Machine Learning with Network Transformer Metadata

Jon Garrity
CEO, Tagup
jon@tagup.io

Project Goal:

Apply new machine learning methods to predict network transformer failures.

The starting point: transformer metadata

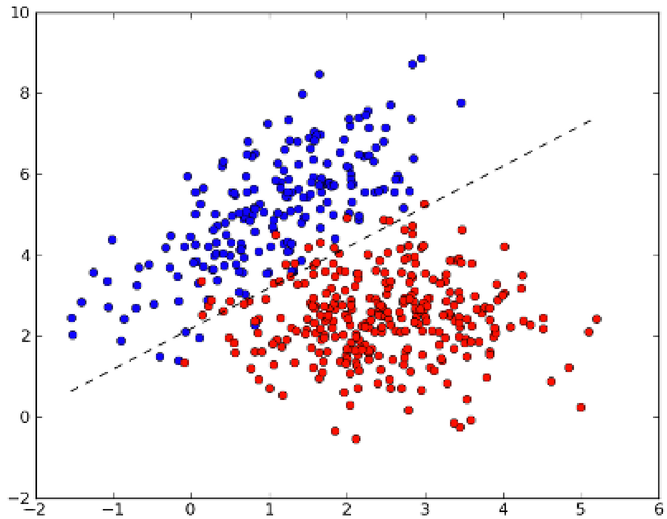
Can we use metadata and machine learning to prioritize assets for replacement?

- **Metadata includes age and transformer demographics** but excludes operating data (loading, temperature, etc.)
- **Apply labels to transformer state** (operational, non-operational, failure type)
- Project funded by NYSERDA under the EPTD High Performing Grid Program

Will a given transformer fail in 2016 or 2017?

Failure prediction can be framed as classification

- Any transformer that failed is a ***positive*** example; any transformer that did not fail is a ***negative*** example.
- After removing incomplete rows and encoding the categorical data, we have:
  - 26,736 labelled training examples (2,757 positive, 23,979 negative)
  - 9 features, encoded as a 112 dimensional vector



Initial Analysis: Insights & Observations

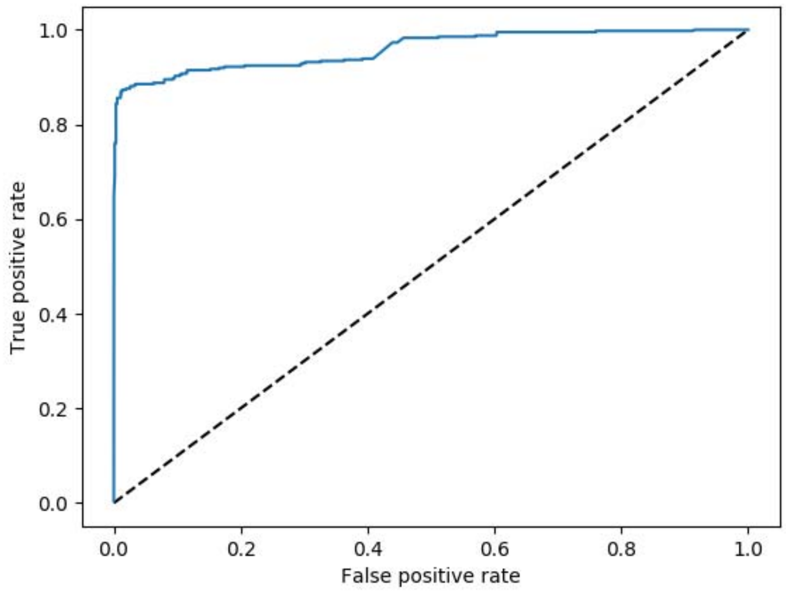
Metadata has some predictive power, but need to explore time series data

- After cleaning and aligning data, **26.7K training examples** with **2.7k failures**
- Metadata is sufficient to predict **69.5% of failures** and **99.8% of non-failures** in 2016/2017
- More information, such as inspection, DGOA or RMS data, is needed to accurately flag failing transformers without also flagging hundreds or thousands of non-failing transformers

Balancing sensitivity and specificity

Need to balance the cost of in-service failure and the cost of unnecessary removal

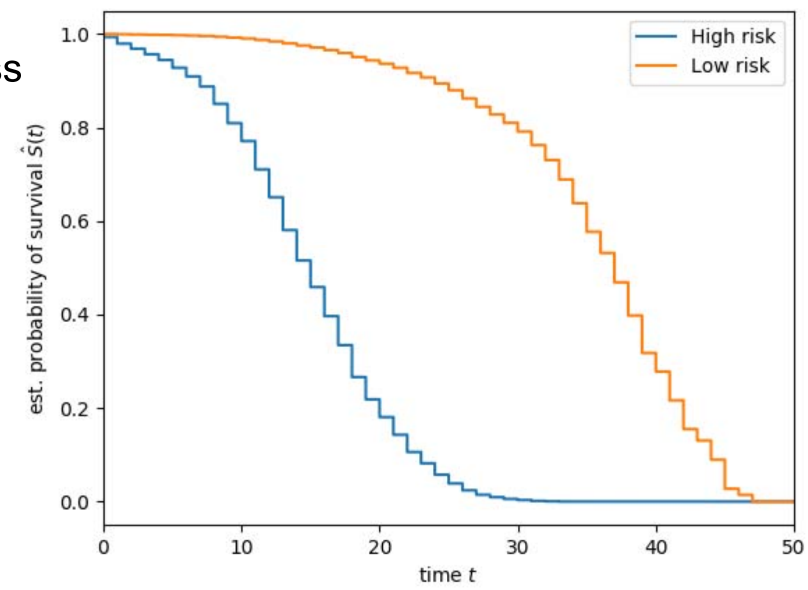
- Because failures are a small fraction (~1%) of the dataset, it is trivial to achieve 99%+ accuracy by assuming no failures will occur
- We have to balance between incorrectly flagging a transformer as a likely failure, and failing to classify a failing transformer.



Building a survivorship model

What is the probability that a given transformer will fail at a certain age?

- Cumulative distribution function for each transformer lets us calculate the “actuarial risk” of failing
- The curves can be updated as the model evolves, and will incorporate sensor/inspection data
- Failure risk can be used to prioritize maintenance and assess install base fragility



Initial Results

Transformer age strong indicator, demographic data improves model

- Unsurprisingly, the most predictive feature is the year of manufacture: older transformers fail more often
- Metadata is sufficient to predict **69.5% of failures** and **99.8% of non-failures** in 2016/2017
- During the test set, the model predicts 285 out of 410 actual failures

TAKEAWAY & NEXT STEPS

Although metadata is useful in predicting failures, **to achieve high accuracy, the models should incorporate time series data**, including loading, temperatures, pressure, DGOA, and maintenance records