Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106

Machine Learning: An Applied Econometric Approach

Sendhil Mullainathan and Jann Spiess

Machines are increasingly doing "intelligent" things: Facebook recognizes faces in photos, Siri understands voices, and Google translates websites. The fundamental insight behind these breakthroughs is as much statistical as computational. Machine intelligence became possible once researchers stopped approaching intelligence tasks procedurally and began tackling them empirically. Face recognition algorithms, for example, do not consist of hard-wired rules to scan for certain pixel combinations, based on human understanding of what constitutes a face. Instead, these algorithms use a large dataset of photos labeled as having a face or not to estimate a function $f(x)$ that predicts the presence y of a face from pixels x. This similarity to econometrics raises questions: Are these algorithms merely applying standard techniques to novel and large datasets? If there are fundamentally new empirical tools, how do they fit with what we know? As empirical economists, how can we use them?[1]

We present a way of thinking about machine learning that gives it its own place in the econometric toolbox. Central to our understanding is that machine learning

[1] In this journal, Varian (2014) provides an excellent introduction to many of the more novel tools and "tricks" from machine learning, such as decision trees or cross-validation. Einav and Levin (2014) describe big data and economics more broadly. Belloni, Chernozhukov, and Hanson (2014) present an econometrically thorough introduction on how LASSO (and close cousins) can be used for inference in high-dimensional data. Athey (2015) provides a brief overview of how machine learning relates to causal inference.

■ *Sendhil Mullainathan is the Robert C. Waggoner Professor of Economics and Jann Spiess is a PhD candidate in Economics, both at Harvard University, Cambridge, Massachusetts. Their email addresses are mullain@fas.harvard.edu and jspiess@fas.harvard.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.31.2.87 doi=10.1257/jep.31.2.87

not only provides new tools, it solves a different problem. Machine learning (or rather "supervised" machine learning, the focus of this article) revolves around the problem of *prediction*: produce predictions of y from x. The appeal of machine learning is that it manages to uncover generalizable patterns. In fact, the success of machine learning at intelligence tasks is largely due to its ability to discover complex structure that was not specified in advance. It manages to fit complex and very flexible functional forms to the data without simply overfitting; it finds functions that work well out-of-sample.

Many economic applications, instead, revolve around *parameter estimation*: produce good estimates of parameters β that underlie the relationship between y and x. It is important to recognize that machine learning algorithms are not built for this purpose. For example, even when these algorithms produce regression coefficients, the estimates are rarely consistent. The danger in using these tools is taking an algorithm built for $\hat{y}$, and presuming their $\hat{\beta}$ have the properties we typically associate with estimation output. Of course, prediction has a long history in econometric research—machine learning provides new tools to solve this old problem.[2] Put succinctly, machine learning belongs in the part of the toolbox marked $\hat{y}$ rather than in the more familiar $\hat{\beta}$ compartment.

This perspective suggests that applying machine learning to economics requires finding relevant $\hat{y}$ tasks. One category of such applications appears when using new kinds of data for traditional questions; for example, in measuring economic activity using satellite images or in classifying industries using corporate 10-K filings. Making sense of complex data such as images and text often involves a prediction pre-processing step. In another category of applications, the key object of interest is actually a parameter β, but the inference procedures (often implicitly) contain a prediction task. For example, the first stage of a linear instrumental variables regression is effectively prediction. The same is true when estimating heterogeneous treatment effects, testing for effects on multiple outcomes in experiments, and flexibly controlling for observed confounders. A final category is in direct policy applications. Deciding which teacher to hire implicitly involves a prediction task (what added value will a given teacher have?), one that is intimately tied to the causal question of the value of an additional teacher.

Machine learning algorithms are now *technically* easy to use: you can download convenient packages in R or Python that can fit decision trees, random forests, or LASSO (Least Absolute Shrinkage and Selection Operator) regression coefficients. This also raises the risk that they are applied naively or their output is misinterpreted. We hope to make them *conceptually* easier to use by providing a crisper

[2] While the ideas we describe as central to machine learning may appear unfamiliar to some, they have their roots and parallels in nonparametric statistics, including nonparametric kernel regression, penalized modeling, cross-validation, and sieve estimation. We refer to Györfi, Kohler, Krzyzak, and Walk (2002) for a general overview, and to Hansen (2014) more specifically for counterparts in sieve estimation.

understanding of how these algorithms work, where they excel, and where they can stumble—and thus where they can be most usefully applied.[3]

How Machine Learning Works

Supervised machine learning algorithms seek functions that predict well out of sample. For example, we might look to predict the value y of a house from its observed characteristics x based on a sample of n houses (y_i, x_i). The algorithm would take a loss function $L(\hat{y}, y)$ as an input and search for a function $\hat{f}$ that has low expected prediction loss $E_{(y,x)}[L(\hat{f}(x), y)]$ on a *new* data point from the same distribution. Even complex intelligence tasks like face detection can be posed this way. A photo can be turned into a vector, say a 100-by-100 array so that the resulting x vector has 10,000 entries. The y value is 1 for images with a face and 0 for images without a face. The loss function $L(\hat{y}, y)$ captures payoffs from proper or improper classification of "face" or "no face."

Familiar estimation procedures, such as ordinary least squares, already provide convenient ways to form predictions, so why look to machine learning to solve this problem? We will use a concrete application—predicting house prices—to illustrate these tools. We consider 10,000 randomly selected owner-occupied units from the 2011 metropolitan sample of the American Housing Survey. In addition to the values of each unit, we also include 150 variables that contain information about the unit and its location, such as the number of rooms, the base area, and the census region within the United States. To compare different prediction techniques, we evaluate how well each approach predicts (log) unit value on a separate hold-out set of 41,808 units from the same sample. All details on the sample and our empirical exercise can be found in an online appendix available with this paper at http://e-jep.org.

Table 1 summarizes the findings of applying various procedures to this problem. Two main insights arise from this table. First, the table highlights the need for a hold-out sample to assess performance. In-sample performance may overstate performance; this is especially true for certain machine learning algorithms like random forests that have a strong tendency to overfit. Second, on out-of-sample performance, machine learning algorithms such as random forests can do significantly better than ordinary least squares, even at moderate sample sizes and with a limited number of covariates. Understanding machine learning, though, requires looking deeper than these quantitative gains. To make sense of how these

[3] This treatment is by no means exhaustive: First, we focus specifically on "supervised" machine learning where prediction is central, and do not discuss clustering or other "unsupervised" pattern recognition techniques. Second, we leave to more specialized sources the more hands-on practical advice, the discussion of computational challenges that are central to a computer-science treatment of the subject, and the overview of cutting-edge algorithms.

Table 1
Performance of Different Algorithms in Predicting House Values

	Prediction performance (R^2)		*Relative improvement over ordinary least squares by quintile of house value*				
Method	*Training sample*	*Hold-out sample*	1st	2nd	3rd	4th	5th
Ordinary least squares	47.3%	41.7% [39.7%, 43.7%]	–	–	–	–	–
Regression tree tuned by depth	39.6%	34.5% [32.6%, 36.5%]	−11.5%	10.8%	6.4%	−14.6%	−31.8%
LASSO	46.0%	43.3% [41.5%, 45.2%]	1.3%	11.9%	13.1%	10.1%	−1.9%
Random forest	85.1%	45.5% [43.6%, 47.5%]	3.5%	23.6%	27.0%	17.8%	−0.5%
Ensemble	80.4%	45.9% [44.0%, 47.9%]	4.5%	16.0%	17.9%	14.2%	7.6%

Note: The dependent variable is the log-dollar house value of owner-occupied units in the 2011 American Housing Survey from 150 covariates including unit characteristics and quality measures. All algorithms are fitted on the same, randomly drawn training sample of 10,000 units and evaluated on the 41,808 remaining held-out units. The numbers in brackets in the hold-out sample column are 95 percent bootstrap confidence intervals for hold-out prediction performance, and represent measurement variation for a fixed prediction function. For this illustration, we do not use sampling weights. Details are provided in the online Appendix at http://e-jep.org.

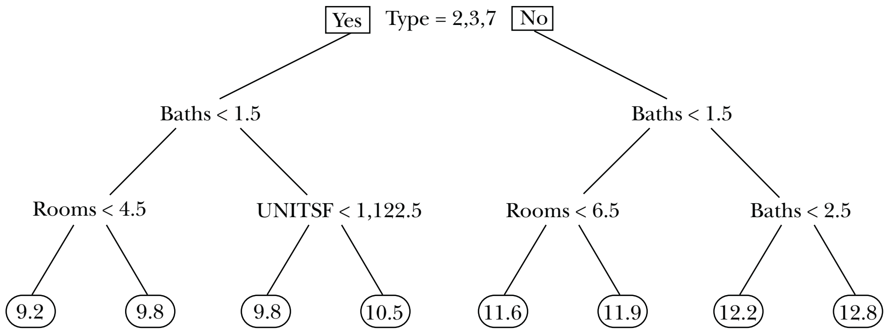
procedures work, we will focus in depth on a comparison of ordinary least squares and regression trees.

From Linear Least-Squares to Regression Trees

Applying ordinary least squares to this problem requires making some choices. For the ordinary least squares regression reported in the first row of Table 1, we included all of the main effects (with categorical variables as dummies). But why not include interactions between variables? The effect of the number of bedrooms may well depend on the base area of the unit, and the added value of a fireplace may be different depending on the number of living rooms. Simply including all pairwise interactions would be infeasible as it produces more regressors than data points (especially considering that some variables are categorical). We would therefore need to hand-curate which interactions to include in the regression. An extreme version of this challenge appears in the face-recognition problem. The functions that effectively combine pixels to predict faces will be highly nonlinear and interactive: for example, "noses" are only defined by complex interactions between numerous pixels.

Machine learning searches for these interactions automatically. Consider, for example, a typical machine learning function class: regression trees. Like a linear function, a regression tree maps each vector of house characteristics to a predicted

Figure 1
A Shallow Regression Tree Predicting House Values



Note: Based on a sample from the 2011 American Housing Survey metropolitan survey. House-value predictions are in log dollars.

value. The prediction function takes the form of a tree that splits in two at every node. At each node of the tree, the value of a single variable (say, number of bathrooms) determines whether the left (less than two bathrooms) or the right (two or more) child node is considered next. When a terminal node—a leaf—is reached, a prediction is returned. An example of a tree is given in Figure 1. We could represent the tree in Figure 1 as a linear function, where each of the leaves corresponds to a product of dummy variables ($x_1 = 1_{TYPE=2,3,7} \times 1_{BATHS<1.5} \times 1_{ROOMS<4.5}$ for the leftmost leaf) with the corresponding coefficient ($\alpha_1 = 9.2$). Trees are thus a highly interactive function class.

The Secret Sauce

How can a tree even be fitted here? A deep enough tree would fit perfectly—each observation would end up in its own leaf. That tree will have perfect *fit,* but of course this is really perfect *overfit:* out of sample, this tree would perform terribly for prediction. The (over)fitting conundrum is not specific to trees. The very appeal of machine learning is high dimensionality: flexible functional forms allow us to fit varied structures of the data. But this flexibility also gives so many possibilities that simply picking the function that fits best in-sample will be a terrible choice. So how does machine learning manage to do out-of-sample prediction?

The first part of the solution is *regularization*. In the tree case, instead of choosing the “best” overall tree, we could choose the best tree among those of a certain depth. The shallower the tree, the worse the in-sample fit: with many observations in each leaf, no one observation will be fit very well. But this also means there will be less overfit: the idiosyncratic noise of each observation is averaged out. Tree depth is an example of a regularizer. It measures the complexity of a function. As we regularize less, we do a better job at approximating the in-sample variation, but for the same reason, the wedge between in-sample and out-of-sample fit will typically increase. Machine learning algorithms typically have a regularizer associated with them. By choosing the level of regularization appropriately, we can have some benefits of flexible functional forms without having those benefits be overwhelmed by overfit.

How do we choose the optimal depth of the tree? In machine learning terms, how do we choose the level of regularization ("tune the algorithm")? This is the second key insight: *empirical tuning*. The essential problem of overfitting is that we would like the prediction function to do well *out of sample,* but we only fit in-sample. In empirical tuning, we create an out-of-sample experiment inside the original sample. We fit on one part of the data and ask which level of regularization leads to the best performance on the other part of the data.[4] We can increase the efficiency of this procedure through cross-validation: we randomly partition the sample into equally sized subsamples ("folds"). The estimation process then involves successively holding out one of the folds for evaluation while fitting the prediction function for a range of regularization parameters on all remaining folds. Finally, we pick the parameter with the best estimated average performance.[5] The second row of Table 1 shows the performance of a regression tree where we have chosen depth in this way.

This procedure works because prediction quality is observable: both predictions $\hat{y}$ and outcomes y are observed. Contrast this with parameter estimation, where typically we must rely on assumptions about the data-generating process to ensure consistency. Observability by itself would not make prediction much easier since the algorithm would still need to sort through a very large function class. But regularization turns this choice into a low-dimensional one—we only need to choose the best tuning parameter. Regularization combines with the observability of prediction quality to allow us to fit flexible functional forms and still find generalizable structure.

Most of Machine Learning in One Expression[6]

This structure—regularization and empirical choice of tuning parameters—helps organize the sometimes bewildering variety of prediction algorithms that one encounters. There is a function class $\mathcal{F}$ (in this case, trees) and a regularizer $R(f)$ (in the specific example, depth of tree) that expresses the complexity of a function

[4] One approach to the tuning problem is deriving the optimal level of regularization analytically for each procedure and under assumptions on the sampling process, such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and SURE (Stein's Unbiased Risk Estimate). This theoretical guidance is helpful when available and applicable, but assumptions may prove hard to verify as the reason for undertaking nonparametric analysis may be that we are unsure about features of the data-generating processes in the first place. In other cases, theoretical results give only asymptotic guidance that remain an unverifiable promise in finite samples.

[5] In some cases, the researcher will adjust the empirical loss minimizer to account for measurement error and/or sample size differences in mapping observed performance to the level of regularization. An example is the "one standard-error rule" for LASSO tuning discussed in Hastie, Tibshirani, and Friedman (2009).

[6] We adapted the title of this section from a post on Francis X. Diebold's "No Hesitations" blog, http://fxdiebold.blogspot.com/2017/01/all-of-machine-learning-in-one.html.

Table 2
Some Machine Learning Algorithms

Function class $\mathcal{F}$ (and its parametrization)	*Regularizer $R(f)$*
Global/parametric predictors	
Linear $\beta' x$ (and generalizations)	Subset selection $\|\|\beta\|\|_0 = \sum_{j=1}^{k} 1_{\beta_j \neq 0}$
	LASSO $\|\|\beta\|\|_1 = \sum_{j=1}^{k} \|\beta_j\|$
	Ridge $\|\|\beta\|\|_2^2 = \sum_{j=1}^{k} \beta_j^2$
	Elastic net $\alpha\|\|\beta\|\|_1 + (1-\alpha)\|\|\beta\|\|_2^2$
Local/nonparametric predictors	
Decision/regression trees	Depth, number of nodes/leaves, minimal leaf size, information gain at splits
Random forest (linear combination of trees)	Number of trees, number of variables used in each tree, size of bootstrap sample, complexity of trees (see above)
Nearest neighbors	Number of neighbors
Kernel regression	Kernel bandwidth
Mixed predictors	
Deep learning, neural nets, convolutional neural networks	Number of levels, number of neurons per level, connectivity between neurons
Splines	Number of knots, order
Combined predictors	
Bagging: unweighted average of predictors from bootstrap draws	Number of draws, size of bootstrap samples (and individual regularization parameters)
Boosting: linear combination of predictions of residual	Learning rate, number of iterations (and individual regularization parameters)
Ensemble: weighted combination of different predictors	Ensemble weights (and individual regularization parameters)

(more precisely the complexity of its representation).[7] Picking the prediction function then involves two steps: The first step is, conditional on a level of complexity, to pick the best in-sample loss-minimizing function.[8] The second step is to estimate the optimal level of complexity using empirical tuning (as we saw in cross-validating the depth of the tree). In Table 2, we give an incomplete overview of methods that follow this pattern.

[7] We write the regularizer as a mapping from the function itself. In cases where functions are not uniquely parametrized (and for practical purposes in many applications), we implicitly refer to a set of parameters that define a function for a given parametrization of the function class. Also, the complexity itself may be estimated from the training data.

[8] We summarize this central step in the expression

$$\text{minimize} \underbrace{\textstyle\sum_{i=1}^{n} L(f(x_i), y_i}_{\text{in-sample loss}}, \text{ over } \overbrace{f \in F}^{\text{function class}} \text{ subject to } \underbrace{R(f) \leq c}_{\text{complexity restriction}}.$$

For example, in our framework, the LASSO (probably the machine learning tool most familiar to economists) corresponds to 1) a quadratic loss function, 2) a class of linear functions (over some fixed set of possible variables), and 3) a regularizer which is the sum of absolute values of coefficients.[9] This effectively results in a linear regression in which only a small number of predictors from all possible variables are chosen to have nonzero values: the absolute-value regularizer encourages a coefficient vector where many are exactly zero. The third row of Table 1 shows the performance of LASSO in predicting house prices. Ridge regression is a close cousin: it simply uses a quadratic regularizer instead of the sum of absolute values.

In some of the most successful machine learning methods, multiple predictors from the same function class are combined into a single prediction function and tuned jointly. The fourth row in Table 1 shows the performance of a random forest; it outperforms ordinary least squares on the hold-out by over 9 percent in terms of overall R^2. The random forest is an average over many (in this case, 700) trees. Each tree is fitted on a bootstrap sample of the original training set and constrained to a randomly chosen subset of variables. The predictions of the trees are then averaged. The regularization variables in this algorithm include the complexity of each individual tree (such as its depth), the number of variables used in each tree, the size of each bootstrap sample, and the number of trees.

The last row in Table 1 lists an ensemble method that runs several separate algorithms (in this case tree, LASSO, and forest) and then averages their predictions, with weights chosen by cross-validation. The fact that the ensemble comes out on top in Table 1—with an out-of-sample R^2 of almost 46 percent—is no isolated case. While it may be unsurprising that such ensembles perform well *on average*—after all, they can cover a wider array of functional forms—it may be more surprising that they come on top in virtually *every* prediction competition.

Other models that we have not estimated in our data also fit this framework. For example, neural nets are popular prediction algorithms for image recognition tasks. For one standard implementation in binary prediction, the underlying function class is that of nested logistic regressions: The final prediction is a logistic transformation of a linear combination of variables ("neurons") that are themselves such logistic transformations, creating a layered hierarchy of logit regressions. The complexity of these functions is controlled by the number of layers, the number of neurons per layer, and their connectivity (that is, how many variables from one level enter each logistic regression on the next).

Econometric Guidance

Viewed this way, there are several choices to be made when using a machine learning approach. First, this approach involves choosing the functions we fit and how we regularize them: Should I use a regression tree or linear functions? If I choose a tree, do I express its complexity by its depth, the minimal number of units

[9] For some readers, a more familiar equation for the LASSO is the Lagrangian dual formulation, where the Lagrange multiplier λ plays the role of the tuning parameter.

in each leaf, or the minimal improvement in prediction quality at every split? Available guidance in the machine learning literature is largely based on a combination of simulation studies and expert intuition. They are complemented by recent theoretical results in econometrics that shed light on the comparative performance of different regularizers, such as Abadie and Kasy (2017) for LASSO and close relatives.

Practically, one must decide how to encode and transform the underlying variables. In our example of house prices, do we include base area *per room* as a variable, or just total area? Should we use logarithmic scales? Normalize to unit variances? These choices about how to represent the features will interact with the regularizer and function class: A linear model can reproduce the log base area per room from log base area and log room number easily, while a regression tree would require many splits to do so. In a traditional estimator, replacing one set of variables by a set of transformed variables from which it could be reconstructed would not change the predictions, because the set of functions being chosen from has not changed. But with regularization, including these variables can improve predictions because—at any given level of regularization—the set of functions might change. If the number of bathrooms *per bedroom* is what we suspect will matter in the price-setting process, creating that variable explicitly lowers the complexity cost for including this variable. Economic theory and content expertise play a crucial role in guiding where the algorithm looks for structure first. This is the sense in which "simply throw it all in" is an unreasonable way to understand or run these machine learning algorithms. For example, in visual tasks, some understanding of geometry proves crucial in specifying the way in which neurons are connected within neural nets.

A final set of choices revolves around the tuning procedure: Should out-of-sample performance be estimated using some known correction for overfitting (such as an adjusted R^2 when it is available) or using cross-validation? How many folds should be used in cross-validation, and how exactly should the final tuning parameter be chosen? While asymptotic results show that cross-validation tuning approximates the optimal complexity (Vaart, Dudoit, and Laan 2006), available finite-sample guidance on its implementation—such as heuristics for the number of folds (usually five to ten) or the "one standard-error rule" for tuning the LASSO (Hastie, Tibshirani, and Friedman 2009)—has a more ad-hoc flavor. Design choices must be made about function classes, regularizers, feature representations, and tuning procedures: there are no definitive and universal answers available. This leaves many opportunities for econometric research.

Quantifying Predictive Performance

While these design choices leave plenty of freedom, having a reliable estimate of predictive performance is a nonnegotiable requirement for which strong econometric guarantees are available. In the house-price example, we divide the sample into a training and a test (hold-out) sample. This implements a *firewall principle*: none of the data involved in fitting the prediction function—which includes cross-validation to tune the algorithm—is used to evaluate the prediction function that is produced. As a result, inference on predictive performance of a fixed predictive function is a straightforward task of mean estimation: the distribution of realized loss in the hold-out (taking any clustering into account) directly yield consistent estimates of performance and confidence intervals.

Econometric theory plays a dual role here. First, econometrics can guide design choices, such as the number of folds or the function class. Guidance in these choices can help improve prediction quality and the power of any test based on it. Second, given the fitted prediction function, it must enable us to make inferences about estimated fit. The hold-out sample exactly allows us to form properly sized tests about predictive properties of the fitted function.

What Do We (Not) Learn from Machine Learning Output?

It is tempting to do more with the fitted function. Why not also use it to learn something about the "underlying model": specifically, why not use it to make inferences about the underlying data-generating process? Even if correlations are not causal, might they not reveal useful underlying structure? The LASSO regression of Table 1 ends up not using the number of dining rooms as a right-hand variable. Does that reveal that the number of dining rooms is an unimportant variable in determining house prices (given the other available variables)? It is tempting to draw such conclusions, and such conclusions could be economically meaningful: for example, in predicting wages, the weight placed on race by a machine learning algorithm seems like it could be a proxy for discrimination. Statistical packages contribute to these inferences by outputting measures of variable importance in the fitted functions.

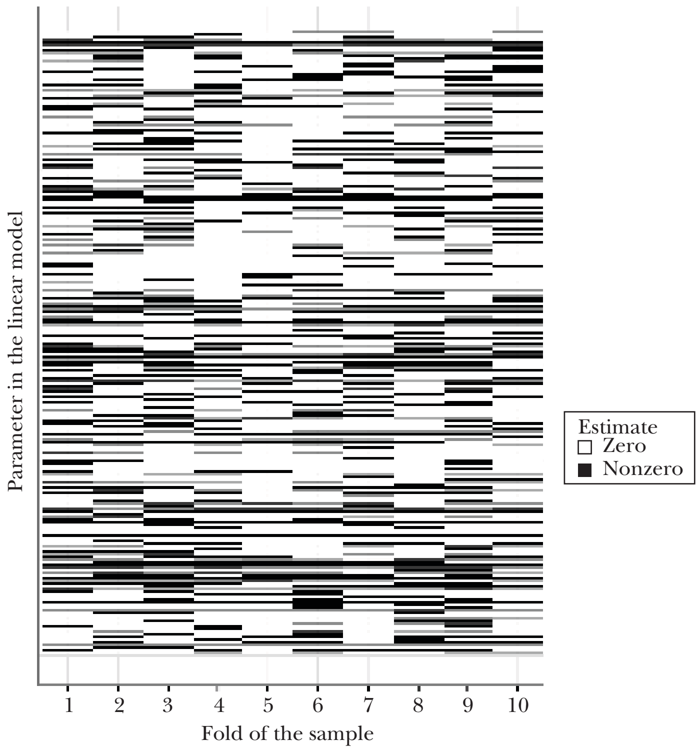
One obvious problem that arises in making such inferences is the lack of standard errors on the coefficients. Even when machine-learning predictors produce familiar output like linear functions, forming these standard errors can be more complicated than seems at first glance as they would have to account for the model selection itself. In fact, Leeb and Pötscher (2006, 2008) develop conditions under which it is impossible to obtain (uniformly) consistent estimates of the distribution of model parameters after data-driven selection.

But there is an even bigger challenge. To illustrate the problem, we repeated the house-value prediction exercise on subsets of our sample from the American Housing Survey. First, we randomly cut the sample into ten partitions of approximately 5,000 units each. On each partition, we re-estimate the LASSO predictor. Through its regularizer, LASSO produces a sparse prediction function, so that many coefficients are zero and are "not used"—in this example, we find that more than half the variables are unused in each run.

Figure 2 shows how the variables that are used vary from partition to partition. Each row represents one of x variables used. Each column represents a different partition. We color each cell black if that variable is used by the LASSO model in that partition. Figure 2 documents a fundamental problem: a variable used in one partition may be unused in another. In fact, there are few stable patterns overall.

These instabilities do not reflect instability in prediction quality—in fact, the R^2 remains roughly constant from partition to partition. The problem arises because if

Figure 2
Selected Coefficients (Nonzero Estimates) across Ten LASSO Regressions



Note: We repeated the house-value prediction exercise on subsets of our sample from the American Housing Survey. First, we randomly cut the sample into ten partitions of approximately 5,000 units each. On each partition, we re-estimate the LASSO predictor, with LASSO regularization parameter fixed. The figure shows how the variables that are used vary from partition to partition. Each row represents one of *x* variables used. Each column represents a different partition. We color each cell black if that variable is used by the LASSO model (has a nonzero coefficient) in that partition. The figure documents a fundamental problem: a variable used in one partition may be unused in another. In fact, there are few stable patterns overall. For details, see discussion in text and online appendix available with this paper at http://e-jep.org.

the variables are correlated with each other (say the number of rooms of a house and its square-footage), then such variables are substitutes in predicting house prices. Similar predictions can be produced using very different variables. Which variables are actually chosen depends on the specific finite sample. In traditional estimation, correlations between observed variables would be reflected in large standard errors that express our uncertainty in attributing effects to one variable over the other.

This problem is ubiquitous in machine learning. The very appeal of these algorithms is that they can fit many different functions. But this creates an Achilles' heel: more functions mean a greater chance that two functions with very different coefficients can produce similar prediction quality. As a result, how an algorithm chooses between two very different functions can effectively come down to the flip of a coin. In econometric terms, while the lack of standard errors illustrates the limitations to making inference *after* model selection, the challenge here is (uniform) model selection consistency itself.

Regularization also contributes to the problem. First, it encourages the choice of less complex, but wrong models. Even if the best model uses interactions of number of bathrooms with number of rooms, regularization may lead to a choice of a simpler (but worse) model that uses only number of fireplaces. Second, it can bring with it a cousin of omitted variable bias, where we are typically concerned with correlations between observed variables and unobserved ones. Here, when regularization excludes some variables, even a correlation between observed variables and other *observed* (but excluded) ones can create bias in the estimated coefficients.

Recovering Structure: Estimation ($\hat{\beta}$) vs Prediction ($\hat{y}$)

We face a challenge. On the one hand, these machine learning algorithms by their very construction—tuning and evaluation out of sample—seek a generalizable structure and are evaluated on their capacity to find it. These algorithms do detect structure in $\hat{y}$: when predictive quality is high, some structure must have been found. Some econometric results also show the converse: when there is structure, it will be recovered at least asymptotically (for example, for prediction consistency of LASSO-type estimators in an approximately sparse linear framework, see Belloni, Chernozhukov, and Hansen 2011). On the other hand, we have seen the dangers of naively interpreting the estimated $\hat{\beta}$ parameters as indicating the discovered structure.

Of course, assumptions about the data-generating process would allow us to take the estimated $\hat{\beta}$ parameters more literally. The discussion above suggests that we must limit the correlations between the observed variables. This is seen clearly in Zhao and Yu (2006) who establish asymptotic model-selection consistency for the LASSO. Besides assuming that the true model is "sparse"—only a few variables are relevant—they also require the "irrepresentable condition" between observables: loosely put, none of the irrelevant covariates can be even moderately related to the set of relevant ones.

In practice, these assumptions are strong. The instability in Figure 2, for example, suggests that they are not realistic in the house price example. But since we know this model is finding some structure, can we characterize it? A key area of future research in econometrics and machine learning is to make sense of the estimated prediction function without making strong assumptions about the underlying true world.

How Machine Learning Can Be Applied

Our starting point for applications of machine learning algorithms is guided by both the strength of machine learning—it provides a powerful, flexible way of making quality predictions—and its weakness: absent strong and mostly unverifiable assumptions, machine learning does not produce stable estimates of the underlying parameters. Therefore, we look for $\hat{y}$ problems, places where improved prediction has large applied value.

New Data

The phrase "big data" emphasizes a change in the scale of data. But there has been an equally important change in the *nature* of this data. Machine learning can deal with unconventional data that is too high-dimensional for standard estimation methods, including image and language information that we conventionally had not even thought of as data we can work with, let alone include in a regression.

Satellites have been taking images of the earth for decades, which we can now use not just as pixelated vectors, but as economically meaningful input. Donaldson and Storeygard (in this journal, 2016) provide an overview of the growing literature in economics using satellite data, including how luminosity at night correlates with economic output (Henderson, Storeygard, and Weil 2012) or estimating future harvest size (Lobell 2013). Satellite images do not directly contain, for example, measures of crop yield. Instead, they provide us with a large x vector of image-based data; these images are then matched (in what we hope is a representative sample) to yield data which form the y variable. This translation of satellite images to yield measures is a prediction problem. Machine learning is the essential tool by which we extract and scale economically meaningful signals from this data.

These new sources of data are particularly relevant where reliable data on economic outcomes are missing, such as in tracking and targeting poverty in developing countries (Blumenstock 2016). Jean et al. (2016) train a neural net to predict local economic outcomes from satellite data in five African countries. Machine learning also yields economic predictions from large-scale network data; for example, Blumenstock, Cadamuro, and On (2015) use cell-phone data to measure wealth, allowing them to quantify poverty in Rwanda at the individual level. Image recognition can of course be used beyond satellite data, and localized prediction of economic outcomes is relevant beyond the developing world: as one example, Glaeser, Kominers, Luca, and Naik (2016) use images from Google Street View to measure block-level income in New York City and Boston.

Language provides another new powerful source of data. As with satellite images, online posts can be made meaningful by labeling them with machine learning. Kang, Kuznetsova, Luca, and Choi (2013) use restaurant reviews on Yelp.com to predict the outcome of hygiene inspections. Antweiler and Frank (2004) classify text on online financial message boards as bullish, bearish, or neither. Their algorithm trains on a small number of manual classifications, and scales these labels up to 1.5 million messages as a basis for the subsequent analysis, which shows that online messages help explain market volatility, with statistically significant, if economically modest, effects on stock returns.

Financial economists rely heavily on corporate financial information, such as that available in Compustat. But companies release detailed reports on their financial positions above and beyond these numbers. In the United States, publicly traded companies must file annual 10-K forms. Kogan, Levin, Routledge, Sagi, and Smith (2009) predict volatility of roughly 10,000 such firms from market-risk disclosure text within these forms, and show that it adds significant predictive information to past volatility. Hoberg and Phillips (2016) extract similarity of firms from their 10-K business description texts, generating new time-varying industry classifications for these firms.

Machine learning can be useful in preprocessing and imputing even in traditional datasets. In this vein, Feigenbaum (2015a, b) applies machine-learning classifiers to match individuals in historical records: he links fathers and sons across censuses and other data sources, which allows him to quantify social mobility during the Great Depression. Bernheim, Bjorkegren, Naecker, and Rangel (2013) link survey responses to observable behavior: A subset of survey respondents take part in a laboratory experiment; a machine learning algorithm trained on this data predicts actual choices from survey responses, giving economists a tool to infer actual from reported behavior.

Prediction in the Service of Estimation

A second category of application is in tasks that we approach as estimation problems. Even when we are interested in a parameter $\hat{\beta}$, the tool we use to recover that parameter may contain (often implicitly) a prediction component. Take the case of linear instrumental variables understood as a two-stage procedure: first regress $x = \gamma' z + \delta$ on the instrument z, then regress $y = \beta' x + \epsilon$ on the fitted values $\hat{x}$. The first stage is typically handled as an estimation step. But this is effectively a prediction task: only the predictions $\hat{x}$ enter the second stage; the coefficients in the first stage are merely a means to these fitted values.

Understood this way, the finite-sample biases in instrumental variables are a consequence of overfitting. Overfitting means that the in-sample fitted values $\hat{x}$ pick up not only the signal $\gamma' z$, but also the noise δ. As a consequence, $\hat{x}$ is biased towards x, and the second-stage instrumental variable estimate $\hat{\beta}$ is thus biased towards the ordinary least squares estimate of y on x. Since overfit will be larger when sample size is low, the number of instruments is high, or the instruments are weak, we can see why biases arise in these cases (Bound, Jaeger, and Baker 1995; Bekker 1994; Staiger and Stock 1997).

This analogy carries through to some of the classical solutions to finite-sample bias. Above, we used hold-out sets (in evaluating the prediction function) or cross-validation (in choosing the tuning parameter) to separate the data used in the fitting of the function from the data used in the forming of predicted values; this ensured, for example, that our evaluations of a function's prediction quality were unbiased. These same techniques applied here result in split-sample instrumental variables (Angrist and Krueger 1995) and "jackknife" instrumental variables (Angrist, Imbens, and Krueger 1999). Overfitting has wider consequences: the flipside of excessive in-sample overfitting is bad out-of-sample prediction. In fact, predicting well requires managing overfitting, which was the goal of both regularization and empirical tuning. These techniques are applicable to the first stage of instrumental variable analysis as well. In particular, a set of papers has already introduced regularization into the first stage in a high-dimensional setting, including the LASSO (Belloni, Chen, Chernozhukov, and Hansen 2012) and ridge regression (Carrasco 2012; Hansen and Kozbur 2014). More recent extensions include nonlinear functional forms, all the way to neural nets (Hartford, Leyton-Brown, and Taddy 2016).

Practically, even when there appears to be only a few instruments, the problem is effectively high-dimensional because there are many degrees of freedom in how instruments are actually constructed. For example, several papers use college proximity as an instrument in estimating returns to schooling (for example, Card 1999, Table 4). How exactly college proximity is used, however, varies. After all, it can be included linearly, logarithmically, or as dummies (and if so, with different thresholds) and can be interacted with other variables (such as demographic groups most likely to be affected). The latitude in making these choices makes it even more valuable to consider the first stage as a prediction problem. It allows us to let the data explicitly pick effective specifications, and thus allows us to recover more of the variation and construct stronger instruments, provided that predictions are constructed and used in a way that preserves the exclusion restriction.[10]

Many other inference tasks also have a prediction problem implicit inside them. In controlling for observed confounders, we do not care about the parameters associated with the control variables as an end in themselves. For example, Lee, Lessler, and Stuart (2010) use machine-learning algorithms to estimate the propensity score. Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, and Newey (2016) take care of high-dimensional controls in treatment effect estimation by solving two simultaneous prediction problems, one in the outcome and one in the treatment equation.

Similar opportunities arise even in cases where we have experimental data. Consider the problem of verifying balance between treatment and control groups (such as when there is attrition). Or consider the seemingly different problem of testing for effects on many outcomes. Both can be viewed as prediction problems (Ludwig, Mullainathan, and Spiess 2017). If treatment assignment can be predicted better than chance from pretreatment covariates, this is a sign of imbalance. If treatment assignment can be predicted from a set of outcomes, the treatment must have had an effect. Estimating heterogeneous treatment effects can also be viewed as a prediction problem, though the parallel is nonobvious and implementing the transformation is a major contribution of the papers in this literature. Typically, heterogeneous treatment effects might be estimated as coefficients on interaction terms in a linear regression. Consider instead the prediction task of mapping unit-level attributes to individual effect estimates. Of course, individual-level treatment effects are not directly observed. Despite this, machine-learning methods have been successfully applied to map out treatment effect heterogeneity. Athey and Imbens (2016) use sample-splitting to obtain valid (conditional) inference on

[10] In particular, we have to avoid "forbidden regressions" (Angrist and Pischke 2008) in which correlation between first-stage residuals and fitted values exists and creates bias in the second stage.

treatment effects that are estimated using decision trees, as previously suggested by Imai and Strauss (2011). Wager and Athey (2015) extend the construction to random forests, while Grimmer, Messing, and Westwood (2016) employ ensemble methods. These heterogenous treatment effects can be used to assign treatments; Misra and Dubé (2016) illustrate this on the problem of price targeting, applying Bayesian regularized methods to a large-scale experiment where prices were randomly assigned.

Expressing parts of these inference tasks as prediction problems also makes clear the limitation on their output. A carefully constructed heterogeneity tree provides valid estimates of treatment effects in every leaf (Athey and Imbens 2016). At the same time, one must be careful in interpreting the particular representation of the chosen tree, as it may suffer from instabilities similar to those in the American Housing Survey data above. Suppose the algorithm chooses a tree that splits on education but not on age. Conditional on this tree, the estimated coefficients are consistent. But that does not imply that treatment effects do not also vary by age, as education may well covary with age; on other draws of the data, in fact, the same procedure could have chosen a tree that split on age instead.

Prediction in Policy

Consider the following policy problem: Shortly after arrest, a judge must decide where defendants will wait for trial, at home or in jail. This decision, by law, must be based on a prediction by the judge: If released, will the defendant return for their court appearance or will they skip court, and will they potentially commit further crimes? Statistical tools have helped improve policies in several ways (such as randomized control trials helping to answer "does the policy work?"). In this case, one might wonder whether a predictive algorithm could similarly help improve the judge's decision (Kleinberg et al. 2017).

Prediction policy problems, such as the bail problem, appear in many areas (Kleinberg, Ludwig, Mullainathan, and Obermeyer 2015). For instance, a large literature estimates the impact of hiring an additional teacher—this is meant to inform the decision of whether to hire more teachers. The decision of *which* teacher to hire, however, requires a prediction: the use of information available at time of hiring to forecast individual teacher quality (Kane and Staiger 2008; Dobbie 2011; Jacob et al. 2016). Chalfin et al. (2016) provide some preliminary evidence of how machine learning may improve predictive accuracy in these and other personnel decisions. Chandler, Levitt, and List (2011) predict highest-risk youth so that mentoring interventions can be appropriately targeted. Abelson, Varshney, and Sun (2014), McBride and Nichols (2016), and Engstrom, Hersh, and Newhouse (2016) use machine learning to improve poverty targeting relative to existing poverty scorecards. These predictive problems intimately relate to questions we already seek to answer: the impact of an extra teacher depends on how that teacher is chosen; the impact of a transfer program depends on how well targeted it is. Given the active research contributing to these policy discussions, expanding the focus to these adjacent prediction questions seems promising.

Economists can play a crucial role in solving prediction policy problems. First, even though prediction is important, machine learning is not enough: familiar econometric challenges arise. In deciding whether an algorithm could improve on the judge, one must resolve a basic counterfactual issue: we only know the crimes committed by those released. Many predictions problems share the feature that the available data is dictated by existing decision rules, and insights from the causal inference could prove helpful in tackling these problems; for example, Kleinberg et al. (2017) use pseudo-random assignment to judges of differing leniency in their application. Second, behavioral issues arise. Even when an algorithm can help, we must understand the factors that determine adoption of these tools (Dawes, Faust, and Meehl 1989; Dietvorst, Simmons, and Massey 2015; Yeomans, Shah, Mullainathan, and Kleinberg 2016). What factors determine faith in the algorithm? Would a simpler algorithm be more believed? How do we encourage judges to use their private information optimally? These questions combine problems of technology diffusion, information economics, and behavioral economics.

Testing Theories

A final application of supervised machine learning is to test directly theories that are inherently about predictability. Within the efficient markets theory in finance, for example, the inability to make predictions about the future is a key prediction. Moritz and Zimmermann (2016) adapt methods from machine learning to show that past returns of US firms do have significant predictive power over their future stock prices.

Machine learning can also be used to construct a benchmark for how well theories are performing. A common concern is that even if a theory is accurate, it may only clear up a little of the systematic variation it aims to explain. The R^2 alone does not address this question, as some of the total variance may not be explainable from what is measured. Kleinberg, Liang, and Mullainathan (2015) propose to compare the predictive power of a theory to that of an optimal predictor. Relatedly, Peysakhovich and Naecker (2015) compare the out-of-sample performance of behavioral economics models for choices under risk and ambiguity to an atheoretical machine-learning benchmark.

Conclusion

The problem of artificial intelligence has vexed researchers for decades. Even simple tasks such as digit recognition—challenges that we as humans overcome so effortlessly—proved extremely difficult to program. Introspection into how our mind solves these problems failed to translate into procedures. The real breakthrough came once we stopped trying to deduce these rules. Instead, the problem was turned into an inductive one: rather than hand-curating the rules, we simply let the data tell us which rules work best.

For empiricists, these theory- and data-driven modes of analysis have always coexisted. Many estimation approaches have been (often by necessity) based on top-down, theory-driven, deductive reasoning. At the same time, other approaches have aimed to simply let the data speak. Machine learning provides a powerful tool to hear, more clearly than ever, what the data have to say.

These approaches need not be in conflict. A structural model of poverty, for example, could be applied to satellite image data processed using machine learning. Theory could guide what variables to manipulate in an experiment; but in analyzing the results, machine learning could help manage multiple outcomes and estimate heterogeneous treatment effects.

In the long run, new empirical tools have also served to expand the kinds of problems we work on. The increased use of randomized control trials has also changed the kinds of questions empirical researchers work on. Ultimately, machine learning tools may also increase the scope of our work—not just by delivering new data or new methods but by focusing us on new questions.

■ *We are grateful to the editors Gordon Hanson, Enrico Moretti, and Timothy Taylor for their feedback and to Alberto Abadie, Peter Ganong, Lawrence Katz, Supreet Kaur, Jens Ludwig, and Andrei Shleifer for thoughtful and detailed comments.*

References

Abadie, Alberto, and Maximilian Kasy. 2017. *The Risk of Machine Learning.* https://ideas.repec.org/p/qsh/wpaper/383316.html.

Abelson, Brian, Kush R. Varshney, and Joy Sun. 2014. "Targeting Direct Cash Transfers to the Extremely Poor." In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 1563–72. ACM.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14(1): 57–67

Angrist, Joshua D., and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business and Economic Statistics* 13(2): 225–35.

Angrist, Joshua D., and Jörn-Steffen S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Antweiler, Werner, and Murray Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance* 59(3): 1259–94.

Athey, Susan. 2015. "Machine Learning and Causal Inference for Policy Evaluation." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 5–6. ACM.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects. *PNAS* 113(27): 7353–60.

Bekker, Paul A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62(3): 657–81.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80(6): 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2011. "Inference for High-Dimensional Sparse Econometric Models." arXiv:1201.0220.

Belloni, Alexandre, Victor Chernozhukov,

and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *Review of Economic Studies* 81(2): 608–650.

Bernheim, Douglas, Daniel Bjorkegren, Jeffrey Naecker, and Anatonio Rangel. 2013. "Non-Choice Evaluations Predict Behavioral Responses to Changes in Economic Conditions." NBER Working Paper 19269.

Blumenstock, Joshua E., Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350(6264): 1073–76.

Blumenstock, Joshua Evan. 2016. "Fighting Poverty with Data." *Science* 353(6301): 753–54.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90(430): 443–50.

Card, David. 1999. "The Causal Effect of Education on Earnings." *Handbook of Labor Economics*, vol. 3A, edited by Orley C. Ashenfelter and David Card, 1801–1863. North-Holland, Elsevier.

Carrasco, Marine 2012. "A Regularization Approach to the Many Instruments Problem." *Journal of Econometrics* 170(2): 383–98.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and Selection of Human Capital with Machine Learning." *American Economic Review* 106(5): 124–27.

Chandler, Dana, Steven D. Levitt, and John A. List. 2011. "Predicting and Preventing Shootings among At-Risk Youth." *American Economic Review* 101(3): 288–92.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. 2016. "Double Machine Learning for Treatment and Causal Parameters." arXiv:1608.00060.

Dawes, Robyn M., David Faust, and Paul E. Meehl. 1989. "Clinical versus Actuarial Judgment." *Science* 243(4899): 1668–74.

Dietvorst, Berkeley J., Joseph P Simmons, and Cade Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err." *Journal of Experimental Psychology: General* 144(1): 114–26.

Dobbie, Will. 2011. "Teacher Characteristics and Student Achievement: Evidence from Teach For America." https://www.princeton.edu/~wdobbie/files/dobbie_tfa_2011.pdf.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30(4): 171–98.

Engstrom, Ryan, Jonathan Hersh, and David Newhouse. 2016. "Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being and Geographic Targeting." Unpublished paper.

Einav, Liran, and Jonathan Levin. 2014. "Economics in the Age of Big Data." *Science* 346(6210): 1243089.

Feigenbaum, James J. 2015a. "Automated Census Record Linking." http://scholar.harvard.edu/jfeigenbaum/publications/automated-census-record-linking.

Feigenbaum, James J. 2015b. "Intergenerational Mobility during the Great Depression." http://scholar.harvard.edu/jfeigenbaum/publications/jmp.

Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nakhil Naik. 2016. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." *Economic Inquiry*, Early View Article, online July 12.

Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2016. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods." https://stanford.edu/~jgrimmer/het.pdf.

Györfi, L., M. Kohler, A. Krzyzak, and H. Walk. 2002. *A Distribution-Free Theory of Nonparametric Regression*. Springer.

Hansen, Bruce E. 2014. "Nonparametric Sieve Regression: Least Squares, Averaging Least Squares, and Cross-Validation." In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics,* edited by Jeffrey S. Racine, Liangjun Su, and Aman Ullah. Oxford Univerity Press.

Hansen, Christian, and Damian Kozbur. 2014. "Instrumental Variables Estimation with Many Weak instruments using Regularized JIVE." *Journal of Econometrics* 182(2): 290–308.

Hartford, Jason, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. 2016. "Counterfactual Prediction with Deep Instrumental Variables Networks." arXiv:1612.09596.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second edition. New York, NY: Springer.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2): 994–1028.

Hoberg, Gerard, and Gordon Phillips. 2016. "Text-based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*

124(5): 1423–65.

Imai, Kosuke, and Aaron Strauss. 2011. "Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-Out-the-Vote Campaign." *Political Analysis* 19(1): 1–19.

Jacob, Bryan, Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen. 2016. "Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools." NBER Working Paper 22054.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353(6301): 790–94.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper 14607.

Kang, Jun Seok, Polina Kuznetsova, Michael Luca, and Yejin Choi. 2013. "Where *Not* to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews." EMNLP 2013: 2013 Conference on Empirical Methods in Natural Language.

Kleinberg, Jon, Annie Liang, and Sendhil Mullainathan. 2015. "The Theory is Predictive, But is It Complete? An Application to Human Perception of Randomness." Unpublished paper.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review* 105(5): 491–95.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." NBER Working Paper 23180.

Kogan, Shimon, Dimitry Levin, Byran R. Routledge, Jacob S. Sagi, and Noah A. Smith. 2009. "Predicting Risk from Financial Reports with Regression." In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics,* pp. 272–80. ACM.

Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart. 2010. "Improving Propensity Score Weighting using Machine Learning." *Statistics in Medicine* 29(3): 337–46.

Leeb, Hannes, and Benedikt M. Pötscher. 2006. "Can One Estimate the Conditional Distribution of Post-Model-Selection Estimators?" *Annals of Statistics* 34(5): 2554–91.

Leeb, Hannes, and Bendikt M. Pötscher. 2008. "Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators?" *Econometric Theory* 24(2): 338–76.

Lobell, David B. 2013. "The Use of Satellite Data for Crop Yield Gap Analysis." *Field Crops Research* 143: 56–64.

Ludwig, Jens, Sendhil Mullainathan, and Jann Spiess. 2017. "Machine Learning Tests for Effects on Multiple Outcomes." Unpublished paper.

McBride, Linden, and Austin Nichols. 2016. "Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning." *World Bank Economic Review,* lhw056.

Misra, Sanjog, and Jean-Pierre Dubé. 2016. "Scalable Price Targeting." Unpublished paper.

Moritz, Benjamin, and Tom Zimmermann. 2016. "Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns." Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740751.

Peysakhovich, Alexander, and Jeffrey Naecker. 2015. "Using Methods from Machine Learning to Evaluate Models of Human Choice."

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3): 557–86.

van der Vaart, Aad W., Sandrine Dudoit, and Mark J. van der Laan. 2006. "Oracle Inequalities for Multi-fold Cross Validation." *Statistics and Decisions* 24(3): 351–71.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2): 3–28.

Wager, Stefan, and Susan Athey. 2015. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." arXiv:1510.04342

Yeomans, Michael, Anuj K. Shah, Sendhil Mullainathan, and Jon Kleinberg. 2016. "Making Sense of Recommendations." Unpublished paper.

Zhao, Peng, and Bin Yu. 2006. "On Model Selection Consistency of Lasso." *Journal of Machine Learning Research* 7: 2541–63.

This article has been cited by:

1. Ah Ran Oh, Jungchan Park, Seo Jeong Shin, Byungjin Choi, Jong-Hwan Lee, Seung-Hwa Lee, Kwangmo Yang. 2023. Prediction model for myocardial injury after non-cardiac surgery using machine learning. *Scientific Reports* **13**:1. . [Crossref]
2. Yogi Sugiawan, Robi Kurniawan, Shunsuke Managi. 2023. Assessing the United Nations sustainable development goals from the inclusive wealth perspective. *Scientific Reports* **13**:1. . [Crossref]
3. Nhung Nghiem, June Atkinson, Binh P. Nguyen, An Tran-Duy, Nick Wilson. 2023. Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets. *Health Economics Review* **13**:1. . [Crossref]
4. Philipp Schade, Monika C. Schuhmacher. 2023. Predicting entrepreneurial activity using machine learning. *Journal of Business Venturing Insights* **19**, e00357. [Crossref]
5. Doohee Chung, Pilwon Jeong, Donghwan Kwon, Hyunsoo Han. 2023. Technology acceptance prediction of robo-advisors by machine learning. *Intelligent Systems with Applications* **18**, 200197. [Crossref]
6. Qin Zhang, He Ni, Hao Xu. 2023. Nowcasting Chinese GDP in a data-rich environment: Lessons from machine learning algorithms. *Economic Modelling* **122**, 106204. [Crossref]
7. Paul M. Anglin, Yanmin Gao. 2023. Value of Communication and Social Media: An Equilibrium Theory of Messaging. *The Journal of Real Estate Finance and Economics* **66**:4, 861-903. [Crossref]
8. Justin J. P. Jansen, Ciaran Heavey, Tom J. M. Mom, Zeki Simsek, Shaker A. Zahra. 2023. Scaling-up: Building, Leading and Sustaining Rapid Growth Over Time. *Journal of Management Studies* **60**:3, 581-604. [Crossref]
9. Evaggelia Siopi, Thomas Poufinas, James Ming Chen, Charalampos Agiropoulos. 2023. Can Regulation Affect the Solvency of Insurers? New Evidence from European Insurers. *International Advances in Economic Research* **29**:1-2, 15-30. [Crossref]
10. Shuiyang Pan, Suwan(Cheng) Long, Yiming Wang, Ying Xie. 2023. Nonlinear asset pricing in Chinese stock market: A deep learning approach. *International Review of Financial Analysis* **87**, 102627. [Crossref]
11. Valentin Zelenyuk. 2023. Aggregation in efficiency and productivity analysis: a brief review with new insights and justifications for constant returns to scale. *Journal of Productivity Analysis* **13**. . [Crossref]
12. Xiaolan Yang, Yidong Huang, Wenchao Li. 2023. Economic preferences and state-level corruption: global evidence and mechanism. *Applied Economics Letters* **30**:7, 892-912. [Crossref]
13. Orçun Moralı, Neslihan Yilmaz. Analysis of Real Estate Prices Using Geospatial Data 180-195. [Crossref]
14. Dario Sansone, Anna Zhu. 2023. Using Machine Learning to Create an Early Warning System for Welfare Recipients*. *Oxford Bulletin of Economics and Statistics* **355**. . [Crossref]
15. Niko Hauzenberger, Florian Huber, Karin Klieber. 2023. Real-time inflation forecasting using non-linear dimension reduction techniques. *International Journal of Forecasting* **39**:2, 901-921. [Crossref]
16. Ephraim Kwashie Thompson, Samuel Buertey. 2023. Which firms opt for corporate social responsibility assurance? A machine learning prediction. *Business Ethics, the Environment & Responsibility* **32**:2, 599-611. [Crossref]
17. Chandler McClellan, Emily Mitchell, Jerrod Anderson, Samuel Zuvekas. 2023. Using machine-learning algorithms to improve imputation in the medical expenditure panel survey. *Health Services Research* **58**:2, 423-432. [Crossref]

18. J. Michelle Brock, Ralph De Haas. 2023. Discriminatory Lending: Evidence from Bankers in the Lab. *American Economic Journal: Applied Economics* **15**:2, 31-68. [Abstract] [View PDF article] [PDF with links]
19. S. W. F. de Rezende, B. P. Barella, J. R. V. Moura, K. M. Tsuruta, A. A. Cavalini, V. Steffen. 2023. ISHM for fault condition detection in rotating machines with deep learning models. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **45**:4. . [Crossref]
20. Joshua S. Gans. 2023. Artificial intelligence adoption in a competitive market. *Economica* **90**:358, 690-705. [Crossref]
21. Jing Hao, Feng He, Feng Ma, Shibo Zhang, Xiaotao Zhang. 2023. Machine learning vs deep learning in stock market investment: an international evidence. *Annals of Operations Research* **6**. . [Crossref]
22. Saahoon Hong, Betty Walton, Hea-Won Kim, Taeho Greg Rhee. 2023. Predicting the Behavioral Health Needs of Asian Americans in Public Mental Health Treatment: A Classification Tree Approach. *Administration and Policy in Mental Health and Mental Health Services Research* **49**. . [Crossref]
23. José Cadena, Manuel Villa, Maira Martínez, Jaime Acurio, Luis Chacón. 2023. An Efficient Technique for Global Facial Recognition using Python and OpenCV in 2D Images. *WSEAS TRANSACTIONS ON SYSTEMS AND CONTROL* **18**, 47-57. [Crossref]
24. Michele Rabasco, Pietro Battiston. 2023. Predicting the deterrence effect of tax audits. A machine learning approach. *Metroeconomica* **36**. . [Crossref]
25. Juergen Deppner, Benedict von Ahlefeldt-Dehn, Eli Beracha, Wolfgang Schaefers. 2023. Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach. *The Journal of Real Estate Finance and Economics* **39**. . [Crossref]
26. S. Salcedo-Sanz, S. Cornejo-Bueno, D. Casillas-Pérez, C. Camacho-Gómez, L. Cuadra, M. I. Chidean, A. J. Caamaño. 2023. Collaboration and competition spatial complex networks in regional science. *Journal of Ambient Intelligence and Humanized Computing* **74**. . [Crossref]
27. Emmanouil Tranos, André Carrascal-Incera, George Willis. 2023. Using the Web to Predict Regional Trade Flows: Data Extraction, Modeling, and Validation. *Annals of the American Association of Geographers* **113**:3, 717-739. [Crossref]
28. Süheyl Uçucu, Talha Karabıyık, Fatih Mehmet Azik. 2023. Machine learning models can predict the presence of variants in hemoglobin: artificial neural network-based recognition of human hemoglobin variants by HPLC. *Turkish Journal of Biochemistry* **48**:1, 5-11. [Crossref]
29. Tobias Götze, Marc Gürtler, Eileen Witowski. 2023. Forecasting accuracy of machine learning and linear regression: evidence from the secondary CAT bond market. *Journal of Business Economics* **34**. . [Crossref]
30. Guan-Yuan Wang. 2023. The effect of environment on housing prices: Evidence from the Google Street View. *Journal of Forecasting* **42**:2, 288-311. [Crossref]
31. PAUL GEERTSEMA, HELEN LU. 2023. Relative Valuation with Machine Learning. *Journal of Accounting Research* **61**:1, 329-376. [Crossref]
32. Philip S. Dale, Alexander Paul, Michael Rosholm, Dorthe Bleses. 2023. Prediction from early childhood vocabulary to academic achievement at the end of compulsory schooling in Denmark. *International Journal of Behavioral Development* **47**:2, 123-134. [Crossref]
33. Majid Majzoubi, Eric Yanfei Zhao. 2023. Going beyond optimal distinctiveness: Strategic positioning for gaining an audience composition premium. *Strategic Management Journal* **44**:3, 737-777. [Crossref]
34. Gert Bijnens, Shyngys Karimov, Jozef Konings. 2023. Does Automatic Wage Indexation Destroy Jobs? A Machine Learning Approach. *De Economist* **171**:1, 85-117. [Crossref]

35. Debdatta Saha, Timothy M. Young, Jessica Thacker. 2023. Predicting firm performance and size using machine learning with a Bayesian perspective. *Machine Learning with Applications* **11**, 100453. [Crossref]

36. Moritz Meister, Annekatrin Niebuhr, Jan Cornelius Peters, Johannes Stiller. 2023. Local attributes and migration balance – evidence for different age and skill groups from a machine learning approach. *Regional Science Policy & Practice* **87**. . [Crossref]

37. Rafael Abrantes Penchel, Ivan Aldaya, Lucas Marim, Mirian Paula dos Santos, Lucio Cardozo-Filho, Veeriah Jegatheesan, José Augusto de Oliveira. 2023. Analysis of Cleaner Production Performance in Manufacturing Companies Employing Artificial Neural Networks. *Applied Sciences* **13**:6, 4029. [Crossref]

38. Natalia Pecorari, Jose Cuesta. Citizen Participation and Political Trust in Latin America and the Caribbean: A Machine Learning Approach **10085**, . [Crossref]

39. Martin Magris, Mostafa Shabani, Alexandros Iosifidis. 2023. Bayesian bilinear neural network for predicting the mid-price dynamics in limit-order book markets. *Journal of Forecasting* **112**. . [Crossref]

40. Yuexi Liu. 2023. A Machine Learning Approach for Selecting Directors of Chinese Listed Company. *Highlights in Business, Economics and Management* **5**, 380-389. [Crossref]

41. Daniela M. Behr, Lixue Chen, Ankita Goel, Khondoker Tanveer Haider, Sandeep Singh, Asad Zaman. Estimating House Prices in Emerging Markets and Developing Economies: A Big Data Approach **4**, . [Crossref]

42. Filipe R Campante, Davin Chor, Bingjing Li. 2023. The Political Economy Consequences of China's Export Slowdown. *Journal of the European Economic Association* **34**. . [Crossref]

43. Daniel Hoang, Kevin Wiegratz. 2023. Machine learning methods in finance: Recent applications and prospects. *European Financial Management* **24**. . [Crossref]

44. Jill Furzer, Maripier Isabelle, Boriana Miloucheva, Audrey Laporte. 2023. Public drug insurance, moral hazard and children's use of mental health medication: Latent mental health risk-specific responses to lower out-of-pocket treatment costs. *Health Economics* **32**:2, 518-538. [Crossref]

45. Luara A. Freitas, Rodrigo P. Savegnago, Anderson A. C. Alves, Ricardo L. D. Costa, Danisio P. Munari, Nedenia B. Stafuzza, Guilherme J. M. Rosa, Claudia C. P. Paz. 2023. Classification Performance of Machine Learning Methods for Identifying Resistance, Resilience, and Susceptibility to Haemonchus contortus Infections in Sheep. *Animals* **13**:3, 374. [Crossref]

46. Rafael Quintana. 2023. Embracing complexity in social science research. *Quality & Quantity* **57**:1, 15-38. [Crossref]

47. Suriyan Jomthanachai, Wai Peng Wong, Khai Wah Khaw. 2023. An Application of Machine Learning to Logistics Performance Prediction: An Economics Attribute-Based of Collective Instance. *Computational Economics* **22**. . [Crossref]

48. Tiantian Zhang, Yongtang Wu, Yuling Chen, Tao Li, Xiaojun Ren. 2023. Collaborative Energy Price Computing Based on Sarima-Ann and Asymmetric Stackelberg Games. *Symmetry* **15**:2, 443. [Crossref]

49. Hai-Van Thi Mai, May Huu Nguyen, Hai-Bang Ly. 2023. Development of machine learning methods to predict the compressive strength of fiber-reinforced self-compacting concrete and sensitivity analysis. *Construction and Building Materials* **367**, 130339. [Crossref]

50. Thema Monroe-White, Jesse Lecy. 2023. The Wells-Du Bois Protocol for Machine Learning Bias: Building Critical Quantitative Foundations for Third Sector Scholarship. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* **34**:1, 170-184. [Crossref]

51. Rashid Zaman. 2023. When corporate culture matters: The case of stakeholder violations. *The British Accounting Review* **48**, 101188. [Crossref]

52. Katherine Hoffmann Pham, Miguel Luengo-Oroz. 2023. Predictive modelling of movements of refugees and internally displaced people: towards a computational framework. *Journal of Ethnic and Migration Studies* **49**:2, 408-444. [Crossref]
53. Mario D. Molina, Nancy Chau, Amanda D. Rodewald, Filiz Garip. 2023. How to model the weather-migration link: a machine-learning approach to variable selection in the Mexico-U.S. context. *Journal of Ethnic and Migration Studies* **49**:2, 465-491. [Crossref]
54. Rajesh Kumar Maurya, Himani Jain, Tarun Kumar Sharma, Surbhi Sharma, Mani Dublish. Novel Framework for Quality Crop Predictions Using Data Mining and Soft Computing Techniques 83-88. [Crossref]
55. Christoph Kern, Bernd Weiß, Jan-Philipp Kolb. 2023. Predicting Nonresponse in Future Waves of a Probability-Based Mixed-Mode Panel with Machine Learning. *Journal of Survey Statistics and Methodology* **11**:1, 100-123. [Crossref]
56. Wen Lin. 2023. The effect of product quantity on willingness to pay: A meta-regression analysis of beef valuation studies. *Agribusiness* **87**. . [Crossref]
57. Xianglong Xu, Eric P. F. Chow, Christopher K. Fairley, Marcus Chen, Ivette Aguirre, Jane Goller, Jane Hocking, Natalie Carvalho, Lei Zhang, Jason J. Ong. 2023. Determinants and prediction of Chlamydia trachomatis re-testing and re-infection within 1 year among heterosexuals with chlamydia attending a sexual health clinic. *Frontiers in Public Health* **10**. . [Crossref]
58. Hamza Heni, S. Arona Diop, Jacques Renaud, Leandro C. Coelho. 2023. Measuring fuel consumption in vehicle routing: new estimation models using supervised learning. *International Journal of Production Research* **61**:1, 114-130. [Crossref]
59. Oleksandr Fomin, Sergii Polozhaenko, Valentyn Krykun, Andrii Orlov, Daria Lys. Interpretation of Dynamic Models Based on Neural Networks in the Form of Integral-Power Series 258-265. [Crossref]
60. Gabriel M. Ahlfeldt, Stephan Heblich, Tobias Seidel. 2023. Micro-geographic property price and rent indices. *Regional Science and Urban Economics* **98**, 103836. [Crossref]
61. Xianfei Hui, Baiqing Sun, Indranil SenGupta, Yan Zhou, Hui Jiang. 2023. Stochastic volatility modeling of high-frequency CSI 300 index and dynamic jump prediction driven by machine learning. *Electronic Research Archive* **31**:3, 1365-1386. [Crossref]
62. Pushpendra Rana, Lav R. Varshney. 2023. Exploring limits to tree planting as a natural climate solution. *Journal of Cleaner Production* **384**, 135566. [Crossref]
63. Anderson Monken, William Ampeh, Flora Haberkorn, Uma Krishnaswamy, Feras A. Batarseh. Assuring AI methods for economic policymaking 371-427. [Crossref]
64. Andrés Alonso, José Manuel Carbó, J. Manuel Marqués. 2023. Machine Learning methods in climate finance: a systematic review. *SSRN Electronic Journal* **225**. . [Crossref]
65. Ragupathy Venkatachalam, Shu G. Wang. Computational Thinking in Economics and Finance: Introductory Remarks 1-12. [Crossref]
66. Gunther Dahm, Frauke Peter. Einfach anders oder vielfältig verschieden? Ein differenzierter Blick auf Hochschulabsolvent*innen mit beruflicher Vorqualifikation 223-262. [Crossref]
67. M. van der Westhuizen, K. H. von Leipzig, V. Hummel. Augmented Reality Combined with Machine Learning to Increase Productivity in Fruit Packing 415-431. [Crossref]
68. Brett R. Gordon, Mitchell J. Lovett, Bowen Luo, James C. Reeder. 2023. Disentangling the Effects of Ad Tone on Voter Turnout and Candidate Choice in Presidential Elections. *Management Science* **69**:1, 220-243. [Crossref]
69. Brian A. Bourquard, Gemma Berenguer, Allan W. Gray, Paul V. Preckel. 2022. Raw material variability in food manufacturing: a data-driven snack food industry case. *Production & Manufacturing Research* **10**:1, 294-320. [Crossref]

70. Hicham Sadok, Fadi Sakka, Mohammed El Hadi El Maknouzi. 2022. Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance* **10**:1. . [Crossref]
71. Omid Rafieian. 2022. Optimizing User Engagement Through Adaptive Ad Sequencing. *Marketing Science* **21**. . [Crossref]
72. Samarra Badrouchi, Mohamed Mongi Bacha, Hafedh Hedri, Taieb Ben Abdallah, Ezzedine Abderrahim. 2022. Toward generalizing the use of artificial intelligence in nephrology and kidney transplantation. *Journal of Nephrology* **3**. . [Crossref]
73. Hannes Mueller, Christopher Rauh. 2022. The Hard Problem of Prediction for Conflict Prevention. *Journal of the European Economic Association* **20**:6, 2440-2467. [Crossref]
74. Erlend Eide Bø, Per Medby, Odd Erik Nygård, Mona Takle. 2022. Hedonisk verdsettelse av bolig med fleksibel geografisk inndeling. *Tidsskrift for boligforskning* **5**:2, 70-87. [Crossref]
75. Hazal Colak Oz, Çiçek Güven, Gonzalo Nápoles. 2022. School dropout prediction and feature importance exploration in Malawi using household panel data: machine learning approach. *Journal of Computational Social Science* **33**. . [Crossref]
76. Muhammed Umar Farrukh, Richard Wainwright, Keeley Crockett, David McLean, Neil Dagnall. Building Actionable Personas Using Machine Learning Techniques 463-472. [Crossref]
77. Pedro Henrique Melo Albuquerque, Yaohao Peng, João Pedro Fontoura da Silva. 2022. Making the whole greater than the sum of its parts: A literature review of ensemble methods for financial time series forecasting. *Journal of Forecasting* **41**:8, 1701-1724. [Crossref]
78. Francesco Decarolis, Cristina Giorgiantonio. 2022. Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science* **11**:1. . [Crossref]
79. Huu Duy Nguyen. 2022. GIS-based hybrid machine learning for flood susceptibility prediction in the Nhat Le–Kien Giang watershed, Vietnam. *Earth Science Informatics* **15**:4, 2369-2386. [Crossref]
80. Netta Barak-Corren, Yoav Kan-Tor, Nelson Tebbe. 2022. Examining the effects of antidiscrimination laws on children in the foster care and adoption systems. *Journal of Empirical Legal Studies* **19**:4, 1003-1066. [Crossref]
81. Yudai Suzuki, Qi Gao, Ken C. Pradel, Kenji Yasuoka, Naoki Yamamoto. 2022. Natural quantum reservoir computing for temporal information processing. *Scientific Reports* **12**:1. . [Crossref]
82. Micha Kaiser, Steffen Otterbach, Alfonso Sousa-Poza. 2022. Using machine learning to uncover the relation between age and life satisfaction. *Scientific Reports* **12**:1. . [Crossref]
83. Minghua Xu, Ziyao Wei, Jiang Wu. 2022. How emotional communication happens in social media: Predicting "Arousal-Homophily-Echo" emotional communication with multi-dimensional features. *Telematics and Informatics Reports* **8**, 100019. [Crossref]
84. TINGBIN BIAN, JIN CHEN, QU FENG, JINGYI LI. 2022. COMPARING ECONOMETRIC ANALYSES WITH MACHINE LEARNING APPROACHES: A STUDY ON SINGAPORE PRIVATE PROPERTY MARKET. *The Singapore Economic Review* **67**:06, 1787-1810. [Crossref]
85. Ebru Çağlayan Akay, Naciye Tuba Yılmaz Soydan, Burcu Kocarık Gacar. 2022. Bibliometric analysis of the published literature on machine learning in economics and econometrics. *Social Network Analysis and Mining* **12**:1. . [Crossref]
86. Qishuo Gao, Vivien Shi, Christopher Pettit, Hoon Han. 2022. Property valuation using machine learning algorithms on statistical areas in Greater Sydney, Australia. *Land Use Policy* **123**, 106409. [Crossref]
87. Elizabeth Jane Casabianca, Michele Catalano, Lorenzo Forni, Elena Giarda, Simone Passeri. 2022. A machine learning approach to rank the determinants of banking crises over time and across countries. *Journal of International Money and Finance* **129**, 102739. [Crossref]

88. Perry Sadorsky. 2022. Using machine learning to predict clean energy stock prices: How important are market volatility and economic policy uncertainty?. *Journal of Climate Finance* **1**, 100002. [Crossref]
89. Haixiang Yao, Shenghao Xia, Hao Liu. 2022. Six-factor asset pricing and portfolio investment via deep learning: Evidence from Chinese stock market. *Pacific-Basin Finance Journal* **76**, 101886. [Crossref]
90. Edward I. Altman, Marco Balzano, Alessandro Giannozzi, Stjepan Srhoj. 2022. Revisiting SME default predictors: The Omega Score. *Journal of Small Business Management* **8**, 1-35. [Crossref]
91. Hernando Grueso. 2022. Unveiling the Causal Mechanisms Within Multidimensional Poverty. *Evaluation Review* **149**, 0193841X2211409. [Crossref]
92. Ibrahim Niankara. Sustainability Through Open Data Sharing and Reuse in The Digital Economy 1-11. [Crossref]
93. Chengyan Gu. 2022. Market segmentation and dynamic price discrimination in the U.S. airline industry. *Journal of Revenue and Pricing Management* **97**. . [Crossref]
94. Lidia Ceriani, Sergio Olivieri, Marco Ranzani. 2022. Housing, imputed rent, and household welfare. *The Journal of Economic Inequality* **38**. . [Crossref]
95. Angela L Duckworth, Katherine L Milkman. 2022. A guide to megastudies. *PNAS Nexus* **1**:5. . [Crossref]
96. Yunhe Lu. 2022. Detecting Imperfect Substitution between Comparably Skilled Immigrants and Natives: A Machine Learning Approach. *International Migration Review* **28**, 019791832211264. [Crossref]
97. Jinwoo Jung, Jihwan Kim, Changha Jin. 2022. DOES MACHINE LEARNING PREDICTION DAMPEN THE INFORMATION ASYMMETRY FOR NON-LOCAL INVESTORS?. *International Journal of Strategic Property Management* **26**:5, 345-361. [Crossref]
98. Byron Botha, Rulof Burger, Kevin Kotzé, Neil Rankin, Daan Steenkamp. 2022. Big data forecasting of South African inflation. *Empirical Economics* **78**. . [Crossref]
99. Ari Hyytinen, Petri Rouvinen, Mika Pajarinen, Joosua Virtanen. 2022. Ex Ante Predictability of Rapid Growth: A Design Science Approach. *Entrepreneurship Theory and Practice* **30**, 104225872211282. [Crossref]
100. Prothit Sen, Phanish Puranam. 2022. Do Alliance portfolios encourage or impede new business practice adoption? Theory and evidence from the private equity industry. *Strategic Management Journal* **43**:11, 2279-2312. [Crossref]
101. Byungjin Choi, Ah Ran Oh, Seung-Hwa Lee, Dong Yun Lee, Jong-Hwan Lee, Kwangmo Yang, Ha Yeon Kim, Rae Woong Park, Jungchan Park. 2022. Prediction Model for 30-Day Mortality after Non-Cardiac Surgery Using Machine-Learning Techniques Based on Preoperative Evaluation of Electronic Medical Records. *Journal of Clinical Medicine* **11**:21, 6487. [Crossref]
102. Hector O. Zapata, Supratik Mukhopadhyay. 2022. A Bibliometric Analysis of Machine Learning Econometrics in Asset Pricing. *Journal of Risk and Financial Management* **15**:11, 535. [Crossref]
103. Guido de Blasio, Alessio D'Ignazio, Marco Letta. 2022. Gotham city. Predicting 'corrupted' municipalities with machine learning. *Technological Forecasting and Social Change* **184**, 122016. [Crossref]
104. Xue-Zhong He, Shen Lin. 2022. Reinforcement Learning Equilibrium in Limit Order Markets. *Journal of Economic Dynamics and Control* **144**, 104497. [Crossref]
105. Christian Leuz. 2022. Towards a design-based approach to accounting research. *Journal of Accounting and Economics* **74**:2-3, 101550. [Crossref]
106. Ian Lundberg, Jennie E. Brand, Nanum Jeon. 2022. Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research* **108**, 102807. [Crossref]

107. Rama K. Malladi. 2022. Application of Supervised Machine Learning Techniques to Forecast the COVID-19 U.S. Recession and Stock Market Crash. *Computational Economics* **8**. . [Crossref]
108. Khondhaker Al Momin, Saurav Barua, Omar Faruqe Hamim, Subrata Roy. 2022. Modeling the Behavior in Choosing the Travel Mode for Long-Distance Travel Using Supervised Machine Learning Algorithms. *Communications - Scientific letters of the University of Zilina* **24**:4, A187-A197. [Crossref]
109. Sathya Uma Lakshmi Kandasamy, Piyush Kumar Singh, Dillip Kumar Swain. 2022. Determination of Factors Affecting the Adoption of Integrated Farming System in Dryland Areas of Southern India by Using Supervised Learning Techniques. *Journal of Asian and African Studies* **66**, 002190962211303. [Crossref]
110. David M Reeb, Wanli Zhao. 2022. Disregarding the Shoulders of Giants: Inferences from Innovation Research. *The Review of Corporate Finance Studies* **11**:4, 923-964. [Crossref]
111. Mariia Garkavenko, Tatiana Beliaeva, Eric Gaussier, Hamid Mirisaee, Cédric Lagnier, Agnès Guerraz. 2022. Assessing the Factors Related to a Start-Up's Valuation Using Prediction and Causal Discovery. *Entrepreneurship Theory and Practice* **18**, 104225872211212. [Crossref]
112. Kathleen T. Li, Christophe Van den Bulte. 2022. Augmented Difference-in-Differences. *Marketing Science* . [Crossref]
113. Christian P R Schmid, Nicolas Schreiner, Alois Stutzer. 2022. Transfer Payment Systems and Financial Distress: Insights from Health Insurance Premium Subsidies. *Journal of the European Economic Association* **20**:5, 1829-1858. [Crossref]
114. Steen Nielsen. 2022. Management accounting and the concepts of exploratory data analysis and unsupervised machine learning: a literature study and future directions. *Journal of Accounting & Organizational Change* **18**:5, 811-853. [Crossref]
115. Alessandra Garbero, Marco Letta. 2022. Predicting household resilience with machine learning: preliminary cross-country tests. *Empirical Economics* **63**:4, 2057-2070. [Crossref]
116. Simon Blöthner, Mario Larch. 2022. Economic determinants of regional trade agreements revisited using machine learning. *Empirical Economics* **63**:4, 1771-1807. [Crossref]
117. Werner Kristjanpoller, Nicole Astudillo, Josephine E. Olson. 2022. An empirical application of a hybrid ANFIS model to predict household over-indebtedness. *Neural Computing and Applications* **34**:20, 17343-17353. [Crossref]
118. Nicolas Robette. 2022. Trees and forest. Recursive partitioning as an alternative to parametric regression models in social sciences. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* **156**:1, 7-56. [Crossref]
119. Marina Bonaccolto-Töpfer, Stephanie Briel. 2022. The gender pay gap revisited: Does machine learning offer new insights?. *Labour Economics* **78**, 102223. [Crossref]
120. Braeden Van Deynze, Robert Fonner, Blake E. Feist, Sunny L. Jardine, Daniel S. Holland. 2022. What influences spatial variability in restoration costs? Econometric cost models for inference and prediction in restoration planning. *Biological Conservation* **274**, 109710. [Crossref]
121. Xueqiu Zhuang, Huihua Jiao, Li Kang. 2022. Digital Management and Optimization of Tourism Information Resources Based on Machine Learning. *International Transactions on Electrical Energy Systems* **2022**, 1-12. [Crossref]
122. Ankita Raj, Sunil Kumar Singh. Forecasting GDP of India and its neighbouring countries using Time Series Analysis 1-6. [Crossref]
123. Shahrbanoo Rezaei, Anahita Khojandi, Antora Mohsena Haque, Candace Brakewood, Mingzhou Jin, Christopher Cherry. 2022. Performance evaluation of mode choice models under balanced and imbalanced data assumptions. *Transportation Letters* **14**:8, 920-932. [Crossref]

124. Abdus Samad Azad, Rajalingam Sokkalingam, Hanita Daud, Sajal Kumar Adhikary. Application Water Level Prediction Through Seasonal Autoregressive Integrated Moving Average: Red Hills Reservoir Case Study 1-6. [Crossref]
125. Kaifeng Chen, Kunrong Zeng. 2022. Performance Optimization Model of Molecular Dynamics Simulation Based on Machine Learning and Data Mining Algorithm. *Mobile Information Systems* **2022**, 1-12. [Crossref]
126. Cigdem Kosar Tas, Huseyin Guler, Yeliz Yalcin. 2022. The usage of bridge estimator to determine the order of integration for possibly integrated series as an alternative to Dickey–Pantula unit root test. *Communications in Statistics - Simulation and Computation* **30**, 1-17. [Crossref]
127. Manivel Murugan, Sankaran Marisamynathan. 2022. Estimation of two-wheeler users' mode shift behavior and policy analysis to encourage electric-bike adoption in India. *Case Studies on Transport Policy* **10**:3, 1673-1685. [Crossref]
128. Antonio Pacifico. 2022. Structural Compressed Panel VAR with Stochastic Volatility: A Robust Bayesian Model Averaging Procedure. *Econometrics* **10**:3, 28. [Crossref]
129. João B. Assunção, Pedro Afonso Fernandes. 2022. Nowcasting GDP: An Application to Portugal. *Forecasting* **4**:3, 717-731. [Crossref]
130. Muzaffer Eroğlu, Meltem Karatepe Kaya. 2022. Impact of Artificial Intelligence on Corporate Board Diversity Policies and Regulations. *European Business Organization Law Review* **23**:3, 541-572. [Crossref]
131. Tsun Se Cheong, Guanghua Wan, David Kam Hung Chui. 2022. Unveiling the Relationship between Economic Growth and Equality for Developing Countries. *China & World Economy* **30**:5, 1-28. [Crossref]
132. Marcos Delprato, Alessia Frola, Germán Antequera. 2022. Indigenous and non-Indigenous proficiency gaps for out-of-school and in-school populations: A machine learning approach. *International Journal of Educational Development* **93**, 102631. [Crossref]
133. Klaus Grobys, James W. Kolari, Joachim Niang. 2022. Man versus machine: on artificial intelligence and hedge funds performance. *Applied Economics* **54**:40, 4632-4646. [Crossref]
134. Muneer M. Alshater, Ilias Kampouris, Hazem Marashdeh, Osama F. Atayah, Hasanul Banna. 2022. Early warning system to predict energy prices: the role of artificial intelligence and machine learning. *Annals of Operations Research* **221**. . [Crossref]
135. Liang Lu, Guang Tian, Patrick Hatzenbuehler. 2022. How agricultural economists are using big data: a review. *China Agricultural Economic Review* **14**:3, 494-508. [Crossref]
136. Jing Yao, Yanze Li. 2022. Youth Sports Special Skills' Training and Evaluation System Based on Machine Learning. *Mobile Information Systems* **2022**, 1-10. [Crossref]
137. Chih-Hung Pai, Sai Xu, Jianren Jin, Yunfeng Shang. 2022. Value evaluation of cultural tourism tourists' psychological expectation based on machine learning data mining. *Frontiers in Psychology* **13**. . [Crossref]
138. Mari O. Mamre, Dag Einar Sommervoll. 2022. Coming of Age: Renovation Premiums in Housing Markets. *The Journal of Real Estate Finance and Economics* **30**. . [Crossref]
139. John Kommunuri. 2022. Artificial intelligence and the changing landscape of accounting: a viewpoint. *Pacific Accounting Review* **34**:4, 585-594. [Crossref]
140. Mona Aghdaee, Bonny Parkinson, Kompal Sinha, Yuanyuan Gu, Rajan Sharma, Emma Olin, Henry Cutler. 2022. An examination of machine learning to map non-preference based patient reported outcome measures to health state utility values. *Health Economics* **31**:8, 1525-1557. [Crossref]
141. Furkan Kartal, Uğur Özveren. 2022. Prediction of activation energy for combustion and pyrolysis by means of machine learning. *Thermal Science and Engineering Progress* **33**, 101346. [Crossref]

142. Mehmet Salti, Emel Ciger, Evrim Ersin Kangal, Bilgin Zengin. 2022. Data-driven predictive modeling of Hubble parameter. *Physica Scripta* **97**:8, 085011. [Crossref]

143. Jing Xu, Ren Zhang, Yangjun Wang, Hengqian Yan, Quanhong Liu, Yutong Guo, Yongcun Ren. 2022. Assessing China's Investment Risk of the Maritime Silk Road: A Model Based on Multiple Machine Learning Methods. *Energies* **15**:16, 5780. [Crossref]

144. Alice Bertoletti, Jasmina Berbegal-Mirabent, Tommaso Agasisti. 2022. Higher education systems and regional economic development in Europe: A combined approach using econometric and machine learning methods. *Socio-Economic Planning Sciences* **82**, 101231. [Crossref]

145. Thomas PlÖtz. 2022. Applying Machine Learning for Sensor Data Analysis in Interactive Systems. *ACM Computing Surveys* **54**:6, 1-25. [Crossref]

146. Yunlong Tong. 2022. Assessment of Physical Fitness and Health Status of Athletes Based on Intelligent Medical Treatment under Machine Learning. *Computational Intelligence and Neuroscience* **2022**, 1-13. [Crossref]

147. Lida Kuang, Samruda Pobbathi, Yuri Mansury, Matthew A. Shapiro, Vijay K. Gurbani. 2022. Predicting age and gender from network telemetry: Implications for privacy and impact on policy. *PLOS ONE* **17**:7, e0271714. [Crossref]

148. Matthias Westphal, Daniel A Kamhöfer, Hendrik Schmitz. 2022. Marginal College Wage Premiums Under Selection Into Employment. *The Economic Journal* **132**:646, 2231-2272. [Crossref]

149. Juergen Deppner, Marcelo Cajias. 2022. Accounting for Spatial Autocorrelation in Algorithm-Driven Hedonic Models: A Spatial Cross-Validation Approach. *The Journal of Real Estate Finance and Economics* **11**. . [Crossref]

150. Anja Garbely, Elias Steiner. 2022. Understanding compliance with voluntary sustainability standards: a machine learning approach. *Environment, Development and Sustainability* **26**. . [Crossref]

151. Harold D. Chiang, Kengo Kato, Yukun Ma, Yuya Sasaki. 2022. Multiway Cluster Robust Double/Debiased Machine Learning. *Journal of Business & Economic Statistics* **40**:3, 1046-1056. [Crossref]

152. Sathya Uma Lakshmi Kandasamy, Piyush Kumar Singh, Dillip Kumar Swain. 2022. Climate change vulnerability assessment of dryland farmers and factors Identification using machine learning techniques. *Local Environment* **27**:7, 824-846. [Crossref]

153. Michael Bailey, Drew Johnston, Theresa Kuchler, Johannes Stroebel, Arlene Wong. 2022. Peer Effects in Product Adoption. *American Economic Journal: Applied Economics* **14**:3, 488-526. [Abstract] [View PDF article] [PDF with links]

154. Yang Wang, Hong Zhang, Libing Liu. 2022. Does city construction improve life quality?-evidence from POI data of China. *International Review of Economics & Finance* **80**, 643-653. [Crossref]

155. Samuel Bazzi, Robert A. Blair, Christopher Blattman, Oeindrila Dube, Matthew Gudgeon, Richard Peck. 2022. The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia. *The Review of Economics and Statistics* **104**:4, 764-779. [Crossref]

156. Sun-Youn Shin, Han-Gyun Woo. 2022. Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies* **15**:13, 4880. [Crossref]

157. Sangjin Park, Jae-Suk Yang. 2022. Interpretable deep learning LSTM model for intelligent economic decision-making. *Knowledge-Based Systems* **248**, 108907. [Crossref]

158. Perry Sadorsky. 2022. Forecasting solar stock prices using tree-based machine learning classification: How important are silver prices?. *The North American Journal of Economics and Finance* **61**, 101705. [Crossref]

159. Felix Lorenz, Jonas Willwersch, Marcelo Cajias, Franz Fuerst. 2022. Interpretable machine learning for real estate market analysis. *Real Estate Economics* **37**. . [Crossref]

160. Josip Franic. 2022. What do we really know about the drivers of undeclared work? An evaluation of the current state of affairs using machine learning. *AI & SOCIETY* **1**. . [Crossref]
161. Shixuan Tang. 2022. Applications of Machine Learning in the Industry of Healthcare. *Highlights in Science, Engineering and Technology* **1**, 87-96. [Crossref]
162. Mehmet Güney Celbiş. 2022. Unemployment in Rural Europe: A Machine Learning Perspective. *Applied Spatial Analysis and Policy* **42**. . [Crossref]
163. Zeye Chen, Xin Xia. 2022. The Top-Level Design Strategy of "Curriculum Civics" in the New Era of Universities Based on Machine Learning under the Perspective of Communication Science. *Security and Communication Networks* **2022**, 1-8. [Crossref]
164. Florian M. Artinger, Gerd Gigerenzer, Perke Jacobs. 2022. Satisficing: Integrating Two Traditions. *Journal of Economic Literature* **60**:2, 598-635. [Abstract] [View PDF article] [PDF with links]
165. Raphael H. Heiberger. 2022. Applying Machine Learning in Sociology: How to Predict Gender and Reveal Research Preferences. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* **74**:S1, 383-406. [Crossref]
166. P. Uday Ashish, Rashtra Vibhuti Sharma, Sindhu Hak Gupta, Asmita Rajawat. 2022. Classification of limb movements using different predictive analysis algorithms. *International Journal of System Assurance Engineering and Management* **13**:3, 1385-1395. [Crossref]
167. Oren Barkan, Jonathan Benchimol, Itamar Caspi, Eliya Cohen, Allon Hammer, Noam Koenigstein. 2022. Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting* **36**. . [Crossref]
168. Saqib Aziz, Michael Dowling, Helmi Hammami, Anke Piepenbrink. 2022. Machine learning in finance: A topic modeling approach. *European Financial Management* **28**:3, 744-770. [Crossref]
169. Nate Breznau. 2022. Integrating Computer Prediction Methods in Social Science: A Comment on Hofman et al. (2021). *Social Science Computer Review* **40**:3, 844-853. [Crossref]
170. Arne Steinkraus. 2022. Asking Why? – Das Spannungsfeld zwischen Ökonometrie und Data Science. *Wirtschaftsinformatik & Management* **14**:3, 186-191. [Crossref]
171. Monica Andini, Michela Boldrini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Andrea Paladini. 2022. Machine learning in the service of policy targeting: The case of public credit guarantees. *Journal of Economic Behavior & Organization* **198**, 434-475. [Crossref]
172. Brian Blankenship, Michaël Aklin, Johannes Urpelainen, Vagisha Nandan. 2022. Jobs for a just transition: Evidence on coal job preferences from India. *Energy Policy* **165**, 112910. [Crossref]
173. Lothar Walter, Nils M. Denter, Jan Kebel. 2022. A review on digitalization trends in patent information databases and interrogation tools. *World Patent Information* **69**, 102107. [Crossref]
174. Pushpendra Rana, Forrest Fleischman, Vijay Ramprasad, Kangjae Lee. 2022. Predicting wasteful spending in tree planting programs in Indian Himalaya. *World Development* **154**, 105864. [Crossref]
175. Jason W. Miller, Travis Kulpa. 2022. Econometrics and archival data: Reflections for purchasing and supply management (PSM) research. *Journal of Purchasing and Supply Management* **28**:3, 100780. [Crossref]
176. Alexandre Rubesam. 2022. Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market. *Emerging Markets Review* **51**, 100891. [Crossref]
177. Ran Liu. 2022. Leveraging machine learning methods to estimate heterogeneous effects: father absence in China as an example. *Chinese Sociological Review* **54**:3, 223-251. [Crossref]
178. Samira Bounid, Mohammed Oughanem, Salman Bourkadi. Advanced Financial Data Processing and Labeling Methods for Machine Learning 1-6. [Crossref]

179. Boubacar Diallo. 2022. Machine learning approaches to testing institutional hypotheses: the case of Acemoglu, Johnson, and Robinson (2001). *Empirical Economics* **62**:5, 2587-2600. [Crossref]
180. Mustafa Savci, Ahmet Tekin, Jon D. Elhai. 2022. Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology* **41**:5, 2755-2764. [Crossref]
181. Matteo Alpino, Karen Evelyn Hauge, Andreas Kotsadam, Simen Markussen. 2022. Effects of dialogue meetings on sickness absence—Evidence from a large field experiment. *Journal of Health Economics* **83**, 102615. [Crossref]
182. Keita Abe, Christopher M. Anderson. 2022. A Dynamic Model of Endogenous Fishing Duration. *Journal of the Association of Environmental and Resource Economists* **9**:3, 425-454. [Crossref]
183. MIAO LIU. 2022. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. *Journal of Accounting Research* **60**:2, 607-651. [Crossref]
184. XI CHEN, YANG HA (TONY) CHO, YIWEI DOU, BARUCH LEV. 2022. Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research* **60**:2, 467-515. [Crossref]
185. Joey Blumberg, Gary Thompson. 2022. Nonparametric segmentation methods: Applications of unsupervised machine learning and revealed preference. *American Journal of Agricultural Economics* **104**:3, 976-998. [Crossref]
186. Yang Yi, Le Wen, Shan He. 2022. Partitioning for "Common but Differentiated" Precise Air Pollution Governance: A Combined Machine Learning and Spatial Econometric Approach. *Energies* **15**:9, 3346. [Crossref]
187. Valentas Gružauskas, Aurelija Burinskienė. 2022. Managing Supply Chain Complexity and Sustainability: The Case of the Food Industry. *Processes* **10**:5, 852. [Crossref]
188. Pierre-Philippe Combes, Laurent Gobillon, Yanos Zylberberg. 2022. Urban economics in a historical perspective: Recovering data with machine learning. *Regional Science and Urban Economics* **94**, 103711. [Crossref]
189. Eric Yanfei Zhao. Optimal Distinctiveness **31**, . [Crossref]
190. Sendhil Mullainathan, Ziad Obermeyer. 2022. Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. *The Quarterly Journal of Economics* **137**:2, 679-727. [Crossref]
191. Steve J. Bickley, Ho Fai Chan, Benno Torgler. 2022. Artificial intelligence in the field of economics. *Scientometrics* **127**:4, 2055-2084. [Crossref]
192. Shinya Sugawara. 2022. What composes desirable formal at-home elder care? An analysis for multiple service combinations. *The Japanese Economic Review* **73**:2, 373-402. [Crossref]
193. Manivel Murugan, Sankaran Marisamynathan. 2022. Mode shift behaviour and user willingness to adopt the electric two-wheeler: A study based on Indian road user preferences. *International Journal of Transportation Science and Technology* **16**. . [Crossref]
194. Khaled Obaid, Kuntara Pukthuanthong. 2022. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics* **144**:1, 273-297. [Crossref]
195. Satish Chand, Yu Zhang. 2022. Learning from machines to close the gap between funding and expenditure in the Australian National Disability Insurance Scheme. *International Journal of Information Management Data Insights* **2**:1, 100077. [Crossref]
196. Joshua D. Angrist, Brigham Frandsen. 2022. Machine Labor. *Journal of Labor Economics* **40**:S1, S97-S140. [Crossref]
197. Kim Anh Thi Nguyen, Tram Anh Thi Nguyen, Brice M. Nguelifack, Curtis M. Jolly. 2022. Machine Learning Approaches for Predicting Willingness to Pay for Shrimp Insurance in Vietnam. *Marine Resource Economics* **37**:2, 155-182. [Crossref]

198. Gonul Colak, Mengchuan Fu, Iftekhar Hasan. 2022. On modeling IPO failure risk. *Economic Modelling* **109**, 105790. [Crossref]

199. S. Surya, Sumeet Gupta, Abolfazl Mehbodniya, Jeidy Panduro-Ramirez, Prabhakara Rao Kapula, Tanweer Alam, Karthikeyan Kaliyaperumal. 2022. Addressing the Real World Problem of Managing Wireless Communication Systems Using Explainable AI-Based Models through Correlation Analysis. *Mathematical Problems in Engineering* **2022**, 1-6. [Crossref]

200. Mehmet Güney Celbiş, Pui-hang Wong, Karima Kourtit, Peter Nijkamp. 2022. Impacts of the COVID-19 outbreak on older-age cohorts in European Labor Markets: A machine learning exploration of vulnerable groups. *Regional Science Policy & Practice* **30**. . [Crossref]

201. Rani Nooraeni, Jimmy Nickelson, Eko Rahmadian, Nugroho Puspito Yudho. 2022. New recommendation to predict export value using big data and machine learning technique. *Statistical Journal of the IAOS* **38**:1, 277-290. [Crossref]

202. Michael Greenstone, Guojun He, Ruixue Jia, Tong Liu. 2022. Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution. *American Economic Review: Insights* **4**:1, 54-70. [Abstract] [View PDF article] [PDF with links]

203. Avinash Kaur, Parminder Singh, Ranbir Singh Batth, Chee Peng Lim. 2022. Deep-Q learning-based heterogeneous earliest finish time scheduling algorithm for scientific workflows in cloud. *Software: Practice and Experience* **52**:3, 689-709. [Crossref]

204. Md Anisur Rahman, Mirko Duradoni, Andrea Guazzini. 2022. Identification and prediction of phubbing behavior: a data-driven approach. *Neural Computing and Applications* **34**:5, 3885-3894. [Crossref]

205. Yicai Huang, Jiayuan Chen, Qiannan Duan, Yunjin Feng, Run Luo, Wenjing Wang, Fenli Liu, Sifan Bi, Jianchao Lee. 2022. A fast antibiotic detection method for simplified pretreatment through spectra-based machine learning. *Frontiers of Environmental Science & Engineering* **16**:3. . [Crossref]

206. Nidhi Rajesh Mavani, Jarinah Mohd Ali, Suhaili Othman, M. A. Hussain, Haslaniza Hashim, Norliza Abd Rahman. 2022. Application of Artificial Intelligence in Food Industry—a Guideline. *Food Engineering Reviews* **14**:1, 134-175. [Crossref]

207. Helmut Wasserbacher, Martin Spindler. 2022. Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. *Digital Finance* **4**:1, 63-88. [Crossref]

208. Takuya Takata, Hajime Sasaki, Hiroko Yamano, Masashi Honma, Mayumi Shikano. 2022. Study on Horizon Scanning with a Focus on the Development of AI-Based Medical Products: Citation Network Analysis. *Therapeutic Innovation & Regulatory Science* **56**:2, 263-275. [Crossref]

209. Joshua O. S. Hunt, James N. Myers, Linda A. Myers. 2022. Improving Earnings Predictions and Abnormal Returns with Machine Learning. *Accounting Horizons* **36**:1, 131-149. [Crossref]

210. Florian Wozny. 2022. The Impact of COVID-19 on Airfares—A Machine Learning Counterfactual Analysis. *Econometrics* **10**:1, 8. [Crossref]

211. Jan Abrell, Mirjam Kosch, Sebastian Rausch. 2022. How effective is carbon pricing?—A machine learning approach to policy evaluation. *Journal of Environmental Economics and Management* **112**, 102589. [Crossref]

212. Ranik Raaen Wahlstrøm, Florentina Paraschiv, Michael Schürle. 2022. A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. *Computational Economics* **59**:3, 967-1004. [Crossref]

213. Haiming Feng, Ye Zhao, Weijian Yan, Xiaoping Wei, Junping Lin, Peng Jiang, Cheng Wang, Bin Li. 2022. Identification of Signature Genes and Characterizations of Tumor Immune Microenvironment and Tumor Purity in Lung Adenocarcinoma Based on Machine Learning. *Frontiers in Medicine* **9**. . [Crossref]

214. Vedant Bhardwaj, Param Bhavsar, Debasis Patnaik. Forecasting GDP per capita of OECD countries using machine learning and deep learning models 1-6. [Crossref]
215. A. V. L. N. Sujith, Naila Iqbal Qureshi, Venkata Harshavardhan Reddy Dornadula, Abinash Rath, Kolla Bhanu Prakash, Sitesh Kumar Singh. 2022. A Comparative Analysis of Business Machine Learning in Making Effective Financial Decisions Using Structural Equation Model (SEM). *Journal of Food Quality* **2022**, 1-7. [Crossref]
216. T. S. McElroy, Thomas Trimbur. 2022. Variable targeting and reduction in large vector autoregressions with applications to workforce indicators. *Journal of Applied Statistics* **57**, 1-23. [Crossref]
217. Ti-Ching Peng, Chun-Chieh Wang. 2022. The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area. *Social Science Japan Journal* **25**:1, 3-28. [Crossref]
218. Jieying Gao, Huan Guo, Xin Xu. 2022. Multifactor Stock Selection Strategy Based on Machine Learning: Evidence from China. *Complexity* **2022**, 1-17. [Crossref]
219. Morgan Henderson, Fei Han, Chad Perman, Howard Haft, Ian Stockwell. 2022. Predicting avoidable hospital events in Maryland. *Health Services Research* **57**:1, 192-199. [Crossref]
220. ANDREAS FUSTER, PAUL GOLDSMITH-PINKHAM, TARUN RAMADORAI, ANSGAR WALTHER. 2022. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance* **77**:1, 5-47. [Crossref]
221. Miaoying Shi, Jintao Xu, Shilei Liu, Zhenci Xu. 2022. Productivity-Based Land Suitability and Management Sensitivity Analysis: The Eucalyptus E. urophylla × E. grandis Case. *Forests* **13**:2, 340. [Crossref]
222. Gema Sakti Raspati, Stian Bruaset, Camillo Bosco, Lars Mushom, Birgitte Johannessen, Rita Ugarelli. 2022. A Risk-Based Approach in Rehabilitation of Water Distribution Networks. *International Journal of Environmental Research and Public Health* **19**:3, 1594. [Crossref]
223. Abdus Samad Azad, Rajalingam Sokkalingam, Hanita Daud, Sajal Kumar Adhikary, Hifsa Khurshid, Siti Nur Athirah Mazlan, Muhammad Babar Ali Rabbani. 2022. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. *Sustainability* **14**:3, 1843. [Crossref]
224. Jason Chun Yu Wong, Brian Blankenship, Johannes Urpelainen, Kanika Balani, Karthik Ganesan, Kapardhi Bharadwaj. 2022. Understanding electricity billing preferences in rural and urban India: Evidence from a conjoint experiment. *Energy Economics* **106**, 105735. [Crossref]
225. Steffen Kinkel, Marco Baumgartner, Enrica Cherubini. 2022. Prerequisites for the adoption of AI technologies in manufacturing – Evidence from a worldwide sample of manufacturing companies. *Technovation* **110**, 102375. [Crossref]
226. Thaynã França, Arthur Martins Barbosa Braga, Helon Vicente Hultmann Ayala. 2022. Feature engineering to cope with noisy data in sparse identification. *Expert Systems with Applications* **188**, 115995. [Crossref]
227. Dweepobotee Brahma, Debasri Mukherjee. 2022. Early warning signs: targeting neonatal and infant mortality using machine learning. *Applied Economics* **54**:1, 57-74. [Crossref]
228. Philipp Kugler. 2022. The role of wage beliefs in the decision to become a nurse. *Health Economics* **31**:1, 94-111. [Crossref]
229. Gareth Macartney. Econometrics in Litigation: Challenges at Class Certification 311-346. [Crossref]
230. Jingyi Qiu. Data Selection and Machine Learning Algorithm Application Under the Background of Big Data 96-103. [Crossref]

231. Daniel Hain, Roman Jurowetzki. Introduction to Rare-Event Predictive Modeling for Inferential Statisticians—A Hands-On Application in the Prediction of Breakthrough Patents 49-83. [Crossref]
232. Yinping Ji, Deepmala Karki. Construction of a Personalized English Learning System Based on Machine Learning 503-510. [Crossref]
233. Bui Thanh Khoa, Ho Nhat Anh, Nguyen Minh Ly, Nguyen Xuan Truong. A Study on Buying Attitude on Facebook in the Digital Transformation Era: A Machine Learning Application 497-510. [Crossref]
234. Bui Thanh Khoa, Nguyen Thi Trang Oanh, Vo Thi Thao Uyen, Dang Cuu Hanh Dung. Customer Loyalty in the Covid-19 Pandemic: The Application of Machine Learning in Survey Data 419-429. [Crossref]
235. Zhigang Qiu, Xiaolin Huo, Yue Dai. Development of FinTech in Academia 71-84. [Crossref]
236. Dieudonné Tchuente, Serge Nyawa. 2022. Real estate price estimation in French cities using geocoding and machine learning. *Annals of Operations Research* **308**:1-2, 571-608. [Crossref]
237. Hans Visser, Niels Evers, Arjan Bontsema, Jasmijn Rost, Arie de Niet, Paul Vethman, Sido Mylius, Annelotte van der Linden, Joost van den Roovaart, Frank van Gaalen, Roel Knoben, Hendrika J. de Lange. 2022. What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. *Water Research* **208**, 117851. [Crossref]
238. Abigail R. Cartus, Ashley I. Naimi, Katherine P. Himes, Marian Jarlenski, Sara M. Parisi, Lisa M. Bodnar. 2022. Can Ensemble Machine Learning Improve the Accuracy of Severe Maternal Morbidity Screening in a Perinatal Database?. *Epidemiology* **33**:1, 95-104. [Crossref]
239. Li-Chin Chen, Ji-Tian Sheu, Yuh-Jue Chuang, Yu Tsao. 2022. Predicting the Travel Distance of Patients to Access Healthcare Using Deep Neural Networks. *IEEE Journal of Translational Engineering in Health and Medicine* **10**, 1-11. [Crossref]
240. Simone Plak, Ilja Cornelisz, Martijn Meeter, Chris Klaveren. 2022. Early warning systems for more effective student counselling in higher education: Evidence from a Dutch field experiment. *Higher Education Quarterly* **76**:1, 131-152. [Crossref]
241. Mark D. Verhagen. 2022. A Pragmatist's Guide to Using Prediction in the Social Sciences. *Socius: Sociological Research for a Dynamic World* **8**, 237802312210817. [Crossref]
242. Montserrat González Garibay, Andrej Srakar, Tjaša Bartolj, Jože Sambt. 2022. Does Machine Learning Offer Added Value Vis-à-Vis Traditional Statistics? An Exploratory Study on Retirement Decisions Using Data from the Survey of Health, Ageing, and Retirement in Europe (SHARE). *Mathematics* **10**:1, 152. [Crossref]
243. M. EMEC, M. H. OZCANHAN. 2022. A Hybrid Deep Learning Approach for Intrusion Detection in IoT Networks. *Advances in Electrical and Computer Engineering* **22**:1, 3-12. [Crossref]
244. Benjamin Bluhm, Jannic Alexander Cutura. 2022. Econometrics at Scale: Spark up Big Data in Economics. *Journal of Data Science* **49**, 413-436. [Crossref]
245. Emilio Lehoucq. 2022. Do Americans Think the Digital Economy is Fair? Using Supervised Learning to Explore Evaluations of Predictive Automation. *Journal of Data Science* **35**, 381-399. [Crossref]
246. Baojun Yu, Changming Li, Nawazish Mirza, Muhammad Umar. 2022. Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. *Technological Forecasting and Social Change* **174**, 121255. [Crossref]
247. Augusto Cerqua, Marco Letta. 2022. Local inequalities of the COVID-19 crisis. *Regional Science and Urban Economics* **92**, 103752. [Crossref]
248. Gianluca Gabrielli, Alice Medioli, Paolo Andrei. 2022. Accounting and Big Data: Trends, opportunities and direction for practitioners and researchers. *FINANCIAL REPORTING* :2, 89-112. [Crossref]

249. Béatrice Boulu-Reshef. Possible in Economics 1096-1103. [Crossref]
250. Zheng Li, Bo Zhou, David A. Hensher. 2022. Forecasting automobile gasoline demand in Australia using machine learning-based regression. *Energy* **239**, 122312. [Crossref]
251. S. Karthik, Robin Singh Bhadoria, Jeong Gon Lee, Arun Kumar Sivaraman, Sovan Samanta, A. Balasundaram, Brijesh Kumar Chaurasia, S. Ashokkumar. 2022. Prognostic Kalman Filter Based Bayesian Learning Model for Data Accuracy Prediction. *Computers, Materials & Continua* **72**:1, 243-259. [Crossref]
252. Philipp Adämmer, Rainer Alexander Schüssler. 2022. Detecting the Sparsity Levels of Economic Time Series: On the Impact of Noise. *SSRN Electronic Journal* **19**. . [Crossref]
253. Andreas Born, Aljoscha Janssen. 2022. Does a district mandate matter for the behavior of politicians? An analysis of roll-call votes and parliamentary speeches. *European Journal of Political Economy* **71**, 102070. [Crossref]
254. Luca Coraggio, Marco Pagano, Annalisa Scognamiglio, Joacim Tåg. 2022. JAQ of All Trades: Job Mismatch, Firm Productivity and Managerial Quality. *SSRN Electronic Journal* **101**. . [Crossref]
255. Falco Bargagli Stoffi, Massimo Riccaboni, Armando Rungi. 2022. Machine Learning for Zombie Hunting: Predicting Distress from Firms' Accounts and Missing Values. *SSRN Electronic Journal* **83**. . [Crossref]
256. Yacine Ait-Sahalia, Jianqing Fan, Lirong Xue, Yifeng Zhou. 2022. How and When are High-Frequency Stock Returns Predictable?. *SSRN Electronic Journal* **217**. . [Crossref]
257. Wei Chien Ng, Yu Qing Soong, Sin Yin Teh. Machine Learning in Food Security and Sustainability 1-17. [Crossref]
258. Elliott Isaac, Haibin Jiang. 2022. Tax-Based Marriage Incentives in the Affordable Care Act. *SSRN Electronic Journal* **48**. . [Crossref]
259. Kathleen Li, Christophe Van den Bulte. 2022. Augmented Difference-in-Differences. *SSRN Electronic Journal* **59**. . [Crossref]
260. Veli Andirin, Yusuf Neggers, Mehdi Shadmehr, Jesse Shapiro. 2022. Measuring the Tolerance of the State: Theory and Application to Protest. *SSRN Electronic Journal* **96**. . [Crossref]
261. Emanuel Kohlscheen, Richhild Moessner. 2022. Changing Electricity Markets: Quantifying the Price Effects of Greening the Energy Matrix. *SSRN Electronic Journal* **31**. . [Crossref]
262. Jens Ludwig, Sendhil Mullainathan. 2022. Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific Discovery. *SSRN Electronic Journal* **18**. . [Crossref]
263. Marcelin Joanis, Andrea Lodi, Igor Sadoune. 2022. On Deep Generative Modeling in Economics: An Application with Public Procurement Data. *SSRN Electronic Journal* **31**. . [Crossref]
264. Yacine Ait-Sahalia, Jianqing Fan, Lirong Xue, Yifeng Zhou. 2022. How and When are High-Frequency Stock Returns Predictable?. *SSRN Electronic Journal* **217**. . [Crossref]
265. Chun-Cheng Wei, Li Sun, Shih-Pang Tseng. The Demand Analysis and Forecast of APP-based Taxi Service via Machine Learning 1-3. [Crossref]
266. Francesco Bloise, Paolo Brunori, Patrizio Piraino. 2021. Estimating intergenerational income mobility on sub-optimal data: a machine learning approach. *The Journal of Economic Inequality* **19**:4, 643-665. [Crossref]
267. Furkan Kartal, Uğur Özveren. 2021. An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass. *Carbohydrate Polymer Technologies and Applications* **2**, 100148. [Crossref]

268. Martin Hirche, Paul W. Farris, Luke Greenacre, Yiran Quan, Susan Wei. 2021. Predicting Under- and Overperforming SKUs within the Distribution–Market Share Relationship. *Journal of Retailing* **97**:4, 697-714. [Crossref]
269. Onder Ozgur, Erdal Tanas Karagol, Fatih Cemil Ozbugday. 2021. Machine learning approach to drivers of bank lending: evidence from an emerging economy. *Financial Innovation* **7**:1. . [Crossref]
270. Mudabbir Ali, Asad Masood Khattak, Zain Ali, Bashir Hayat, Muhammad Idrees, Zeeshan Pervez, Kashif Rizwan, Tae-Eung Sung, Ki-Il Kim. 2021. Estimation and Interpretation of Machine Learning Models with Customized Surrogate Model. *Electronics* **10**:23, 3045. [Crossref]
271. Jungsun Kim, Jaewoong Won, Hyeongsoon Kim, Joonghyeok Heo. 2021. Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea. *Sustainability* **13**:23, 13088. [Crossref]
272. Mehmet Güney Celbiş, Pui-Hang Wong, Karima Kourtit, Peter Nijkamp. 2021. Innovativeness, Work Flexibility, and Place Characteristics: A Spatial Econometric and Machine Learning Approach. *Sustainability* **13**:23, 13426. [Crossref]
273. Wenjing Lyu, Jin Liu. 2021. Artificial Intelligence and emerging digital technologies in the energy sector. *Applied Energy* **303**, 117615. [Crossref]
274. Jiaming Zhang, Zhanfeng Li, Xinyuan Song, Hanwen Ning. 2021. Deep Tobit networks: A novel machine learning approach to microeconometrics. *Neural Networks* **144**, 279-296. [Crossref]
275. Antonio Rodríguez Andrés, Abraham Otero, Voxi Heinrich Amavilah. 2021. Using deep learning neural networks to predict the knowledge economy index for developing and emerging economies. *Expert Systems with Applications* **184**, 115514. [Crossref]
276. Zach Warner, J. Andrew Harris, Michelle Brown, Christian Arnold. 2021. Hidden in plain sight? Irregularities on statutory forms and electoral fraud. *Electoral Studies* **74**, 102411. [Crossref]
277. Guan-Yuan Wang. 2021. The Brand Effect: A Case Study in Taiwan Second-Hand Smartphone Market. *Journal of Social and Economic Statistics* **10**:1-2, 30-42. [Crossref]
278. Maja Micevska. 2021. Revisiting forced migration: A machine learning perspective. *European Journal of Political Economy* **70**, 102044. [Crossref]
279. Steven Lehrer, Tian Xie, Tao Zeng. 2021. Does High-Frequency Social Media Data Improve Forecasts of Low-Frequency Consumer Confidence Measures?. *Journal of Financial Econometrics* **19**:5, 910-933. [Crossref]
280. Deon Filmer, Vatsal Nahata, Shwetlena Sabarwal. Preparation, Practice, and Beliefs: A Machine Learning Approach to Understanding Teacher Effectiveness **1**, . [Crossref]
281. Sree Teja Buddaraju, Ananya Bardhan, Ramya Sri Boddu, Simranjit Kaur, Thangarajah Akilan. Remote Sensing-based Socioeconomic Analysis using Task-driven Transfer Learning and Regression 1-7. [Crossref]
282. Chengge Wu. Empowering Financial Technical Analysis using Computer Vision Techniques 179-184. [Crossref]
283. Alessandra Garbero, Bia Carneiro, Giuliano Resce. 2021. Harnessing the power of machine learning analytics to understand food systems dynamics across development projects. *Technological Forecasting and Social Change* **172**, 121012. [Crossref]
284. Javier Ortega-Bastida, Antonio Javier Gallego, Juan Ramón Rico-Juan, Pedro Albarrán. 2021. A multimodal approach for regional GDP prediction using social media activity and historical information. *Applied Soft Computing* **111**, 107693. [Crossref]
285. Stelios Michalopoulos, Melanie Meng Xue. 2021. Folklore. *The Quarterly Journal of Economics* **136**:4, 1993-2046. [Crossref]
286. Matthew C. Harding, Carlos Lamarche. 2021. Small Steps with Big Data: Using Machine Learning in Energy and Environmental Economics. *Annual Review of Resource Economics* **13**:1, 469-488. [Crossref]

287. Bryan Kelly, Asaf Manela, Alan Moreira. 2021. Text Selection. *Journal of Business & Economic Statistics* **39**:4, 859-879. [Crossref]

288. Prodosh E. Simlai. 2021. Predicting owner-occupied housing values using machine learning: an empirical investigation of California census tracts data. *Journal of Property Research* **38**:4, 305-336. [Crossref]

289. Yonghui Dai, Tao Wang. 2021. Prediction of customer engagement behaviour response to marketing posts based on machine learning. *Connection Science* **33**:4, 891-910. [Crossref]

290. Francesco Decarolis, Gabriele Rovigatti. 2021. From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising. *American Economic Review* **111**:10, 3299-3327. [Abstract] [View PDF article] [PDF with links]

291. M. Salti, E.E. Kangal, O. Aydogdu. 2021. Evolution of CMB temperature in a Chaplygin gas model from deep learning perspective. *Astronomy and Computing* **37**, 100504. [Crossref]

292. Jesús Gonzalo, Jean-Yves Pitarakis. 2021. Spurious relationships in high-dimensional systems with strong or mild persistence. *International Journal of Forecasting* **37**:4, 1480-1497. [Crossref]

293. İlhan KOYUNCU, Abdullah Faruk KILIÇ. 2021. Classification of Scale Items with Exploratory Graph Analysis and Machine Learning Methods. *International Journal of Assessment Tools in Education* **8**:4, 928-947. [Crossref]

294. Seyedehmaryam Moosavi, Otilia Manta, Yaser A. El-Badry, Enas E. Hussein, Zeinhom M. El-Bahy, Noor fariza Binti Mohd Fawzi, Jaunius Urbonavičius, Seyed Mohammad Hossein Moosavi. 2021. A Study on Machine Learning Methods' Application for Dye Adsorption Prediction onto Agricultural Waste Activated Carbon. *Nanomaterials* **11**:10, 2734. [Crossref]

295. Muhammad Ali Musarat, Wesam Salah Alaloul, Muhammad Babar Ali Rabbani, Mujahid Ali, Muhammad Altaf, Roman Fediuk, Nikolai Vatin, Sergey Klyuev, Hamna Bukhari, Alishba Sadiq, Waqas Rafiq, Waqas Farooq. 2021. Kabul River Flow Prediction Using Automated ARIMA Forecasting: A Machine Learning Approach. *Sustainability* **13**:19, 10720. [Crossref]

296. Keith Leavitt, Kira Schabram, Prashanth Hariharan, Christopher M. Barnes. 2021. Ghost in the Machine: On Organizational Theory in the Age of Machine Learning. *Academy of Management Review* **46**:4, 750-777. [Crossref]

297. Heather A. Haveman, Joseph T. Mahoney, Elizabeth Mannix. 2021. The Evolving Science of Organization: Theory Matters. *Academy of Management Review* **46**:4, 660-666. [Crossref]

298. Vitor Miguel Ribeiro. 2021. Sulfur dioxide emissions in Portugal: Prediction, estimation and air quality regulation using machine learning. *Journal of Cleaner Production* **317**, 128358. [Crossref]

299. David Bholat, Daniel Susskind. 2021. The assessment: artificial intelligence and financial services. *Oxford Review of Economic Policy* **37**:3, 417-434. [Crossref]

300. Bonnie G Buchanan, Danika Wright. 2021. The impact of machine learning on UK financial services. *Oxford Review of Economic Policy* **37**:3, 537-563. [Crossref]

301. Giovanni Cerulli. 2021. Improving econometric prediction by machine learning. *Applied Economics Letters* **28**:16, 1419-1425. [Crossref]

302. Ebru Tomris AYDOĞAN, Esra KARADENİZ, Mehmet Güney CELBİŞ. 2021. Türkiye'de Girişimcilik ve Sürdürülebilir Bölgesel Kalkınma: Makine Öğrenmesi Yaklaşımlarından Elde Edilen Bulgular. *Ekonomi, Politika & Finans Araştırmaları Dergisi* **6**:3, 882-911. [Crossref]

303. Hing Ling Chan, Minling Pan. 2021. Fishing trip cost modeling using generalized linear model and machine learning methods – A case study with longline fisheries in the Pacific and an application in Regulatory Impact Analysis. *PLOS ONE* **16**:9, e0257027. [Crossref]

304. Dragos-Cristian Vasilescu, Michael Filzmoser. 2021. Machine invention systems: a (r)evolution of the invention process?. *AI & SOCIETY* **36**:3, 829-837. [Crossref]

305. Joshua O.S. Hunt, David M. Rosser, Stephen P. Rowe. 2021. Using machine learning to predict auditor switches: How the likelihood of switching affects audit quality among non-switching clients. *Journal of Accounting and Public Policy* **40**:5, 106785. [Crossref]
306. Christoph Breunig, Iuliia Grabova, Peter Haan, Felix Weinhardt, Georg Weizsäcker. 2021. Long-run expectations of households. *Journal of Behavioral and Experimental Finance* **31**, 100535. [Crossref]
307. Virgilio Galdo, Yue Li, Martin Rama. 2021. Identifying urban areas by combining human judgment and machine learning: An application to India. *Journal of Urban Economics* **125**, 103229. [Crossref]
308. Feras A. Batarseh, Munisamy Gopinath, Anderson Monken, Zhengrong Gu. 2021. Public policymaking for international agricultural trade using association rules and ensemble machine learning. *Machine Learning with Applications* **5**, 100046. [Crossref]
309. Yaohao Peng, Pedro Henrique Melo Albuquerque, Herbert Kimura, Cayan Atreio Portela Bárcena Saavedra. 2021. Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications* **5**, 100060. [Crossref]
310. Shiho Kino, Yu-Tien Hsu, Koichiro Shiba, Yung-Shin Chien, Carol Mita, Ichiro Kawachi, Adel Daoud. 2021. A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM - Population Health* **15**, 100836. [Crossref]
311. Frederick Andrés Mendoza-Lozano, Jose Wilmar Quintero-Peña, Oscar Leonardo Acevedo-Pabón, Jose Félix García-Rodríguez. 2021. Fundamentación teórica para la creación de un programa académico de ingeniería y ciencia de datos: una aplicación bibliométrica. *Aibi revista de investigación, administración e ingeniería* **9**:3, 49-58. [Crossref]
312. Rafael Quintana. 2021. Who Belongs in School? Using Statistical Learning Techniques to Identify Linear, Nonlinear and Interactive Effects. *The Quantitative Methods for Psychology* **17**:3, 312-328. [Crossref]
313. James T. Bang, Atin Basuchoudhary, Aniruddha Mitra. 2021. Validating Game-Theoretic Models of Terrorism: Insights from Machine Learning. *Games* **12**:3, 54. [Crossref]
314. Ka Shing Cheung, Julian TszKin Chan, Sijie Li, Chung Yim Yiu. 2021. Anchoring and Asymmetric Information in the Real Estate Market: A Machine Learning Approach. *Journal of Risk and Financial Management* **14**:9, 423. [Crossref]
315. Mohamed Elhag Mohamed Abo, Norisma Idris, Rohana Mahmud, Atika Qazi, Ibrahim Abaker Targio Hashem, Jaafar Zubairu Maitama, Usman Naseem, Shah Khalid Khan, Shuiqing Yang. 2021. A Multi-Criteria Approach for Arabic Dialect Sentiment Analysis for Online Reviews: Exploiting Optimal Machine Learning Algorithm Selection. *Sustainability* **13**:18, 10018. [Crossref]
316. Ruixin Liang, Joanne Yip, Winnie Yu, Lihua Chen, Newman Lau. 2021. Finite Element-Based Machine Learning Method to Predict Breast Displacement during Running. *AATCC Journal of Research* **8**:1_suppl, 69-74. [Crossref]
317. Mengying Shang, Yonghua Zhou, Hamido Fujita. 2021. Deep reinforcement learning with reference system to handle constraints for energy-efficient train control. *Information Sciences* **570**, 708-721. [Crossref]
318. Alexandre Godzinski, Milena Suarez Castillo. 2021. Disentangling the effects of air pollutants with many instruments. *Journal of Environmental Economics and Management* **109**, 102489. [Crossref]
319. Akash Malhotra. 2021. A hybrid econometric–machine learning approach for relative importance analysis: prioritizing food policy. *Eurasian Economic Review* **11**:3, 549-581. [Crossref]
320. Mehmet Güney CELBİŞ. 2021. Social Networks, Female Unemployment, and the Urban-Rural Divide in Turkey: Evidence from Tree-Based Machine Learning Algorithms. *Sosyoekonomi* . [Crossref]

321. Claire S.H. Lim, James M. Snyder. 2021. What Shapes the Quality and Behavior of Government Officials? Institutional Variation in Selection and Retention Methods. *Annual Review of Economics* **13**:1, 87-109. [Crossref]
322. Tiffany Jiang. 2021. Using Machine Learning to Analyze Merger Activity. *Frontiers in Applied Mathematics and Statistics* **7**. . [Crossref]
323. Molly McCann-Pineo, Julia Ruskin, Rehana Rasul, Eugene Vortsman, Kristin Bevilacqua, Samantha S. Corley, Rebecca M. Schwartz. 2021. Predictors of emergency department opioid administration and prescribing: A machine learning approach. *The American Journal of Emergency Medicine* **46**, 217-224. [Crossref]
324. Marja-Liisa Halko, Olli Lappalainen, Lauri Sääksvuori. 2021. Do non-choice data reveal economic preferences? Evidence from biometric data and compensation-scheme choice. *Journal of Economic Behavior & Organization* **188**, 87-104. [Crossref]
325. Sugato Chakravarty, Douglas J. Cumming, Samuele Murtinu, Vittoria G. Scalera, Christian Schwens. 2021. Exploring the next generation of international entrepreneurship. *Journal of World Business* **56**:5, 101229. [Crossref]
326. Mehmet Güney Celbiş. 2021. A machine learning approach to rural entrepreneurship. *Papers in Regional Science* **100**:4, 1079-1104. [Crossref]
327. Paola Bertoli, Veronica Grembi. 2021. Territorial differences in access to prenatal care and health at birth. *Health Policy* **125**:8, 1092-1099. [Crossref]
328. Andreu Arenas, Caterina Calsamiglia, Annalisa Loviglio. 2021. What is at stake without high-stakes exams? Students' evaluation and admission to college at the time of COVID-19. *Economics of Education Review* **83**, 102143. [Crossref]
329. Misuk Kim. 2021. Adaptive trading system integrating machine learning and back-testing: Korean bond market case. *Expert Systems with Applications* **176**, 114767. [Crossref]
330. Levi Altringer, Jordan Navin, Michael J. Begier, Stephanie A. Shwiff, Aaron Anderson. 2021. Estimating wildlife strike costs at US airports: A machine learning approach. *Transportation Research Part D: Transport and Environment* **97**, 102907. [Crossref]
331. Praphula Kumar Jain, Rajendra Pamula, Gautam Srivastava. 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review* **41**, 100413. [Crossref]
332. Dong-sup Kim, Seungwoo Shin. 2021. THE ECONOMIC EXPLAINABILITY OF MACHINE LEARNING AND STANDARD ECONOMETRIC MODELS-AN APPLICATION TO THE U.S. MORTGAGE DEFAULT RISK. *International Journal of Strategic Property Management* **25**:5, 396-412. [Crossref]
333. Jian Chen, Ani L. Katchova, Chenxi Zhou. 2021. Agricultural loan delinquency prediction using machine learning methods. *International Food and Agribusiness Management Review* **24**:5, 797-812. [Crossref]
334. Jake M. Hofman, Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J. Salganik, Simine Vazire, Alessandro Vespignani, Tal Yarkoni. 2021. Integrating explanation and prediction in computational social science. *Nature* **595**:7866, 181-188. [Crossref]
335. Luca Barbaglia, Sebastiano Manzan, Elisa Tosetti. 2021. Forecasting Loan Default in Europe with Machine Learning. *Journal of Financial Econometrics* **29**. . [Crossref]
336. Devesh Raval, Ted Rosenbaum, Nathan E. Wilson. 2021. How do machine learning algorithms perform in predicting hospital choices? evidence from changing environments. *Journal of Health Economics* **78**, 102481. [Crossref]

337. Ronald Richman. 2021. AI in actuarial science – a review of recent advances – part 1. *Annals of Actuarial Science* **15**:2, 207-229. [Crossref]
338. Mark Musumba, Naureen Fatema, Shahriar Kibriya. 2021. Prevention Is Better Than Cure: Machine Learning Approach to Conflict Prediction in Sub-Saharan Africa. *Sustainability* **13**:13, 7366. [Crossref]
339. Benjamin Lucas, R. Elena Francu, James Goulding, John Harvey, Georgiana Nica-Avram, Bertrand Perrat. 2021. A Note on Data-driven Actor-differentiation and SDGs 2 and 12: Insights from a Food-sharing App. *Research Policy* **50**:6, 104266. [Crossref]
340. Ya Chen, Mike G. Tsionas, Valentin Zelenyuk. 2021. LASSO+DEA for small and big wide data. *Omega* **102**, 102419. [Crossref]
341. David Easley, Marcos López de Prado, Maureen O'Hara, Zhibai Zhang. 2021. Microstructure in the Machine Age. *The Review of Financial Studies* **34**:7, 3316-3363. [Crossref]
342. Itay Goldstein, Chester S Spatt, Mao Ye. 2021. Big Data in Finance. *The Review of Financial Studies* **34**:7, 3213-3225. [Crossref]
343. Isil Erel, Léa H Stern, Chenhao Tan, Michael S Weisbach. 2021. Selecting Directors Using Machine Learning. *The Review of Financial Studies* **34**:7, 3226-3264. [Crossref]
344. Kai He. 2021. Prediction Model of Juvenile Football Players' Sports Injury Based on Text Classification Technology of Machine Learning. *Mobile Information Systems* **2021**, 1-10. [Crossref]
345. Andrew T. Tredennick, Giles Hooker, Stephen P. Ellner, Peter B. Adler. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology* **102**:6. . [Crossref]
346. Giovanni Di Franco, Michele Santurro. 2021. Machine learning, artificial neural networks and social research. *Quality & Quantity* **55**:3, 1007-1025. [Crossref]
347. Ankit Thakkar, Kinjal Chaudhari. 2021. A Comprehensive Survey on Portfolio Optimization, Stock Price and Trend Prediction Using Particle Swarm Optimization. *Archives of Computational Methods in Engineering* **28**:4, 2133-2164. [Crossref]
348. Xiaolu Zhou, Weitian Tong. 2021. Learning with self-attention for rental market spatial dynamics in the Atlanta metropolitan area. *Earth Science Informatics* **14**:2, 837-845. [Crossref]
349. Daniel Jacob. 2021. CATE meets ML. *Digital Finance* **3**:2, 99-148. [Crossref]
350. Chiara Binelli. 2021. Estimating Causal Effects When the Treatment Affects All Subjects Simultaneously: An Application. *Big Data and Cognitive Computing* **5**:2, 22. [Crossref]
351. Shenhao Wang, Qingyi Wang, Nate Bailey, Jinhua Zhao. 2021. Deep neural networks for choice analysis: A statistical learning theory perspective. *Transportation Research Part B: Methodological* **148**, 60-81. [Crossref]
352. Wenchao Li. 2021. The "miseries" of sex imbalance: Evidence using subjective well-being data. *Journal of Development Economics* **151**, 102634. [Crossref]
353. Luca Panzone, Alistair Ulph, Francisco Areal, Valeria Grippo. 2021. A ridge regression approach to estimate the relationship between landfill taxation and waste collection and disposal in England. *Waste Management* **129**, 95-110. [Crossref]
354. Xianxu Bai, Chao Tang. 2021. Dynamic RC operator-based hysteresis model of MR dampers. *Smart Materials and Structures* . [Crossref]
355. Nika Haghtalab, Matthew O. Jackson, Ariel D. Procaccia. 2021. Belief polarization in a complex world: A learning theory perspective. *Proceedings of the National Academy of Sciences* **118**:19. . [Crossref]
356. Justin Grimmer, Margaret E. Roberts, Brandon M. Stewart. 2021. Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science* **24**:1, 395-419. [Crossref]

357. Yongtong Shao, Tao Xiong, Minghao Li, Dermot Hayes, Wendong Zhang, Wei Xie. 2021. China's Missing Pigs: Correcting China's Hog Inventory Data Using a Machine Learning Approach. *American Journal of Agricultural Economics* **103**:3, 1082-1098. [Crossref]
358. Kristian Bondo Hansen. 2021. Model Talk: Calculative Cultures in Quantitative Finance. *Science, Technology, & Human Values* **46**:3, 600-627. [Crossref]
359. Clement Nyamekye, Samuel Anim Ofosu, Richard Arthur, Gabriel Osei, Linda Boamah Appiah, Samuel Kwofie, Benjamin Ghansah, Dieter Bryniok. 2021. Evaluating the spatial and temporal variations of aquatic weeds (Biomass) on Lower Volta River using multi-sensor Landsat Images and machine learning. *Heliyon* **7**:5, e07080. [Crossref]
360. Yash Raj Shrestha, Vivianna Fang He, Phanish Puranam, Georg von Krogh. 2021. Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize?. *Organization Science* **32**:3, 856-880. [Crossref]
361. Prasanna Tantri. 2021. Fintech for the Poor: Financial Intermediation Without Discrimination*. *Review of Finance* **25**:2, 561-593. [Crossref]
362. Arthur Charpentier, Romuald Élie, Carl Remlinger. 2021. Reinforcement Learning in Economics and Finance. *Computational Economics* **70**. . [Crossref]
363. Indah Fitri Astuti, Lisa Faizah, Dyna Marisa Khairina, Dedy Cahyadi. A fuzzy Mamdani approach on community business loan feasibility assessment 438-442. [Crossref]
364. Rainer Schulz, Martin Wersing. 2021. Automated Valuation Services: A case study for Aberdeen in Scotland. *Journal of Property Research* **38**:2, 154-172. [Crossref]
365. Andreu Arenas. 2021. Human capital portability and international student migration. *Journal of Economic Geography* **21**:2, 195-229. [Crossref]
366. Shunqin Chen, Zhengfeng Guo, Xinlei Zhao. 2021. Predicting mortgage early delinquency with machine learning methods. *European Journal of Operational Research* **290**:1, 358-372. [Crossref]
367. Yogesh K. Dwivedi, Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan, Rohita Dwivedi, John Edwards, Aled Eirug, Vassilis Galanos, P. Vigneswara Ilavarasan, Marijn Janssen, Paul Jones, Arpan Kumar Kar, Hatice Kizgin, Bianca Kronemann, Banita Lal, Biagio Lucini, Rony Medaglia, Kenneth Le Meunier-FitzHugh, Leslie Caroline Le Meunier-FitzHugh, Santosh Misra, Emmanuel Mogaji, Sujeet Kumar Sharma, Jang Bahadur Singh, Vishnupriya Raghavan, Ramakrishnan Raman, Nripendra P. Rana, Spyridon Samothrakis, Jak Spencer, Kuttimani Tamilmani, Annie Tubadji, Paul Walton, Michael D. Williams. 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* **57**, 101994. [Crossref]
368. Ahmed Sameer El Khatib. 2021. Machine Learning and Finance. *International Journal for Innovation Education and Research* **9**:4, 29-55. [Crossref]
369. Hamidreza Taghvaee, Akshay Jain, Xavier Timoneda, Christos Liaskos, Sergi Abadal, Eduard Alarcón, Albert Cabellos-Aparicio. 2021. Radiation Pattern Prediction for Metasurfaces: A Neural Network-Based Approach. *Sensors* **21**:8, 2765. [Crossref]
370. Joseph Price, Kasey Buckles, Jacob Van Leeuwen, Isaac Riley. 2021. Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History* **80**, 101391. [Crossref]
371. Shenhao Wang, Baichuan Mo, Jinhua Zhao. 2021. Theory-based residual neural networks: A synergy of discrete choice models and deep neural networks. *Transportation Research Part B: Methodological* **146**, 333-358. [Crossref]

372. María E. Pérez-Pons, Javier Parra-Dominguez, Sigeru Omatu, Enrique Herrera-Viedma, Juan Manuel Corchado. 2021. Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. *Journal of Artificial Intelligence and Soft Computing Research* **12**:2, 79-100. [Crossref]

373. Joseph S Shapiro. 2021. The Environmental Bias of Trade Policy*. *The Quarterly Journal of Economics* **136**:2, 831-886. [Crossref]

374. Joshua Gallin, Raven Molloy, Eric Nielsen, Paul Smith, Kamila Sommer. 2021. Measuring aggregate housing wealth: New insights from machine learning #. *Journal of Housing Economics* **51**, 101734. [Crossref]

375. Oliver Lock, Michael Bain, Christopher Pettit. 2021. Towards the collaborative development of machine learning techniques in planning support systems – a Sydney example. *Environment and Planning B: Urban Analytics and City Science* **48**:3, 484-502. [Crossref]

376. Nino Antulov-Fantulin, Raffaele Lagravinese, Giuliano Resce. 2021. Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization* **183**, 681-699. [Crossref]

377. Francesco Bloise, Massimiliano Tancioni. 2021. Predicting the spread of COVID-19 in Italy using machine learning: Do socio-economic factors matter?. *Structural Change and Economic Dynamics* **56**, 310-329. [Crossref]

378. Chirag Kumar, Guillermo Podestá, Katherine Kilpatrick, Peter Minnett. 2021. A machine learning approach to estimating the error in satellite sea surface temperature retrievals. *Remote Sensing of Environment* **255**, 112227. [Crossref]

379. Renji George Amballoor, Shankar B. Naik. Utility-based Frequent Itemsets in Data Streams using Sliding Window 108-112. [Crossref]

380. Clement Nyamekye, Benjamin Ghansah, Emmanuel Agyapong, Samuel Kwofie. 2021. Mapping Changes in Artisanal and Small-Scale Mining (ASM) Landscape using Machine and Deep Learning algorithms. - A Proxy Evaluation of the 2017 Ban on ASM in Ghana. *Environmental Challenges* **5**, 100053. [Crossref]

381. Domonkos F. Vamossy. 2021. Investor emotions and earnings announcements. *Journal of Behavioral and Experimental Finance* **16**, 100474. [Crossref]

382. Toshiaki Aizawa. 2021. Decomposition of Improvements in Infant Mortality in Asian Developing Countries Over Three Decades. *Demography* **58**:1, 137-163. [Crossref]

383. Perry Sadorsky. 2021. A Random Forests Approach to Predicting Clean Energy Stock Prices. *Journal of Risk and Financial Management* **14**:2, 48. [Crossref]

384. Stuart Gabriel, Matteo Iacoviello, Chandler Lutz. 2021. A Crisis of Missed Opportunities? Foreclosure Costs and Mortgage Modification During the Great Recession. *The Review of Financial Studies* **34**:2, 864-906. [Crossref]

385. Daniele Bianchi, Matthias Büchner, Andrea Tamoni. 2021. Bond Risk Premiums with Machine Learning. *The Review of Financial Studies* **34**:2, 1046-1089. [Crossref]

386. Lan Gao, Wei Luo, Utsana Tonmukayakul, Marj Moodie, Gang Chen. 2021. Mapping MacNew Heart Disease Quality of Life Questionnaire onto country-specific EQ-5D-5L utility scores: a comparison of traditional regression models with a machine learning technique. *The European Journal of Health Economics* **174**. . [Crossref]

387. Andreas Fagereng, Martin Blomhoff Holm, Kjersti Næss Torstensen. 2021. Housing wealth in Norway, 1993–20151. *Journal of Economic and Social Measurement* **45**:1, 65-81. [Crossref]

388. Qingyuan Zhao, Trevor Hastie. 2021. Causal Interpretations of Black-Box Models. *Journal of Business & Economic Statistics* **39**:1, 272-281. [Crossref]

389. Marcelo C. Medeiros, Gabriel F. R. Vasconcelos, Álvaro Veiga, Eduardo Zilberman. 2021. Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics* **39**:1, 98-119. [Crossref]
390. Prithwiraj Choudhury, Ryan T. Allen, Michael G. Endres. 2021. Machine learning for pattern discovery in management research. *Strategic Management Journal* **42**:1, 30-57. [Crossref]
391. Donna B. Gilleskie. 2021. In sickness and in health, until death do us part: A case for theory. *Southern Economic Journal* **87**:3, 753-768. [Crossref]
392. Agostino Valier. Evaluating AVMs Performance. Beyond the Accuracy 1155-1164. [Crossref]
393. Achim Ahrens, Christopher Aitken, Jan Ditzen, Erkal Ersoy, David Kohns, Mark E. Schaffer. A Theory-Based Lasso for Time-Series Data 3-36. [Crossref]
394. Achim Ahrens, Christopher Aitken, Mark E. Schaffer. Using Machine Learning Methods to Support Causal Inference in Econometrics 23-52. [Crossref]
395. Reinaldo Padilha França, Ana Carolina Borges Monteiro, Rangel Arthur, Yuzo Iano. The Fundamentals and Potential for Cybersecurity of Big Data in the Modern World 51-73. [Crossref]
396. Reinaldo Padilha França, Ana Carolina Borges Monteiro, Rangel Arthur, Yuzo Iano. An Overview of the Machine Learning Applied in Smart Cities 91-111. [Crossref]
397. Peter Romero, Stephen Fitz. The Use of Psychometrics and Artificial Intelligence in Alternative Finance 511-587. [Crossref]
398. Falco J. Bargagli-Stoffi, Jan Niederreiter, Massimo Riccaboni. Supervised Learning for the Prediction of Firm Dynamics 19-41. [Crossref]
399. Ruixin Liang, Joanne Yip, Kai-Tsun Michael To, Yunli Fan. Machine Learning Approaches to Predict Scoliosis 116-121. [Crossref]
400. Béatrice Boulu-Reshef. Possible in Economics 1-8. [Crossref]
401. Jaehyun Yoon. 2021. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics* **57**:1, 247-265. [Crossref]
402. Louis J. Catania. The science and technologies of artificial intelligence (AI) 29-72. [Crossref]
403. Louis J. Catania. Current AI applications in medical therapies and services 199-291. [Crossref]
404. Samane Zare, Michael R. Thomsen, Rodolfo M. Nayga, Anthony Goudie. 2021. Use of Machine Learning to Determine the Information Value of a BMI Screening Program. *American Journal of Preventive Medicine* **319**. . [Crossref]
405. Christophre Georges, Javier Pereira. 2021. Market stability with machine learning agents. *Journal of Economic Dynamics and Control* **122**, 104032. [Crossref]
406. Lily Shen, Stephen Ross. 2021. Information value of property description: A Machine learning approach. *Journal of Urban Economics* **121**, 103299. [Crossref]
407. Michael Yeomans. 2021. A concrete example of construct construction in natural language. *Organizational Behavior and Human Decision Processes* **162**, 81-94. [Crossref]
408. Heidi A. Hanson, Claire L. Leiser, Gretchen Bandoli, Brad H. Pollock, Margaret R. Karagas, Daniel Armstrong, Ann Dozier, Nicole G. Weiskopf, Maureen Monaghan, Ann M. Davis, Elizabeth Eckstrom, Chunhua Weng, Jonathan N. Tobin, Frederick Kaskel, Mark R. Schleiss, Peter Szilagyi, Carrie Dykes, Dan Cooper, Shari L. Barkin. 2021. Charting the life course: Emerging opportunities to advance scientific approaches using life course research. *Journal of Clinical and Translational Science* **5**:1. . [Crossref]
409. Shuyin Li, Yang Liu. 2021. News Video Title Extraction Algorithm Based on Deep Learning. *IEEE Access* **9**, 12143-12157. [Crossref]

410. Weihong Zhou, Yingjie Wang, Xiaoping Gu, Zhong-Ping Feng, Kang Lee, Yuzhu Peng, Andrew Barszczyk. 2021. Importance of general adiposity, visceral adiposity and vital signs in predicting blood biomarkers using machine learning. *International Journal of Clinical Practice* **75**:1. . [Crossref]
411. Elizabeth W. Diemer, James I. Hudson, Kristin N. Javaras. 2021. More (Adjustment) Is Not Always Better: How Directed Acyclic Graphs Can Help Researchers Decide Which Covariates to Include in Models for the Causal Relationship between an Exposure and an Outcome in Observational Research. *Psychotherapy and Psychosomatics* **90**:5, 289-298. [Crossref]
412. Kristof Lommers, Ouns El Harzli, Jack Kim. 2021. Confronting Machine Learning with Financial Research. *SSRN Electronic Journal* **1**. . [Crossref]
413. Amitabh Chandra, Evan Flack, Ziad Obermeyer. 2021. The Health Costs of Cost-Sharing. *SSRN Electronic Journal* **21**. . [Crossref]
414. Itay Goldstein, Chester S. Spatt, Mao Ye. 2021. Big Data in Finance. *SSRN Electronic Journal* **5**. . [Crossref]
415. Daniel Jacob. 2021. CATE Meets ML - Conditional Average Treatment Effect and Machine Learning. *SSRN Electronic Journal* **47**. . [Crossref]
416. Deni Mazrekaj, Vítězslav Titl, Fritz Schiltz. 2021. Identifying Politically Connected Firms: A Machine Learning Approach. *SSRN Electronic Journal* **322**. . [Crossref]
417. Albert Gyamfi. Determining Appropriate Social Media Sites for Knowledge Sharing 16-42. [Crossref]
418. Michael Peichl, Stephan Thober, Luis Samaniego, Bernd Hansjürgens, Andreas Marx. 2021. Machine-learning methods to assess the effects of a non-linear damage spectrum taking into account soil moisture on winter wheat yields in Germany. *Hydrology and Earth System Sciences* **25**:12, 6523-6545. [Crossref]
419. Hazik Mohamed. I-FinTech and Its Value Proposition for Islamic Asset and Wealth Management 249-266. [Crossref]
420. Hakan Kahyaoglu. The Impact of Artificial Intelligence on Central Banking and Monetary Policies 83-98. [Crossref]
421. Mohammadsaleh Saadatmand, Tuğrul U. Daim. Technology Intelligence Map: Finance Machine Learning 337-356. [Crossref]
422. Tony Liu, Lyle Ungar, Konrad Kording. 2021. Quantifying causality in data science with quasi-experiments. *Nature Computational Science* **1**:1, 24-32. [Crossref]
423. Mehmet Güney Celbiş. Applications of Machine Learning Models in Regional and Demographic Economic Analysis: A Literature Survey 219-229. [Crossref]
424. Lucia Alessi, Roberto Savona. Machine Learning for Financial Stability 65-87. [Crossref]
425. Kaustubh Arun Bhavsar, Ahed Abugabah, Jimmy Singla, Ahmad Ali AlZubi, Ali Kashif Bashir, Nikita. 2021. A Comprehensive Review on Medical Diagnosis Using Machine Learning. *Computers, Materials & Continua* **67**:2, 1997-2014. [Crossref]
426. Tim de Silva, David Thesmar. 2021. The Complementarity Between Man and Machine in Forecasting. *SSRN Electronic Journal* **64**. . [Crossref]
427. Christophe Hurlin, Christophe Perignon, Sébastien Saurin. 2021. The Fairness of Credit Scoring Models. *SSRN Electronic Journal* **11**. . [Crossref]
428. J. Michelle Brock, Ralph De Haas. 2021. Discriminatory Lending: Evidence from Bankers in the Lab. *SSRN Electronic Journal* **30**. . [Crossref]
429. Jake Krupa, Miguel Minutti-Meza. 2021. Regression and Machine Learning Methods to Predict Discrete Outcomes in Accounting Research. *SSRN Electronic Journal* **3466478**. . [Crossref]

430. Xiang Zheng. 2021. How can Innovation Screening be Improved? A Machine Learning Analysis with Economic Consequences for Firm Performance. *SSRN Electronic Journal* **31**. . [Crossref]
431. Madhura Dasgupta, Samarth Gupta. 2021. Determinants of Self-Help Groups lending to Enterprises in India: A Predictive Assessment using Supervised Machine Learning Algorithms. *SSRN Electronic Journal* **23**. . [Crossref]
432. Gerard Domènech-Arumí. 2021. Neighborhoods, Perceived Inequality, and Preferences for Redistribution: Evidence from Barcelona. *SSRN Electronic Journal* **95**. . [Crossref]
433. Paul Hünermund, Jermain Kaminski, Carla Schmitt. 2021. Causal Machine Learning and Business Decision Making. *SSRN Electronic Journal* **33**. . [Crossref]
434. Francesca Micocci, Armando Rungi. 2021. Predicting Exporters with Machine Learning. *SSRN Electronic Journal* **83**. . [Crossref]
435. Mario Hendriock. 2021. Forecasting Earnings with Predicted, Conditional Probability Density Functions. *SSRN Electronic Journal* **11**. . [Crossref]
436. Simon Blöthner, Mario Larch. 2021. Economic Determinants of Regional Trade Agreements Revisited Using Machine Learning. *SSRN Electronic Journal* **7**. . [Crossref]
437. James T. E. Chapman, Ajit Desai. 2021. Macroeconomic Predictions using Payments Data and Machine Learning. *SSRN Electronic Journal* **29**. . [Crossref]
438. Martin Thomas Hibbeln, Raphael M. Kopp, Noah Urban. 2021. Credit Risk Modeling in the Age of Machine Learning. *SSRN Electronic Journal* **44**. . [Crossref]
439. Mobina Shafaati, Don M. Chance, Robert E. Brooks. 2021. The Cross-Section of Individual Equity Option Returns. *SSRN Electronic Journal* **61**. . [Crossref]
440. Grazia Cecere, Nicoletta Corrocher, Clara Jean. 2021. Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. *SSRN Electronic Journal* **13**. . [Crossref]
441. Emanuel Kohlscheen. 2021. What does machine learning say about the drivers of inflation?. *SSRN Electronic Journal* **45**. . [Crossref]
442. Andrew Tsang. 2021. Uncovering Heterogeneous Regional Impacts of Chinese Monetary Policy. *SSRN Electronic Journal* **51**. . [Crossref]
443. Andrew Tsang. 2021. Uncovering Heterogeneous Regional Impacts of Chinese Monetary Policy. *SSRN Electronic Journal* **51**. . [Crossref]
444. Gabriel Loumeau, Christian Stettler. 2021. Fiscal Autonomy and Self-Determination. *SSRN Electronic Journal* **83**. . [Crossref]
445. Archana Das, Saswat Kumar Das, Naveen Rathee. 2021. Roles of Big Data, Data Science, Artificial Intelligence in Entrepreneurships. *SSRN Electronic Journal* **55**. . [Crossref]
446. Man Cho, Seongwuk Moon, Inbok Rhee. 2021. A Sectoral Approach in Assessing the Data Economy Ecosystem: Focusing on the Finance, Real Estate, and Medical Service Sectors in Korea. *SSRN Electronic Journal* **122**. . [Crossref]
447. Benjamin Ghansah, Clement Nyamekye, Seth Owusu, Emmanuel Agyapong. 2021. Mapping flood prone and Hazards Areas in rural landscape using landsat images and random forest classification: Case study of Nasia watershed in Ghana. *Cogent Engineering* **8**:1. . [Crossref]
448. Gharad Bryan, Dean S. Karlan, Adam Osman. 2021. Big Loans to Small Businesses: Predicting Winners and Losers in an Entrepreneurial Lending Experiment. *SSRN Electronic Journal* **51**. . [Crossref]
449. YongKi Hong. 2021. Related-Party Trades in Vertical Integration. *SSRN Electronic Journal* **64**. . [Crossref]

450. Sandra Achten, Christian Lessmann. 2020. Spatial inequality, geography and economic activity. *World Development* **136**, 105114. [Crossref]
451. Giorgio Tripodi, Francesca Chiaromonte, Fabrizio Lillo. 2020. Knowledge and social relatedness shape research portfolio diversification. *Scientific Reports* **10**:1. . [Crossref]
452. Julian M. Saad, James O. Prochaska. 2020. A philosophy of health: life as reality, health as a universal value. *Palgrave Communications* **6**:1. . [Crossref]
453. Arjun Remadevi Somanathan, Suprabha Kudigrama Rama. 2020. A Bibliometric Review of Stock Market Prediction: Perspective of Emerging Markets. *Applied Computer Systems* **25**:2, 77-86. [Crossref]
454. Manuel J. García Rodríguez, Vicente Rodríguez Montequín, Francisco Ortega Fernández, Joaquín M. Villanueva Balsera. 2020. Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain. *Complexity* **2020**, 1-20. [Crossref]
455. Carolina Castagnetti, Luisa Rosti, Marina Töpfer. The Age Pay Gap between Young and Older Employees in Italy: Perceived or Real Discrimination against the Young? 195-221. [Crossref]
456. Michael T. Kiley. 2020. Financial Conditions and Economic Activity: Insights from Machine Learning. *Finance and Economics Discussion Series* **2020**:095, 1-40. [Crossref]
457. Avinash Kaur, Pooja Gupta, Parminder Singh, Manpreet Singh. 2020. Data Placement Oriented Scheduling Algorithm for Scheduling Scientific Workflow in Cloud: A Budget-Aware Approach. *Recent Advances in Computer Science and Communications* **13**:5, 871-883. [Crossref]
458. Kristian Jönsson. 2020. Machine Learning and Nowcasts of Swedish GDP. *Journal of Business Cycle Research* **16**:2, 123-134. [Crossref]
459. Fiona Burlig, Christopher Knittel, David Rapson, Mar Reguant, Catherine Wolfram. 2020. Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists* **7**:6, 1181-1217. [Crossref]
460. YA-HUI LIN, SHAO-WEN CHIU, YING-CHE LIN, CHIEN-CHUNG LIN, LUNG-KWANG PAN. 2020. INVERSE PROBLEM ALGORITHM APPLICATION TO SEMI-QUANTITATIVE ANALYSIS OF 272 PATIENTS WITH ISCHEMIC STROKE SYMPTOMS: CAROTID STENOSIS RISK ASSESSMENT FOR FIVE RISK FACTORS. *Journal of Mechanics in Medicine and Biology* **20**:09, 2040021. [Crossref]
461. Markku Maula, Wouter Stam. 2020. Enhancing Rigor in Quantitative Entrepreneurship Research. *Entrepreneurship Theory and Practice* **44**:6, 1059-1090. [Crossref]
462. Omar Isaac Asensio, Ximin Mi, Sameer Dharur. 2020. Using Machine Learning Techniques to Aid Environmental Policy Analysis. *Case Studies in the Environment* **4**:1. . [Crossref]
463. Winky K.O. Ho, Bo-Sin Tang, Siu Wai Wong. 2020. Predicting property prices with machine learning algorithms. *Journal of Property Research* **147**, 1-23. [Crossref]
464. Jens Frankenreiter, Michael A. Livermore. 2020. Computational Methods in Legal Analysis. *Annual Review of Law and Social Science* **16**:1, 39-57. [Crossref]
465. Songul Cinaroglu. 2020. Modelling unbalanced catastrophic health expenditure data by using machine-learning methods. *Intelligent Systems in Accounting, Finance and Management* **27**:4, 168-181. [Crossref]
466. Vivianna Fang He, Phanish Puranam, Yash Raj Shrestha, Georg Krogh. 2020. Resolving governance disputes in communities: A study of software license decisions. *Strategic Management Journal* **41**:10, 1837-1868. [Crossref]

467. Jens Prüfer, Patricia Prüfer. 2020. Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands. *Small Business Economics* **55**:3, 651-672. [Crossref]

468. Jermain C. Kaminski, Christian Hopp. 2020. Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Business Economics* **55**:3, 627-649. [Crossref]

469. Eugenio Levi, Fabrizio Patriarca. 2020. An exploratory study of populism: the municipality-level predictors of electoral outcomes in Italy. *Economia Politica* **37**:3, 833-875. [Crossref]

470. Wanling Qiu, Simon Rudkin, Paweł Dłotko. 2020. Refining understanding of corporate failure through a topological data analysis mapping of Altman's Z-score model. *Expert Systems with Applications* **156**, 113475. [Crossref]

471. Elyria Kemp, Gregory N. Price, Nicole R. Fuller, Edna Faye Kemp. 2020. African Americans and COVID-19: Beliefs, behaviors and vulnerability to infection. *International Journal of Healthcare Management* **13**:4, 303-311. [Crossref]

472. Ali Hayek, Zaher Khraibani, Dana Radwan, Nabil Tabaja, Samir Abbad Andaloussi, Joumana Toufaily, Evelyne Garnier-Zarli, Tayssir Hamieh. 2020. Analysis of the extreme and records values for temperature and precipitation in Lebanon. *Communications in Statistics: Case Studies, Data Analysis and Applications* **6**:4, 411-428. [Crossref]

473. Heidi Webber, Gunnar Lischeid, Michael Sommer, Robert Finger, Claas Nendel, Thomas Gaiser, Frank Ewert. 2020. No perfect storm for crop yield failure in Germany. *Environmental Research Letters* **15**:10, 104012. [Crossref]

474. SeyedSoroosh Azizi, Kiana Yektansani. 2020. Artificial Intelligence and Predicting Illegal Immigration to the USA. *International Migration* **58**:5, 183-193. [Crossref]

475. Michael F. Lohrer, Yang Liu, Darrin M. Hanna, Kang-Hsin Wang, Fu-Tong Liu, Ted A. Laurence, Gang-yu Liu. 2020. Determination of the Maturation Status of Dendritic Cells by Applying Pattern Recognition to High-Resolution Images. *The Journal of Physical Chemistry B* **124**:39, 8540-8548. [Crossref]

476. Berkay Eren, Mehmet Ali Guvenc, Selcuk Mistikoglu. 2020. Artificial Intelligence Applications for Friction Stir Welding: A Review. *Metals and Materials International* **54**. . [Crossref]

477. Joshua B. Smith, Matthew Shew, Omar A. Karadaghy, Rohit Nallani, Kevin J. Sykes, Gregory N. Gan, Jason A. Brant, Andrés M. Bur. 2020. Predicting salvage laryngectomy in patients treated with primary nonsurgical therapy for laryngeal squamous cell carcinoma using machine learning. *Head & Neck* **42**:9, 2330-2339. [Crossref]

478. Liye Ma, Baohong Sun. 2020. Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing* **37**:3, 481-504. [Crossref]

479. C. Kokkotis, S. Moustakidis, E. Papageorgiou, G. Giakas, D.E. Tsaopoulos. 2020. Machine learning in knee osteoarthritis: A review. *Osteoarthritis and Cartilage Open* **2**:3, 100069. [Crossref]

480. Huamao Wang, Yumei Yao, Said Salhi. 2020. Tension in big data using machine learning: Analysis and applications. *Technological Forecasting and Social Change* **158**, 120175. [Crossref]

481. Shenhao Wang, Qingyi Wang, Jinhua Zhao. 2020. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C: Emerging Technologies* **118**, 102701. [Crossref]

482. Tobias Götze, Marc Gürtler, Eileen Witowski. 2020. Improving CAT bond pricing models via machine learning. *Journal of Asset Management* **21**:5, 428-446. [Crossref]

483. Louis R. Nemzer, Florence Neymotin. 2020. How words matter: machine learning & movie success. *Applied Economics Letters* **27**:15, 1272-1276. [Crossref]

484. Mitsuru Igami. 2020. Artificial intelligence as structural estimation: Deep Blue, Bonanza, and AlphaGo. *The Econometrics Journal* **23**:3, S1-S24. [Crossref]
485. Jeppe Druedahl, Anders Munk-Nielsen. 2020. Higher-order income dynamics with linked regression trees. *The Econometrics Journal* **23**:3, S25-S58. [Crossref]
486. Gary Smith. 2020. Data mining fool's gold. *Journal of Information Technology* **35**:3, 182-194. [Crossref]
487. Ryan Engstrom, David Newhouse, Vidhya Soundararajan. 2020. Estimating small-area population density in Sri Lanka using surveys and Geo-spatial data. *PLOS ONE* **15**:8, e0237063. [Crossref]
488. Louis R. Nemzer, Florence Neymotin. 2020. Concierge care and patient reviews. *Health Economics* **29**:8, 913-922. [Crossref]
489. Matthew A. Cole, Robert J R Elliott, Bowen Liu. 2020. The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. *Environmental and Resource Economics* **76**:4, 553-580. [Crossref]
490. Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek, Omer Berat Sezer. 2020. Deep learning for financial applications : A survey. *Applied Soft Computing* **93**, 106384. [Crossref]
491. Arsen Sheverdin, Francesco Monticone, Constantinos Valagiannopoulos. 2020. Photonic Inverse Design with Neural Networks: The Case of Invisibility in the Visible. *Physical Review Applied* **14**:2. . [Crossref]
492. Michelle Sapitang, Wanie M. Ridwan, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafie. 2020. Machine Learning Application in Reservoir Water Level Forecasting for Sustainable Hydropower Generation Strategy. *Sustainability* **12**:15, 6121. [Crossref]
493. Anita M. Cassard, Brian W. Sloboda. AI and AR 216-231. [Crossref]
494. Rohit Sharma, Sachin S. Kamble, Angappa Gunasekaran, Vikas Kumar, Anil Kumar. 2020. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers & Operations Research* **119**, 104926. [Crossref]
495. Xilei Zhao, Xiang Yan, Alan Yu, Pascal Van Hentenryck. 2020. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society* **20**, 22-35. [Crossref]
496. Crocker H Liu, Adam D Nowak, Patrick S Smith. 2020. Asymmetric or Incomplete Information about Asset Values?. *The Review of Financial Studies* **33**:7, 2898-2936. [Crossref]
497. Christopher Munch, Mehdi B. Tahoori. Defect Characterization of Spintronic-based Neuromorphic Circuits 1-4. [Crossref]
498. Jonathan A. Cook, Saad Siddiqui. 2020. Random forests and selected samples. *Bulletin of Economic Research* **72**:3, 272-287. [Crossref]
499. Alette Tammenga. 2020. The application of Artificial Intelligence in banks in the context of the three lines of defence model. *Maandblad Voor Accountancy en Bedrijfseconomie* **94**:5/6, 219-230. [Crossref]
500. Hugo Storm, Kathy Baylis, Thomas Heckelei. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* **47**:3, 849-892. [Crossref]
501. Jian-qiang Guo, Shu-hen Chiang, Min Liu, Chi-Chun Yang, Kai-yi Guo. 2020. CAN MACHINE LEARNING ALGORITHMS ASSOCIATED WITH TEXT MINING FROM INTERNET DATA IMPROVE HOUSING PRICE PREDICTION PERFORMANCE?. *International Journal of Strategic Property Management* **24**:5, 300-312. [Crossref]
502. Philip Hans Franses, Thomas Wiemann. 2020. Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping. *Computational Economics* **56**:1, 59-75. [Crossref]
503. Gary Smith. 2020. The paradox of big data. *SN Applied Sciences* **2**:6. . [Crossref]

504. Gonul Colak, Mengchuan Fu, Iftekhar Hasan. 2020. Why are some Chinese firms failing in the US capital markets? A machine learning approach. *Pacific-Basin Finance Journal* **61**, 101331. [Crossref]

505. Robert Northcott. 2020. Big data and prediction: Four case studies. *Studies in History and Philosophy of Science Part A* **81**, 96-104. [Crossref]

506. Andrew B. Martinez. 2020. Forecast Accuracy Matters for Hurricane Damage. *Econometrics* **8**:2, 18. [Crossref]

507. Zhao Tong, Xiaomei Deng, Hongjian Chen, Jing Mei, Hong Liu. 2020. QL-HEFT: a novel machine learning scheduling scheme base on cloud computing environment. *Neural Computing and Applications* **32**:10, 5553-5570. [Crossref]

508. Nanae Kaneko, Yu Fujimoto, Satoshi Kabe, Motonari Hayashida, Yasuhiro Hayashi. 2020. Sparse modeling approach for identifying the dominant factors affecting situation-dependent hourly electricity demand. *Applied Energy* **265**, 114752. [Crossref]

509. Christophe Croux, Julapa Jagtiani, Tarunsai Korivi, Milos Vulanovic. 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization* **173**, 270-296. [Crossref]

510. Kevin A. Bryan, Yasin Ozcan, Bhaven Sampat. 2020. In-text patent citations: A user's guide. *Research Policy* **49**:4, 103946. [Crossref]

511. Omer Berat Sezer, Mehmet Ugur Gudelek, Ahmet Murat Ozbayoglu. 2020. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing* **90**, 106181. [Crossref]

512. Marcos M. López de Prado. Machine Learning for Asset Managers **44**, . [Crossref]

513. Georges Bresson. Comments on "An Econometrician's Perspective on Big Data" by Cheng Hsiao 431-443. [Crossref]

514. Matthew J. Salganik, Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gao, Andrew Halpern-Manners, Sonia P. Hashim, Sonia Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, David Jurgens, Patrick Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang, Katy M. Pinto, Ethan Porter, Kristin E. Porter, Crystal Qian, Tamkinat Rauf, Anahit Sargsyan, Thomas Schaffner, Landon Schnabel, Bryan Schonfeld, Ben Sender, Jonathan D. Tang, Emma Tsurkov, Austin van Loon, Onur Varol, Xiafei Wang, Zhi Wang, Julia Wang, Flora Wang, Samantha Weissman, Kirstie Whitaker, Maria K. Wolters, Wei Lee Woon, James Wu, Catherine Wu, Kengran Yang, Jingwen Yin, Bingyu Zhao, Chenyun Zhu, Jeanne Brooks-Gunn, Barbara E. Engelhardt, Moritz Hardt, Dean Knox, Karen Levy, Arvind Narayanan, Brandon M. Stewart, Duncan J. Watts, Sara McLanahan. 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences* **117**:15, 8398-8403. [Crossref]

515. Filiz Garip. 2020. What failure to predict life outcomes can teach us. *Proceedings of the National Academy of Sciences* **117**:15, 8234-8235. [Crossref]

516. José-Luis Alfaro-Navarro, Emilio L. Cano, Esteban Alfaro-Cortés, Noelia García, Matías Gámez, Beatriz Larraz. 2020. A Fully Automated Adjustment of Ensemble Methods in Machine Learning for Modeling Complex Real Estate Systems. *Complexity* **2020**, 1-12. [Crossref]
517. Raymond Duch, Denise Laroze, Thomas Robinson, Pablo Beramendi. 2020. Multi-modes for Detecting Experimental Measurement Error. *Political Analysis* **28**:2, 263-283. [Crossref]
518. Agostino Valier. 2020. Who performs better? AVMs vs hedonic models. *Journal of Property Investment & Finance* **38**:3, 213-225. [Crossref]
519. Thorsten Sellhorn. 2020. Machine Learning und empirische Rechnungslegungsforschung: Einige Erkenntnisse und offene Fragen. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* **72**:1, 49-69. [Crossref]
520. Natalya V Kontsevaya. 2020. On the specificity of modeling market dynamics. *Journal of Physics: Conference Series* **1479**:1, 012032. [Crossref]
521. Achim Ahrens, Christian B. Hansen, Mark E. Schaffer. 2020. lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal: Promoting communications on statistics and Stata* **20**:1, 176-235. [Crossref]
522. Mathias Bärtl, Simone Krummaker. 2020. Prediction of Claims in Export Credit Finance: A Comparison of Four Machine Learning Techniques. *Risks* **8**:1, 22. [Crossref]
523. David Easley, Eleonora Patacchini, Christopher Rojas. 2020. Multidimensional diffusion processes in dynamic online networks. *PLOS ONE* **15**:2, e0228421. [Crossref]
524. Giovanni Cicceri, Giuseppe Inserra, Michele Limosani. 2020. A Machine Learning Approach to Forecast Economic Recessions—An Italian Case Study. *Mathematics* **8**:2, 241. [Crossref]
525. Jianhua Zhang, Mohammad Shahidul Islam. 2020. The Heterogeneous Impacts of R&D on Innovation in Services Sector: A Firm-Level Study of Developing ASEAN. *Sustainability* **12**:4, 1643. [Crossref]
526. Christophe Hurlin, Christophe Pérignon. 2020. Machine learning et nouvelles sources de données pour le scoring de crédit. *Revue d'économie financière* **N° 135**:3, 21-50. [Crossref]
527. Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, Ahmed J. Aljaaf. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science 3-21. [Crossref]
528. David Watson. The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence 45-65. [Crossref]
529. Petrus H. Potgieter. Machine Learning and Forecasting: A Review 193-207. [Crossref]
530. Matthew F. Dixon, Igor Halperin, Paul Bilokon. Feedforward Neural Networks 111-166. [Crossref]
531. Mitja Kovač. Introduction to the Autonomous Artificial Intelligence Systems 47-63. [Crossref]
532. Agostino Valier. The Cross Validation in Automated Valuation Models: A Proposal for Use 585-596. [Crossref]
533. Chuan Liao, Jiangxiao Qiu, Bin Chen, Deliang Chen, Bojie Fu, Matei Georgescu, Chunyang He, G. Darrel Jenerette, Xia Li, Xiaoyan Li, Xin Li, Bading Qiuying, Peijun Shi, Jianguo Wu. 2020. Advancing landscape sustainability science: theoretical foundation and synergies with innovations in methodology, design, and application. *Landscape Ecology* **35**:1, 1-9. [Crossref]
534. Muhammad Bilal, Lukumon O. Oyedele. 2020. Guidelines for applied machine learning in construction industry—A case of profit margins estimation. *Advanced Engineering Informatics* **43**, 101013. [Crossref]

535. Marcus H. Böhme, André Gröger, Tobias Stöhr. 2020. Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics* **142**, 102347. [Crossref]
536. Ludovic Rheault, Christopher Cochrane. 2020. Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis* **28**:1, 112-133. [Crossref]
537. Steven Buck, Maximilian Auffhammer, Hilary Soldati, David Sunding. 2020. Forecasting Residential Water Consumption in California: Rethinking Model Selection. *Water Resources Research* **56**:1. . [Crossref]
538. Kristian Bondo Hansen. 2020. The virtue of simplicity: On machine learning models in algorithmic trading. *Big Data & Society* **7**:1, 205395172092655. [Crossref]
539. Jorge Iván Pérez Rave, Favián González Echavarría, Juan Carlos Correa Morales. 2020. Modeling of apartment prices in a Colombian context from a machine learning approach with stable-important attributes. *DYNA* **87**:212, 63-72. [Crossref]
540. Lily Shen, Stephen L. Ross. 2020. Information Value of Property Description: A Machine Learning Approach. *SSRN Electronic Journal* . [Crossref]
541. Elena-Ivona Dumitrescu, Sullivan Hué, Christophe Hurlin, sessi tokpavi. 2020. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds. *SSRN Electronic Journal* . [Crossref]
542. Falco Bargagli Stoffi, Massimo Riccaboni, Armando Rungi. 2020. Machine Learning for Zombie Hunting. Firms' Failures and Financial Constraints. *SSRN Electronic Journal* . [Crossref]
543. Ranik Raaen Wahlstrøm, Florentina Paraschiv, Michael Schürle. 2020. A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Models. *SSRN Electronic Journal* . [Crossref]
544. Luca Barbaglia, Sebastiano Manzan, Elisa Tosetti. 2020. Forecasting Loan Default in Europe with Machine Learning. *SSRN Electronic Journal* **36**. . [Crossref]
545. Christopher Hennessy, Charles A.E. Goodhart. 2020. Goodhart's Law and Machine Learning. *SSRN Electronic Journal* **23**. . [Crossref]
546. Kevin Bauer, Nicolas Pfeuffer, Benjamin Abdel-Karim, Oliver Hinz, Michael Kosfeld. 2020. The Terminator of Social Welfare? The Economic Consequences of Algorithmic Discrimination. *SSRN Electronic Journal* . [Crossref]
547. Yucheng Yang, Zhong Zheng, Weinan E. 2020. Interpretable Neural Networks for Panel Data Analysis in Economics. *SSRN Electronic Journal* . [Crossref]
548. Elliott Ash, Sergio Galletta, Tommaso Giommoni. 2020. A Machine Learning Approach to Analyzing Corruption in Local Public Finances. *SSRN Electronic Journal* **156**. . [Crossref]
549. Domonkos F. Vamossy. 2020. Investor Emotions and Earnings Announcements. *SSRN Electronic Journal* **16**. . [Crossref]
550. Yulin Liu, Luyao Zhang. 2020. Cryptocurrency Valuation and Machine Learning. *SSRN Electronic Journal* **6**. . [Crossref]
551. Paul G. Geertsema, Helen Lu. 2020. Relative Valuation with Machine Learning. *SSRN Electronic Journal* **58**. . [Crossref]
552. Xi Chen, Yang Ha Cho, Yiwei Dou, Baruch Itamar Lev. 2020. Fundamental Analysis of XBRL Data: A Machine Learning Approach. *SSRN Electronic Journal* **58**. . [Crossref]
553. Juan Díaz, Erwin Hansen, Gabriel Cabrera. 2020. Predictive Regressions for Aggregate Stock Market Volatility with Machine Learning. *SSRN Electronic Journal* **9**. . [Crossref]

554. Huy Duc Dang, Au Hai Thi Dam, Thuyen Thi Pham, Tra My Thi Nguyen. 2019. Determinants of credit demand of farmers in Lam Dong, Vietnam. *Agricultural Finance Review* **80**:2, 255-274. [Crossref]
555. Stefan P. Penczynski. 2019. Using machine learning for communication classification. *Experimental Economics* **22**:4, 1002-1029. [Crossref]
556. Yang Zhou, Shuaishuai Zhang, Libo Wu, Yingjie Tian. 2019. Predicting sectoral electricity consumption based on complex network analysis. *Applied Energy* **255**, 113790. [Crossref]
557. E.E. Kangal, M. Salti, O. Aydogdu. 2019. Machine learning algorithm in a caloric view point of cosmology. *Physics of the Dark Universe* **26**, 100369. [Crossref]
558. Qi Wu, Xing Yan. 2019. Capturing deep tail risk via sequential learning of quantile dynamics. *Journal of Economic Dynamics and Control* **109**, 103771. [Crossref]
559. Ramkumar Harikrishnakumar, Alok Dand, Saideep Nannapaneni, Krishna Krishnan. Supervised Machine Learning Approach for Effective Supplier Classification 240-245. [Crossref]
560. Elham Buxton, Kenneth Kriz, Matthew Cremeens, Kim Jay. An Auto Regressive Deep Learning Model for Sales Tax Forecasting from Multiple Short Time Series 1359-1364. [Crossref]
561. Sen Yang, Dave Towey, Zhi Quan Zhou, T.Y. Chen. Preparing Software Quality Assurance Professionals: Metamorphic Exploration for Machine Learning 1-7. [Crossref]
562. Manuel Huber, Christoph Kurz, Reiner Leidl. 2019. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Medical Informatics and Decision Making* **19**:1. . [Crossref]
563. Manav Raj, Robert Seamans. 2019. Primer on artificial intelligence and robotics. *Journal of Organization Design* **8**:1. . [Crossref]
564. Shengying Zhai, Qihui Chen, Wenxin Wang. 2019. What Drives Green Fodder Supply in China?—A Nerlovian Analysis with LASSO Variable Selection. *Sustainability* **11**:23, 6692. [Crossref]
565. Erik Brynjolfsson, Xiang Hui, Meng Liu. 2019. Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. *Management Science* **65**:12, 5449-5460. [Crossref]
566. Ron Adner, Phanish Puranam, Feng Zhu. 2019. What Is Different About Digital Strategy? From Quantitative to Qualitative Change. *Strategy Science* **4**:4, 253-261. [Crossref]
567. Jorge Mejia, Shawn Mankad, Anandasivam Gopal. 2019. A for Effort? Using the Crowd to Identify Moral Hazard in New York City Restaurant Hygiene Inspections. *Information Systems Research* **30**:4, 1363-1386. [Crossref]
568. Alessandro Roncaglia. The Age of Fragmentation **2**, . [Crossref]
569. . Technological Change 136-157. [Crossref]
570. Prithwiraj Choudhury, Dan Wang, Natalie A. Carlson, Tarun Khanna. 2019. Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal* **40**:11, 1705-1732. [Crossref]
571. Vasilios Plakandaras, Periklis Gogas, Theophilos Papadimitriou, Rangan Gupta. 2019. A re-evaluation of the term spread as a leading indicator. *International Review of Economics & Finance* **64**, 476-492. [Crossref]
572. David McKenzie, Dario Sansone. 2019. Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria. *Journal of Development Economics* **141**, 102369. [Crossref]
573. Arno Parolini, Wei Wu Tan, Aron Shlonsky. 2019. Decision-based models of the implementation of interventions in systems of healthcare: Implementation outcomes and intervention effectiveness in complex service environments. *PLOS ONE* **14**:10, e0223129. [Crossref]

574. Eszter Czibor, David Jimenez-Gomez, John A. List. 2019. The Dozen Things Experimental Economists Should Do (More of). *Southern Economic Journal* **86**:2, 371-432. [Crossref]
575. Nicolas Huck. 2019. Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research* **278**:1, 330-342. [Crossref]
576. Yijiang Liu, Yinghong Wan, Xiao Su. 2019. Identifying individual expectations in service recovery through natural language processing and machine learning. *Expert Systems with Applications* **131**, 288-298. [Crossref]
577. Adel Daoud, Rockli Kim, S.V. Subramanian. 2019. Predicting women's height from their socioeconomic status: A machine learning approach. *Social Science & Medicine* **238**, 112486. [Crossref]
578. Joseph R. Cimpian, Jennifer D. Timmer. 2019. Large-Scale Estimates of LGBQ-Heterosexual Disparities in the Presence of Potentially Mischievous Responders: A Preregistered Replication and Comparison of Methods. *AERA Open* **5**:4, 233285841988889. [Crossref]
579. Kamran Soomro, Muhammad Nasir Mumtaz Bhutta, Zaheer Khan, Muhammad A. Tahir. 2019. Smart city big data analytics: An advanced review. *WIREs Data Mining and Knowledge Discovery* **9**:5. . [Crossref]
580. David Watson. 2019. The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds and Machines* **29**:3, 417-440. [Crossref]
581. Austin M. Williams. 2019. Understanding the micro-determinants of defensive behaviors against pollution. *Ecological Economics* **163**, 42-51. [Crossref]
582. Ian Hoffman, Evan Mast. 2019. Heterogeneity in the effect of federal spending on local crime: Evidence from causal forests. *Regional Science and Urban Economics* **78**, 103463. [Crossref]
583. Meenakshi Meenakshi, Satpal. A Novel Approach Web Services Based Long Tail Web Services Using Deep Neural Network 1-9. [Crossref]
584. Tien-Tung Nguyen, Jong-Ho Lee, Minh-Tuan Nguyen, Yong-Hwa Kim. 2019. Machine Learning-Based Relay Selection for Secure Transmission in Multi-Hop DF Relay Networks. *Electronics* **8**:9, 949. [Crossref]
585. R. Arjun, K.R. Suprabha. 2019. Forecasting banking sectors in Indian stock markets using machine intelligence. *International Journal of Hybrid Intelligent Systems* **15**:3, 129-142. [Crossref]
586. Susan Athey, Guido W. Imbens. 2019. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics* **11**:1, 685-725. [Crossref]
587. Vincent P. Crawford. 2019. Experiments on Cognition, Communication, Coordination, and Cooperation in Relationships. *Annual Review of Economics* **11**:1, 167-191. [Crossref]
588. Xiaolu Zhou, Weitian Tong, Dongying Li. 2019. Modeling Housing Rent in the Atlanta Metropolitan Area Using Textual Information and Deep Learning. *ISPRS International Journal of Geo-Information* **8**:8, 349. [Crossref]
589. Lei Dong, Carlo Ratti, Siqi Zheng. 2019. Predicting neighborhoods' socioeconomic attributes using restaurant data. *Proceedings of the National Academy of Sciences* **116**:31, 15447-15452. [Crossref]
590. Mario Molina, Filiz Garip. 2019. Machine Learning for Sociology. *Annual Review of Sociology* **45**:1, 27-45. [Crossref]
591. Patrick Doupe, James Faghmous, Sanjay Basu. 2019. Machine Learning for Health Services Researchers. *Value in Health* **22**:7, 808-815. [Crossref]
592. Charles P. Martin-Shields, Wolfgang Stojetz. 2019. Food security and conflict: Empirical challenges and future opportunities for research and policy making on food security and conflict. *World Development* **119**, 150-164. [Crossref]

593. Michael J. Weir, Thomas W. Sproul. 2019. Identifying Drivers of Genetically Modified Seafood Demand: Evidence from a Choice Experiment. *Sustainability* **11**:14, 3934. [Crossref]
594. Anja Lambrecht, Catherine Tucker. 2019. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* **65**:7, 2966-2981. [Crossref]
595. Ilaria Gandin, Claudio Cozza. 2019. Can we predict firms' innovativeness? The identification of innovation performers in an Italian region through a supervised learning approach. *PLOS ONE* **14**:6, e0218175. [Crossref]
596. Veni Arakelian, Petros Dellaportas, Roberto Savona, Marika Vezzoli. 2019. Sovereign risk zones in Europe during and after the debt crisis. *Quantitative Finance* **19**:6, 961-980. [Crossref]
597. Hamed Ghoddusi, Germán G. Creamer, Nima Rafizadeh. 2019. Machine learning in energy economics and finance: A review. *Energy Economics* **81**, 709-727. [Crossref]
598. Nanae Kaneko, Yasuhiro Hayashi, Yu Fujimoto. Toward Data-Driven Identification of Essential Factors Causing Seasonal Change in Daily Electricity Demand Curves 1-6. [Crossref]
599. Henry E. Brady. 2019. The Challenge of Big Data and Data Science. *Annual Review of Political Science* **22**:1, 297-323. [Crossref]
600. Michael Mayer, Steven C. Bourassa, Martin Hoesli, Donato Scognamiglio. 2019. Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research* **12**:1, 134-150. [Crossref]
601. Ajay Agrawal, Joshua S. Gans, Avi Goldfarb. 2019. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives* **33**:2, 31-50. [Abstract] [View PDF article] [PDF with links]
602. Jens Ludwig, Sendhil Mullainathan, Jann Spiess. 2019. Augmenting Pre-Analysis Plans with Machine Learning. *AEA Papers and Proceedings* **109**, 71-76. [Abstract] [View PDF article] [PDF with links]
603. Israt Jahan, Sayeed Z. Sajal, Kendall E. Nygard. Prediction Model Using Recurrent Neural Networks 1-6. [Crossref]
604. Cristobal Young. 2019. The Difference Between Causal Analysis and Predictive Models: Response to "Comment on Young and Holsteen (2017)". *Sociological Methods & Research* **48**:2, 431-447. [Crossref]
605. Gyeongcheol Cho, Jinyeong Yim, Younyoung Choi, Jungmin Ko, Seoung-Hwan Lee. 2019. Review of Machine Learning Algorithms for Diagnosing Mental Illness. *Psychiatry Investigation* **16**:4, 262-269. [Crossref]
606. Morgan R. Frank, David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, Dashun Wang, Hyejin Youn, Iyad Rahwan. 2019. Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences* **116**:14, 6531-6539. [Crossref]
607. Paolo Brunori, Vito Peragine, Laura Serlenga. 2019. Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare* **52**:4, 635-661. [Crossref]
608. Yogi Sugiawan, Shunsuke Managi. 2019. New evidence of energy-growth nexus from inclusive wealth. *Renewable and Sustainable Energy Reviews* **103**, 40-48. [Crossref]
609. Dario Sansone. 2019. Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics* **81**:2, 456-485. [Crossref]
610. Hilal Atasoy, Brad N. Greenwood, Jeffrey Scott McCullough. 2019. The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization. *Annual Review of Public Health* **40**:1, 487-500. [Crossref]
611. Jeanine Miklós-Thal, Catherine Tucker. 2019. Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?. *Management Science* **65**:4, 1552-1561. [Crossref]

612. Colin F. Camerer, Gideon Nave, Alec Smith. 2019. Dynamic Unstructured Bargaining with Private Information: Theory, Experiment, and Outcome Prediction via Machine Learning. *Management Science* **65**:4, 1867-1890. [Crossref]

613. John Gathergood, Neale Mahoney, Neil Stewart, Jörg Weber. 2019. How Do Individuals Repay Their Debt? The Balance-Matching Heuristic. *American Economic Review* **109**:3, 844-875. [Abstract] [View PDF article] [PDF with links]

614. Alois Weigand. 2019. Machine learning in empirical asset pricing. *Financial Markets and Portfolio Management* **33**:1, 93-104. [Crossref]

615. Wen Jiang, Xianjun Xing, Shan Li, Xianwen Zhang, Wenquan Wang. 2019. Synthesis, characterization and machine learning based performance prediction of straw activated carbon. *Journal of Cleaner Production* **212**, 1210-1223. [Crossref]

616. Jinu Lee. 2019. A Neural Network Method for Nonlinear Time Series Analysis. *Journal of Time Series Econometrics* **11**:1. . [Crossref]

617. Georg Rauter, Nicolas Gerig, Roland Sigrist, Robert Riener, Peter Wolf. 2019. When a robot teaches humans: Automated feedback selection accelerates motor learning. *Science Robotics* **4**:27. . [Crossref]

618. Nicholas Apergis, Ioannis Pragidis. 2019. Stock Price Reactions to Wire News from the European Central Bank: Evidence from Changes in the Sentiment Tone and International Market Indexes. *International Advances in Economic Research* **25**:1, 91-112. [Crossref]

619. Shangran Li. 2019. Research on Data Mining Technology Based on Machine Learning Algorithm. *Journal of Physics: Conference Series* **1168**, 032132. [Crossref]

620. Pushpendra Rana, Daniel C Miller. 2019. Machine learning to analyze the social-ecological impacts of natural resource policy: insights from community forest management in the Indian Himalaya. *Environmental Research Letters* **14**:2, 024008. [Crossref]

621. Arun Balakrishna, Tom Gross. Towards Optimum Integration of Human and Car Navigation System 1-8. [Crossref]

622. Hao Li, Haw Yang. 2019. Statistical Learning of Discrete States in Time Series. *The Journal of Physical Chemistry B* **123**:3, 689-701. [Crossref]

623. Yan Liu, Tian Xie. 2019. Machine learning versus econometrics: prediction of box office. *Applied Economics Letters* **26**:2, 124-130. [Crossref]

624. Jorge Iván Pérez-Rave, Juan Carlos Correa-Morales, Favián González-Echavarría. 2019. A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research* **36**:1, 59-96. [Crossref]

625. Peng Yaohao, Pedro Henrique Melo Albuquerque. 2019. Non-Linear Interactions and Exchange Rate Prediction: Empirical Evidence Using Support Vector Regression. *Applied Mathematical Finance* **26**:1, 69-100. [Crossref]

626. Wen Jiang, Xianjun Xing, Xianwen Zhang, Mengxing Mi. 2019. Prediction of combustion activation energy of NaOH/KOH catalyzed straw pyrolytic carbon based on machine learning. *Renewable Energy* **130**, 1216-1225. [Crossref]

627. Halgurd S. Maghdid. 2019. Web News Mining Using New Features: A Comparative Study. *IEEE Access* **7**, 5626-5641. [Crossref]

628. K. John McConnell, Stephan Lindner. 2019. Estimating treatment effects with machine learning. *Health Services Research* **54**:6, 1273. [Crossref]

629. Christopher Münch, Rajendra Bishnoi, Mehdi B. Tahoori. Reliable in-memory neuromorphic computing using spintronics 230-236. [Crossref]

630. Manuel J. García Rodríguez, Vicente Rodríguez Montequín, Francisco Ortega Fernández, Joaquín M. Villanueva Balsera. 2019. Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning. *Complexity* **2019**, 1. [Crossref]
631. Ian Lundberg, Arvind Narayanan, Karen Levy, Matthew J. Salganik. 2019. Privacy, Ethics, and Data Access: A Case Study of the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World* **5**, 237802311881302. [Crossref]
632. Alexander T. Kindel, Vineet Bansal, Kristin D. Catena, Thomas H. Hartshorne, Kate Jaeger, Dawn Koffman, Sara McLanahan, Maya Phillips, Shiva Rouhani, Ryan Vinh, Matthew J. Salganik. 2019. Improving Metadata Infrastructure for Complex Surveys: Insights from the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World* **5**, 237802311881737. [Crossref]
633. Thomas Davidson. 2019. Black-Box Models and Sociological Explanations: Predicting High School Grade Point Average Using Neural Networks. *Socius: Sociological Research for a Dynamic World* **5**, 237802311881770. [Crossref]
634. Daniel E. Rigobon, Eaman Jahani, Yoshihiko Suhara, Khaled AlGhoneim, Abdulaziz Alghunaim, Alex "Sandy" Pentland, Abdullah Almaatouq. 2019. Winning Models for Grade Point Average, Grit, and Layoff in the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World* **5**, 237802311882041. [Crossref]
635. Matthew J. Salganik, Ian Lundberg, Alexander T. Kindel, Sara McLanahan. 2019. Introduction to the Special Collection on the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World* **5**, 237802311987158. [Crossref]
636. Hellen Geremias dos Santos, Carla Ferreira do Nascimento, Rafael Izbicki, Yeda Aparecida de Oliveira Duarte, Alexandre Dias Porto Chiavegatto Filho. 2019. Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil. *Cadernos de Saúde Pública* **35**:7. . [Crossref]
637. Khashayar Khosravi, Gregory Lewis, Vasilis Syrgkanis. 2019. Non-Parametric Inference Adaptive to Intrinsic Dimension. *SSRN Electronic Journal* . [Crossref]
638. Jens Prufer, Patricia Prufer. 2019. Data Science for Entrepreneurship Research: Studying Demand Dynamics for Entrepreneurial Skills in the Netherlands. *SSRN Electronic Journal* . [Crossref]
639. Saqib Aziz, Michael M. Dowling, Helmi Hammami, Anke Piepenbrink. 2019. Machine Learning in Finance: A Topic Modeling Approach. *SSRN Electronic Journal* . [Crossref]
640. Ajay Agrawal, Joshua S. Gans, Avi Goldfarb. 2019. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *SSRN Electronic Journal* . [Crossref]
641. Costanza Naguib. 2019. Estimating the Heterogeneous Impact of the Free Movement of Persons on Relative Wage Mobility. *SSRN Electronic Journal* . [Crossref]
642. Andreas Joseph. 2019. Shapley Regressions: A Framework for Statistical Inference on Machine Learning Models. *SSRN Electronic Journal* . [Crossref]
643. Dashan Huang, Fuwei Jiang, Guoshi Tong, Guofu Zhou. 2019. Scaled PCA: A New Approach to Dimension Reduction. *SSRN Electronic Journal* . [Crossref]
644. Marcos López de Prado. 2019. Ten Applications of Financial Machine Learning. *SSRN Electronic Journal* . [Crossref]
645. Christophre Georges, Javier Pereira. 2019. Market Stability with Machine Learning Agents. *SSRN Electronic Journal* . [Crossref]
646. George G. Judge. 2019. Combining the Information From Econometrics Learning (EL) and Machine Learning (ML). *SSRN Electronic Journal* . [Crossref]
647. Qi Wu, Xing Yan. 2019. Quantile Forecast via Serial Dependence Learning. *SSRN Electronic Journal* **25**. . [Crossref]

648. Nicolas Apfel. 2019. Relaxing the Exclusion Restriction in Shift-Share Instrumental Variable Estimation. *SSRN Electronic Journal* . [Crossref]
649. Aparna Gupta, Abena Owusu. 2019. Identifying the Risk Culture of Banks Using Machine Learning. *SSRN Electronic Journal* . [Crossref]
650. Stefania Albanesi, Domonkos Vamossy. 2019. Predicting Consumer Default: A Deep Learning Approach. *SSRN Electronic Journal* . [Crossref]
651. Simon Schmickler. 2019. Asset Fire Sales or Assets on Fire?. *SSRN Electronic Journal* . [Crossref]
652. Yanhao Wei, Zhenling Jiang. 2019. Estimating Parameters of Structural Models Using Neural Networks. *SSRN Electronic Journal* . [Crossref]
653. Brett R. Gordon, Mitch Lovett, Bowen Luo, James Reeder. 2019. Disentangling the Effects of Ad Tone on Voter Turnout and Candidate Choice in Presidential Elections. *SSRN Electronic Journal* **47**. . [Crossref]
654. Prasanna L. Tantri. 2019. Does Skillful Use of Hard Information by Machines Outperform a Combination of Hard and Soft Information of Loan Officers in Lending Decisions?. *SSRN Electronic Journal* . [Crossref]
655. Vasilios Plakandaras, Periklis Gogas, Theophilos Papadimitriou. 2019. The Effects of Geopolitical Uncertainty in Forecasting Financial Markets: A Machine Learning Approach. *Algorithms* **12**:1, 1. [Crossref]
656. Stephane Helleringer, Chong You, Laurence Fleury, Laetitia Douillot, Insa Diouf, Cheikh Tidiane Ndiaye, Valerie Delaunay, Rene Vidal. 2019. Improving age measurement in low- and middle-income countries through computer vision: A test in Senegal. *Demographic Research* **40**, 219-260. [Crossref]
657. Alexandre Rubesam. 2019. Machine Learning Portfolios with Equal Risk Contributions. *SSRN Electronic Journal* **61**. . [Crossref]
658. Alex Xi He, Daniel le Maire. 2019. Mergers and Managers: Manager-Specific Wage Premiums and Rent Extraction in M&As. *SSRN Electronic Journal* **67**. . [Crossref]
659. Oleg Rytchkov, Xun Zhong. 2019. Macroeconomic Content of Characteristics-Based Asset Pricing Models: A Machine Learning Analysis. *SSRN Electronic Journal* **69**. . [Crossref]
660. Xinni Cai, Fuxiu Jiang, Jun-Koo Kang. 2019. Remote Board Meetings and Board Monitoring Effectiveness: Evidence from China. *SSRN Electronic Journal* **46**. . [Crossref]
661. Jennifer Ifft, Ryan Kuhns, Kevin Patrick. 2018. Can machine learning improve prediction – an application with farm survey data. *International Food and Agribusiness Management Review* **21**:8, 1083-1098. [Crossref]
662. Fritz Schiltz, Paolo Sestito, Tommaso Agasisti, Kristof De Witte. 2018. The added value of more accurate predictions for school rankings. *Economics of Education Review* **67**, 207-215. [Crossref]
663. Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini. 2018. Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization* **156**, 86-102. [Crossref]
664. Marco Pangallo, Michele Loberto. 2018. Home is where the ad is: online interest proxies housing demand. *EPJ Data Science* **7**:1. . [Crossref]
665. Alex Coad, Dominik Janzing, Paul Nightingale. 2018. Tools for causal inference from cross-sectional innovation surveys with continuous or discrete variables: Theory and applications. *Cuadernos de Economía* **37**:75, 779-808. [Crossref]
666. Joseph R. Cimpian, Jennifer D. Timmer, Michelle A. Birkett, Rachel L. Marro, Blair C. Turner, Gregory L. Phillips. 2018. Bias From Potentially Mischievous Responders on Large-Scale Estimates of Lesbian, Gay, Bisexual, or Questioning (LGBQ)–Heterosexual Youth Health Disparities. *American Journal of Public Health* **108**:S4, S258-S265. [Crossref]

667. William P. Erchul, Aaron J. Fischer, Melissa A. Collier-Meek, Bradley S. Bloomfield. 2018. Highlighting the Utility of the Consultation Analysis Record for Consultation Research and Training. *Journal of Educational and Psychological Consultation* **28**:4, 445-459. [Crossref]
668. Simon Willcock, Javier Martínez-López, Danny A.P. Hooftman, Kenneth J. Bagstad, Stefano Balbi, Alessia Marzo, Carlo Prato, Saverio Sciandrello, Giovanni Signorello, Brian Voigt, Ferdinando Villa, James M. Bullock, Ioannis N. Athanasiadis. 2018. Machine learning for ecosystem services. *Ecosystem Services* **33**, 165-174. [Crossref]
669. M. Hino, E. Benami, N. Brooks. 2018. Machine learning for environmental monitoring. *Nature Sustainability* **1**:10, 583-588. [Crossref]
670. Gilles Bastin, Paola Tubaro. 2018. Le moment big data des sciences sociales. *Revue française de sociologie* **Vol. 59**:3, 375-394. [Crossref]
671. Chiara Masci, Geraint Johnes, Tommaso Agasisti. 2018. Student and school performance across countries: A machine learning approach. *European Journal of Operational Research* **269**:3, 1072-1085. [Crossref]
672. Alberto Abadie, Matias D. Cattaneo. 2018. Econometric Methods for Program Evaluation. *Annual Review of Economics* **10**:1, 465-503. [Crossref]
673. Jiaying Kou, Xiaoming Fu, Jiahua Du, Hua Wang, Geordie Z. Zhang. Understanding Housing Market Behaviour from a Microscopic Perspective 1-9. [Crossref]
674. Francesco Lamperti, Andrea Roventini, Amir Sani. 2018. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control* **90**, 366-389. [Crossref]
675. Daniel P. Tabor, Loïc M. Roch, Semion K. Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph H. Montoya, Shyam Dwaraknath, Muratahan Aykol, Carlos Ortiz, Hermann Tribukait, Carlos Amador-Bedolla, Christoph J. Brabec, Benji Maruyama, Kristin A. Persson, Alán Aspuru-Guzik. 2018. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials* **3**:5, 5-20. [Crossref]
676. Andrew Crane-Droesch. 2018. Technology Diffusion, Outcome Variability, and Social Learning: Evidence from a Field Experiment in Kenya. *American Journal of Agricultural Economics* **100**:3, 955-974. [Crossref]
677. Tom Wilson, Huw Brokensha, Francisco Rowe, Ludi Simpson. 2018. Insights from the Evaluation of Past Local Area Population Forecasts. *Population Research and Policy Review* **37**:1, 137-155. [Crossref]
678. Michał Żejmo, Marek Kowal, Józef Korbicz, Roman Monczak. Nuclei Recognition Using Convolutional Neural Network and Hough Transform 316-327. [Crossref]
679. Yong Yoon. Spatial Choice Modeling Using the Support Vector Machine (SVM): Characterization and Prediction 767-778. [Crossref]
680. William Herlands, Edward McFowland III, Andrew Gordon Wilson, Daniel B. Neill. Automated Local Regression Discontinuity Design Discovery 1512-1520. [Crossref]
681. Guanhao Feng, Nick Polson, Yuexi Wang, Jianeng Xu. 2018. Sparse Regularization in Marketing and Economics. *SSRN Electronic Journal* . [Crossref]
682. Jens Prufer, Patricia Prufer. 2018. Data Science for Institutional and Organizational Economics. *SSRN Electronic Journal* . [Crossref]
683. Shinya Sugawara. 2018. What Comprises Effective Formal Elder Care at Home? Estimating Effects for Combinations of Multiple Services. *SSRN Electronic Journal* . [Crossref]
684. Phanish Puranam, Yash Raj Shrestha, Vivianna Fang He, Georg von Krogh. 2018. Algorithmic Induction Through Machine Learning: Opportunities for Management and Organization Research. *SSRN Electronic Journal* . [Crossref]

685. Marcelo Cunha Medeiros, Gabriel Vasconcelos, Alvaro Veiga, Eduardo Zilberman. 2018. Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *SSRN Electronic Journal* . [Crossref]
686. Anthony Niblett. 2018. Regulatory Reform in Ontario: Machine Learning and Regulation. *SSRN Electronic Journal* . [Crossref]
687. Marianne Bertrand, Emir Kamenica. 2018. Coming Apart? Cultural Distances in the United States Over Time. *SSRN Electronic Journal* **117**. . [Crossref]
688. Erik Brynjolfsson, Xiang Hui, Meng Liu. 2018. Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. *SSRN Electronic Journal* . [Crossref]
689. Ronald Richman. 2018. AI in Actuarial Science. *SSRN Electronic Journal* . [Crossref]
690. Benjamin Bluhm. 2018. Time Series Econometrics at Scale: A Practical Guide to Parallel Computing in (Py)Spark. *SSRN Electronic Journal* . [Crossref]
691. Daniele Bianchi, Matthias Büchner, Andrea Tamoni. 2018. Bond Risk Premia with Machine Learning. *SSRN Electronic Journal* . [Crossref]
692. Isil Erel, Lea Henny Stern, Chenhao Tan, Michael S. Weisbach. 2018. Selecting Directors Using Machine Learning. *SSRN Electronic Journal* . [Crossref]
693. Yingying Fan, Jinchi Lv, Jingbo Wang. 2018. DNN: A Two-Scale Distributional Tale of Heterogeneous Treatment Effect Inference. *SSRN Electronic Journal* . [Crossref]
694. Dimitris Korobilis. 2018. Machine Learning Macroeconometrics: A Primer. *SSRN Electronic Journal* **201**. . [Crossref]
695. Prithwiraj Choudhury, Ryan Allen, Michael Endres. 2018. Developing Theory Using Machine Learning Methods. *SSRN Electronic Journal* . [Crossref]
696. Marco Pangallo, Michele Loberto. 2018. Home Is Where the Ad Is: Online Interest Proxies Housing Demand. *SSRN Electronic Journal* . [Crossref]
697. Tingting Liu, Zhongjin Lu, Tao Shu, Fengrong Wei. 2018. Distinct Relatedness and Synergies in Mergers and Acquisitions. *SSRN Electronic Journal* . [Crossref]
698. Andres Liberman, Christopher Neilson, Luis Opazo, Seth D. Zimmerman. 2018. The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets. *SSRN Electronic Journal* . [Crossref]
699. Jeanine Miklós-Thal, Catherine E. Tucker. 2018. Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?. *SSRN Electronic Journal* . [Crossref]
700. Jack Blundell, Erling Risa. 2018. Do Rank-Rank Income Mobility Measures Fully Capture Broader Parental Influence on Child Income?. *SSRN Electronic Journal* . [Crossref]
701. Germán G. Creamer, Hamed Ghoddusi, Nima Rafizadeh. 2018. Machine Learning in Energy Economics and Finance: A Review. *SSRN Electronic Journal* . [Crossref]
702. Julian TszKin Chan, Weifeng Zhong. 2018. Reading China: Predicting Policy Change with Machine Learning. *SSRN Electronic Journal* **130**. . [Crossref]
703. Yannan (Lily) Shen. 2018. Information Value of Property Description: A Machine Learning Approach. *SSRN Electronic Journal* . [Crossref]
704. Chinmay Kakatkar, Volker Bilgram, Johann Füller. 2018. Innovation Analytics: Leveraging Artificial Intelligence in the Innovation Process. *SSRN Electronic Journal* . [Crossref]
705. Khaled Obaid, Kuntara Pukthuanthong. 2018. A Picture Is Worth a Thousand Words: Market Sentiment From Photos. *SSRN Electronic Journal* . [Crossref]
706. Michael Mayer, Steven C. Bourassa, Martin Edward Ralph Hoesli, Donato Flavio Scognamiglio. 2018. Estimation and Updating Methods for Hedonic Valuation. *SSRN Electronic Journal* . [Crossref]

707. Derek Snow. 2018. Predicting Restaurant Facility Closures. *SSRN Electronic Journal* . [Crossref]
708. Andrew Bell, Jennifer Zavaleta Cheek, Frazer Mataya, Patrick Ward. 2018. Do As They Did: Peer Effects Explain Adoption of Conservation Agriculture in Malawi. *Water* **10**:1, 51. [Crossref]
709. Futoshi Narita, Rujun Yin. 2018. In Search of Information:. *IMF Working Papers* **18**:286, 1. [Crossref]
710. Xuezhong He, Shen Lin. 2018. Rational Learning and Trading Behavior in Limit Order Markets. *SSRN Electronic Journal* **104**. . [Crossref]
711. David M. Reeb, Wanli Zhao. 2018. Dissecting Innovation. *SSRN Electronic Journal* **103**. . [Crossref]
712. Oz Shy. 2018. Alternative Methods for Studying Consumer Payment Choice. *SSRN Electronic Journal* **105**. . [Crossref]
713. Etan A Green. 2017. Bayesian Instinct. *SSRN Electronic Journal* **136**. . [Crossref]
714. Francesco Lamperti, Amir Sani. 2017. Agent-Based Model Calibration Using Machine Learning Surrogates. *SSRN Electronic Journal* . [Crossref]
715. Charles Martin-Shields, Wolfgang Stojetz. 2017. Food Security As Peacebuilding: Analyzing the Relationship between Food Security and Conflict Data to Support Empirical Policy Making. *SSRN Electronic Journal* **93**. . [Crossref]
716. Ethan M.J. Lieber. 2017. Targeting with In-Kind Transfers: Evidence from Medicaid Home Care. *SSRN Electronic Journal* . [Crossref]
717. Christos Andreas Makridis. 2017. Does Culture Pay? Compensating Differentials, Job Satisfaction, and Organizational Practices. *SSRN Electronic Journal* **4**. . [Crossref]
718. Chiranjit Chakraborty, Andreas Joseph. 2017. Machine Learning at Central Banks. *SSRN Electronic Journal* . [Crossref]
719. Jonathan Aaron Cook, Saad Siddiqui. 2017. Random Forests and Selected Samples. *SSRN Electronic Journal* . [Crossref]
720. Andrew B. Martinez. 2017. How Quickly Can We Adapt to Change? An Assessment Of Hurricane Damage Mitigation Efforts Using Forecast Uncertainty. *SSRN Electronic Journal* . [Crossref]
721. Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther. 2017. Predictably Unequal? The Effects of Machine Learning on Credit Markets. *SSRN Electronic Journal* . [Crossref]
722. Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, Viola Salvestrini. 2017. Targeting Policy-Compliers with Machine Learning: An Application to a Tax Rebate Programme in Italy. *SSRN Electronic Journal* . [Crossref]
723. Daniel Martin Katz, Michael James Bommarito, Josh Blackman. 2017. Crowdsourcing Accurately and Robustly Predicts Supreme Court Decisions. *SSRN Electronic Journal* . [Crossref]
724. Arno Parolini, Wei Wu Tan, Aron Shlonsky. 2017. A Blueprint for Causal Inference in Implementation Systems. *SSRN Electronic Journal* . [Crossref]
725. Stuart A. Gabriel, Matteo M. Iacoviello. 2016. A Crisis of Missed Opportunities? Foreclosure Costs and Mortgage Modification During the Great Recession. *SSRN Electronic Journal* . [Crossref]
726. Anja Lambrecht, Catherine E. Tucker. 2016. Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *SSRN Electronic Journal* **3**. . [Crossref]
727. Hui Chen, Winston Wei Dou, Leonid Kogan. 2013. Measuring the 'Dark Matter' in Asset Pricing Models. *SSRN Electronic Journal* **71**. . [Crossref]