

## Feature Engineering Techniques for Binary IoT Sensors

Nidal AlBeiruti 29 July 2021

# Internet of Things (IoT)



### IoT solutions are complex & multidimensional





### What customers are doing using AWS IoT



Improve the performance and productivity of industrial processes



Grow healthier crops with greater efficiencies



Remotely monitor patient health & wellness applications



Manage energy resources more efficiently



Track inventory levels and manage warehouse operations



Transform transportation with connected and autonomous vehicles



Build smarter products & user experiences in homes, buildings, and cities



Enhance safety in the home, the office, and the factory floor



# Microcontrollers and microprocessors are powering billions of IoT devices



#### **AWS Partner Device Catalog**



Development Kit Sensor Gateway / Router Asset Tracker Industrial PC Camera Cellular Modem

• • •

#### devices.amazonaws.com



# **Binary Sensors**



#### **Binary sensors**

• Report the states and conditions of monitored entities.



- In each event, they report only one of two mutually exclusive values.
- Examples:
  - 'True' or 'False'
  - 'On' or 'Off'
  - 'Open' or 'Closed'
  - '0' or '1'







### **Binary sensors - Examples**



Passive Infrared sensors (PIR sensors) Motion detectors



Pressure sensors









# Machine Learning: Feature Engineering



### Data preparation process for machine learning



Collect data



Prepare data



Transform data



Pipeline data prep



#### What is a feature?

• Feature: An attribute useful for your modelling task



• All coming from 'raw data': variable, attribute or field





#### **Examples of features**

Images: colours, textures, contours, ...

Signals: frequency, phase, samples, spectrum, ...

Time series: ticks, trends, self-similarities, ...

Biomed: DNA sequence, genes, ...

Text: words, PoS tags, grammatical dependencies, ...









### What is Feature Engineering?

The process of representing a problem domain to make it amenable for learning techniques.

#### Feature Engineering:

- Helps improve results
- Bring out performance gains



Source: The Art of Feature Engineering

### Sub-problems of Feature Engineering





## Suggested Feature Engineering Techniques



#### **Feature Engineering techniques**

- Imputation
- Handling outliers
- Binning
- Log transform
- One-hot Encoding
  - </>

- Grouping operations
- Feature split
- Scaling
- Extracting date



#### How do we make sense of data from sensors?



Transformation: For example, we need to factor in the specifics of the frequencies rather than how numerous they are.



- Bag of features
- Time-window
- Generalisation
  - Dominant value / Median
  - Presence of at least one value









Localisation in regions:

- Presence or absence in a region.
- Localisation functionality.
- Areas are partitioned into sub-regions.





New features that can be extracted:

- Location/area
- Total elapsed time
- Time-window aggregation



• Merging other sensors events to add context



Specific examples for PIR sensors:

• The activity level.



- Elapsed time between sensor events:
  - Two consecutive same value events from the same sensor (inactivity time).
  - Two consecutive same value events from different sensors.
  - The frequency of movement events (activity time vs no-activity time).
- No sensors events at all from all sensors.







## Amazon SageMaker Data Wrangler

#### Start by – Preparing data

#### Amazon SageMaker

#### PREPARE —

SageMaker Ground Truth Label training data for machine learning

SageMaker Data Wrangler NEW Aggregate and prepare data for machine learning

SageMaker Processing Built-in Python, BYO R/Spark

SageMaker Feature Store NEW Store, update, retrieve, and share features

SageMaker Clarify NEW Detect bias and understand model predictions

#### BUILD -

SageMaker Studio Notebooks Jupyter notebooks with elastic compute and sharing

Built-in and Bring your-own Algorithms Dozens of optimized algorithms or bring your own

**Local Mode** Test and prototype on your local machine

SageMaker Autopilot Automatically create machine learning models with full visibility

SageMaker JumpStart NEW Pre-built solutions for common use cases

#### TRAIN & TUNE —

Managed Training Distributed infrastructure management

SageMaker Experiments Capture, organize, and compare every step

Automatic Model Tuning Hyperparameter optimization

**Distributed Training NEW** Training for large datasets and models

SageMaker Debugger NEW Debug and profile training runs

Managed Spot Training Reduce training cost by 90%

#### DEPLOY & MANAGE \_\_\_\_\_

Managed Deployment Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration Simplify Kubernetes-based machine learning

**Multi-Model Endpoints** Reduce cost by hosting multiple models per instance

SageMaker Model Monitor Maintain accuracy of deployed models

SageMaker Edge Manager NEW Manage and monitor models on edge devices

SageMaker Pipelines NEW Workflow orchestration and automation

SageMaker Studio



# Amazon SageMaker Data Wrangler is the fastest way to prepare data





Data selection

#### Data transforms



See data, spot inconsistencies, diagnose, and fix



Export your data



### How Amazon SageMaker Data Wrangler works







### Thank you!