Machine Learning in Management

Elizabeth Degefe, Duke University, Krishna Savani, Management and Business, The Hong Kong Polytechnic University, and Abhishek Sheetal, The Hong Kong Polytechnic University

https://doi.org/10.1093/acrefore/9780190224851.013.389
Published online: 13 December 2023

Summary

Although artificial intelligence and machine learning have been around for more than 70 years, their use in academia and industry has grown exponentially in the last decade.

Keywords: machine learning, management, text classification, topic modeling, natural language processing, decision trees, neural networks, abduction

Subjects: Business Policy and Strategy, Human Resource Management, Organizational Behavior

Introduction

Machine learning is an analytical method that researchers have been using since the mid-20th century (Samuel, 1959) that "studies how to use computers to simulate human learning activities, and to study self-improvement methods of computers" (Wang et al., 2009, p. 1). Machine learning seeks to imitate human learning by using self-improving algorithms, which are computational processes that use input data to complete a task without being programmed to produce a specific outcome. To do so, algorithms go through a repetitive process in which they identify patterns in the data and, over successive iterations, adjust the model to better complete their desired task (El Naqa & Murphy, 2015). Using such advanced techniques, machine learning has enabled academics to conduct analyses that they would not have been able to perform using conventional analytical techniques.

The scope of this article is to provide a general overview of research using machine learning in management to give interested management researchers a sense of what kinds of data machine-learning models can assess and how commonly used machine-learning models are built. This article will begin with a general review of the types of data that machine learning can assess and provide a general overview on how machine-learning models are built. For further information on the specifics of model building, a table of supplementary readings is provided in Further Reading. The article then proceeds to assess how machine learning has been used in management and ends with common critiques of machine learning and considerations of new methods. A glossary of terms has been provided in the Appendix; terms included will be italicized at first mention in the article.

Overview of Machine-Learning Approaches

Decades of management research have focused on identifying best practices using traditional analytical methods, such as regressions, structural equation models, multilevel models, and so on. Although powerful, these models come with a set of assumptions that, if violated, damage their ability to accurately represent the underlying phenomena (Berry, 1993; Osborne & Waters, 2019). These regression-based models assume that outcome variables are normally distributed, that error variances are homogeneous across all predictor variables, that predictor variables are not highly correlated, and, unless clearly specified, predictors and outcome variables are linearly related (Erceg-Hurn & Mirosevich, 2008). However, in the data typically used in management research, these assumptions are likely to be violated, resulting in unreliable estimates of the relationships between predictors and the outcome (Berry, 1993). Moreover, regression-based models do not automatically consider complex relationships. Thus, should researchers wish to identify interactions or nonlinear effects, they must manually specify them in the model. This need to prespecify all effects is a severe limitation as it restricts the scope of the analysis to the researcher's knowledge. Given the natural limits of human cognition, this reliance likely results in many complex relationships being overlooked. Given these difficulties, regression-based methods are often not well suited for examining complex data with large sample sizes and many intercorrelated and interacting variables.

However, machine learning is the perfect tool for examining such complex data as it does not share these limitations. First, most machine-learning models make no assumptions about the nature of the data, such as normality, homoskedasticity, absence of multicollinearity, and independently and identically distributed errors (Goldstein et al., 2017; Yaworsky et al., 2020). Consequently, these models can process a large number of highly correlated predictors and nonnormally distributed variables while still accurately representing the underlying phenomenon. Second, because machine-learning models make no assumptions about the relationships between variables, they can assess nonlinear effects and interactions between variables automatically (Breiman, 2001; Friedman, 2002; Goodfellow et al., 2016). By considering all possible effects, machine learning also overcomes the constraints of human cognition, in which important relationships may not be assessed and are therefore overlooked. In these ways, machine learning is capable of allowing researchers to examine incredibly complex datasets that conventional statistical analyses would be ill-suited for.

It should be noted that while machine-learning models are a great way to handle complex data, the use of machine-learning methods is not always called for. Table A1

provides a list of the pros and cons to both machine-learning methods and traditional regression-based methods. In essence, while machine learning allows researchers to assess complex data, it does come with a number of practical difficulties (e.g., greater time investments, the necessity of large datasets and high-performance computers) and results in models that are often difficult to interpret. Thus, in situations where data is not complex and all statistical assumptions are satisfied, traditional methods, such regression analyses, are the ideal tool. The practical inconveniences of building a machine-learning model are not worth the effort if a simpler method will suffice.

Type of Input Data and Models

Machine learning has been used in management to analyze both text and numeric data. Both types of data can measure a wide range of phenomena, and most phenomena can be represented using either form of data. This input data can be assessed using two broad classes of machine-learning methods: *supervised* and *unsupervised*. Supervised machine-learning models analyze data with a predetermined outcome variable in order to predict the outcome based on the input data provided; unsupervised machine learning analyzes data to freely build models without a specific dependent variable (Leavitt et al., 2021). Examples of unsupervised machine learning are automatic clustering of observations or language translation.

When analyzing numeric data, management researchers typically use supervised machine-learning methods to predict an outcome. When analyzing qualitative data (primarily text but sometimes also images and videos), management researchers have used both supervised and unsupervised machine-learning algorithms. Specifically, management studies use supervised machine-learning models for text classification; that is, to detect the frequency of known constructs in the text (Kang et al., 2020; Pitigala & Li, 2015). Unsupervised machine learning is often used for topic modeling, which seeks to discover themes within a collection of text by building a model that freely examines how words cluster around topics within the larger corpus of documents (Blei et al., 2003). Next, the general process of building machine-learning models using both numeric and text data will be reviewed, focusing on models that are commonly used in management research (see Table A2 for an overview).

Basic Process of Building a Machine-Learning Model With Numeric Data

Although there are many different machine-learning models, several basic steps are common to most. Before building their model, researchers first often engage in *feature engineering* in which they exclude, clean, and process variables before analysis (Dong & Liu, 2018). For example, researchers can exclude variables that are irrelevant to the model and *one-hot-encode* categorical variables (see Table A1 for a glossary of definitions). Second, researchers must choose which method to use to impute any missing data. They can use traditional multiple imputation (Lee & Simpson, 2014) or use machine learning-based imputation (Tang & Ishwaran, 2017). The latter is usually more efficient, has narrower confidence intervals, and has less imputation bias (Shah et al., 2014).

Researchers next need to decide which algorithm they will use to build their model, a decision whose options will be discussed later in the chapter. Researchers also need to decide the model's *loss* function, which assesses the gap between the true values of the outcome and the predicted values the model estimates. The loss function can be the same as that used in traditional analyses (e.g., mean square error for continuous outcome variables). However, dozens of loss functions have been proposed in the machine-learning literature for different purposes (e.g., dice loss and focal loss for unbalanced data with very few positive cases; Li et al., 2020). Once a loss function is chosen, machine-learning algorithms then work on minimizing loss as much as possible by iteratively adjusting the model's parameters.

However, machine-learning models often overfit the data while trying to minimize the loss function, which means that the model would fail to make accurate predictions when presented with new data. The first step to detecting overfitting is *unseen testing* (Browne, 2000; Cabitza & Zeitoun, 2019). In this testing, a small representative sample of the overall data, the *unseen data*, is hidden away during the model-building process. The model is then built on the remaining data, known as the *seen data*, which holds a large majority of the observations from the original dataset. Once the model is built on the seen data, it is presented with the unseen data. The model's *accuracy* in the unseen data is the key metric that should be reported (with the assumption that accuracy on the seen data is most likely exaggerated due to overfitting and is thus not reliable). If the model has acceptable accuracy in the unseen data, then the model has passed the unseen test. If the gap between the model's accuracy in the seen data and the unseen data is small, then the model has minimized overfitting. In practice, the model's unseen accuracy tends to be lower than the seen accuracy; however, if the proper precautions are taken during model building, it is possible to see higher accuracy in the unseen testing.

Researchers can reduce overfitting by taking preemptive steps during model building. A simple approach is *cross-validation* (Browne, 2000), such as k-fold cross-validation (Fushiki, 2011). A commonly used value of k is 10. Here, the seen data is randomly split into 10 parts. The model is then built on nine parts and used to predict the outcome in the 10th part. The loss value is generated from this 10th part. This loop is repeated until the loss value is generated from each of the 10 parts. The average loss value across the ten parts is the key metric of interest that needs to be minimized. Another common approach is the leave-p-out technique, in which the data is split into a training set on which the model is built, and a validation set, which is used to test the predictions of the model built on the training set; the two sets are reshuffled after every iteration of the model (Browne, 2000). Cross-validation reduces overfitting because it assesses how well the model generalizes to new data during the model-building process. In practice, cross-validation is now a normal option in most common algorithms and the user guide for that specific algorithm will describe how to enable it. Another method to reduce overfitting is using a regularization parameter. When a model is overfitted, it often relies on a large number of predictors (Hawkins, 2004). *Regularization parameters* penalize the model for relying on too many variables, thus reducing overfitting (Ghojogh & Crowley, 2019).

In addition to reducing overfitting, researchers should also be wary of underfitting; that is, their model is less accurate than it can be because of suboptimal parameters. Many machine-learning model parameters are often set to default values that may not be the best set of values for a given analysis. To find a relatively optimal set of parameters, researchers can conduct a *hyperparameter search* to test how well the model performs with many different combinations of parameter values. There are three different hyperparameter search methods—grid search, random search, and Bayesian search (see Yu & Zhu, 2020 for details; also see Table A1). Across all hyperparameter search methods, the set of parameters that generate the lowest loss value in the seen data is deemed the best set of parameters. This final set of parameters are the ones to use to build the final model.

Once the hyperparameter search process is complete, a final model is built on the seen data using the selected set of parameters. Next, a few specific machine-learning models that are commonly used in management are reviewed: *decision trees* and *neural networks*.

Decision Trees

Decision tree-based algorithms are commonly used to build machine-learning models with numeric data. These models build classification trees that identify variables that best split the data into the categories of interest (Kotsiantis, 2013). This basic approach is used in multiple machine-learning algorithms, such as random forest, Bayesian networks, generalized boosted models, and extreme gradient boosting (XGBoost). Random forest models create multiple independent trees and make their final prediction by aggregating the results across all trees (Breiman, 2001). A variation on the random forest algorithm is the boosted tree algorithm, in which the trees are serially dependent—successive trees learn from the errors of previous trees before all trees are aggregated to make the final prediction (Natekin & Knoll, 2013). A more powerful variation of this method is called extreme gradient boosting (XGBoost; Chen & Guestrin, 2016; Cho et al., 2020). XGBoost differs from the boosted tree algorithm as it uses both the lasso and ridge regularization methods (Hastie et al., 2009) to penalize the model for using too many predictor variables, thus reducing the chances of overfitting the data. Additionally, XGBoost utilizes the parallel computational power of graphics cards, allowing for faster model building.

Neural Networks

Neural network models have been responsible for most recent breakthroughs using machine learning in academics and industry in the 21st century (e.g., Gil et al., 2014). Neural network algorithms mimic the structure of neurons in the human brain, in which *perceptrons*, similar to neurons, are connected in a neural network (Rosenblatt, 1958). A perceptron takes weighted inputs from multiple other perceptrons. Each perceptron has an activation function that generates an output, which is passed on to subsequent perceptrons in the chain to which the focal perceptron is connected. In the final neural network model, each input weight on each perceptron is optimized to predict the dependent variable (Schmidhuber, 2015).

A neural network has three sections: the input layer (i.e., the predictor variables), the intermediate layer, and the output layer (which predicts the outcome variable). Deep-learning neural networks contain multiple intermediate layers, which help build more complex models (Schmidhuber, 2015). Deep-learning machine-learning models work through forward and backward propagation to optimally weigh predictors in the multiple hidden layers of perceptrons. This optimizing of weights involves an iterative process in which the algorithm adjusts the weights and tests the accuracy of the model until either the model's error is no longer significantly reduced across successive iterations or until the model has completed a maximum number of iterations (Goodfellow et al., 2016).

There are numerous other variations on neural networks (Islam et al., 2019). For instance, recurrent neural network algorithms are used to analyze inputs that are sequential in nature, such as text and time series data (Zaremba et al., 2015), and convolutional neural networks are used to analyze multidimensional input data, such as images and videos (O'Shea & Nash, 2015).

Evaluation Metrics for Machine-Learning Models Using Numeric Data

Once the model is built, researchers must then test the accuracy of the model in the unseen dataset. To determine the performance of a model in unseen testing, researchers rely on a number of metrics. There are dozens of other metrics that researchers have created but it would be beyond the scope of the current article to cover all of them; thus, only the most prevalent will be presented. To assess the performance of classification models (i.e., in which the outcome variable is binary or categorical), researchers should examine the *confusion matrix* in the unseen data. This is a table that shows the number of observations that the model classified correctly versus incorrectly for each category in the outcome variable. The information from this confusion matrix can be used to then calculate a number of other metrics.

The first performance metric is *accuracy*, which measures the proportion of observations that were accurately classified (i.e., the sum of true positives and true negatives divided by the total number of observations). This metric is often compared to the model's *no information rate*, representing the accuracy if the model merely guessed the most prevalent class for all observations. Note that if an outcome variable is unbalanced with only a small number of observations in a single class, accuracy is unlikely to adequately represent the model's performance. For instance, when examining corporate misconduct, Bao et al. (2020) found that only 3% of the observations committed fraud. A model using such data would have a no information rate of 97%, so the model could have deceptively high accuracy while still misclassifying the majority of fraud cases.

In classification models, numerous metrics other than accuracy are important and should be reported. *Sensitivity*, defined as the ratio of true positives to the number of positive cases in the data, measures a model's ability to identify positive cases. Sensitivity is a particularly important metric when the goal is to pick up positive cases (e.g., identifying cases of fraud). *Specificity*, defined as the ratio of true negatives to the number of negative cases in the data, measures the model's ability to identify negative cases. When an outcome is unbalanced, these metrics can provide a more holistic insight into the model's accuracy, allowing the researcher to see whether the accuracy is primarily driven by the accurate classification of positive or negative cases. With unbalanced classes, other useful metrics are balanced accuracy, which is the average of sensitivity and specificity, and the F1 score, which is the geometric mean of sensitivity and precision. *Precision* is the ratio of true positives to predicted positives (i.e., the sum of true and false positives). The final classification metric of note is the Area Under the Receiver Operating Characteristics Curve, or simply Area-Under-the-Curve (AUC). AUC measures a model's ability to rank order positive and negative cases. It does so by examining the probability that, when given a random positive and negative case, the model would assign a higher probability of being positive to the positive case than to the negative case.

Machine-learning regression models, that is, machine-learning models that predict a continuous outcome are metrics identical to those already commonly used for evaluating linear regressions, such as R^2, which measures the percentage of the variance in the outcome that the model explains (Botchkarev, 2019). As a percentage, it is also a metric equivalent across models regardless of how the data are scaled, making it useful for comparing models.

While there are many formal performance metrics for assessing supervised machine-learning models, assessing the performance of an unsupervised model is generally not possible via simple mathematical constructs (Palacio-Niño & Berzal, 2019). Instead, secondary field testing is an appropriate method to determine the performance of unsupervised models. In this testing, researchers conduct follow-up studies to validate the model's findings using traditional methods.

Basic Process of Building a Machine-Learning Model With Text Data

While machine-learning models that use numeric data in management are most commonly supervised, those that use text data often use both supervised and unsupervised learning to detect and discover topics and themes within texts.

Topic Classifiers

Text is unstructured data. However, text can be converted into structured data using *topic classifiers*, which use machine learning-based linguistic techniques to mark whether or not various topics of interest are present in each unit of text. This data can then be analyzed using traditional methods, such as logistic regression, for hypothesis testing. Such models are called topic classifiers and are used to automatically analyze text and assign observations to predefined classes or groups (Sunagar et al., 2021). Although there are a number of unsupervised methods for text classification, researchers often rely on supervised machine learning for *text classification* (Pitigala & Li, 2015). Researchers tend to use supervised learning for text classification because they use such models to categorize responses into classes that are predefined by researchers; hence, the data is often labeled by researchers and the model assesses a specified outcome variable, making the learning process supervised (Kadhim, 2019; Sebastiani, 2002).

If a researcher wishes to test whether more agreeable employees get paid less (e.g., Judge et al., 2012), they could run a preexisting text classification model that codes personality from text (e.g., Keh & Cheng, 2019) in employees' emails and then use the personality scores from this model to test whether there is a negative correlation between employees' agreeableness scores and their compensation. This reliance on preexisting algorithms is one of the most common approaches to text classification in management. However, if no preexisting personality topic classifier exists, researchers can build one independently.

To do this, they would first have to select a subset of the text and manually code the outcome of interest (e.g., personality traits). After coding the subset of text, researchers need to convert each text observation (e.g., email) into a set of numbers representing the content of the text using tools such as *Word2Vec* (Church, 2017) or *word embeddings* (Liu et al., 2015). These tools essentially act as dictionaries that allow researchers to score the frequency of words in each entry.

This coded, quantified text data is then split into seen and unseen datasets. Researchers can then choose a supervised machine-learning algorithm to predict personality scores from the quantified text. Once the model is built on the seen data, it is tested on unseen data. If the model has adequate accuracy in the unseen data, researchers can then use it to classify all remaining text entries in the remainder of the dataset (after converting the remaining text into a numeric format). Text classification can be done through various methods using algorithms such as decision trees, random forest, support vector machines (SVM), and neural networks (Kang et al., 2020). To evaluate such models, researchers simply use the same metrics that they would with numeric supervised-learning classification models.

Topic Modeling

Whereas text classification requires researchers to predefine topics that they wish to code, *topic modeling* identifies naturally occurring themes within a collection of text (Kang et al., 2020). For example, suppose a researcher wishes to identify the topics employees most commonly discuss when evaluating the organization. In this case, they can use a topic-modeling algorithm that identifies common themes in employees' feedback, along with the words that define each theme; the researcher can then assess the relative frequency of the topics in the text and the content of each topic.

Latent dirichlet allocation (LDA; Blei et al., 2003) is the most common topic-modeling method and is the only form of unsupervised machine learning commonly used in management. Before starting an LDA analysis, researchers must decide the number of topics present within their data. Although there are indexes to help with this decision, there is no clear method to determine the number of topics that should be in a model (Grimmer & Stewart, 2013). Thus, it is ultimately up to the researcher's judgement. One method to guide this decision is for researchers to fit multiple LDA models differing in the number of topics and select the model with the best fit as judged by a variety of indexes such as *held-out likelihood*, *residuals*, *lower bound*, and *semantic coherence* (Roberts et al., 2019; Taddy, 2012; Wallach et al., 2009).

Once the number of topics is determined, the algorithm determines which words define each topic by first randomly assigning words to topics and then assessing the placement of each word one at a time. When examining each word, the algorithm calculates the probability of the word belonging to the given topic and assigns the target word to a new topic based on this probability, repeating the process until the improvement in probability is minimal or until a set number of iterations has been run. Once the model is built, researchers can assess the content of each topic by examining the words with the highest probability of occurring within each topic. The performance of the model is based on the subjective judgment of researchers on how comprehensible the topics are.

Other Machine-Learning Models

There are a number of other algorithms used within management. For instance, the SVM learning algorithm finds the best hyperplane that partitions observations belonging to different levels of the outcome variable (Vapnik & Lerner, 1963). This hyperplane could be linear or nonlinear. SVM has been used in management to predict firms' deceptiveness (Deng et al., 2021) or to detect public sentiment toward the firm (Yiu et al., 2022).

The k-nearest-neighbors (kNN) algorithm follows the "birds of a feather" approach. It predicts the outcome variable for each observation based on the outcome value of most of its "neighbors," that is, based on other similar observations (Fix & Hodges, 1989). Management researchers have also used the kNN machine-learning algorithm to detect customers' emotional responses based on their images (Pantano et al., 2021). Although useful, kNN and SVM algorithms are not as frequently used in management as decision trees and neural networks.

A commonly used method—lasso regression—straddles the boundary between regression and machine learning. The lasso model aims to identify relevant and irrelevant predictors (Tibshirani, 1996). The relevant predictors identified by the lasso are then entered into a regular regression. Lasso uses iterative learning techniques to determine the optimal value of a key parameter. Lasso can be a default method if its assumptions of sparsity and linearity hold in the dataset; that is, if only a few predictor variables are relevant, and the predictors and outcome variables are related linearly. If either assumption does not hold, lasso is not optimal for analyzing the given data. However, to the extent there are complex relationships in the dataset, properly tuned decision trees or neural networks are likely to have higher accuracy than lasso regressions.

Current State of Machine Learning in the Management Literature

Machine learning is a powerful and versatile method. Although management researchers have only recently begun to take advantage of it, there are examples of how machine learning has allowed management scholars to analyze complex data. For this review, the Web of Science database was accessed and searched for the keywords *machine learning* or *artificial intelligence*, with search results restricted to management journals. Only papers that actually used an algorithm that did learn from the data were included, which only occurred in a minority of the hits from this search.

Machine Learning to Analyze Text Data in Management

Management researchers have often used machine learning to analyze text data. Machine learning can capture complex social constructs such as corporate culture (Li et al., 2021) and the communication styles of chief executive officers (CEOs; Choudhury et al., 2019). Machine learning is ideal for such analyses because it can easily code constructs of interest in text data and can discover conceptual themes within text using topic modeling.

Topic Classifiers

In management, topic classifiers have been commonly used to identify individual differences. Malhotra et al. (2018) used text classification to identify extroverted CEOs based on transcripts of their conference calls. They first ran the transcripts through known lexicons such as the Linguistic Inquiry and Word Count (LIWC) and the MRC Psycholinguistic database. These lexicons contain semantic and grammatical information about words that researchers can use to convert text into numeric data, such as word counts, number of phonemes, and word type. They then used a support vector machine algorithm to determine which linguistic features best predicted CEO extraversion in a subset of data already coded for CEO extroversion. Once their model had adequate accuracy, they used the model to classify all CEOs as either low or high in extroversion.

Harrison et al. (2019) used the same approach to measure CEOs' Big Five personality traits. Using the lexicon Word2Vec, they quantified transcripts of CEOs' earning calls and then used a gradient boosted machine-learning model to predict the Big Five in a subset of CEOs before using it to detect the remaining CEOs' personalities. Hickman et al. (2021) used a similar approach to predict job applicants' Big Five traits. After coding the verbal (e.g., word count), paraverbal (e.g., tone), and nonverbal information (e.g., facial expressions) in video interviews into a quantitative form, they used machine learning to optimally weigh this input to predict applicants' self-reported and interviewer-reported personality scores.

Management researchers have also used topic classifiers to code individual differences outside the Big Five. Akstinaite et al. (2021) developed a machine learning-based measure of CEO hubris (i.e., excessive pride and confidence). They first examined interviews of CEOs both high and low in hubris, processed the transcripts with LIWC, and then built a random forest machine-learning model to best classify hubris within a new set of CEOs. Marshall et al. (2022) used topic classifiers to develop a measure of leaders' charismatic, ideological, or pragmatic styles (Yammarino et al., 2013). Using historical records of U.S. Presidents' interviews, they first built a machine-learning model that predicted each U.S. president's charismatic, ideological, or pragmatic leadership score (Yammarino et al., 2013). They then used this model to measure the leadership style of U.S. governors during COVID-19 based on their COVID-related press briefings, thus gaining insight into how societal challenges affect the leadership styles used to comfort and manage citizens.

Text classification has also been used beyond measuring individual differences, such as to identify the emotional valence of text. Researchers have used emotional valence detectors to determine how people of different gender and rank vary in their use of emotional language in organizational communications (Gallus & Bhatia, 2020) and how the COVID-19 pandemic changed people's sentiment toward work-from-home arrangements on Twitter (Min et al., 2021). Organizational theorists have also used text classification to identify the "linguistic signature" of employees with good organizational fit by identifying the differences in the stylistic, topical, and emotional characteristics of emails from employees with high and low levels of organizational fit (Srivastava & Goldberg, 2017).

Researchers have also used text classification to classify product evaluations. Norinder and Norinder (2022) used deep learning to predict Amazon product ratings based on written reviews. Zhang et al. (2022) used a machine-learning model to detect nuanced sentiments in customer reviews and identify how relevant products and services can be improved. Nauhaus et al. (2021) used a machine-learning approach to classify positive and negative sentiments toward technological products within professional magazines and trade press publications, which predicted corporate capital allocation. Yiu et al. (2022) used machine learning to measure public sentiment toward nations to assess how international acquisitions affect how citizens of a nation feel toward the foreign nations that acquired domestic businesses within their market. They examined 410 acquisitions from 22 foreign nations into the Chinese market from 2010 to 2017. To assess public sentiment, they examined 100,902 social media posts in China regarding the foreign nations involved in these acquisition deals. After quantifying the data and creating a subset of manually coded posts, they then used a machine-learning algorithm known as support-vector machine to create a text classification model that marked if posts had a positive or negative sentiment. They then used regression analyses to test how various factors predicted sentiment toward the acquiring nations. They found that local public sentiment toward nations that acquired Chinese businesses was more positive when the foreign firms had greater levels of ownership postacquisition.

Lastly, topic classifiers have also been used to develop measures of job applicants’ work experience, tenure history, turnover, and approach to obtaining jobs (Sajjadiani et al., 2019) and to identify the strategic focus of top-level managers within annual reports (Kabanoff & Brown, 2008).

Topic Modeling

Within management, topic modeling is a popular approach for discovering and tracking themes in texts. For example, Choudhury et al. (2019) discovered different CEO oral communication styles through topic modeling. Using latent dirichlet allocation (LDA) to analyze transcripts of CEOs’ interviews, they found that CEOs tended to adopt five different communication styles in their speech: excitable, stern, dramatic, rambling, and melancholic. They then queried the LDA model for the words most associated with each style and used those to conduct a text classification analysis.

Similarly, Song et al. (2022) also used LDA to discover the different topics covered in technology patents and then assessed how each topic predicted patent transferability (i.e., whether patents would be sold to another party). Theme identification has also been used to examine progress in academic literature. Devinney and Hohberger (2017) assessed themes in the international business literature related to culture and how the themes were altered by Kirkman et al.’s (2006) influential review paper on cultural influences. They found that the themes covered in published articles had not shifted before versus after Kirkman et al.’s (2006) publication.

Similarly, Croidieu and Kim (2018) used such a topic-modeling approach to uncover the legitimization process of U.S. radio operators. In the late 1800s, radio operators were not seen as legitimate professionals as the radio was a recently invented form of entertainment. However, over a few decades, radio broadcasts and their social influence gained societal recognition, making radio operators a legitimized career choice. Croidieu and Kim used LDA to examine this process by identifying themes within historical texts that pertained to U.S. amateur radio operators from 1899 to 1927. By tracking how these themes differed across time, they could assess how public perceptions of radio operators shifted over time. These shifts in perceptions indicated the nature of the legitimization process through which amateur radio operators became viewed as professional experts. To become legitimized, amateurs needed to engage in four major activities: building collective competence and mastery, operating in public spaces to display mastery, contributing to societal changes and reforms, and establishing a collective identity with expected roles.

Giorgi et al. (2019) used an identical approach to investigate how public understanding of auto safety evolved over time by examining auto safety themes within congressional hearings, firm reports, and popular media. They used this information as a contextual background to understand how firms responded to legal and social pressures, finding that firms' responses to legal pressure ultimately depended on whether the cultural context constrained them or allowed them to potentially alter the law. Furman and Teodoridis (2020) also used this temporal topic-modeling approach to discover themes to assess how the content of engineering research changes and diversifies after the introduction of innovative technologies.

Another common use of topic modeling is assessing individual observations' uniqueness. Van Angeren et al. (2022) assessed the uniqueness of mobile applications (apps) by using LDA to examine apps' descriptions, which specify their function and features. This analysis created a list of common topics in apps and quantified the concentration of each topic within each app. They then calculated the uniqueness of each app by calculating the difference between the concentration of each topic in each app and the average topic concentration across all apps within the same app category. This method measured how much each app deviated from its category norms regarding its function and features. They then related this app distinctiveness to subsequent app performance and found a U-shaped relationship for free apps and an inverted U-shape for paid apps.

Giorgi and Weber (2015) used this approach to assess distinct framing styles in analysts' reports. Using LDA, they found 14 themes within analysts' investment advice. They then measured analysts' deviation from the average framing themes to assess the uniqueness of each analyst's framing. They found that deviating from the average frame moderately improved investors' evaluations of analysts. Topic deviation has even been used to measure firms' distinctiveness using public information available on firms' websites (Haans, 2019) and corporate annual reports (Choi et al., 2021).

Researchers can also take the outputs of the LDA model and analyze them using traditional regression-based methods. For example, researchers have used LDA to model key themes in employees' reviews on Glassdoor and then assessed the relationship between interpersonal and intrapersonal heterogeneity in themes to predict firm-level outcomes using regressions (Corritore et al., 2019). Similarly, researchers have modeled topics in the abstracts of patent filings (Kaplan & Vakili, 2015) and scientific articles (Antons et al., 2019) and used these to predict future citation counts (see also Doldor et al., 2019; Sun & Slepian, 2020 for additional examples). Researchers have used LDAs to model specific themes of interest occurring in companies' financial statements and used these to predict financial misreporting (Brown et al., 2020). In these cases, machine learning is a complement to traditional methods. The machine-learning model does not directly predict the outcome variable; it merely provides numeric values associated with constructs of interest.

While there are many uses of topic modeling in management, one commonality among these studies is that they tend to use a single algorithm—LDA.

Machine Learning to Analyze Numeric Data in Management

Management researchers have used machine learning to analyze numeric datasets with a large collection of intercorrelated variables, which cannot be analyzed using regressions due to multicollinearity and other violations of assumptions. For instance, Kumar et al. (2022) used machine learning to discover firms' alliance networks based on 48,104 distinct alliances in each year from 1975 to 1996. Such a dataset is immensely large and full of complicated and interrelated variables. However, machine learning is the ideal tool for handling such complex data and creating impactful models. Although dozens of methods can be used for such analyses, decision trees and neural networks are some of the most common approaches in management.

Decision Trees

Within management, decision trees have been used to find important predictors of outcome variables of interest. For instance, using a decision tree machine-learning model known as the Bayesian network, Hosseini (2021) was able to predict product sales based on customers' characteristics, thus determining which customer characteristics had the greatest impact on sales. They did so by first building a model that used customer attributes to predict sale outcomes and then querying the model to assess the degree to which each attribute predicted a successful sale. Similarly, Tidhar and Eisenhardt (2020) used a random forest algorithm to find which revenue features best distinguish between high and low performing apps and subsequently identified optimal revenue models for popular and unpopular digital products.

Sometimes researchers are interested in a single predictor variable; however, for complex phenomena determined by a large number of factors, they need to control for several variables, which machine learning can do easily. For instance, Miric and Jeppesen (2020) used a random forest algorithm to assess how piracy affected product innovation. Their study examined a specific instance of piracy in which a major hacking incident resulted in thousands of mobile apps being copied from online repositories and made freely available to illegally download on a third party app. Their model assessed how product innovation, operationalized as the frequency of app innovations and revisions, differed before and after the hacking event across apps that were pirated, unpirated, and pirated in the past. They did so while controlling for a large number of factors, such as app price, size, description, rating, price, market niche, the number of downloads before piracy, and developer identity. Their model found that although piracy reduced small product innovations, such as bug fixes, it did not affect substantial product revisions, such as feature updates, and increased the creation of new apps. Similarly, using a generalized boosted model, Jabbari et al. (2022) found that participating in technology-mediated human capital investments, such as online learning and working in the gig economy, increased individuals' entrepreneurial intentions; they did so while controlling for a slew of administrative tax data. Schulz et al. (2022) also used a random forest algorithm to detect the true shape of the relationship between organizational income inequality and employees' trust in their managers while accounting for employee collective voice, tenure, supervisory role, gender, occupation, employment contract, salary, firm size, age, and employee legal status; their model identified an inverted U-shaped relationship between the key predictor and the outcome.

Lastly, decision tree models have been used to create important outcome variables from data whose input is too complex to understand. Such an approach has been used with game-based assessments, that is, employee selection tools used to measure applicant aptitude and personality through a game format (Auer et al., 2022). These models use hundreds of raw trace data variables collected during game assessments, such as mouse movement, objects interacted with, and time spent playing, to predict applicants' cognitive ability, goal orientation, and personality traits (Auer et al., 2022; Wu et al., 2022).

Neural Networks

Neural networks are powerful and versatile but require considerable programming skill and computational power. Management researchers have used these algorithms to model complex processes. Grand (2020) used a neural network algorithm to examine how trace data within situational judgment tests predict respondents' work experience and expertise and whether they evaluate response options in the test based on the option's objective behavior or its potential consequences. Thus, the researcher modeled respondents' cognitive processes while they were making judgments and decisions during the assessment. Gibson (2000) used neural networks to simulate human learning based on experiments in which participants made hypothetical judgments about how to increase firm production in environments of varying uncertainty; the researcher was able to model human learning within dynamic environments using the neural network model.

Management scholars have also used neural networks to model complex social phenomenon. Kennedy and McComb (2014) used machine learning to model how teams' processes predict their performance. First, Kennedy and McComb ran a series of laboratory experiments in which 180 lab participants were grouped into 60 teams and charged with creating a work schedule that minimized cost. As the teams set about their task, they were recorded, thus documenting their team processing. These recordings were then manually rated to quantify when teams were focused on the following team processes: mission analysis (e.g., discussion of task objectives), goal specification (e.g., discussion of task goals), tactical strategy (e.g., discussion of planned courses of action), operational strategy (e.g., discussion of the individual roles and responsibilities of team members), and action process (e.g., discussion of task performance). The researchers then marked process shifts, that is, when teams transitioned in their focus from one team process to the next; thus, marking the temporal progression of the team. The teams' performances were then measured based on the cost of their work schedules and the time it took them to complete the task. Kennedy and McComb then used a neural network algorithm to model how process shifts predicted team performance, allowing them to model the nonlinear relationships between the variables. They then used the relationships found within the machine-learning model to create a prototype of an optimal team with ideal process shifts and ran simulations on how such a team would perform under the effects of various interventions.

Neural networks have also been used for simpler purposes. For instance, Halim et al. (2021) used neural networks to model how firms' liquidity and financial efficiency predicted firms' financial distress over time, thus using machine learning to complete a time series analysis. He et al. (2020) used neural networks to generate new data while analyzing factors that predict the resolution of governance disputes. Sen and Puranam (2022) used a neural network algorithm, called an artificial neural network, along with a decision tree-based algorithm, called random forest, to predict firms' adoption of new business practices. They used the Preqin dataset containing 4,505 private equity firms across more than 60 industrial sectors from 1990 to 2016. They then used their two machine-learning models to assess their incredibly complex dataset and determine which factors were most relevant to the adoption of new practices. They did so by first building two machine-learning models with adequate accuracy in predicting adoption. They then assessed both models to determine how much each variable contributed to either models' accuracy. The variables with the highest impact were determined to be the most relevant to adoption. Then, to further interpret the effects of these variables, the researchers ran a lasso regression examining how the most relevant variables from their neural network and random forest models related to adoption. This analysis allowed the researchers to further narrow down the top predictors and obtain regression coefficients that are simpler to interpret. In their analyses, they found that firms' number of unique coinvestors negatively related to the adoption of new business practices presumably due to a limited capacity to form new alliances needed for adoption and a greater reliance on partners who may have rivalries with new adoption-relevant partners. While neural network-based models are uncommon in management, neural networks are the machine-learning model of choice in industry, natural sciences, and engineering (Adeli, 2001; Meireles et al., 2003; Tadeusiewicz, 2015; van der Baan & Jutten, 2000).

Criticisms of Machine Learning

Despite its strengths, there are many common criticisms against the use of machine learning in management studies. For one, many researchers worry that machine-learning models are overfitted. They worry that the models are perfectly weighted to maximally predict an outcome in a single dataset and that they will not generalize to new contexts. However, as mentioned in section “Basic Process of Building a Machine-Learning Model With Numeric Data”, numerous procedures can be used to prevent overfitting, such as regularization, cross-validation, and unseen testing (Ghojogh & Crowley, 2019).

Another common criticism is that machine learning is an atheoretical approach. This criticism fundamentally misunderstands the purposes of machine learning. Machine-learning models are not meant to be guided by theory. Instead, they are guided by empirical patterns in the data that can be used to inform theory. Thus, by their very nature, they are data-driven rather than theory-driven. As will be discussed in the next section “Future Research Directions for Machine Learning in Management”, machine learning can be used to discover patterns in *abductive* research. Such a process can then be used to develop and test new theoretical perspectives using machine learning. Though theory has little role in creating a machine-learning model, its insights can lead to greater theory refinement.

Next, within management, researchers might worry that creating machine-learning models requires more time and effort than the analyses are worth. Researchers not only need to build their knowledge on machine-learning approaches and programming, but they must also find hardware with enough processing power to complete the intense computations required. Moreover, machine-learning models require large datasets that are hard to find and are often incomplete. It cannot be denied that there are difficulties in using machine learning for management scholars. If a research question can be answered with simpler methods, researchers can conserve their energy and use a simpler approach. If there are no large datasets available, then even considering machine learning is purposeless as it is not even feasible. However, if researchers are dealing with large and complex datasets, machine learning can extract much more information than regressions. Moreover, most machine-learning tools and related tutorials are in the open source domain, and many universities provide free high-powered computing, thereby creating a low barrier to entry if researchers wish to learn machine learning.

Lastly, scholars have raised concerns for the possibility of “machine bias” in which machine learning perpetuates prejudice and discrimination (e.g., racism, sexism, ageism, and classism; Angwin et al., 2016). However, it is important to remember that machine learning is a tool. It is the responsibility of the tool’s creator to test for and correct any biases. Bias in machine-learning systems is a result of a variety of missteps during data curation, data verification, model building, and model implementation (for a more thorough description, see Suresh & Guttag, 2021). For instance, some models are biased because they were built on nondiverse samples (Prates et al., 2020). A dataset that does not adequately represent minorities forces the model to focus on learning patterns from majority groups (e.g., White Americans, men) in order to optimize its accuracy. Thus, the resulting model functions well for majority groups but has little information on how to make predictions for minority samples. This becomes an issue if the nature of the phenomenon differs across social groups, leading to more misclassification for minority groups. Alternatively, a machine-learning model may be biased if it is trained on data that reflects a real world bias. Machine-learning models learn to mimic human decisions. If the human decisions on which the model was trained were biased, then the model will inevitably learn to exhibit that bias. Any system built with such fundamental flaws would be biased regardless of the method used.

There are many methods to mitigate bias in machine-learning systems. (For a more thorough description, see van Giffen et al., 2022.) For instance, researchers can ensure that the training data is diverse and represents the underlying population so that the model identifies patterns and relationships that are relevant across majority and minority groups. Also, researchers need to carefully evaluate whether the dataset has any built-in biases. For instance, when examining criminal justice systems, researchers need to ask whether the judicial system is rife with racial biases, as any model built on such data will mimic that bias. Additionally, when choosing variables to represent key phenomenon, researchers need to consider whether the variables are equally appropriate and accurately measured across various groups. During model building, researchers can include a special model parameter (i.e., a regularizer to the loss function) that can help assess differences in classification across social groups (Kamishima et al., 2012). Researchers could also build multiple models, one for each social group, and then combine these models to avoid the disproportionate prioritization of certain groups (Calders & Verwer, 2010). Once models are built, researchers can assess model evaluation metrics by group to ensure that it performs equally well across various groups. Lastly, after having built the model, researchers can incorporate humans into the machine learning-based system, encouraging them to question and analyze the machine-learning model's recommendations, thus helping avoid blind adherence to or confirmation bias in favor of a potentially biased system (Provost & Fawcett, 2013).

It is important to note that management researchers' concern that machine learning can be used for nefarious purposes, such as discrimination (Kleinberg et al., 2018), ignores a key point: machine learning is simply a tool (Ahmed et al., 2022). Machine-learning algorithms are mathematical algorithms that seek to make the best predictions possible based on all the information available. In this sense, machine-learning algorithms are no different from regressions. Both regressions and machine learning can be misused by being applied on biased data to make biased conclusions. The predictive power of machine learning can be used for nefarious purposes. For instance, there are machine-learning models that can analyze people's faces and classify their ethnicity with high accuracy (Leibold, 2020). Those with malevolent intentions can use these models to discriminate against and target people from particular ethnicities. The crux of this problem is in the user of machine learning, not of the machine-learning method per se. Just because Oppenheimer (1948) used Einstein's $E = mc^2$ equation to create the atomic bomb does not mean that Einstein's equation had a "dark side"; instead, it merely shows that scientific tools can be used for both useful and morally unsound purposes.

Future Research Directions for Machine Learning in Management

Much impressive work in management focuses on using machine learning to measure and predict phenomena. However, as with all methods, there are still uses of machine learning that management researchers have not fully utilized.

Unsupervised Machine Learning for Quantitative Data

Unsupervised machine learning remains underutilized in management research. The only predominant use of unsupervised machine learning is topic modeling. While topic modeling has allowed researchers to creatively engage, analyze, and measure phenomena using qualitative data, unsupervised quantitative data analyses are noticeably less common. A variety of archival databases could already be used for such an analysis. These datasets provide information on firm-level phenomena, such as measures of financial performance from Compustat (Uotila et al., 2009), corporate social responsibility from the MSCI KLD 400 Social Index (Godfrey et al., 2009), and firms' mergers, acquisitions, and alliances from the Securities Data Company (Villalonga & McGahan, 2005). There are also datasets that examine individual-level phenomena, such as measures of social values and attitudes from the World Values Survey (WVS) (Cullen et al., 2004), leadership preferences and beliefs from Global Leadership and Organizational Behavior Effectiveness (House et al., 2002), and employee job satisfaction and perceptions of the organization from Glassdoor (Dineen & Allen, 2016).

Using unsupervised machine learning would allow researchers to identify how phenomena, such as firm behavior and employee characteristics, cluster and group together (Gentleman & Carey, 2008). For instance, many management researchers use clustering approaches to assess employees' biodata and create employee profiles (e.g., Cortina & Wasti, 2005; Schmitt et al., 2007; Woo et al., 2020). Such profiles can be generated using unsupervised machine-learning methods, such as k-means clustering, which will allow researchers to automatically evaluate far larger collections of complex biodata and devote more effort to interpreting the processed data and model to generate meaningful insights (Sinaga & Yang, 2020). Another example of the utility of unsupervised learning is Kumar et al. (2022) who used unsupervised machine learning to conduct a clustering analysis to discover firms' alliance networks based on a large dataset containing 48,104 distinct alliances in each year from 1975 to 1996.

Unsupervised machine learning also allows researchers to engage in dimensionality reduction by excluding irrelevant variables from a dataset while still keeping variables that provide valuable information and boost model performance (Gentleman & Carey, 2008). Unsupervised learning's unrestricted data analysis also allows the algorithm to uncover novel patterns that the human mind might not consider and can indicate problem spaces within the literature that need further investigation and development (Leavitt et al., 2021). Perhaps the model discovers clusters of job characteristics that indicate the emergence of a new form of work. Such suggestions can lead researchers to new theoretical discoveries as they investigate the problem spaces indicated by the model.

Algorithm Supported Abduction

In the past decade, editors of top management journals have called for the use of machine-learning methods and have recognized the potential of machine learning (George et al., 2014; George, Howard-Grenville, et al., 2016; George, Osinga, et al., 2016). Specifically, researchers have started to recognize machine learning as a powerful tool to discover new relationships and generate insights in the organizational sciences above and beyond traditional methods (Choudhury et al., 2021; Shrestha et al., 2021).

Machine learning is the ideal tool for discovering empirical relationships and patterns, a necessary first step for theory development. First, as an automated learning process that creates models to predict a given outcome, machine-learning models improve upon traditional analytical methods through their greater accuracy (Ware, 1955). However, regarding pattern discovery, the greatest benefits of machine-learning models lie in their ability to identify complex relationships that are generally robust and replicable (Choudhury et al., 2021; Shrestha et al., 2021). They can do so due to their sophisticated functional form, which better handles complex relationships between variables than regression-based methods (Shrestha et al., 2021).

Second, machine learning has better protection against overfitting than most analytical methods, improving the reliability of identified patterns (Choudhury et al., 2021). As mentioned in "Basic Process of Building a Machine-Learning Model With Numeric Data", model overfit is caused by excessive model complexity, a result of the overinclusion of predictors, or by excessive sample dependence, a result of the cherry-picking of variables (Hawkins, 2004). Within both cases, the models created fit exceptionally well to the current sample but fail to generalize to new observations due to a high level of prediction error (Choudhury et al., 2021). Machine learning protects against these risks through regularization, a model parameter that penalizes model complexity, and cross-validation, which tests how well the model generalizes across samples (Ghojogh & Crowley, 2019).

Lastly, given algorithms' greater computational ability, machine learning can assess and identify patterns and variables that the human mind would have difficulty comprehending while also being free of the cognitive biases common to human reasoning (Shrestha et al., 2021). All of these capabilities allow machine learning-based pattern discovery to be unconstrained by the methodological and cognitive limits common in other quantitative and human judgement-based approaches, making it the perfect tool to engage in abductive research (Choudhury et al., 2021; Shrestha et al., 2021).

Abductive research involves a cycle of theory building and theory testing. It starts with theory building, wherein the researcher assesses data to discover empirical relationships and patterns. These discoveries are then used to engage in theory testing by creating hypotheses and testing them in follow-up studies (Behfar & Okhuysen, 2018; Haig, 2005). In algorithm-supported abduction, machine-learning algorithms' superior computational abilities and pattern detection skills are used to complete the first step of the abductive process: discovering empirical relationships and patterns.

Algorithm-supported abduction starts with machine learning-based pattern discovery (Shrestha et al., 2021). To do so, researchers first build a model of their phenomenon (Choudhury et al., 2021). Once the model is built, researchers must then determine if the model can adequately predict the outcome—a decision that can be made by comparing the machine-learning model's performance to a baseline regression-based model (Choudhury et al., 2021).

If the machine-learning model's accuracy surpasses the baseline, researchers can then query it to understand the effects of individual variables. First, researchers can query the model to determine which variable is most influential in predicting the outcome by examining individual variables' total effect on the outcome (Biecek, 2018). Thus, researchers can know which variable is most closely related to the outcome, even when accounting for all the other variables in the model. Next, researchers can examine the nature of the effect of individual variables through partial dependence plots (Greenwell, 2017). These plots illustrate the relationship between the individual variable and the outcome while holding the effects of all the other variables in the model constant. This approach best captures the shape of the effect as it calculates the value of the outcome at each value of the independent variable with no assumption on the shape of the effect.

Understanding the impact of variables and the nature of their effect, researchers can then consult the literature to explain the empirical relationships discovered (Shrestha et al., 2021). This stage is vital to algorithm-supported abduction as it allows machine learning and theory to play complementary roles in advancing the field (Leavitt et al., 2021). By consulting and considering the theory, machine learning goes from an identifier of empirical relationships to an identifier of theoretical areas that require further consideration. Having already known the effect, the theorist can then determine which theoretical perspectives are informed by the model, describe the theoretical mechanism behind the machine-learning findings, and contribute to theory building within the field. With a set theory, the researcher can then create formal hypotheses that will subsequently be tested using conventional methods in follow-up studies (Leavitt et al., 2021; Shrestha et al., 2021).

How hypotheses are tested in follow-up studies depends on the nature of the research. In primarily archival research where primary data is difficult or impossible to obtain, hypothesis testing can be done by splitting the data at the beginning of the abduction process (Shrestha et al., 2021). The data is split into two equal subsamples, with the first sample used to build the machine-learning model and create hypotheses and the second sample used to deductively test the hypotheses generated. Such an approach has been used in strategic management with Sen and Puranam (2022) using algorithm-supported abduction to identify antecedents of new business practice adoption. Sen and Puranam used the Preqin dataset containing 4,505 private equity firms across more than 60 industrial sectors from 1990 to 2016. They then equally split the data into two subsamples: Sample I and Sample II. They built their model using Sample I and empirically found a negative relationship between firms' unique coinvestors and the adoption of new business practices. Consulting the literature, they theorized that the more partners firms had in their alliance portfolios, the less likely they were to adopt new business practices due to a limited capacity to form new alliances needed for adoption and due to a greater reliance on partners who may have rivalries with new adoption-relevant partners. They then tested and found support for their formal hypotheses using regression modeling in their Sample II data. Through algorithm-supported abduction, Sen and Puranam were able to identify a counterintuitive phenomenon relevant to the strategic interest of firms.

With regard to lower-level phenomena for which primary data can be collected, algorithm-based hypotheses can be tested using a variety of observational and experimental methods (Leavitt et al., 2021). Within such research, all of the archival data is used to conduct the machine-learning analysis. Then new data is collected to test the abductive hypotheses, ensuring that the effect is not specific to the dataset. Within the domain of psychology, Sheetal et al. (2020) used such algorithm-supported abduction to find predictors of ethicality. Based on WVS respondents' answers to a series of 700 questions, Sheetal et al. built a deep-learning model to classify individuals' ethicality based on their attitudes, values, and beliefs. They found that respondents' optimism about the future of humanity was one of the strongest predictors of ethicality, a variable previously unexamined in the past literature on unethical behavior. They verified their machine-learning insight of optimism reducing unethical behavior with a correlational and an experimental study, thereby verifying the abductive hypothesis.

Algorithm-supported abduction is an innovative method used to develop theory within management. It allows researchers to balance theoretical insights with the computational power of machine learning, resulting in highly impactful and relevant theoretical discoveries. However, to fully utilize these advantages, researchers must first understand how machine-learning models function.

Conclusion

Machine learning is a long-established analytical method that has started making headway in management research. Despite numerous calls to use machine learning in management (e.g., George et al., 2014; George, Howard-Grenville, et al., 2016; George, Osinga, et al., 2016), management researchers have only used a narrow range of machine-learning methods to address a relatively narrow set of research questions. This article suggests that management scholars can consider adopting a wider range of machine-learning methods to answer a wider set of research questions. Over the past decade, machine learning-based predictive models have largely replaced regression-based predictive models in industry. Over the coming decades, a similar trend is likely to unfold in the management sciences. Thus, a deeper understanding of the many machine-learning algorithms available and their uses can help researchers capitalize on this shift.

Further Reading

Papers on Using Machine Learning for Scientific Discovery

Friederich, P., Krenn, M., Tamblyn, I., & Aspuru-Guzik, A. (2021). Scientific intuition inspired by machine learning-generated hypotheses. *Machine Learning: Science and Technology*, *2*(2), 025027.

Papers for Details on Model Building

Janocha, K., & Czarnecki, W. M. (2017). On loss functions for deep neural networks in classification *<https://doi.org/10.4467/20838476SI.16.004.6185>*. *Schedae Informaticae*, *25*, 49–59.

Refaeilzadeh, P., Tang, L., & Liu, H. (2016). Cross-validation <https://doi.org/10.1007/978-1-4899-7993-3_565-2>. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 1–7). Springer.

Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2022). A comprehensive survey of loss functions in machine learning <https://doi.org/10.1007/s40745-020-00253-5>. *Annals of Data Science*, *9*(2), 187–212.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice <https://doi.org/10.1016/j.neucom.2020.07.061>. *Neurocomputing*, *415*, 295–316.

Papers on Evaluation Metrics

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview <https://doi.org/10.48550/arXiv.2008.05756>. *arXiv*.

Papers on Specific Algorithms

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms <https://doi.org/10.1007/s10462-020-09896-5>. *Artificial Intelligence Review*, *54*(3), 1937–1967.

Schopf, T. (2022). *Unsupervised text classification with Lbl2Vec* <https://towardsdatascience.com/unsupervised-text-classification-with-lbl2vec-6c5e040354de>. Towards Data Science.

References

Adeli, H. (2001). Neural networks in civil engineering: 1989–2000 <https://doi.org/10.1111/0885-9507.00219>. *Computer-Aided Civil and Infrastructure Engineering*, *16*(2), 126–142.

Ahmed, M. A., Chatterjee, M., Dadure, P., & Pakray, P. (2022, May). *The role of biased data in computerized gender discrimination* [Conference session]. 2022 IEEE/ACM 3rd International Workshop on Gender Equality, Diversity and Inclusion in Software Engineering (GEICSE), Pittsburgh, PA.

Akstinaite, V., Garrard, P., & Sadler-Smith, E. (2021). Identifying linguistic markers of CEO hubris: A machine learning approach <https://doi.org/10.1111/1467-8551.12503>. *British Journal of Management*, *33*(3), 1163–1178.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. *ProPublica*.

Antons, D., Joshi, A. M., & Salge, T. O. (2019). Content, contribution, and knowledge consumption: Uncovering hidden topic structure and rhetorical signals in scientific texts. *Journal of Management*, *45*(7), 3035–3076.

Auer, E. M., Mersy, G., Marin, S., Blaik, J., & Landers, R. N. (2022). Using machine learning to model trace behavioral data from a game-based assessment <https://doi.org/10.1111/ijsa.12363>. *International Journal of Selection and Assessment*, *30*(1), 82–102.

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach <https://doi.org/10.1111/1475-679X.12292>. *Journal of Accounting Research*, *58*(1), 199–235.

Behfar, K., & Okhuysen, G. A. (2018). Perspective—Discovery within validation logic: Deliberately surfacing, complementing, and substituting abductive reasoning in hypothetico-deductive inquiry <https://doi.org/10.1287/orsc.2017.1193>. *Organization Science*, *29*(2), 323–340.

Berry, W. D. (1993). *Understanding regression assumptions* (1st ed.). SAGE.

Biecek, P. (2018). DALEX: Explainers for complex predictive models <http://arxiv.org/abs/1806.08915>. *arXiv*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms <https://doi.org/10.28945/4184>. *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 45–76.

Breiman, L. (2001). Random forests <https://doi.org/10.1023/A:1010933404324>. *Machine Learning*, *45*(1), 5–32.

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, *58*(1), 237–291.

Browne, M. W. (2000). Cross-validation methods <https://doi.org/10.1006/jmps.1999.1279>. *Journal of Mathematical Psychology*, *44*(1), 108–132.

Cabitza, F., & Zeitoun, J.-D. (2019). The proof of the pudding: In praise of a culture of real-world validation for medical artificial intelligence <https://doi.org/10.21037/atm.2019.04.07>. *Annals of Translational Medicine*, *7*(8), 161.

Calders, T., & Verwer, S. (2010). Three naive Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, *21*, 277–292.

Chen, T., & Guestrin, C. (2016, August). *XGBoost: A scalable tree boosting system* <https://doi.org/10.1145/2939672.2939785> [Conference session]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA.

Cho, P. H., Chen, T., Kalia, R., Britten, R., Yuan, J., Cross, J. H., Schneider, G. W., Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly*, *65*(2), 359–394.

Choi, J., Menon, A., & Tabakovic, H. (2021). Using machine learning to revisit the diversification–performance relationship <https://doi.org/10.1002/smj.3317>. *Strategic Management Journal*, *42*(9), 1632–1661.

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research <https://doi.org/10.1002/smj.3215>. *Strategic Management Journal*, *42*(1), 30–57.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles <https://doi.org/10.1002/smj.3067>. *Strategic Management Journal*, *40*(11), 1705–1732.

Church, K. W. (2017). Word2Vec <https://doi.org/10.1017/S1351324916000334>. *Natural Language Engineering*, *23*(1), 155–162.

Corritore, M., Goldberg, A., & Srivastava, S. B. (2019). Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly*, *65*(2), 359–394.

Cortina, L. M., & Wasti, S. A. (2005). Profiles in coping: Responses to sexual harassment across persons, organizations, and cultures. *Journal of Applied Psychology*, *90*(1), 182–192.

Croidieu, G., & Kim, P. H. (2018). Labor of love: Amateurs and lay-expertise legitimation in the early U.S. radio field <https://doi.org/10.1177/0001839216686531>. *Administrative Science Quarterly*, *63*(1), 1–42.

Cullen, J. B., Parboteeah, K. P., & Hoegl, M. (2004). Cross-national differences in managers' willingness to justify ethically suspect behaviors: A test of institutional anomie theory. *Academy of Management Journal*, *47*(3), 411–421.

Deng, Z., Liesch, P. W., & Wang, Z. (2021). Deceptive signaling on globalized digital platforms: Institutional hypnosis and firm internationalization. *Journal of International Business Studies*, *52*(6), 1096–1120.

Devinney, T. M., & Hohberger, J. (2017). The past is prologue: Moving on from culture's consequences <https://doi.org/10.1057/s41267-016-0034-z>. *Journal of International Business Studies*, *48*(1), 48–62.

Dineen, B. R., & Allen, D. G. (2016). Third party employment branding: Human capital inflows and outflows following "best places to work" certifications. *Academy of Management Journal*, *59*(1), 90–112.

Doldor, E., Wyatt, M., & Silvester, J. (2019). Statesmen or cheerleaders? Using topic modeling to examine gendered messages in narrative developmental feedback for leaders. *The Leadership Quarterly*, *30*(5), 101308.

Dong, G., & Liu, H. (Eds.). (2018). *Feature engineering for machine learning and data analytics*. CRC Press.

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? <https://doi.org/10.1007/978-3-319-18305-3_1> In I. El Naqa, R. Li, & M. Murphy (Eds), *Machine learning in radiation oncology* (pp. 3–11). Springer.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research <https://doi.org/10.1037/0003-066X.63.7.591>. *American Psychologist*, *63*(7), 591–601.

Fix, E., & Hodges, J. L. (1989). Discriminatory analysis: Nonparametric discrimination; consistency properties. *Revue Internationale de Statistique* [International Statistical Review], *57*(3), 238–247.

Friedman, J. H. (2002). Stochastic gradient boosting <https://doi.org/10.1016/S0167-9473(01)00065-2>. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Furman, J. L., & Teodoridis, F. (2020). Automation, research technology, and researchers' trajectories: Evidence from computer science and electrical engineering <https://doi.org/10.1287/orsc.2019.1308>. *Organization Science*, *31*(2), 330–354.

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation <https://doi.org/10.1007/s11222-009-9153-8>. *Statistics and Computing*, *21*(2), 137–146.

Gallus, J., & Bhatia, S. (2020). Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of Wikipedia editor conversations <https://doi.org/10.1016/j.obhdp.2020.02.003>. *Organizational Behavior and Human Decision Processes*, *160*, 115–130.

Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning <https://doi.org/10.1007/978-0-387-77240-0_10>. In *Bioconductor case studies: Use R!* Springer.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management <https://doi.org/10.5465/amj.2014.4002>. *Academy of Management Journal*, *57*(2), 321–326.

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research <https://doi.org/10.5465/amj.2016.4007>. *Academy of Management Journal*, *59*(6), 1880–1895.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research <https://doi.org/10.5465/amj.2016.4005>. *Academy of Management Journal*, *59*(5), 1493–1507.

Ghojogh, B., & Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial <https://doi.org/10.48550/arXiv.1905.12787>. *arXiv*.

Gibson, F. P. (2000). Feedback delays: How can decision makers learn not to buy a new car every time the garage is empty? <https://doi.org/10.1006/obhd.2000.2906> *Organizational Behavior and Human Decision Processes*, *83*(1), 141–166.

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence <https://doi.org/10.1126/science.1259439>. *Science*, *346*(6206), 171–172.

Giorgi, S., Maoret, M., & Zajac, E. J. (2019). On the relationship between firms and their legal environment: The role of cultural consonance <https://doi.org/10.1287/orsc.2018.1250>. *Organization Science*, *30*(4), 803–830.

Giorgi, S., & Weber, K. (2015). Marks of distinction: Framing and audience appreciation in the context of investment advice <https://doi.org/10.1177/0001839215571125>. *Administrative Science Quarterly*, *60*(2), 333–367.

Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, *30*(4), 425–445.

Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges <https://doi.org/10.1093/eurheartj/ehw302>. *European Heart Journal*, *38*(23), 1805–1814.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Grand, J. A. (2020). A general response process theory for situational judgment tests <https://doi.org/10.1037/apl0000468>. *Journal of Applied Psychology*, *105*(8), 819–862.

Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *The R Journal*, *9*(1), 421–436.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297.

Haans, R. F. J. (2019). What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories <https://doi.org/10.1002/smj.2978>. *Strategic Management Journal*, *40*(1), 3–27.

Haig, B. D. (2005). An abductive theory of scientific method <https://doi.org/10.1037/1082-989X.10.4.371>. *Psychological Methods*, *10*(4), 371–388.

Halim, Z., Shuhidan, S. M., & Sanusi, Z. M. (2021). Corporation financial distress prediction with deep learning: Analysis of public listed companies in Malaysia <https://doi.org/10.1108/BPMJ-06-2020-0273>. *Business Process Management Journal*, *27*(4), 1163–1178.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool <https://doi.org/10.1002/smj.3023>. *Strategic Management Journal*, *40*(8), 1316–1330.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hawkins, D. M. (2004). The problem of overfitting <https://doi.org/10.1021/ci0342472>. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12.

He, V. F., Puranam, P., Shrestha, Y. R., & von Krogh, G. (2020). Resolving governance disputes in communities: A study of software license decisions <https://doi.org/10.1002/smj.3181>. *Strategic Management Journal*, *41*(10), 1837–1868.

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2021). Automated video interview personality assessments: Reliability, validity, and generalizability investigations <https://doi.org/10.1037/apl0000695>. *Journal of Applied Psychology, 107*(8), 1323–1351.

Hosseini, S. (2021). A decision support system based on machined learned Bayesian network for predicting successful direct sales marketing <https://doi.org/10.1080/23270012.2021.1897956>. *Journal of Management Analytics*, *8*(2), 295–315.

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: An introduction to project GLOBE. *Journal of World Business*, *37*(1), 3–10.

Islam, M., Chen, G., & Jin, S. (2019). An overview of neural network <https://doi.org/10.11648/j.ajnna.20190501.12>. *American Journal of Neural Networks and Applications*, *5*(1), 7.

Jabbari, J., Roll, S., Bufe, S., & Chun, Y. (2022). Cut me some slack! An exploration of slack resources and technology-mediated human capital investments in entrepreneurship <https://doi.org/10.1108/IJEBR-10-2020-0731>. *International Journal of Entrepreneurial Behavior & Research*, *28*(5), 1310–1346.

Judge, T. A., Livingston, B. A., & Hurst, C. (2012). Do nice guys—And gals—Really finish last? The joint effects of sex and agreeableness on income <https://doi.org/10.1037/a0026021>. *Journal of Personality and Social Psychology*, *102*(2), 390–407.

Kabanoff, B., & Brown, S. (2008). Knowledge structures of prospectors, analyzers, and defenders: Content, structure, stability, and performance <https://doi.org/10.1002/smj.644>. *Strategic Management Journal*, *29*(2), 149–171.

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification <https://doi.org/10.1007/s10462-018-09677-1>. *Artificial Intelligence Review*, *52*(1), 273–292.

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012, September). *Fairness-aware classifier with prejudice remover regularizer* [Conference session]. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, Proceedings, Part II 23.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review <https://doi.org/10.1080/23270012.2020.1756939>. *Journal of Management Analytics*, *7*(2), 139–172.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457.

Keh, S. S., & Cheng, I.-T. (2019). Myers-Briggs personality classification and personality-specific language generation using pre-trained language models <https://doi.org/10.48550/arXiv.1907.06333>. *arXiv*.

Kennedy, D. M., & McComb, S. A. (2014). When teams shift among processes: Insights from simulation and optimization. *Journal of Applied Psychology*, *99*(5), 784–815.

Kirkman, B., Lowe, K., & Gibson, C. (2006). A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework <https://doi.org/10.1057/palgrave.jibs.8400202>. *Journal of International Business Studies*, *37*, 285–320.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, *10*, 113–174.

Kotsiantis, S. B. (2013). Decision trees: A recent overview <https://doi.org/10.1007/s10462-011-9272-4>. *Artificial Intelligence Review*, *39*(4), 261–283.

Kumar, P., Liu, X., & Zaheer, A. (2022). How much does the firm's alliance network matter? <https://doi.org/10.1002/smj.3379> *Strategic Management Journal*, *43*(8), 1433–1468.

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). Ghost in the machine: On organizational theory in the age of machine learning <https://doi.org/10.5465/amr.2019.0247>. *Academy of Management Review*, *46*(4), 750–777.

Lee, K. J., & Simpson, J. A. (2014). Introduction to multiple imputation for dealing with missing data. *Respirology*, *19*(2), 162–167.

Leibold, J. (2020). Surveillance in China's Xinjiang region: Ethnic sorting, coercion, and inducement. *Journal of Contemporary China*, *29*(121), 46–60.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning <https://doi.org/10.1093/rfs/hhaa079>. *The Review of Financial Studies*, *34*(7), 3265–3315.

Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., & Li, J. (2020). Dice loss for data-imbalanced NLP tasks <https://doi.org/10.48550/arXiv.1911.02855>. *arXiv*.

Liu, Y., Liu, Z., Chua, T.-S., & Sun, M. (2015). Topical word embeddings <https://doi.org/10.1609/aaai.v29i1.9522>. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1), 2418–2424.

Malhotra, S., Reus, T. H., Zhu, P., & Roelofsen, E. M. (2018). The acquisitive nature of extraverted CEOs <https://doi.org/10.1177/0001839217712240>. *Administrative Science Quarterly*, *63*(2), 370–408.

Marshall, J. D., Yammarino, F. J., Parameswaran, S., & Cheong, M. (2022). Using CATA and machine learning to operationalize old constructs in new ways: An illustration using U.S. governors' COVID-19 press briefings <https://doi.org/10.1177/10944281221098607>. *Organizational Research Methods*.

Meireles, M. R. G., Almeida, P. E. M., & Simoes, M. G. (2003). A comprehensive review for industrial applicability of artificial neural networks <https://doi.org/10.1109/TIE.2003.812470>. *IEEE Transactions on Industrial Electronics*, *50*(3), 585–601.

Min, H., Peng, Y., Shoss, M., & Yang, B. (2021). Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic <https://doi.org/10.1037/apl0000886>. *Journal of Applied Psychology*, *106*(2), 214–229.

Miric, M., & Jeppesen, L. B. (2020). Does piracy lead to product abandonment or stimulate new product development? Evidence from mobile platform-based developer firms <https://doi.org/10.1002/smj.3208>. *Strategic Management Journal*, *41*(12), 2155–2184.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial <https://doi.org/10.3389/fnbot.2013.00021>. *Frontiers in Neurorobotics*, *7*.

Nauhaus, S., Luger, J., & Raisch, S. (2021). Strategic decision making in the digital age: Expert sentiment and corporate capital allocation <https://doi.org/10.1111/joms.12742>. *Journal of Management Studies*, *58*(7), 1933–1961.

Norinder, U., & Norinder, P. (2022). Predicting Amazon customer reviews with deep confidence using deep learning and conformal prediction <https://doi.org/10.1080/23270012.2022.2031324>. *Journal of Management Analytics*, *9*(1), 1–16.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks <https://doi.org/10.48550/arXiv.1511.08458>. *arXiv*.

Oppenheimer, J. R. (1948). Physics in the contemporary world. *Bulletin of the Atomic Scientists*, *4*(3), 65–86.

Osborne, J., & Waters, E. (2019). Four assumptions of multiple regression that researchers should always test <https://doi.org/10.7275/r222-hv23>. *Practical Assessment, Research, and Evaluation*, *8*(2), 1–5.

Palacio-Niño, J.-O., & Berzal, F. (2019). Evaluation metrics for unsupervised learning algorithms <https://doi.org/10.48550/arXiv.1905.05667>. *arXiv*.

Pantano, E., Dennis, C., & Alamanos, E. (2021). Retail managers' preparedness to capture customers' emotions: A new synergistic framework to exploit unstructured data with new analytics <https://doi.org/10.1111/1467-8551.12542>. *British Journal of Management*, *33*(3), 1179–1199.

Pitigala, S., & Li, C. (2015). *Classification based filtering for personalized information retrieval* [Conference session]. Proceedings of the International Conference on Information and Knowledge Engineering (IKE).

Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google translate. *Neural Computing and Applications*, *32*(10), 6363–6381.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models <https://doi.org/10.18637/jss.v091.i02>. *Journal of Statistical Software*, *91*(1), 1–40.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain <https://doi.org/10.1037/h0042519>. *Psychological Review*, *65*(6), 386–408.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover <https://doi.org/10.1037/apl0000405>. *Journal of Applied Psychology*, *104*(10), 1207–1225.

Samuel, A. (1959). Machine learning. *The Technology Review*, *62*(1), 42–45.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview <https://doi.org/10.1016/j.neunet.2014.09.003>. *Neural Networks*, *61*, 85–117.

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A., & Shivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology*, *92*(1), 165–179.

Schulz, F., Valizade, D., & Charlwood, A. (2022). The effect of intra-workplace pay inequality on employee trust in managers: Assessing a multilevel moderated mediation effect model <https://doi.org/10.1177/0018726721994193>. *Human Relations*, *75*(4), 705–733.

Sebastiani, F. (2002). Machine learning in automated text categorization <https://doi.org/10.1145/505282.505283>. *ACM Computing Surveys*, *34*(1), 1–47.

Sen, P., & Puranam, P. (2022). Do alliance portfolios encourage or impede new business practice adoption? Theory and evidence from the private equity industry <https://doi.org/10.1002/smj.3399>. *Strategic Management Journal*, *43*(11), 1–34.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, *179*(6), 764–774.

Sheetal, A., Feng, Z., & Savani, K. (2020). Using machine learning to generate novel hypotheses: Increasing optimism about COVID-19 makes people less willing to justify unethical behaviors <https://doi.org/10.1177/0956797620959594>. *Psychological Science*, *31*(10), 1222–1235.

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? <https://doi.org/10.1287/orsc.2020.1382> *Organization Science*, *32*(3), 856–880.

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727.

Song, K., Ran, C., & Yang, L. (2022). A digital analysis system of patents integrating natural language processing and machine learning <https://doi.org/10.1080/09537325.2022.2035349>. *Technology Analysis & Strategic Management*, 1–17.

Srivastava, S. B., & Goldberg, A. (2017). Language as a window into culture <https://doi.org/10.1177/0008125617731781>. *California Management Review*, *60*(1), 56–69.

Sun, K. Q., & Slepian, M. L. (2020). The conversations we seek to avoid. *Organizational Behavior and Human Decision Processes*, *160*, 87–105.

Sunagar, P., Kanavalli, A., Nayak, S. S., Mahan, S. R., Prasad, S., & Prasad, S. (2021). News text classification using machine learning techniques <https://doi.org/10.1007/978-981-33-4909-4_35>. In V. Bindhu, J. M. R. S. Tavares, A.-A. A. Boulogeorgos, & C. Vuppalapati (Eds.), *International conference on communication, computing and electronics systems* (pp. 461–474). Springer.

Suresh, H., & Guttag, J. (2021, October). A framework for understanding sources of harm throughout the machine learning life cycle <https://doi.org/10.1145/3465416.3483305> [Conference session]. Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21), New York, NY.

Taddy, M. (2012). *On estimation and selection for topic models* <https://proceedings.mlr.press/v22/taddy12.html> [Conference session]. Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands.

Tadeusiewicz, R. (2015). Neural networks as a tool for modeling of biological systems <https://doi.org/10.1515/bams-2015-0021>. *Bio-Algorithms and Med-Systems*, *11*(3), 135–144.

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(6), 363–377.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: Analysis of S&P 500 corporations. *Strategic Management Journal, 30*(2), 221–231.

Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying . . . finding revenue model fit using machine learning and multiple cases <https://doi.org/10.1002/smj.3142>. *Strategic Management Journal*, *41*(7), 1245–1273.

van Angeren, J., Vroom, G., McCann, B. T., Podoynitsyna, K., & Langerak, F. (2022). Optimal distinctiveness across revenue models: Performance effects of differentiation of paid and free products in a mobile app market <https://doi.org/10.1002/smj.3394>. *Strategic Management Journal*, *43*(10), 2066–2100.

van der Baan, M., & Jutten, C. (2000). Neural networks in geophysical applications <https://doi.org/10.1190/1.1444797>. *Geophysics*, *65*(4), 1032–1047.

van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, *144*, 93–106.

Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, *24*, 774–780.

Villalonga, B., & McGahan, A. M. (2005). The choice among acquisitions, alliances, and divestitures. *Strategic Management Journal, 26*(13), 1118–1208.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). *Evaluation methods for topic models <https://doi.org/10.1145/1553374.1553515>* [Conference session]. Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Canada.

Wang, H., Ma, C., & Zhou, L. (2009, December). *A brief review of machine learning and its application <https://doi.org/10.1109/ICIECS.2009.5362936>* [Conference session]. 2009 International Conference on Information Engineering and Computer Science, Wuhan, China.

Ware, W. H. (1955, December). Introduction to session on learning machines [International Workshop]. Managing Requirements Knowledge, IEEE Computer Society.

Woo, S. E., Tay, L., Jebb, A. T., Ford, M. T., & Kern, M. L. (2020). Big data for enhancing measurement quality. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), Big data in psychological research (pp. 59–85). American Psychological Association.

Wu, F. Y., Mulfinger, E., Alexander, L., III, Sinclair, A. L., McCloy, R. A., & Oswald, F. L. (2022). Individual differences at play: An investigation into measuring big five personality facets with game-based assessments *<https://doi.org/10.1111/ijsa.12360>*. *International Journal of Selection and Assessment*, *30*(1), 62–81.

Yammarino, F. J., Mumford, M. D., Serban, A., & Shirreffs, K. (2013). Assassination and leadership: Traditional approaches and historiometric methods *<https://doi.org/10.1016/j.leaqua.2013.08.004>*. *The Leadership Quarterly*, *24*(6), 822–841.

Yaworsky, P. M., Vernon, K. B., Spangler, J. D., Brewer, S. C., & Codding, B. F. (2020). Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the grand staircase-escalante national monument *<https://doi.org/10.1371/journal.pone.0239424>*. *PLOS ONE*, *15*(10), e0239424.

Yiu, D. W., Wan, W. P., Chen, K. X., & Tian, X. (2022). Public sentiment is everything: Host-country public sentiment toward home country and acquisition ownership during institutional transition *<https://doi.org/10.1057/s41267-020-00380-4>*. *Journal of International Business Studies, 53,* 1202–1227.

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications *<http://arxiv.org/abs/2003.05689>*. *arXiv*.

Zaremba, W., Sutskever, I., & Vinyals, O. (2015). Recurrent neural network regularization *<https://doi.org/10.48550/arXiv.1409.2329>*. *arXiv*.

Zhang, Z., Yang, K., Zhang, J. Z., & Palmatier, R. W. (2022). Uncovering synergy and Dysergy in consumer reviews: A machine learning approach *<https://doi.org/10.1287/mnsc.2022.4443>*. *Management Science, 69*(4), 2339–2360.

Appendix

Table A1. Pros and Cons of Traditional Regression-Based Methods and Machine-Learning Methods

	Pro	Con
Traditional regression-based analyses	• Easy to use • Do not require large datasets • Easy to interpret	• Have many assumptions that limit its effectiveness • Cannot handle complex data well • All nonlinear or interactive effects need to be specified prior to analysis
Machine-learning analyses	• Handle complex data very well • Create models with higher predictive accuracy • Not limited by assumptions on the nature of data • Automatically assess complex effects	• Hard to learn/steep learning curve • Building machine-learning models takes a long time • Require large datasets • Require high performance computers

Table A2. Summary of Machine-Learning Approaches

	Purpose	Type of input data	Type of learning	Output	Example algorithms	Notes
Decision tree-based models	Regression or classification	Numeric	Supervised	Prediction of an outcome or classification of observations	Random forest, generalized boosted models, XGBoost	Easier to explain how it functions, often more interpretable.
Neural network-based models	Regression or classification	Numeric	Supervised	Prediction of an outcome or classification of observations	Deep-learning model	Can automatically capture complex relationships. Hard to explain functioning; seen as a Black Box method. Long training time.
Topic classifiers	Identify topics in text	Text	Commonly supervised	Indication of the presence of a topic in text entries	Supervised: Random forest, naïve Bayes Unsupervised: Lbl2Vec	Many preexisting classifiers are available.
Topic modeling	Discover topics in text	Text	Unsupervised	A list of topics and the words that define them found in collections of text	Latent dirichlet allocation, structural topic modeling	Requires user to state number of topics before analysis.

Table A3. Supplementary Reading

Topic	Readings
General machine-learning reviews	❖ Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research *<https://doi.org/10.1002/smj.3215>*. *Strategic Management Journal*, *42*(1), 30–57. ❖ Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *<https://doi.org/10.1287/orsc.2020.1382>* *Organization Science*, *32*(3), 856–880. ❖ Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review *<https://doi.org/10.1080/23270012.2020.1756939>*. *Journal of Management Analytics*, *7*(2), 139–172.
Further on model building	❖ Browne, M. W. (2000). Cross-validation methods *<https://doi.org/10.1006/jmps.1999.1279>*. *Journal of Mathematical Psychology*, *44*(1), 108–132. ❖ Ghojogh, B., & Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial *<https://doi.org/10.48550/arXiv.1905.12787>*. *arXiv*. ❖ Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation *<https://doi.org/10.1007/s11222-009-9153-8>*. *Statistics and Computing*, *21*(2), 137–146. ❖ Janocha, K., & Czarnecki, W. M. (2017). On loss functions for deep neural networks in classification *<https://doi.org/10.4467/20838476SI.16.004.6185>*. *Schedae Informaticae, 25*, 49–59. ❖ Refaeilzadeh, P., Tang, L., & Liu, H. (2016). Cross-validation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* *<https://doi.org/10.1007/978-1-4899-7993-3_565-2>* (pp. 1–7). Springer. ❖ Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2022). A comprehensive survey of loss functions in machine learning *<https://doi.org/10.1007/s40745-020-00253-5>*. *Annals of Data Science*, *9*(2), 187–212.Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice *<https://doi.org/10.1016/j.neucom.2020.07.061>*. *Neurocomputing*, *415*, 295–316. ❖ Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications *<http://arxiv.org/abs/2003.05689>*. *arXiv*.
Further on evaluation metrics	❖ Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms *<https://doi.org/10.28945/4184>*. *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 45–76. ❖ Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview *<https://doi.org/10.48550/arXiv.2008.05756>*. *arXiv*. ❖

Topic	Readings
	Palacio-Niño, J.-O., & Berzal, F. (2019). Evaluation metrics for unsupervised learning algorithms <https://doi.org/10.48550/arXiv.1905.05667>. *arXiv*.
Further on specific algorithms	❖ Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms <https://doi.org/10.1007/s10462-020-09896-5>. *Artificial Intelligence Review*, *54*(3), 1937–1967. ❖ Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, *3*, 993–1022. ❖ International Conference on Knowledge Discovery and Data Mining. ❖ Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. ❖ Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models <https://doi.org/10.18637/jss.v091.i02>. *Journal of Statistical Software*, *91*, 1–40. ❖ Schopf, T. (2022). *Unsupervised text classification with Lbl2Vec* <https://towardsdatascience.com/unsupervised-text-classification-with-lbl2vec-6c5e040354de>. Towards Data Science. ❖ Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727.

Table A4. Glossary

Term	Definition
abduction	A form of logical inference that seeks to find the most likely conclusion from a set of observations.
accuracy	Performance metric for machine-learning classification models. It measures the proportion of observations that were accurately classified (i.e., the sum of true positives and true negatives divided by the total number of observations).
confusion matrix	A table used to assess the performance of a machine-learning classification model. The table shows the number of observations that the model classified correctly versus incorrectly for each category in the outcome variable.
cross-validation	During the model training stage, the model is repeatedly tested on how well it predicts the outcome variable in a subset of seen data. Common methods include *k*-fold cross-validation and leave-p-out technique.

Term	Definition
decision tree-based models	Machine-learning models that are based on the premise of using classification trees to identify variables that best split the data into the categories of interest. Example algorithms include random forest, Bayesian networks, generalized boosted models, and extreme gradient boosting (XGBoost).
feature engineering	The step before building a machine-learning model in which researchers clean and process the data.
held-out likelihood	In topic modeling using LDA, it is a metric to help determine the number of topics in the data that measures a topic model's ability to represent the phenomenon in a subset of unseen data.
hyperparameter search	The process in which researchers determine the optimal values for their model's parameters. This involves building a model for different combinations of parameter values and selecting the model that generates the lowest value of the loss function. Common methods include grid search, random search, and Bayesian search.
hyperparameter tuning	Most machine-learning models have free parameters that are set to their default values. These default values are based on simulations and are generally optimal; however, these parameters are not meant to be final. Researchers should vary these free parameters to find an optimal set for their specific dataset. They must tune their model's parameters.
loss values	A metric that measures the machine-learning model's ability to predict the outcome variable by assessing the gap between the true values and the predicted values of the outcome.
lower bound	In LDA, it is a metric to help determine the number of topics in the data that measures the lower bound of the log likelihood of the model.
neural network-based models	Machine-learning models that mimic the structure of neurons in the human brain to create a neural network that optimally weighs input data to predict an outcome variable. Example algorithms include deep-learning neural networks, recurrent neural network, and convolutional neural networks.
no information rate	A metric that represents the accuracy of a machine-learning classification model if it merely guessed the most prevalent class for all observations.
one-hot-encoding	Transforming a categorical variable into a set of variables that each represent a single category of the original variable.
perceptrons	The figurative "neurons" of the neural network.
regularization parameters	A machine-learning model parameter that protects against overfitting by penalizing the model for relying on too many predictors.

Term	Definition
residuals	In LDA, it is a metric to help determine the number of topics in the data that measures the model's error.
seen data	The portion of data in unseen testing on which the model is built.
semantic coherence	In topic modeling using LDA, it is a metric that represents how semantically similar key scoring words are within each topic in the model, helping researchers understand how interpretable the topics are.
sensitivity	Performance metric for machine-learning classification models. Defined as the ratio of true positives to the number of positive cases in the data, it measures a model's ability to identify positive cases.
specificity	The ratio of true negatives to the number of negative cases in the data; it measures the model's ability to identify negative cases.
supervised machine learning	Machine-learning models that analyze data with a predetermined outcome variable in order to predict the outcome based on the input data provided. Common algorithms include deep learning, random forest, support vector machine learning (SVM), *k*-nearest-neighbors, and XGBoost.
text classification models	Machine-learning models that are used to identify prespecified topics in text data. These models are often built with supervised algorithms such as decision trees, random forest, SVM, and neural networks.
topic-modeling models	Machine-learning models that are used to discover topics or themes in text data. These models are often built with supervised algorithms such as latent dirichlet allocation (LDA).
unseen data	The portion of data in unseen testing on which the model is tested.
unseen testing	A test that assesses the accuracy of a machine-learning model in a new set of data. In this test, a certain proportion of the data are kept aside and presented to the model after the model training is complete. The model is then asked to predict the dependent variable from the independent variables in this data. The model is judged on how well it predicts the dependent variable in the unseen data.
unsupervised machine learning	Machine-learning model that analyzes data to freely build models without a specific dependent variable. Common algorithms include latent dirichlet allocation (LDA) and *k*-means clustering.
word embeddings	A tool for converting text data into a numerical format.
word2vec	A tool for converting text data into a numerical format.

Related Articles

Content and Text Analysis Methods for Organizational Research

Artificial Intelligence and Entrepreneurship Research