# Artificial Intelligence in Industry

Last update: 17 September 2024

Academic Year 2024 – 2025

Alma Mater Studiorum · University of Bologna

# Contents

**1 Anomaly detection: Taxi calls** **1**
1.1 Data . . . . . 1
1.2 Approaches . . . . . 1
1.2.1 Gaussian assumption . . . . . 1
1.2.2 Characterize data distribution . . . . . 1

# 1 Anomaly detection: Taxi calls

**Anomaly** Event that deviates from the usual pattern. Anomaly

**Time series** Data with an ordering (e.g., chronological). Time series

## 1.1 Data

The dataset is a time series and it is a `DataFrame` with the following fields:

`timestamp` with a 30 minutes granularity.

`value` number of calls.

The label is a `Series` containing the timestamps of the anomalies.
An additional `DataFrame` contains information about the time window in which the anomalies happen:

`begin` acceptable moment from which an anomaly can be detected.

`end` acceptable moment from which there are no anomalies anymore.

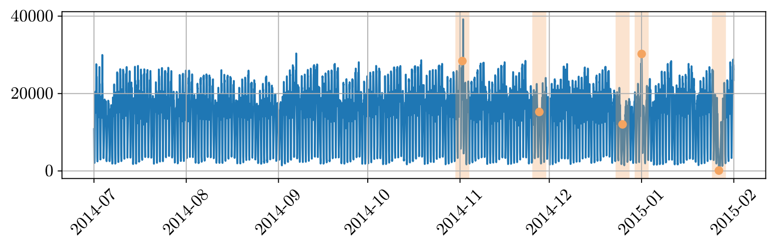


Figure 1.1: Plot of the time series, anomalies, and windows

## 1.2 Approaches

### 1.2.1 Gaussian assumption

Assuming that the data follows a Gaussian distribution, mean and variance can be used to determine anomalies through a threshold. $z$-score can also be used.

### 1.2.2 Characterize data distribution

Classify a data point as an anomaly if it is too unlikely.

**Formalization** Given a random variable $X$ with values $x$ to represent the number of taxi calls, we want to find its probability density function (PDF) $f(x)$.

An anomaly is determined whether:

$$f(x) \leq \varepsilon$$

where $\varepsilon$ is a threshold.

**Remark.** The PDF can be reasonably used even though the dataset is discrete if its data points are sufficiently fine-grained.