Machine Learning and Diagnostics for Electricity Theft Detection and Industrial IoT

Background

- Manage the data science team at C3 IoT. Previously at UC Berkeley working on the FORCES project.
- C3 IoT provides a development platform for largest scale enterprise. Application examples:
    - Revenue protection
    - Predictive maintenance
    - Supply chain optimization
    - Sensor health
    - Energy management
    - Inventory optimization
    - …

Revenue Protection (for Utilities)

- What are utilities doing today?
- How can people steal from meters?
- Back at UC Berkeley, Lillian Ratliff and I worked on electricity theft and anomaly detection. At that time, we used a mix of Matlab and Python to run offline batch analysis of cases.

Revenue Protection (for Utilities)

- What are utilities doing today?
- How can people steal from meters?
- Back at UC Berkeley, Lillian Ratliff and I worked on electricity theft and anomaly detection. At that time, we used a mix of Matlab and Python to run offline batch analysis of cases.

3 years later

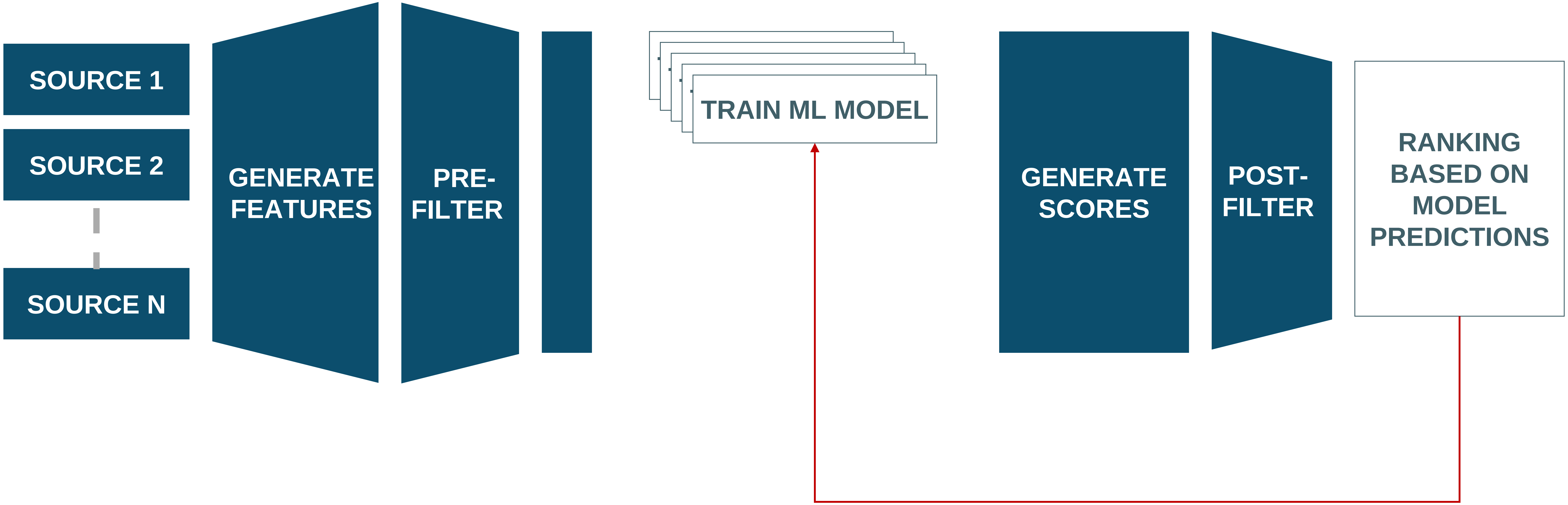
- **Online** (results get updated as soon as new data arrives & algorithms are automatically updated as new inspection results come in), runs in **parallel**, and converted it to **trade off** between maximizing energy recovery (**$**) and **precision** (related to some of the work that Amin et al. has done).

- Twice as good as the experts.

What is next?

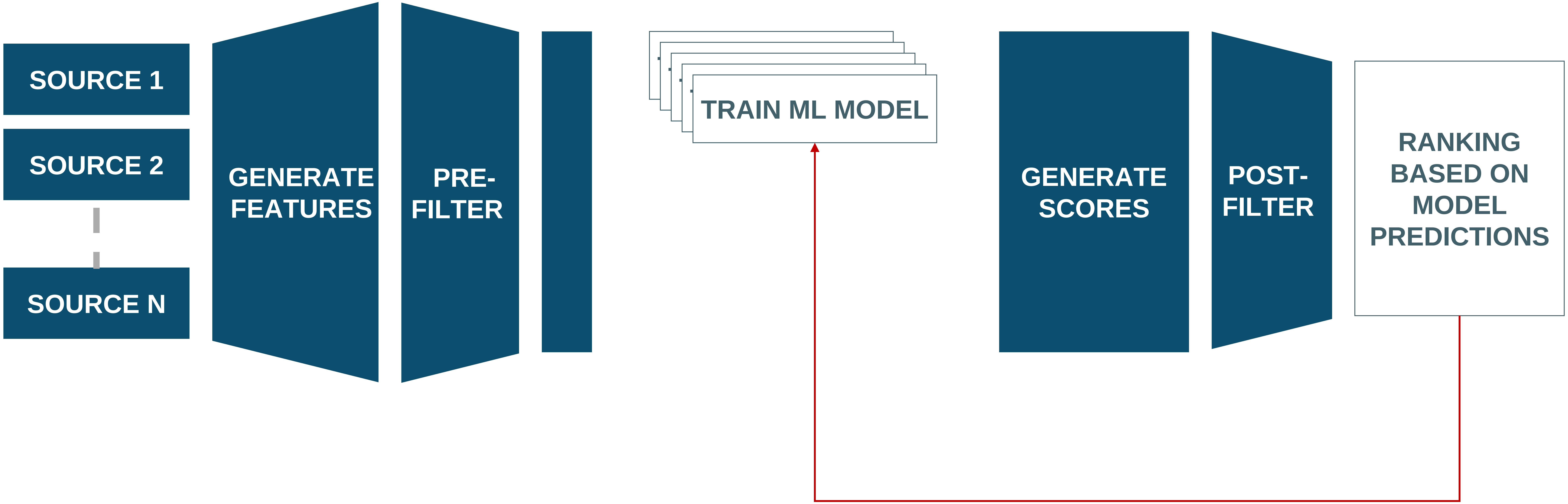
“C3 finds this meter suspicious because there is abnormally large consumption drop in March 2016. Please inspect this meter and look for a direct connect.”

Understanding Complex Machine Learning Pipelines



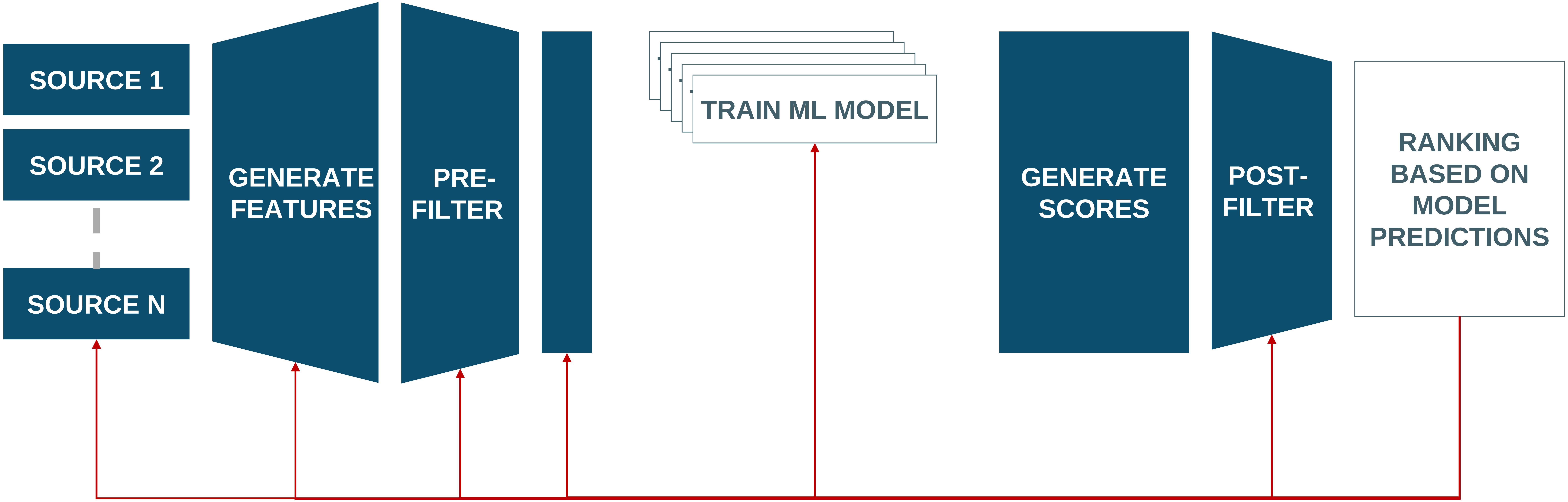
What part of the machine learning model explains the score?

Understanding Complex Machine Learning Pipelines



What part of the machine learning ~~model~~ **pipeline** explains the score?

Understanding Complex Machine Learning Pipelines



What part of the machine learning **pipeline** explains the score?

Motivation – *“We should strive to make models explain their decisions”*

General

- Explaining predictions
- Providing algorithmic transparency
- Debugging and optimizing ML pipelines
- Explaining drift in model accuracy
- Allowing more complexity

For C3’s customers – the above plus:

- Have ML questioned answered in the UI
- Improve trust in C3’s models and predictions

Product Specific use cases

- Predictive Maintenance – Anomaly Diagnostics:
  - Rank contributions of different measurements to prioritize further inspection
  - Understand how much each measurement deviates from expected values

CUSTOMER INTERFACE

- Revenue Protection:
  - Explain ranking of individual service points
  - Understand why features are important in model

DATA SCIENCE TOOL

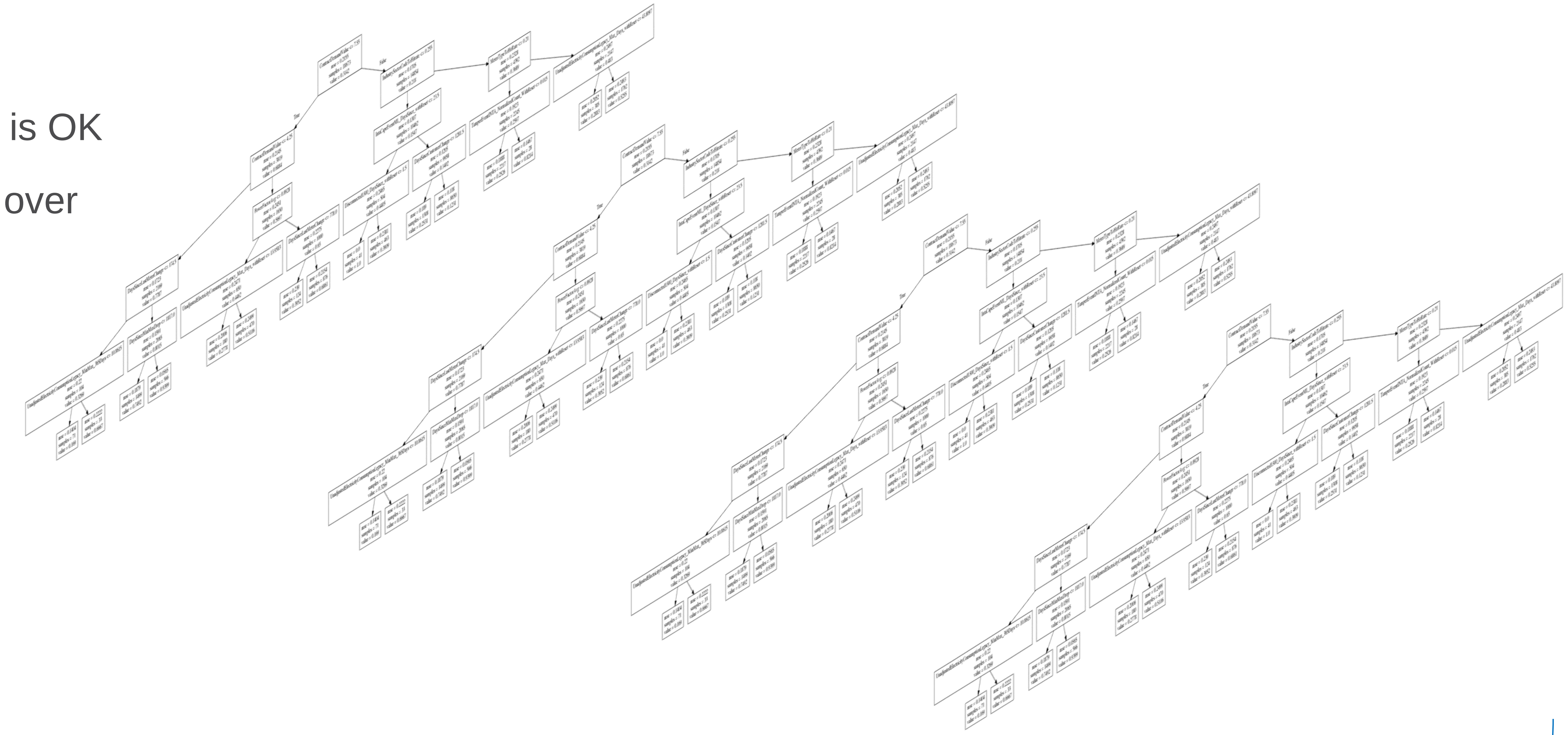
Tools: Tree Interpretation for Revenue Protection

Idea

- Fit a random forest to predicted scores
- Goal is to explain, not to generalize → overfitting is OK
- Retrieve the linear contributions for each feature over all trees in forest
- Rank features based on absolute contributions

Results

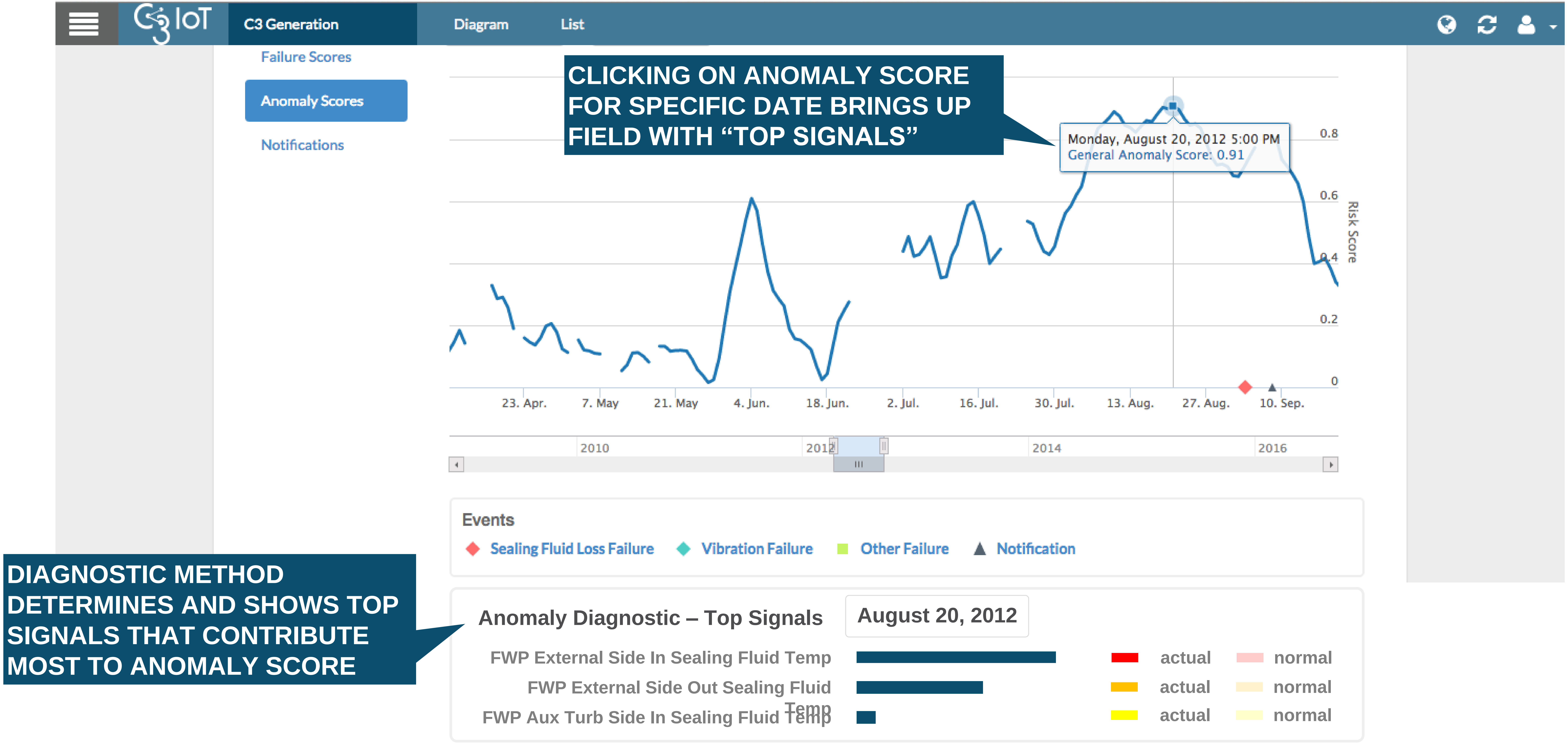
- Allows assessing feature importance for individual predictions
- Consistently ranked important features for Enel PM Generation
  - tool implemented with customer



Score = Bias + Contribution *1* + Contribution *2* + … + Contribution *m*

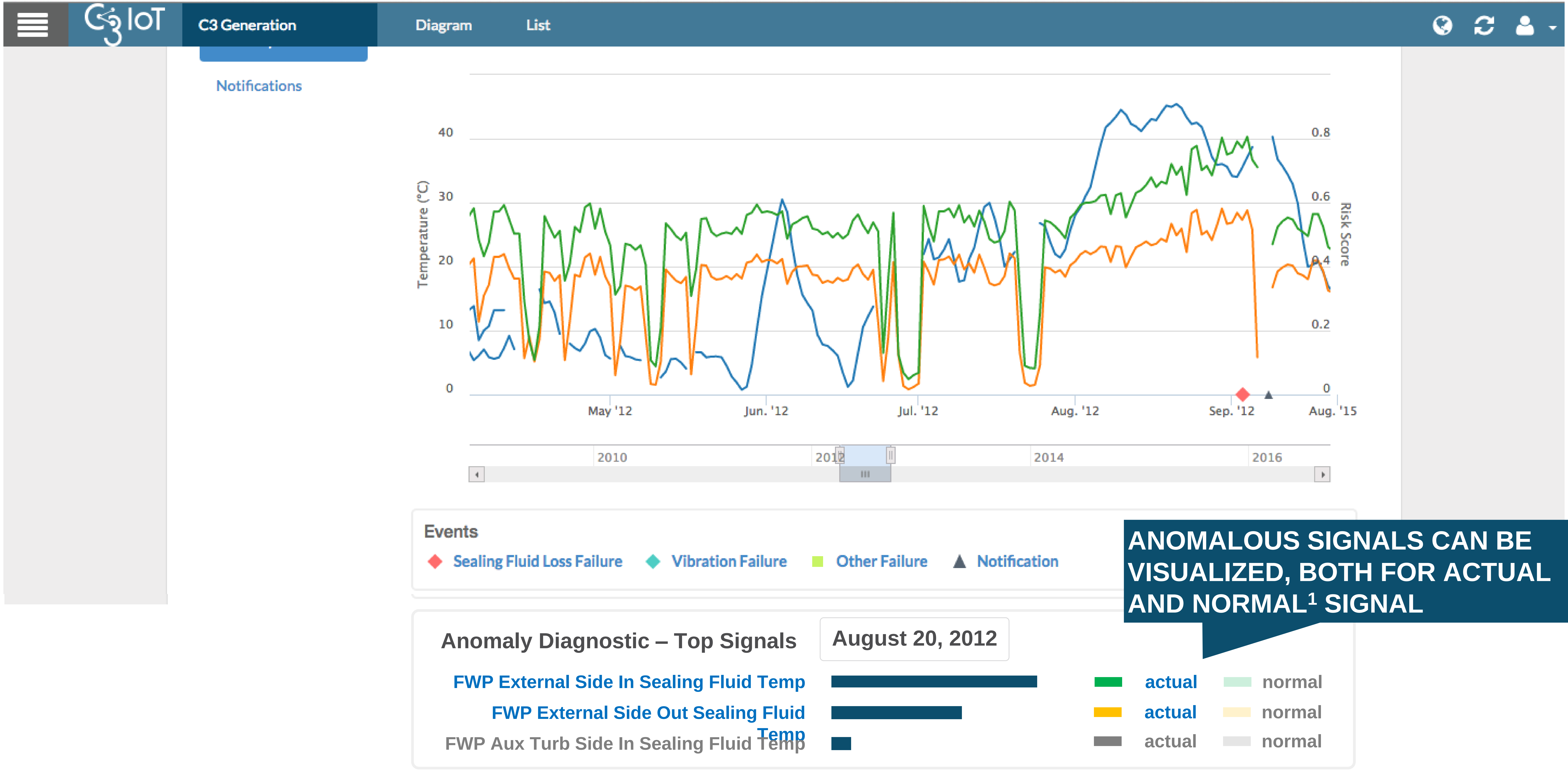
For all *m* features

Use Case – Anomaly Diagnostic



C3 IoT

Use Case – Anomaly Diagnostic



ANOMALOUS SIGNALS CAN BE VISUALIZED, BOTH FOR ACTUAL AND NORMAL[1] SIGNAL

1. Normal signal denotes the value of normal operation as estimated by machine learning model

C3 IoT