

An open-source, data-science toolkit for power and data engineers

August 3 2022, by Stephen Wampler



Credit: CC0 Public Domain

As of 2020, 102.9 million smart meters—devices that record and communicate electric consumption, voltage and current to consumers and grid operators—have been installed in the United States.

As the number of smart meters and the demand for energy is expected to increase by 50% by 2050, so will the amount of data those smart meters produce.

While [energy standards](#) have enabled large-scale data collection and storage, maximizing this data to mitigate costs and [consumer demand](#) has been an ongoing focus of energy research.

To help make the most of all this data, a Lawrence Livermore National Laboratory (LLNL) team has developed GridDS—an open-source, data-science toolkit for power and data engineers that will provide an integrated energy data storage and augmentation infrastructure, as well as a flexible and comprehensive set of state-of-the-art machine-learning models.

"Until now, no open-source platforms have provided data integration or machine learning models. The few existing platforms have been proprietary and not available to the broader research community," said principal investigator and data scientist Indra Chakraborty at the Laboratory's Center for Applied Scientific Computing (CASC). "As an [open-source](#) toolkit, GridDS opens the door to data and power scientists everywhere who are working on these challenges and want to make the most of this data."

By providing an integrative software platform to train and validate machine learning models, GridDS will help improve the efficiency of distributed [energy resources](#), such as [smart meters](#), batteries and solar photovoltaic units.

GridDS also is designed to leverage advanced metering infrastructure, outage management systems data, supervisory control data acquisition and [geographic information systems](#) to forecast energy demands and detect incipient grid failures.

GridDS features a modular, generalizable Python software library for these multiple streams of data. In adapting to disparate datasets recorded by various devices, GridDS provides a range of unique functionalities not presently implemented in current advanced distribution management systems, which tend to have highly specific software infrastructure by design.

"Previous experiments have demonstrated that when it comes to applying the best machine learning model for a given energy problem, one shoe does not fit all. Each scenario is different, and context is key," said Vaibhav Donde, associate program lead for Energy Infrastructure Modernization.

"We have found that researchers are better off trying several approaches to see what works best. With GridDS, you can make small tweaks to task designs, such as horizon or history in an autoregression, or carry over machine learning models between datasets, which enables learning transfer and broader model validation. GridDS can take general approaches, apply them to highly specific energy tasks and evaluate and validate their performance," Donde added.

GridDS also can rapidly and efficiently test several approaches to [energy](#) and sensor time-series problems and train model hyperparameters.

GridDS is now available via [Github](#).

Provided by Lawrence Livermore National Laboratory

Citation: An open-source, data-science toolkit for power and data engineers (2022, August 3) retrieved 23 April 2024 from <https://techxplore.com/news/2022-08-open-source-data-science-toolkit-power.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.