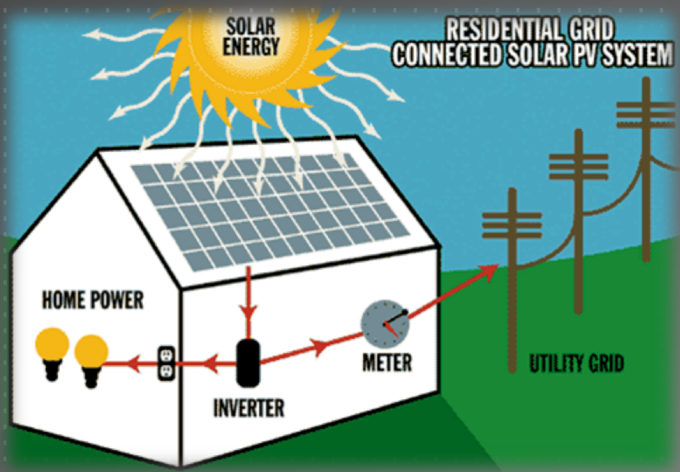
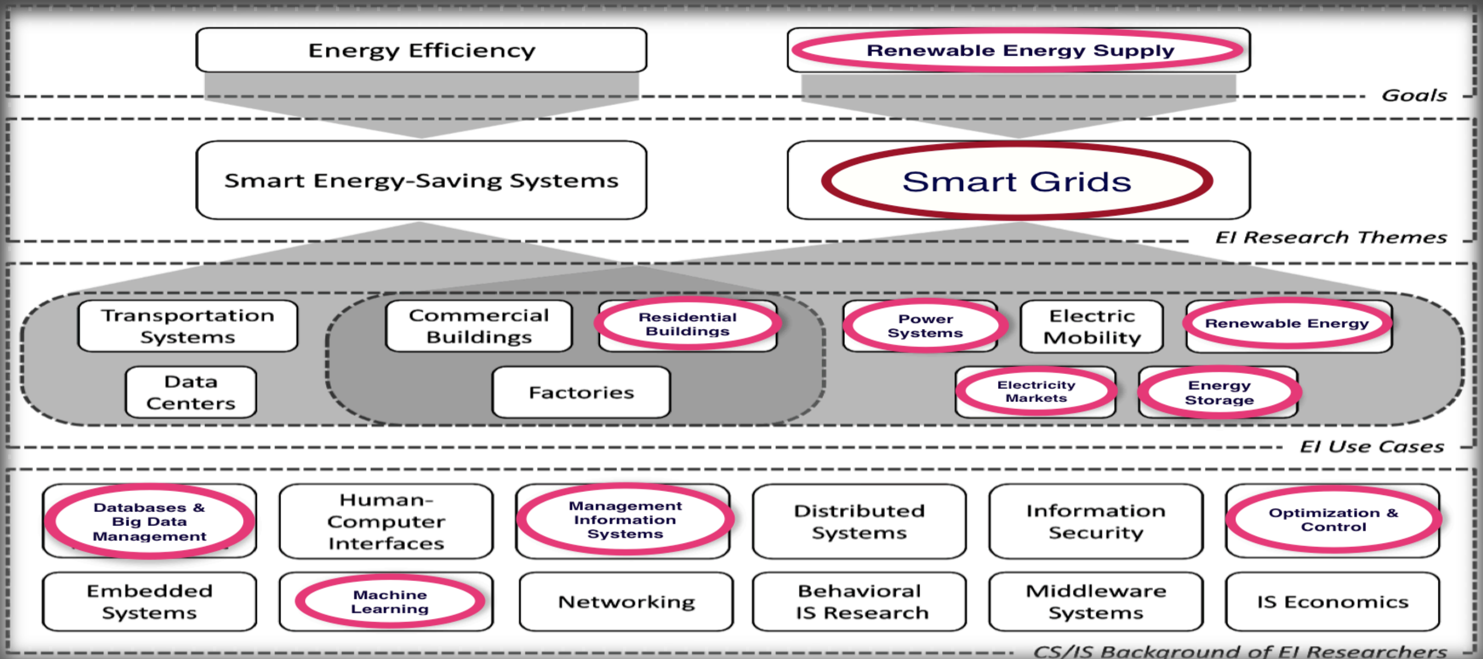
A PREDICTIVE MODEL TO FORECAST CUSTOMER ADOPTION OF ROOFTOP SOLAR



Vivian Sultan
Professor of Information Systems and Business Management
California State University, Los Angeles
College of Business and Economics
5151 State University Dr., ST F617
Los Angeles, CA 90032-8126
Email: vsultan3@calstatela.edu
Phone: (818) 419-4666

ENERGY INFORMATICS RESEARCH (GOEBEL ET AL. 2014)



PUBLICATIONS 2016

- "Geographic decision support systems to optimize the placement of distributed energy resources," International Journal of Smart Grid and Clean Energy, 5(3)
- "Is California's aging infrastructure the principal contributor to the recent trend of power outage?" Journal of Communication and Computer, USA, 13 (5)
- "Exploring Geographic Information Systems To Mitigate America's Electric Grid Traffic Congestion Problem," Proceedings of the 4th International Symposium on Computational and Business Intelligence, Switzerland, September 5-7
- "A Predictive Model to Forecast Customer Adoption of Rooftop Solar," Proceedings of the 4th International Symposium on Computational and Business Intelligence, Switzerland, September 5-7
- "Geographic Decision Support Systems To Optimize The Placement Of Distributed Energy Resources," Proceedings of the 22nd Americas Conference on Information Systems, San Diego, August 11-13
- "Is California's aging infrastructure the principal contributor to the recent trend of power outage?" Proceedings of the 22nd Annual California GIS Conference, Anaheim, May 10-12

BACKGROUND AND FOUNDATION

- Worldwide energy demand growing significantly
- Fossil fuels cause irreparable damage to the environment, climate change, air pollution, and a dramatic rise in carbon emissions
- RES is a viable solution
- Studies to pinpoint the most likely customers to invest in renewable energy technology
- Factors included household income level, house ownership, education level, and other financial factors
- A need of a predictive model for PV rooftop solar adoption

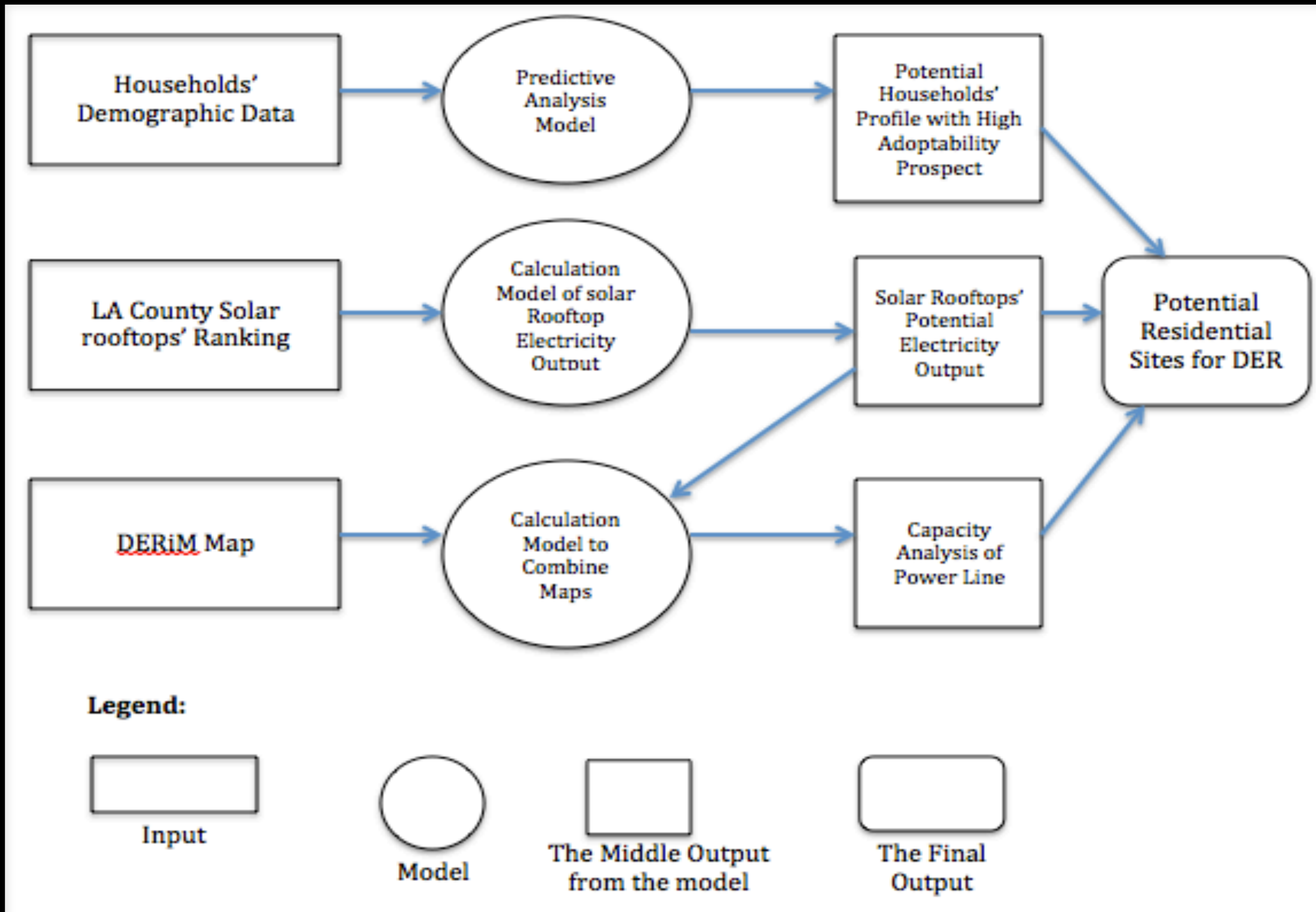
GDSS SOLUTION TO MEET UTILITIES' SPECIFIC GEO-PROCESSING NEEDS

Research Question: How can we forecast residential solar rooftop adoption?

Geographic Decision Support System

An interactive, computer-based system designed to assist in decision making while solving a semi-structured spatial problem

Geographic Decision Support System Model To Optimize DERs' Placement



DATA SOURCES

Solar Parcel Data

- Total roof area and area suitable for solar
- Potential solar system size
- Potential solar annual output
- Potential Cost Savings

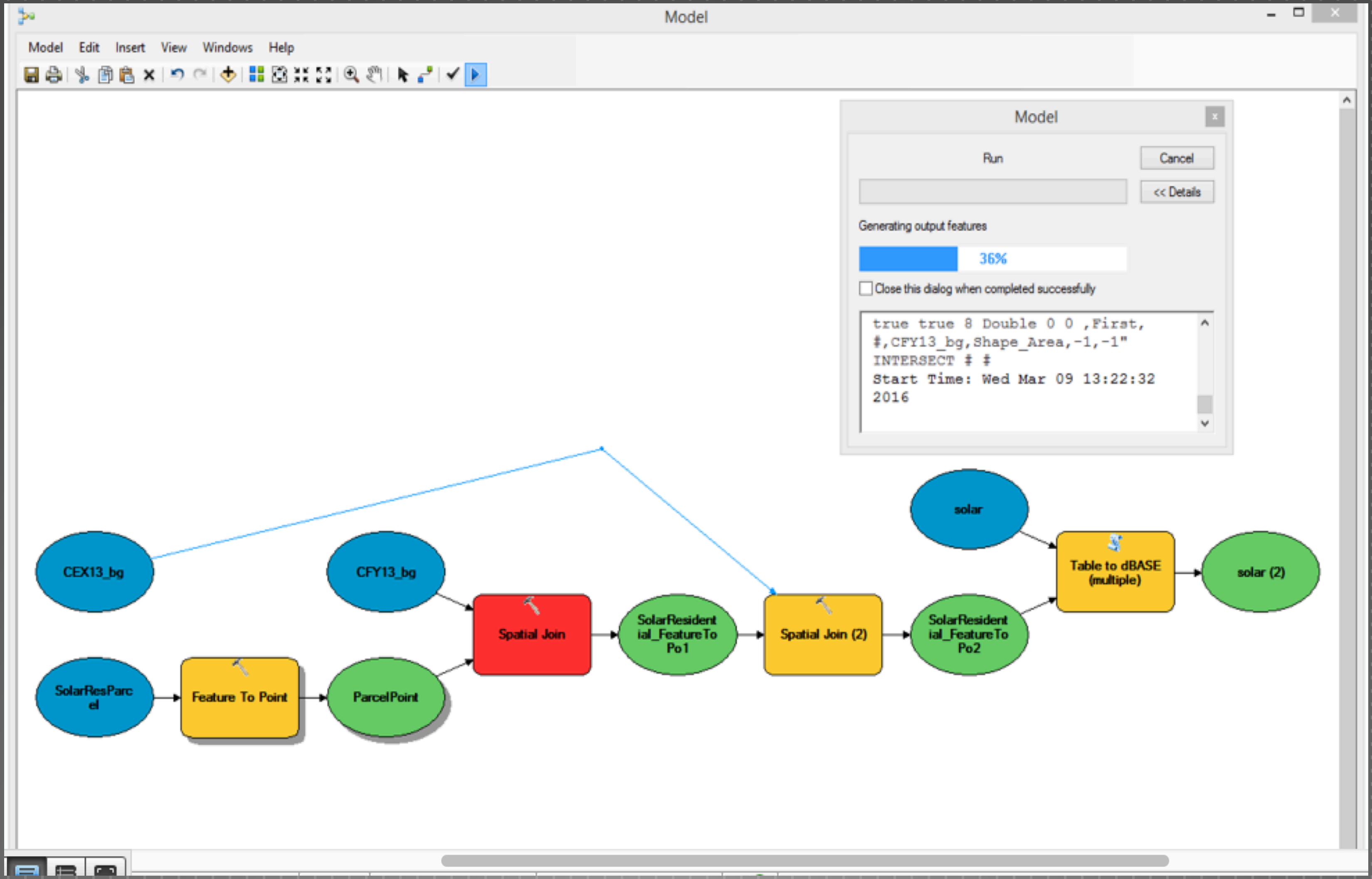
LA County's Solar Map
http://solarmap.lacounty.gov

Customer Expenditure & Demographic Data

- Total roof area and area suitable for solar
- Household expenditures
- Spending Potential Index
- Population demographic data
- Household data
- Income data

ESRI reports data at the block level

ArcMap Model Builder to add demographic and expenditure data to each parcel



KEY VARIABLES

Dependent variable

- The likelihood of a household to adopt solar energy panels

Independent variables

- Parcel information
  - parcel age, parcel value, etc.
- Customer demographics
  - household income, household size, etc.
- Expenditure data
  - electricity usage, mortgage value, etc.

Logistic regression

To predict the probability that a particular observation will fall into one of the dependent variable's two groups (Adopt/Not adopt solar system)

FINDINGS

Factors emerged as significant predictors

- ✓ Parcel age
- ✓ Total area suitable for solar roof top
- ✓ Average home value
- ✓ Average household size
- ✓ Total building area square feet
- ✓ Average household income
- ✓ Utility company – only 3 companies showed significant differences between solar adoption groups
- ✓ LA county cities

Probability of event, P (event) =

$$\frac{1}{1+e^{(a+b1*x1+b2*x2+\cdots+b10*x10)}} =$$

$$\frac{1}{1+2.718^{(1.832+-2.516*x1+.193*x2+-.543*x3+0.543*x4+-1.592*x5+.297*x6+-1.039*x7+.261*x8+.006*x9+-0.579*x10)}}$$

Cox and Snell and the Nagelkerke R^2

		B	Wald	df	Sig.	R2
Step 10	Pasadena Water and Power(1)	-2.516	144.542	1	.000	.092 - .137
	zscoreforAverageHomeValue	.193	6.288	1	.012	.172 - .255
	zscoreforAverageHouseholdSize	-.546	67.045	1	.000	.183 - .271
	zscoreforAverageHouseholdIncome	.543	54.491	1	.000	.195 - .289
	zscoreforTotalbuildingareasquarefeet	-1.592	32.886	1	.000	.216 - .319
	zscoreforTotalaresuitableforsolarrooftopr1r2	.297	29.183	1	.000	.223 - .330
	LONGBEACH(1)	-1.039	31.094	1	.000	.229 - .339
	Southern California Edison(1)	.261	5.759	1	.016	.231 - .342
	ParcelAge	.006	7.290	1	.007	.233 - .345
	Glendale Water and Power(1)	-.579	4.928	1	.026	.234 -. 346
	Constant	1.832	19.300	1	.000	

The study revealed a total of ten variables that significantly impact homeowners' decisions regarding whether or not to install solar rooftops. These ten variables accounted for around 23 percent of solar adoption variances. Future research is needed to expand on this analysis and include the whole state of California, as to apply this analysis to other Energy Informatics cases

Testing The Model

Classification Table[a,b]

Observed			Predicted: InstallYN10 0	Predicted: InstallYN10 1	Percentage Correct
Step 0	InstallYN10	0	2414	0	100.0
		1	807	0	0.0
	Overall Percentage				74.9

a. Constant is included in the model.
b. The cut value is .500

Classification Table[a]

Step 10	InstallYN10	0	2270	144	94.0
		1	521	286	35.4
	Overall Percentage				79.4

The model successfully classified 94% of households that did not install the solar panels and 35.4% of total cases that did install the solar panel. Overall success rate of 79.4 %

SOLAR ROOFTOP ELECTRICITY OUTPUT

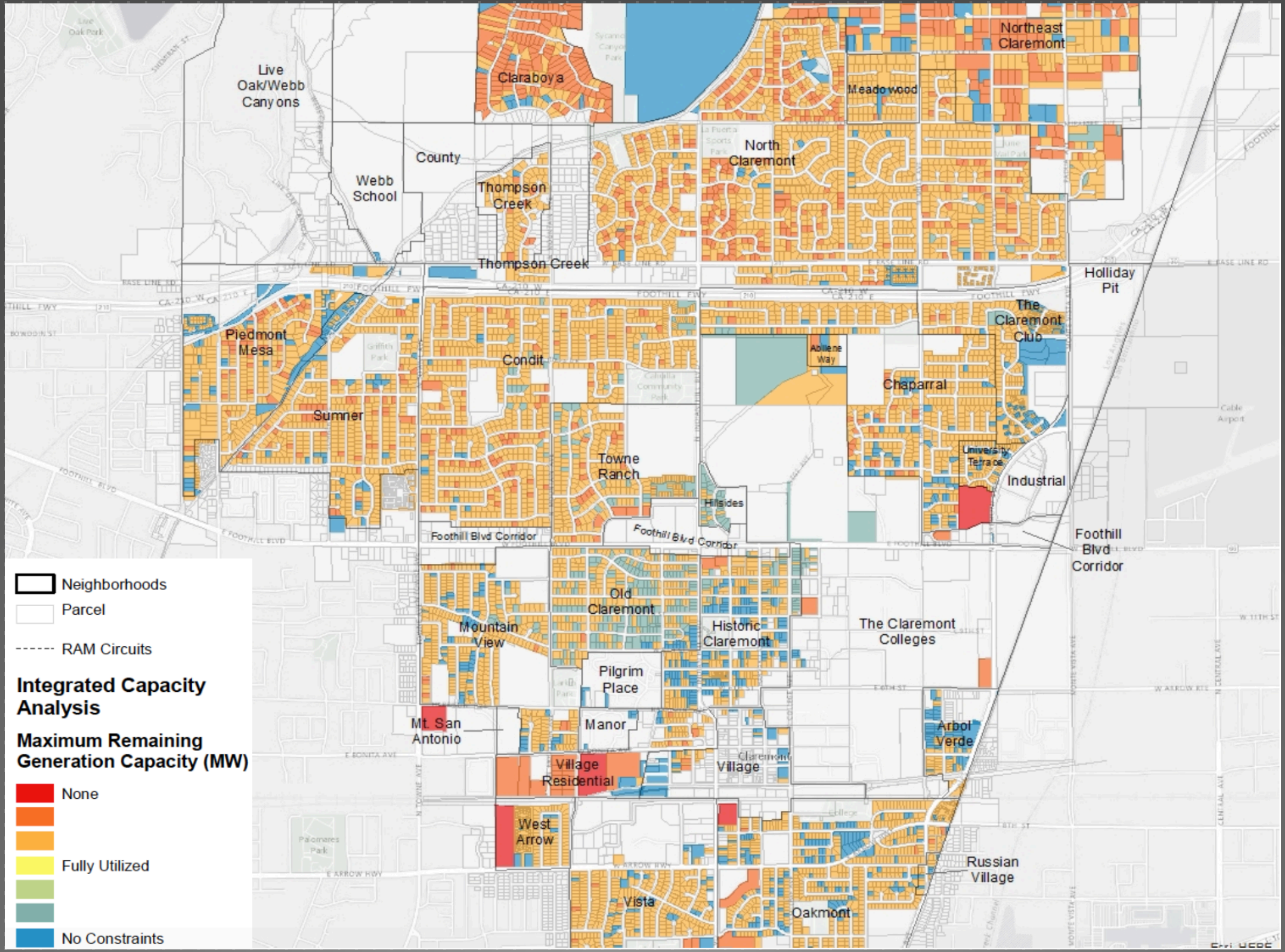
$$E=A*r*H*PR$$

E = DER's total energy output
A = Total solar panel Area
r = solar panel yield
H = Annual average solar radiation on tilted panels
PR = Performance ratio

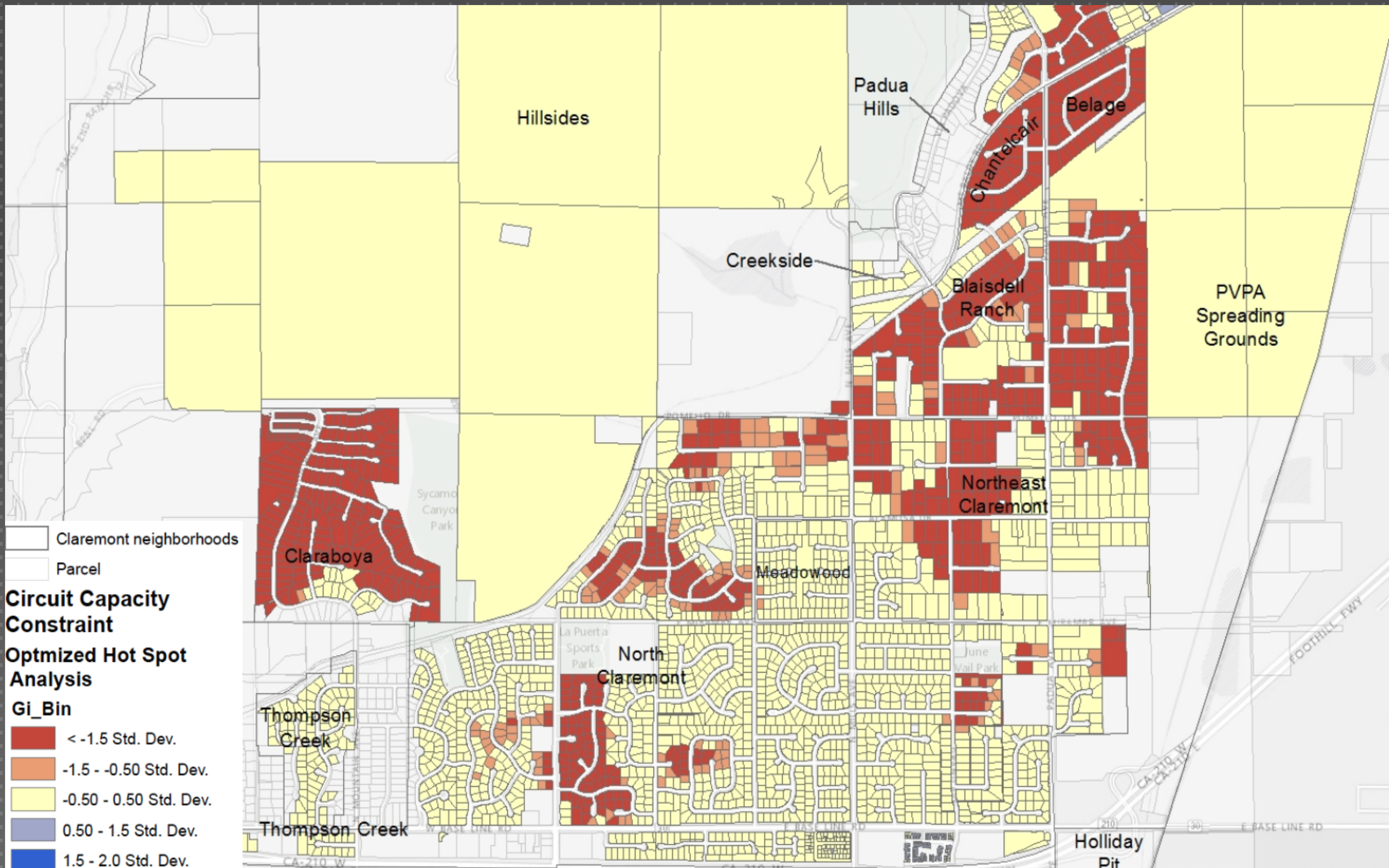
ELECTRIC CIRCUIT CAPACITY CONSTRAINT

Maximum Remaining Generation Capacity = Current Capacity Load - (Solar Potential + Existing Generation)

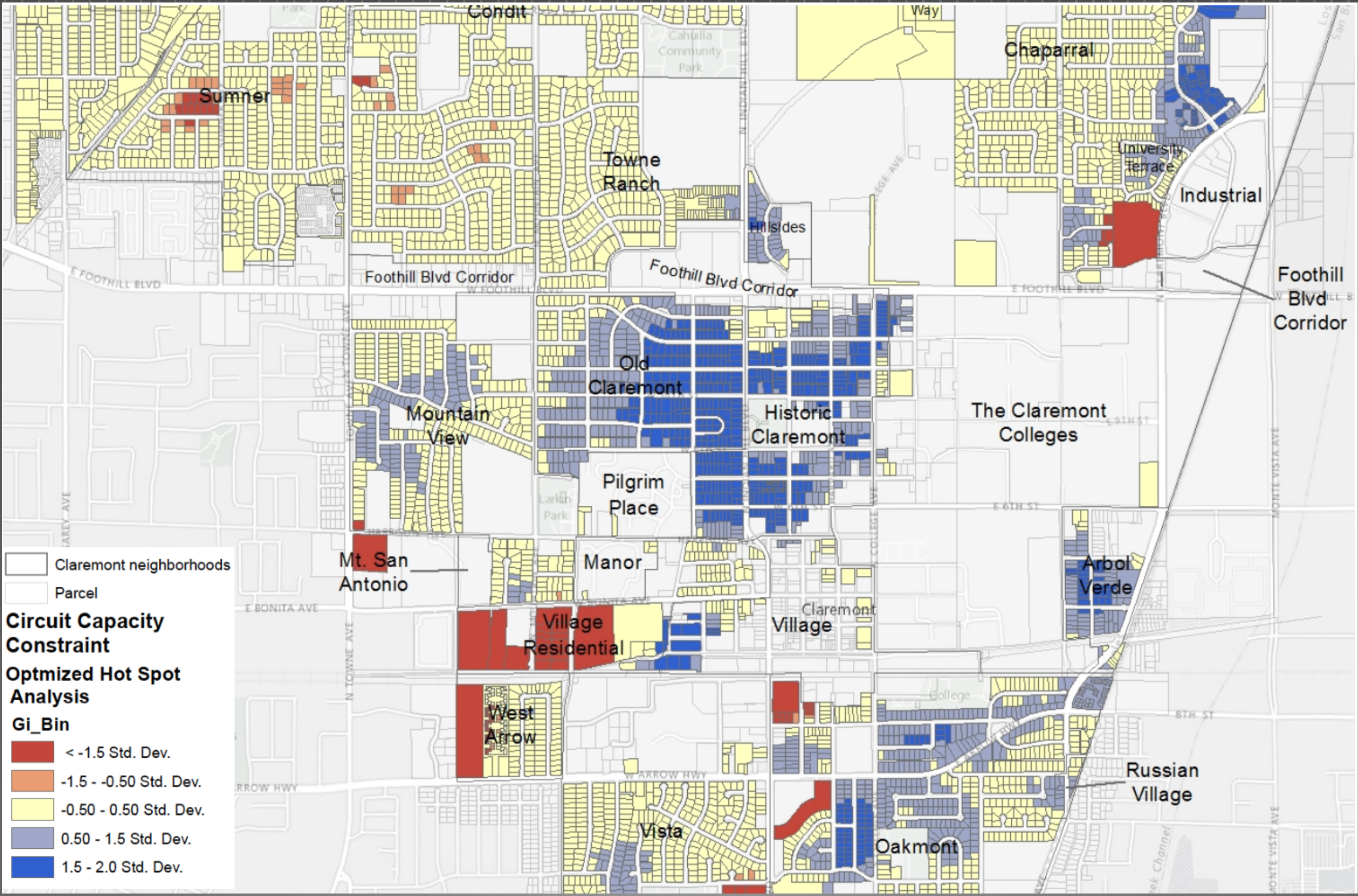
Claremont GDSS Model



Capacity Constraint Hot Spots in City of Claremont



Hosting Capacity Cold Spots in City Of Claremont



Infrastructure Work Prioritization Finding

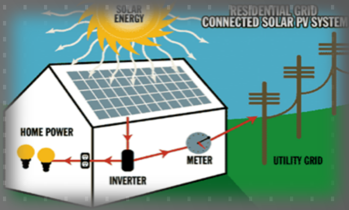
Circuit Name	Infrastructure Work Priority
PADOVA	1
BIG CONE	2
CALSPAR	3
FORBES	4
ANAWALT	5
NEIBEL	6
ALAMOSA	7
ROCK	8
BONTANIC	9
KINGSLEY	10
PITZER	11
WINTHROP	12
LIMBER	13
BASELINE	14
POMALL	15
LEHIGH	16
PALMER	17
MOAB	18
AVENIDA	19

Table 1: The Maximum Residential Solar Rooftops Adoption Scenario

Circuit Name	Infrastructure Work Priority
MOAB	1
BIG CONE	2
PADOVA	3
PALMER	4
LEHIGH	5
KINGSLEY	6
CALSPAR	7
AVENIDA	8
WINTHROP	9
BASELINE	10
LIMBER	11
BONTANIC	12
FORBES	13
NEIBEL	14
POMALL	15
PITZER	16
ALAMOSA	17
ROCK	18
ANAWALT	19

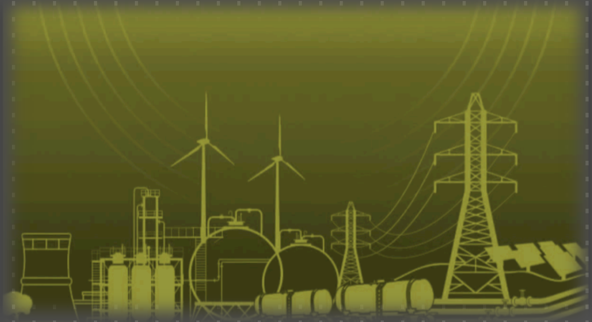
Table 2: The Existing Scenario

The output of this analysis can be utilized by a utility in several ways. It would result in a different prioritization in infrastructure work than those based solely on current or known installation projects. Knowing where solar installations are more likely to occur will impact energy planning.

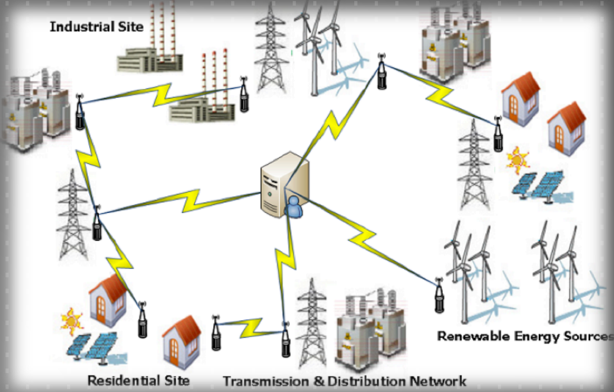


Thanks!

Q and A



APPENDIX


Industrial Site
Renewable Energy Sources
Residential Site
Transmission & Distribution Network

DATA PREPARATION STEPS

- Data extracted and loaded into ArcMap
- ArcMap model builder to spatially join three map layers
- Unnecessary and null fields eliminated
- Output DBF file saved as a geodatabase and exported as a txt file
- Text file converted to CSV and loaded into Microsoft Azure Machine Learning platform
- SQL query to pull current installations data
- Output JSON file converted to a CSV file format using Python code
- ArcMap Model Builder to obtain a map layer showing solar installations and their respective expenditure and demographics data
- Dbf output file converted to a CSV file
- Installation data merged with the entire dataset

STATISTICAL TESTS

Testing the four assumptions of logistic regression

- ✓ Sample Representativeness
- ✓ The dependent variable to be measured on a dichotomous scale
- ✓ One or more independent variables can be either continuous or categorical
- ✓ Multicollinearity or rather non multicollinearity of your data

The Hosmer-Lemeshow test

- ✓ A statistical test for goodness of fit for logistic regression models

STATISTICAL TESTS

Independent-samples t-test

- ✓ To determine whether or not the mean different between two groups is significantly different

Chi Square test

- ✓ To reveal whether or not there is a relationship between two categorical variables

The variance inflation factor (VIF) and tolerance

- ✓ To determine if two or more predictor variables in a multiple regression model are correlated