# Chosen Algorithm: Z-Score Anomaly Detection

In this project, Z-score anomaly detection is implemented, a statistical method used to flag outliers based on the number of standard deviations a data point is from the rolling mean of a data stream. The Z-score is calculated as follows:

Where:

* X is the current data point,
* μ is the rolling mean,
* σ is the rolling standard deviation.

If the absolute Z-score (∣Z∣) exceeds a certain threshold (typically 3), the point is considered anomalous. The algorithm utilizes a sliding window to maintain a fixed number of recent points for calculating the rolling mean and standard deviation. This approach allows the method to dynamically adapt to changes in the data stream while identifying extreme deviations.

## Key Features of the Z-Score Method:

* **Real-Time Detection**: The algorithm processes each data point as it arrives, flagging anomalies on the fly without waiting for the entire stream to finish.
* **Adaptability**: By using a rolling window for the mean and standard deviation, the algorithm can adjust to slow trends and periodic patterns (such as seasonality) while still identifying sharp deviations.
* **Concept Drift Handling**: By regularly updating the mean and standard deviation based on recent data, the method effectively handles concept drift, which refers to slow changes in the statistical properties of the data stream.

## Effectiveness:

* **Simplicity**: Z-score anomaly detection is computationally simple, making it suitable for real-time applications. The incremental calculation of the rolling mean and standard deviation ensures that the algorithm is fast, even for large data streams.
* **Anomaly Precision**: The method is effective at detecting anomalies in relatively stable environments where deviations from the mean are rare. It performs well when anomalies are significant deviations from the recent normal pattern (e.g., sudden spikes).
* **Limitations**: Z-score detection may be less effective when anomalies are subtle or when the underlying distribution of the data is not Gaussian. In such cases, more advanced methods, such as machine learning models (e.g., Isolation Forest, LSTM), might offer better performance.

## Why Z-Score?

Z-score method is chosen because it is well-suited for this task, where anomalies are expected to be noticeable deviations (e.g., spikes or drops in the data stream). The simplicity and efficiency of the algorithm make it ideal for real-time anomaly detection in continuous data streams, such as financial transactions or system metrics