Realtime Sensor Anomaly Detection with Scikit-Learn and the Azure Stack

Aaron (Ari) Bornstein
Open Source Engineer
About Me

• Website: pythiccoder.com

• Twitter: @pythiccoder

• Github: aribornstein
Intro To Municipal Water Management
That seems boring…
Why should we care?
Boring .... Why do we care?
Well we care when it all goes wrong!
The leading causes of sewer overflow:

**Inflow** water that enters sewers from inappropriate connections.

**Infiltration** water that enters sewers through defective pipe joints and broken pipes.

Image: Environment - Arlington County Muddy runoff from a construction site fills a street side drain.

Image: MWRA Infiltration in a sanitary sewer
Inflow and Infiltration are hard to predict.
Traditional Rules of Thumb

1. Place **flow meters** and **rain gauges** along the sewer pipe
Traditional Rules of Thumb

1. Place **flow meters** and **rain gauges** along the sewer pipe

2. Monitor during dry season
Traditional Rules of Thumb

1. Place **flow meters** and **rain gauges** along the sewer pipe

2. Monitor during dry season

3. Monitor during a period of high seasonal groundwater
Traditional sensor reading is a manual process.
Recently this has started to change.
IoT enables Realtime flow monitoring allowing for more accurate predictions.
Predictive Analytics

FlowWorks replaces manual time-intensive tasks by providing machine learning models to use when looking for patterns. Machine Learning can dramatically simplify what was previously an extremely tedious task. Automation allows you to monitor city infrastructure and water management systems more efficiently to protect municipal water infrastructure.
However sensors, especially, when submerged water, can sometimes fail.
It can be hard to differentiate between flow anomalies and sensor error which can severely disrupt prediction accuracy.
Our open source team partnered with Carl Data to solve this problem using Azure and Sci-Kit Learn.
Solution Architecture
Azure Machine Learning

✓ ML studio – Easily define experiments
✓ Build entire ML Apps & Deploy as Cloud APIs
✓ ML Applications Marketplace
✓ Integrates with other Azure Services
Solution Architecture
Azure Event Hubs

✓ Receive telemetry from millions of devices.
✓ Enables tracking in mobile apps,
✓ Collecting of stream data
Solution Architecture
Azure Stream Analytics

Get & Process Real-time Insight from

✓ Devices,
✓ Sensors,
✓ Web sites,
✓ Social media,
✓ Applications,
✓ Infrastructure systems
Solution Architecture
Power BI Embedded

- Machine Learning
- Stream Analytics
- Event Hubs
- Storage
- SQL Database
- HDInsight

✓ Enables Realtime Visualization of your data.
print ("That’s enough slides, let's see some code!")
Resources

