Chosen Algorithm and Effectiveness

For real-time anomaly detection, I implemented a sliding window-based **Z-Score** algorithm. This method dynamically computes the mean and standard deviation using a set window size (50 points in this case) to identify anomalies as the data stream updates.

When a new data point is generated, the Z-score is computed to measure how much the data point deviates from the most recent mean. If the Z-score exceeds the set threshold (3 in this case), the point is flagged as an anomaly.

Reasons for Effectiveness:

- **Adaptability:** Updates the mean and standard deviation on every data point entry, which adapts the algorithm to slight changes and seasonal variations.
- **Efficiency:** Restricts computation to a set number of recent data points exclusively, which makes the sliding window approach optimize memory usage and enable the algorithm to handle continuous data in real-time.
- **Accuracy:** Detects significant deviations and naturally filters out minor fluctuations through threshold-based detection and adaptive calculations, which makes it robust against noise and reducing false positives.

Isolation Forest (IF) and **Local Outlier Factor (LOF)** are algorithms that could've been used for this project. However, I chose Z-Score detection for these reasons:

- **Computational Cost:** The computational cost of the Z-Score algorithm is lower compared to Isolation Forest and Local Outlier Factor, requiring only a few calculations per point, which makes it ideal for real-time applications.
- **Interpretability:** Z-Score's threshold is intuitive and interpretable compared to Isolation Forest which depends on isolation and LOF whose accuracy is significantly parameter-sensitive (e.g., number of neighbours).
- **Adaptability to Drift:** The Z-Score algorithm adapts well to concept drift using the sliding window mechanism. However, IF needs to be retrained for drift, while LOF's adaptability comes with heavy computational cost.
- **Seasonal Data: The w**indow-based Z-Score is the most effective algorithm for handling seasonal data, as it continuously updates the mean and standard deviation, allowing it to accurately identify anomalies without requiring retraining, unlike IF and LOF, which may struggle with seasonal variations.

Overall, this Z-Score anomaly detection strikes a balance between accuracy and computational efficiency, ensuring the system can swiftly identify anomalies in a high-throughput data stream while adapting to changing patterns.