Opening the Black Box: Explainable AI for Greater Energy Saving

Biodun Iwayemi PhD., Jeffrey Hullstrung, Nick Neverisky, Dan Fredman PhD.

About VEIC

- Nonprofit founded in 1986 with a mission **to generate the energy solutions the world needs –** focus is **reducing GHGs and improving energy equity**
- **National consulting practice** advising states, utilities, Federal agencies, nonprofit organizations, and private industry
- Program design & implementation for **energy efficiency and clean energy programs** serving customers, including program administrator for Efficiency Vermont & the DC Sustainable Energy Utility; on administration team for California TECH, Hawaii Energy, and Focus on Energy (Wisconsin)
- **Clean transportation** program design, evaluation, technical analysis/assistance, data collection/analysis, education, fleet studies, and implementation for LD/MD/HD Electric Vehicles

Now


THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Podcasts | Latest Headlines

Home World U.S. Politics E... ...y Business Tech Markets Opinion Books & Arts Real Estate Life & Work WSJ. M

Why S... ...e Is Eating The World

...Marc Andreessen

August 20, 2011

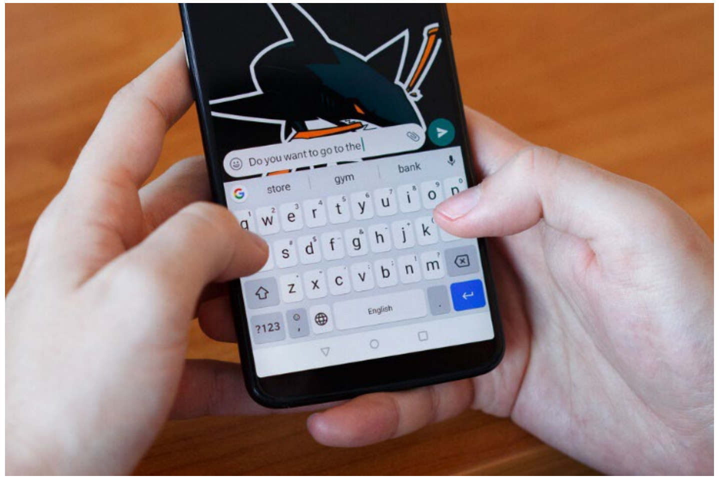
PRINT TEXT

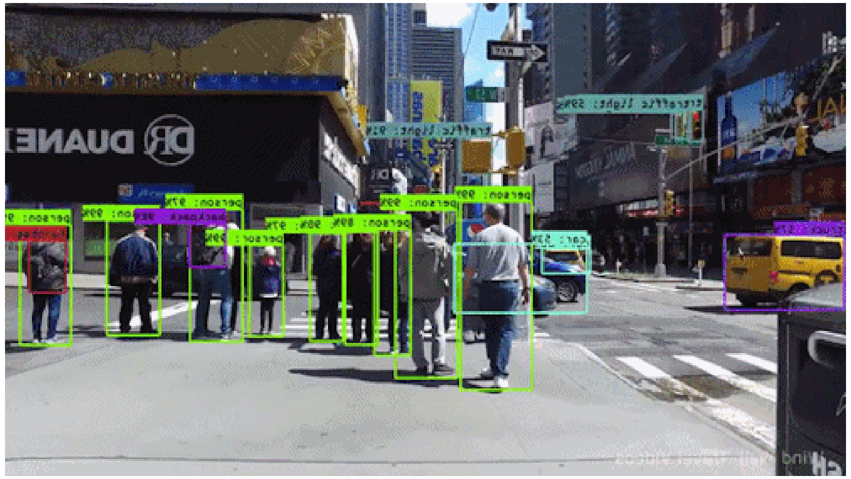
This week, Hewlett-Packard (where I am on the board) announced that it is exploring

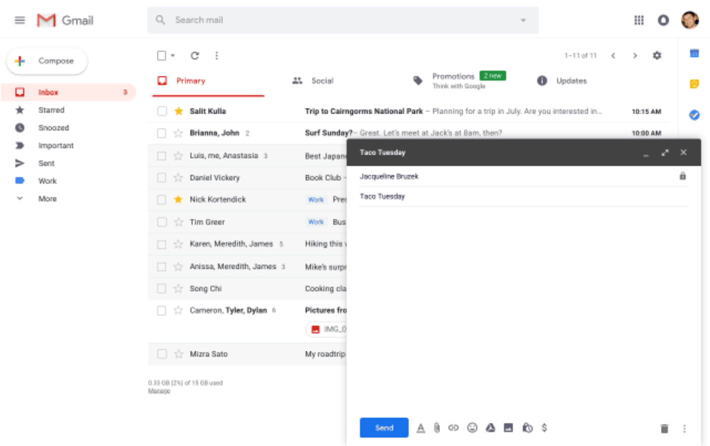
MOST PO

1. Torn Kent With Fear

Machine learning in everyday life







Voice assistants
https://www.etsy.com/market/change_my_diaper

Predictive text and email

Computer vision: Object detection and tracking
https://pythonawesome.com/yolo-rcnn-object-detection-and-multi-object-tracking

Definitions

Artificial Intelligence: the ability of computers and robots to exhibit behavior that mimics human abilities like reasoning and problem-solving.

Machine learning: *"[A] set of methods that computers use to make and improve predictions or behaviors based on data"* – Christoph Molnar

Evolution of Machine Learning for Energy Modeling

1986

Linear Regression (PRISM)

Model parameters: daily temperature, monthly billing date, heating degree days relative to building balance point

2011

Time of week and temperature (TOWT)

Model parameters: hour-of-week, outdoor temperature, type of day (regular, holiday)

2018-date

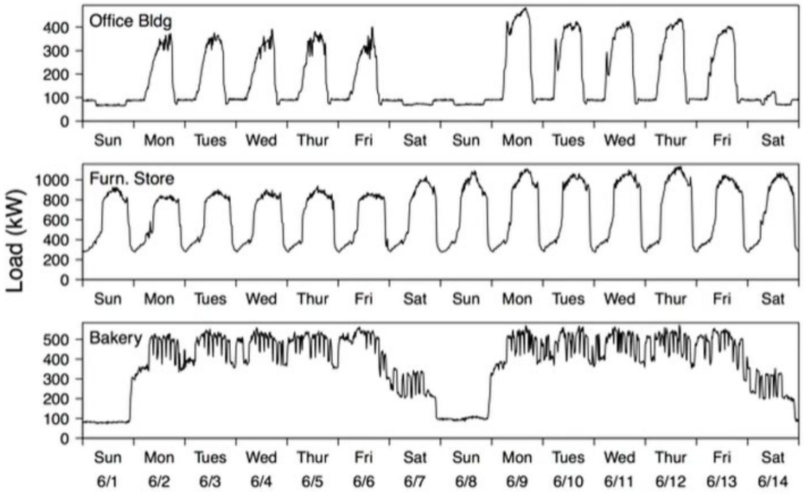
Gradient Boosting Machines / Deep Neural networks

Model parameters (time-of-day, day-of-week, month, outdoor air temperature, outdoor relative humidity, type of day)

Time of Week and Temperature (TOWT) model

Time of week and temperature (TOWT) modeling*:

- Published by LBNL team in 2011, refined and tested
- One of the CALTRACK evaluation methods
- Performs a continuous, piecewise temperature regression with 5 segments
- Introduced hour-of-week as a key variable



*Quantifying Changes in Building Electricity Use, with Application to Demand Response"; Johanna L. Mathieu, Phillip N. Price, Sila Kiliccote, Mary Ann Piette; Environmental Energy Technologies Division; April 2011; Submitted to IEEE Transactions on Smart Grid

Time of Week and Temperature (TOWT) model

Model inputs

- Temperature (outdoor dry bulb)

Powerful model but a regression model with 175 parameters is not easy to digest or explain

- Tim
  - Hour of week (categorical, 168 possibilities)
- Day Type
  - Regular or Holiday (categorical, 2 possibilities)

Gradient Boosting Machines

Overview

- Ensemble models that combine many “weak”/mediocre models to create a powerful and accurate model
- Each weak model learns from it’s predecessors mistakes
- Excellent performance on tabular data and able to capture non-linear relationships
- We used LightGBM*, Microsoft’s implementation of GBM’s

LightGBM: https://github.com/Microsoft/LightGBM

Gradient Boosting Machine

Model inputs:

- Temperature (outdoor dry bulb)
- Relative humidity (outdoor)
- H
- D
- M
- Re

Also a powerful model but no closed form equation explaining the model

Modeling 8 Grocery Stores

Image: efficiencyvermont.com

- 8 grocery stores in Vermont
- Average size: 50,000 sq. ft
- Annual consumption: 14,000 MWh across all stores
- Previously analyzed for large multi-measure projects using our standard weather normalization process
- LBNL Time-of-week Temperature (TOWT) regression provided good to excellent results

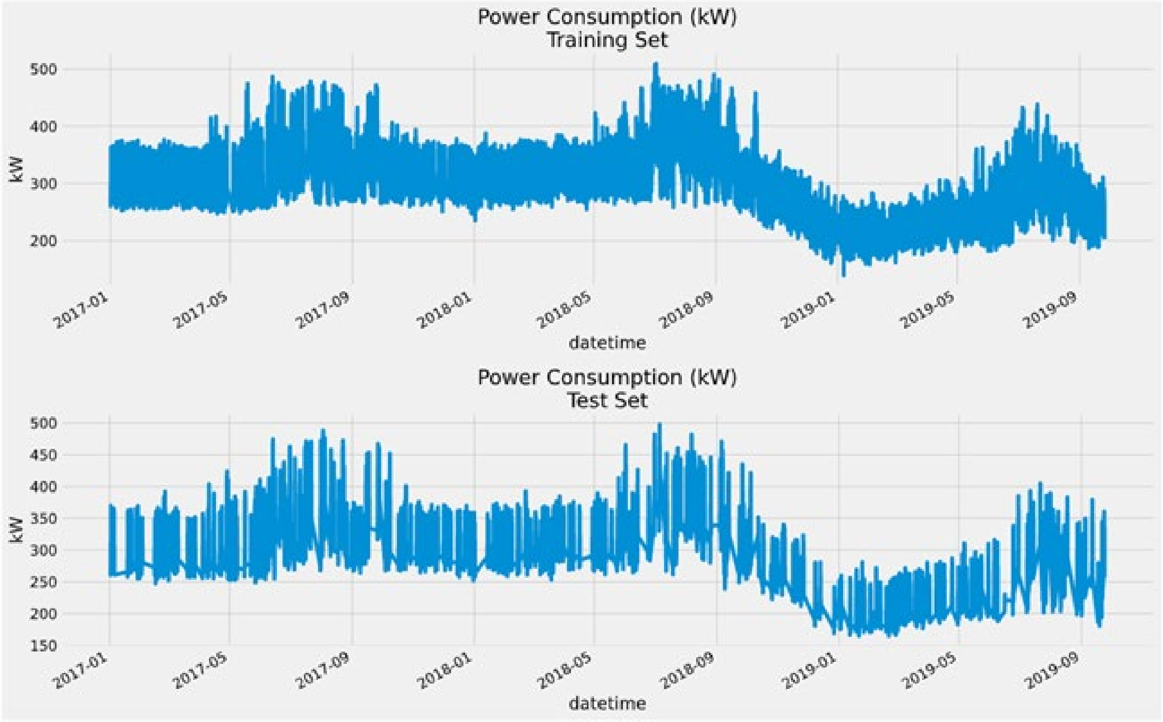
Model development and evaluation

Data preparation

- 70% training data / 30% test data
- Used SKLearnstratified split to balance data from each season
- Used daily “chunks” of data to test model with unseen days

Evaluation metrics

- R^2
- CV-RMSE



Training and Test sets

Results: LightGBM vs TOWT

parameter	value	value	pct diff
model_name	Grocery1_lgbm	Grocery1_towt	
cvrmse_train	2.9%	4.3%	
rsquared_train	94.7%	88.1%	
cvrmse_test	3.8%	4.8%	21%
rsquared_test	91.9%	87.0%	6%

parameter	value	value	pct diff
model_name	Grocery2_lgbm	Grocery2_towt	
cvrmse_train	3.6%	7.7%	
rsquared_train	97.3%		
cvrmse_test	5.3%		
rsquared_test	94.7%		

parameter	value		
model_name	Grocery3_lgbm		
cvrmse_train	3.7%		
rsquared_train	92.4%		
cvrmse_test	4.9%	6.1%	19%
rsquared_test	88.3%	82.0%	8%

parameter	value	value	pct diff
model_name	Grocery4_lgbm	Grocery4_towt	
cvrmse_train	3.3%	4.6%	
rsquared_train	96.3%	92.7%	
cvrmse_test	4.9%	5.6%	13%
rsquared_test	92.2%	89.8%	3%

Store by store comparison (locations 1 – 4)

Overall results

Average reduction in CV-RMSE test	15.0%
Average increase in R-squared test	3.7%

parameter	value	value	pct diff
model_name	Grocery5_lgbm	Grocery5_towt	
cvrmse_train	3.0%	4.0%	
rsquared_train	95.3%	91.8%	
cvrmse_test	4.1%	4.2%	4%
rsquared_test	91.8%	91.1%	1%

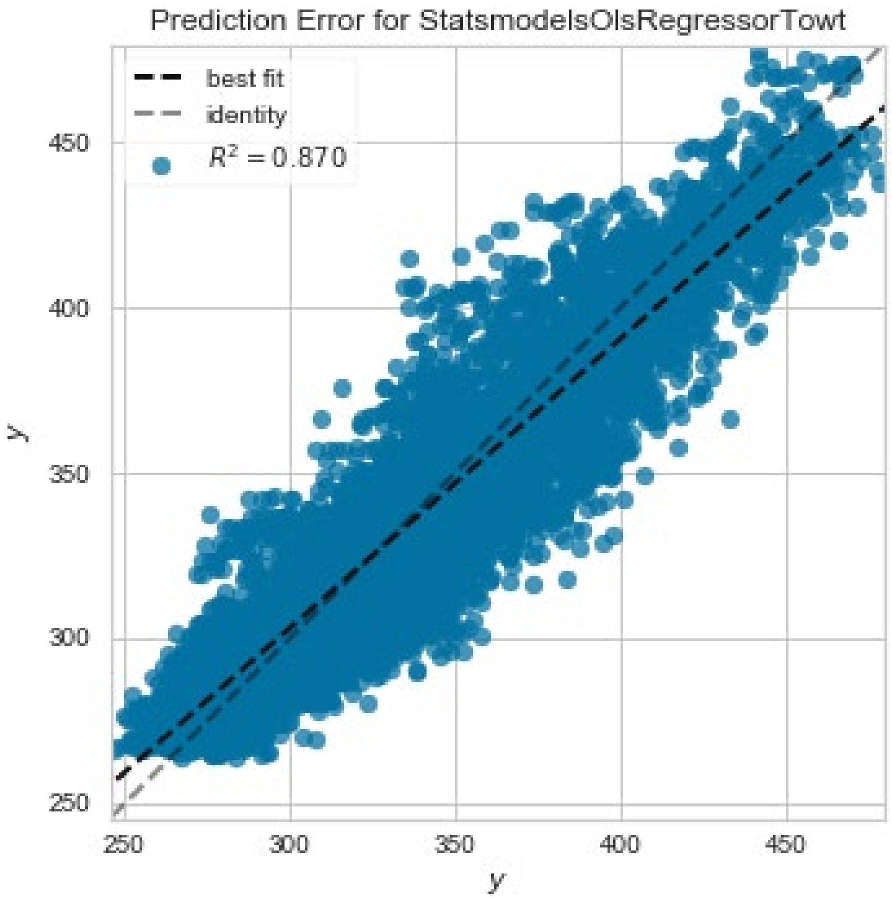
parameter	value	value	pct diff
model_name	Grocery6_lgbm	Grocery6_towt	
cvrmse_train	3.2%	4.0%	
		95.2%	
		4.7%	6%
		93.6%	1%

		value	pct diff
		rocery7_towt	
		3.3%	
		92.9%	
cvrmse_test	3.4%	3.6%	7%
rsquared_test	93.4%	92.3%	1%

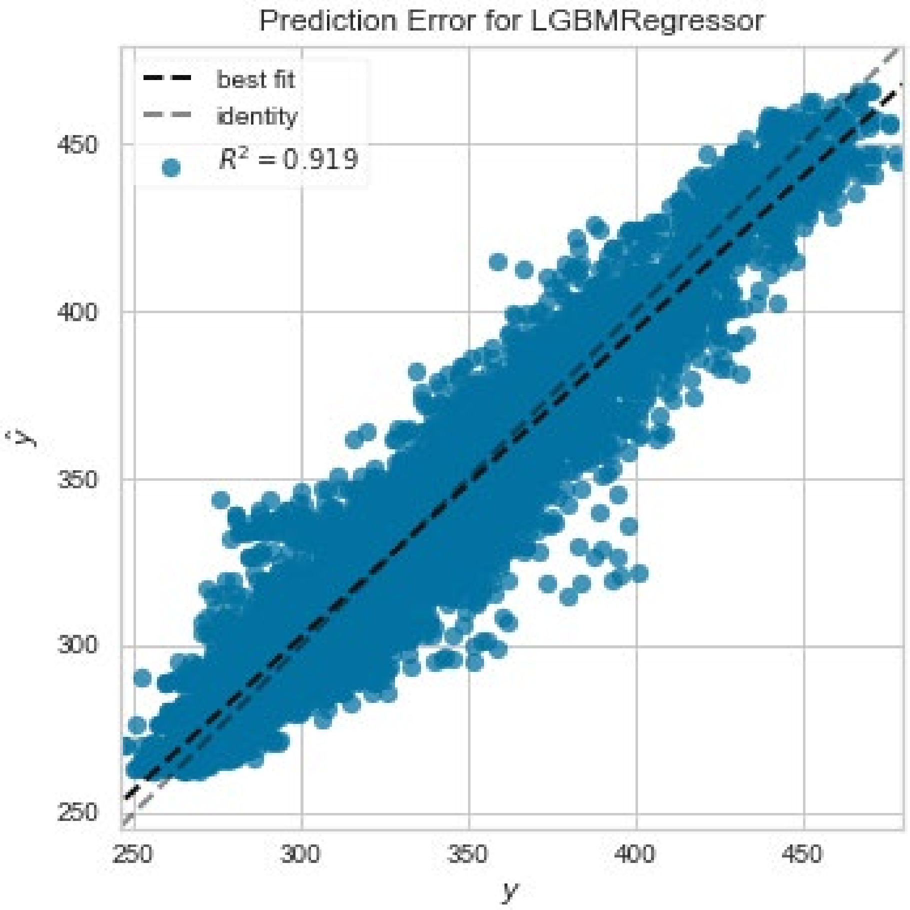
parameter	value	value	pct diff
model_name	Grocery8_lgbm	Grocery8_towt	
cvrmse_train	2.6%	3.7%	
rsquared_train	97.6%	94.9%	
cvrmse_test	3.5%	4.0%	12%
rsquared_test	95.7%	94.4%	1%

Store by store comparison (locations 5 – 8)

Model diagnostic charts: Grocery store #1 prediction

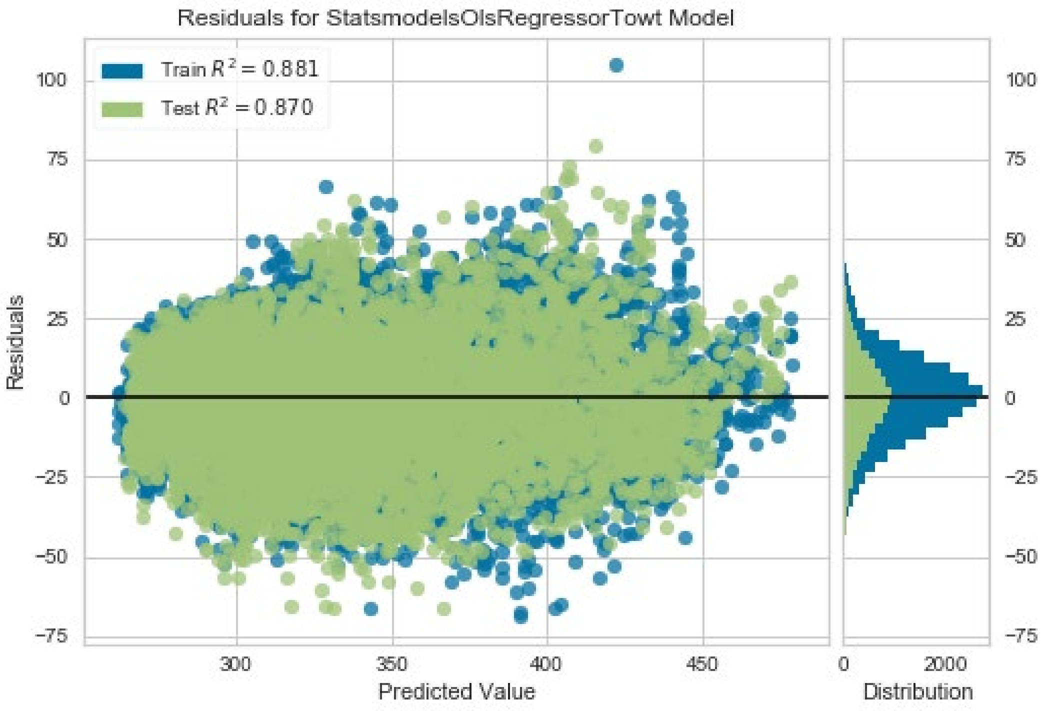


TOWT Prediction Error

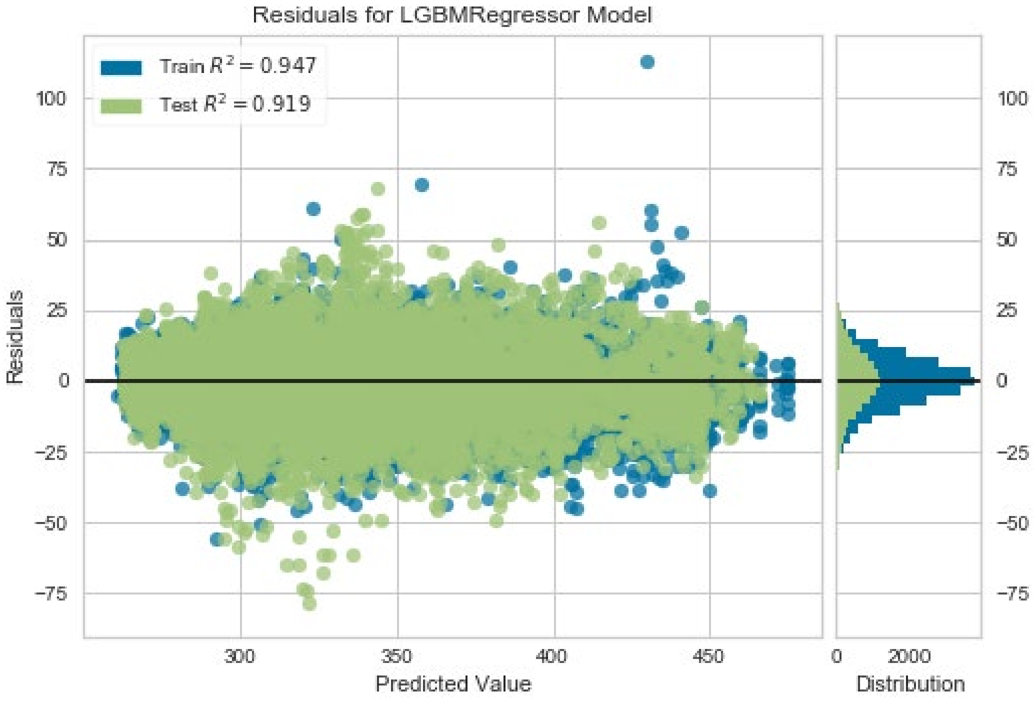


LightGBM Prediction Error

Model diagnostic charts: Grocery store #1 prediction



TOWT Residuals and distribution



LightGBM Residuals and distribution

Great, we know we have good models, but.....

- What are the drivers of energy usage in these facilities?
- Does the model make sense?
- How do we explain the model results to our customers?

The case against advanced regression models

veic

Opacity

Efficiency Vermont

www.efficiencyvermont.com
888-921-5990 | 802-860-4095

Performance Period Update Report

[redacted]

model: Power (kW) = 166 - 0.04*kWh_gen

Reporting Period Start Date	2019-09-01
Reporting Period End Date	2020-07-01
Number of points	305
Average power actual (kW)	108
Predicted usage during period, energy (kWh)	862,155
Actual usage during period, energy (kWh)	789,893
Savings during period, energy (kWh)	72,262
Savings uncertainty (kWh)	+/-14,334
Savings fraction	8.4%
Uncertainty percentage points	+/-1.7%
Fractional savings uncertainty (target<50%)	19.8%

EVT Customer Report

More flexible (accurate models) are usually less interpretable

Advanced ML Models like GBM's don't provide a regression equation

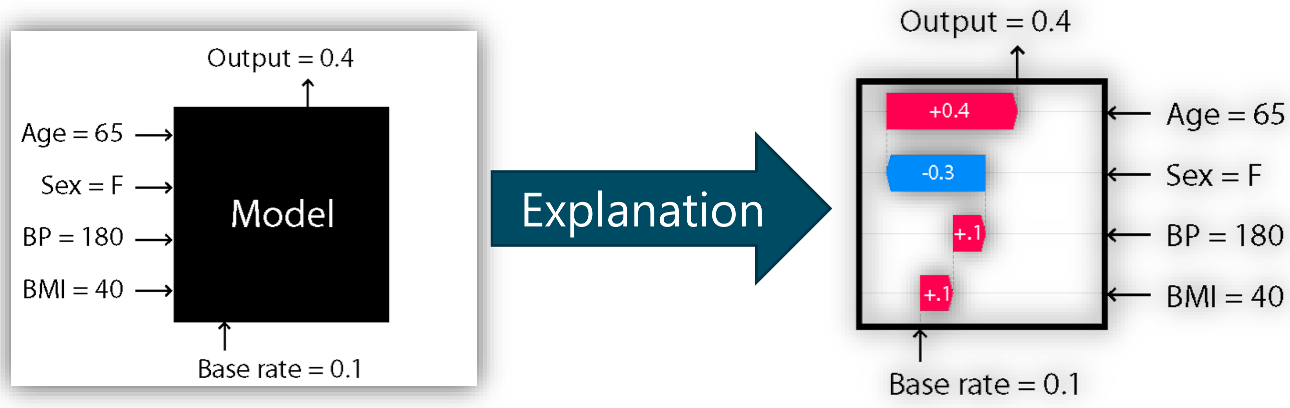
Black box models

Not ideal for:

- Highly regulated environments where **transparency** is critical
- When you want to understand **why**

Explainable AI (XAI) for explaining machine model outputs

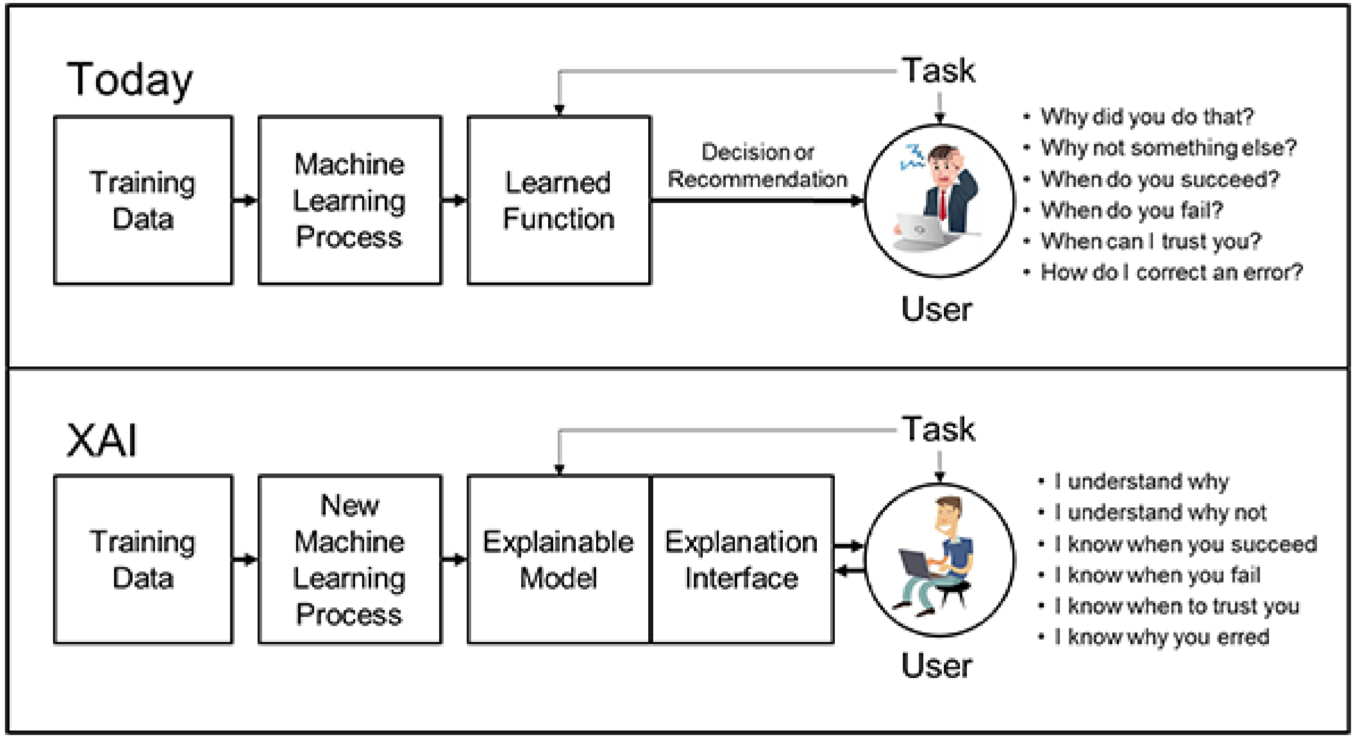
SHAP



veic

[GitHub - slundberg/shap: A game theoretic approach to explain the output of any machine learning model.]

eXplainable AI (XAI)



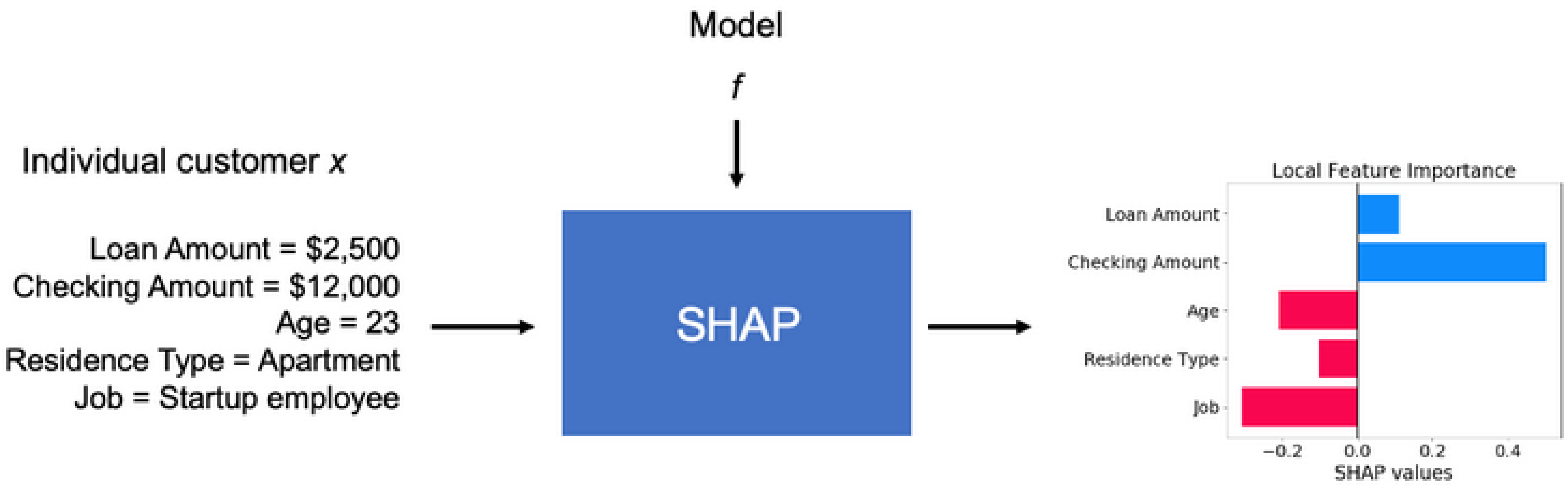
https://www.darpa.mil/program/explainable-artificial-intelligence

Explaining Machine Learning Models using SHAP

- SHapley Additive exPlanation (SHAP) provides a way to fairly attribute contributions from multiple participants
- Based on Nobel prize winning work in cooperative game theory by Lloyd Shapely
- Post-hoc explanation that can be applied to a variety of machine learning models

Explaining Machine Learning Models using SHAP

Why was your loan application approved/denied?



Attributing impacts to model inputs

https://iancovert.com/blog/understanding-shap-sage/

XAI Visualizations from SHAP

Local models

- Provide model explanations for a snapshot in time

Global models

- Provide model explanations over multiple readings over an extended period of time (e.g. a quarter, year etc.)

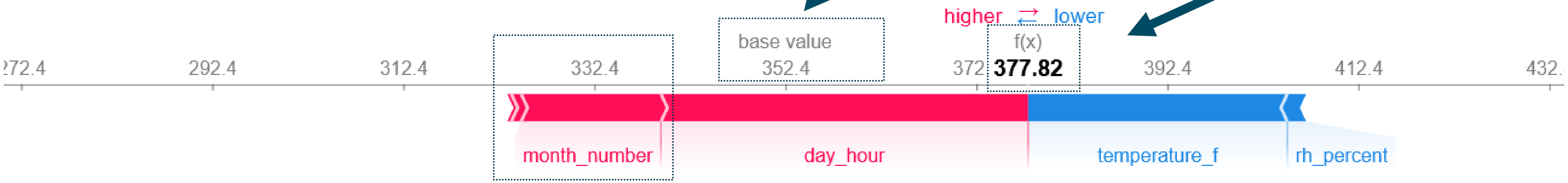
XAI Visualizations from SHAP

Local model explanation

- A snapshot in time

Average power value across entire dataset

Predicted value for this instance



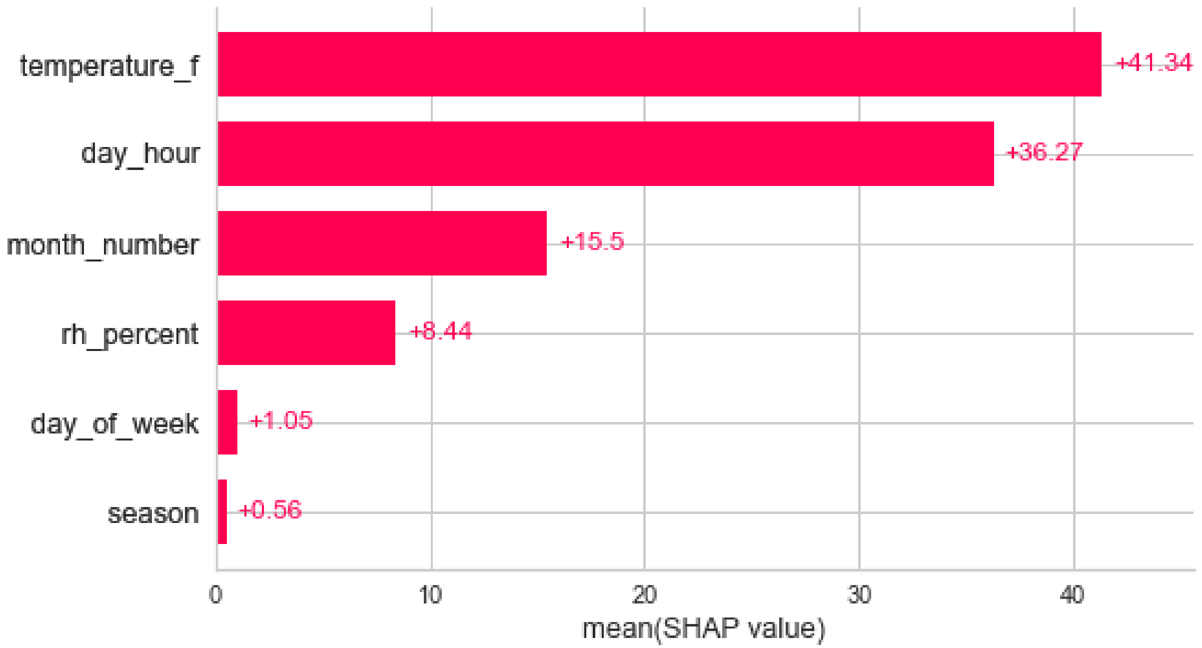
Contribution of month To predicted value

A SHAP force-plot

XAI Visualizations from SHAP

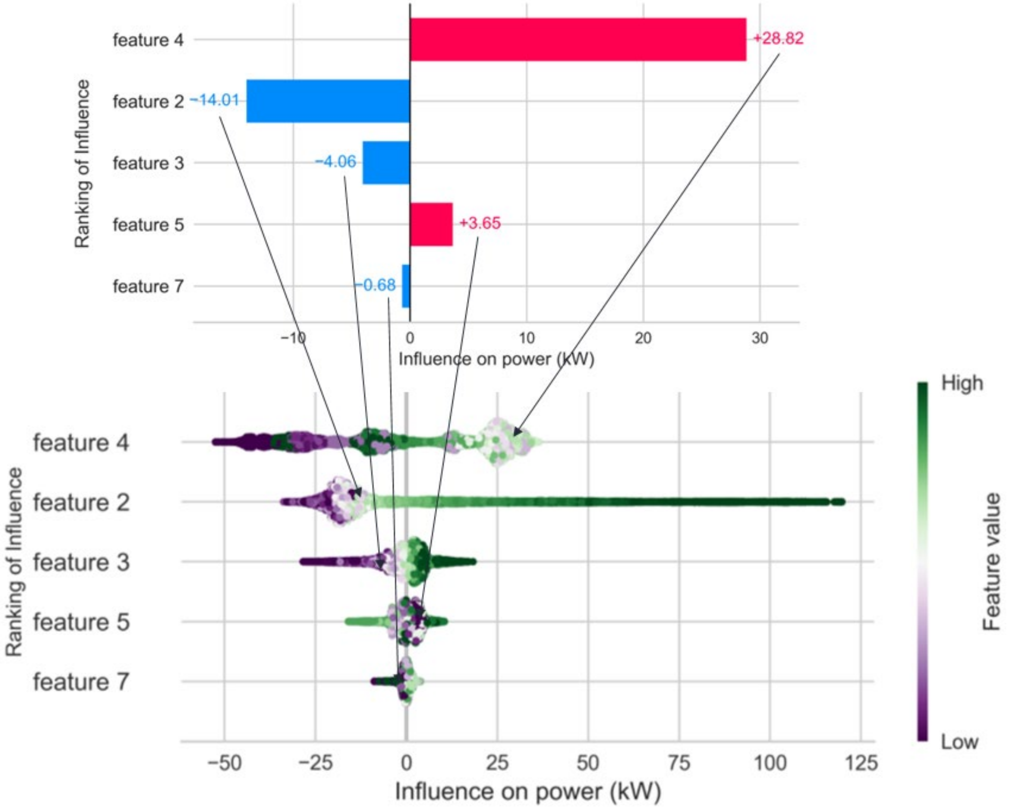
Global model explanations

- Aggregation over time (e.g. a quarter, year etc.)



XAI Visualizations from SHAP

Global model explanations



Mapping from local to global explanations

User testing XAI Visualizations

Photo by David Travis on Unsplash

User testing goal

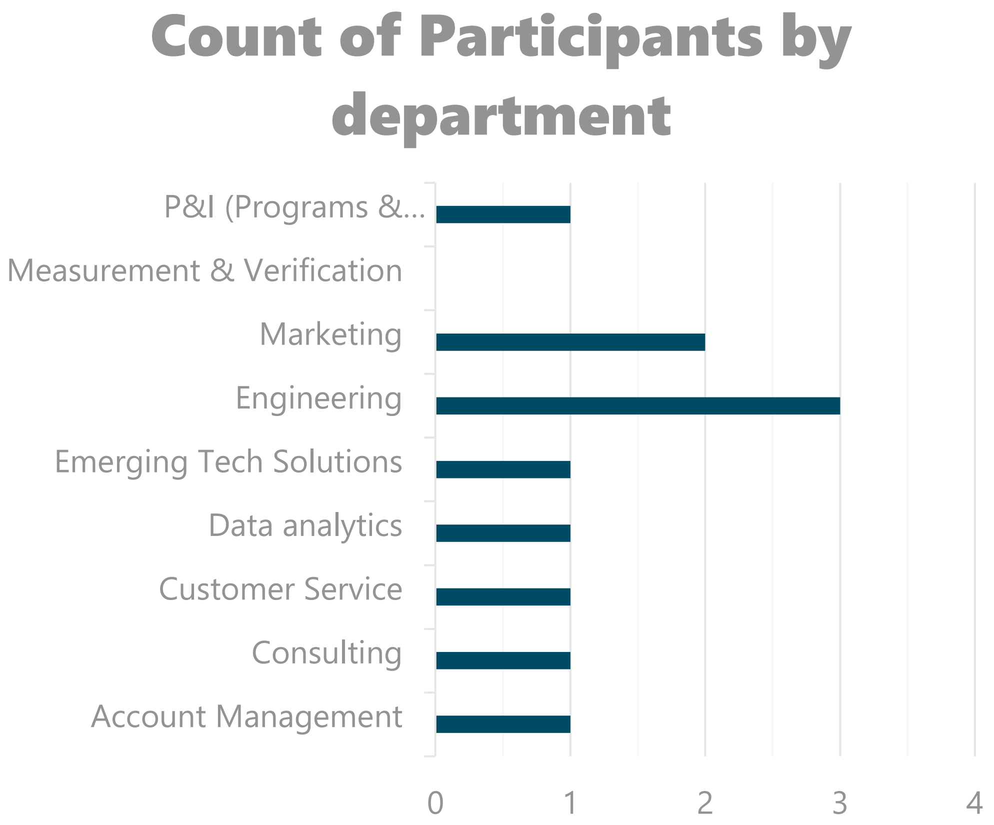
High-level goal

- Identify methods for communicating outputs of machine learning algorithms that model power use to audiences with a range of experience with energy management and modeling

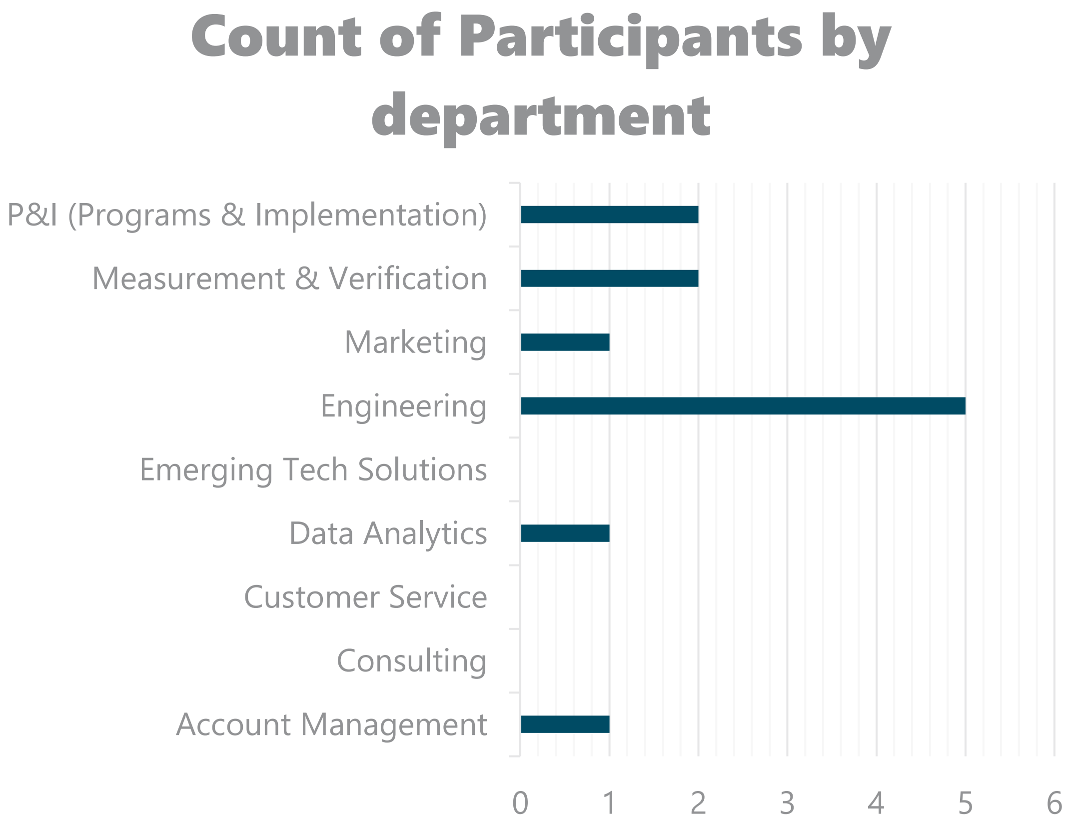
Sub-goals

- Identify chart types that effectively communicate drivers of power use
- Identify chart elements that aid in communication
- Identify chart elements that cause confusion

User Testing Participants



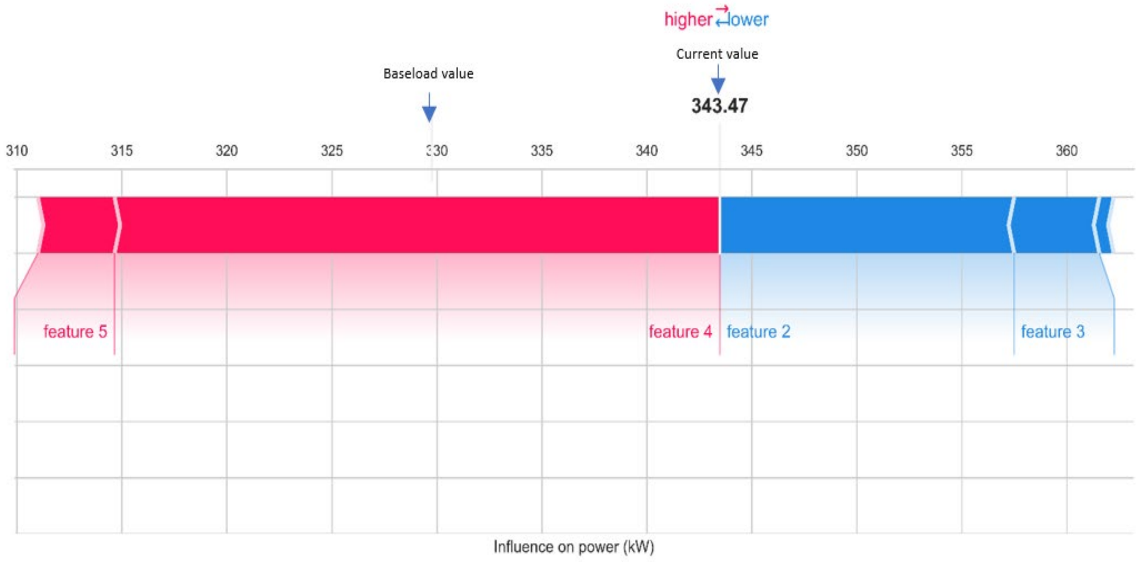
Group 1



Group 2

Force plot (Local model)

- SHAP force plot showing correlations learned from a snapshot of specific model.
- Participants found the force plot the hardest chart to interpret.

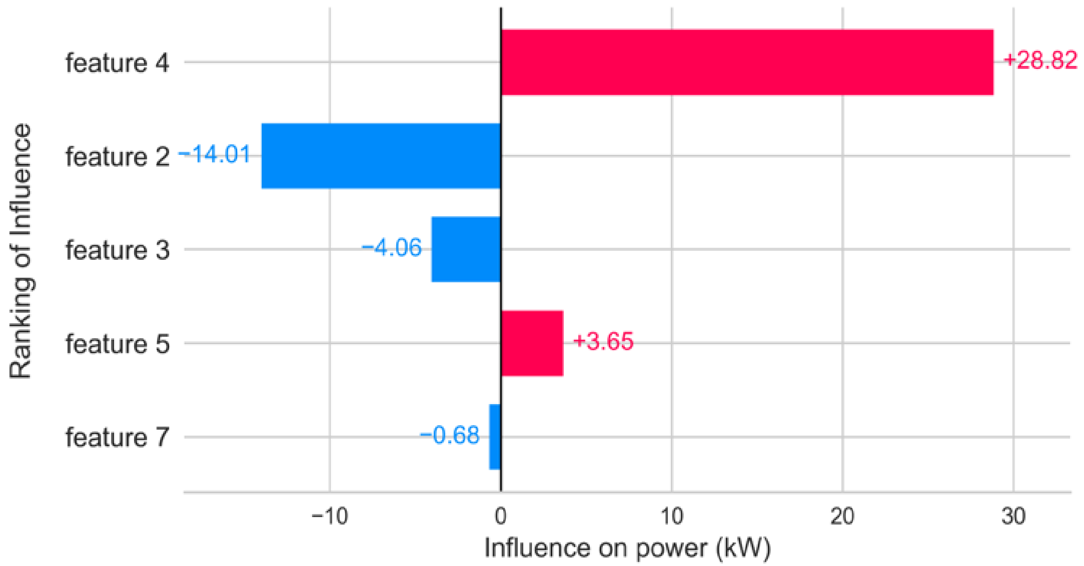


	All participants	Session #1	Session #2
Feature most increasing power use	78%	80%	77%
Feature most decreasing power use	63%	56%	69%
Feature with largest effect on power use	92%	100%	85%
Power use	75%	55%	92%
Average	77%	73%	81%
% indicating difficulty	66%	90%	46%

Values other than "% indicating difficulty" are the percent of respondents who answered correctly

Bar chart (Local model)

- Shows correlations learned from a snapshot of specific model inputs
- Users found the bi-directional bar charts the most intuitive of all the chart types



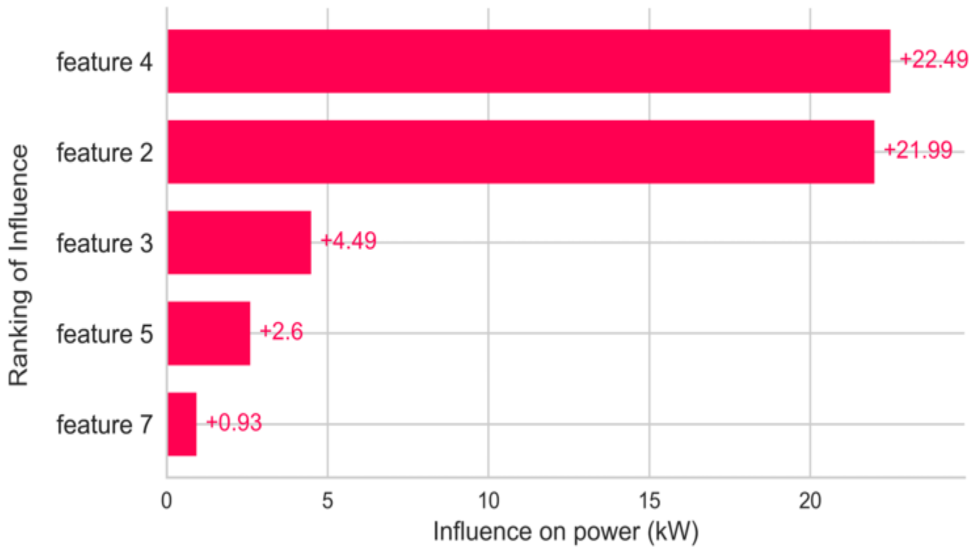
	All participants	Session #1	Session #2
Feature most increasing power use	87%	82%	92%
Feature most decreasing power use	92%	82%	100%
Feature with largest effect	96%	91%	100%
Average	92%	85%	97%
% indicating difficulty	5%	10%	0%

Values other than "% indicating difficulty" are the percent of respondents who answered correctly

Bar chart (Global model)

Shows correlations learned over several months of data:

- Users found this chart easy to interpret
- It also had the highest rate of accurate interpretations across all the charts:



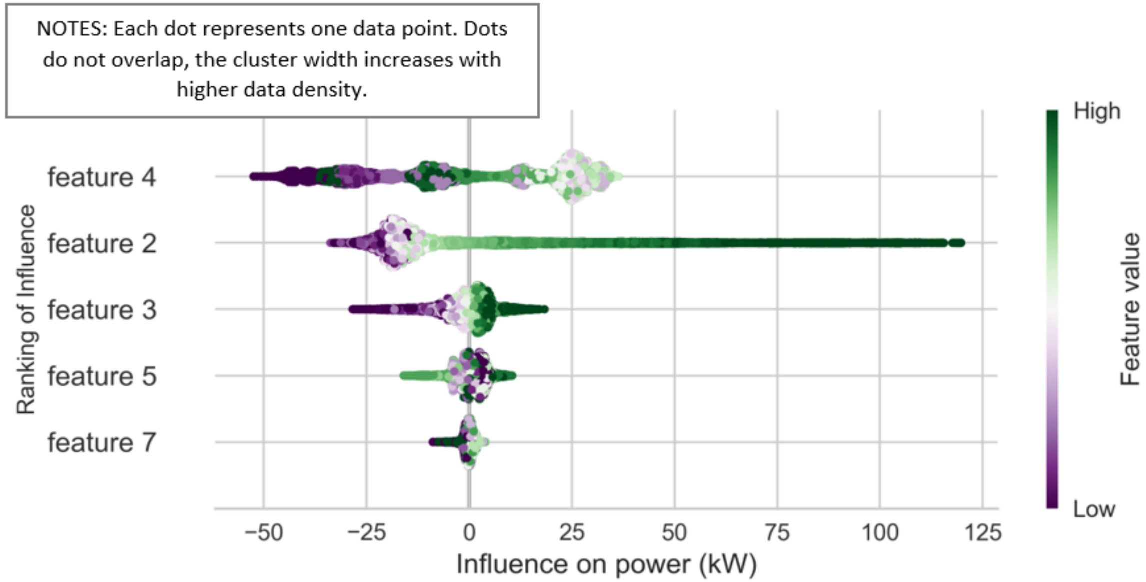
	All participants	Session #1	Session #2
Feature with largest effect on power use	96%	91%	100%
Feature with smallest effect on power use	100%	100%	100%
Average	98%	96%	100%
% indicating difficulty	5%	10%	0%

Values other than "% indicating difficulty" are the percent of respondents who answered correctly

Bee Swarm (Global model)

Most complex visualization, but captured participants' interest the most.

- Feedback was that this chart would provide the most information if users were guided on how to use/interpret it.



	All participants	Session #1	Session #2
Feature with largest potential to affect power use	70%	45%	92%
Feature with smallest potential to affect power use	92%	91%	92%
Identify features with potential to not affect power use			
Average	81%	68%	92%
% indicating difficulty	52%	50%	54%

Values other than "% indicating difficulty" are the percent of respondents who answered correctly

Findings

- The charts most likely to resonate with focus group participants were simple bar charts and the global model bee-swarm chart.
- The highest rate of misinterpretation and reported difficulty were on the local model force plot.
- Participants report that charts like these are **useful representations of power data**

Recommendations/Learnings

- Visually easy-to-understand charts offer good representations of data about power use.
- Clarity of chart labels and design elements (e.g. helper text, color-blind-friendly color schemes) helped making charts easily understood and accessible to a broader audience
- Present power use in dollars terms (rather than energy use in kWh) to relate to operating costs and drive investments

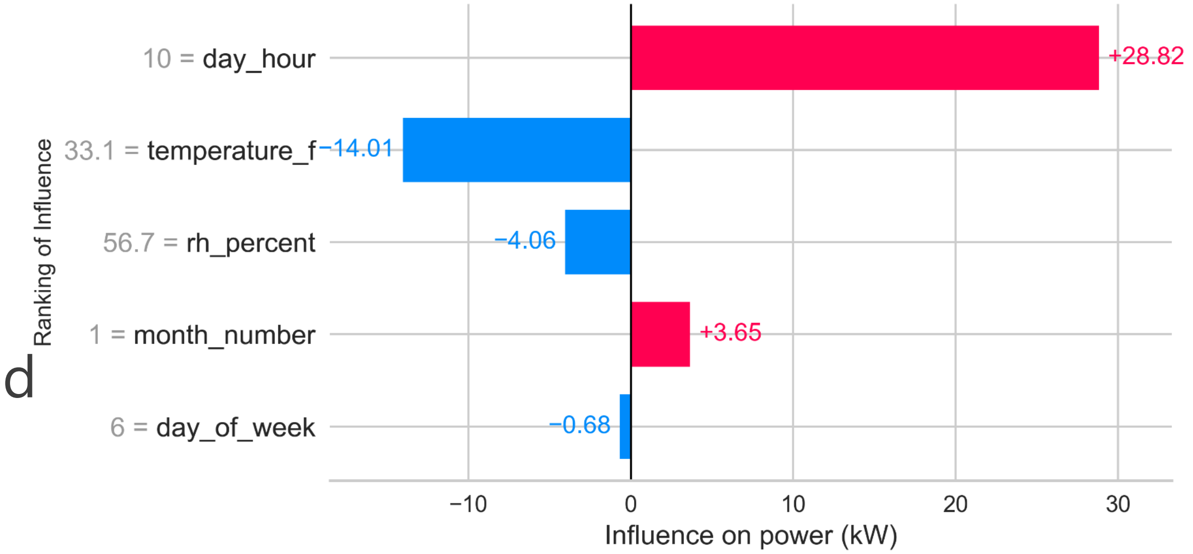
XAI on Grocery store regression models

Image: efficiencyvermont.com

Model Interpretation

Given an average power draw of 330 kW:

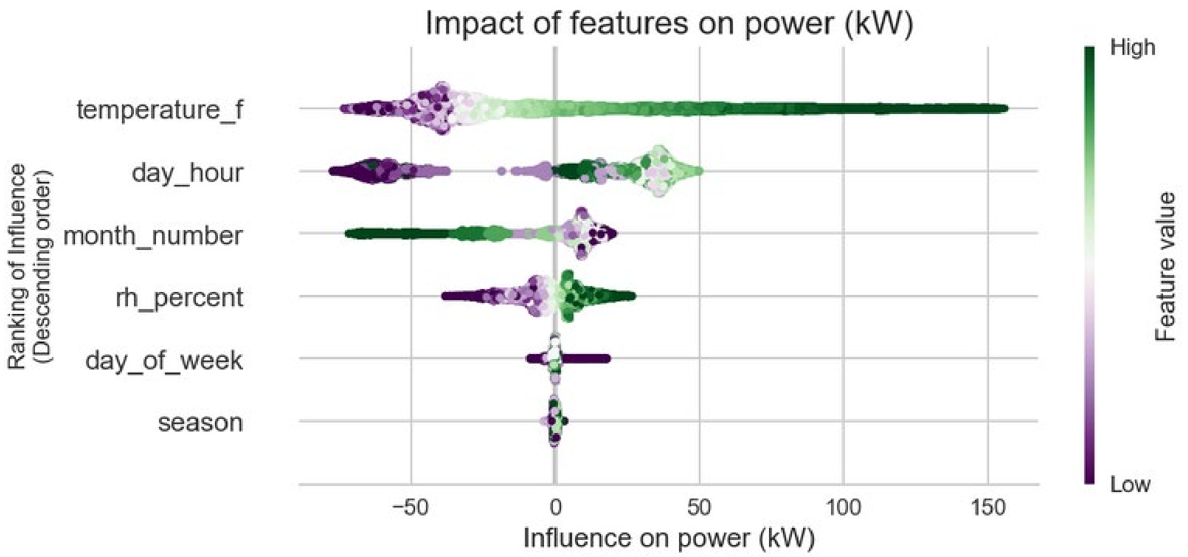
- 10am is correlated with a power increase of 28.82 kW above the 330 kW average
- A temperature of 33.1ºF is correlated with a reduction in power draw of 14 kW
- A relative humidity reading of 56.7% is correlated with a power reduction of 4 kW



Local model for grocery store #1

Global Interpretation

- Low temperatures are correlated with **decreases** in power consumption
- Power draw **decreases** by 50 kW during the early hours of the day (midnight onwards)
- Higher humidity is correlated with power **increases** of 1-30 kW



Global model for Grocery store #2

Predicting is not explaining

- The SHAP plots show only the correlations learned by the model: **Correlation is not causation.**
- The learned models are valid only for relationships learned within the dataset the models have seen.
- These correlations could change as modifications are made to the building, or if building operation patterns change.
- Determining causal relationships between model inputs and the model output requires more rigor, and is best achieved through causal modeling, experiments and causal hypotheses.

Take-aways

- **Determine if you have sufficient data**. These methods require 15-minute or hourly interval data. The team also recommends at least one year of data for modeling, with data that capture all the seasons experienced at the building's location.
- **Benchmark your models.** Modelers should compare the performance of existing modeling strategies with GBRTs. The VEIC team advises switching to GBRTs only if the performance gain on the data is significant.
- **Customize your plots.** The team recommends using simple SHAP plot types, with added helper text and color-blind-friendly colors.
- **Be clear about model limitations.** These are predictive, not causal, models

veic

Questions?