Voting Ensemble for Anomaly Detection

Algorithm Overview

The **Voting Ensemble** for anomaly detection combines the predictions of multiple algorithms to improve overall performance and robustness. This approach leverages the strengths of individual models—**RCTreeForest**, **QuantileBasedAnomalyDetector**, **EMAAnomalyDetector**, and **SHESDAnomalyDetector**—which utilize a sliding window mechanism to adapt to concept drift and seasonal variations in the data.

1. **RCTreeForest**:
    - A random cut tree-based algorithm that creates an ensemble of decision trees to identify anomalies. It effectively detects anomalies based on data distributions and is robust to noise.
2. **QuantileBasedAnomalyDetector**:
    - This detector uses quantiles within a sliding window to establish dynamic thresholds for identifying anomalies. By adapting to changes in data distribution, it effectively captures seasonal patterns and concept drift.
3. **EMAAnomalyDetector**:
    - The Exponential Moving Average (EMA) method smooths the data over time, providing a dynamically adjusted baseline for anomaly detection. It is particularly effective for detecting gradual changes and shifts in data trends.
4. **SHESDAnomalyDetector**:
    - This method employs a Seasonal Hybrid Extreme Studentized Deviate (SH-ESD) approach to identify anomalies in seasonal data. It effectively combines statistical tests to detect outliers while considering the seasonal context.

Effectiveness

The Voting Ensemble approach provides several advantages:

- **Increased Robustness**: By aggregating predictions from multiple models, the ensemble reduces the likelihood of false positives and false negatives, leading to more reliable anomaly detection.
- **Adaptability**: Using sliding windows across all detectors allows the ensemble to dynamically adjust to concept drift and seasonal variations, making it suitable for real-time applications.
- **Improved Accuracy**: The combination of diverse algorithms captures different aspects of the data, enhancing the overall detection accuracy. Each algorithm's unique strengths help mitigate the weaknesses of others.
- **Real-Time Performance**: The ensemble can efficiently process continuous data streams while maintaining low latency, making it suitable for applications in finance, cybersecurity, and system monitoring.

In summary, the Voting Ensemble leverages the strengths of various anomaly detection algorithms to provide a comprehensive, adaptive, and accurate solution for identifying anomalies in data streams in real time.