group of methods: Convolutional ×

□ Select all

Select a window length: 1024 ×

Accuracy Evaluation

Overall evaluation of 125 classification algorithm used for model selection for anoamly detection. We use the 496 randomly selected time series from the TSB-UAD benchmark. Measure used: VUS_PR

Pick a dataset: OPPORTUNITY

Pick a time series: S1-ADL1.test....

Pick a method: inception_ti...

Pick a window length: 16

Detector selected by inception_time_default_16_score : LOF

S1-ADL1.test.csv@78.out time series snippet (40k points maximum)

Figure : Detection page of website

4. PURPOSE

Anomaly detection in time series datasets serves several important purposes across various domains and applications:
Early Warning Systems: Anomaly detection can act as an early warning system, alerting stakeholders to potential issues or abnormalities in a timely manner. This proactive approach allows organizations to take preventive actions and mitigate risks before they escalate into larger problems.
Fault Detection and Diagnosis: In industrial settings, anomaly detection helps identify faults or anomalies in machinery or processes. By detecting deviations from normal behavior, it enables maintenance teams to diagnose the root cause of issues and perform timely repairs, minimizing downtime and maximizing productivity.
Cybersecurity: Time series anomaly detection plays a crucial role in cybersecurity by identifying unusual patterns or activities in network traffic, system logs, or user behavior. Detecting anomalies indicative of security breaches or unauthorized access allows security teams to respond swiftly and safeguard sensitive data and systems.
Financial Fraud Detection: In the finance sector, anomaly detection helps detect fraudulent activities such as unauthorized transactions, money laundering, or insider trading. By flagging suspicious patterns in financial transactions or market data, anomaly detection systems protect against financial losses and ensure regulatory compliance.
Healthcare Monitoring: Anomaly detection is utilized in healthcare for monitoring patient vital signs, disease progression, or medical device data. Detecting anomalous patterns can help healthcare providers identify critical health issues, such as irregular heartbeats or abnormal physiological parameters, and intervene promptly to provide appropriate medical care.

Infrastructure Monitoring: In infrastructure management, anomaly detection monitors critical systems such as power grids, transportation networks, or telecommunications infrastructure. By identifying anomalies in data such as sensor readings or network traffic, it helps prevent infrastructure failures, optimize performance, and ensure uninterrupted service delivery.

Predictive Maintenance: Anomaly detection aids in predictive maintenance by identifying deviations from normal equipment behavior that may indicate impending failures or degradation. By scheduling maintenance activities based on actual equipment condition rather than predetermined schedules, organizations can minimize downtime and reduce maintenance costs.

Quality Control: In manufacturing and production processes, anomaly detection ensures product quality by identifying deviations from expected standards or specifications. Detecting anomalies in sensor data or production metrics allows organizations to take corrective actions and maintain consistent product quality levels.

5.FUTURE WORK

In the realm of time series anomaly detection, the development of advanced models stands out as a primary avenue for future research. Novel architectures and algorithms leveraging the capabilities of deep learning, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), present promising opportunities to capture intricate temporal patterns and dependencies within data. These models can be further enhanced with attention mechanisms to focus on relevant parts of the time series or by incorporating techniques like generative adversarial networks (GANs) for anomaly generation and detection.

Moreover, there is a pressing need for anomaly detection models that not only flag anomalies but also provide interpretable explanations for their detections. Interpretability is crucial for gaining trust and acceptance of anomaly detection systems in real-world applications. Methods for creating interpretable models may involve attention mechanisms, saliency mapping, or rule-based post-processing to explain the rationale behind anomaly detections in a human-understandable manner.

Anomaly localization is another area ripe for exploration. Techniques such as attention-based mechanisms can highlight specific time points or features contributing to anomaly detection, providing valuable insights into the context and severity of anomalies. Additionally, segmentation methods can identify anomalous segments within time series data, while visualization techniques can offer intuitive representations of detected anomalies, aiding in their interpretation and understanding.

For anomaly detection to be successful, it is also essential to address the issue of imbalanced datasets. In example, when anomalies are rare in comparison to normal data points, techniques like oversampling, undersampling, or synthetic data generation can assist balance the distribution of anomalies and normal instances, boosting the effectiveness of anomaly detection algorithms. By advancing research in these key areas, the field of time series anomaly detection can make significant strides towards more accurate, interpretable, and actionable anomaly detection solutions across various domains and applications.

References

i. H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate Time-Series Anomaly Detection via Graph Attention Network," in 2020 IEEE International Conference on Data Mining (ICDM), 2020.
https://arxiv.org/pdf/2009.02040.pdf

ii. R. Wu and E. Keogh, "Current Time Series Anomaly Detection Benchmarks Are Flawed and Are Creating the Illusion of Progress," arXiv, 2020.
https://arxiv.org/ftp/arxiv/papers/2009/2009.13807.pdf

iii. S. Wang, Y. Zeng, X. Liu, E. Zhu, J. Yin, C. Xu, and M. Kloft, "Effective End-to-End Unsupervised Outlier Detection via Inlier Priority of Discriminative Network," in Advances in Neural Information Processing Systems, 2019.
https://proceedings.neurips.cc/paper_files/paper/2019/file/6c4bb406b3e7cd5447f7a76fd7008806-Paper.pdf

iv. D. Park, Y. Hoshi, and C. C. Kemp, "A Multimodal Anomaly Detector for Robot- Assisted Feeding Using an LSTM-Based Variational Autoencoder," in IEEE Robotics and Automation Letters, 2018.
https://arxiv.org/pdf/1711.00614.pdf

v. A. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T.
G. Dietterich, and K.-R. Muller, "A Unifying Review of Deep and Shallow Anomaly Detection," arXiv, 2020.
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9347460

vi. M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnt: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," in IEEE Access, 2019.
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8581424

vii. F. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in 2008 Eighth IEEE International Conference on Data Mining, 2008.
https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf?q=isolation-forest

viii. B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection," in International Conference on Learning Representations, 2018.
https://openreview.net/pdf?id=BJJLHbb0-

ix. Y. He and J. Zhao, "Temporal Convolutional Networks for Anomaly Detection in Time Series," in Journal of Physics: Conference Series, 2019.
https://iopscience.iop.org/article/10.1088/1742-6596/1213/4/042050/pdf

x. P. de Haan and S. Lowe, "Contrastive Predictive Coding for Anomaly Detection," arXiv, 2021. https://arxiv.org/pdf/2107.07820.pdf

xi. M. U. Gutmann and A. Hyvarinen, "Noise-Contrastive Estimation of Unnormalized Statistical Models," in Journal of Machine Learning Research, 2012.
https://www.jmlr.org/papers/volume13/gutmann12a/gutmann12a.pdf

xii. M. Solch, J. Bayer, M. Ludersdorfer, and P. van der Smagt, "Variational Inference for Online Anomaly Detection in High-Dimensional Time Series," arXiv, 2016.
https://arxiv.org/pdf/1602.07109.pdf

xiii. I. Golan and R. El-Yaniv, "Deep Anomaly Detection Using Geometric Transformations," in Advances in Neural Information Processing Systems, 2018.
https://proceedings.neurips.cc/paper_files/paper/2018/file/5e62d03aec0d17facfc5355dd90d441c-Paper.pdf

xiv. K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in ACM SIGKDD, 2018.
https://arxiv.org/pdf/1802.04431.pdf

xv. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection in Time Series Data: A Survey," in ACM Computing Surveys, 2009.
https://conservancy.umn.edu/bitstream/handle/11299/215731/07-017.pdf?sequence=1