RESEARCH BRIEF

Machine Learning for Solar Accessibility: Implications for Low-Income Solar Expansion and Profitability

Sruthi Davuluri, René García Franceschini, Christopher R. Knittel, Chikara Onda, and Kelly Roache

The solar industry in the United States typically uses a standard credit score, such as a FICO score, as an indicator of consumer utility payment performance and credit worthiness to approve customers for new solar installations. Using data on over 800,000 utility payment performance and over 5,000 demographic variables, we compare machine learning and econometric models to predict the probability of default to credit-score cutoffs. We compare these models across a variety of measures, including how they affect consumers of different socio-economic backgrounds and profitability. Relative to FICO, the machine learning model increases the number of low-to-moderate income consumers approved for community solar by 1.1% to 4.2% depending on the stringency used for evaluating potential customers, while decreasing the default rate by 1.4 to 1.9 percentage points. Using electricity utility repayment as a proxy for solar installation repayment, shifting from a FICO score cutoff to the machine learning model increases profits by 34% to 1882% depending on the stringency used for evaluating potential customers.

Most solar companies currently use credit scores to determine whom to approve for solar installations. Despite their widespread use, credit scores consider many aspects of a consumer's credit history that are not directly related to utility payment; therefore, the

FICO score is an imperfect proxy for predicting utility payment performance. This implies that traditional credit score cutoffs exclude people with low credit scores and those with insufficient credit history, which disproportionately hurts low-to-moderate (LMI) income

households.

The goal of this research is: (1) to develop an alternative prediction model of default based on machine learning algorithms, specifically LASSO, SVM, and random forests; and (2) to compare its overall forecasting performance, as well as its implications for LMI consumers, to traditional credit metrics. We do so by developing a model that predicts the probability of non-delinquency of utility bill payments using a large data set of utility repayment and other financial data obtained from a credit reporting agency (CRA). We find that a traditional regression analysis using a small number of variables specific to utility repayment performance greatly increases accuracy and LMI inclusivity relative to FICO score, and that using machine learning techniques further enhances model performance.

A number of regression and machine learning techniques were used to predict utility bill delinquency. Among the variety of models that we explored, the random forest algorithm was clearly superior in terms of accuracy. Moreover, the random forest algorithm not only has better accuracy, but it also requires less data pre-processing. Finally, it is easier to interpret and runs more quickly.

The alternative scoring methods developed with traditional regression analysis and machine learning techniques were compared to standard FICO cutoffs, with a number of different metrics, including accuracy, default rate, and LMI inclusion.

For example, Figure 1 displays the probabilities of non-delinquency using the random forest algorithm against the individual's FICO Score. There are many individuals who have a high probability of non-delinquency with the random forest algorithm, but do not have a very high FICO score, which demonstrates the amount of people that would have been rejected with the FICO cutoff, but accepted according to the random forest algorithm ("false negatives"). Additionally, there are quite a few data points with high

FICO scores but do not have a very high probability with the random forest algorithm, who would be erroneously accepted ("false positives"). Figure 1 suggests that there are a high numbers of false negatives and false positives under traditional FICO scoring. Though the FICO Score is one variable used by the random forest algorithm, there are many other variables as well.

Importantly, the random forest algorithm, when tested with both 30 and 90 day definitions of delinquency, increase the number of LMI applicants approved. The random forest algorithm using a 30 day definition increases the number of LMI accounts approved by 11.4% to 14.0% depending on the stringency, while that using a 90 day definition increases LMI customers by 1.1% to 4.2%.
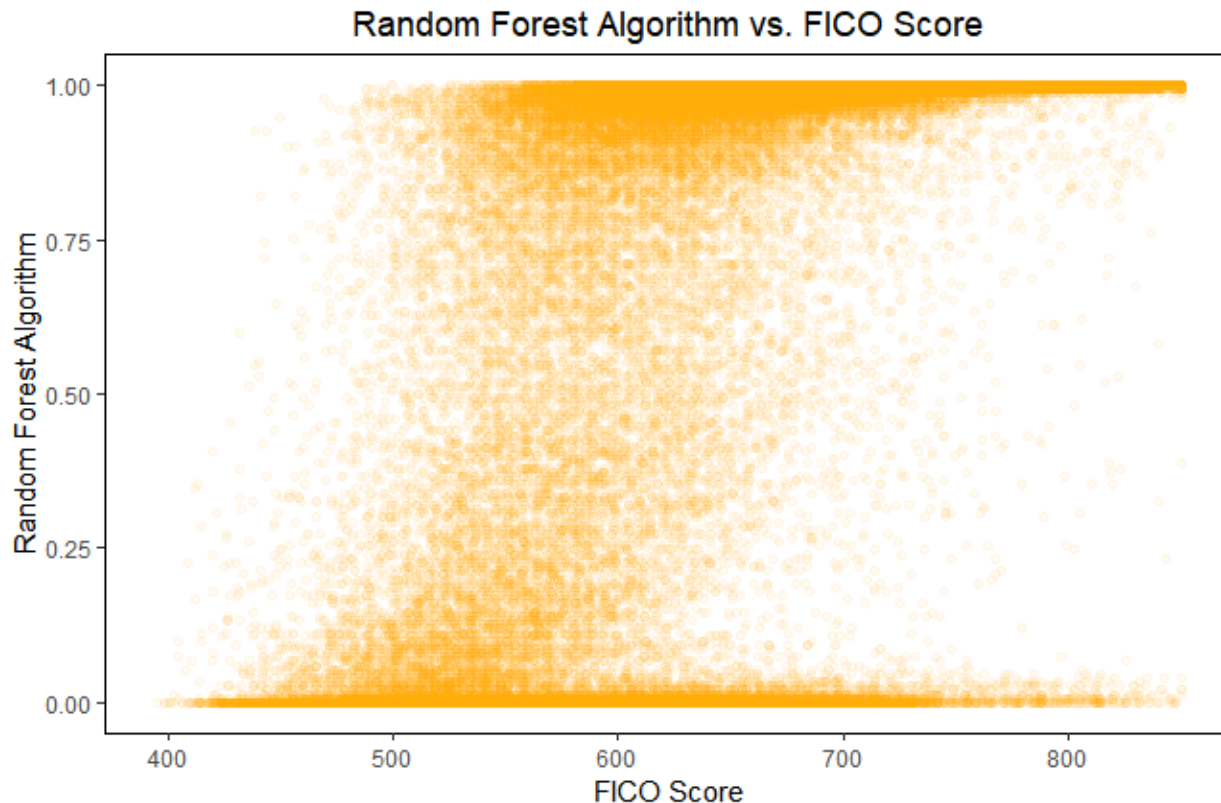
Finally, the impact of the alternative scoring methods on the profitability is estimated. The results shown in the paper demonstrate that the random forest algorithm leads to an increase in profits for the firm, which is a very significant result from our study. The random forest algorithm both benefits the customers, by accepting more LMI customers, and benefits the firms, by increasing profits.

We can decompose the increase in profits from the random forest algorithm to two sources. First is the increase in profits due to accepting new customers who would have been denied under the FICO score cutoff, or a decrease in false negatives (π from New Customers). Second is a reduction in losses from rejecting those who are accepted under the FICO Score cutoff but whom the random forest algorithm identifies as high-risk, or a decrease in false positives ("π from Less Delinquents").

Overall, the random forest algorithm improves accuracy when compared to the FICO Score, offers access to solar energy for more LMI customers, and leads to an increase in profits when compared to the FICO score cutoff, regardless of the stringency of the industry standard.

Random Forest Algorithm vs. FICO Score

References

Davuluri, Sruthi, René García Franceschini, Christopher R. Knittel, Chikara Onda, and Kelly Roache (2019) "Machine Learning for Solar Accessibility: Implications for Low-Income Solar Expansion and Profitability", MIT CEEPR Working Paper 2019-020.

About the Authors

Sruthi Davuluri recently earned her M.S. in the Technology and Policy Program at MIT, while working as a Research Assistant for Professor Christopher Knittel. While at CEEPR, Sruthi collaborated with solar startups, worked on a demand response project, and performed research on how electric distribution systems may adapt to heightened DER penetration. Sruthi is now a consultant at an energy consulting firm, E3, in San Francisco, CA.

About the Center for Energy and Environmental Policy Research (CEEPR)

Since 1977, CEEPR has been a focal point for research on energy and environmental policy at MIT. CEEPR promotes rigorous, objective research for improved decision making in government and the private sector, and secures the relevance of its work through close cooperation with industry partners from around the globe. CEEPR is jointly sponsored at MIT by the MIT Energy Initiative (MITEI), the Department of Economics, and the Sloan School of Management.

René García Franceschini is a current MIT Master's student in the Technology and Policy Program at the MIT Institute for Data, Systems, and Society. He is also working with Solstice, a company that aims to expand solar energy access. René is interested in combining renewable energy with social equity and social entrepreneurship.

Christopher R. Knittel is the George P. Shultz Professor of Applied Economics at the Sloan School of Management, Director of the Center for Energy and Environmental Policy Research, and Co-Director of the MITEI Low-Carbon Energy Center for Electric Power Systems Research at the Massachusetts Institute of Technology. He joined the faculty at MIT in 2011, having taught previously at UC Davis and Boston University. Professor Knittel received his B.A. in economics and political science from the California State University, Stanislaus in 1994 (summa cum laude), an M.A. in economics from UC Davis in 1996, and a Ph.D. in economics from UC Berkeley in 1999. His research focuses on environmental economics, industrial organization, and applied econometrics.

Chikara Onda is a Ph.D. candidate in the Emmett Interdisciplinary Program in Environment and Resources. His research focuses on the distributional impacts of climate and energy policies, with a focus on jobs and worker mobility. In his dissertation work, he looks at how factors such as industry-specific experience and education level might drive how readily workers can move across industries after being laid off and the wage cuts they might face in their new jobs. Chikara worked for the White House Council on Environmental Quality in 2016, where he worked on the U.S. Mid-Century Strategy for Deep Decarbonization, a long-term planning document submitted to the UNFCCC pursuant to the Paris Accord. He holds an MPA in Economics and Public Policy from the Woodrow Wilson School at Princeton University, and a BA in Economics and Environmental Science from Columbia.

Kelly Roache is a research and communications specialist at the Energy and Policy Institute. She has previously worked as an organizer and advocate for energy democracy and social justice in the US and internationally and as Director of Inclusion at Solstice.

While at Solstice, a community solar-focused social enterprise, Kelly led the development and implementation of alternative qualification and financing mechanisms to increase low-income and environmental justice communities' access to shared solar energy in New York and California.

About the Center for Energy and Environmental Policy Research (CEEPR)

Since 1977, CEEPR has been a focal point for research on energy and environmental policy at MIT. CEEPR promotes rigorous, objective research for improved decision making in government and the private sector, and secures the relevance of its work through close cooperation with industry partners from around the globe. CEEPR is jointly sponsored at MIT by the MIT Energy Initiative (MITEI), the Department of Economics, and the Sloan School of Management.