Analysing PMP data

Ljubomir Hrboka (CARNET), Krešimir Šodan (CARNET)

WP6-T3 GÉANT GN5-2

Network Performance and Monitoring Workshop, Prague, Czech Republic

1 April 2025

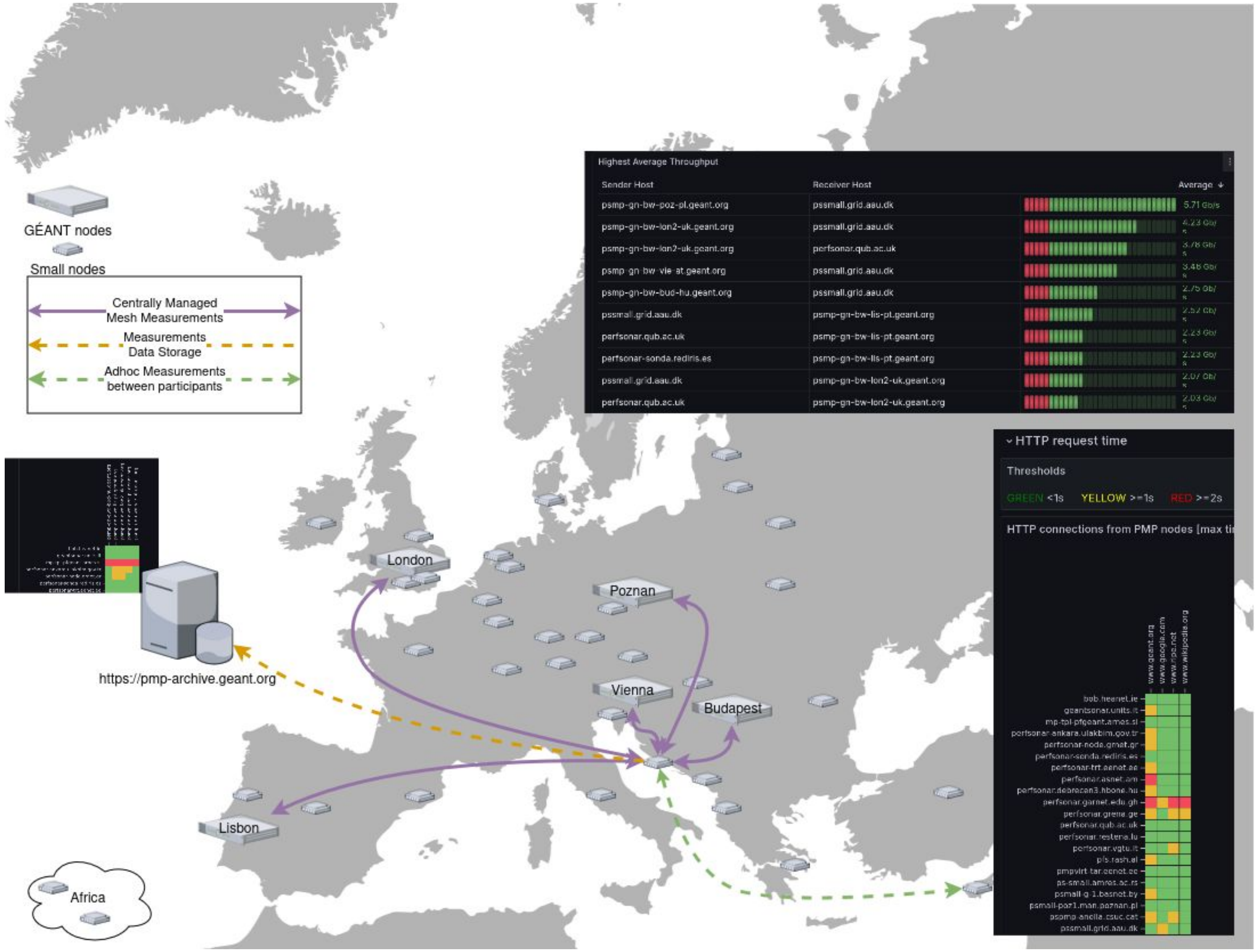
GN5-2

Agenda

- PMP
- Anomaly Detection
- Pipeline
- Two ML Approaches

GÉANT Performance Management Platform - PMP

- Latency – 1 histogram of 600 values every minute
- Jitter - 1 value every minute
- RTT - 5 values every ten minutes
- Throughput - four values a day
- HTTP, DNS, ...

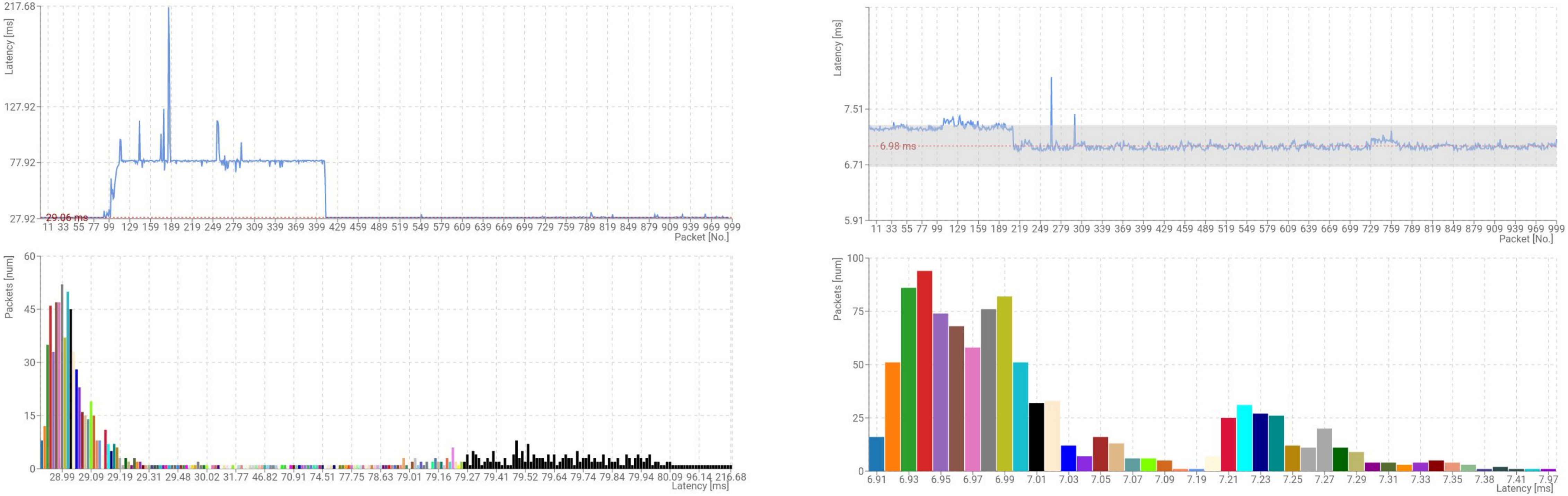


Rationale for the data analysis and anomaly detection

- Implementing anomaly detection programs on existing measurement data is a low-cost high-value investment
- Detection of the small anomalies is relevant for the application depending on low network latency
- Detecting any type of anomalies early can prevent them from escalating into major issues
- All modern data analysis software are highly data-dependent, especially for the purpose of identifying correlations

Using Latency Data for Anomaly Detection

- The substantial amount of available latency data prompted us to focus on detecting anomalies within latency measurements.



Anomaly Detection

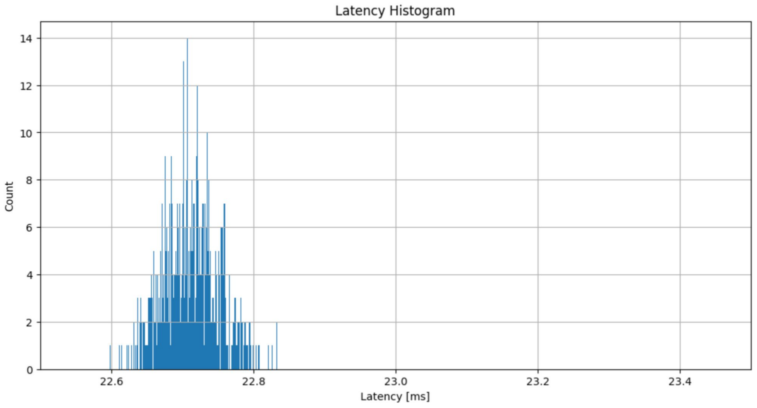
- Identifying data points that deviate significantly from the norm
- What is a norm?

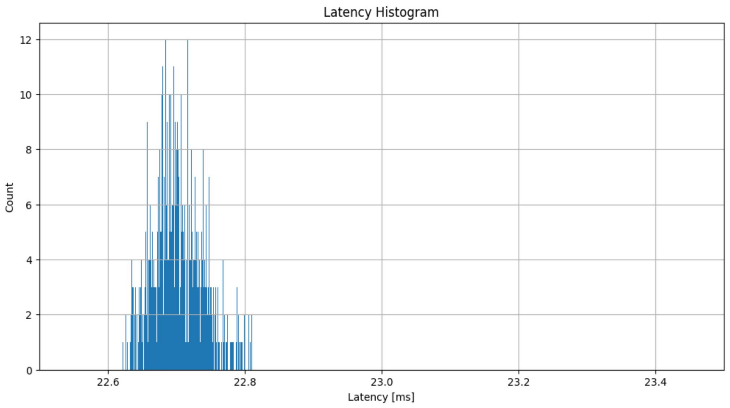
  The common shape or distribution of latencies that can be seen during healthy network conditions
- What is significant?

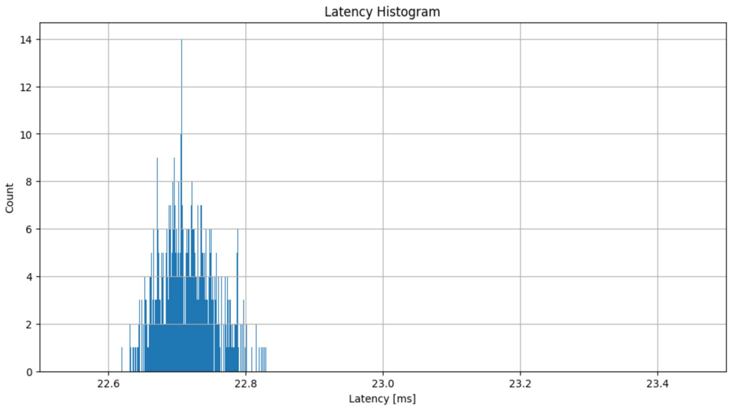
  Current observation is far enough from the normal pattern that it’s unlikely to be just noise or randomness.

Norm examples

- Most packets fall within 20–30 ms
- Little variation between histograms
- Low or consistent statistical values (mean, variance, skewness, kurtosis, etc.)

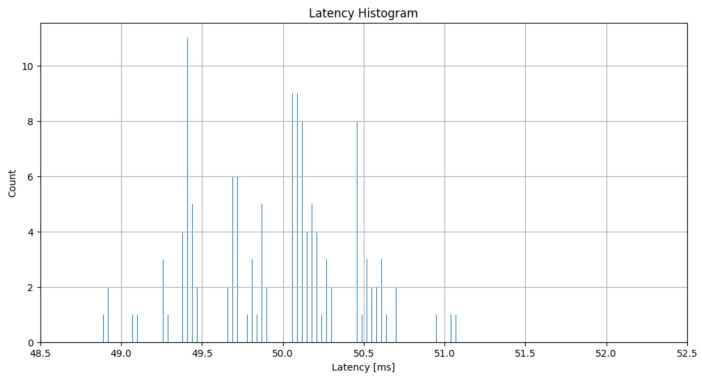


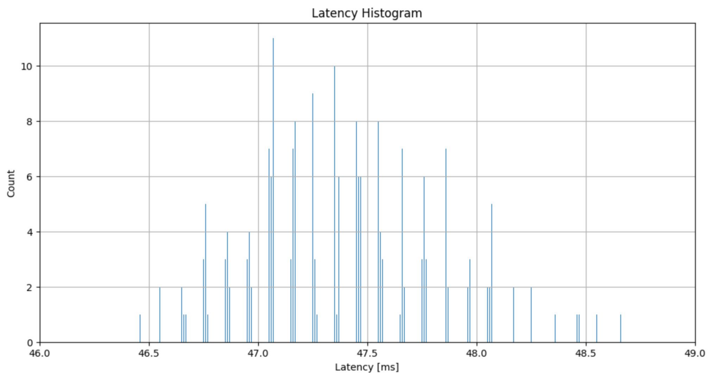


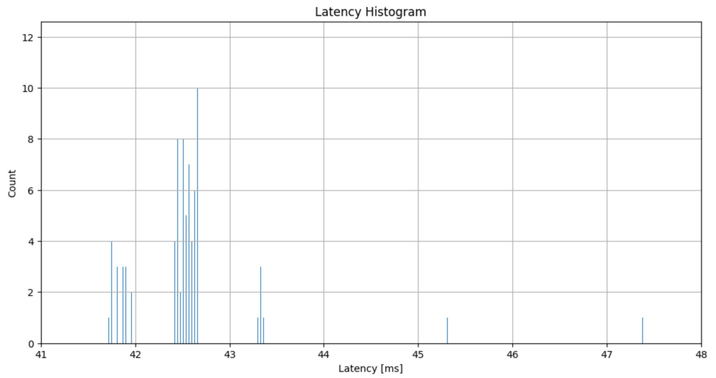


Anomaly examples

- Latencies are spread out or shifted outside the normal 20-30 ms range
- Histogram shapes vary significantly
- Shifts in the mean, increased variance, irregular skewness, abnormal kurtosis or higher entropy







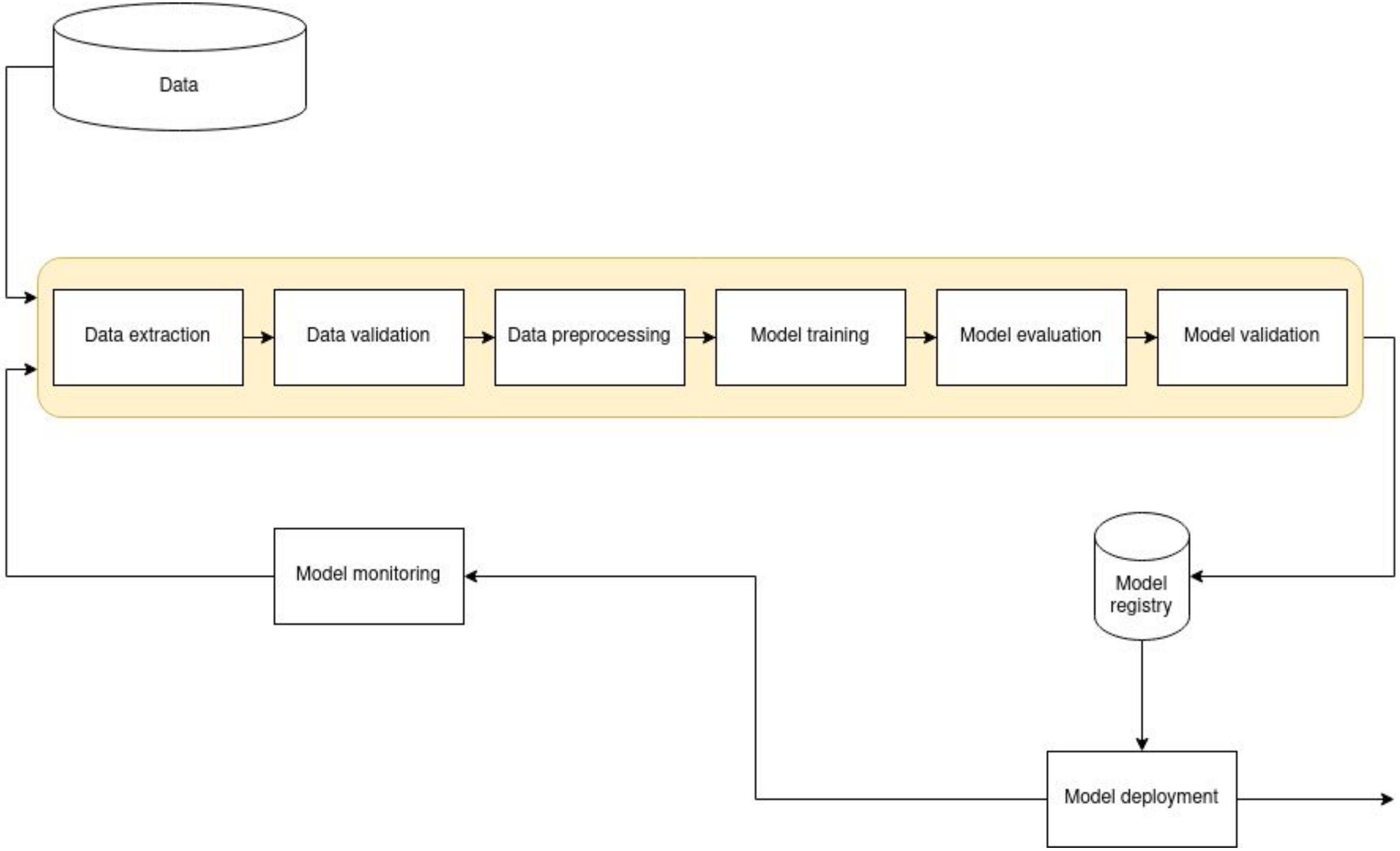
Using unsupervised machine learning

- Working with untagged data
- Finding hidden patterns and anomalies
- Generating labeled data for future analysis
- Treating every link and every direction as a new dataset

Autoencoders

- Detect anomalies by learning and reconstructing normal behavior
- Work without labeled data, making them great for untagged datasets
- Extract important features while filtering out noise
- Scalable solution for handling large volumes of data

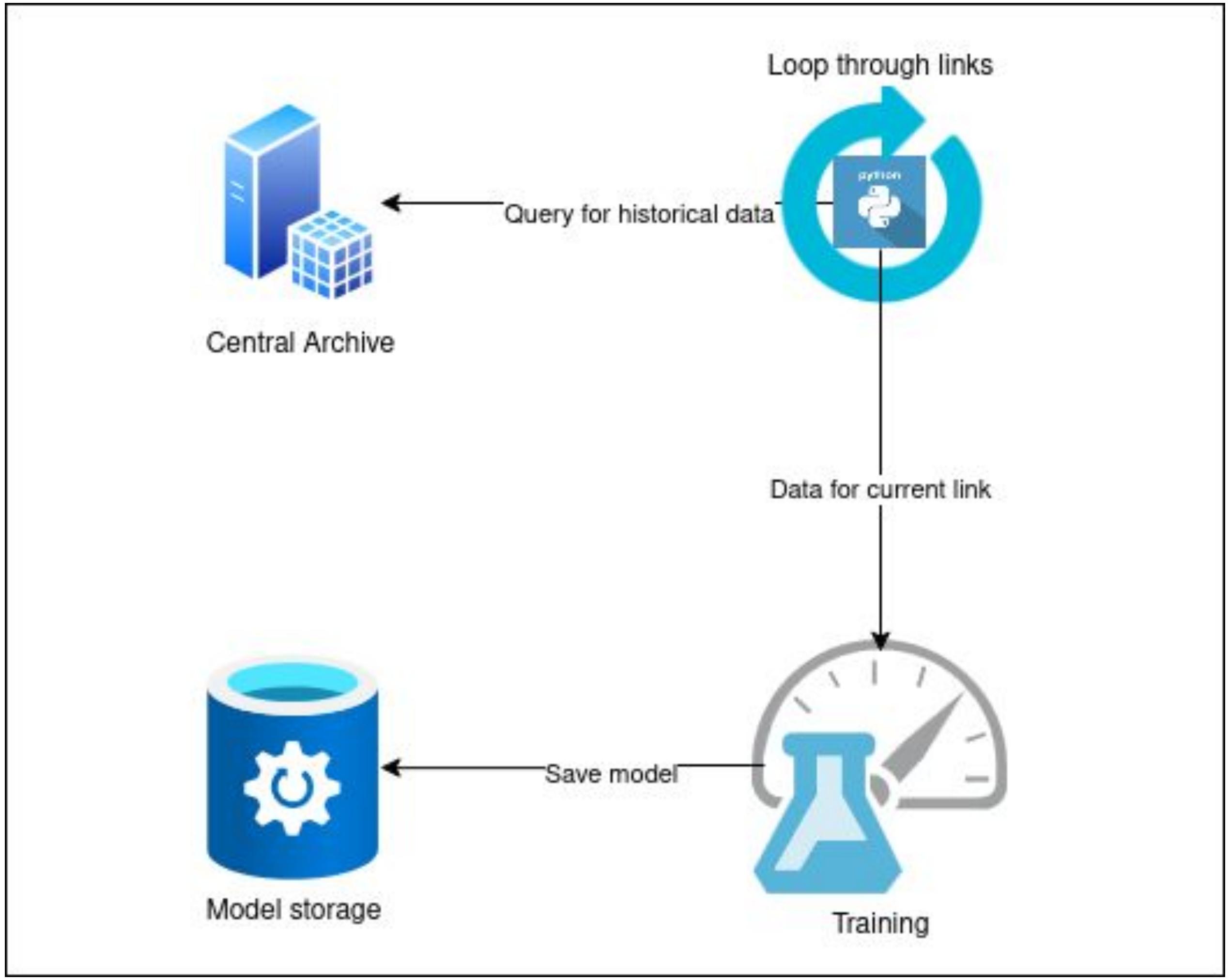
ML Model Pipeline



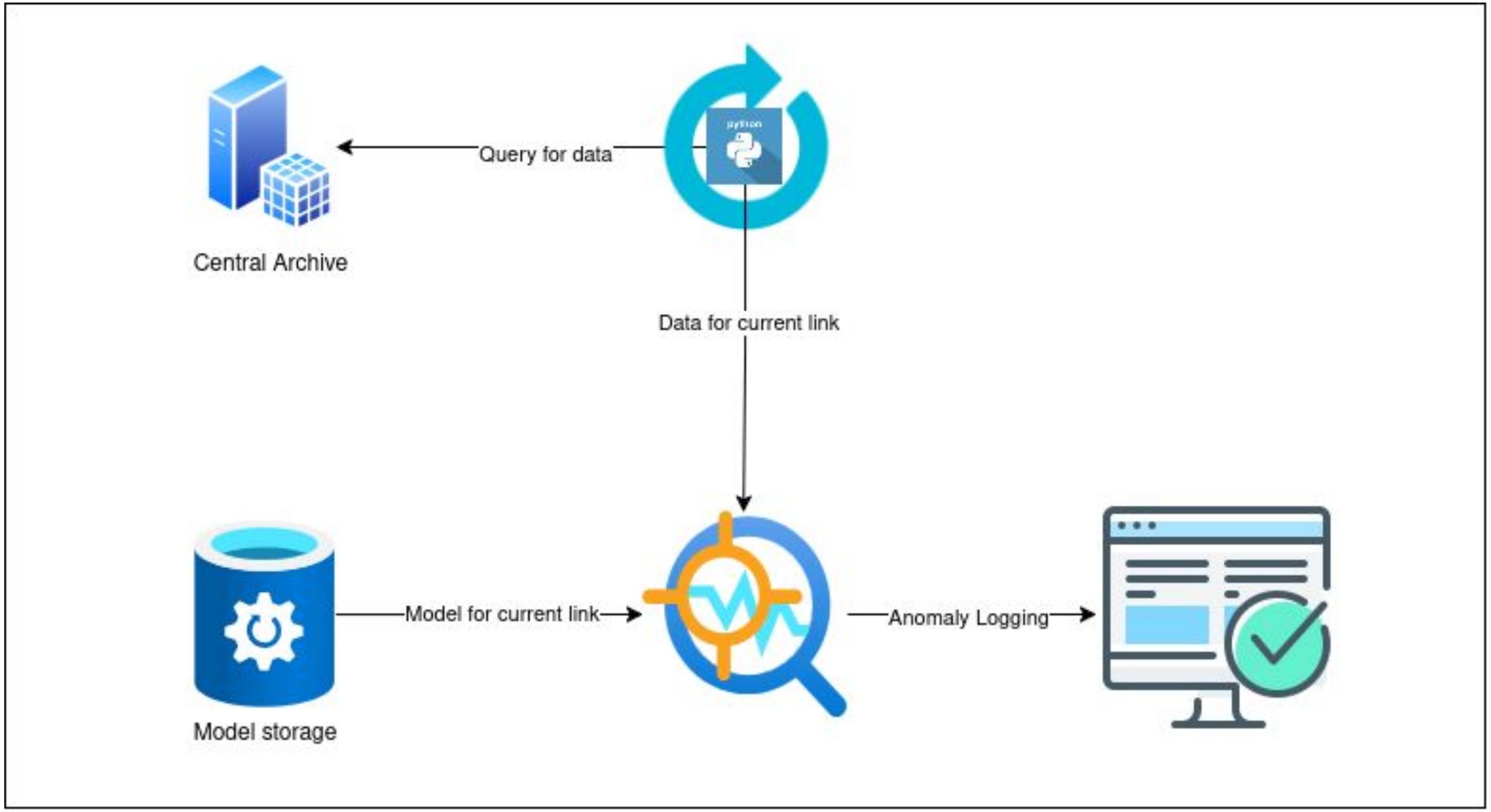
Action plan

- Choose three measurement links
- Train ML model for each of them
- Implement app to loop through data from central archive
- Perform anomaly detection
- Log results for verification and observation

Training phase



Inference phase



Two Approaches

- First approach: Resample histograms to have the same input size
  - Fixing histograms to 256 bins
  - Statistical distribution of the data is preserved
  - Histograms with different resolutions become comparable
  - Model can identify patterns across histograms, regardless of their original binning scheme
  - Allows use of predefined architecture
- Second approach: Statistical Features Extraction
  - Calculate various statistical features of the histogram and pass them to ML model
  - Mean, Variance, Skewness, Kurtosis, Entropy, Median, Interquartile Range, ...

Autoencoder

- Different size of linear layers and activation layers
- ReLU, LeakyReLU activation functions
- MSELoss, L1Loss, SmoothL1Loss loss functions
- Adam optimizer
- 30-60 epoch training

Data preprocessing

- Better Results: Clean data helps the model learn patterns more effectively.
- Removes Noise: Gets rid of irrelevant or incorrect data.
- Speeds Up Training: Simplified data makes training faster.
- Ensures Fairness: Makes sure all features are treated equally.
- Keeps Things Consistent: Standardized data improves reliability.

OWD measurements – raw data

- 'ts': 1706918450, 'val': {'126.09': 1, '126.12': 1, '126.13': 1, '126.18': 1, '126.21': 2, '126.22': 1, '126.23': 3, '126.24': 4, '126.25': 1, '126.26': 1, '126.27': 1, '126.28': 2, '126.29': 1, '126.30': 1, '126.32': 5, '126.33': 1, '126.34': 1, '126.35': 1, '126.36': 3, '126.37': 1, '126.38': 4, '126.39': 3, '126.40': 6, '126.41': 4, '126.42': 6, '126.43': 1, '126.44': 1, '126.45': 6, '126.46': 4, '126.47': 3, '126.48': 5, '126.49': 5, '126.50': 7, '126.51': 3, '126.52': 4, '126.53': 5, '126.54': 2, '126.55': 4, '126.56': 2, '126.57': 5, '126.58': 5, '126.59': 5, '126.60': 3, '126.61': 2, '126.62': 5, '126.63': 4, '126.64': 4, '126.65': 5, '126.66': 1, '126.67': 7, '126.68': 8, '126.69': 7, '126.70': 6, '126.71': 2, '126.72': 3, '126.73': 6, '126.74': 7, '126.75': 4, '126.76': 6, '126.77': 8, '126.78': 4, '126.79': 4, '126.80': 5, '126.81': 5, '126.82': 6, '126.83': 11, '126.84': 4, '126.85': 5, '126.86': 6, '126.87': 9, '126.88': 6, '126.89': 9, '126.90': 10, '126.91': 9, '126.92': 2, '126.93': 10, '126.94': 4, '126.95': 5, '126.96': 5, '126.97': 8, '126.98': 5, '126.99': 5, '127.00': 5, '127.01': 7, '127.02': 4, '127.03': 7, '127.04': 5, '127.05': 6, '127.06': 3, '127.07': 7, '127.08': 2, '127.09': 5, '127.10': 5, '127.11': 1, '127.12': 5, '127.13': 7, '127.14': 2, '127.15': 3, '127.16': 6, '127.17': 8, '127.18': 5, '127.19': 7, '127.20': 4, '127.21': 2, '127.22': 4, '127.23': 3, '127.24': 4, '127.25': 3, '127.26': 2, '127.27': 4, '127.28': 5, '127.29': 6, '127.30': 4, '127.31': 5, '127.32': 1, '127.33': 5, '127.34': 7, '127.35': 3, '127.36': 3, '127.37': 1, '127.38': 3, '127.39': 4, '127.40': 1, '127.41': 1, '127.42': 1, '127.43': 3, '127.44': 5, '127.45': 1, '127.46': 3, '127.47': 4, '127.48': 4, '127.49': 2, '127.50': 1, '127.51': 1, '127.52': 2, '127.53': 4, '127.54': 2, '127.55': 1, '127.56': 2, '127.57': 2, '127.58': 1, '127.59': 2, '127.60': 4, '127.61': 2, '127.62': 2, '127.63': 2, '127.66': 2, '127.67': 4, '127.68': 3, '127.69': 2, '127.72': 1, '127.74': 1, '127.75': 1, '127.77': 1, '127.80': 1, '127.85': 1, '127.86': 2, '127.88': 1, '127.90': 1, '127.93': 1, '128.00': 1, '128.24': 1, '128.41': 1, '129.51': 1}}

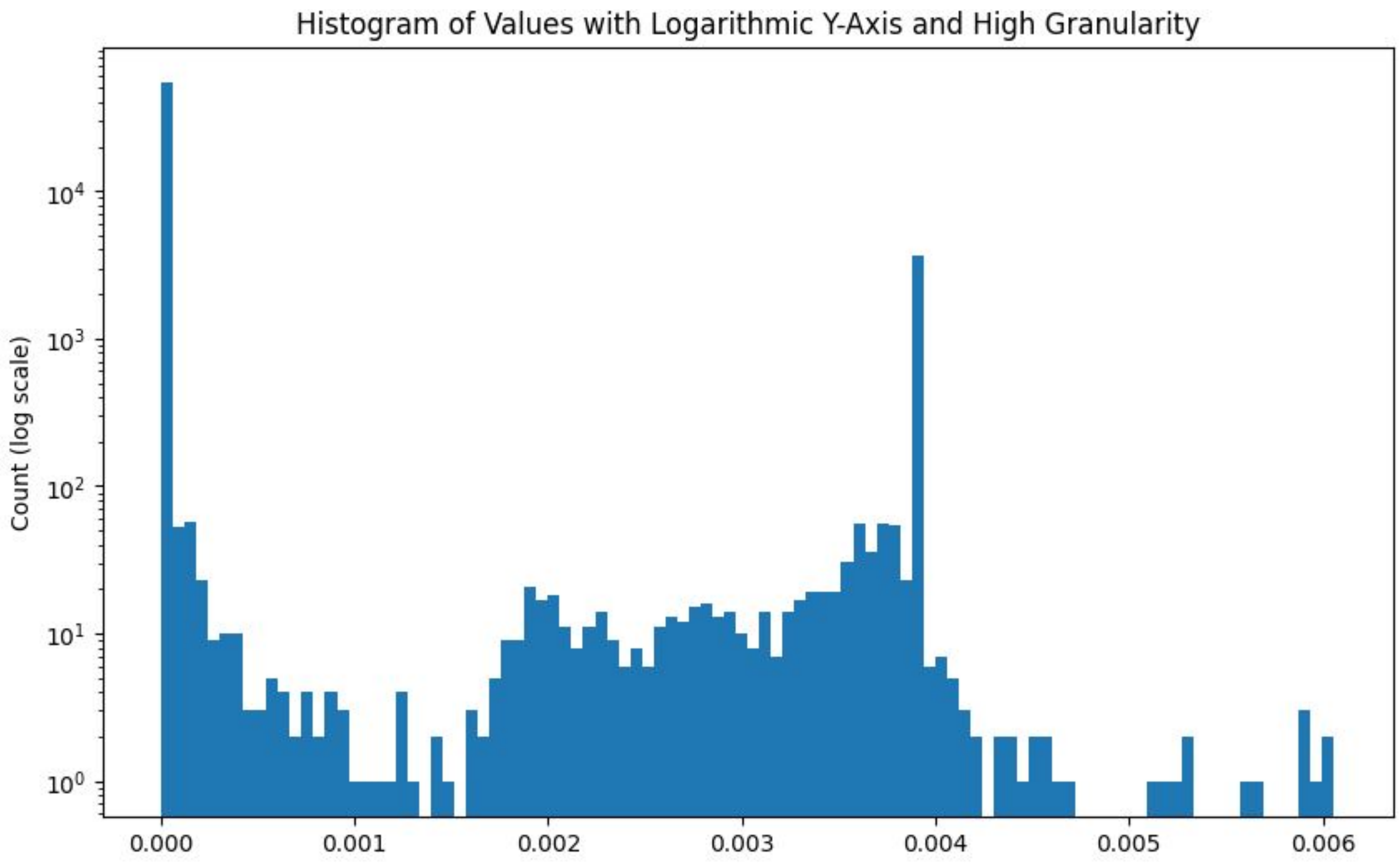
- 600 points in one measurement each minute

OWD measurements – processed data

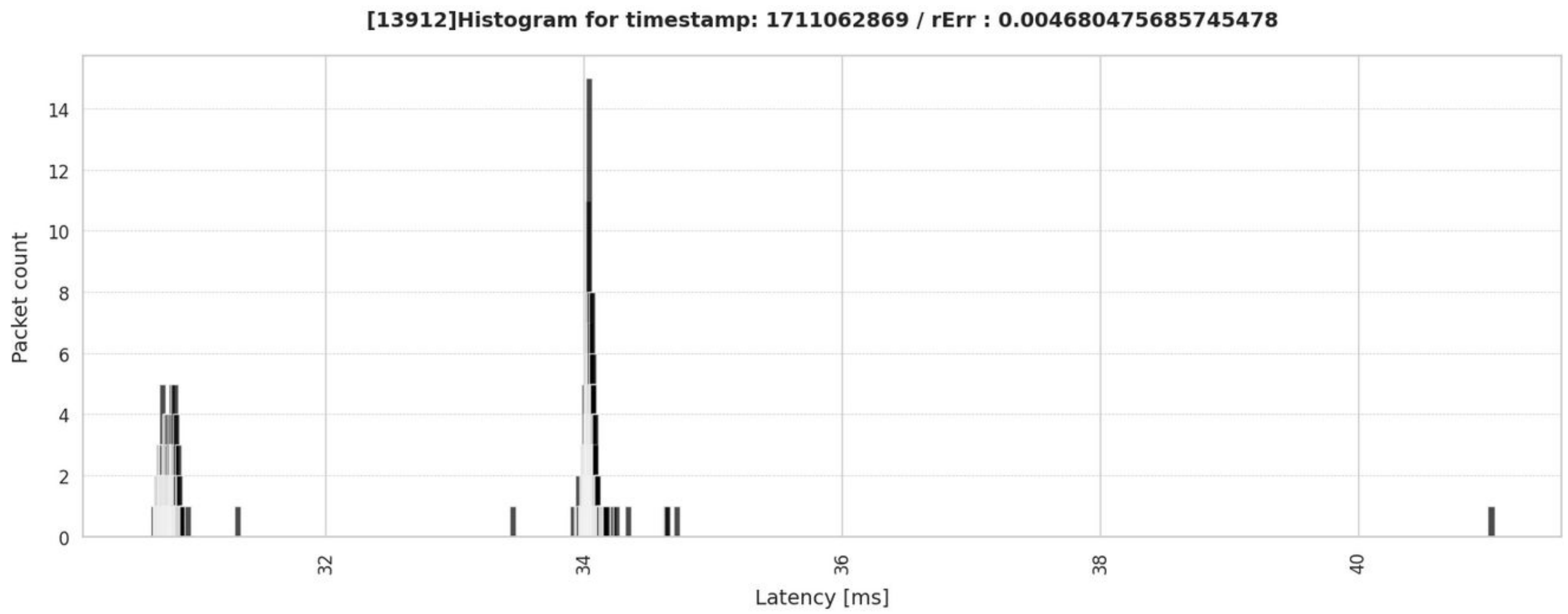
- Each measurement is fixed-size vector of 256 numbers
- Normalized [0,1]

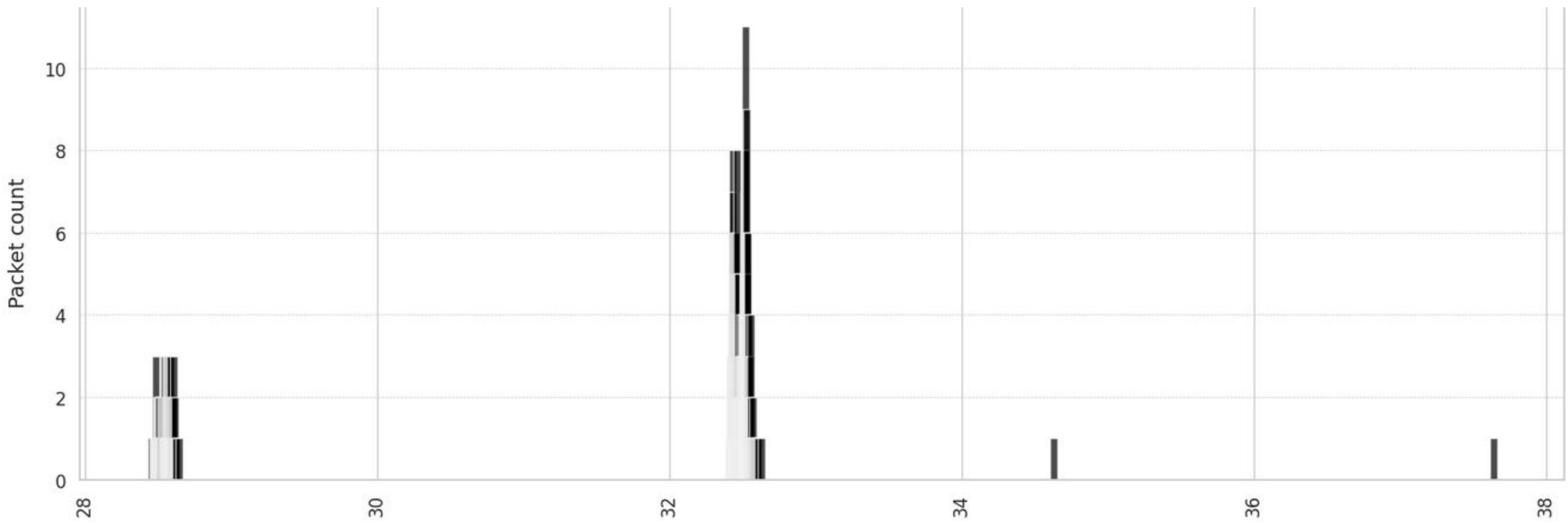
```
tensor([0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0017, 0.4350, 0.5433, 0.0183, 0.0017, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000,
        0.0000, 0.0000, 0.0000, 0.0000])
```

First approach: OWD measurements - loss distribution

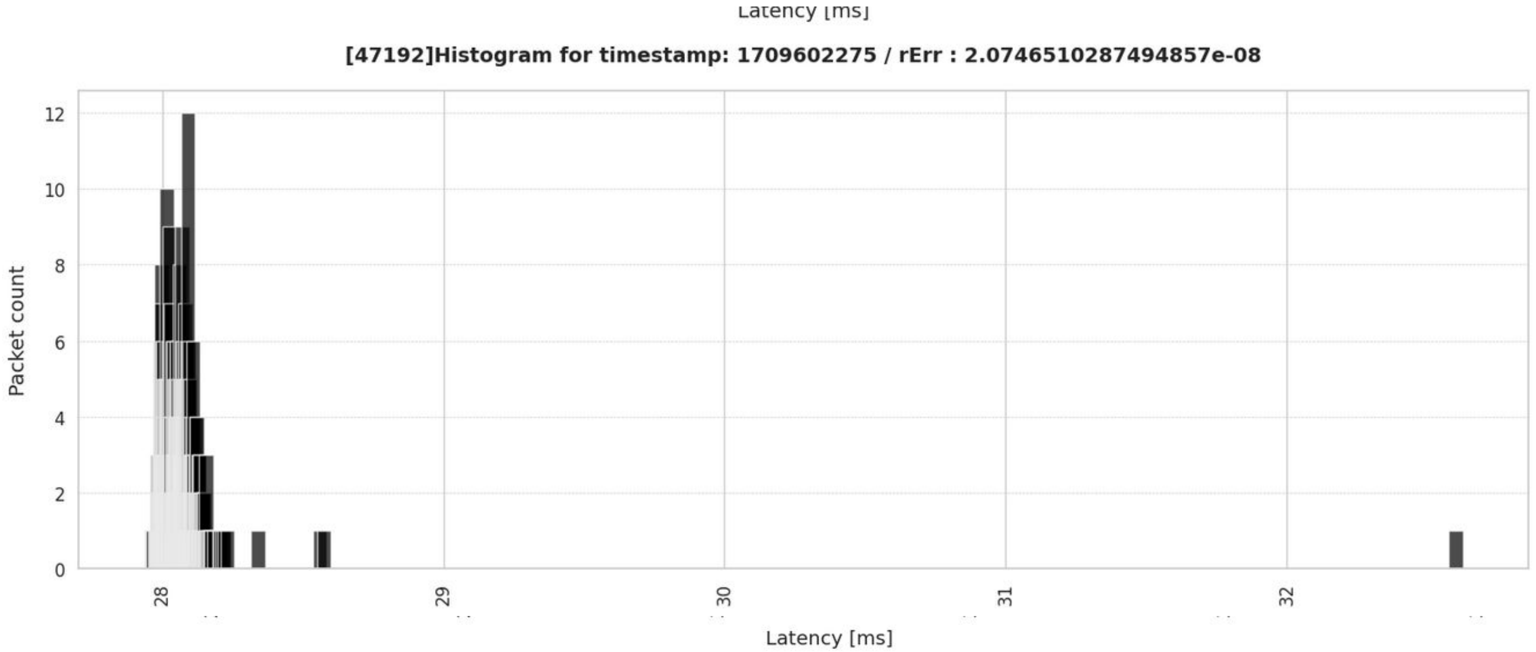


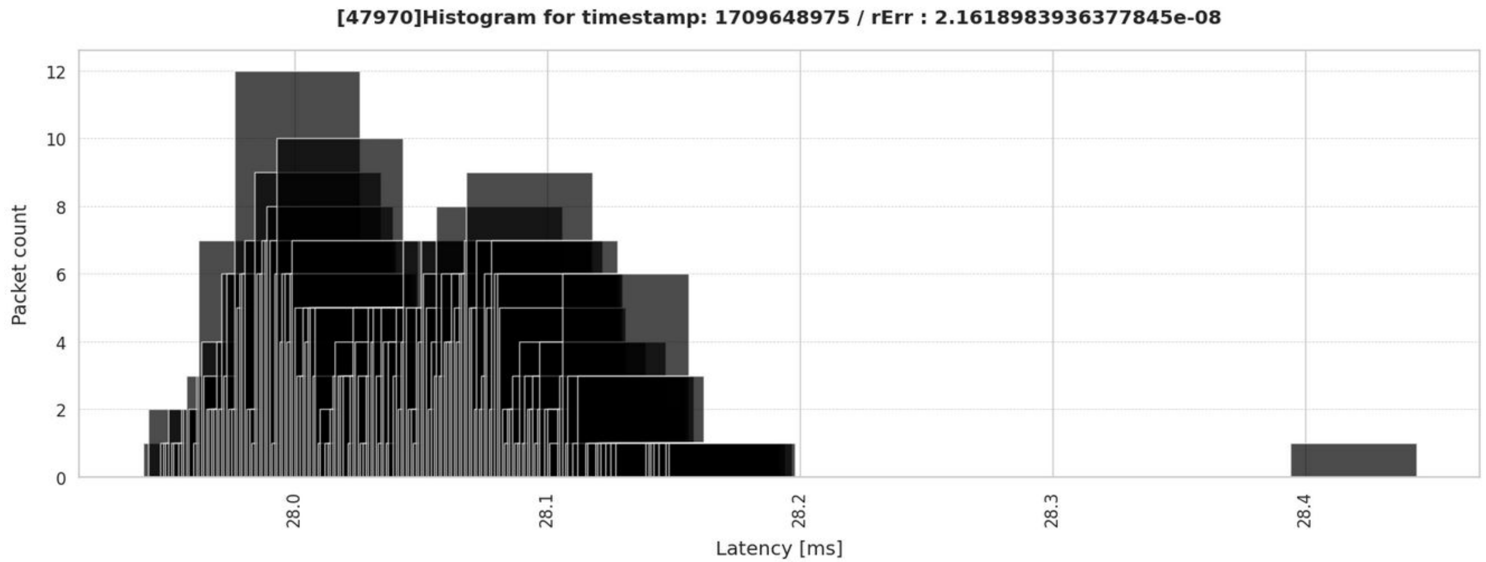
First approach: OWD measurements – anomalies



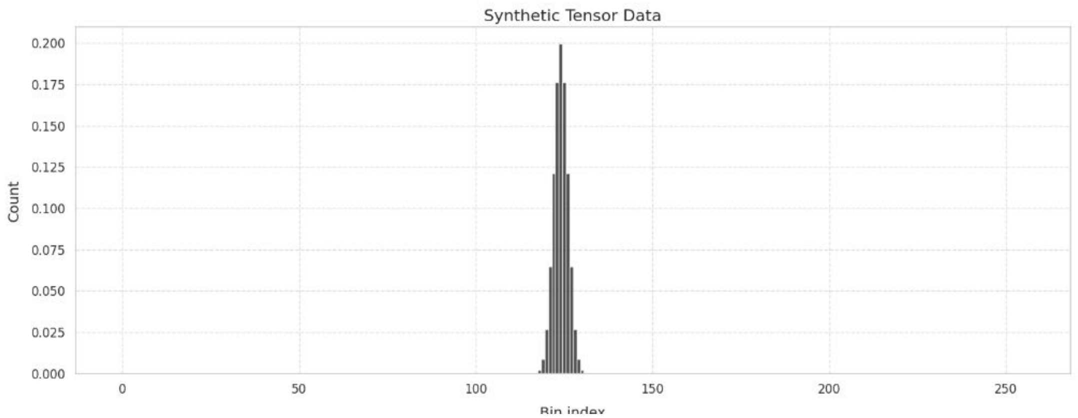


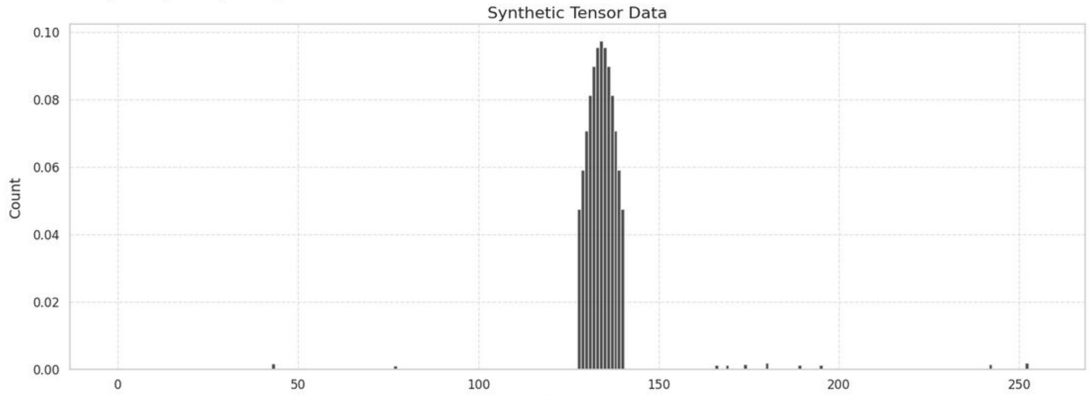
First approach: OWD measurements – normal data

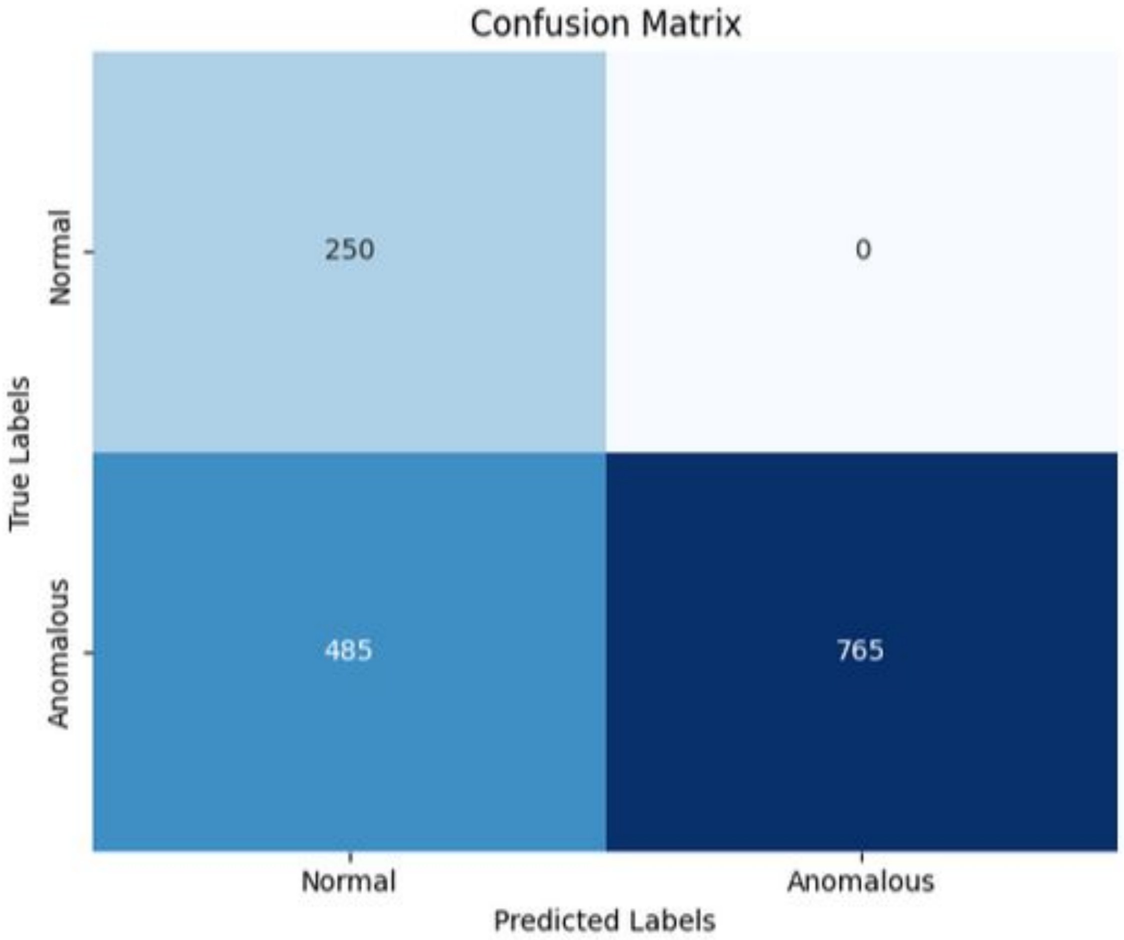




First approach: Model Evaluation







First approach: Analysis of Results

- Accuracy - 67.67%
- Precision - 100%
- Sensitivity - 61.2%
- Specificity – 100%
- F1 Score - 75.8%

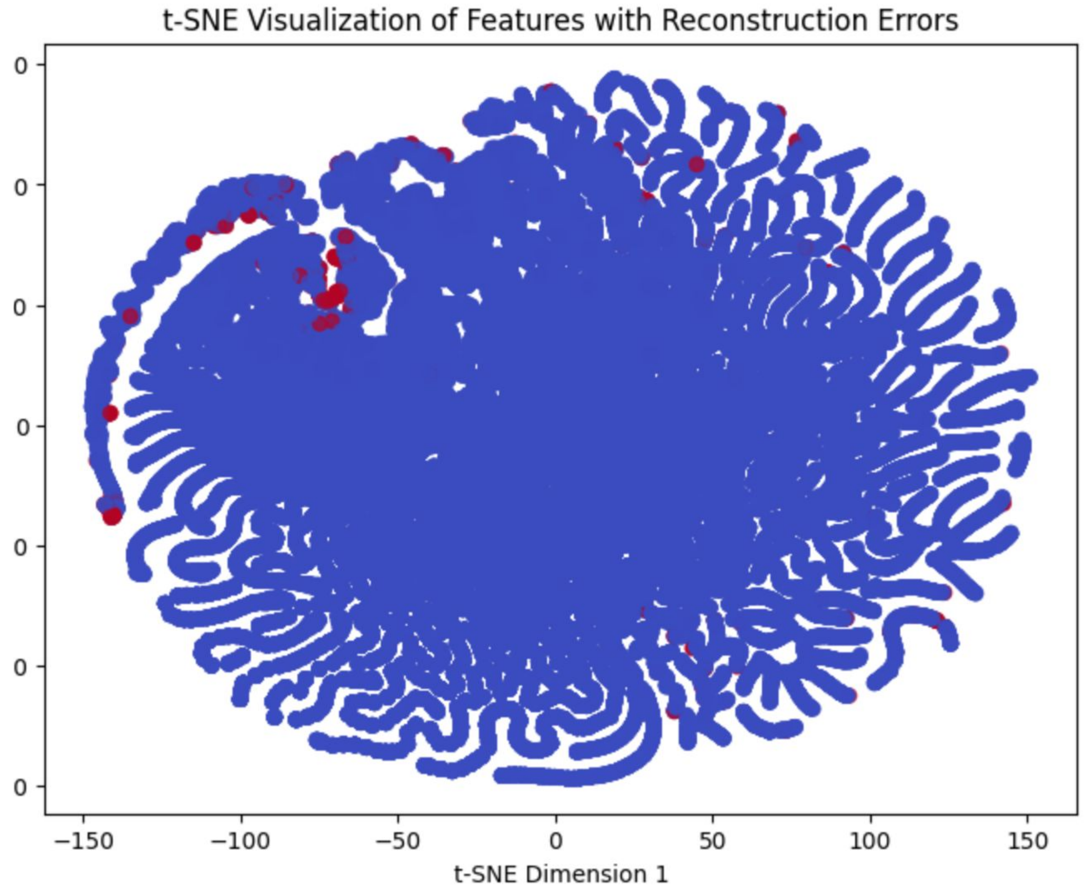
- Moderate Accuracy - many False Negatives
- High Precision - avoiding false alarms for anomalies
- Low Sensitivity - significant number of actual anomalies missed
- High Specificity - no misclassification of Normal samples

Second approach: Statistical Features

- Mean: Average value of the bins.
- Variance: Measure of spread in the data.
- Skewness: Measure of the asymmetry of the distribution.
- Kurtosis: Measure of how heavy the tails are.
- Entropy: Quantifies randomness in the histogram.
- 25th Percentile: The value below which 25% of the data falls.
- Median: The middle value, dividing the data into two halves.
- 75th Percentile: The value below which 75% of the data falls.
- Interquartile Range: Measures the spread of the middle 50% of the data

Second approach - Summary

- Many red points are distinctly separated from the blue clusters, which suggests that the features and reconstruction errors provide a meaningful distinction between normal and anomalous data.
- This separation is a strong indicator of the effectiveness of feature extraction and dimensionality reduction.
- Red points within or very close to the blue clusters, suggest reduction in detection accuracy



Thank You

www.geant.org

Co-funded by
the European Union