# Machine Learning Use Cases in Electricity Distribution

Adrian Omambia Onsare

Utility Management & Consultancy Department, KPLC

adrianonsare@gmail.com

**Abstract**
The electricity distribution sector in Kenya has witnessed exponential growth over the last decade. The growth is underpinned by the Kenya National Electricity Strategy, which seeks to achieve universal access to electricity by 2022. The increased grid intensification has place increasing demands on power system reliability, quality and efficiency. In addition, the increased penetration of distributed generation and regional interchanges, have given rise to the need for improvements in power system planning, operation and control. Consequently, there are rapid advances in adoption of information and communication technologies. Some of these include: advanced metering infrastructure, distribution and asset condition monitoring sensors. The proliferation of these technologies has led to the rapid generation of vast amounts of variable energy data. To unleash the full value of this data, there is immense potential for the application of big data and machine learning techniques in electricity distribution. This paper not only offers an overview of these tools, but also presents sample use cases and proposed use cases. Further, it presents overview of technical challenges, change readiness and risks of adoption of these tools.
**Keywords: machine learning, big data, supervised learning, unsupervised learning, data streaming**

## Introduction

The Kenyan electricity sector is experiencing significant growth in both the scale of grid extensions and system complexity. The key underpinnings of the growth include: increased geographical coverage of the grid, increased grid customer base, proliferation of smart grid technology in grid operation and customer relationship management, increased penetration of non-programmable renewable energy systems, interconnection to neighboring utility grids and regional power pools and the increased uptake of distributed generation. The result of all these features is the generation of huge amounts of variable, high velocity data from all facets of the electricity sector; from electricity generation, transmission, distribution and retail.

Traditional power system analysis has been the sole standard for techno-economic planning and analysis of the utility grid. However, as mentioned before, the complexity of the grid is increasing. As such, traditional power system simulation is severely limited when subjected to high volumes of high speed, variable data. Power system simulation tools lack flexibility to extensively interface with other data sources that offer more analytical value. As a consequence, efficient and suitable data management and analysis is required to leverage this large amounts of structured and unstructured data to meet the demands of planning, operating and maintaining a modern power grid. This is the gap that Big Data technology and machine learning analytics can fill.

## Big Data and Machine Learning

### 2.1 Big Data

Big data generally refers to vast sets of structured and unstructured data [1]. For data to be classified as big data, it should have the following characteristics [2]:

I. Volume – The name “Big Data” in itself implies enormous data. Size of data is a key determinant in obtaining valuable insights out of data. In the context of the power system, real time monitoring of equipment, customer transactions and energy data inherently creates massive volumes of data.
II. Variety – This refers to both the source and nature of the data. Aside from structured spreadsheets and databases, the big data context includes a wide variety of source including photos, videos, sensor data, audio etc. The variability of unstructured data presents certain issues for storage, mining and analysis of data
III. Velocity – The term 'velocity' refers to the speed of generation of data. In the context of the power system, meter data, customer management systems and SCADA are examples of sources of high speed data.
IV. Veracity – This refers to the trustworthiness of the data

The big data ecosystem in the power system is summarized in the figure below:

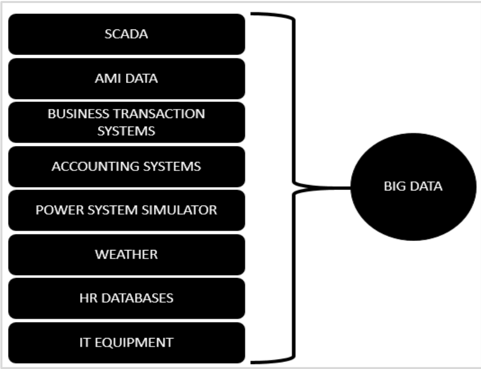


Figure 1: Big Data and Power Systems

Obtaining value from big data using traditional database and business intelligence approaches has over time proven to be technically and financially challenging enterprise. As such, technology has evolved, demanding specific technology for the storage, processing and analysis of big data. This is the province of big data technologies.

## 2.2 Machine learning

Machine learning is an important branch of artificial intelligence [3]. It uses largely obscure statistical tools that allow machines to improve on tasks with experience. Machine Learning algorithms enable the computers to learn from data, and even improve themselves, without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available.

Machine learning implementations are classified into three major categories, which are outlined below in turn.

I. Supervised learning: An algorithm learns from example data with provided categories or classes. The algorithm learns from existing data, and is then able to correctly classify new data into the predefined classes. An example would be to predict whether an action lead to a power outage or not.
II. Unsupervised learning: In this case, the algorithm is left to learn and decide the data patterns on its own. It achieves this by observing the degree of similarity between data points, and then deciding on the optimum clustering approach. Applications include: recommender systems for online purchases, electricity customer segmentation and fraud detection.
III. Semi-supervised learning: In this case the model is fed a dataset with some of the target outputs missing from the training dataset.
IV. Reinforcement learning: The machine seeks to optimize the solutions after each cycle of learning using a feedback loop. Applications include computer chess engines, self-driving cars and self-healing power systems

## 2.3 Big data and machine learning in Power systems

Within the power distribution sector, machine learning is used within a big data context to offer solutions in descriptive, predictive and prescriptive analytics. Descriptive analytics uses historical data to explain observations, predictive analytics leverages machine learning to predict future occurrences based on historical data. Prescriptive analytics seeks to offer recommendations based on insights gained from analytics. The interaction between big data and machine learning is illustrated below.

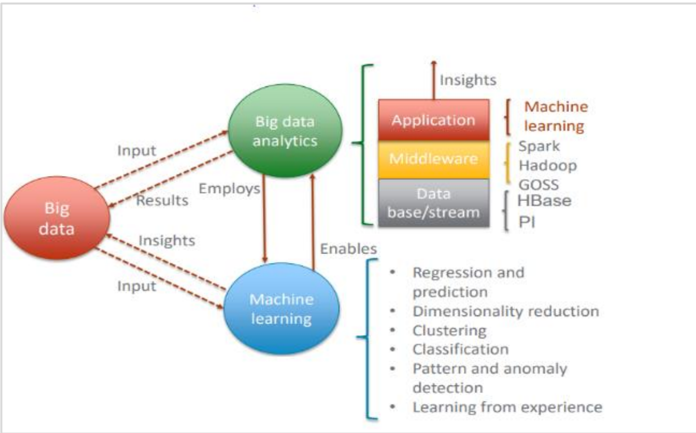


Figure 2: Big Data, Machine Learning in Context [4]

In the context of the electricity value chain, the following possible use cases are identified.

| Power System Data | Power Generation | Power Transmission | Power Distribution | Demand Side Management |
|---|---|---|---|---|
| • Equipment Data<br>• Environmental Data<br>• Management Data<br>• Operating Data | • Generation Efficiency Improvement<br>• Power Plant Operations<br>• Electricity Price Modeling<br>• Renewable Energy Planning management<br>• Economic Load dispatch<br>• Generation Market modeling<br>• Generation planning and optimization | • Power Transmission Planning<br>• Grid Loss Identification<br>• Islanding Detection and Isolation<br>• Asset management and Lifecyle Planning<br>• Outage Detection and Restoration<br>• Transmission Line Fault Detection and Monitoring | • Fault Detection and Identification<br>• Transformer Health Monitoring<br>• Outage detection and restoration<br>• Loss reductionPower Retail | • Customer Segmentation<br>• Revenue protection and loss detection<br>• Customer load forecasting<br>• Distributed Generation Monitoring |

Figure 3: Big Data and Machine Learning Use Cases In the Electricity Value Chain

The core advantages of machine learning in big data analysis is the ability of a machine learn automatically, handle large varieties of data, continuous self-improvement, and the breadth of its applicability. Sample use cases are presented below in turn.

## Machine Learning- Sample Use Cases

### 3.1 Case 1: Unsupervised Learning: Anomaly/Fraud Detection

In detecting anomalies, the model seeks to establish a baseline for what it considers anomalous, and what if considers normal. It achieves this by learning from both historic and live data. With anomalies detected, it is easy to flag an anomalous occurrence in real time and precipitate corrective action, which will in turn reduce overall equipment downtime and failure. This case study involved analysing energy consumption data from a customer with a history of suspicious variations. An unsupervised learning model – Isolation Forests, was used to identify anomalies in the time series current data. With the anomaly range identified, it is possible to flag such issues ahead of time, before a business loss is incurred.

The plot below visualizes normal and anomalous readings from energy meter readings.

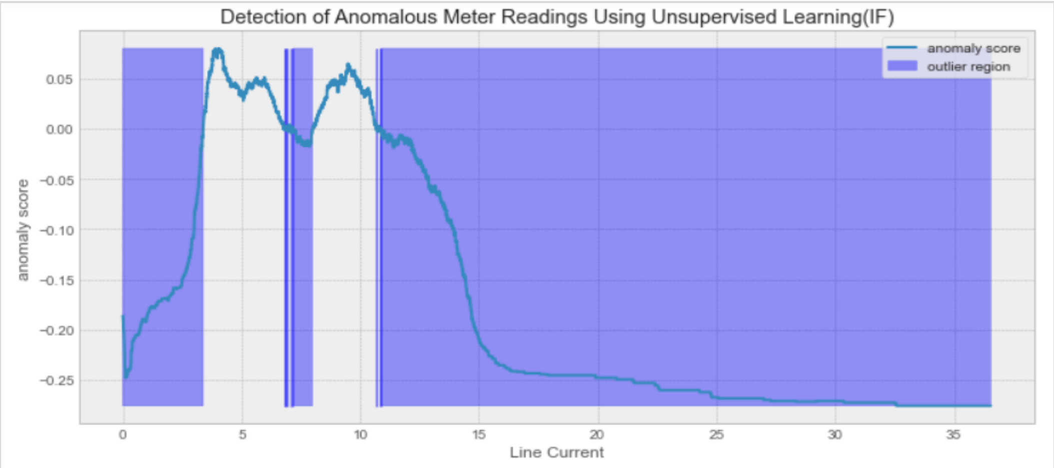


Figure 4 :Anomaly Detection Using Unsupervised Learning

### 3.2 Use Case II: Supervised Learning

**Customer Demand Forecasting**

It is of interest for a power distributor and consumers as well, to wish to not only keep tabs on their electricity costs, but also to afford themselves a level of predictability in consumption. Machine learning finds its use in this context. This case involved mining of data from a meter reading portal, for a certain customer, over a 1-year period, and deploying a Decision Tree Regressor model to identify consumption trends. Decision tree belongs to a class of models known as "Tree Based Models", which builds regression or classification models form of tree structures, by using statistical tools to breakdown data into subsets.

The meter data was pre-processed, split into a "training" dataset and a "testing" dataset. The training dataset is used by the model to identify the underlying patterns in consumption and attempt predictions on it, while the test dataset is used to validate the model. If the accuracy score of the model is high, then the model can be deployed in a real use case.

The dataset in question was pre-processed, and then used to train and evaluate the model. The model predictions are then visualized versus the actuals

.

Figure 5: Customer Load Prediction Using Supervised Learning

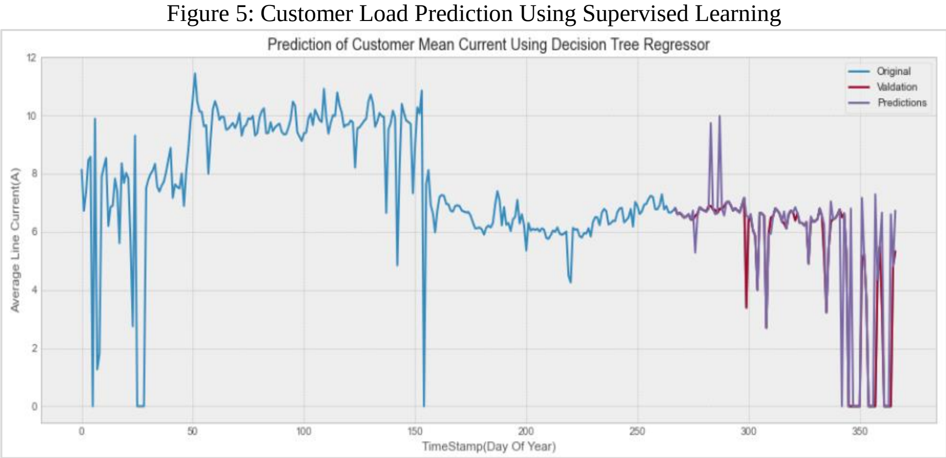


Figure 6 :Customer Demand Forecasting

**Solar PV Plant: Prescriptive Maintenance**

Kenya presently has over 100MW of installed grid-scale solar pv systems. The sheer size of these plants, as well as the non-programmable nature of the resource, presents challenges for both grid operators and solar power plant developers. A key element, is the need to monitor performance of the plant, by comparing it to an ideal "digital twin". This comparison allows for identification of performance deviations due to component failure, shading or dust. This case utilized production data from a real solar pv power plant in South Africa, and compared its performance to an ideal representation of the plant. The digital twin utilized a linear regression model, which employed component data and weather/climate data obtained from weather APIs.

The actual power output was the resulting predictions was then compared to the expected ideal, and the anomalies from each date identified for prescription of the adequate maintenance intervention.

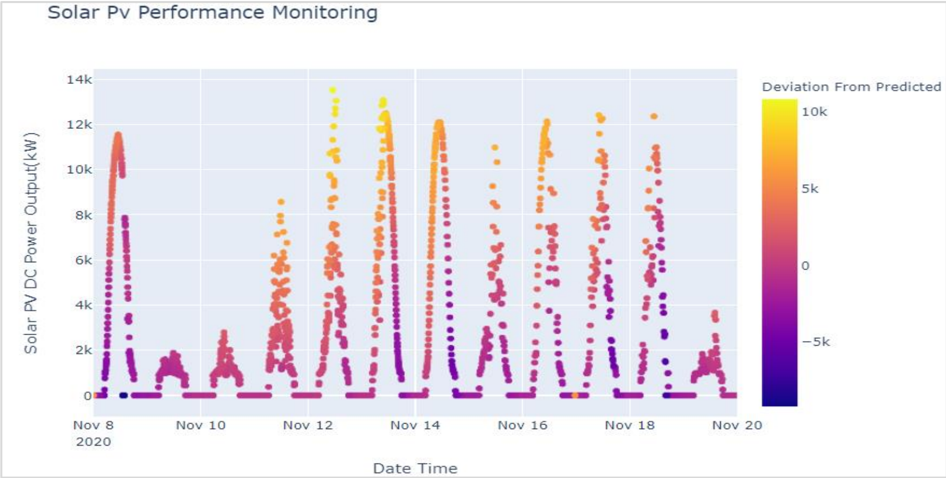


**Figure 7: Solar Pv Plant Performance Monitoring**

## Risks and Change Readiness

Plea The use cases above have presented a strong case for leveraging machine learning and big data tools in power systems planning, operations and maintenance. However, there are requirements for adoption and attendant risks. These are summarized as below:

- Data Security: Some of the more efficient and powerful tools leverage cloud solutions. Cloud services bring with them serious concerns with information security
- Data Architecture and Availability: Machine Learning algorithms require massive training datasets. These datasets should be of good quality and unbiased. This is a challenge in a sector that is yet to fully automate processes.
- Value chain synergies. For instance, implementation of Automatic Generation Control (AGC) would require synergies between the System Operator, the TSO and Generating entities. AGC can then leverage machine learning to optimize dispatch and ensure system security.
- Capacity Gaps: There is a dearth of engineers in the sector capable of leveraging big data and machine learning in grid planning, operations and maintenance. There would have to be deliberate capacity building in this area for active uptake to occur.

## Conclusions

The global transition to smart utilities is inevitable, and so is the widespread application of big data and machine learning in the planning, operation and maintenance of the Kenyan utility grid. The utility sector needs to create synergies that would ensure the mainstreaming if this technology in grid operation. The expected benefits of the technology outweigh the risks. The result would be a more efficient, reliable, secure and affordable electricity supply.

## References


1. AW. Lidong and C. A. Alexande, "Machine Learning in Big Data," International Journal of Mathematical, Engineering and Management Sciences, vol. Vol1 , no. No 2, pp. 52-61, 2016.
2. Q. Huang, "Review and Prospects of Machine Learning for Power System Analytics and Operation," in Workshop On Data Analytics for the Smart Grid (DASG), 2017.
3. Guru 99, "What is BIG DATA? Introduction, Types, Characteristics & Example," Guru 99, 2020. [Online]. Available: https://www.guru99.com/what-is-big-data.html. [Accessed 01 11 2020].
4. Udacity, "Machine Learning for Big Data," Udacity, 14 August 2014. [Online]. Available: https://blog.udacity.com/2020/08/machine-learning-for-big-data.html. [Accessed 01 November 2020].