Appendix C. Load and DER forecasting supplemental information

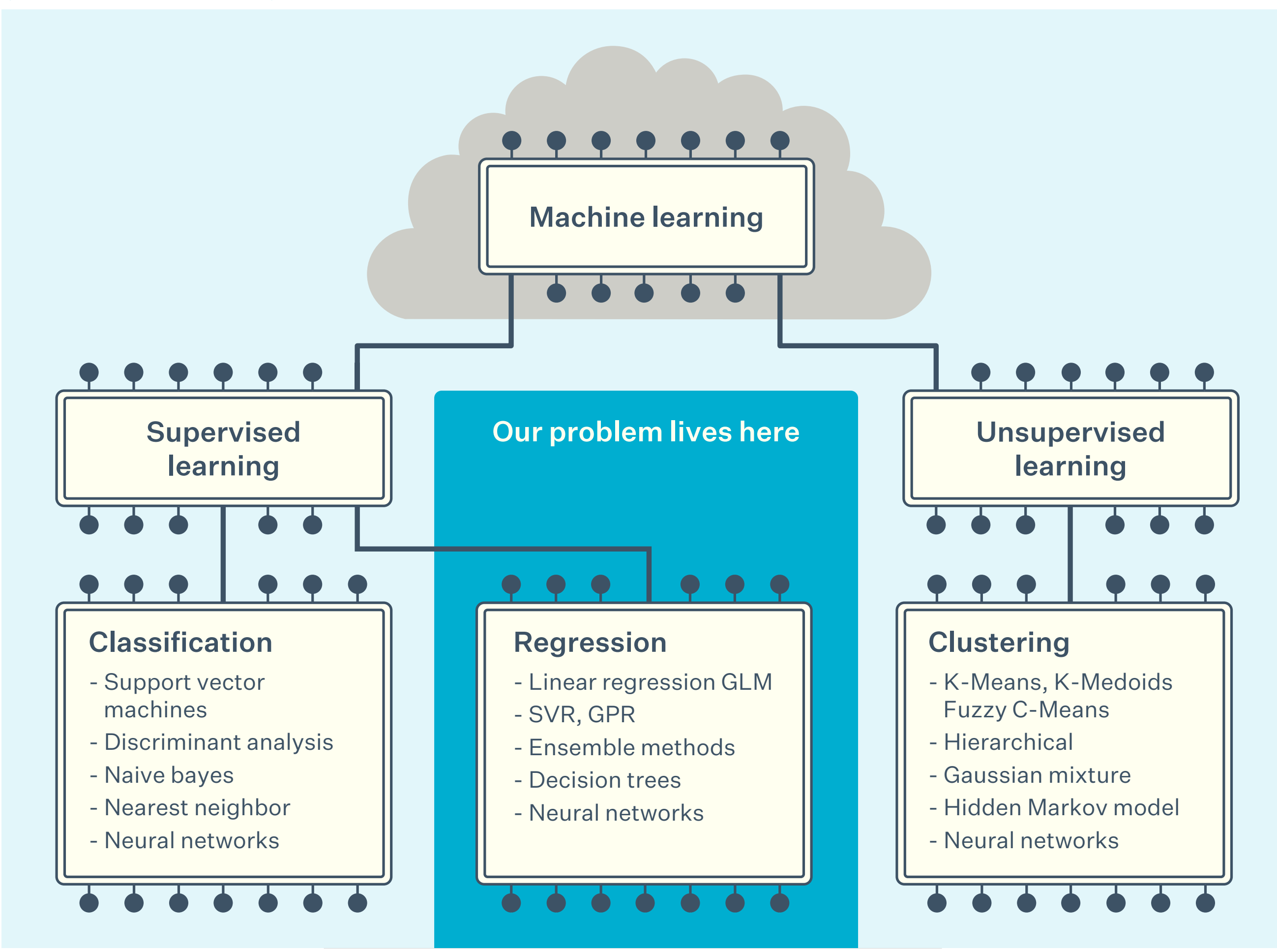
C.1 Statistical model detailed methodology

The overarching goal of the statistical models is to rank-order adoption probability for selected DER measures, not to develop the most sophisticated model. We considered the following requirements when constructing the statistical models:

- Model must be scorable in AdopDER for every customer, measure, and year. This means that considerations of model run time place natural upper limits on the scoring algorithm's complexity.
- We need to be able to use the statistical model to adjust adoption probability for each customer and measure.
- Model must have locational and temporal awareness.

The selected methodology to develop these models was the scorecard model. A scorecard model is a type of regression model, as shown in **Figure 56**.

Figure 56. Machine learning taxonomy



Moreover, the scorecard model fits our selection criteria for model characteristics:

- Predicts a binary outcome (Adopt: yes/no)
- Uses binning for continuous variables
- Able to work around missing data
- Applies transformation to assign score points
- Provides a high degree of transparency, used in financial services
- "Easy" to implement in AdopDER

We used a structured modeling framework for statistical modeling. For all DER types modeled with a statistical modeling approach, we follow the steps in **Figure 56** to:

- Select variables
- Test the strength of the model, and
- Apply to the full population

Figure 57 shows the workflow used for developing each separate model.

For the statistical models, we take all potential candidate variables identified in the literature review that may potentially help explain differences in adoption and then create a training model. We train the model on 70% of past adopters and test different combinations of variables for their ability to "predict" adoption for the remaining 30% of the sample that was withheld from the model training. This method is a commonly applied industry practice called "out of model validation".

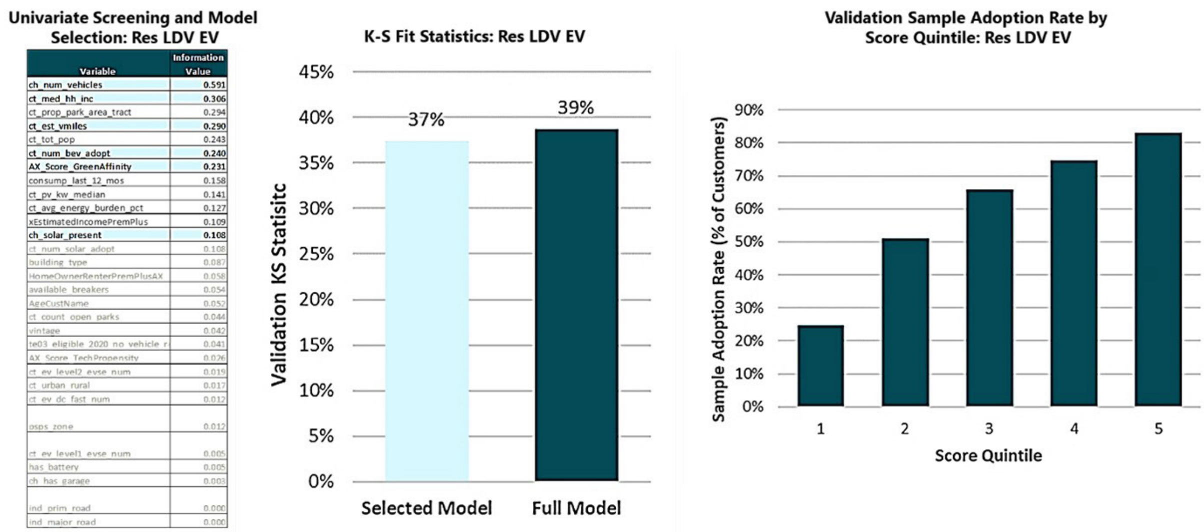
Once we select the candidate variables and develop the final model specification, we conduct one last validation step (KS scoring) before deploying the model into AdopDER to disaggregate the DER adoption into locational granularity. At the end of this process, we have a process to feed into AdopDER and develop site-level adoption estimates for each year, and these are then aggregated up to the feeder or substation level for reporting purposes.

Figure 57. Structured modeling framework for statistical models

Acquire data	Combine	Sample	Train	Validate	Deploy
• Identify candidate variables	• Join all candidate variables into single dataset	• Train model on 70% of premises • Validate model on 30% of premises	• Variable selection • Model specification	• Test rank order with validation KS	• Add scorecard to AdopDER • Dynamic scoring in AdopDER

The full variable list, specification results, and resulting EV LDV adoption propensity quintile rankings are shown in **Figure 58**. The selected model is shown as the model with blue-shaded variables in the univariate screening table, while the full model includes all variables that pass the univariate screen. Variables that were considered but had weak correlation (i.e., did not pass univariate screening) are shown in gray text.

Figure 58. Residential LDV adoption — model creation process



Univariate Screening and Model Selection: Res LDV EV

Variable	Information Value
ch_num_vehicles	0.591
ct_med_hh_inc	0.306
ct_prop_park_area_tract	0.294
ct_est_vmiles	0.290
ct_tot_pop	0.243
ct_num_bev_adopt	0.240
AX_Score_GreenAffinity	0.231
consump_last_12_mos	0.158
ct_pv_kw_median	0.141
ct_avg_energy_burden_pct	0.127
xEstimatedIncomePremPlus	0.109
ch_solar_present	0.108
ct_num_solar_adopt	0.108
building_type	0.087
HomeOwnerRenterPremPlusAX	0.058
available_breakers	0.054
AgeCustName	0.052
ct_count_open_parks	0.044
vintage	0.042
te03_eligible_2020_no_vehicle_r	0.041
AX_Score_TechPropensity	0.026
ct_ev_level2_evse_num	0.019
ct_urban_rural	0.017
ct_ev_dc_fast_num	0.012
osps_zone	0.012
ct_ev_level1_evse_num	0.005
has_battery	0.005
ch_has_garage	0.003
ind_prim_road	0.000
ind_major_road	0.000

The relative contribution that each of the final variables has on increasing or decreasing the adoption propensity away from the overall average is shown in the scorecard. **Table 51** shows how the selected model variables were binned and what their score was. Note that a score higher than zero means higher adoption probability compared to the baseline, whereas a score less than zero means lower adoption probability compared to the baseline adoption.

Table 51. Residential LDV EV adoption scorecard

Variable	Bin	Score Points
base points	NA	552
ch_num_vehicles	[-Inf,2)	-213
ch_num_vehicles	[2,3)	-17
ch_num_vehicles	[3,4)	84
ch_num_vehicles	[4, Inf)	115
ct_med_hh_inc	missing	3
ct_med_hh_inc	[-Inf,50000)	-4
ct_med_hh_inc	[50000,75000)	-1
ct_med_hh_inc	[75000,90000)	1
ct_med_hh_inc	[90000, Inf)	5
ct_est_vmiles	missing	-25
ct_est_vmiles	[-Inf,30)	19
ct_est_vmiles	[30,37)	7
ct_est_vmiles	[37,49)	-7
ct_est_vmiles	[49,52)	24
ct_est_vmiles	[52, Inf)	-17
ct_num_bev_adopt	missing	44
ct_num_bev_adopt	[-Inf,40)	-84
ct_num_bev_adopt	[40,90)	-23
ct_num_bev_adopt	[90,130)	10
ct_num_bev_adopt	[130,310)	39
ct_num_bev_adopt	[310, Inf)	103
xEstimatedIncomePremPlus	missing	-5
xEstimatedIncomePremPlus	$100,000 - $124,999	8
xEstimatedIncomePremPlus	$15,000 - $19,999%,%$20,000 -	-13
xEstimatedIncomePremPlus	$40,000 - $49,999%,%$50,000 -	-3
xEstimatedIncomePremPlus	$75,000 - $99,999	5
xEstimatedIncomePremPlus	Greater than $124,999	13
xEstimatedIncomePremPlus	Less than $15,000%,%Unknown	-19
AX_Score_GreenAffinity	missing	-16
AX_Score_GreenAffinity	[-Inf,3)	-74
AX_Score_GreenAffinity	[3,5)	-31
AX_Score_GreenAffinity	[5,6)	7
AX_Score_GreenAffinity	[6, Inf)	56

Figure 59 and **Table 52** show the same model selection process and scorecard results for the residential Solar PV model.

Figure 59. Residential solar PV — model creation process

Univariate Screening and Model Selection: Res Solar

Variable	Information Value
building_type	0.788
ct_med_hh_inc	0.637
ct_num_solar_adopt	0.554
ct_tot_pop	0.492
HomeOwnerRenterPremPlusAX	0.438
ct_num_bev_adopt	0.365
xEstimatedIncomePremPlus	0.327
ch_num_vehicles	0.302
AgeCustName	0.256
AX_Score_GreenAffinity	0.242
consump_last_12_mos	0.240
ct_pv_kw_median	0.231
vintage	0.176
AX_Score_TechPropensity	0.084
has_battery	0.058
ct_avg_energy_burden_pct	0.040
ct_urban_rural	0.014
psps_zone	0.011

Selected Variable
Fails Univariate Screen

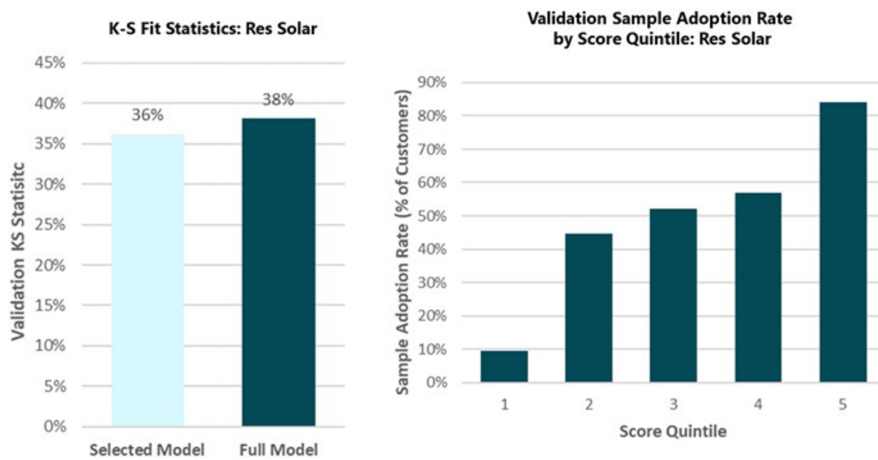


Table 52. Residential solar PV — adoption scorecard

Variable	Bin	Score Points
base points		493
building_type	MF	-325
	MH%,%SF	31
ct_med_hh_inc	missing	-17
	[-Inf,40000)	-26
	[40000,50000)	-13
	[50000,65000)	-2
	[65000, Inf)	7
ct_num_solar_adopt	missing	-80
	[-Inf,10)	-169
	[10,20)	-64
	[20,25)	-25
	[25,75)	22
	[75, Inf)	95
HomeOwnerRenterPremPlusAX	missing	-97
	O	34
	R	-112

C.2 Heuristic model detailed methodology

For the heuristic models, variables and weighting assignments were developed based on a combination of literature review and subject area expert judgment by Cadeo and Brattle.

The single largest driver for residential storage adoption probability is whether or not a customer resides in a public safety power shutoff (PSPS) zone. Following that, there are high adoption probabilities for customers with solar, those residing in single-family dwellings, and/or those with high household incomes.

Table 53 shows the variables considered and the relative "points" used to score their impact on raising or lowering adoption propensity.

Table 53. Residential behind-the-meter energy storage scorecard

Variable	Value	Approx % of customers	Points
Baseline		100%	500
In PSPS Zone	Yes	2%	300
	No	98%	-30
Presence of Solar	Yes	2%	200
	No	98%	-20
Type of Building	SF	73%	50
	MF or MH	27%	-100
Household Income	0-40k	18%	-50
	40k-100k	40%	0
	100k+	34%	50
	Unknown or Missing	19%	-10

For non-residential storage, we sorted non-residential premises according to the type of business (using North American Industry Classification System or NAICS codes), their "green score," and their load factor based on analysis of customer load profiles. Similar to residential storage, being located in a PSPS zone drives the highest adoption probability. Otherwise, high probability tends to reflect customers with a high load factor, such as manufacturing and health care customers. The NAICS classification and ranking we used aligns with recent CA Self-Generation Incentive Program (SGIP) reported data. **Table 54**, **Table 55** and **Table 56** show the scorecard development process for non-residential storage, including the categorization and contribution of the principal components.

Table 54. Non-residential behind-the-meter energy storage scorecard

Variable	Value	Approx % of customers	Points
Base		100%	500
In PSPS Zone	Yes	2%	300
	No	98%	-30
Type of Business	Tier A	10%	250
	Tier B	20%	100
	Tier C	15%	0
	Tier D	55%	-100
Green Score	5	40%	80
	3 or 4	45%	0
	0 or 2	5%	-50
	Unknown	10%	-20
Load Factor	LF > 0.25	20%	0
	0 < LF <= 0.25	20%	50
	Missing	60%	-50

Table 55. Non-residential behind-the-meter storage NAICS groupings

Tier	NAICS2	Count
A	Manufacturing	5229
A	Health Care	5688
B	Education	2727
B	Professional Services	7097
B	Public Admin	4159
B	Accommodation and Food	6613
B	Retail	13160
B	Transport and Warehouse	2169
B	Wholesale	3915
C	Admin and Waste Management	2199
C	Ag Forestry Fishing	3017
C	Arts Entertainment Recreation	2540
C	Corp Management	470
C	Financial	2428
C	Mining and Extraction	53
D	Construction	5804
D	Information	1924
D	Other Services	8520
D	Real Estate	14730
D	Unknown	22085
D	(blank)	1092
D	Utilities	216

Table 56. Non-residential behind-the-meter energy storage profile by score quintile

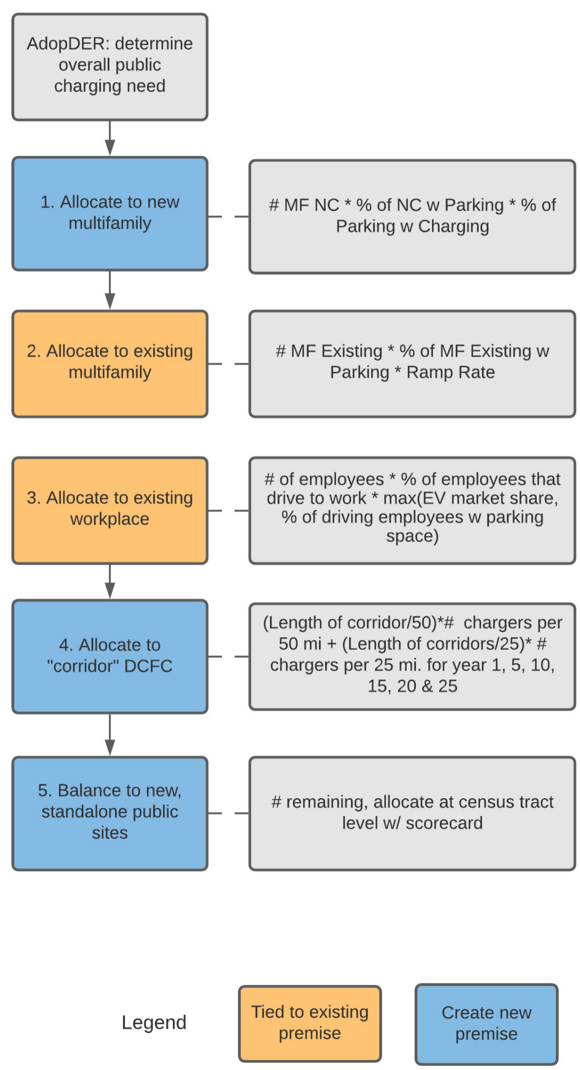
Metric	Quintile 1 (Lowest Score)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (Highest Score)
Min Score	270	320	420	520	620
Max Score	320	420	520	620	1180
% in PSPS Zone	0%	0%	0%	0%	15%
% Tier A NAICS*	0%	0%	0%	1%	47%
% Tier B NAICS*	0%	0%	33%	68%	27%
% w/ Green Score = 5	0%	54%	47%	48%	35%
% w/ Green Score < 3	5%	1%	11%	7%	2%
% w/ Load Factor > 0.25	0%	14%	21%	25%	43%
% w/ Load Factor between 0 and 0.25	2%	21%	30%	27%	26%
% w/ missing Load Factor (i.e. no demand charge	98%	65%	49%	49%	31%

For public EV charging needs,[80] AdopDER determines public EVSE need based on EV adoption and on-site EVSE adoption. We allocate public EVSE based on premise and census-tract level data within AdopDER by considering the following factors:

- Presence of multifamily buildings
- Workplace charging requirements
- Corridor DCFC needs
- Equity considerations

Figure 60 shows the heuristic allocation process by which we assign public charging needs in AdopDER. The overall public charging need is an output of the Phase I DER forecast and is accounts for the amount of unmet total charging energy across all vehicles and across segments.[81]. Both AdopDER and TEINA use NREL's EVI-Pro Lite tool in order to determine EV charging needs, but AdopDER is considering both private charging and public charging needs. Therefore, the TEINA study is a helpful benchmark, but is by itself insufficient for understanding the overall charging need of our customers.

Figure 60. Non-residential public charging process flow



80. "Public" = any EV charging not directly tied to the premise of a customer that has adopted an EV.
81. For a discussion of how AdopDER determines the overall public charging need, see chapter 4 of PGE DER and Flex Load Potential Study – Phase I Report, submitted as Appendix G to the DSP Part I filing and available at: https://portlandgeneral.com/about/who-we-are-planning/distribution-system-planning

The balance of new, standalone EV charging sites (Step 5 in **Figure 60**) are then allocated by census tract using criteria shown in **Table 57**. We differentiate between census tract median income levels to reflect the greater need of public charging infrastructure in areas where there may not be high accessibility for home charging, either because of higher multi-unit dwellings or no presence of garage/driveway for single-family sites. The greater need for public charging in these areas can help inform program design efforts aimed at improving equity of access to EV charging infrastructure.

Table 57. Non-residential standalone public charging scorecard

Variable	Value	Points
Base		500
Census tract Median Income	> 85,000	-100
	45,000 to 85,000	0
	< 45,000	100
NAICS Tier 1	> 100 premises in (Public Admin, Accom+Food, Arts+Ent, Retail)	50
	> 20 premises in (Public Admin, Accom+Food, Arts+Ent, Retail)	0
	Otherwise	-50
Unmet charging	> 1000 vehicles	25
	200 to 1000	0
	< 200 vehicles	-25

C.3 Detailed energy efficiency locational methodology

Given the nature of energy efficiency programs, the Proportional Allocation Method was recommended by the California Working Group on Distribution-level DER forecasting. This method consists of three steps:

1. Using the service territory EE forecast
2. Allocating to circuits based on allocation factors (calculated as ratio of sector-level energy or peak at the individual circuit-level to the overall sector energy or peak)
3. Making adjustments to this allocation to account for local information, such as large known projects

PGE hopes to continue working with Energy Trust to refine the method used in this initial DSP and better account for specific program and measure offerings included in the long-run Energy Trust forecast and how they align with geographic and customer characteristics, and past adoption of EE measures. PGE sees potential for greater planning integration along the following general areas:

- More refined modeling of new construction code impacts within Energy Trust's New Homes residential program. Currently, PGE provides a system-wide forecast of residential customer additions based on Population Estimates from PSU's Population Research Center that inform Energy Trust's long-run potential assessment for above-code energy savings. PGE sees potential to allocate these residential new construction savings forecasts into more granular elements by developing shared assumptions of location-specific population growth estimates, impact of local reach codes, and market knowledge of builder practices and customer demand preferences.
- Greater coordination on impact of low-to-moderate income programs on changes to measure adoption rates. Income is a key variable for our solar PV statistical model and is likely an important indicator of relative adoption for more expensive energy efficiency retrofits like shell upgrades (windows and insulation), HVAC and water heating equipment upgrades, and other higher cost measures. Although past Energy Trust studies have shown that more impactful measures do tend to be clustered among higher income groups, there is potential to improve the equitable adoption of these measures by continued refinement of LMI program offerings and combination with other potential funding sources (e.g., Portland Clean Energy Fund, low-income weatherization funds, and federal infrastructure bill dollars).
- Identify commercial and industrial EE potential by key market segments and drivers

Historically, the linkage between PGE's load forecast for business customers and Energy Trust's EE forecast for commercial and industrial programs has been difficult to align. The current method of allocating by proportion of annual kWh deliveries by revenue class and substation does not account for the relative measure mix included in Energy Trust's forecast as it applies to building- and equipment-level baselines. In future iterations, identifying how the EE potential differs by market sub-segment could potentially allow greater insights about locational impacts of EE on the distribution grid.