TED4LAT

PhD candidate Roberts Volkovičs
Vidzeme University of Applied Sciences

«Anomaly detection: methods and use cases»

(Professor, Dr. sc. ing. Artis Teilāns)

PhD progress report 2025

TED4LAT, No. 101079206

Publications related to PhD work [1]

TED4LAT

Environment. Technology. Resources. Rezekne, Latvia

Proceedings of the 14th International Scientific and Practical Conference. Volume 2, 105-112

«Anomaly Detection - Review of Methods, Tools and Algorithms»

Roberts Volkovičs

Modelling of sociotechnical systems

Vidzeme University of Applied Sciences

Publications related to PhD work [2]

TED4LAT

Environment. Technology. Resources. Rezekne, Latvia

Proceedings of the 15th International Scientific and Practical Conference. Volume II, 317-323

«Environmental Data and Digital Twins for Road Traffic Safety in Rural Areas»

Roberts Volkovičs, Viesturs Bambāns, Dr. sc. ing. Ginta Majore

Modelling of sociotechnical systems

Vidzeme University of Applied Sciences

Publications related to PhD work [3]

TED4LAT

Environment. Technology. Resources. Rezekne, Latvia

Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 375-382

«Data Centre Monitoring Model Utilizing Artificial Intelligence, Machine Learning and Anomaly Detection Algorithms»

Roberts Volkovičs

Modelling of sociotechnical systems

Vidzeme University of Applied Sciences

Publications related to PhD work [4]

TED4LAT

Environment. Technology. Resources. Rezekne, Latvia

Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 383-388

«Pollution Detection Model Utilizing Anomaly Detection Techniques Performing Analysis of Satellite Data»

Roberts Volkovičs

Modelling of sociotechnical systems

Vidzeme University of Applied Sciences

Publications related to PhD work [5]

TED4LAT

Environment. Technology. Resources. Rezekne, Latvia

Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 389-394

«Model for Processing Historical Timeseries and Establishing Ruleset for Anomaly Detection in Current Sensor Data and General-Purpose Forecasting for Smart Farming in Latvia»

Roberts Volkovičs

Modelling of sociotechnical systems

Vidzeme University of Applied Sciences

Classification of the research and guidelines

- **The category of science – system analysis, modelling and design**
- **The problem** – Is it possible to utilize nowadays available anomaly detection tools and techniques to improve existing results in modern work environment?
- **Scientific novelty/innovation – prove of concept models (digital twins) utilising anomaly detection techniques in the area of big and dense data analysis allowing to implement decision support systems.**
- **Practical significance** – to provide the way or angle of information analysis and decision making which is not available in the existing tools available in the market.
- **Object of the study** – anomaly detection, tools and techniques, including AI algorithms and machine learning.
- Subject of the study – the way the anomaly detection and the related tools and techniques could be used to improve the work results in modern work environment.
- Materials and methods – open data (CSDD, NASA, ESA), Python, Python anomaly detection libraries, Oracle database, SQL, C#, Cacti, MATLAB.
- Discussion …

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. [8]

Concepts – anomaly detection

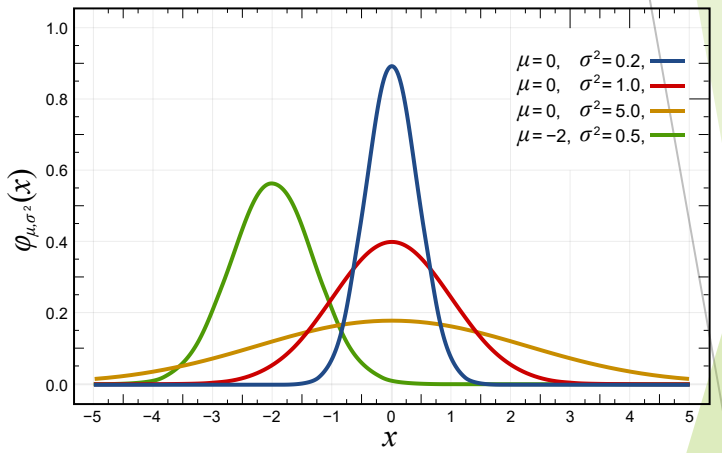
Anomaly detection finds application in many domains including **cyber security, medicine, machine vision, statistics, neuroscience, law enforcement** and **financial fraud** to name only a few.

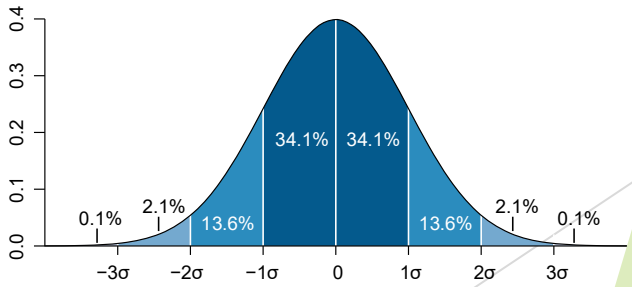
Anomalies were initially searched for **clear rejection or omission** from the data to aid **statistical analysis**, for example to compute the mean or standard deviation. They were also **removed to better predictions** from models such as linear regression, and more recently their removal aids the **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. [8]

PhD work main focus areas – theoretical basis of anomaly detection

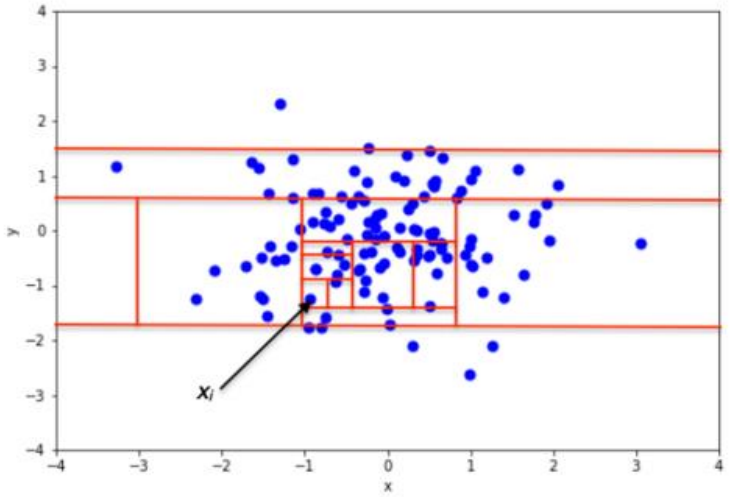
- Theoretical basis of anomaly detection
  - Percent calculations;
  - Proportions;
  - Ratios;
  - Probability theory;
  - Statistics [6].





PhD work main focus areas – anomaly detection algorithms and machine learning

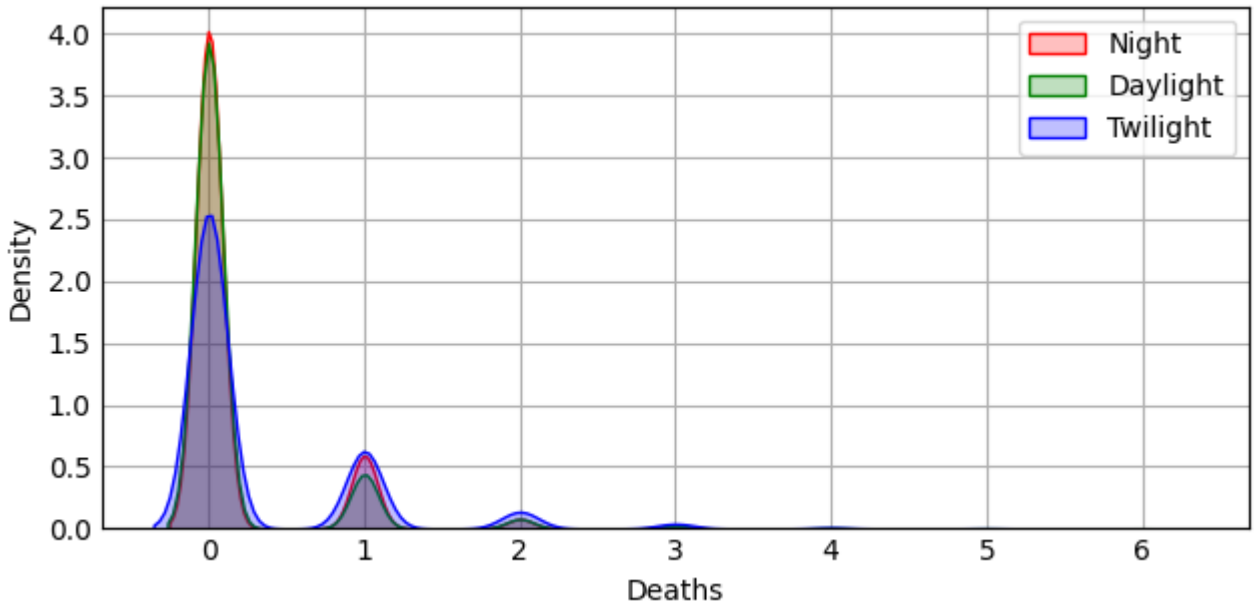
- Isolation forests [7];
- K nearest neighbours;
- Supervised learning;
- Unsupervised learning.




```
Working with model: IForest
[0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0
 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0
 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
seconds:  1.4327132149999215

Working with model: LOCI
[1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 1 0
 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 1 0 0 0
 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
seconds:  20.55228206700008

Working with model: LOF
[1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0
 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0
 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
seconds:  0.0040239280001515

Working with model: ECOD
[0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
seconds:  0.11084409600016443
```

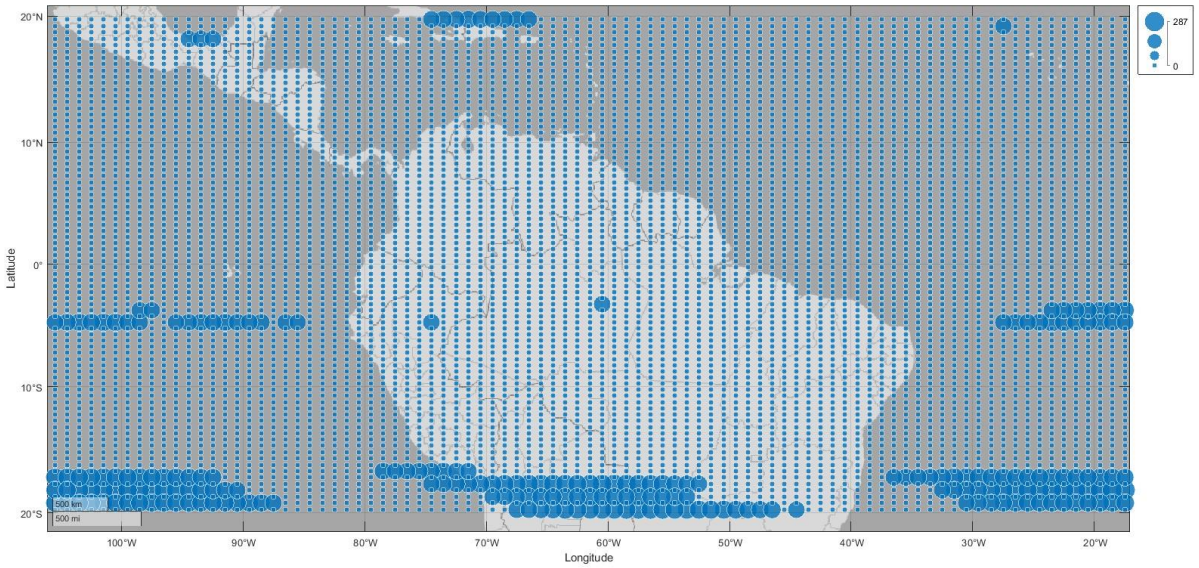
Using Anomaly Detection on open data from Latvian road meteo stations and traffic accidents

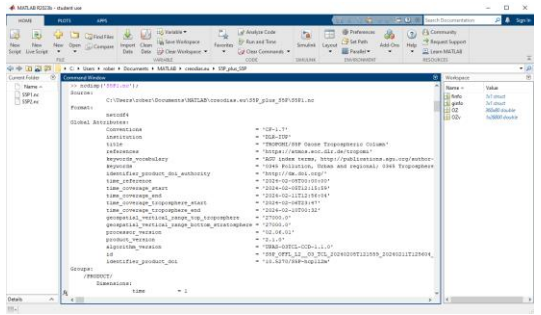
- 2 from more than 20 meteo stations do work correctly others need technical repair of sensors;
- wind of 300 km/h detected;
- finding root cause critical factors of traffic accidents [2].

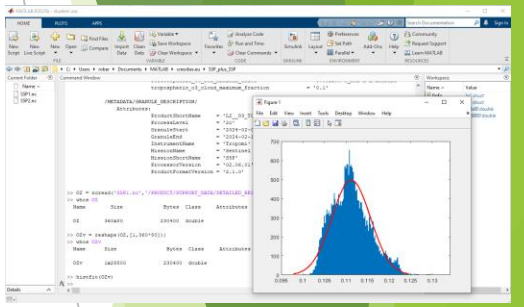


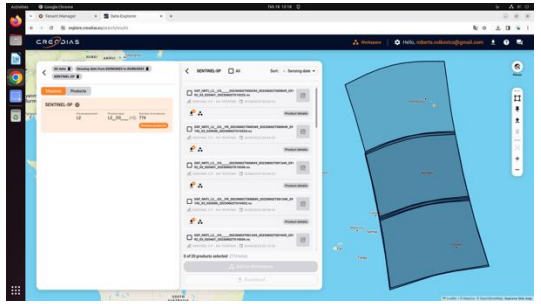
PhD work main focus areas – development of anomaly detection models to solve problems of particular type

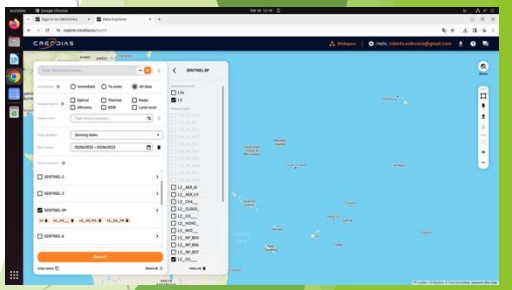
- Pollution detection [4] – analysis of 5 years of data;









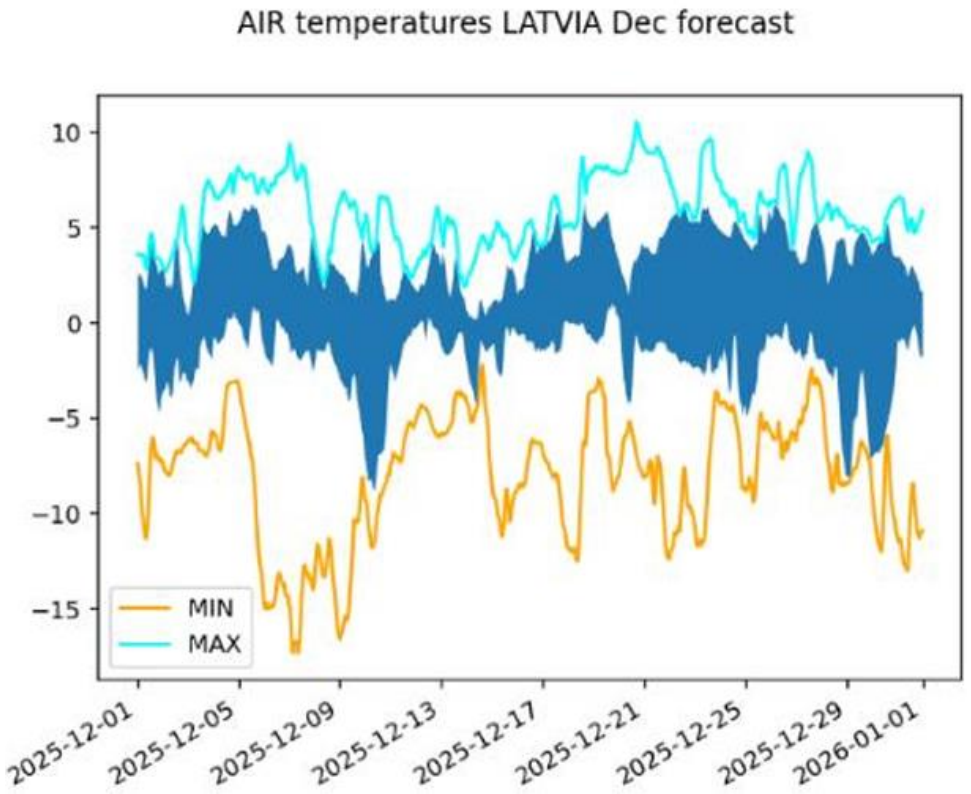


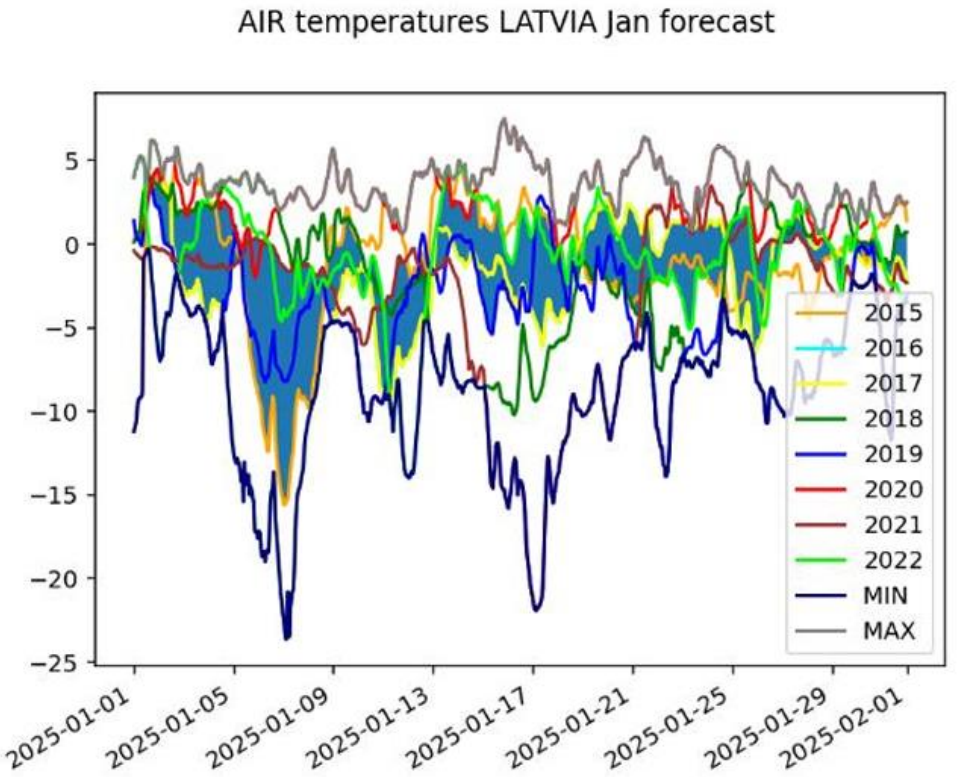
PhD work main focus areas – development of anomaly detection models to solve problems of particular type

- Data centre monitoring model [3] - analysis of 6 months of data;

PhD work main focus areas – development of anomaly detection models to solve problems of general type

- Establishing ruleset for anomaly detection in seasonal and periodic timeseries [5] - analysis of 5 years of data;





Conclusions

- **The objectives of the research were achieved**;
- It is possible to utilize anomaly detection techniques in modern work environment:
  - Big data analysis;
  - Dense data analysis;
  - Big and dense data analysis;
  - Analysis of time series;
  - Analysis of multi dimensional timeseries and spatial timeseries;
  - For building sophisticated models to perform computations in different areas.
- Digital Twin for road accident analysis;
- Data centre monitoring model;
- Pollution detection model;
- **Method for anomaly detection based on historical timeseries in seasonal data.**

TED4LAT

Thank you!

Q and A

References

1. https://doi.org/10.17770/etr2023vol2.7283
2. https://doi.org/10.17770/etr2024vol2.8012
3. https://doi.org/10.17770/etr2025vol2.8565
4. https://doi.org/10.17770/etr2025vol2.8566
5. https://doi.org/10.17770/etr2025vol2.8569
6. https://en.wikipedia.org/wiki/Normal_distribution
7. https://en.wikipedia.org/wiki/Isolation_forest
8. https://en.wikipedia.org/wiki/Anomaly_detection