Artificial Intelligence and the Internet of Energy (IoE)

In the past few years, there has been a lot of buzz surrounding artificial intelligence (AI) and machine learning (ML) in the news. From Google's self-driving cars to Apple's Siri to Netflix's movie recommendation engine to smart buildings to Nest thermostats, the list goes on of applications utilizing ML to automatically derive value from numerous input sources.

But how do these use cases apply to more conservative, industrial industries like energy and utilities that are often seen as technology laggards?

Contrary to what one might expect, AI and ML are a perfect fit for the energy and utilities sector. Given the vast amounts of data these industries generate—particularly with all the extra sensors added by the Industrial Internet of Things (IIoT)—the Internet of Energy (IoE) is the perfect environment for ML applications.

This is why SparkCognition has been working with many of the world's leading utilities to help revolutionize how energy is produced.

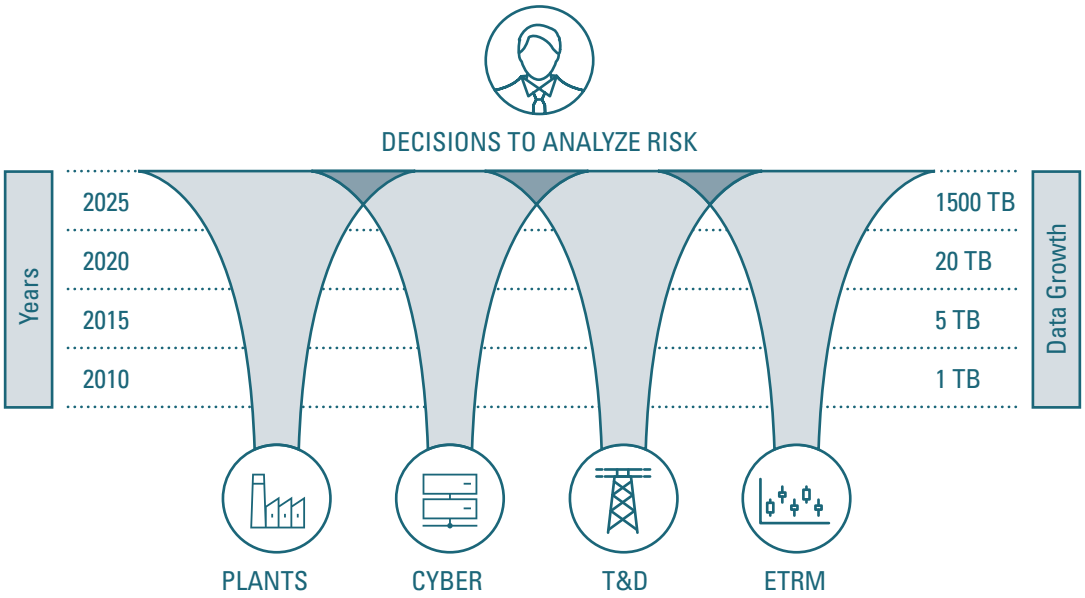
> *"One of the biggest things we're going to get out of it (machine learning) is learning what we don't know already. These systems are going to identify issues and opportunities that we've never even thought about looking for.*
>
> *That's the beauty of a machine being able to do it.... It's going to go find things that we hadn't even thought of and lead us to conclusions that we wouldn't have reached on our own."*
>
> ***—Jim Taylor, CTO of Tucson Electric Power***

Explosion of Data and the AI Approach

A significant obstacle to better ML integration among utilities is the traditional siloing of data across business units. Figure 1 represents the way many utilities are currently operating. There are several sources of siloed data across the major business units (Generation, Transmission and Distribution, Energy Trading and Risk Management, and Cybersecurity), and the data is growing at exponential rates.

–Figure 1–

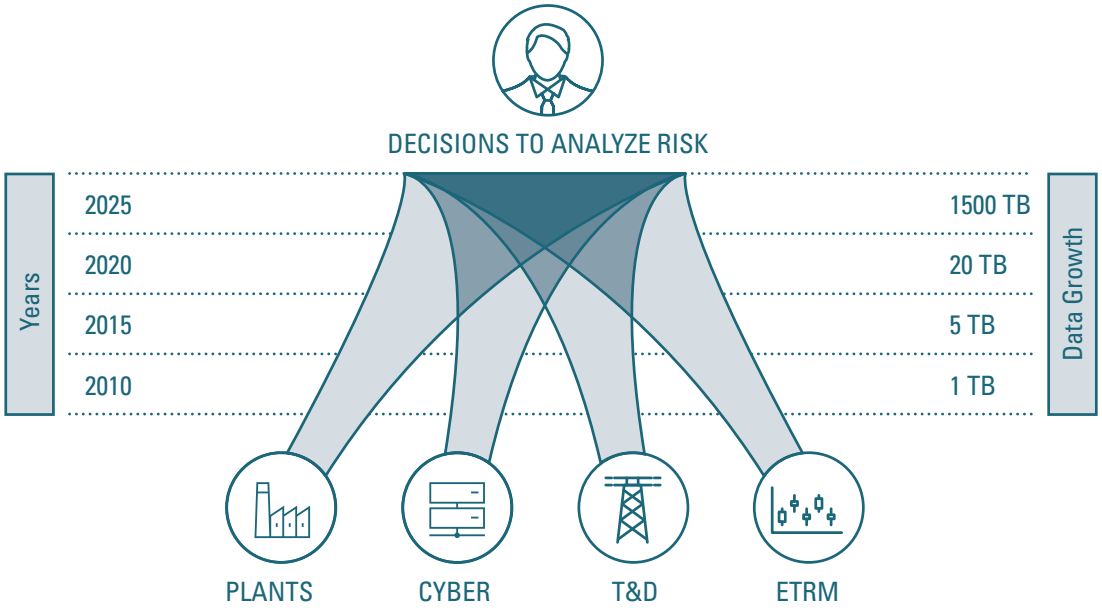
Sources of Siloed Data in Utilities



A method is needed of using big data technologies like Hadoop, Cassandra, and Spark to unlock the data, freeing it from historical stovepipes or silos. From there, AI capabilities can be incorporated to gain insight from all the data uniformly across use cases and the organization. This business transformation will undoubtedly lead to more powerful discoveries and insights, allowing what were once separate business units to merge as one.

–Figure 2–

Collapsing of Silos



To this point, much research has been done on how Artificial Intelligence and Machine Learning can impact an organization, and what benefits they can provide. For example, a major survey of utilities sited the top two benefits of Machine Learning to be 1) increased cybersecurity and 2) providing better data driven decision making.

Getting these benefits out of ML, however, means getting the wealth of data out of each of the silos (Generation, Transmission and Distribution, Energy Trading and Risk Management, and Cybersecurity) and putting that data to work to promote a better IoE experience.

The Internet of Energy

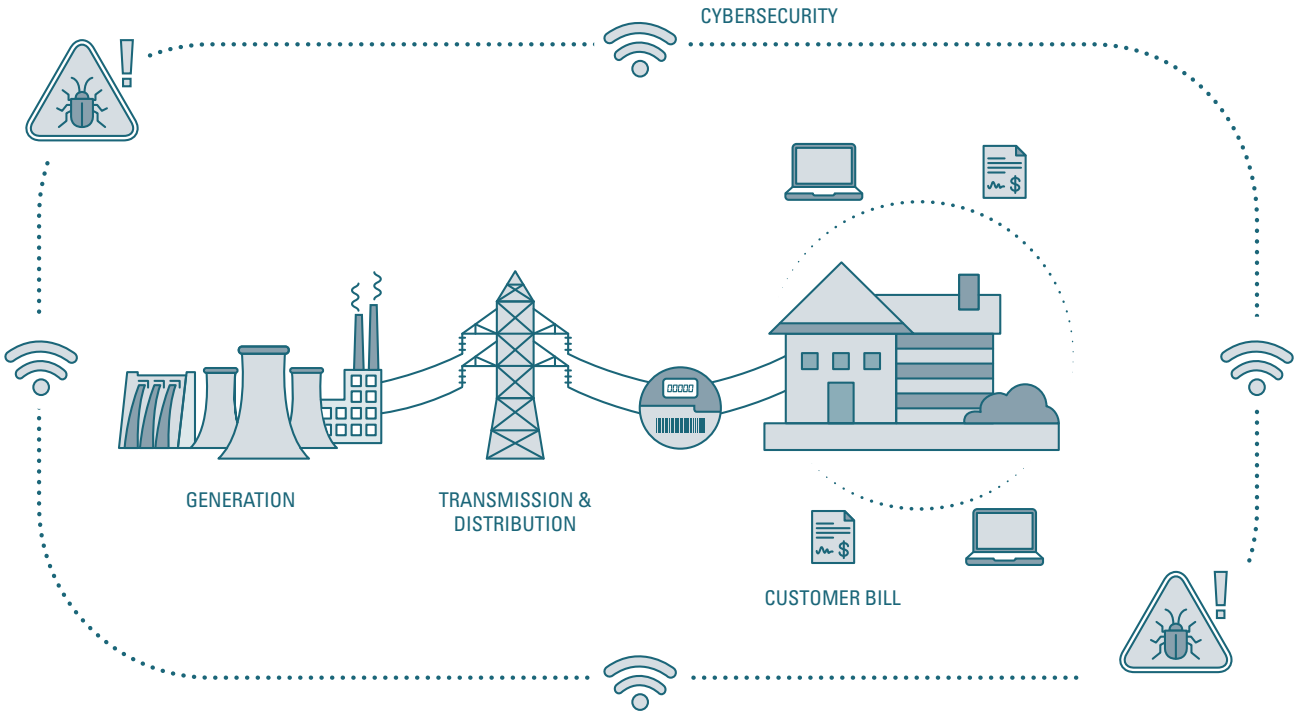
The Internet of Energy (IoE) can be broadly defined as the upgrading and automating of electricity infrastructures, making energy production more clean and efficient, and putting more power in the hands of the consumer. This paper discusses how to apply ML analytics to the four silos in the utilities industry to create the IoE.

From Figure 1, a typical utility is broken into four areas across silos. Figure 3, however, displays a different representation where the silos are broken down. Here, the IoE functions as one system where data is shared and analyzed, producing targeted, efficient results to utilities and consumers.

> *"There's a lot going on with analytics, but we continue to just build more and more silos of information. We need to start homogenizing it all. It all has to be integrated into the systems that will run the utility of the future."*
>
> —***Jeffrey Akers, Chief Strategist at Hewlett Packard Enterprise***

–Figure 3–

Internet of Energy



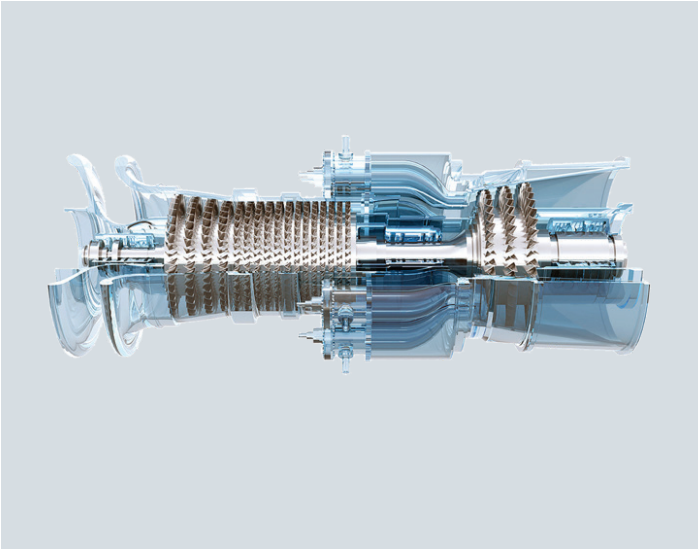
Generation: Prescriptive Maintenance of Turbines

The first major silo in the utilities sector is Generation, which relies heavily on the work of turbines. Turbines, whether they be fueled by natural gas, steam, nuclear, or coal, are massive engineering marvels from a mechanical standpoint. There are thousands of moving parts with extreme tolerances, and minute disturbances in the system can lead to major problems, causing downtime, loss of power, safety concerns, and more.

SparkCognition is applying machine learning and artificial intelligence to the problem of preventing catastrophic breakdowns and unplanned downtime for a large number of major utilities. Our proven technologies monitor hundreds or even thousands of sensors from turbine fleets (very much a big data problem), providing operators a real-time view of the system.

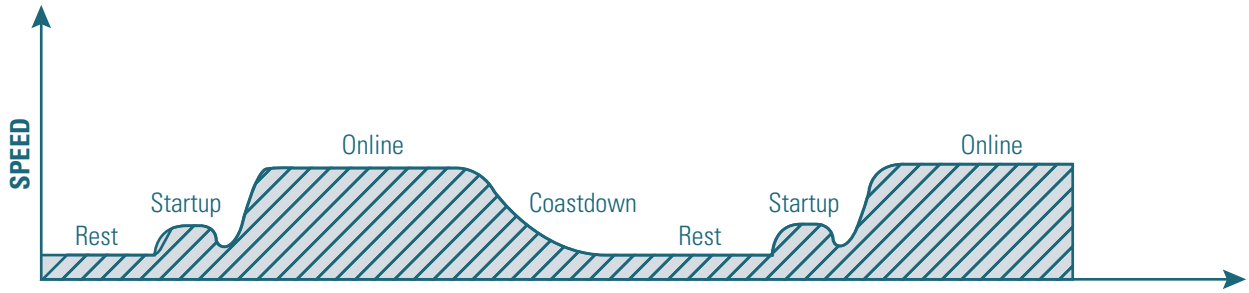
The traditional, standard approach forces subject-matter experts (SMEs) to spend time developing static, first principle models. Given the shortage of SMEs in the industry today, the standard approaches put a tremendous amount of burden on organizations to maintain and constantly update these models. And because the models are static, SMEs are constantly needing to adjust, maintain, and track them.

Due to the nature and mathematics behind these traditional models, operators are only able to look at steady state operation. However, more interesting and meaningful information is contained in an untapped source—transient events like startups and coastdowns (Figure 4).



–Figure 4–

Turbine Operation



Transient conditions are where many critical issues first materialize and can be identified. One challenge to transient analysis is these events occur over indeterminate lengths of time, so there needs to be an approach to normalize Startup A through Startup Z across time. This is an issue model-based systems can't solve, but AI approaches can.

SparkCognition employs a semi-supervised AI approach to the problem. On day one, the system starts analyzing data and providing meaningful information on abnormal startups with no user input or SME-developed models. The system then goes one step further and, similar to Facebook's "Like" button, learns and adapts based on SME input, improving its own accuracy and effectiveness. Additional benefits of AI in transient operations are highlighted in Figure 5.

–Figure 5–

AI Compared to a Traditional Analytics Approach

	Traditional Analytics	Artificial Intelligence
TIME AS A VARIABLE	• Single point event prediction • Good at steady state, non-time varying events • Can't handle transient events, where time is a variable rather than an index	• Complete time series treated as single event • Very good at comparing both steady state & transient events (Startup, coastdown, thermal vector, cooldown, etc.)
USER SKILL REQUIRED	• Knowing the relevant data is key to building a model • Model effectiveness depends on the skill of the user	• Relevant data and key features identified automatically • The information is inherent to the process • Value can be captured on day one
DIAGNOSTIC DATA IDENTIFICATION	• Fault diagnostic library must be defined explicitly • Key features are user calculated and added to the model for diagnosis	• Fault diagnostic data is inherent to the event • 1st, 2nd, & 3rd order features are automatically generated • User input integrated into system to "learn" from SME experience
DATA CLEANSING & MODEL TRAINING	• Model training is done by manually identifying and excluding bad points from a data set	• Model identifies normal (median) behavior and 3 standard deviations automatically

While the information from SMEs can be input manually, IoT and IoE make this unnecessary. Imagine the utility industry in the not so distant future where operators can approach a machine (using virtual reality), get an immediate list of alarms and readouts, diagnose issues with appropriate manuals and work order history integrations, and schedule necessary repairs. This is what the IoE is predicted to look like for utility systems.

Transmission and Distribution: More Than Just "Smart Meters"

AI has the power to tackle much larger problems for the second silo of Transmission and Distribution. There has been a great deal of excitement recently about implementing smart meters and end user control of in-home appliances to do everything from getting better usage reports and billing to unobtrusively turning off/on AC units. While these are certainly big data problems (96 million readings per day/million meters[7]), the analytics utilized are rudimentary.

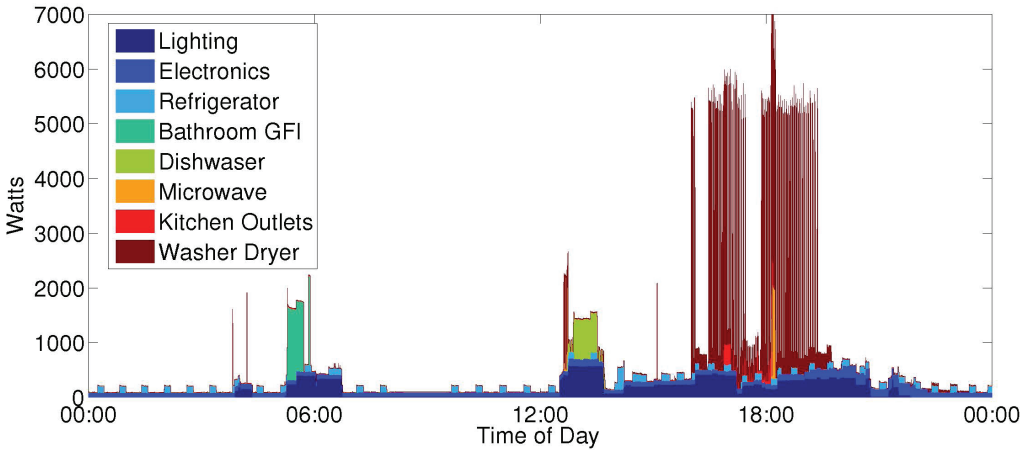
Specific areas where artificial intelligence is playing a key role include:

ENERGY DISAGGREGATION

Energy disaggregation is where patterns of usage per appliance are gathered by deconstructing information from a single home sensor. This application requires the utilization of ML because thousands of energy "signatures" must be analyzed to find patterns of usage (Figure 6). The savings are immense, as homeowners can discover how much every appliance contributes to their energy bill. On the operation side, analysis of energy signatures predicts suspicious consumption values due to physically or digitally manipulated devices, sophisticated thefts, meter malfunctions, and more.

–Figure 6–

Energy Disaggregation Example



POWER VOLTAGE INSTABILITY MONITORING

A phasor measurement unit (PMU) is an instrument installed within the grid, capturing high-speed, electrical waveforms to detect minute instabilities which propagate through the system. Unchecked, these instabilities can eventually cause the grid to collapse. This can be visualized to a similar phenomenon of the Tacoma Narrows Bridge failure of 1940, wherein small instabilities caused resonances resulting in a bridge which eventually collapsed in high winds[3].

A single PMU (of which there are thousands installed) can generate **4GB of data per day**, creating a true big data problem for which human analysis falls short. Today's analyses focus primarily on static methods, reading Power-to-Voltage curves and accounting only for steady-state power flow (like in our gas turbine example above), but the real value is in the dynamic, or transient, operations. The explosion of this dynamic data is leading researchers to utilize techniques like artificial neural networks, support vector machines, and multivariate regression analysis to identify voltage instabilities[4]—thus preventing brownouts and blackouts on the grid.

GRID MAINTENANCE

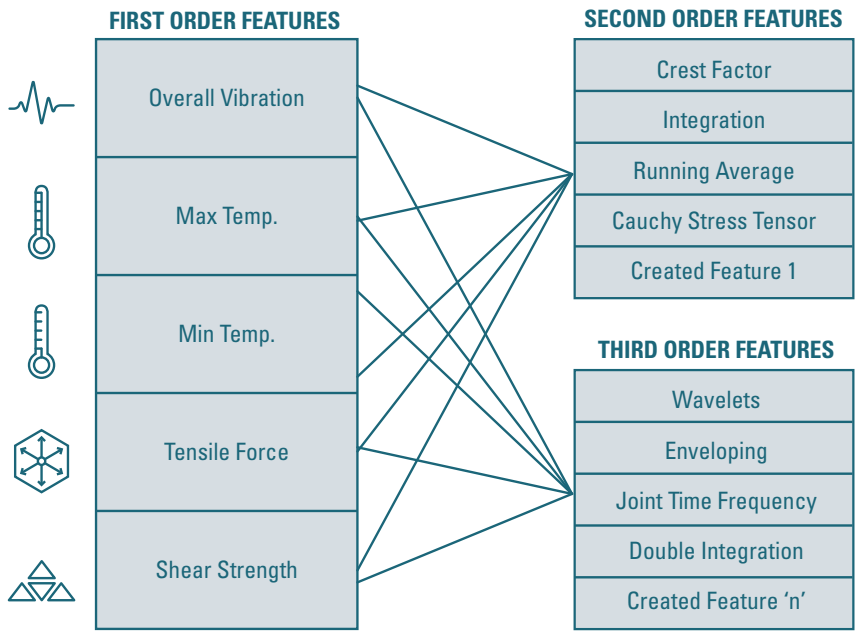
Many utilities struggle to use even the data they are collecting from devices such as SCADA systems. Data related to faults in breaker operations (a major grid component) often go unnoticed until major faults occur. But the data, sometimes called features in the ML world, is there. Features like vegetation index, ambient conditions, kVA loading, and historical maintenance can be used in ML algorithms to classify and ultimately predict failures well in advance. The savings associated are impactful, with reductions in capital expenses reduced by 4.5% and operational costs decreasing by 2%[5].

SparkCognition recognizes feature elevation and selection are major elements in a successful Machine Learning or Artificial Intelligence approach, and are developing patented techniques for this. One such technique is SparkArtemis™. It ingests first order features many analysts are used to trending on, and then continues to automatically expand each of those features into second and third order features—creating thousands or millions of new features where initially there might have been tens. This process is illustrated in Figure 7.

Traditional modeling history and experiences teach us we should be analyzing a very fixed set of features, but this is generally because of a human lack of multidimensional understanding. Because machine learning algorithms can automatically break features down into additional features and analyze them at machine speed, correlations which were never inferred historically are now being discovered.

–Figure 7–

SparkArtemis™ Feature Selection



Cybersecurity: The Modern Battleground

The third major data silo in utilities is Cybersecurity. In a recent poll, SAS and Zpryme[2] surveyed over 200 North American utilities, and identified the top five ML benefits. They found the top stated benefit of ML and AI for utilities was cybersecurity, because "the volume of network data exceeds that which can be analyzed by the human eyes, so self-learning algorithms will compliment business rules to strengthen critical infrastructure protection"[6].

The recent onset of attacks to critical infrastructure—with new malware such as Havex, BlackEnergy, Flame, and with Kaspersky Labs currently counting over 11 million unique virus strains with 28,000 being added daily—makes the need for new cybersecurity methods vital.

SparkCognition utilizes AI to identify, categorize, and remediate a variety of threats including loss of personal identifiable information, zero-day malware, and advanced persistent threat attacks.

Another form of threat, usually used as a medium for exploit delivery, is the socially engineered phishing attack. In this form of attack, employees with trusted access to a secure network may receive email or other digital communication pretending to be from a secure or known source. However, when employees take the action suggested by the communication, an exploit is triggered. The complexity of these attacks is increasing significantly, and this trend is expected to continue into the future.

Is cybersecurity so different from the previous examples of grid automation, prescriptive maintenance, PMUs, and the like? To a mathematical algorithm, there is little difference. Whether the input is a temperature, pressure, or vibration sensor versus a web proxy, a firewall log, or an IP address, they are all simply pieces of information with unique patterns to an algorithm.

> *"Signature-based anti-virus is dead; get over it!"*
>
> ***— Tung, 2008***

To combat the cyber front of industrial threats, SparkCognition developed the SparkSecure® execution engine. SparkSecure augments the workload of security experts by using proprietary ML software to analyze network data and identify and prioritize all types of threats. SparkSecure automates the threat research process, prioritizes threats based on confidence, and displays corroborating evidence to the analyst, significantly reducing both time to threat remediation and overall risk (Figure 8).

–Figure 8–

Traditional vs. Machine Learning approaches to Cybersecurity

Feature	Traditional Security Applications	SPARKSECURE®
THREAT DETECTION	Simple signature-based threat detection	Cognitive, signature-free threat detection
THREAT RESEARCH	Analysts must manually research potential threats after detection	Threats are automatically researched using natural language processing (NLP) technology
THREAT COMPARISON	Analysts must manually compare threats to online threat databases (e.g. CVE and NVD)	Threats are automatically checked against multiple threat repositories
THREAT PRIORITIZATION	Analysts must manually prioritize threats and corresponding workload, mistakes possible	Threats are automatically prioritized using file analysis, research, and NLP
THREAT ACTION	In most cases, analysts will have to kick off workflow to remediate	Integrates with workflow process to automatically initiate remediation

SparkSecure works by analyzing aggregated log data forwarded from either existing logging sources (i.e. a proxy server, firewall, web server, etc.) or a security information and event management software (SIEM). The analysis process consists of numerous ML pipelines, each of which features a variety of different threat-detection algorithms and techniques.

Finally, SparkSecure leverages SparkCognition's proprietary natural language processing API (DeepNLP®) in many of its pipelines. Like a human security analyst, DeepNLP not only searches the internet for threat evidence, but understands the written context of each threat. In doing so, SparkSecure separates, with confidence, truly malicious content from everything which is simply anomalous.

NLP can also be utilized in prescriptive maintenance, providing operators not just information that an asset is going to fail, but includes how to fix it, what parts are necessary, and provides technical information when the SME needs it.

Energy Trading and Risk Management

The final data silo is Energy Trading and Risk Management. As Figure 3 indicates, the very last stage of energy distribution is when customers receive their bill. In the highly competitive and regulated utility business, there is a clear link between the company's bottom line and forecast accuracy and reliability. However, legacy load forecasting solutions are not designed to handle the variability, complexity, and volume of data emerging in the utility landscape. Why is this important to consumers? If new techniques can provide more accurate forecasting, utilities can begin to offer better pricing to their customers.

> *"[Artificial Intelligence] is going to go find things that we hadn't even thought of and lead us to conclusions that we wouldn't have reached."*

As we've previously seen, AI techniques are providing insight into this process. With thousands of features from hundreds of sources (including weather, time of day, time of year, holidays, price of gas, wind speed, customer sentiment via Twitter/Facebook, etc.), there are infinite ways to combine and correlate information. Looking for subtle, transient movements of price data on an hourly or even a second-by-second basis with millions of combinations is where AI excels.

Because utility companies need to buy oil, gas, coal, nuclear fuel, and electricity, they are constantly at the mercy of volatile commodity prices. For this reason, utilities are using AI techniques to develop methodologies for market and credit risk aggregation. As an example, Old Dominion Electric Cooperative attributes four rate decreases in a year to the capabilities of advanced forecasting[7].

Conclusion

As the Internet of Energy (IoE) will undoubtedly continue its rapid growth over the coming years, the energy and utilities industry needs to be prepared. With current and upcoming improvements in the sharing of data from data silos, the entire industry can reap the wealth of new knowledge.

The siloing of information must be disrupted. When this occurs, AI systems "…are going to identify issues and opportunities that we've never even thought about looking for," says Jim Taylor, Chief Technology Officer of Tucson Electric Power. "That's the beauty of a machine being able to do it. If it can learn on its own, then we don't have to tell it what to go look for. It's going to go find things that we hadn't even thought of and lead us to conclusions that we wouldn't have reached"[2].

From prescriptive maintenance to energy trading to cybersecurity, the Internet of Analytics will play an important role in how energy is produced and provided to consumers long into the future. For utilities, it is recommended to start with pilot projects to demonstrate quick wins and value to an organization. As adoption increases, AI techniques will continue to learn and adapt, providing more value in the Internet of Energy.

References

1. *http://www.digitalistmag.com/iot/2016/05/04/how-the-internet-of-things-helps-utility-companies-04180965*
2. *https://www.sas.com/en_us/whitepapers/autonomous-grid-machine-learning-iot-utilities-108227.html*
3. *https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)*
4. *http://www.remotemagazine.com/main/articles/machine-learning-in-power-voltage-instability-monitoring/*
5. *http://www.elp.com/articles/powergrid_international/print/volume-20/issue-1/features/improving-grid-reliability-through-machine-learning.html*
6. *http://www.elp.com/articles/powergrid_international/print/volume-21/issue-9/features/part-ii-machine-learning-continues-journey-to-becoming-a-utilities-iot-ninja.html*
7. *http://analytics-magazine.org/utilities-dust-off-the-forecasting-playbook/*