**Executive Summary:**

This model and dataset were created by IBM in order to demonstrate the feasibility of detecting anomalous signals or momentary drops in voltage as could be observed in Power transformers outputs. The model allows for a single hyperparameter, a threshold, to be adjusted to select the degree of sensitivity to how far the data lies outside of normal parameters. The model will identify the points in time where anomalous signals/momentaries were suspected.

This model could be used as a demo for Energy vertical or as a good first step for Power companies hoping to provide automated detection of such signals to incorporate into failure analysis or equipment maintenance schedules.

**Motivation:**

This project was motivated by discussions with clients in the energy vertical who described potential anomaly detection use cases for transformer and other signals. These clients keep their data in protected states which makes modeling on real data impossible. So a search was made to determine realistic data from whitepapers, journals, and other public data sources and then to create a model to simulate the signals so that many more experiments can be run. Subsequent to the simulation data or generated data project, anomalies (momentaires) matching client descriptions were in injected into the simulated signals. A second anomaly detection model was created that runs in near real time.

**FAQ:**

This short paper answers the following questions:

* Why is this a problem that requires an AI model to detect the anomaly?
  + Simple rules based approaches fall very short. The anomaly can easily hide in random noise inherent in the signal and be impossible to detect using simple rules based approaches such as setting thresholds
* Does the model need to be built/trained every time you change the hyperparameter or is it a variable that can be set when running the model?
  + Once the model is trained it can be run on any transformer signal similar to what is shown. The human adjustable hyperparameter “threshold” can be adjusted on the fly at runtime to increase or decrease the sensitivity to anomalies
* What does DSX do to help make building the model easier?
  + The model can be trained in DSX or DSX local allowing for easy collaboration among team members and easy deployment as a Restful API. DSX will allow easy Spark based ETL on the front end and deployment later as an inferencing model at the backend.

**Datasets:**

Naturally, we could not acquire actual data from client sources, so in order to model anomalies on something similar to what a Power company might see we created a data simulator. This simulator is a model unto itself that uses ML techniques.

The datasets were derived or simulated based on waveform patterns observed in a bulletin published by Toronto Hydro ([PDF here](https://www.elstersolutions.com/assets/downloads/Toronto_Hydro_Transformer_Monitors.pdf)) regarding Transformer Monitors.

The waveforms to be simulated was digitized from these charts from the bulletin.

|  |  |
| --- | --- |
| Toronto Hydro Transformer Waveform Charts | |
|  |  |
| Below: IBM Simulated Data from selected time ranges with/without Momentary injection | |
|  |  |

Based on these very few examples, IBM created a signal generator to simulate these types of signals.

Using this simulator, we generate simulated transformer data as follows with the same statistical distribution properties in time as the original data from Toronto Hydro. Not the vertical scale can be adjusted in the simulator as desired. Here we see something like eleven or twelve days worth of simulated signal

|  |
| --- |
| Twelve or so periods of thousand generated representing simulated waveforms |
|  |

**Momentary Injection:**

Our model allows a user to inject momentaries into the data stream at any time stamp in order to test the how well our anomaly detector works.

|  |
| --- |
| Same simulated waveform with a momentary anomaly injected (circled in red) |
|  |

**Model:**

This preliminary model is capable of detecting the anomalies by investigating lags and principal component analysis in order to display transformed data to estimate a good threshold which specifies anomaly level.

It successfully detects anomalies such as seen above where the momentary drop out is still within the daily variation although the smaller the momentary the more false positives would be detected as the anomaly threshold would have to be adjusted downward to detect smaller momentaries.

The detector will work with even a single time point drop momentary but the graph here used 4 time units to make the effect more visible.

Here is a shot of the PCA analysis with outliers marked and plotted so that threshold values can be chosen.

I used IBM DSX to host my workbooks to allow collaboration with colleagues and to set the stage for Spark transformations for possible future ETL and possible inferencing via Rest API.

To date we have NOT created the rest API for inferencing but this can be extended if the project is picked up for farther exploration by a client.

Next steps include:

1. Turning the model into a python pipeline for easier maintainability and faster inference and deployment as a REST API
2. Creating the Rest API and deploying as a web page

|  |
| --- |
| Detection of anomaly edges using near real time algorithm (anomalies detected by setting a detector threshold near a value of 60 |
|  |

**Results:**

Below is the same signal with the anomaly highlighted in red and contrasted with the simulated signals in blue. This demonstrates that anomaly detection for transformer like signals is possible within the DSX PowerAI framework whereas at least one client indicated frustration with their own attempts to create an anomaly detector. It is hoped that this model will accelerate them in their analysis/detection project for anomaly detection of signals.

|  |
| --- |
| Model Results (near real time detection of anomaly in red) |
|  |

**Summary:**

This preliminary model is computed using PowerAI using Jupyter Notebook and DSX. It successfully detects anomalies in data such as would be seen in transformer output voltages. It could be used as a demo for Energy vertical or as a good first step for Power companies hoping to provide automated detection of such signals to incorporate into failure analysis or equipment maintenance schedules.