

## Smart Meter Data Analytics using Orange - Ankit Mahato

### About the speaker

1. Ankit Mahato, Product Manager, Fuzzy Logix Inc.

Ankit is a Product Manager with 3+ years of industrial experience in machine learning, quantitative modelling, data analytics and visualization. Over the years, he has developed an expertise in handling the entire data analytics pipeline comprising – ingestion, exploration, transformation, modeling and deployment. He is a polyglot programmer with an extensive knowledge of algorithms, statistics and parallel programming. He has shipped multiple releases of DB Lytix, a comprehensive library of over 800 mathematical and statistical functions used widely in data mining, machine learning and analytics applications, including "big data analytics". A die hard Pythonista, Ankit is an open source contributor and a former Google Summer of Code 2013 scholar (under Python Software Foundation). He is also actively involved in the Indian Python community and recently organised the workshop - Making Machine Learning Fruitful and Fun using Orange in PyCon India 2017.

An IIT Kanpur alumnus, Ankit is also an active researcher with publications in international journal and conferences. He is actively working in the domain of IoT Analytics and recently presented his work - "In-database Analytics in the Age of Smart Meters" in the 5th IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence, 2017.

2. Ashita Prasad, Strategy Analyst, Accenture Strategy

Ashita is an IIT Kanpur and IIM Ahmedabad alumnus having a keen sense of technology and its application in business. She was actively involved in research in 3D Microscopy in IIT Kanpur where she developed advanced 3D visualization toolkit using wxPython and Vpython. In IIM Ahmedabad, Ashita utilised her domain knowledge in e-commerce to undertake development of new Batching algorithms in Python for Operations Management. Currently, she is working as a strategy analyst helping firms in deriving value from their data through the strategic application of technology and analytics.

### Abstract

The rapid depletion of energy resources has made energy security and management a major area of interest for the utility industry. In several countries, deployment of smart meters is leading to a widespread adoption of smart systems for energy management. This smart utility service forms a building block of the Internet of Things framework, and the growing amount of data captured by smart meters provide an opportunity to gather various analytical insights. Currently, there is a need to better visualize the high-volume of data captured by smart meters and provide a means to effectively gather various analytical insights which can help in better understanding the energy usage patterns, energy leak/theft and managing peak loads.

This paper demonstrates how Python based data mining and machine learning toolkit – Orange can be used for visualizing and analyzing the smart meter data. The component based toolkit provides APIs to extend it and build new plugins and widgets for custom requirements which can be used for smart meter data visualization and analytics. It even allows the development of widgets capable of in-database analytics which is faster than traditional analytics as the analytics is pushed inside the database.

The article presents a real-life use case of an energy utility firm whose objective is to identify the households which have similar energy usage patterns. As k-means clustering is flat in

nature and the data contains high number of dimensions, a hierarchical clustering based approach is used for cluster division. Orange ML toolkit is used for the calculation of distances and the application of a Hierarchical Clustering model. A custom widget is built using matplotlib to view the selected clusters in real time. The final clusters obtained demonstrate the effectiveness of the unsupervised learning algorithm used in this use case. The identified consumer clusters enable the utility service provider in preparing segment-specific campaigns and tariffs to efficiently shape the future energy usage patterns and distribute the peak loads.