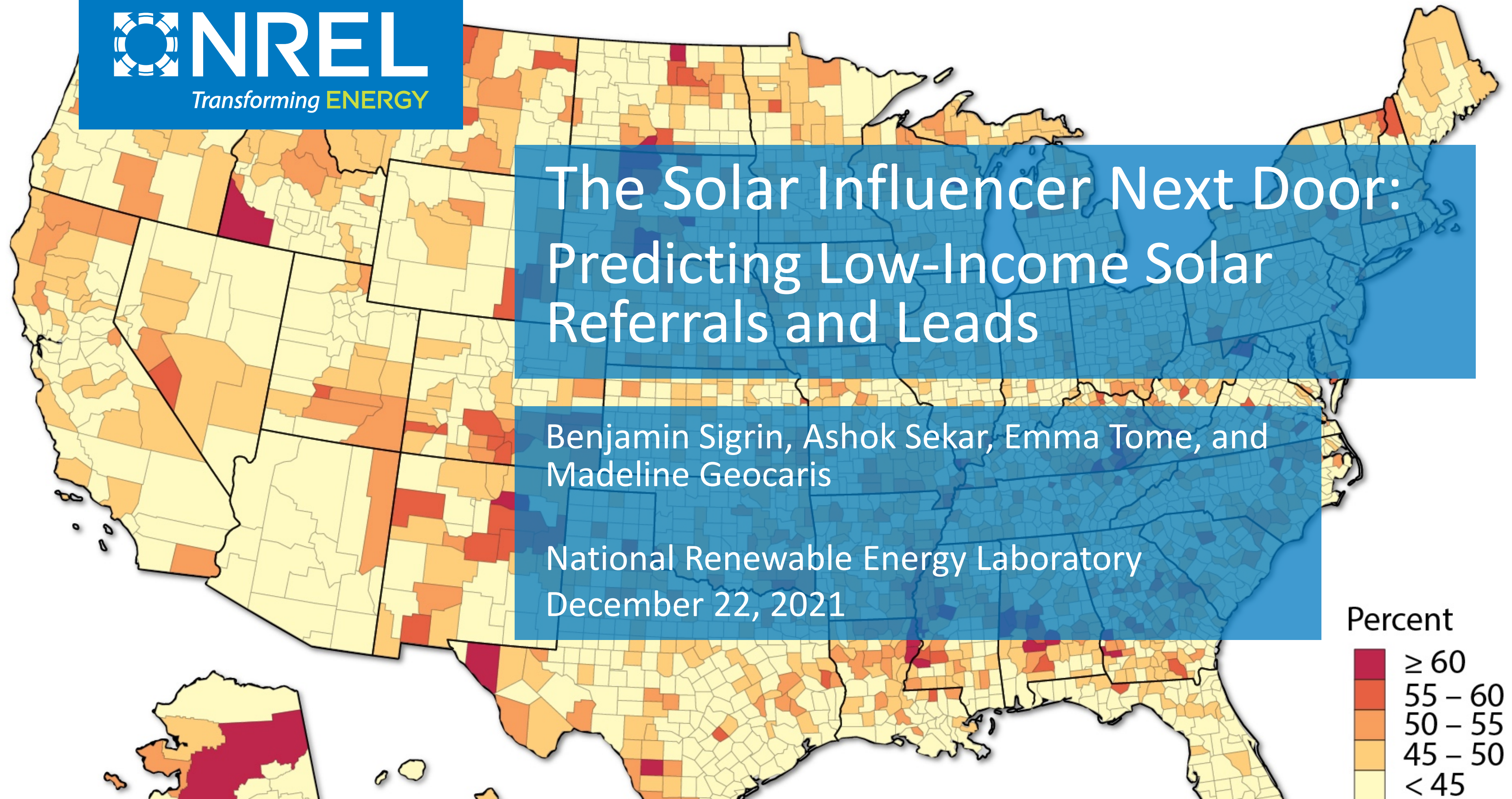
NREL
Transforming ENERGY

The Solar Influencer Next Door: Predicting Low-Income Solar Referrals and Leads

Benjamin Sigrin, Ashok Sekar, Emma Tome, and Madeline Geocaris

National Renewable Energy Laboratory

December 22, 2021

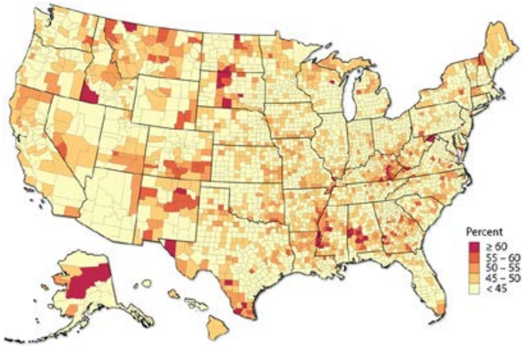


SEEDS II Overview

Solar Energy Evolution and Diffusion Studies, Solar Energy Technologies Office

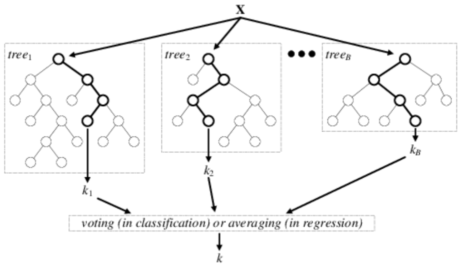
Project Goal:

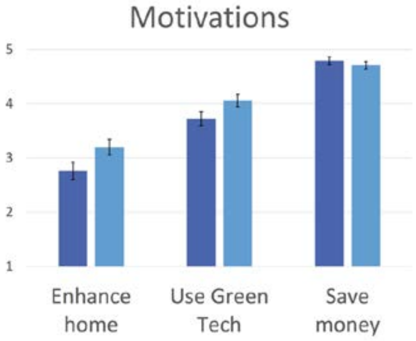
Identify strategies to scale up solar adoption among low-and-moderate income (LMI) communities across the U.S.



Determine technical potential of LMI market and opportunities for new deployment models
https://maps.nrel.gov/solar-for-all

Use historic data to develop predictive models for adoption and referrals





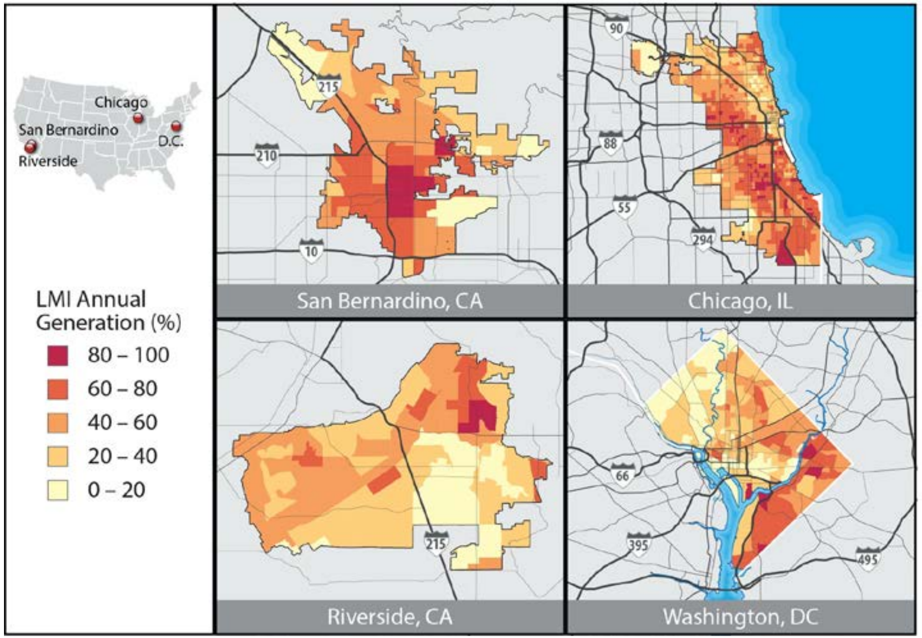
- Survey LMI adopters
- Comparative city case studies
- Pilot test referral elicitation strategies

Why is Solar Adoption in Underserved Communities Important?

- Growing divide in adoption rates by income groups
- A large, unaddressed market: 100s of GW of potential

Extending solar to more communities could:

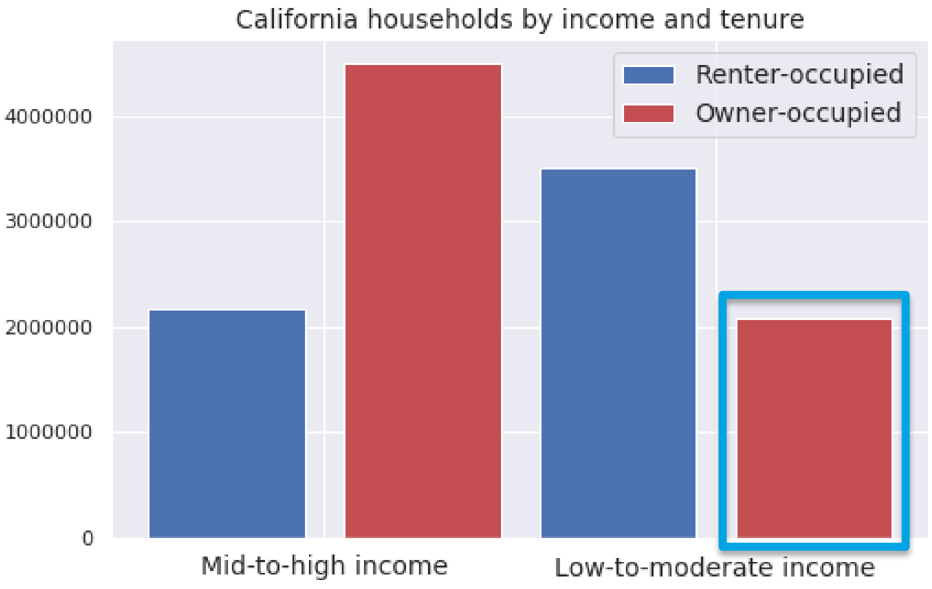
- Reduce energy burden
- Support energy equity
- Accelerate decarbonization
- Meet policy goals



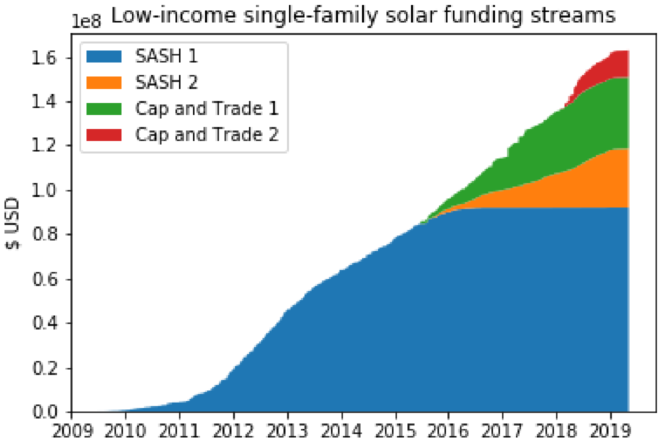
Some Families Face More Barriers to Adopting Solar

- Affordability
- Homeownership
- Financial: ability to secure financing or monetize tax-based incentives
- Structural: Unsuitable roof, electric code compliance
- Information and distrust

=> This analysis studies low-income owner-occupied potential



California's Low-Income Solar Programs



Our study uses anonymized household-level data from these programs:

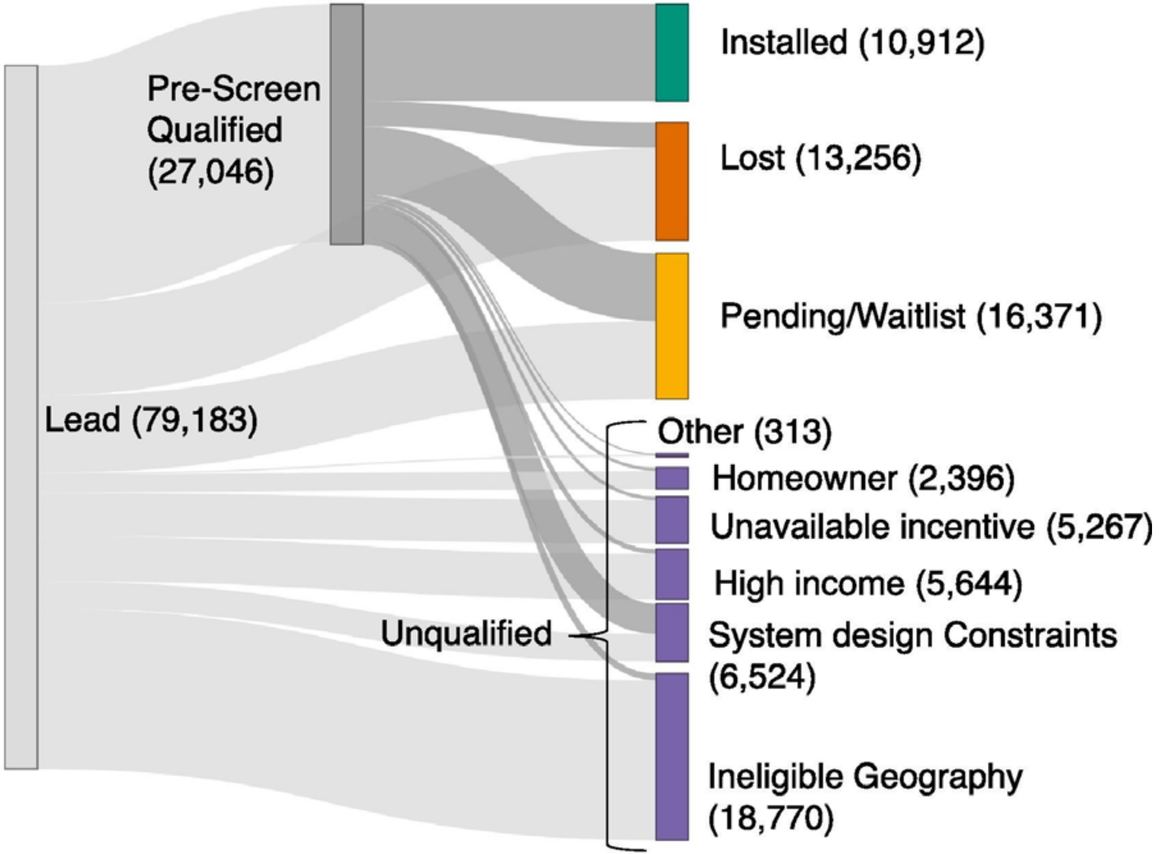
- Largest source of low-income program data in the U.S.
- Single provider means that all leads and referrals are captured in the database

Why do referrals matter?

RESULT:
Program Outcomes

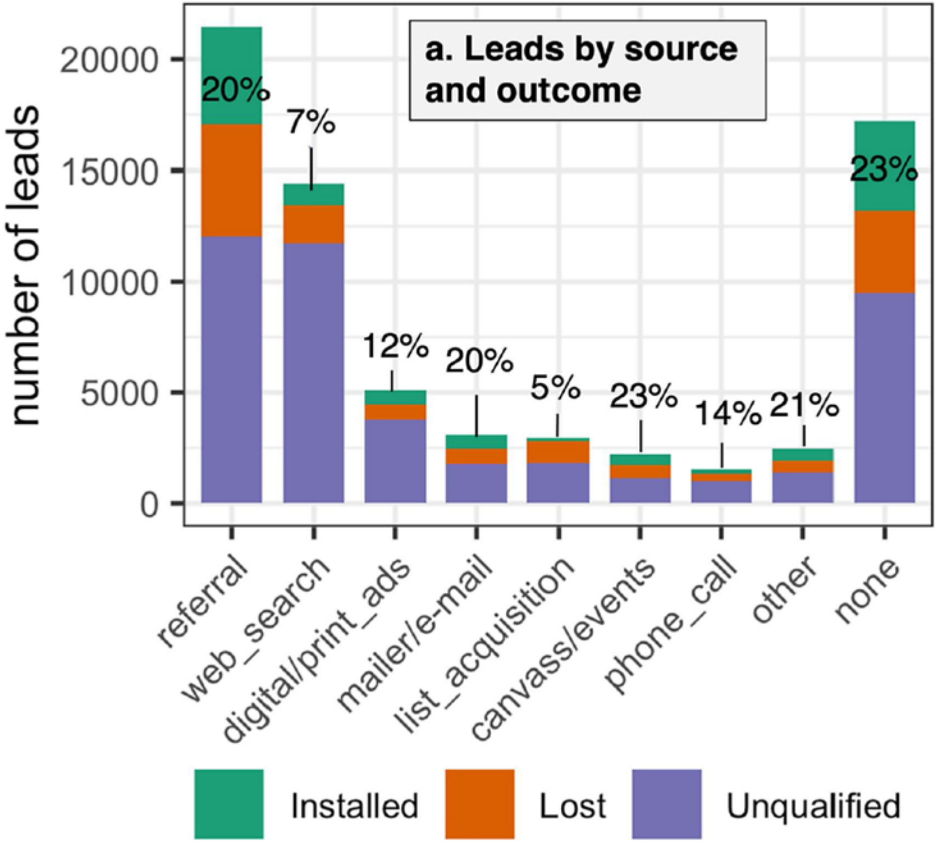
Eligibility criteria:

- Own and live in home
- Income below 80% AMI
- Live in an eligible geography:
  - Qualified census tract
  - Enterprise Zone
  - Targeted Economic Area
  - Opportunity Zone
  - Disadvantaged Community



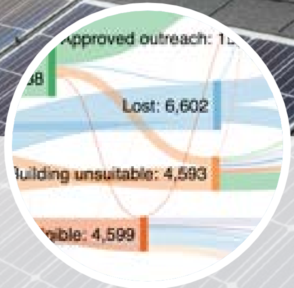
Over 85% of leads contacted do not receive solar. Of these, geographic ineligibility is the largest barrier.

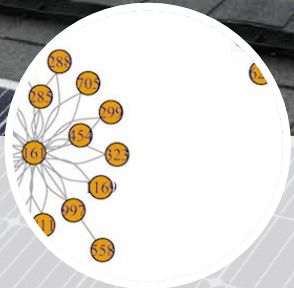
The Biggest Source of Solar Leads Is Right Next Door



- Referrals are the largest sources of leads and clients
- Referrals had nearly highest success rate
- Many other successful methods of lead generation

Predictive Modeling Approach





Tract-level

Within eligible geographies, where do LMI installations occur?

Adoption

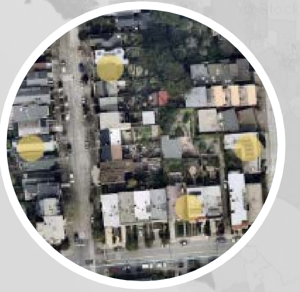
Will a qualified lead become lost or disinterested?

Referrals

Will a former client provide a referral?

Tract data
(Public)

Client data
(Private)



DeepSolar

Predictive estimates of solar uptake and socio-demographic predictor variables

(Yu et Al 2018)

REPLICA

Rooftop technical potential estimates by building, income, tenure.

(Sigrin and Mooney 2018)

Cal EnviroScreen

Environmental and socioeconomic vulnerability indicators

(OEHHA 2017)

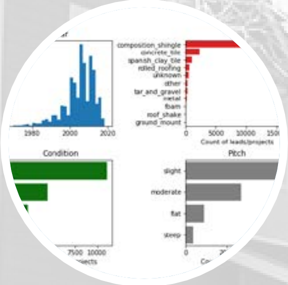
Relational

Installed base

Distance measures

(Graziano, Bollinger, Gillingham 2019)

(LBNL Tracking the Sun)



Household

Demographics

Lead source

Electricity consumption

PV system specs

Modeling approach

1. Training:
Fitting the model to a subset of data

2. Validate
Evaluate hyperparameter fit

"Crossfold" data 5 times to avoid overfitting

3. Test
Evaluate model fit on out-of-sample data

4. Analyze
Feature importance, variable influence

Predictive models can be evaluated in many ways, but ultimately:

- How well does the model predict out-of-sample?
- Does the performance on the Test approximate that of the Training?

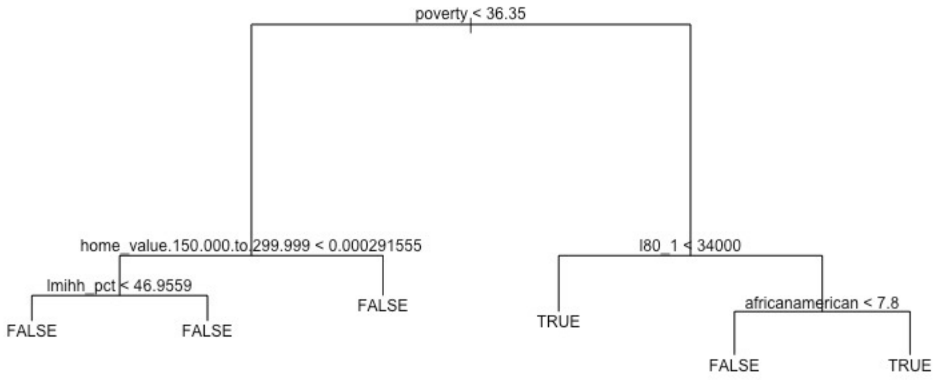
Balanced Accuracy used to assess performance due to imbalanced data, i.e. more non-referrals than referrals

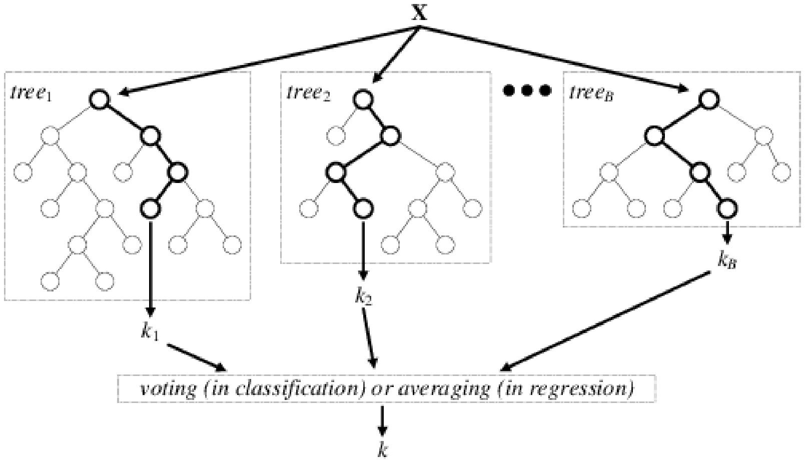
Decision tree

- Partition features according to the outcome of interest

Random forest

- Creates many decision trees from a random subset of available variables
- Averages the predictions from the trees

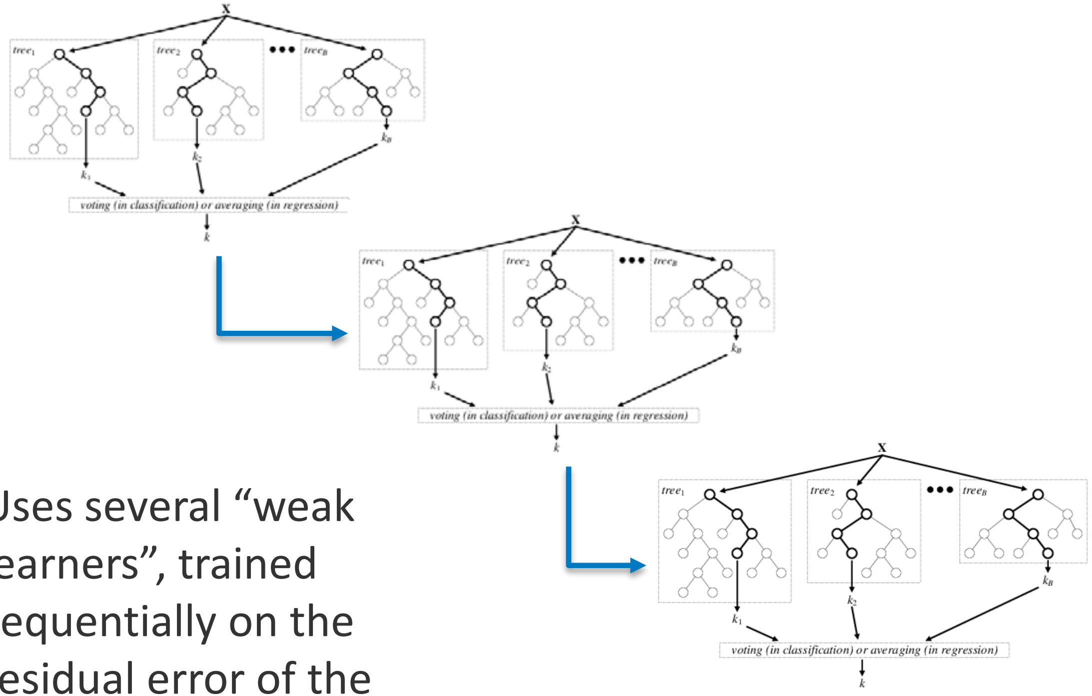




Extreme Gradient Boosting (XGBoost)

Hyperparameters:

- n trees
- max depth
- min samples
- learning rate
- subsampling rows/columns
- scoring



Uses several “weak learners”, trained sequentially on the residual error of the previous model.

The contribution of each tree is weighted by the learning rate. Early stopping reduces overfitting.

Lost Lead and Referral Models

Example: When are predictive models useful?

		Actual: No	Actual: Yes
Predicted	No	90%	5%
	Yes	10%	95%

1. Model sensivity and specificity influence predicted outcomes given actuals

		No	Yes
1000	Clients	880	120

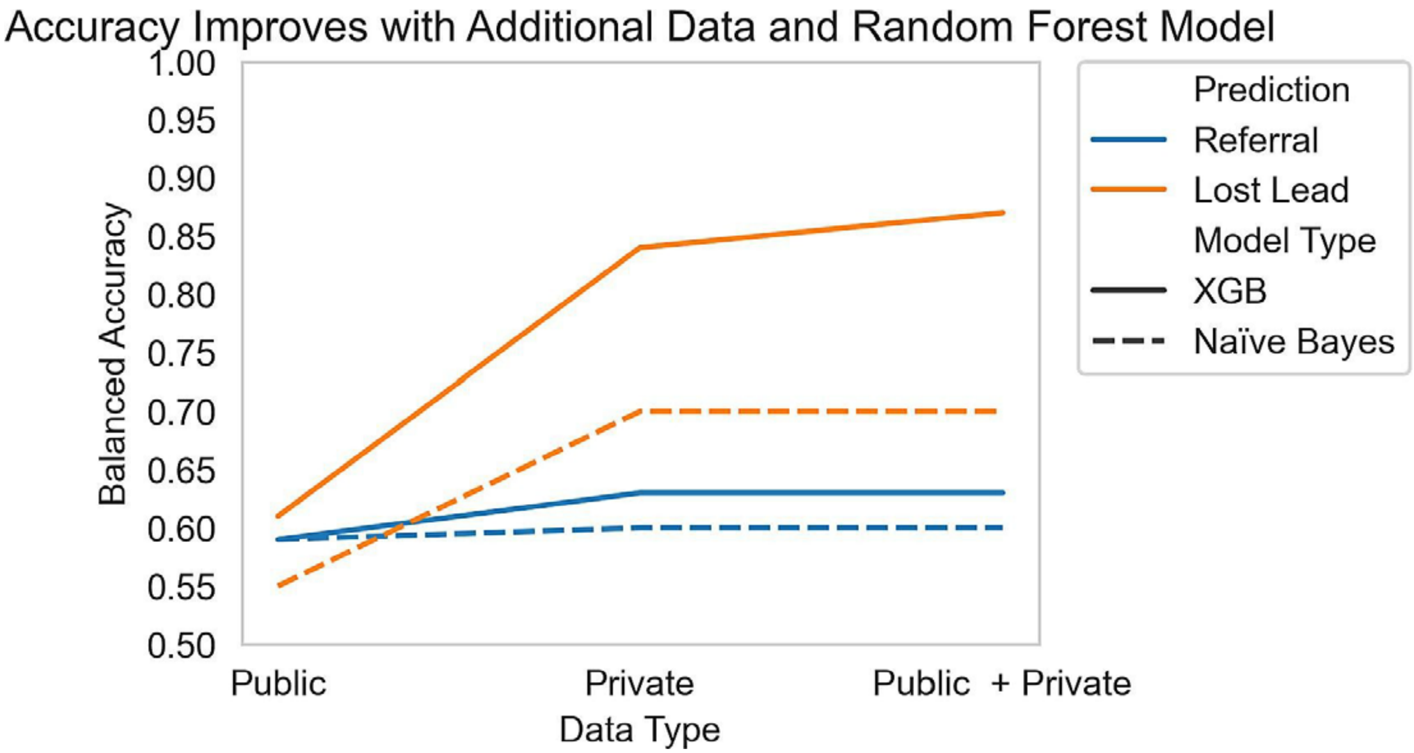
2. For 1,000 clients apply the prevalence of providing a referral

		Actual: No	Actual: Yes
Predicted	No	792	6
	Yes	88	114

3. Yielding 202 clients predicted to be referrers, of whom 114 are referrers. Clients not predicted to refer (798) are not pursued further though 6 are true referrers

Predictive models always imply some households are not pursued.
Used well they reduce time spent for a positive outcome.

Predictive Model Performance



Lost Lead Model Performance

Balanced accuracy is 87%.

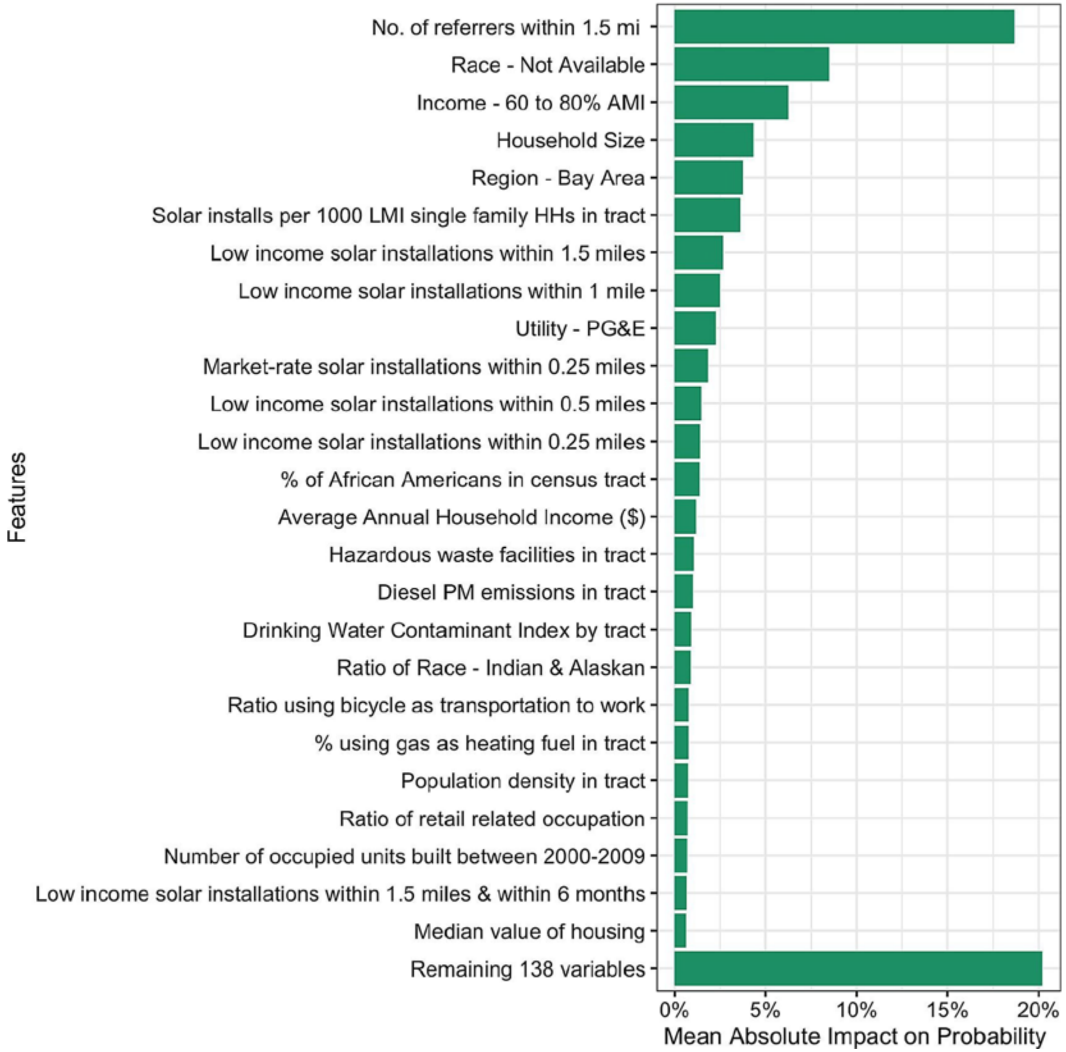
Model Type	Data Type	Prevalence	Sensitivity	Specificity	Precision	F1	Balanced Accuracy
XGB	Public	0.20	0.51	0.71	0.31	0.38	0.61
	All		0.87	0.87	0.63	0.73	0.87
Naïve Bayes	Public		0.55	0.56	0.24	0.33	0.55
	All		0.90	0.50	0.32	0.47	0.70

Lost Lead – Feature Importance

Feature importance plots show the most relevant features for the prediction and the magnitude

Number of referrers within 1.5 miles was the most important feature.

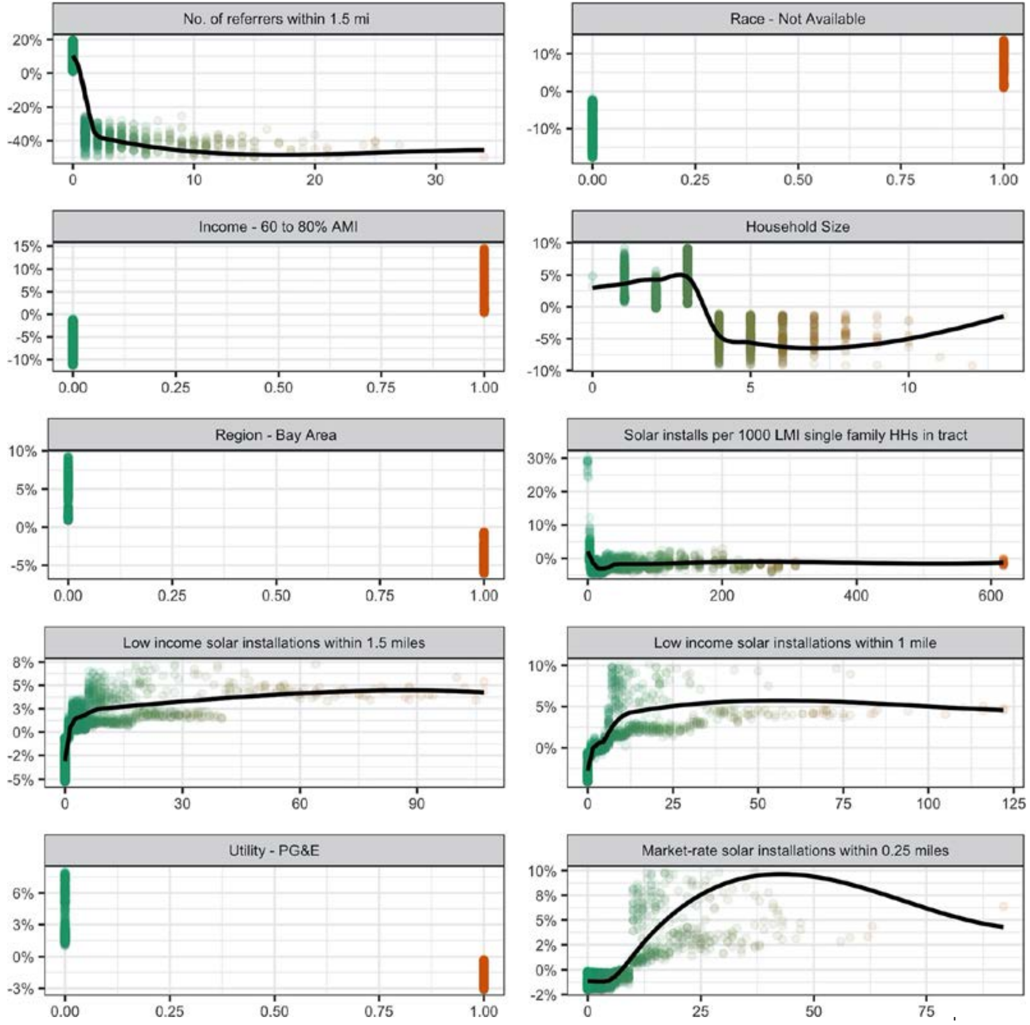
Several other proxies for peer effects were relevant as well as measures of environmental effects



Lost Lead – Conditional Dependence

Conditional dependence plots, based on SHAP analysis, demonstrate the direction and magnitude of influence

For example, more referrers nearby decreases the chance of a prospective solar adopter becoming disinterested.



Referral Lead Model Performance

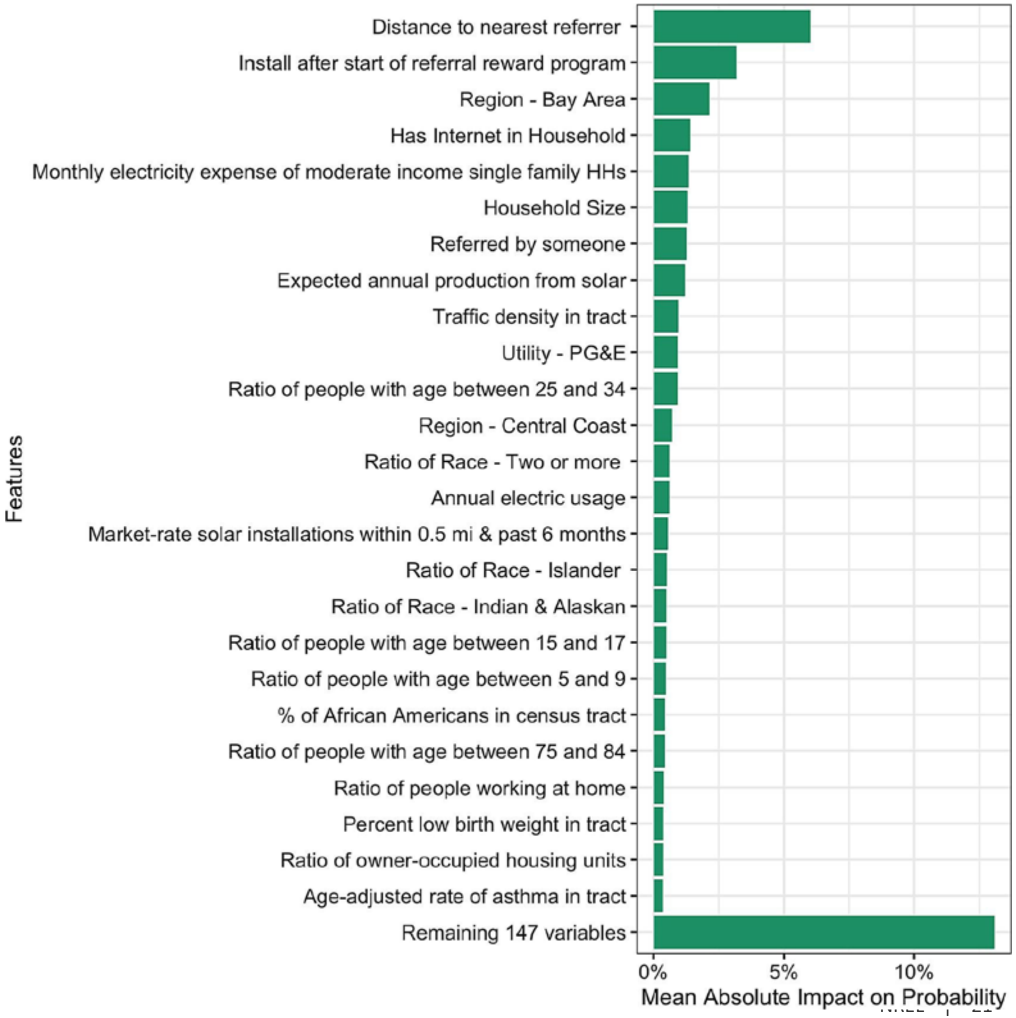
Balanced accuracy is 63%, lower than the lost lead model.

Model Type	Data Type	Prevalence	Sensitivity	Specificity	Precision	F1	Balanced Accuracy
XGB	Public	0.12	0.55	0.65	0.18	0.27	0.59
	All		0.61	0.61	0.19	0.29	0.63
Naïve Bayes	Public		0.55	0.62	0.17	0.26	0.59
	All		0.51	0.70	0.19	0.28	0.60

Referral – Feature Importance

Compared to lose lead model individual features have lower impact on prediction due to lower accuracy.

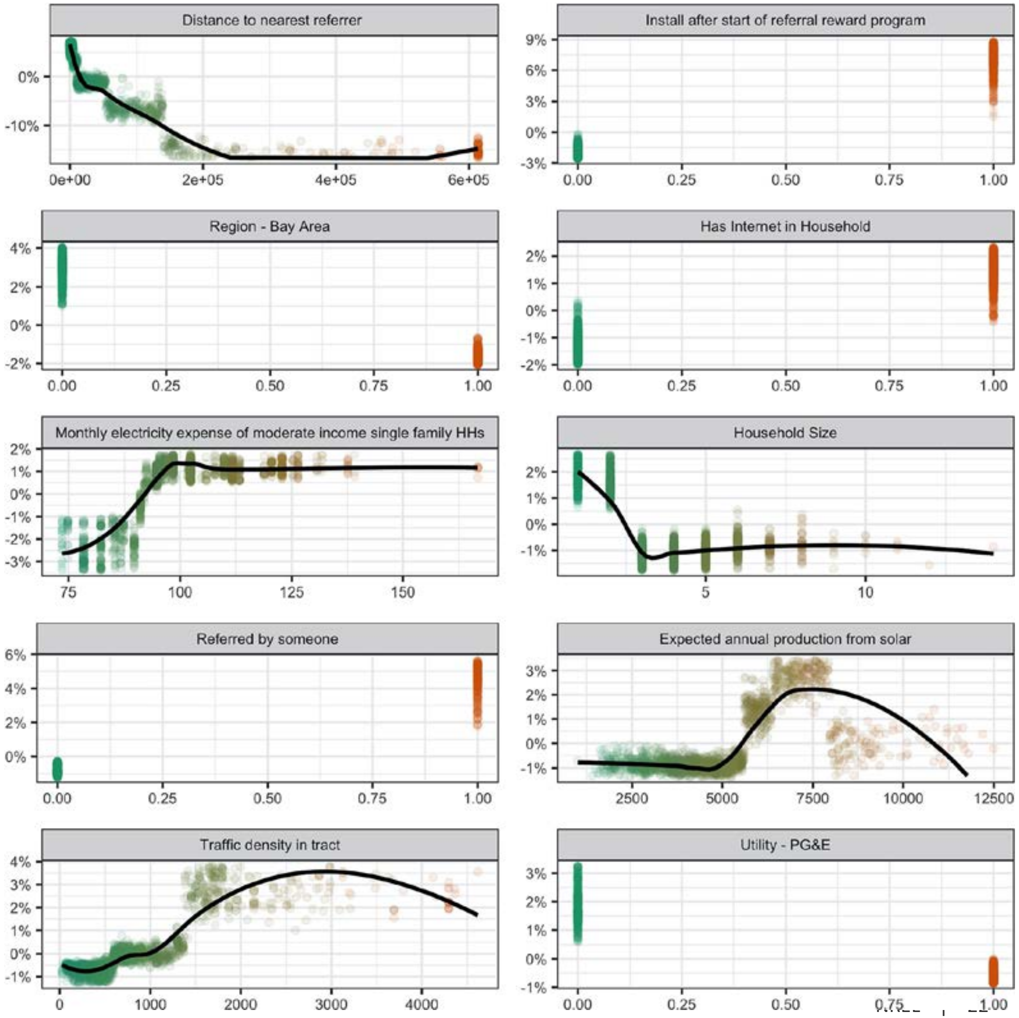
Distance from referrer is again the most impactful variable, as incentivizing referrals, internet access, expected solar bill savings



Referral – Feature Importance

Like lost leads, proximity to referrers and having been referred increases the chance of an adopter making a referral.

However, *unlike* lost leads, knowing the number of installations nearby did not have a large impact.

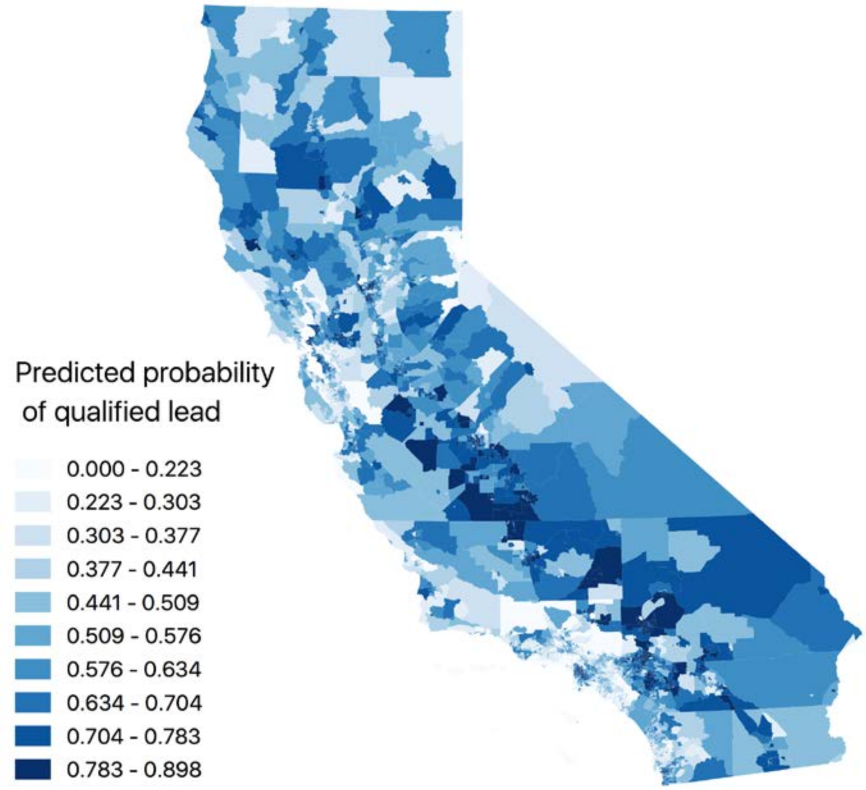


Program Implications

- **LMI adoption models** could support affordable solar program uptake and estimate policy outcomes.
  - Applying the lost lead model could save 20% of staff time to acquire leads.
  - Applying the referral model could save 56% of staff time to acquire referrers.

Program Implications

- **Referral incentives** should be built into program design.
- **More flexible eligibility requirements** could extend solar to more communities and reduce administrative burden.



Thank you

www.nrel.gov

NREL/PR-7A40-81771

Contacts: bsigrin@nrel.gov / ashok.sekar@nrel.gov
Project website: https://www.nrel.gov/solar/seeds/2017-2019-study.html
Read the *Energy Research & Social Science* article.
Subscribe to NREL SEEDS news.

NREL Transforming ENERGY