If It Ain't Broke, Fix It?

Predictive Maintenance for BP's Gas and Oil Wells

By Ben Doremus

GOAL

Predict the likelihood of a part failing on any of BP's ~8,000 land-based wells in the lower 48 states.

DATA

I received data on Friday, Sept 1. There are three different data sets I have looked at:

- Surface Failures: ~23,000 entries with 37 features
  Failures with specific part models, geographic location of well, and date of failure

- Surface Failure Actions: ~30,000 entries with 47 features
  Actions taken, includes those with no component failures. Part category, failure reason, and text fields

- Work Management: ~69,000 entries with 179 features
  Field technician notes for all issues that they responded to

Additional data I would like to have:

- Weather at each well around the time of each part failure (temp and precipitation in particular)
- Cost to replace each part
- Well information at each location – find some proxy for how intensely the part gets used
- How many of each part exist in the field – find the failure rate, not just the number of failures

PROJECT PROGRESSION

Minimum Viable Product:

1. Combine Surface Failure, Surface Failure Action, and Weather data sets together.
   Ignore Work Management for now. Do some small amounts of feature engineering.
2. Find one part category that fails often, and focus only on it. (Compressors look promising)
3. Cluster to find groups of failures to see what may be causing that part to fail
4. Create a very simple predictive model

Improvement 1: **Look deeper into the selected part**

1. Engineer new features: part age, recent failures at the same well, usage intensity for the part, etc.
2. See how the new features change the clustering of part failures.
3. Create a more advanced model for the selected part type.

Improvement 2: **Expected Profit**

1. Using the cost of replacing the part and the number of parts in the field, find a threshold predictive value that warrants preventative maintenance.

Improvement 3: Increase the amount of data used

1. Incorporate the Work Management data set
2. Engineer new features: length of time since part was serviced, severity of prior service, etc.
3. See how the new features change the clustering of part failures.
4. Update the model for the selected part.

Improvement 4: **Draw deeper conclusions**

1. Find wells or part manufacturers that are outliers (good or bad). Explore reasons why they may be outliers, and make suggestions based on findings.

Improvement 5: Extend the scope of the project

1. Predict the probability of failure for other parts