PhD candidate Roberts Volkovičs
Vidzeme University of Applied Sciences
Socio-technical Systems Modelling
(Prof., Dr.sc.ing. Artis Teilāns Rezekne Academy of Technologies)

«Methods and AI solutions for anomaly detection and their use cases in modern work environment»

PhD research report 2024

TED4LAT
Funded by the European Union

Relationship of anomaly detection and AI

First let’s go through already known methods of anomaly detection then we can find **relationship** of modern framework of **anomaly detection tool** which is part of Cognitive Services (Microsoft AI solution in Azure cloud) and **machine learning**.

TED4LAT
Funded by the European Union

Concepts – anomaly detection

In data analysis, **anomaly detection** (also referred to as **outlier detection** and sometimes as **novelty detection**) is generally understood to be the **identification of rare items, events or observations** which **deviate significantly from the majority of the data** and do not conform to a well defined notion of normal behavior.

Such examples may arouse suspicions of being generated by a different mechanism, or appear inconsistent with the remainder of that set of data. [1]

TED4LAT

Funded by
the European Union

Concepts – anomaly detection

Anomaly detection finds application in many domains including **cyber security, medicine, machine vision, statistics, neuroscience, law enforcement** and **financial fraud** to name only a few.

Anomalies were initially searched for **clear rejection or omission** from the data to aid **statistical analysis**, for example to compute the mean or standard deviation. They were also **removed to better predictions** from models such as linear regression, and more recently their removal aids the **performance of machine learning algorithms**.

However, in many applications **anomalies themselves are of interest** and are the observations most desirous in the entire data set, which need to be identified and separated from noise or irrelevant outliers. [1]

TED4LAT
Funded by the European Union

Concepts – Azure Cognitive Services

Azure Cognitive Services provide solutions in following AI domains: [2]

Speech:

- Speech to text;
- Text to speech;
- Speech translation;
- Speaker recognition;

Language:

- Entity recognition;
- Sentiment analysis;
- Question answering;
- Conversational language understanding;

TED4LAT
Funded by the European Union

Concepts – Azure Cognitive Services

Azure Cognitive Services provide solutions in following AI domains: [2]

Vision:

- Computer vision;
- Custom vision;
- Face API;

Decision:

- **Anomaly detector;**
- Content moderator;
- Personalizer;

Open AI service.

TED4LAT
Funded by the European Union

Concepts – time-series

In **mathematics**, a **time series** is a series of **data points** indexed (or listed or graphed) in **time order**. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are **heights of ocean tides**, **counts of sunspots**, and the **daily closing value** of the Dow Jones Industrial Average. [3]

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. [3]

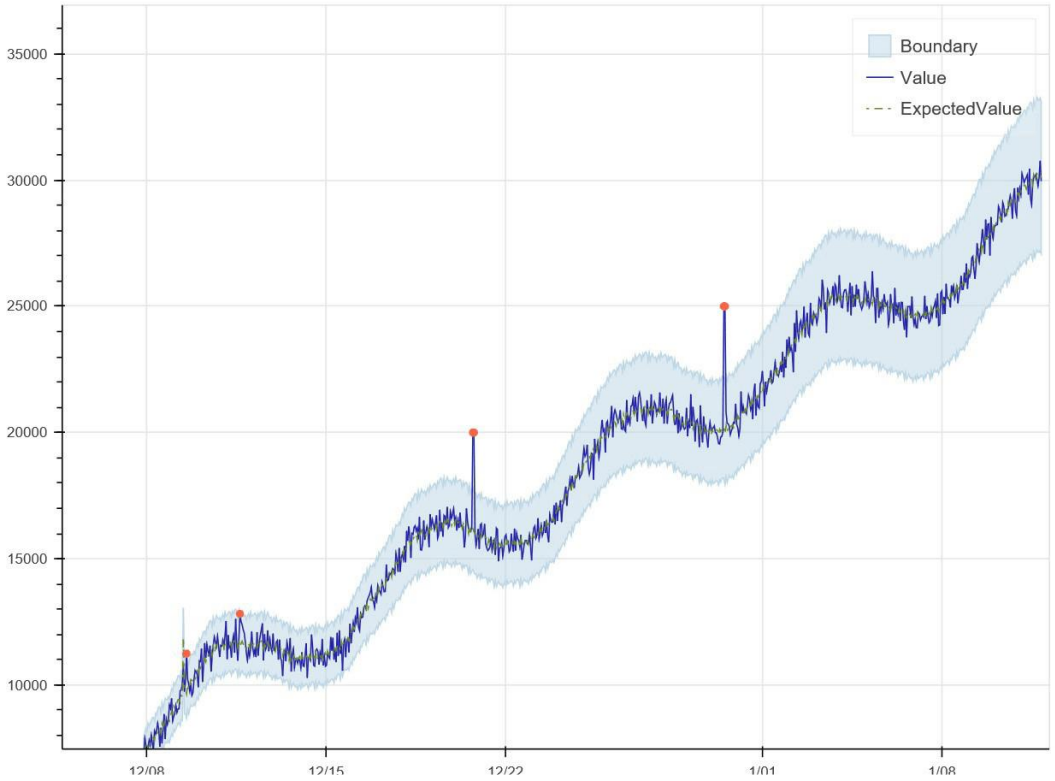
Time series forecasting is the use of a model to predict future values based on previously observed values. [3]

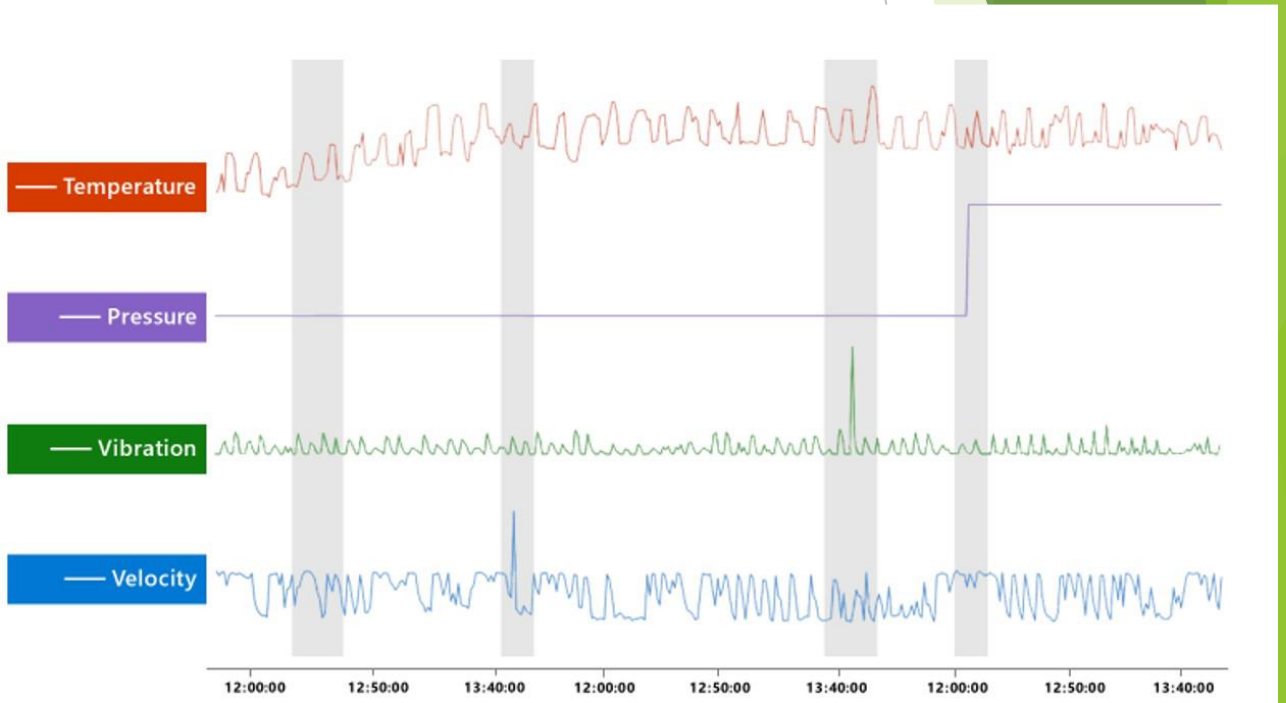
TED4LAT
Funded by the European Union

Azure Cognitive Services Anomaly detector

Anomaly detector service helps to detect anomalies in:

- **univariate** time series data;
- **multivariate** time series data;



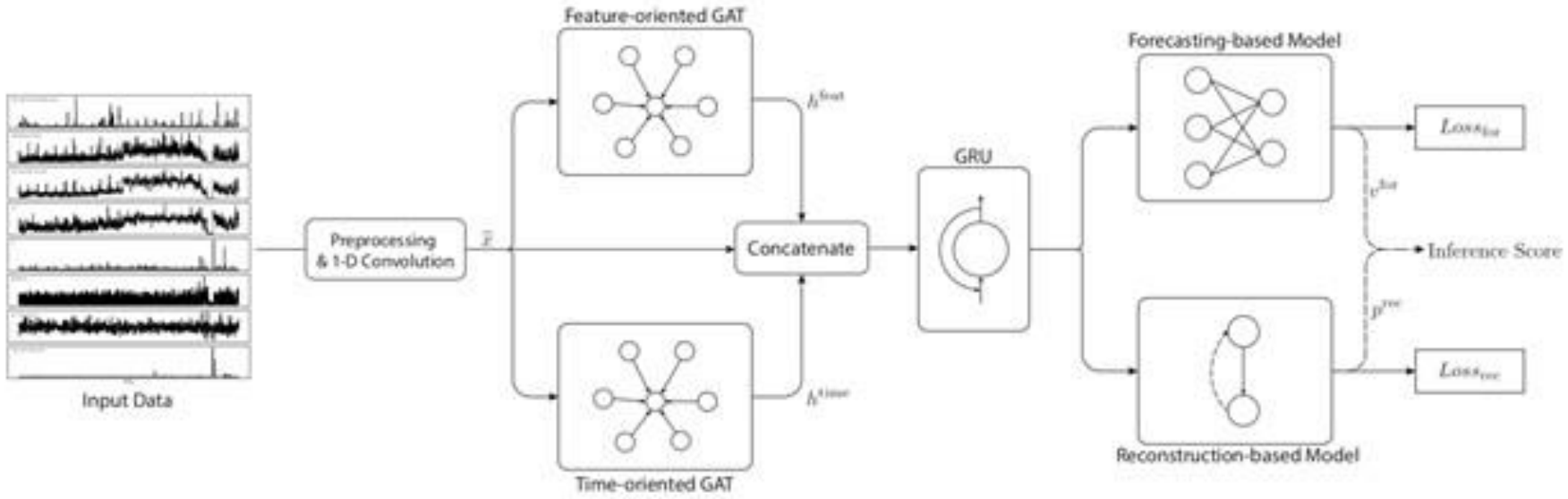


TED4LAT
Funded by the European Union

Azure Cognitive Services Anomaly detector

According to **Microsoft documentation machine learning algorithms are used only** when **working with multivariate time series**. Univariate timeseries Anomaly detector is processing using many different algorithms and mathematical methods which could be a part of processing multivariate time series as well.

Recently **Microsoft has released** the **graphical representation of the framework** which they do use in the process of multivariate anomaly detection. It was done by one of Anomaly detector project leaders **via blog** [6]. This is **not the part of Anomaly detector documentation** it is rather as description of the framework for customers and explanation on how the results are achieved.



TED4LAT
Funded by the European Union

Azure Cognitive Services Anomaly detector

Tony Xing describes the **multivariate anomaly detection process:** “In this newly introduced feature, we productized a novel framework - **MTAD-GAT (Multivariate Time-series Anomaly Detection via Graph Attention Network)**, to tackle the limitations of previous solutions. Our method considers each univariate time-series as an individual feature and tries to model the correlations between different features explicitly, while the temporal dependencies within each time-series are modeled at the same time.

The **key ingredients** in our model are **two graph attention layers**, namely the **feature-oriented** graph attention layer and the **time-oriented** graph attention layer. The **feature-oriented** graph attention layer captures the **causal relationships between multiple features**, and the **time-oriented** graph attention layer **underlines the dependencies along the temporal dimension**. In addition, we **jointly train** a **forecasting-based** model and a **reconstruction-based** model for better representations of time-series data. The two models can be optimized simultaneously by a joint objective function.” [6]

TED4LAT
Funded by the European Union

GAT – Graph Attention Networks

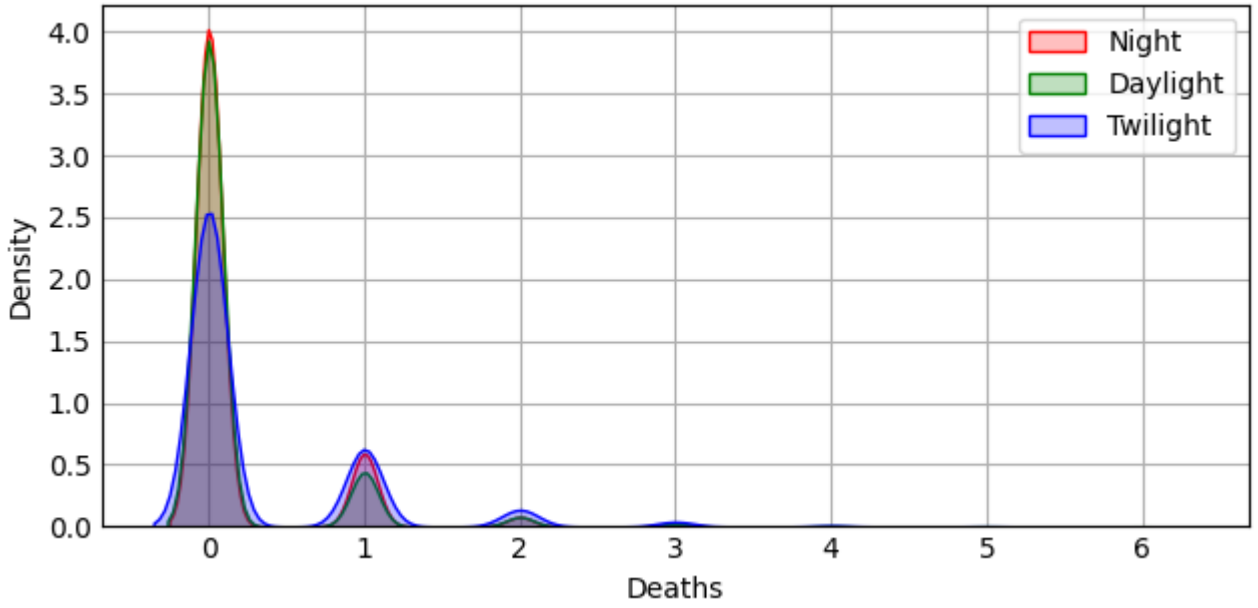
Graph Attention Networks is a **subclass of Graph Neural Networks** first presented to wider audience as a conference paper at **ICLR 2018 (Sixth International Conference on Learning Representations, Vancouver Convention Center, Vancouver CANADA)**. [7]

The **idea is** that **for GANs during matrix computations** behind the scenes **the weight of attention is calculated** for a node on **how much it should take into account other nodes in the neighborhood** during the decision making process and that is **used as learning mechanism**. Weights could change during the calculation process as new information comes from the neighborhood.

TED4LAT
Funded by the European Union

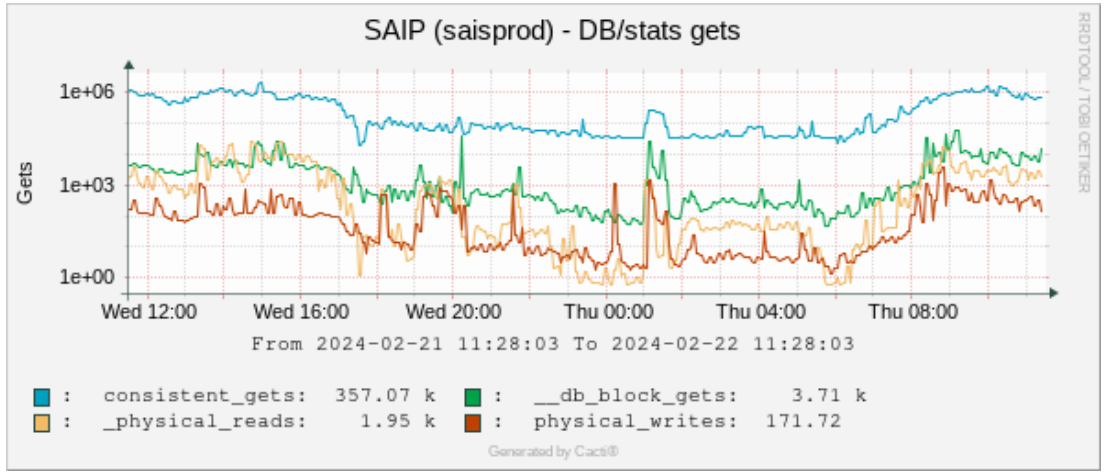
Using Anomaly Detection on open data from Latvian road meteo stations and traffic accidents

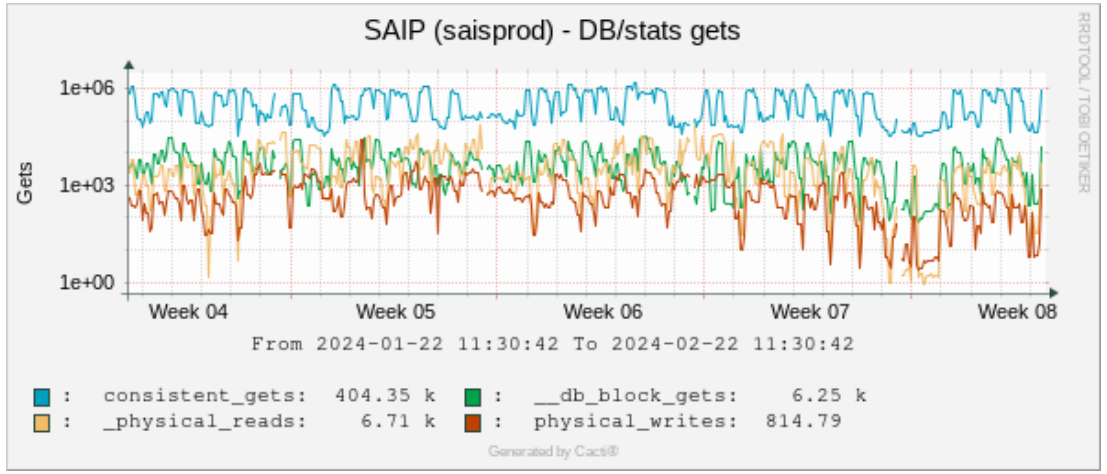
- 2 from more than 20 meteo stations do work correctly others need technical repair of sensors;
- wind of 300 km/h detected;
- finding root cause critical factors of traffic accidents.



TED4LAT
Funded by the European Union

Using Microsoft Anomaly detector to monitor Oracle DB IO operations





TED4LAT

Funded by the European Union

Using Microsoft Anomaly detector for processing monitoring data of potato field (agriculture)

Purpose: monitoring of potato field for anomalies and crop forcasting

Processed timeseries data:

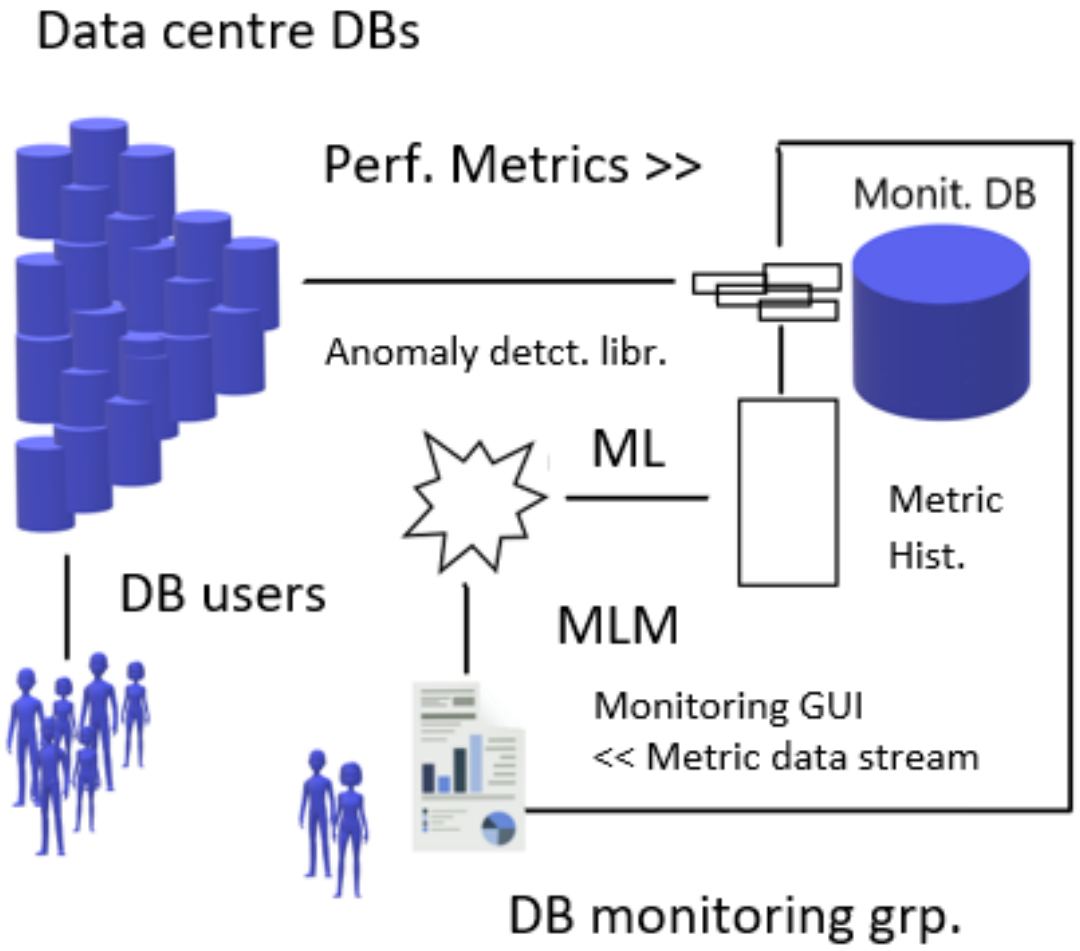
- Soil moisture at depth 1;
- Soil moisture at depth 2;
- Soil temperature;
- Air temperature vision;

One season of data processed and conclusion is that we need data from more than one season to teach AI algorithms to take seasonality into account. It appeared that even three of summer month in Latvia (June, July and August) are different from monitored factor perspective.

TED4LAT

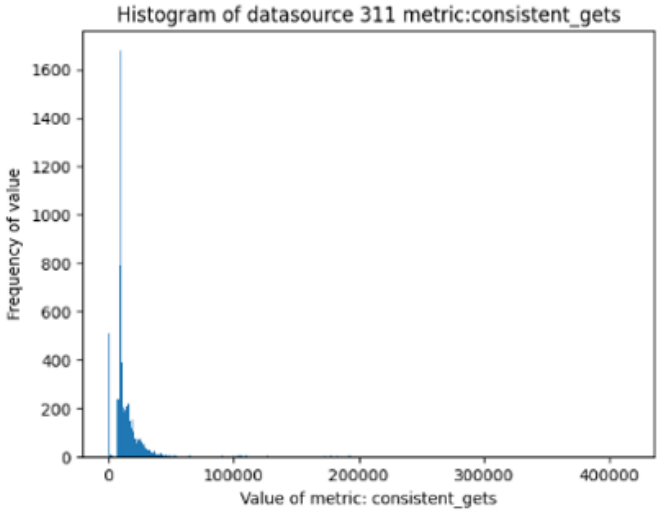
Funded by the European Union

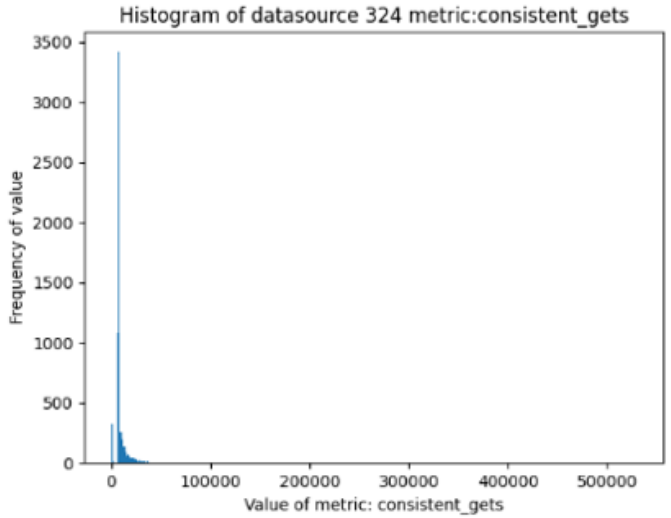
AI model of anomaly detection for data centre monitoring

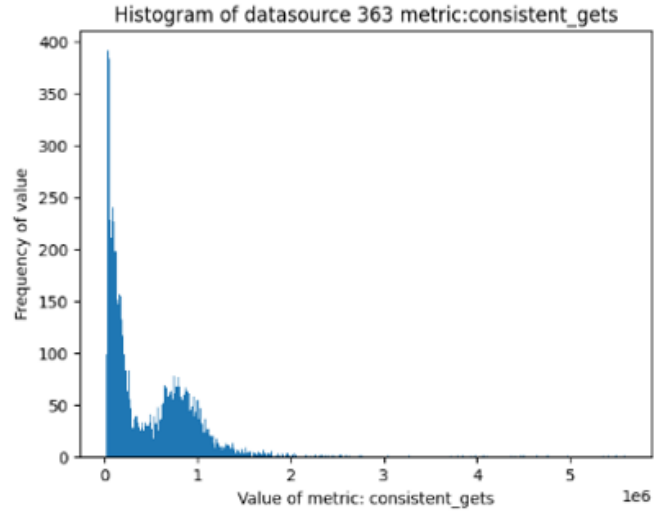


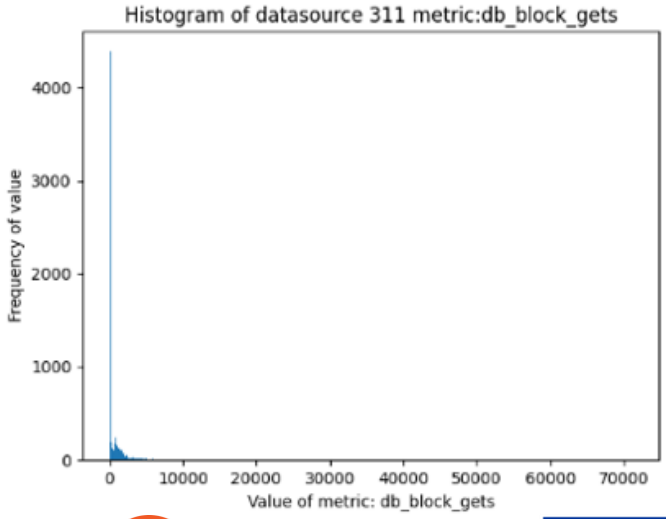
TED4LAT
Funded by
the European Union

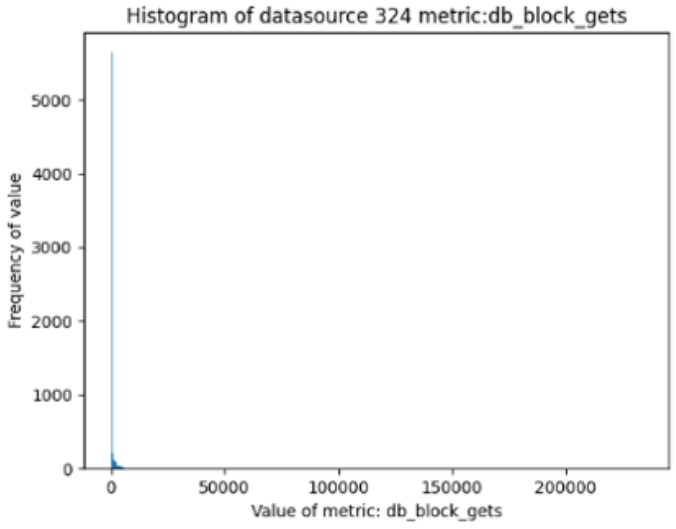
Oracle performance metric logical reads

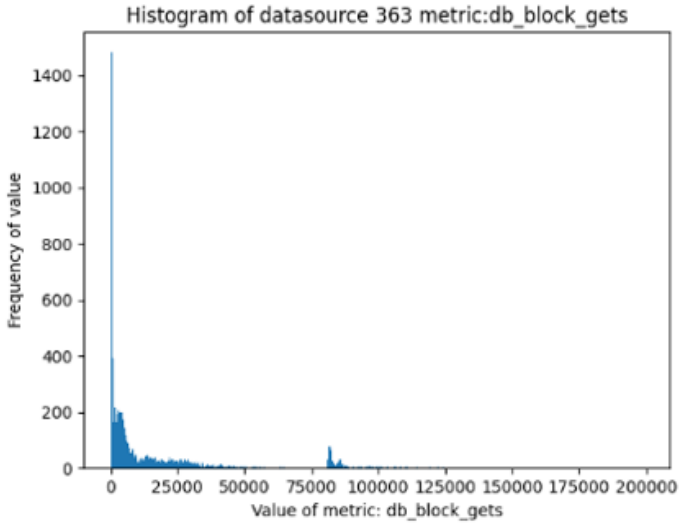








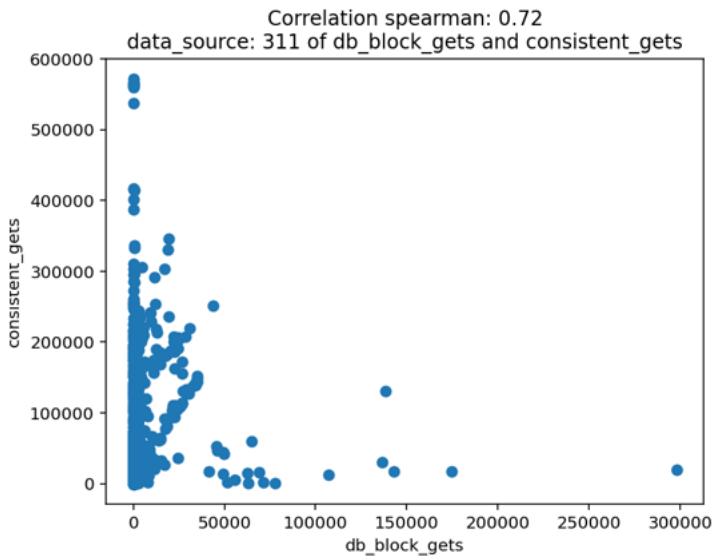


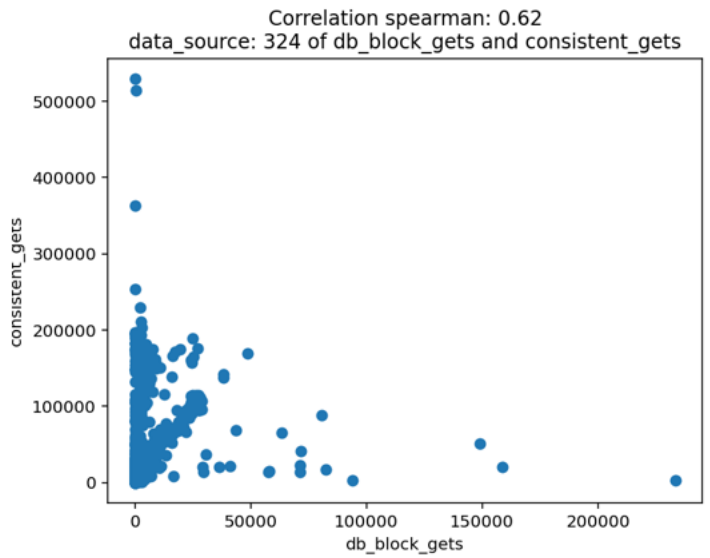


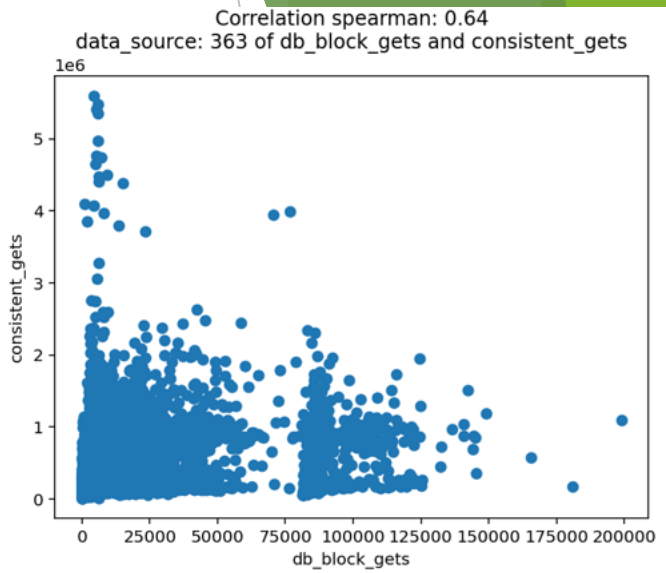
TED4LAT

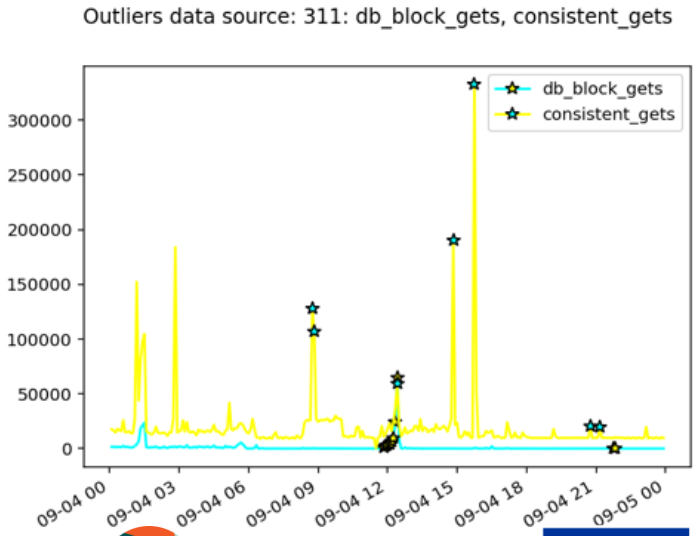
Funded by
the European Union

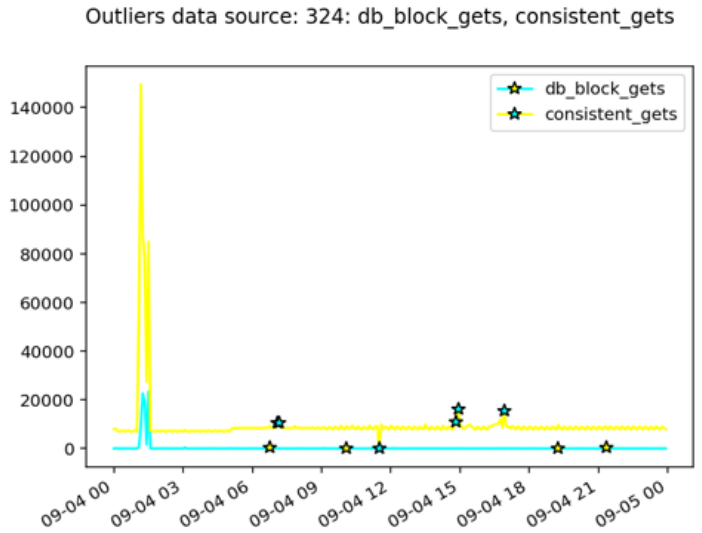
Correlation of metrics of logical reads

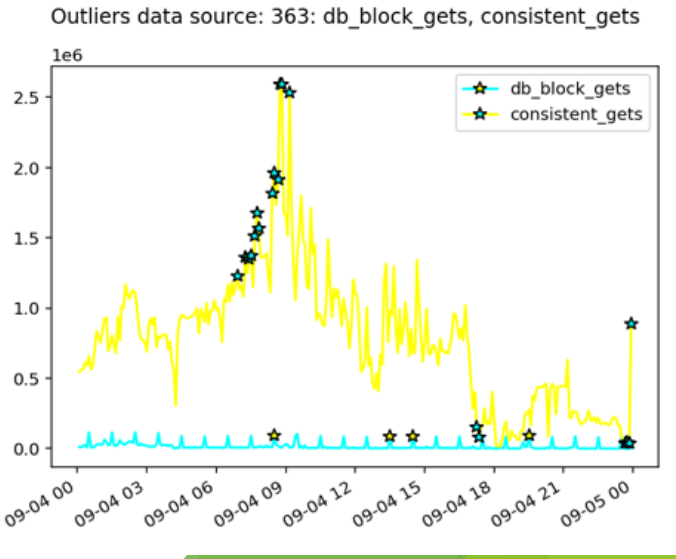










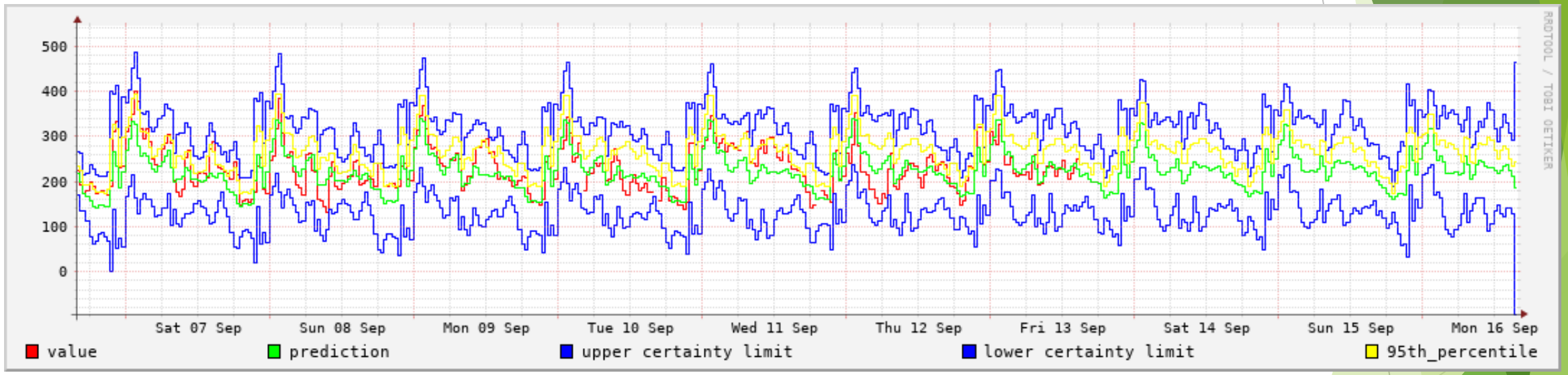


TED4LAT
Funded by
the European Union

Measurements of outlierness level for Oracle database logical reads over data centre

TED4LAT
Funded by the European Union

Prediction of outlierness level – normal vs not normal state of data centre



TED4LAT
Funded by the European Union

References

1. https://en.wikipedia.org/wiki/Anomaly_detection

2. https://azure.microsoft.com/en-us/services/cognitive-services/#overview

3. https://en.wikipedia.org/wiki/Time_series

4. https://azure.microsoft.com/en-us/services/cognitive-services/anomaly-detector/#features

5. https://techcommunity.microsoft.com/t5/ai-customer-engineering-team/introducingazure-anomaly-detector-api/ba-p/490162

6. https://techcommunity.microsoft.com/t5/ai-cognitive-services-blog/introducingmultivariate-anomaly-detection/ba-p/2260679

7. Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, GRAPH ATTENTION NETWORKS, Sixth International Conference on Learning Representations (ICLR), 2018.

TED4LAT

Funded by the European Union