Case Study

COMPLEX STATISTICAL MODELING FOR ACCURATE SERVICE LIFE PREDICTIONS

Highlights

Industry

- Energy

Outcomes

- Reliable statistical model to predict battery life and failures
- Prototype of a workflow to lifetime component (battery) analysis
- World-wide data consolidation for descriptive/predictive analytics
- Statistical model using physical properties and maintenance dates to anticipate sales

Technical Areas

- Data Science
- Predictive Analytics
- Component Lifetime Analysis
- Data Visualization
- Big Data

Technologies

- Hadoop
- Python
- SQLServer
- MapReduce
- Hive
- C#
- Sqoop
- Hue
- PowerBI

Statistics

- 3 datasets
- 10 million records

Data

- Sensor readings
- Batteries (age, resistance, failures)
- Maintenance actions
- Sales

Applying Data Science to Volumes of Disparate Information to Model and Forecast Known Unknowns

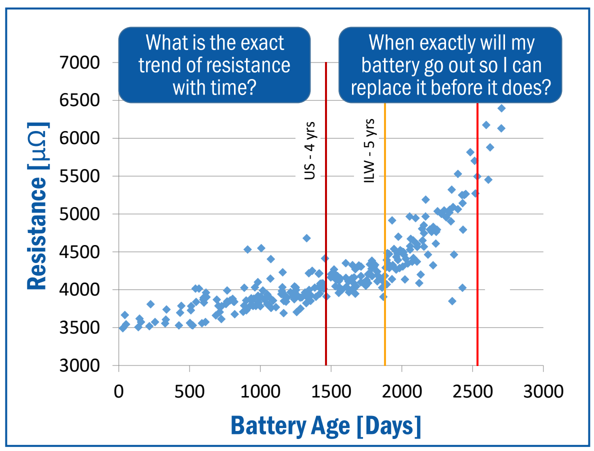
Business Challenge

A power, thermal, and infrastructure management company required a means to identify expected battery failure from historical measurements of battery resistance and other factors

The company had 10 years of maintenance records proprietarily formatted in several SQLServer and other databases, and wanted to pull this data together in a common data lake and link to other sources (sales, industry standards, marketing, etc.) to gain novel insights, improve decision making, and ultimately reduce costs.

Innovative Solution

ILW created a repeatable statistical model to accurately predict battery lifetime and failures, adding significant value to client decision making



ILW moved, copied, extracted, and analyzed millions of maintenance records and sensor logs for predictive insights. Data was ingested into a common pool in Hadoop from static files and enterprise-scale relational database management systems. An array of statistical, analytical, and modeling techniques were employed to understand battery lifetimes and enable novel descriptive and predictive insights. Python was used to facilitate data modeling, anomaly detection, and forecasts. Quality checks and updates were applied to sensor logs to maximize the value of time series data collected, isolate sample biases, and facilitate the modeling and interpretation of results. Visualization technologies were then used for understanding and communicating the meaning of the maintenance data and forecasts identified. These novel insights gained lead to significant cost savings and reduced downtime.

ILLUMINATION WORKS

CIO Review 100 MOST PROMISING BIG DATA SOLUTION PROVIDERS-2017

www.illuminationworksllc.com