Predictive Modeling for Consumers' Adoption of Solar PV

W201 (Section 2) Final Project – Week 15

Aditya Nemani
Jan Forslow
Guillaume De Roo
Vincent Chu

Table of Contents

Client Industry Background 3

Current Landscape of Data Analytics 3

Proposed Research Question 4

Research Design 5

- Overview 5
- Data Collection and Survey Design 5
- Analysis, Presentation and Application 6

From Model to Decisions 7

Risks and Mitigations 7

Future Work 9

Appendix A: Sample Predictors and Corresponding Data Sources 10

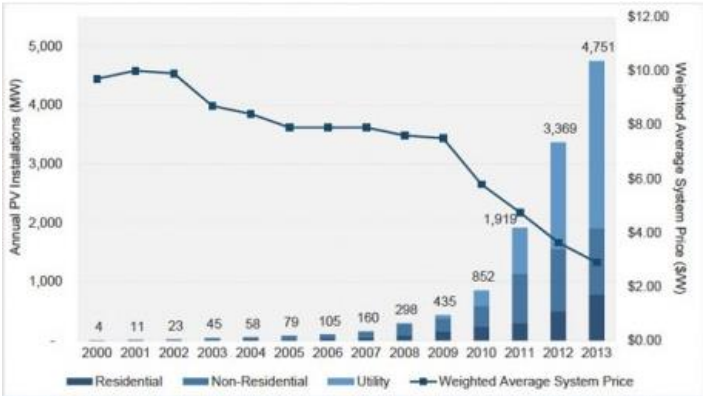
Appendix B: Sample Interview/Survey Questions 10

Works Cited 11

Client Industry Background

Solar photovoltaic (PV) started its entrance into the US residential market around year 2000 as part of the initial green energy wave coming from consumers and regulators. The average system price was above $10/Watt, making a 5kW residential solar installation an expensive investment for homeowners [1]. This high price point opened up the market for solar leasing companies that offer financing options of residential solar installations, e.g. SolarCity, Vivint and SunRun.

The evolution of PV technology has brought down average system cost to about $3/Watt and solar is expected to hit grid parity by the end of 2016. This cost reduction is challenging the original solar leasing business model, prompting SolarCity and others to reposition themselves as full-service solar companies with capabilities to provide optimal operation and maintenance of installed PV systems. The fast market growth[1] has also created another challenge for industry leaders. While the first million US residential customers (i.e., early adopters) were relatively easy to acquire, the next wave of customers will be harder to identify.



Installations (MW)	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Residential	1	5	11	15	24	27	38	58	82	164	246	304	494	792
Non-Residential	2	3	9	27	32	51	67	93	200	213	339	831	1,072	1,112
Utility	0	3	2	3	2	1	0	9	16	58	267	784	1,803	2,847
Total Installations	4	11	23	45	58	79	105	160	298	435	852	1,919	3,369	4,751

Figure 1: US PV Installations and Average System Price, 2000-2013 [2].

Current Landscape of Data Analytics

While players in the solar industry[2] have been utilizing customer research and analytical modeling techniques for various operational purposes, customer acquisitions have largely relied on referrals as well as broad direct-response marketing and channel/partner events. The limited set of marketing data analysis in use tends to focus on financial indicators (e.g. income and credit score), geographic indicators (e.g. weather and average hours of sunlight), and policy driven indicators (e.g. state and federal rebate programs). While these types of indicators

[1] GreenTech Media expects that we will approach 22% penetration of the overall 50 million potential rooftops solar market in the next 5 years [5].

[2] Key players in the solar industry include solar providers, manufacturers of solar panels and inverters, utility companies, etc.

deliver high-level, directional trends in sales, they don't provide granular details that will enable solar providers to predict consumer behavior and shape their marketing strategies accordingly.

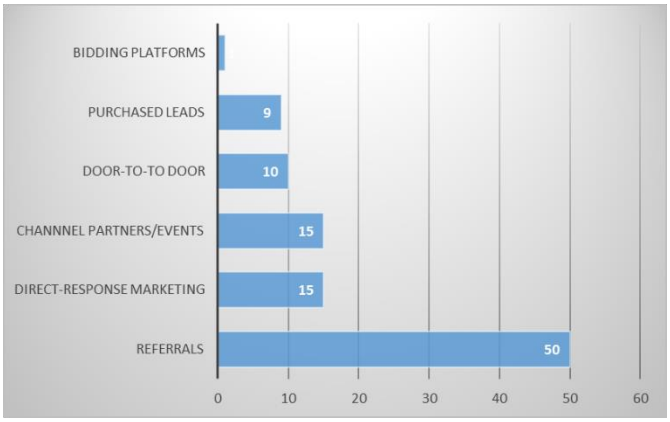


Figure 2: U.S. Market-Wide Sources of Customers in %-age, 2013 [2].

With its continuous leverage of IT technologies, the solar PV industry has a lot of potential to be a data-driven industry. Companies such as Geostellar and EagleView are able to systematically compute solar potential for neighborhoods and cities by leveraging cutting edge techniques such as ray-tracing, aerial photographs from drones, and 3D rendered CAD drawings[3] [3]. Another example is Google's ability to predict potential solar generation roof by roof. Leveraging data from these technologies will be a great step towards improving granularity in predicting new solar customer prospects for solar providers.

Proposed Research Question

As the historically defragmented solar market continues to consolidate, it becomes crucial for solar providers to be one step ahead of their competitors in customer acquisition. Predictive models with higher degree of accuracy will offer an edge to solar providers who want to be key players in the future. Not only do the low success rates of marketing campaigns[4] impose big costs for solar providers[5], they hurt confidence of utility companies to invest in cross-promotional campaigns. Moreover, the sales model will be evolving from leasing to individual ownership and most likely to ownership by communities[6] in the future. This will alter the way that

[3] These techniques can be used to simulate atmospheric conditions and alongside high-resolution rooftop models take into account factors such as utilizable surface area, obstructions to sunlight (i.e., other structures or trees), roof angles and pitch and azimuth orientation to robustly predict solar power output at scale [6].
[4] Success rate of marketing campaigns for solar PV is typically <= 5%, according to Customer Care Department at NorCal Energy.
[5] According to GTM Research, it costs 49 cents per watt to acquire a solar customer in the U.S., which equates to about $3,000 for the typical 6-kilowatt residential rooftop [7].
[6] Community solar refers to solar systems installed off-premises and are intended to offer solar access to renters, homeowners with shaded or shared roofs, and those who does not desire to install solar panels on their own roofs [8].

prospective customers are identified: from considering only limited characteristics of "early bird" individual homeowners to including social factors affecting potential community clients.

The focus of this project will be to advise solar providers on the following research question: "How can companies use consumer data to accurately predict solar adoption in support of marketing analysis?" Both solar providers and utility companies are part of the target client base as the landscape of the solar industry continues to go through changes like the introduction of community solar.

Research Design

Overview

Because of the current distribution of solar installations in the U.S., this research will focus on California. A logistic regression model will be created with the decision variable being a binary variable indicating solar adoption. The variable "solar_adoption" will have value of 0 or 1, indicating whether an individual will adopt solar services in the next 5 years. A mixed research method will be deployed to build the regression model, using qualitative interviews to understand consumer behavior and identify potential predictors, and collecting quantitative data for the predictors through surveying and mining data from Utility companies and public sources.

Data Collection and Survey Design

Leveraging a relationship with NorCal Energy and an agreement to share findings, one of the main data sources will be responses to a survey intended for heads of households and distributed by NorCal Energy. Prior to sending out the survey, a set of interviews will be conducted with a random sample of customers who have already adopted Solar PV to observe any notable consumer pattern and behavior. The aim is to perform ~50 qualitative interviews to cover participants of different education levels (as a proxy for opinion towards solar), occupying different property types (as a proxy for decision making process), etc. Insights gained from these interviews will help identify potential predictors for solar adoption and shape survey questions[7].

Stratification of NorCal Energy's service territory will be performed to focus the survey on areas with high growth potential for solar PV. The 20 divisions of NorCal Energy's service territory[8] will be ranked based on a scoring system that considers annual number of clear-sky days, monthly irradiance index, PV penetration and year-over-year growth in solar PV. Focus will initially be on

[7] See Appendix B for sample interview/survey questions.

[8] Divisions are NorCal Energy's geographical groupings that determined where locations of local operational offices (called yards). These geographic groupings represent either a major city plus its surrounding areas (e.g., Sacramento, San Francisco, San Jose, etc.) or a collection of cities that make up a well-known geographic region (e.g., Peninsula, Napa/Sonoma Wine Country, North Sierra, etc.). We can assume that the climate within the same division is rather uniform in most cases.

the top 3 divisions, where customers will be further stratified into NEM[9] and non-NEM. A random sample of customers will be drawn from each strata and NorCal Energy will distribute the survey to 5,000 solar and 5,000 non-solar customers[10]. The survey will be a cross-sectional, self-administered questionnaire.

NorCal Energy will also provide access to data of their electric customers such as city of residence, usage pattern, rate plan, etc. Additionally, public data, such as weather and social networking data, will be used (e.g. number of friends with solar installations to test effect of "social pressure")[11].

Analysis, Presentation and Application

A hierarchical approach to build a logistic regression model will be used based on the set of predictors identified through the qualitative interviews. Data from the survey, NorCal Energy's databases and other public sources[12] will be used to train and finalize the model in terms of statistically and practically significant predictors. The expected result is a robust model that can predict whether a particular customer will adopt solar services.

Once the significant predictor variables have been determined, the target is to generate input data by zip code in the targeted divisions. For example, we will generate *n* data points for a zip code with *n* households and estimate values of the other predictors based on data obtained from financial / marketing research firms (such as Experian) and public sources such as GSS[13]. The final outcome of will be number of potential new solar customers by zip code.

The analysis results will be presented as heat maps. In contrast to existing heat maps focusing on irradiance[14], these customer acquisition heat maps will show a concentration of potential future solar customers. Different shades of red will represent high, medium and low number of potential customers. Geographic coordinates and contact information (if available) of prospective households will be provided with the heat maps. Solar providers can consume this data into their marketing tools, including door-to-door and call center marketing activities.

[9] NEM stand for Net Metering, which is the way how electric customers benefit from installations. Customers are refunded based on the energy generated by their solar installations and pushed back into the utility company's grid.

[10] According to NorCal Energy's Customer Care group, the average response rate of surveys sent out to their customers is about 50%.

[11] A data mining module has been developed by the team to mine public social networking data such as Tweets and Facebook posts in order to understand one's social network with solar installations (i.e., number of friends with solar).

[12] Public data sources include climate/irradiance (from National Weather Service), estimated electric generation (from Geostellar / EagleView), social network with solar (from mining data from social network such as Facebook, Tweeter, etc.), political affiliation (from district election results), property characteristics (from US public property records), etc.

[13] GSS (General Social Survey) is a sociological survey used to collect information and keep a historical record of the concerns, experiences, attitudes, and practices of residents of the United States.

[14] See Figure 3 as an example.

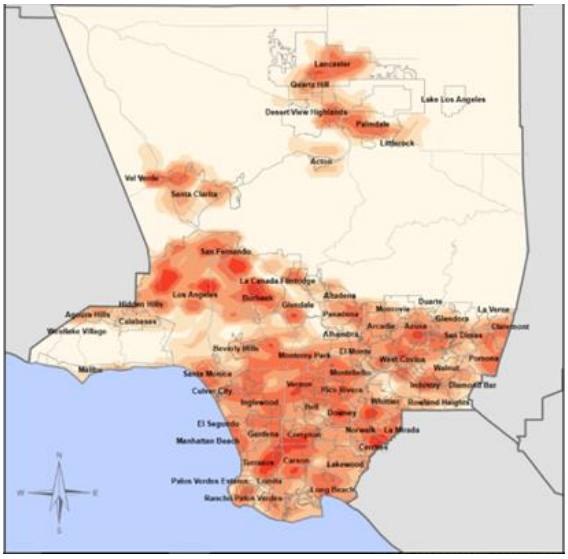


Figure 3: Heat Map of LA County Solar Potential [4].

From Model to Decisions

Results from the model will help solar providers to focus their marketing campaigns (i.e., financial and human resources) on the "low-hanging fruits", i.e., geographic areas with more potential for new solar adoption. It will also help solar providers to estimate solar generation and corresponding energy/dollar savings[15] to further entice customers. By referencing the heat maps, solar providers can be more focused when soliciting support from utility companies for running cross-promotions. Furthermore, the research findings from this project can be leveraged in an ongoing consulting service to advise solar providers on which indicators are the most predictive (statistically and practically significant) and how to optimize their current marketing campaigns continuously.

The model can also help utility companies make important operational and compliance-related decisions. For example, they can use this information to plan for any infrastructural and asset upgrades required to support the increase in solar PV penetration, especially considering the future trend of community solar. Utility companies can also have a better plan to hit green energy commitments mandated by the utility commission, for example, the State of California's mandate to achieve 50% renewable energy by 2030.

Risks and Mitigations

The following table classifies the major challenges during the study and associated mitigation strategies.

[15] Tools such as Geostellar and EagleView will be used to estimate solar generation and corresponding energy/dollar savings data.

Type	Description	Mitigation
Ethical / Legal	Unnecessary prying: Putting together information that individuals never meant to provide can be perceived negatively (e.g. willing to give different data to separate players without imagining it being combined).	• Check T&Cs[16] from providers to avoid legal entanglements. • Provide opt out option for surveys.
Ethical / Legal	Turning information against the client: Suppose the project finds locations in which the value of the building would be greatly improved by a solar installation. Someone inside the project could decide to buy the building from the unsuspecting owner?	Develop internal compliance processes to avoid purchases by company or individuals based on privileged access to information.
Data Access	Some behavioral data can be hard to obtain outside of the survey, which is not scalable for commercial use.	• Partner with companies who can gain the information, e.g. utility companies. • Use correlated public proxies, e.g. district election results rather than individual political affiliation.
Data Quality	Survey accuracy can negatively impact the model calibration.	• Introduce trick questions (e.g. who is president of USA) and filter based on time to exclude fake respondents. • Go through survey providers, who usually have reliable panels, with incentives attached. • Survey a large enough sample. • Use follow-up phone calls and emails to solicit responses. • Once predictor is used commercially, recalibrate based on actual success data.
Data Quality	External database can be hard to cross-reference between one another.	• Use several “unique identifiers” (address, phone number, Social Security Number, etc.).
Organizational / Cultural	Internal cultural biases could affect the ability to perform good qualitative interviews and discover the most relevant predictors (e.g. assuming political bias and failing to see property management).	Design interview guide with survey provider, and make external person conduct interview while listening in.
Organizational / Cultural	Companies that have invested a lot of money into very different and expensive marketing systems and approaches could suffer from sunk cost bias.	• Partner with newer companies or new executives in more established companies. • Use successes with initial customers to show values of our research.

[16] T&C stands for Terms and Conditions

Future Work

In order to improve our predictive model to be applicable across California (and the U.S. in the future), the same research process could be extended to NorCal Energy's other divisions. With data points covering a larger geographical area, this initiative will be able to help its clients predict solar adoption even better and continue to be competitive as the "low-hanging fruits" runs out and the solar market becomes more saturated.

With the initial set of data collected, more insights can be drawn on customer behavioral information regarding pricing and promotion. This can open up for creating more specific models to provide insights on the following questions:

- What types of cross-promotions are effective (e.g. advertising on utility bills and with home improvement retailers)?
- What types of incentives do customers better react to (e.g. promotions, revenue sharing or refunds)?

Appendix A: Sample Predictors and Corresponding Data Sources

Predictor		**Source**
Usual household characteristics	City / Zip Code	Utility Company
	Household income	Survey
	Decision maker / head of household demographics	Survey
	Size of household	Survey
Solar installation potential	Average temperature by month	National Weather Service
	Average number of sunny days by month	National Weather Service
	Average Irradiance Index	National Weather Service
	Estimated Energy Generation (based on average solar installation capacity)	Geostellar / EagleView
	Average monthly electric bill	Utility Company
	Average monthly energy consumption (KWh)	Utility Company
Environmental awareness	Selection of green energy option	Utility Company
	Membership of environmentally themed non-profit organizations such as Sierra Club	Survey
	Proximity of other solar initiatives (i.e., other installations in neighborhood or social circle)	Survey / Solar provider / Social network data
Receptivity to new technology	Ownership of Electrical Vehicle (EV)	Survey
	Early adoption of new technology	Survey
Property characteristics	Property Management profile	Public property records
	Age of building / last renovation / last sale	Public property records
Other behavioral data	Political affiliation	Survey / District election results

Appendix B: Sample Interview/Survey Questions

a) Do you currently have solar panels on your property? (Binary: 0,1)
b) What is your sex? (Binary: 0,1)
c) What is your age? (Range:18 - 100)
d) What is your annual household income? (Range: 0 - $1,000,000)
e) What is your education level (Categorical: below college, high school, college, advanced degree)?
f) How many individuals are in your household? (Range:1 - 10)
g) Do you own your home? (Binary: 0,1)
h) What type of home do you live in? (Categorical: Condo/Apartment, PUD, Single-family Home)
i) Do you currently own or lease an electric vehicle? (Binary: 0,1)
j) Do you think gas prices are too high? (Binary: 0,1)
k) Does your workplace have outlets for EV's in the garage? (Binary: 0,1)
l) Does anyone in your neighborhood have solar? (Binary: 0,1)
m) How strongly do you feel about environmental issues like global warming? (Scale: 1-10)
n) Are you a part of any environmental groups (e.g. Sierra Club)? (Binary: 0,1)
o) Do you align yourself with any particular political party? (Binary: 0,1)
p) If answer to o) was yes (i.e., 1), which party do you typically align yourself with? (String)
q) Do you consider yourself an early adopter to technology? (Binary: 0,1)

Works Cited

[1] "Photovoltaic (Solar Electric) Issues and Policies," [Online]. Available: http://www.seia.org/policy/solar-technology/photovoltaic-solar-electric. [Accessed Aug 2016].

[2] G. M. R. SEIA, "U.S. Solar Market Insight: 2013 Year-in-Review," 5 March 2014. [Online]. Available: http://www.solarnovus.com/2013-a-record-shattering-year-for-us-solar-industry_N7527.html. [Accessed 13 8 2016].

[3] "Geostellar – Commercial Residential and Utility Scale Briefing," Geostellar, Jan 2011. [Online]. Available: https://geostellar.com/. [Accessed Aug 2016].

[4] W. Norman, "Heat Map of LA County Solar Potential," Los Angeles County Enterprise GIS, 1 11 2011. [Online]. Available: http://egis3.lacounty.gov/eGIS/2010/06/16/heat-map-of-la-county-solar-potential/. [Accessed 13 8 2016].

[5] "GreenTech Media Briefing," in *InterSolar Conference*, San Francisco, 2015.

[6] "Eagle View Sample Residential Solar Report," EagleView, Aug 2010. [Online]. Available: http://www.eagleview.com/. [Accessed Aug 2016].

[7] R. Dumas, "Lowering Solar Customer Acquisition Costs With Phones, Software and the Cloud," GreenTech Media, 23 3 2015. [Online]. Available: http://www.greentechmedia.com/articles/read/Lowering-Solar-Customer-Acquisition-Costs-With-Phones-Software-and-the-Clo. [Accessed 13 8 2016].

[8] "A Guide to Community Solar: Utility, Private, and Non-profit Project Development," National Renewable Energy Laboratory, Jan 2011. [Online]. Available: http://www.nrel.gov/docs/fy11osti/49930.pdf. [Accessed 7 Aug 2016].