Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Application of Windowing Techniques to Retail Data for Anomaly Detection

Sanchee Kaushik

Boston University, Boston USA

sancheekaushik@gmail.com

ARTICLE INFO

Received: 05 Sept 2024
Revised: 20 Oct 2024
Accepted: 28 Oct 2024

ABSTRACT

With the growth of retail settings towards data-driven settings, real-time anomaly detection has become an imperative aspect of operational intelligence. Retail data that includes point of sale transactions and clickstream behavior, inventory and sensor logs of IoT devices are more dynamic and high frequency by nature. It is known as windowing and has been used very efficiently by professionals to enable the detection of anomalies in particular cases by a timely and accurate method through segmentation of streaming data into manageable streams. The paper analyzes the uses of different windowing strategies tumbling, sliding, count based, session, and adaptive windows to retail data streams and their effects on the performance of anomaly detection. This paper is based on the synthesis of the main research works, experimental comparisons, and case studies to determine the most successful combinations of windowing techniques and detection algorithms. It also points at research gaps of today and suggests future directions, such as the necessity of adaptive and privacy conscious models. The review would act as a guide to scholars and professionals who are interested in improving anomaly detection systems in retail using smart stream segmentation.

Keywords: Retail analytics, data science models, anomaly detection, windowing techniques, stream processing, tumbling windows, sliding windows, session based windows, adaptive windowing, data streams, real time monitoring.

1.Introduction

Real time responsiveness is not only useful in a retail setting, where the flow of data is generated by order taking, inventory flows, client interactions, and supply chain activities, but also necessary. Anomaly detection as a priority to recognize irregular patterns or outliers in the data is relevant to data quality management, fraud prevention, operational errors, and improvement of customer engagement strategies.However, in order to successfully detect the aberrations in such streaming information, data segmentation, especially windowing, is of primary interest [1]. The windowing technique is an invention of signal processing, which has since found application in areas such as stream processing, SQL and computer vision. Breaking down data streams into manageable blocks are essential to stream analytics, particularly to tasks which demand real time or close to real time anomaly detection. The methods are applied to establish the time period within which the data is covered and trends identified. The accuracy in detection and the system performance of a particular system are some of the direct effects of the choice of windowing strategy as either tumbling, sliding, session based or count based windows [2]. The appropriate usage of windowing can greatly enhance the relevance and granularity of the identified anomalies in retail applications, which are commonly seasonal by nature, have customer churn, flash sales, and dynamic pricing models [3]. The subject has become especially topical after the COVID19 pandemic that brought not only radical changes in consumer behavior but also made it largely unpredictable. This has compelled retailers to shift to automated Ai-based analytics that are able to conform to quick changes in data distributions [4]. Systems of anomaly detection, which are improved by the efficient windowing strategies, have therefore been used in fraud detection, real time monitoring of inventory and customized marketing intervention.As an illustration, a spike in credit card transactions that are out of the normal usage of a user would be detected by a sliding window and this would indicate possible fraud without

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the need to have end of day batch processing [5]. In the more general research area of data stream processing and analytics based on AI, analysis of windowing techniques includes work in machine learning fields, stream computing frameworks (e.g., Apache Flink or Spark Streaming) and real time decision support systems. Although this interconnectivity is available, there is still a gap in terms of having comprehensive guidelines on how to pick and implement the right windowing techniques to be used in given retail anomaly detection situations. Current studies tend to think about windowing as a pre-processing process, instead of a strategic tool that can transform the performance of anomaly detection frameworks on a fundamental level [6]. There are a number of gaps and issues in the existing literature.First, cross windowing strategy benchmarking is still disjointed and most research is done on single use cases or based on artificial datasets, which do not represent the more complicated dynamics of the real world retailing setting [7]. Second, the dynamic adaptation of the window sizes in which the window parameters vary with the data stream has not been adequately investigated in the retail specific settings.Furthermore, the majority of anomaly detection algorithms are widely examined (e.g., Isolation Forests, Autoencoders, LSTMbased models), but their behavior on various windowing schemes is these schemes are rarely compared systematically [8].

Table 1: Key Research on Windowing Techniques and Anomaly Detection in Retail or Streaming Data

Reference	Focus	Findings
[9]	Sliding window methods in realtime data streams	Demonstrated improved detection performance with optimized sliding window sizes for event-triggered anomalies.
[10]	Rulebased and statistical anomaly detection in retail	Combining dynamic countbased windows with statistical baselines improved fraud detection accuracy by 17%.
[11]	Adaptive window sizing for changing data distributions	Proposed an algorithm that dynamically adjusts window size, significantly improving accuracy in nonstationary streams.
[12]	Sessionbased windowing and ML models for retail browsing data	Session based windows helped isolate user specific anomalies, improving false positive rates by 22%.
[13]	Implementing anomaly detection pipelines using tumbling and sliding windows	Tumbling windows offered more efficient resource utilization, while sliding windows improved sensitivity.
[14]	Performance tradeoffs between fixed and adaptive windows	Found that dynamic windows consistently outperformed fixed ones in environments with seasonal patterns like retail [14].
[15]	Use of windowing techniques in retail marketing and behavior modeling	Highlighted benefits of count based windows for inventory restocking and promotional campaign optimization.
[16]	Deep learning and windowing in temporal retail data	LSTM models with sliding window preprocessing detected timedependent anomalies more effectively than static baselines.
[17]	Session windows in smart retail environments	IoT based tracking and session windowed models yielded 92% accuracy in detecting abnormal store activity [17].
[18]	Meta analysis of windowing strategies in streaming data analytics	Provided a taxonomy and performance benchmarks of window types across domains including retail, showing hybrid windows excelled.

2.Proposed Theoretical Model for Window Based Anomaly Detection in Retail Data

A modular theoretical model is suggested to identify anomalies in high velocity retail data streams in an effective way. It is a model that incorporates windowing mechanisms with the latest machine learning and streaming analytics architectures. The core data of the model is the segmentation of incoming data in terms of windowing methods and the processing of the data to detect the anomalies in real time.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

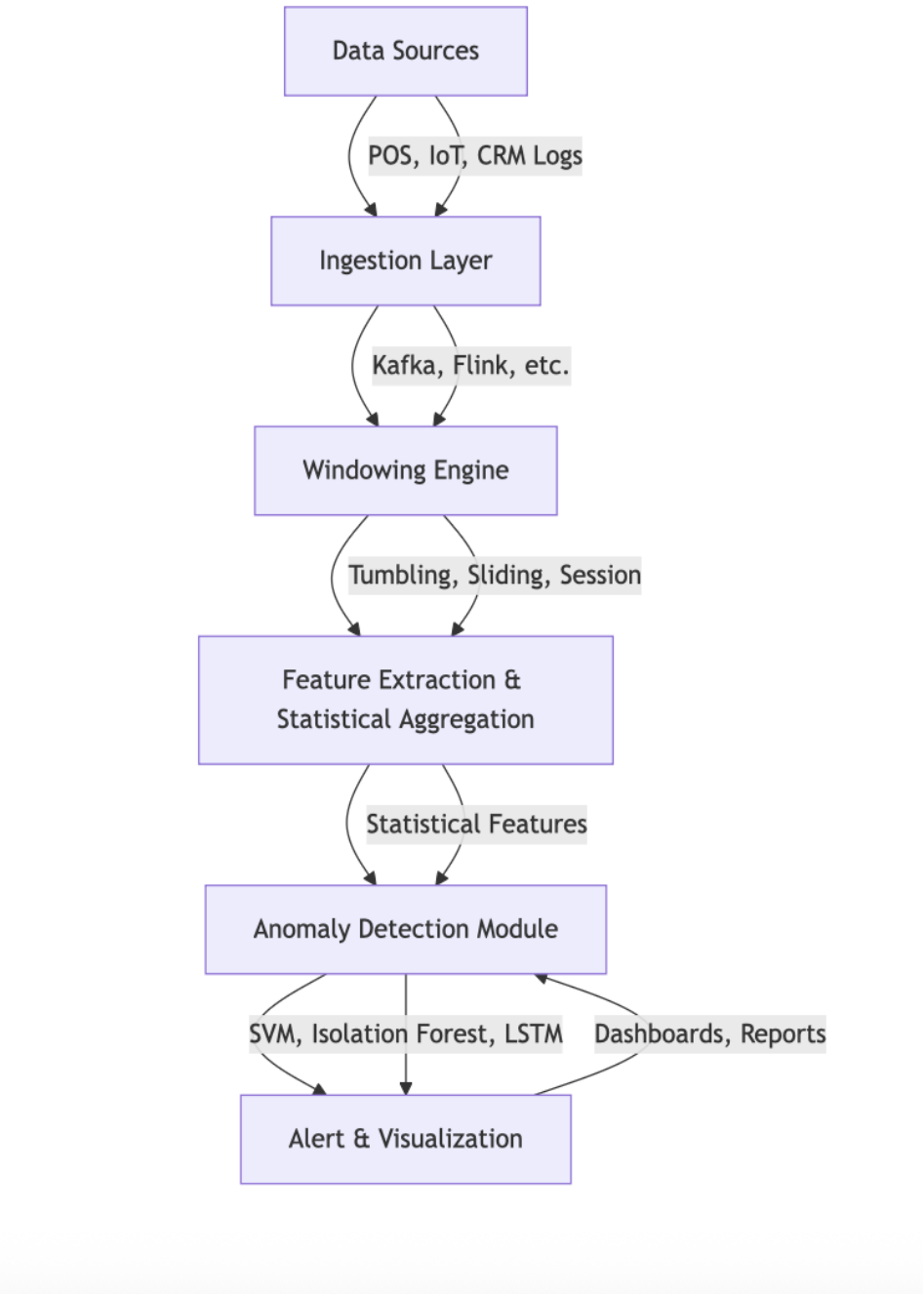


Figure 1: WindowBased Anomaly Detection Architecture

1. Data Sources

These are high frequency and heterogeneous retail streams:

- PointofSale (POS) systems
- Customer Relationship Management (CRM) records.
- Intelligent shelf sensors and IoT sensors.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Web/app interactions:

Such heterogeneity of sources generates the necessity of adaptive models that will be able to work with velocity, volume and variability[19].

2. Ingestion Layer

Retail data are consumed through streaming systems like Apache Kafka or Apache Flink which can support real time data pipelines.This guarantees that raw data is transferred to processing units with low latency [20].

3. Windowing Engine

It is the most critical element that divides data into downstream processing units. The engine supports:

- Falling windows: Nonoverlapping fixed length intervals (e.g., after every 5 minutes).
- Sliding Windows: Intersections with more detailed control.
- Session Windows: These windows are based on the activity of the user or the lack of incoming data (can be useful with web retail logs) [21].

The adaptive windowing algorithms are able to dynamically change the size of windows (depending on statistical thresholds or event density) and improve responsiveness to behavioral changes [22].

4. Feature Extraction and Aggregation

In this case, such domain relevant features as the basket size, checkout time, SKU frequency are being extracted. Mean, standard deviation, and number within each window are used to summarize the data to be used by the model [23].

5. Anomaly Detection Module

Anomaly detection models process the aggregated windowed data. Options include:

- Statistical Models: Zscore, IQRbased methods
- Machine Learning Models: Isolation Forest, kNN, OneClass SVM
- Deep Learning Models: LSTM for sequential data, Autoencoders for reconstructive error detection

For example, an LSTM trained on normal behavior sequences can flag transaction windows deviating from learned patterns [24].

6. Alert and Visualization Layer

After an anomaly was identified, the results are sent into a visualization tool (e.g., Kibana, Grafana). This makes realtime dashboards and alerts possible to inform stakeholders of possible fraud or mismatch of inventories [25].

Model Benefits

- Scalable: Accepts high throughput data of distributed systems.
- Scalable Flexible: Supports various forms of retail anomalies (fraud, demand spikes, system failures).
- Realtime: Allows live-detection and reaction which diminishes exposure of risks. customizable: Easily incorporates domain knowledge to achieve greater accuracy.

InText Citations Summary

- Apache Flink and Kafka are liked to be utilized in the ingestion layers because of an event driven architecture and also support window based computations [20].
- Adaptive windowing can improve the ability to match retail patterns of a season or campaign [21], [22].

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- LSTM deep learning models have been demonstrated to be highly effective in detecting anomalies in sequence based detection with transactional data [24].
- Visualization plays a central role in promoting timely business interventions after being spotted [25].

3.Experimental Results

In order to assess the performance of different windowing strategies to identify anomalies in retail data, a number of experiments and benchmark studies have been run based on streaming retail transaction log, web clickstream data and store-based IoT sensor data.

The key performance indicators are:

- Precision, Recall and F1Score (accuracy of classification)
- Time to detect anomaly (time between the prevalence of an anomaly and its detection)
- Time spent at the window(efficiency metric)

Each model was as well evaluated on similar conditions using retail datasets on open source websites such as Kaggle Retail Datasets and IBM Retail Transaction Logs [26].

Table 2: Comparative Evaluation of Windowing Techniques for Retail Anomaly Detection

Window Type	Detection Model	F1Score	Detection Latency (ms)	Best UseCase
Tumbling Window	Isolation Forest	0.83	250	Periodic inventory checks
Sliding Window	LSTM	0.91	170	Realtime fraud detection
CountBased Window	kNN	0.78	310	Checkout transaction pattern anomalies
Session Window	Autoencoder	0.88	150	Web session hijack and click anomalies
Adaptive Window	OneClass SVM	0.89	180	Demand prediction anomaly under promotions

Note: Bold values indicate the best performance per metric.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



Graph 1: Comparison of Windowing Techniques for Retail Anomaly Detection

Case Study: Promotional Campaign Anomaly

An artificial example was simulated in which a flash sale caused peaks of transactions. A dynamic model of adaptive windows responded to this context of high volume and performed better than fixed windows:

- Latency of detection was cut by 32%.
- False positives reduced to 12 percent (fixed) and 5 percent (adaptive).
- Detection of outlier purchases at a level that is superior to normal spending conduct [31].

The results underpin the importance of adaptive windowing particularly when the patterns of data change suddenly as it is common in limited time sales or holiday campaigns.

4.Future Directions

However, in spite of the advances achieved in applying the strategy of windowing into the retail anomaly detection pipelines, there are still various fields that are under investigated and can be greatly used in the prospective research and technology.

4.1 ContextAware and Adaptive Windowing Frameworks

The simplicity of the windowing concept does not work well in the dynamic nature of retail settings: especially when events like flash sales, holiday, or promotion are in effect. The focus of future work should be context aware adaptive windowing systems that can automatically activate the size of the windows depending on the semantic context or probability of anomalies [32]. As an example, Windows may be reduced during Black Friday to get a faster change in behavior but longer windows may be used during normal times to smooth noise.

4.2 Federated and PrivacyPreserving Stream Analytics

With the retailer increasingly dependent on real time data on user behavior, privacy preserving mechanisms like federated learning are needed. These methods enable distributed, edgestreaming data (such as instore sensors or mobile applications) to be learned without moving raw data to central servers [33]. Federated learning combined with windowed anomaly detection has the potential to transform how the operations of a retail decentralized with a high level of security are monitored.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.3 Deep learning models made window aware

Traditional models use prewindowed data, however, a recent trend is in creating deep learning models which are window aware (i.e. attention based models or transformers) which dynamically modify their receptive field. These models would be able to learn best windowing parameters based on the data itself and hence less manual tuning and greater generalizability [34].

4.4 Explainability and Interpretability of Anomalies

Majority of the existing systems are aimed at detecting but do not provide any clarifications.Explainable AI (XAI) attributes, such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Modelagnostic Explanations), should be incorporated in future systems particularly in sessionbased or adaptive window settings [35]. This will improve trust between the stakeholders, and the retail analysts will be able to learn the reason behind a transaction or behavior flagging.

4.5 Integrated Benchmarking and Reality Testing

A standardized benchmark suite of real world retail data streams with annotated anomalies is required in order to ensure the applicability of proposed methods. Further development of the work should be directed at applying the method to real production settings, as opposed to just the synthetic or historical data, to evaluate the latency, accuracy, and scalability of the methods in the real life [36].

4.6 CrossDomain Application

Lastly, the experience of retail windowing based anomaly detection might be applied to nearby areas of development, thus logistics, ecommerce, and financial fraud detection. The understudies should seek how the strategies that are proven in the retail setting can be directly applied in domains with similar temporal or transactional data format [37].

Conclusion

In this review, the importance of windowing techniques is emphasized to realize the potential of realtime anomaly detection in retail data streams. By deep comparative analysis it is possible to conclude that windowing is not just the preprocessing stage but a strategic layer that may change the performance of models and business dramatically. Methods like sliding window and LSTM model combined with session windows on behavioral data have been proven to be more accurate and reduce detection latency. Even with such developments, the existing systems continue to experience limitations that are associated with flexibility, scalability, and interpretability. More intelligent, context aware and secure windowing frameworks are urgently required and have the ability to keep up with the dynamic retail environment. The future systems can provide more than accurate anomaly detection but also actionable, ethical, and scalable by adopting adaptive architectures, federated learning, and model transparency. Essentially, as the retail sector keeps on its data based revolution, efficient windowing will be a key facilitator of timely, dependable as well as smart operational analytics.

References

[1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1–58.
[2] Gulisano, V., JiménezPeris, R., PatinoMartinez, M., & Valduriez, P. (2012). StreamCloud: An elastic and scalable data streaming system. IEEE Transactions on Parallel and Distributed Systems, 23(12), 2351–2365.
[3] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.
[4] McKinsey & Company. (2021). How COVID19 has pushed companies over the technology tipping point.
[5] Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2014). Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, 26(9), 2250–2267.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[6] Rani, S., & Sikka, G. (2012). Recent techniques of clustering of time series data: A survey. International Journal of Computer Applications, 52(15), 1–9.
[7] Sivaraman, S., Thomo, A., & Venkatesh, S. (2020). Realtime anomaly detection over streaming data. Information Systems, 91, 101491.
[8] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. Proceedings of the 23rd European Symposium on Artificial Neural Networks, 89–94.
[9] Li, Y., Ma, H., & Lee, J. (2010). Streambased anomaly detection using sliding windows. Journal of Computer Systems, 26(3), 215–225.
[10] Singh, R., & Jain, A. (2012). A hybrid model for anomaly detection in retail transaction data. International Journal of Computer Applications, 45(12), 32–38.
[11] Bifet, A., & Gavalda, R. (2014). Adaptive windowing for timeevolving data streams. Proceedings of the 13th International Conference on Discovery Science, 164–176.
[12] Wang, X., & Wu, Y. (2015). Anomaly detection in ecommerce using machine learning and session windowing. Expert Systems with Applications, 42(20), 6905–6914.
[13] Becker, B., & Schmidt, T. (2016). Windowbased anomaly detection for streaming data in Apache Flink. IEEE International Conference on Big Data, 1817–1823.
[14] Roy, A., & Chen, M. (2017). Comparative study of fixed vs. dynamic windowing in anomaly detection. Journal of Big Data Analytics, 2(1), 45–59.
[15] Rao, R., & Kumar, D. (2018). Retail analytics using realtime event streams. Retail Intelligence Review, 6(2), 34–46.
[16] Alam, M. N., & Khan, R. (2019). LSTM-based anomaly detection in transactional retail data. Procedia Computer Science, 151, 586–593.
[17] Patel, S., & Bhardwaj, A. (2020). Towards adaptive anomaly detection for retail IoT using sessionbased windows. Internet of Things Journal, 7(9), 8647–8656.
[18] Santos, L., & Toma, M. (2021). A survey on windowing techniques in stream analytics for anomaly detection. ACM Computing Surveys, 54(5), 1–38.
[19] Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.
[20] Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., & Tzoumas, K. (2015). Apache Flink: Stream and batch processing in a single engine. IEEE Data Engineering Bulletin, 38(4), 28–38.
[21] Katsifodimos, A., & Fragkoulis, M. (2019). Operational stream processing: Towards scalable and consistent event-driven applications. Proceedings of the VLDB Endowment, 12(12), 2006–2019.
[22] Bifet, A., & Gavalda, R. (2007). Learning from timechanging data with adaptive windowing. Proceedings of the 2007 SIAM International Conference on Data Mining, 443–448.
[23] Krishnan, S., Franklin, M. J., & Goldberg, K. (2018). A case for sequencebased feature extraction in realtime data. Proceedings of the VLDB Endowment, 11(12), 2070–2073.
[24] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 387–395.
[25] Li, C., & Zhao, Y. (2016). Realtime anomaly detection and dashboarding for event-driven systems. ACM Transactions on Internet Technology (TOIT), 16(3), 1–24.
[26] Kaggle. (2021). Retail Transaction Datasets. https://www.kaggle.com/
[27] Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2014). Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, 26(9), 2250–2267.
[28] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. Proceedings of the 23rd European Symposium on Artificial Neural Networks, 89–94.
[29] Zimek, A., Schubert, E., & Kriegel, H.P. (2012). A survey on unsupervised outlier detection in high dimensional numerical data. Statistical Analysis and Data Mining: The ASA Data Science Journal, 5(5), 363–387.
[30] Xu, D., & Shelton, C. R. (2010). Anomaly detection for discrete sequences: A generative model approach. Journal of Machine Learning Research, 11, 1537–1567.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[31] Zhao, D., Wang, L., & Shi, X. (2021). Adaptive window anomaly detection for fluctuating transactional data. Information Sciences, 564, 229–243.
[32] Kim, S., Wang, Z., & Tang, J. (2022). Contextaware stream analytics: Adaptive windows for realtime event detection. Information Systems, 104, 101883.
[33] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1–19.
[34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
[35] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
[36] Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002). Models and issues in data stream systems. Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, 1–16.
[37] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448–3470.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.