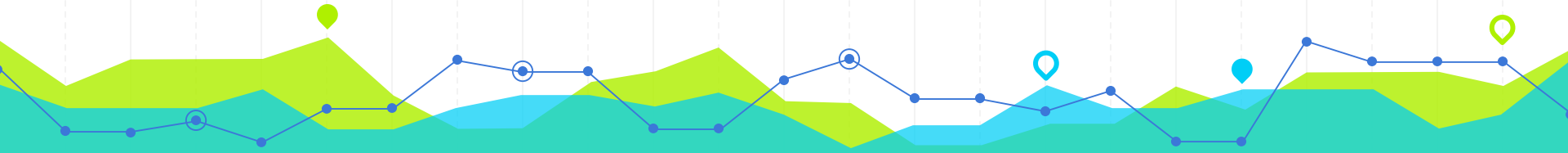
SciPy India

Smart Meter Data Analytics using Orange

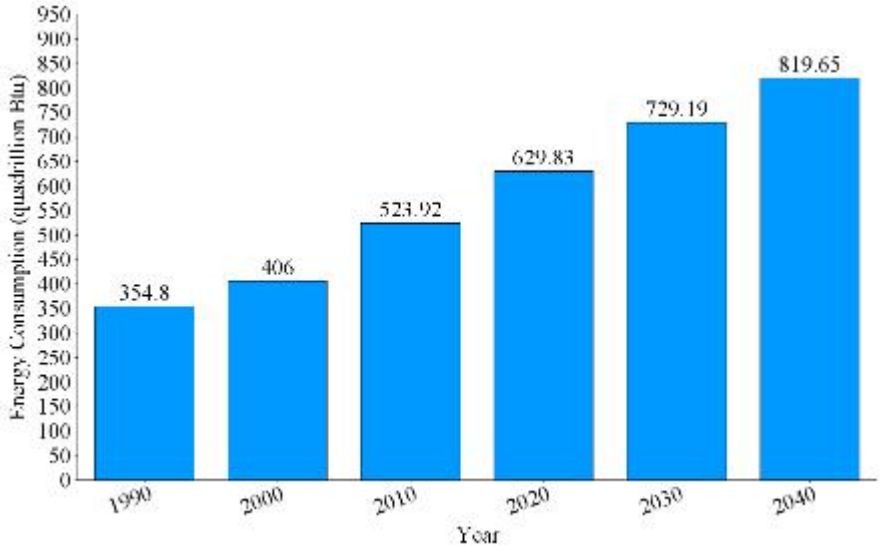
Ankit Mahato
Ashita Prasad



Smart Meter Data Analytics using Orange 1

PRESENT SCENARIO

- By 2040, world energy consumption will grow by 57%.
- Coal, oil and gas form 80% of the world energy supply and their sources are depleting rapidly.



Application of Advanced Analytics

In Smart Grid Infrastructure by Energy Utility Industry

Leads to Greater Energy Security and Management

Smart grid infrastructure forms the core component of the Internet of Things (IoT) framework, where devices across various industries are inter-connected utilizing the internet, peer-to-peer networks and closed networks to provide valuable information to the consumers and the product/service providers.

IoT 50,000,000,000

connected devices by 2020!

IoT

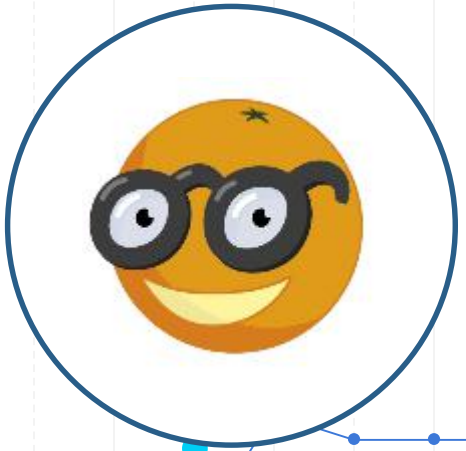
- Enables efficient management of smart grids to reduce losses and meet the energy demand distribution.

HOW?

- Through large scale deployment of smart meters.

WHAT?

- Smart meter records the energy consumed at every 15-minute interval allowing quantitative measurement, calibration and control, communication and power management.
- Recorded data is communicated back to the data warehouse and it is used for analysis of energy consumption patterns and future demand prediction.
- Smart meter analytics help customers in better understanding their energy usage patterns and obtain personalized feedback and incentives to save energy and money.



Smart Meter Data Analytics using Orange 2

Orange - History

- 1997: Development started by Janez Demšar & Blaž Zupan.
- Development continued at the Artificial Intelligence Laboratory, University of Ljubljana and is currently undertaken at Laboratory of Bioinformatics, University of Ljubljana.
- Conceived as a C++ library.
- Aimed at constructing a platform where advanced users would write their own components in C++. Also, the components were packed into programs with command line interfaces.

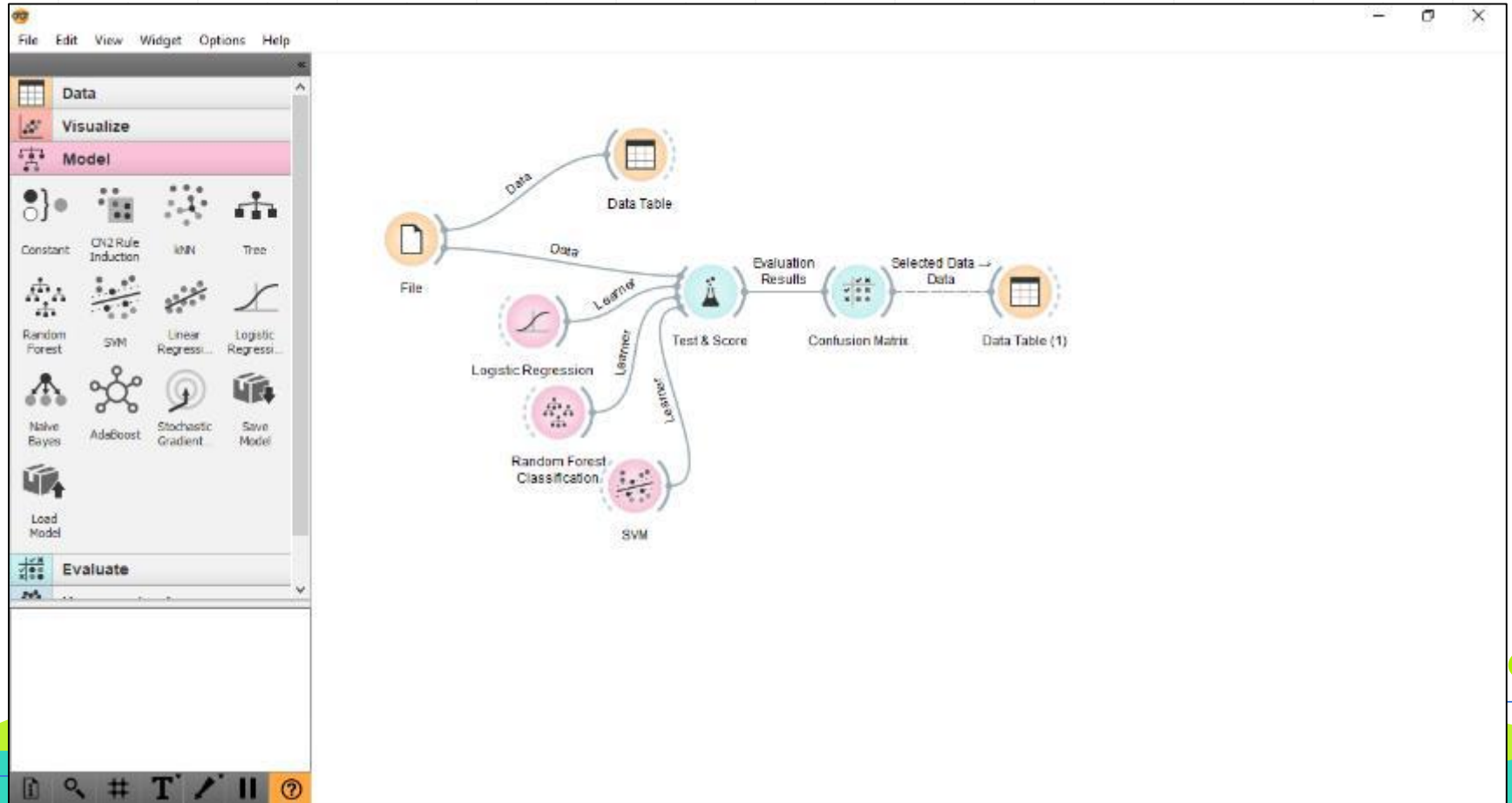
Orange – Transition to Python

- 1999: Orange is migrated and used exclusively as a Python module.
- Clean and simple syntax that is easy to learn and understand, enabled wider adoption and collaboration.
- Migration to Python also simplified the development of a graphical user interface.
- Benefited users by providing a simple interface to rapidly prototype new algorithms and test custom components.

Orange - Today

- Open-source (GPL license) framework that features both scripting and visual programming.
- Provides data visualisation and data analysis for novice and expert, through interactive workflows.
- Large widget toolbox and several add-ons.
- Possibility to use it programmatically via GUI (Orange canvas, PyQt).

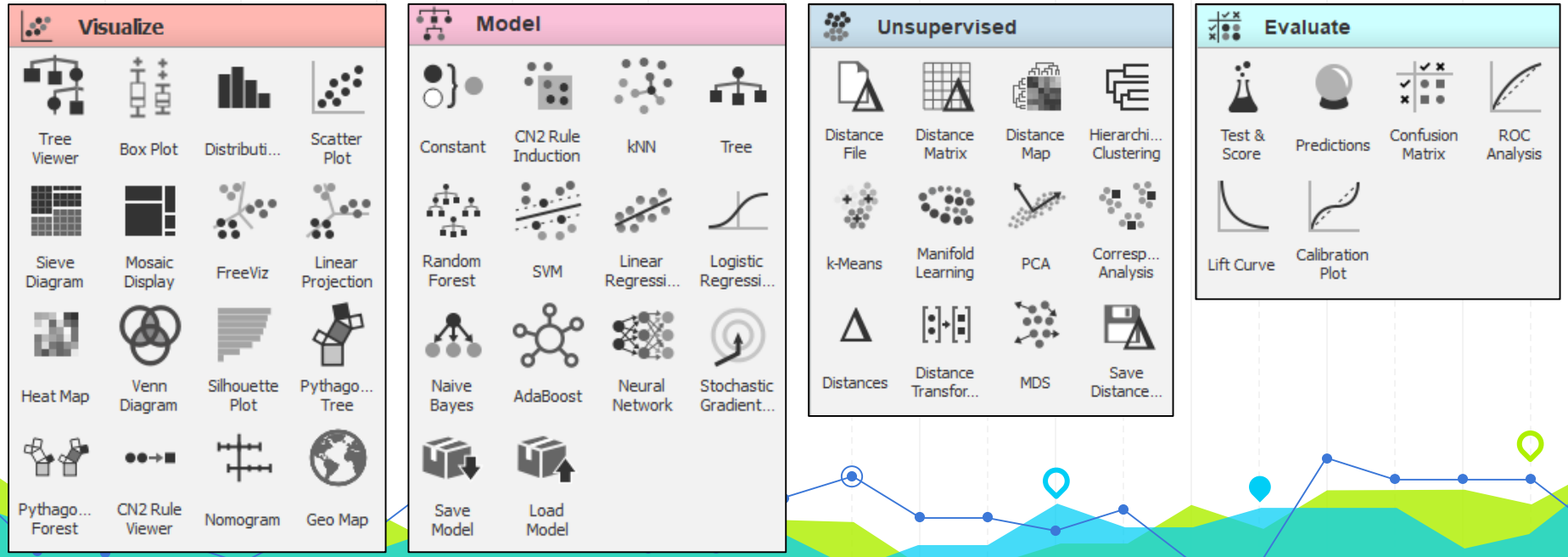
Orange Canvas - GUI

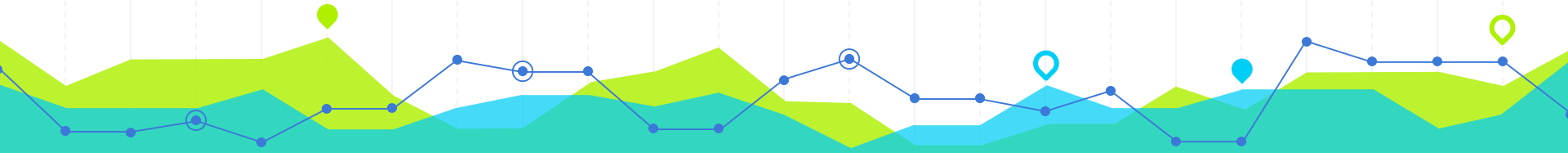


Smart Meter Data Analytics using Orange 3

Orange & Analytics

- Orange provides a rich library of widgets that can be used for visualization, data mining and machine learning.





Use Case: Clustering Energy Usage Patterns 4

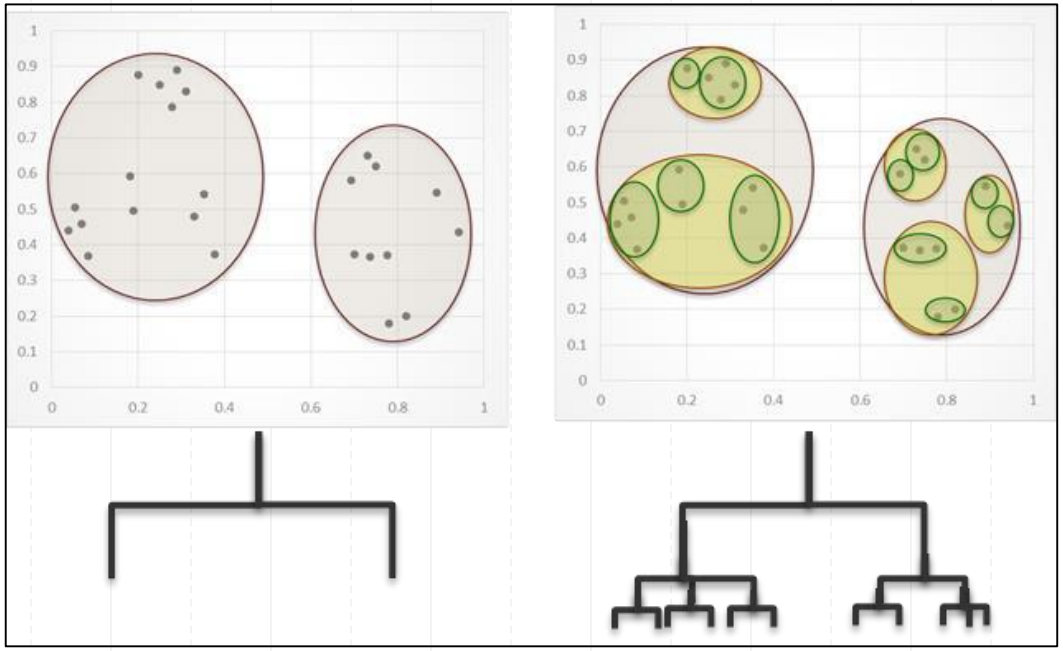
Smart Meters Data

- Time series data of 3,000 smart meters.
- 96 energy consumption values (in kW) captured at every 15-minute interval beginning from 00:15 hrs to 24:00 hrs for a particular day.

Clustering

- Clustering methods are useful in exploring data, identifying anomalies and predicting future outcomes (Arora et al. 2016).
- These statistical techniques help analyze time series data such as those generated by smart-meters.
- K-means clustering of residential load patterns, registered by smart meters, has been reported in some recent studies (Lavin and Klabjan 2015; Arora et al. 2016; Al-Wakeel and Wu 2016; Al-Wakeel and Wu in press) where the households are clustered based on their energy usage patterns to design targeted incentive schemes and design the tariff based on Time of Use.
- The benefit of time-variable electricity tariff in improving pricing and reducing peak loads have been described in some previous studies (Troxel 1938; Eckel 1985; Stephenson et al. 2001; and Parmesano 2007; Flath et al. 2012).
- However, not much has been reported regarding segment specific rate tariff based on the consumer load profile which have certain advantages as the complexity of rates designed at individual customer level is reduced significantly due to the classification of energy usage patterns into groups (Chicco et al. 2003; Flath et al. 2012; Wijaya et al. 2014).

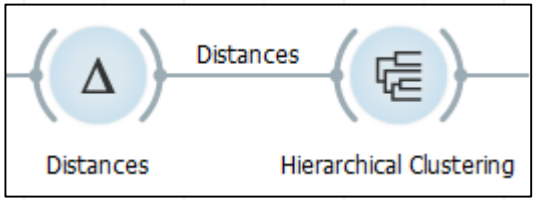
Why k-means fails?



- Traditional K-means clustering (Lloyd 1982) seeks to classify similar observations into distinct clusters.
- Hierarchical clustering shows the relationship between clusters by placing them within a hierarchy through the recursive splitting of data to form a top-down hierarchy of clusters.
- Hierarchical clustering algorithm yields better clusters as compared to the standard k-means algorithm, especially in the presence of a high number of dimensions (Jain and Dubes 1988; Cutting et al. 1992; Larsen and Aone 1999).

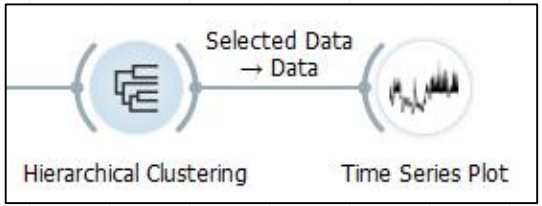
Algorithm

Algorithm - Agglomerative Hierarchical Clustering
Distance Measure – Euclidean
Type of Linkage – Complete Link



Result Visualization

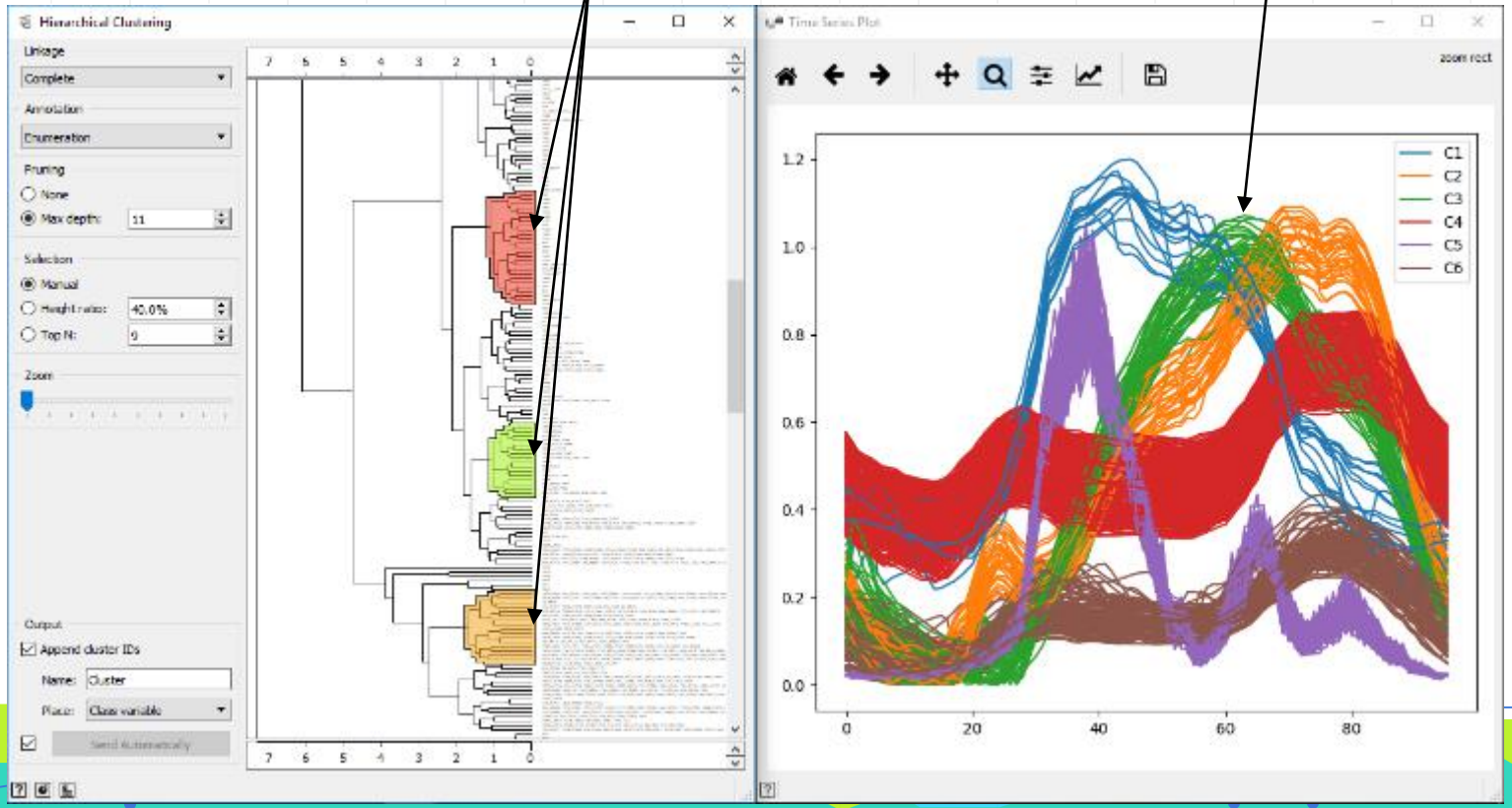
A new widget is developed which utilizes matplotlib to plot time series data in real-time for clusters selected in the hierarchical clustering widget.



Results

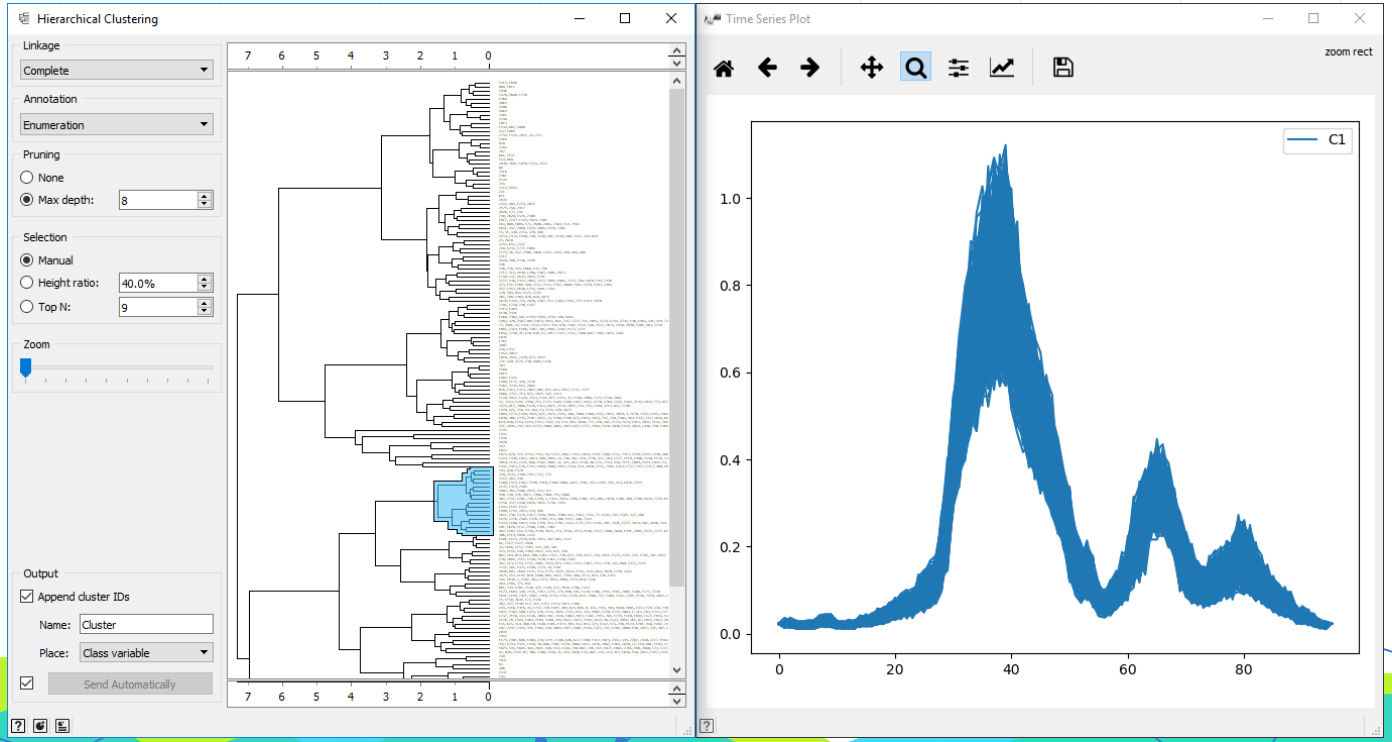
3 of 6 selected clusters

A new cluster is dynamically drawn upon selection



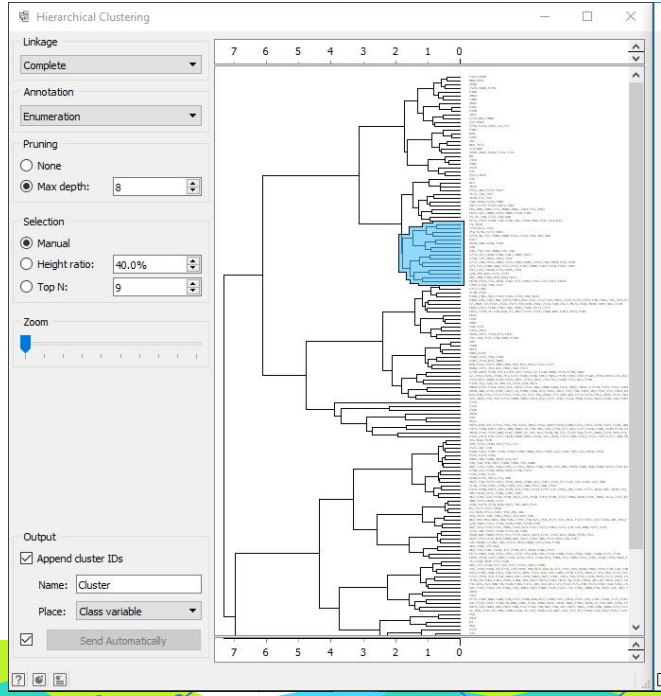
Results (Figures)

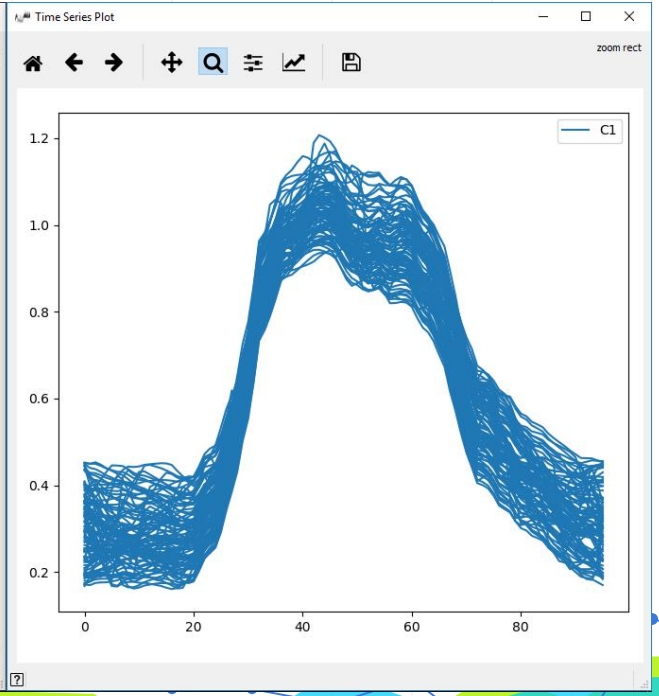
Terminal cluster demonstrates households having similar energy usage patterns



Results (Figures)

Terminal cluster demonstrates households having similar energy usage patterns





CONCLUSION

Orange
Open-source python framework which provides a rich library of widgets that can be used for visualization, data mining and machine learning.

Growing volume of sensor data
Growing usage of smart meters leads to the generation of large volume of sensor data which can be analyzed to identify the various energy consumption patterns useful for managing the smart grid.

Use Case
Energy usage patterns are analyzed using the hierarchical clustering algorithm.

Benefit of Analytics
Identified consumer clusters enable the utility companies in preparing segment-specific campaigns and tariffs to distribute the peak energy loads. This also helps them in meeting the electricity supply-demand more efficiently, thereby limiting energy loss and preventing grid over-load in future.

THANKS!

Any questions?