



Developing your inner Spidey Sense

Anomaly Detection for apps

RON DAGDAG

Spidey Sense?

- tingling Sensation on the back of Peter Parker's skull
- ability to sense and react to danger before it happens.

Uses

- Increases his ability to detect evil (and even clones)
- Helps him navigate if he is impaired (disoriented or unable to see/hear)
- Aids him in discovering secret passageways and find hidden/lost objects
- Helps fire his Web Shooters and swing instinctively



This Photo by Unknown Author is licensed under CC BY-SA

[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

Real Spider Sense

“hyper-awareness”

long, thin hairs, *trichobothria*

- low-level vibrations through their web
- can detect the vibrations of faint sounds
- small insects moving up to 3 meters away



This Photo by Unknown Author is licensed under [CC BY-SA](#)



Any new web developers here?

Spidey Sense?

Gut feeling

Vibe

Feeling

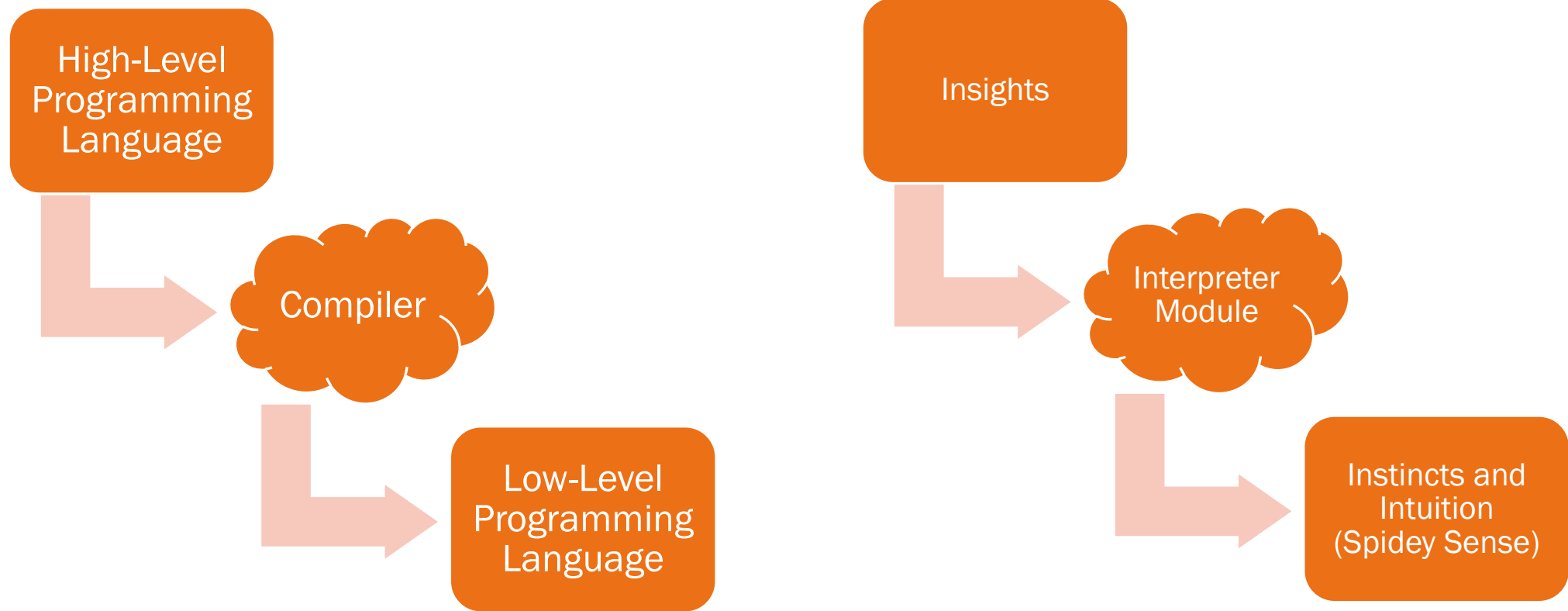
Intuition

Discover Blind Spots

Learning from the past



IDE



Agenda

What is Anomaly Detection?

Time Series Anomaly
Detection

Demo

Takeaways

Anomaly Detection

Identifying unexpected items or events in data sets, which differ from the norm

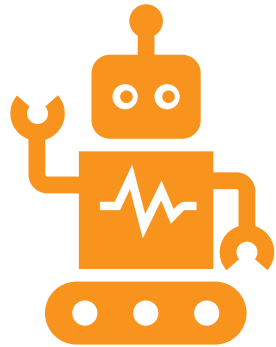
An Outlier

Assumptions:

- Anomalies only occur very rarely in the data.
- Their features differ from the normal instances significantly.



Causes of Outliers



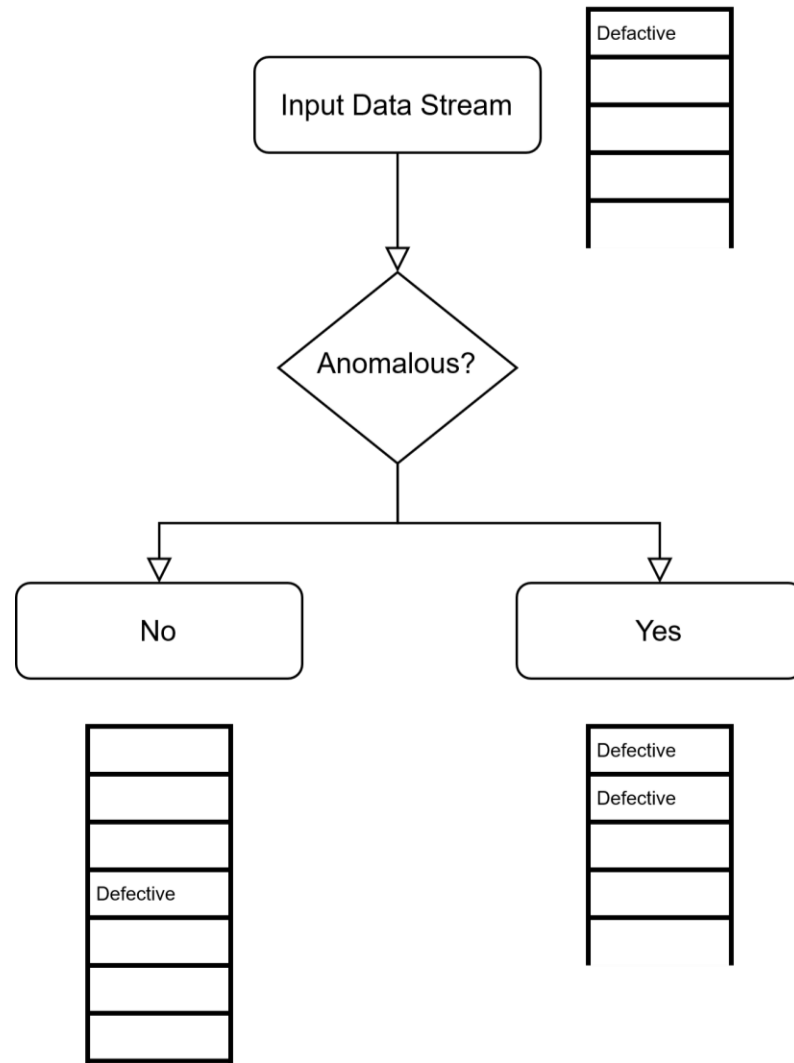
Artificial (Error) / Non-natural



Natural

Causes of Outliers

- Data Entry Errors: 100,000 vs 1,000,000 - fat fingered
- Measurement Error: common
- Experimental Error: start late in sprint
- Intentional Outlier: underreporting alcohol consumption
- Data Processing Error: extraction errors
- Sampling Error: reporting height for all athletes and included few basketball players
- Natural Outlier: When it's not artificial



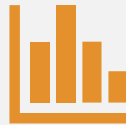
A large, round, golden-brown hay bale sits in a field of tall green grass. A long, thin, silver needle is stuck into the side of the bale, angled upwards and to the left. In the background, another smaller hay bale is visible under a blue sky with scattered white clouds.

Needle in a
haystack

Methods



Rule-based Systems



Statistical Techniques



Machine Learning

Rule-based Systems



Specific Rules



Assign Threshold and
limits



Experience of Industry
Experts to detect
“known anomalies”



Doesn't Adapt as
patterns change



Data Labeling



“Experience is the
teacher of all things.”

Julius Caesar

Statistical Techniques



flags the data points => deviate from common statistical properties (mean, median, mode, & quantiles)



a rolling average or a moving average



n-period simple moving average
"low pass filter." e.g. Kalman Filters



Histogram-based Outlier Detection (HBOS)



More Interpretable and sometimes more useful than ML methods

ANOMALY DETECTION

- Very small number of positive examples
- Large number of negative examples
- Many different “types” of anomalies. Hard to learn from positive examples
- Future anomalies may not be discovered yet.

SUPERVISED LEARNING

- Large number of positive and negative examples
- Enough positive examples for algorithm to learn.
- Future positive examples likely to be similar to training set

ANOMALY DETECTION

- Fraud Detection
- Manufacturing (engines/machineries)
- Monitoring Data Center
- Internet of Things

SUPERVISED LEARNING

- Email spam classification
- Weather prediction
- Cancer classification



Supervised
(e.g. Decision Tree, SVM, LSTM
Forecasting)



Unsupervised
(e.g. K-Means, Hierarchical
Clustering, DBSCAN)



Self-Supervised
(e.g. LSTM Autoencoder)

Machine Learning Methods

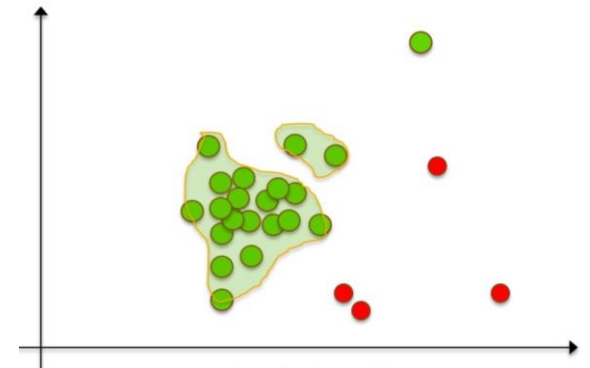
Machine Learning

Density-Based Anomaly Detection

- based on the k-nearest neighbors algorithm.
- *Assumption:* Normal data points occur around a dense neighborhood and abnormalities are far away.

Clustering-Based Anomaly Detection

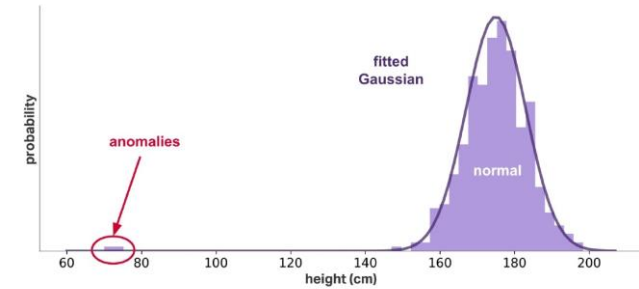
- *Assumption:* Data points that are similar tend to belong to similar groups or clusters, as determined by their distance from local centroids.
- K-means



Machine Learning

Gaussian Distribution

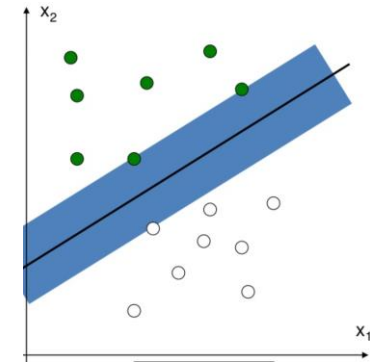
- Gaussian Distribution and given a new data-point,
- Compute the probability of the data-point
- If the probability is below a threshold => outlier or anomalous.



Machine Learning

Support Vector Machine-Based Anomaly Detection

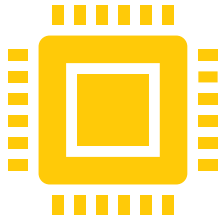
- *OneClassSVM*
- *>100 features, aggressive boundary*
- find a function that is positive for regions with high density of points, and negative for small densities



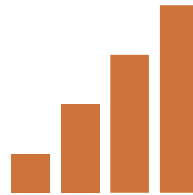
PCA-Based Anomaly Detection

- analyzing available features to determine what constitutes a "normal" class
- applying distance metrics
- Fast training

Internet of Things



Increasing Data Volume
(sensors are cheaper)



Increased Data Speed
(improved networking)

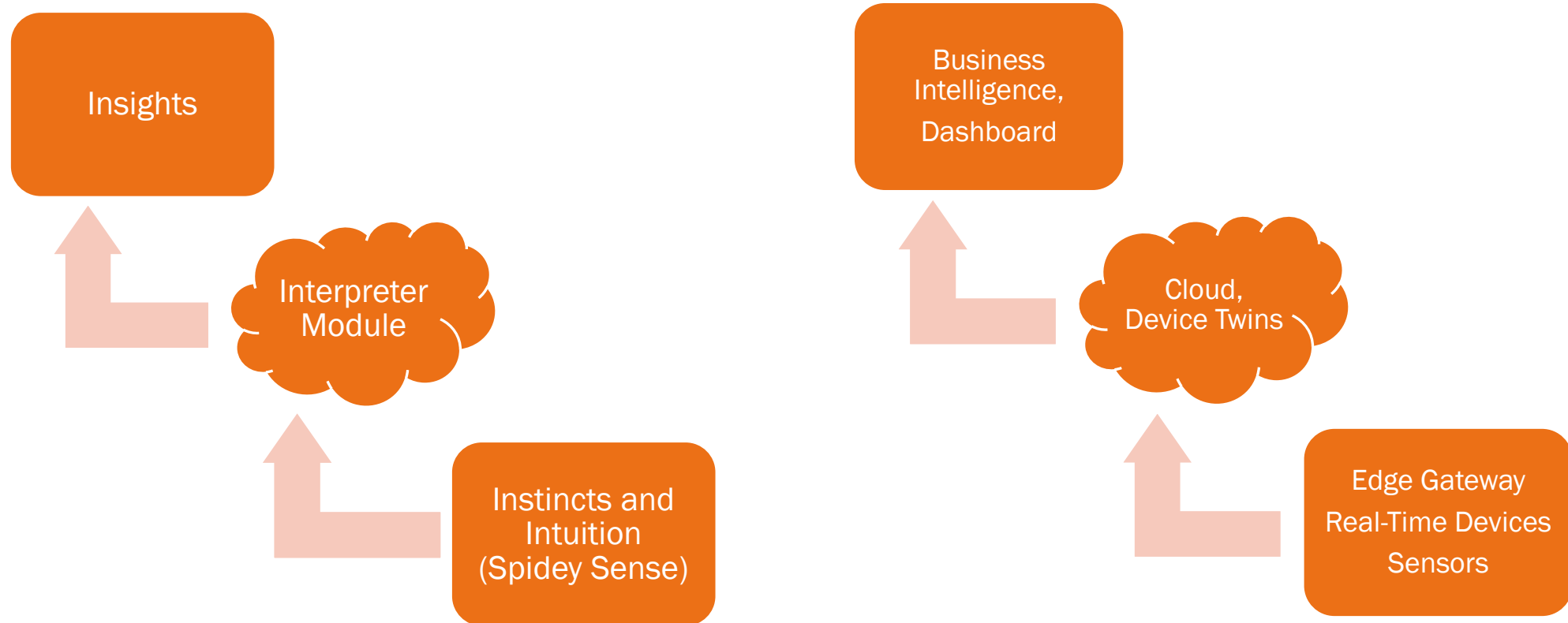


Risk environment that are
moving very fast but failures
are not tolerated.

Internet of Broken Things



Artificial Intelligence of Things



Time Series Anomaly Types



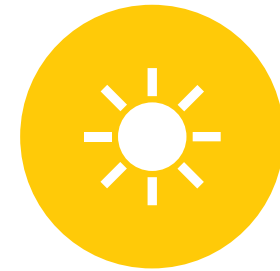
OUTLIER



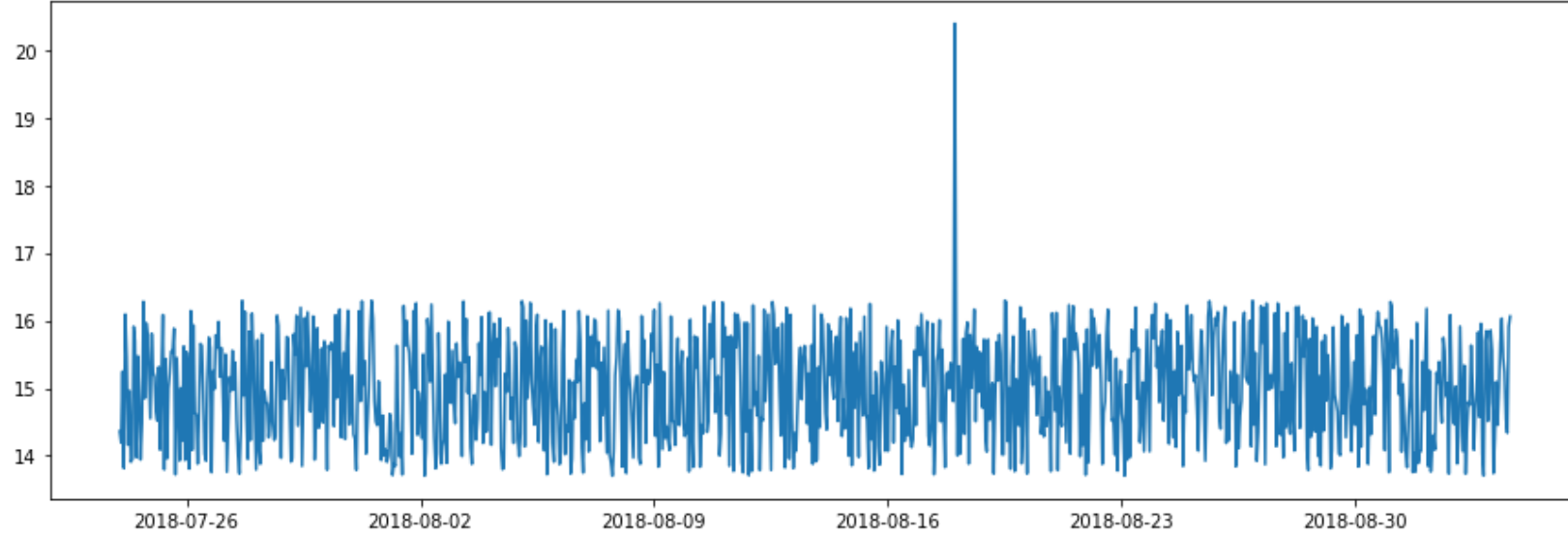
SPIKE AND
LEVEL SHIFT



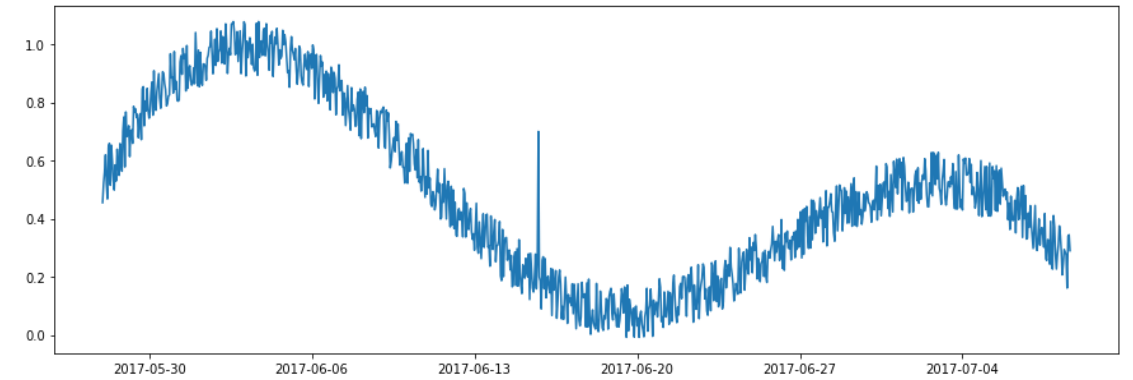
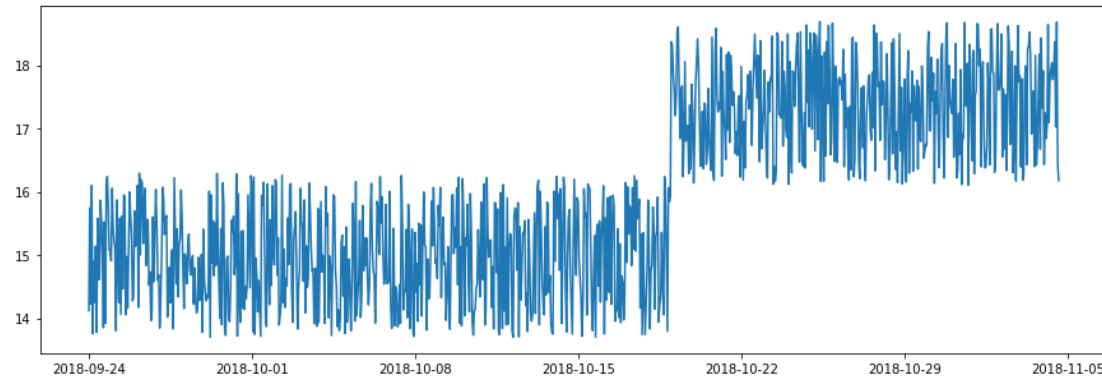
PATTERN
CHANGE



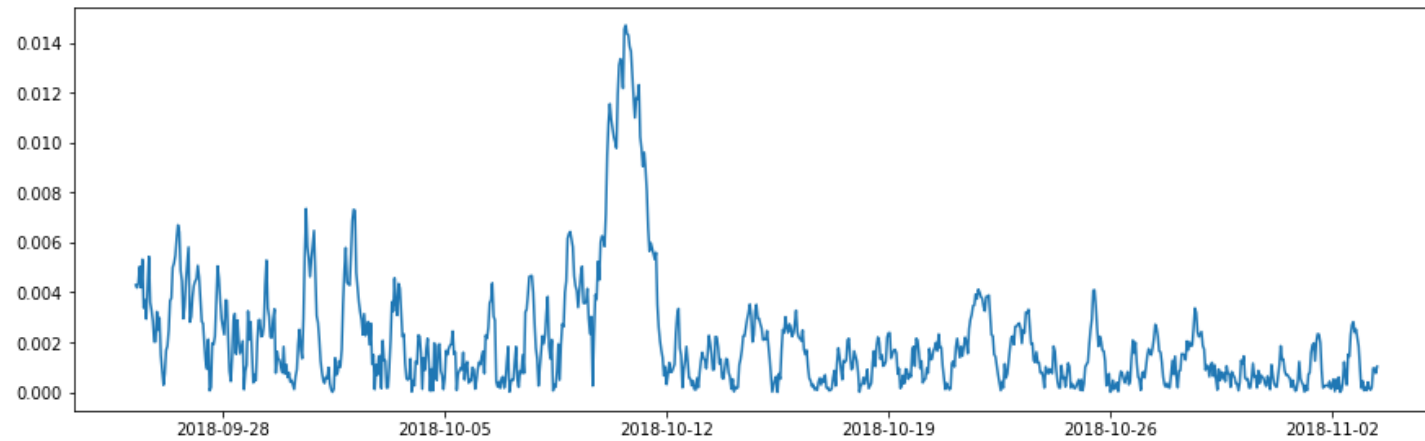
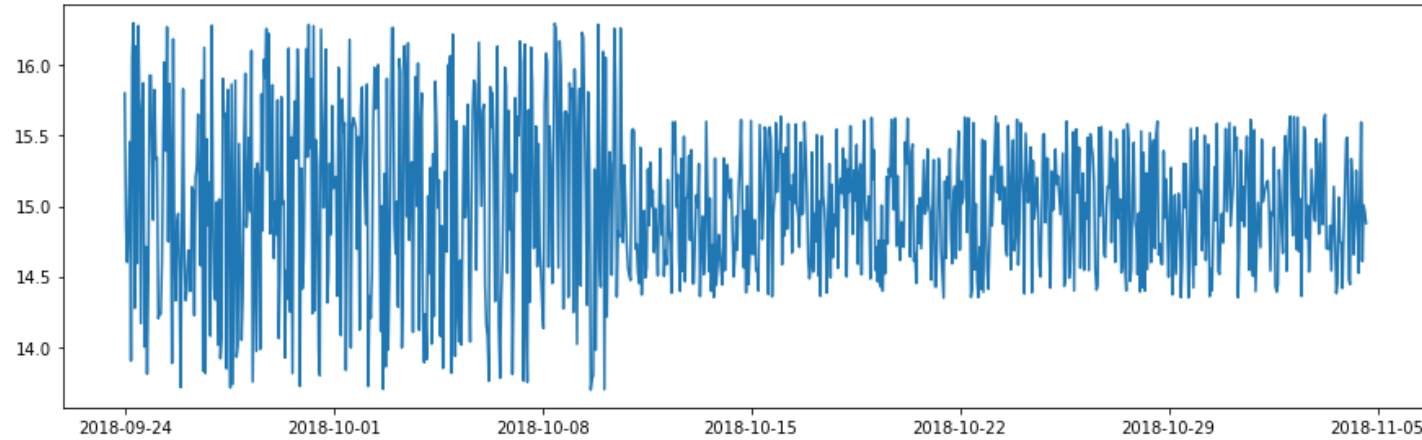
SEASONALITY



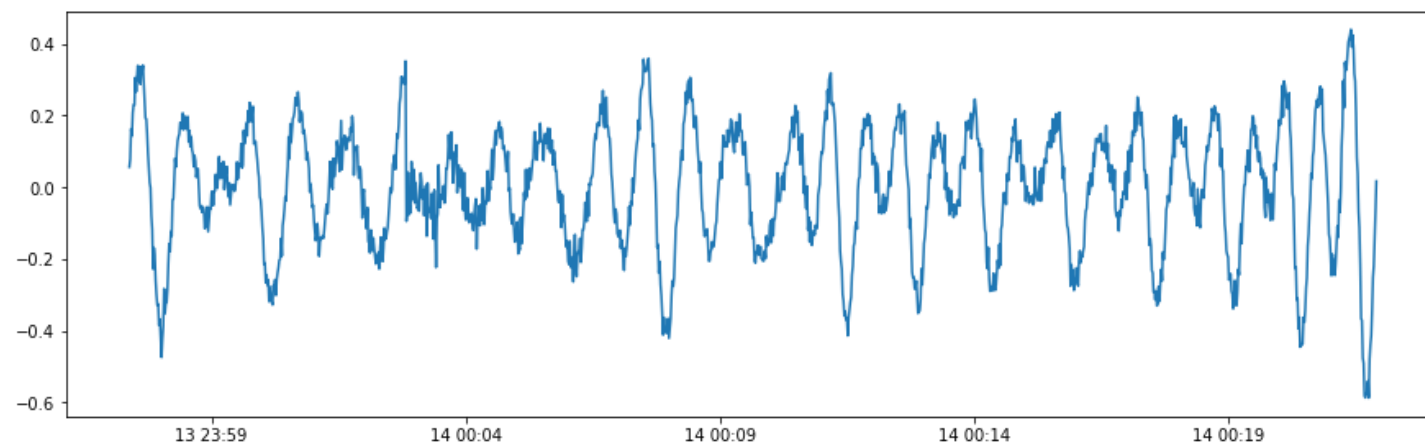
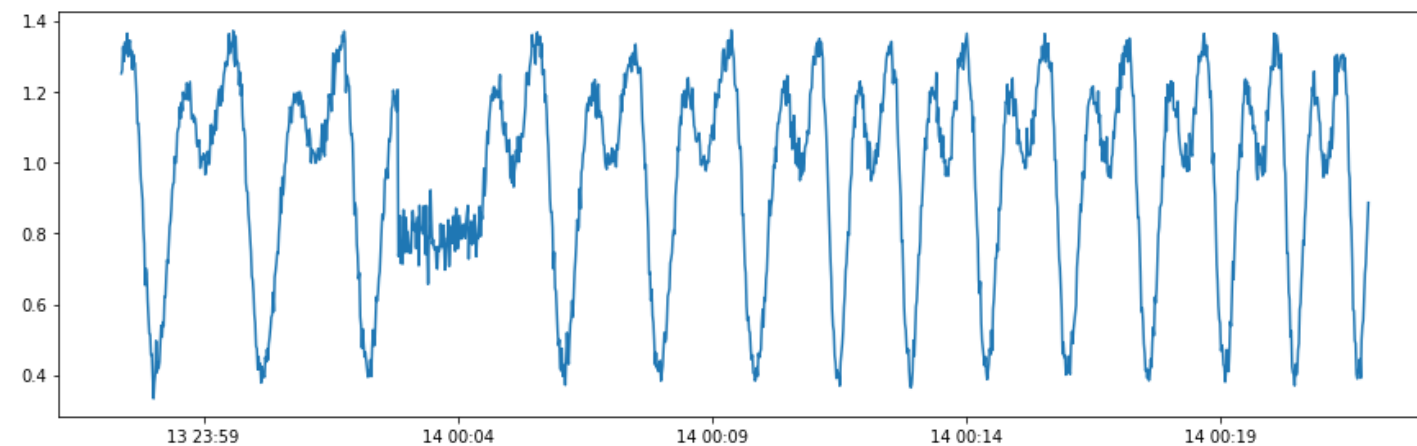
Outlier



Spike and Level Shift



Pattern Change



Seasonality

Production Issues?



IID datasets

Identically Distributed

- no overall trends – the distribution doesn't fluctuate
- all items in the sample are taken from the same probability distribution

Independent

- Items are all independent events.
- Not connected to each other in any way.



Time Series Anomaly Detection

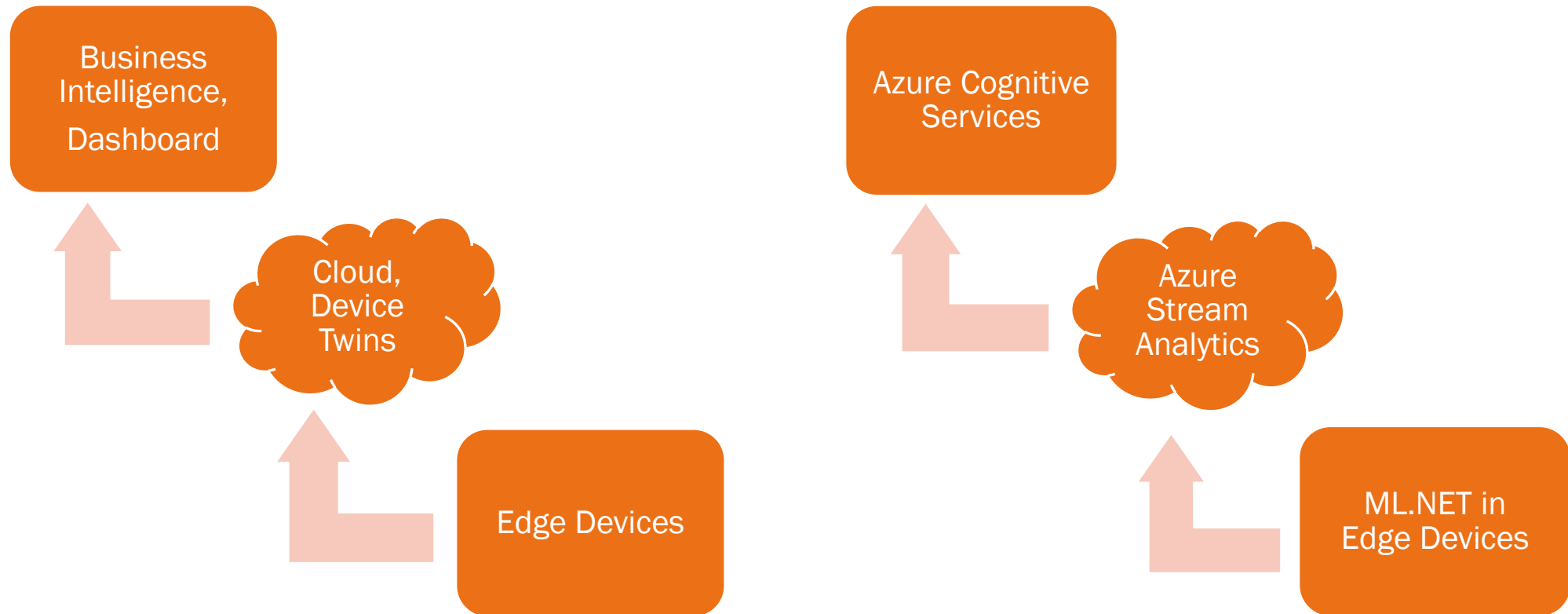
Spikes

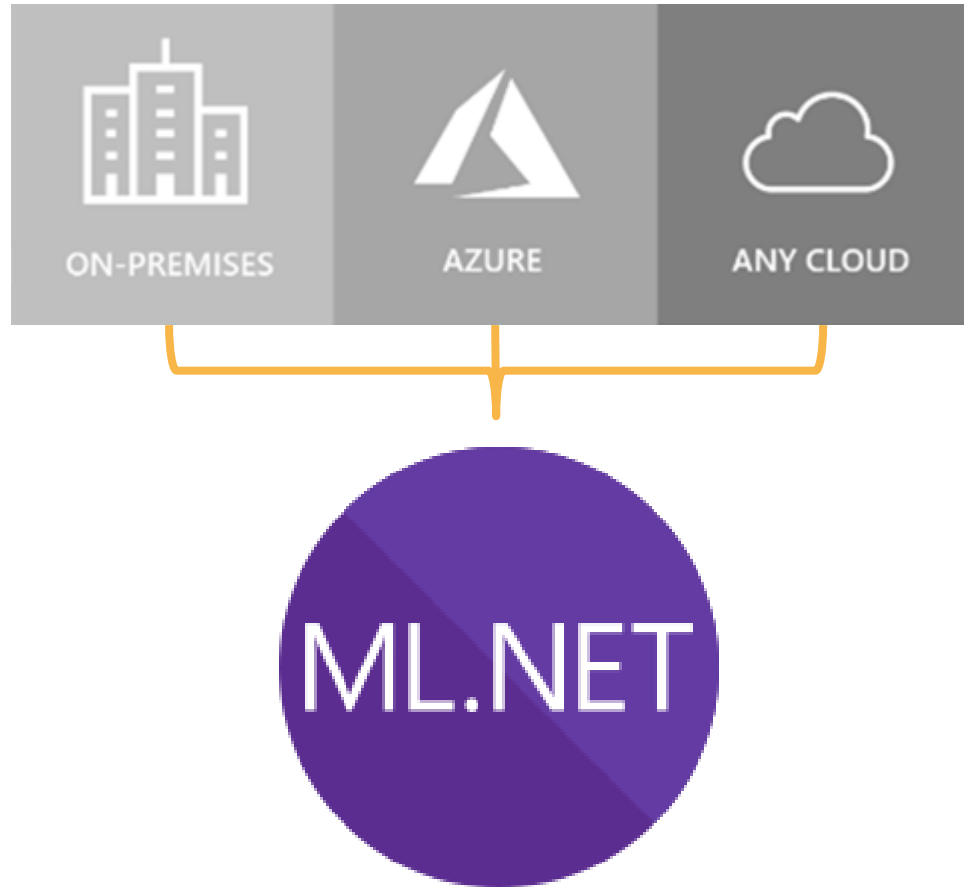
- temporary bursts of anomalous behavior in the system.

Change points

- indicate the beginning of persistent changes over time in the system.
- level changes and trends

Anomaly Detection in IoT





**An open source and cross-platform
machine learning framework for .NET**



Built for .NET Developers

Can use existing
C# and F# skills
to integrate ML
into .NET apps

Data science &
ML experience
not required

ML.NET Time Series Catalog



DetectAnomalyBySrCnn

- detects anomalies with the Spectral Residual Convolutional Neural Network (SRCNN) algorithm



DetectEntireAnomalyBySrCnn

- detects timeseries anomalies for entire input using SRCNN algorithm.



DetectChangePointBySsa

- detects anomalies with the Singular Spectrum Analysis (SSA) algorithm in an independent identically distributed (i.i.d.) time series-based algorithm.

ML.NET Time Series Catalog



DetectIidSpike

– detects changes with an i.i.d. algorithm but predicts spikes instead of change points



DetectSpikeBySsa

– detects spikes in time series using Singular Spectrum Analysis (SSA).



ForecastBySsa

– Uses Singular Spectrum Analysis (SSA) model for singular variable (univariate) based time-series

ML.NET Time Series Catalog



DetectIidSpike

– detects changes with an i.i.d. algorithm but predicts spikes instead of change points



DetectSpikeBySsa

– detects spikes in time series using Singular Spectrum Analysis (SSA).



ForecastBySsa

– Uses Singular Spectrum Analysis (SSA) model for singular variable (univariate) based time-series

ML.NET 1.5.1 – Time Series

Detecting seasonality in time series

Removing seasonality from time series prior to anomaly detection

Threshold for root cause analysis

RCA for anomaly detection can now return multiple dimensions



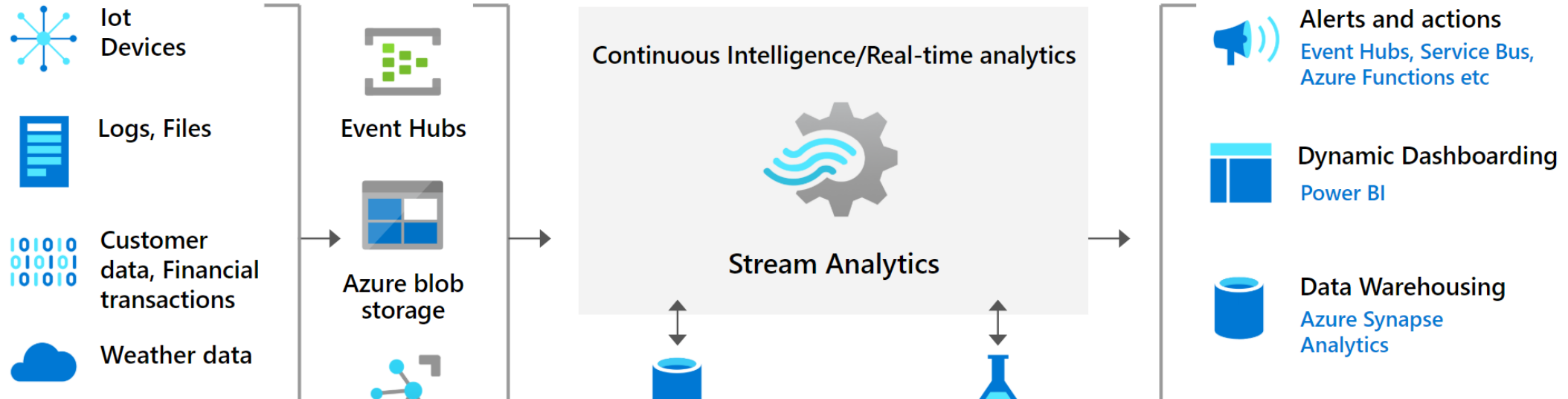


ML.NET DEMO

Ingest

Analyze

Deliver



Azure Stream Analytics

Spike and Dips

WITH AnomalyDetectionStep AS

(

SELECT

EVENTENQUEUEDUTCTIME AS time,

CAST(temperature AS float) AS temp,

AnomalyDetection_SpikeAndDip(CAST(temperature AS float), 95, 120, 'spikesanddips')

OVER(LIMIT DURATION(second, 120)) AS SpikeAndDipScores

FROM input

)

SELECT

time,

temp,

CAST(GetRecordPropertyValue(SpikeAndDipScores, 'Score') AS float) AS

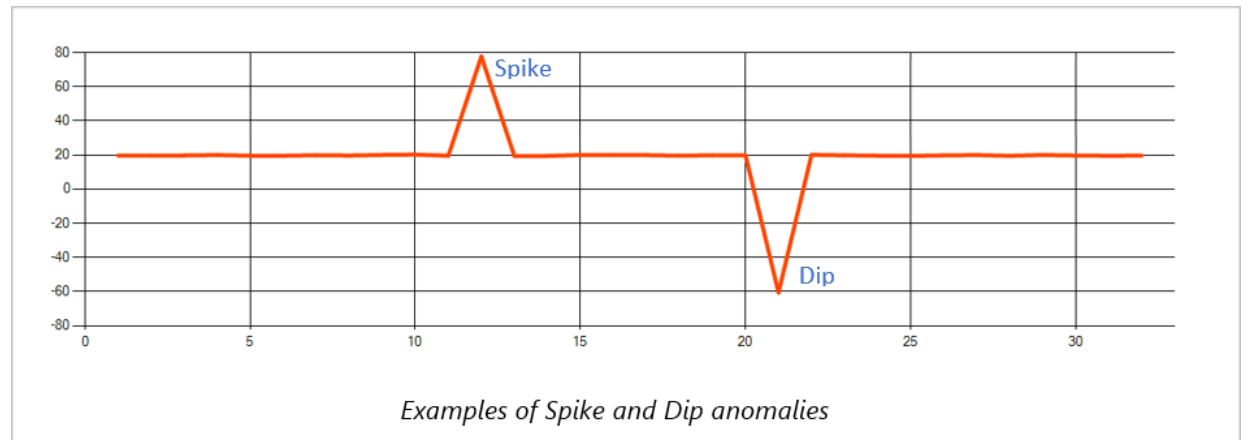
SpikeAndDipScore,

CAST(GetRecordPropertyValue(SpikeAndDipScores, 'IsAnomaly') AS bigint) AS

IsSpikeAndDipAnomaly

INTO output

FROM AnomalyDetectionStep



Change Point

WITH AnomalyDetectionStep AS

(

SELECT

EVENTENQUEUEDUTCTIME AS time,

CAST(temperature AS float) AS temp,

AnomalyDetection_ChangePoint(CAST(temperature AS float), 80, 1200)

OVER(LIMIT DURATION(minute, 20)) AS ChangePointScores

FROM input

)

SELECT

time,

temp,

CAST(GetRecordPropertyValue(ChangePointScores, 'Score') AS float) AS

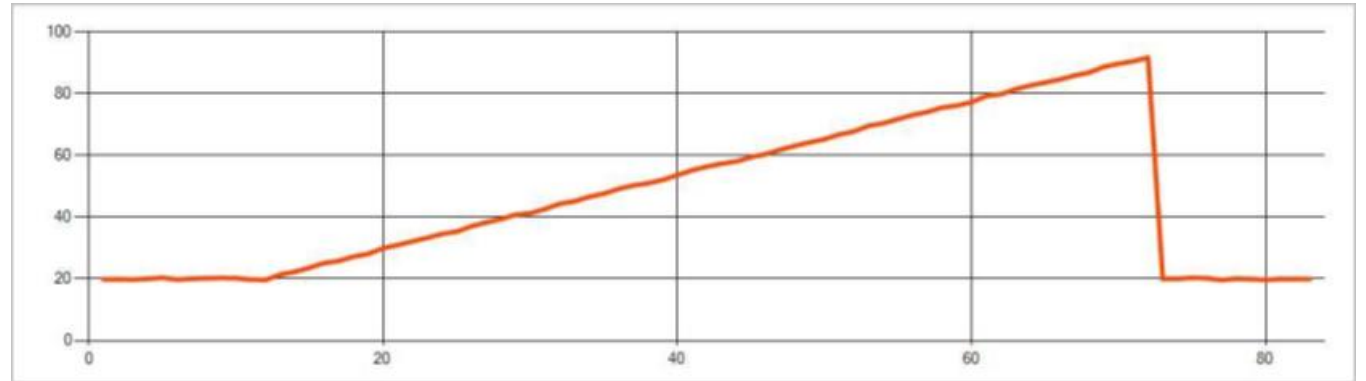
ChangePointScore,

CAST(GetRecordPropertyValue(ChangePointScores, 'IsAnomaly') AS bigint) AS

IsChangePointAnomaly

INTO output

FROM AnomalyDetectionStep





This Photo by Unknown Author is licensed under CC BY-NC

Azure Streaming Analytics DEMO

Azure Cognitive Services

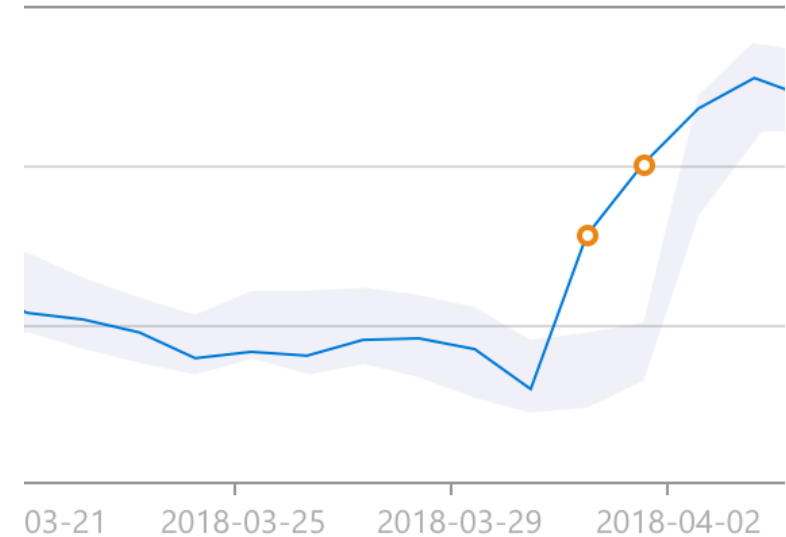
- AI for every developer—without requiring machine-learning expertise.
- Just an API call



Decision	Make smarter decisions faster
Language	Anomaly Detector <small>PREVIEW</small> Identify potential problems early on.
Speech	Content Moderator Detect potentially offensive or unwanted content.
Vision	Personalizer Create rich, personalized experiences for every user.
Web search	

Anomaly Detector

- Identify potential problems early on
- RESTful API
- monitor and detect abnormalities
- no machine learning expertise needed
- automatically identify and apply the best-fitting models
- Identify boundaries for anomaly detection
- expected values
- Eliminates the need for labeled training data
- Fine-tune sensitivity
- Used by 200 Microsoft product teams



Anomaly Detector Features



Detect anomalies as they occur in real-time.



Detect anomalies throughout your data set as a batch.

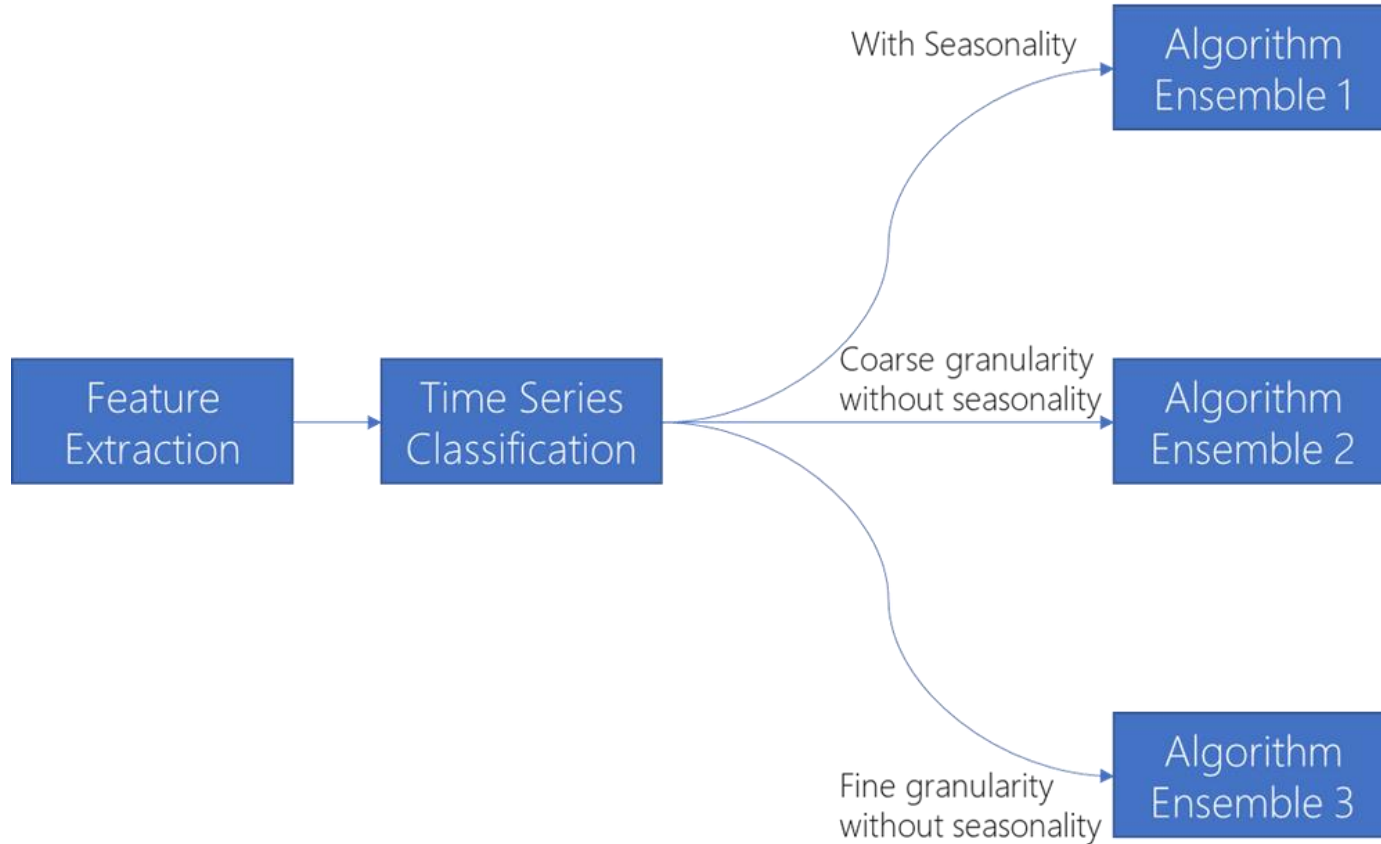


Get additional information about your data.



Adjust anomaly detection boundaries.

Gallery of Algorithms



Fourier Transformation

Extreme Studentized Deviate
(ESD)

[STL Decomposition](#)

Dynamic Threshold

Z-score detector

[SR-CNN](#)

Anomaly Detector demo



<https://github.com/microsoft/Cognitive-Samples-IntelligentKiosk>

Where can you use this?

C#, Javascript, Python

<https://docs.microsoft.com/en-us/azure/cognitive-services/anomaly-detector/quickstarts/client-libraries?pivots=programming-language-csharp&tabs=linux>

Docker Containers

<https://docs.microsoft.com/en-us/azure/cognitive-services/anomaly-detector/anomaly-detector-container-howto>

Power BI

<https://docs.microsoft.com/en-us/azure/cognitive-services/anomaly-detector/tutorials/batch-anomaly-detection-powerbi>

Azure Databricks for streaming data

<https://docs.microsoft.com/en-us/azure/cognitive-services/anomaly-detector/tutorials/anomaly-detection-streaming-databricks>

The best superpower you can give to your IoT project is “spidey-sense”.





<https://bit.ly/spideysense-anomaly>

About Me

Ron Dagdag



Ron Lyle Dagdag

Immersive Experience Developer

Cell: 682-560-3988

ron@dagdag.net



Experience AR

www.dagdag.net

[@rondagdag](https://twitter.com/rondagdag)

<http://ron.dagdag.net>

Lead Software Engineer / AI Edge Specialist

4th year Microsoft MVP awardee

Personal Projects
www.dagdag.net

Email: ron@dagdag.net
Twitter [@rondagdag](https://twitter.com/rondagdag)

Connect me via Linked In
www.linkedin.com/in/rondagdag/

Feedback appreciated, help improve my presentation skills