Business Horizons (2020) 63, 157—170

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY

www.journals.elsevier.com/business-horizons

Machine learning for enterprises: Applications, algorithm selection, and challenges

In Lee [a], Yong Jae Shin [b,*]

[a] School of Computer Sciences, Western Illinois University, Macomb, IL, U.S.A.
[b] Hankyong National University, Anseong 17579, South Korea

KEYWORDS
Machine learning;
Artificial intelligence;
Deep learning;
Big data;
Neural networks;
Chatbot;
Innovation capability;
Resources and capabilities

Abstract Machine learning holds great promise for lowering product and service costs, speeding up business processes, and serving customers better. It is recognized as one of the most important application areas in this era of unprecedented technological development, and its adoption is gaining momentum across almost all industries. In view of this, we offer a brief discussion of categories of machine learning and then present three types of machine-learning usage at enterprises. We then discuss the trade-off between the accuracy and interpretability of machine-learning algorithms, a crucial consideration in selecting the right algorithm for the task at hand. We next outline three cases of machine-learning development in financial services. Finally, we discuss challenges all managers must confront in deploying machine-learning applications.

1. Rise of machine learning

Machine learning is regarded as one of the most disruptive innovations for businesses today and a strong enabler of competitive advantage. While machine learning has been around for more than 60 years, it has only recently showed significant potential for disrupting economies and societies. Healthcare, banking, manufacturing, and transportation are just some of the industries affected by machine learning. According to Accenture (2018), machine learning can reduce costs by between 20%—25% across banking, IT operations, infrastructure, and maintenance; generate new revenues across products and services; and improve customer retention and acquisition. By transforming human processes into intelligent, automated processes, enterprises can focus their resources toward higher-value activities, such as offering better products and services to their customers and improving customer acquisition and retention.

* Corresponding author
E-mail addresses: i-lee@wiu.edu (I. Lee), yjshin@hknu.ac.kr (Y.J. Shin)

https://doi.org/10.1016/j.bushor.2019.10.005

Anecdotal evidence of the benefits of investing in machine learning abounds. Banks use machine learning to analyze various deal scenarios, match product prices to value, and increase revenue (Rizzi, Wang, & Zielinski, 2018). Machine learning is widely used for sentiment analysis of online review data (Lee, 2017); it enables investment banks to process large data sets at high speeds and to make instant predictions for various automated trading activities, such as buying and selling shares, commodities, and derivatives. Machine learning helps retail banks automate key processes, including mortgage, loan, and customer services; for example, Erica, a Bank of America chatbot, helps customers perform routine banking transactions while offering simple insights on improving financial management (Bank of America, 2019). Target, a major U.S. retailer, employs several machine-learning applications that use large data sets of millions of customers to predict their shopping behavior (Gottsegen, 2019).

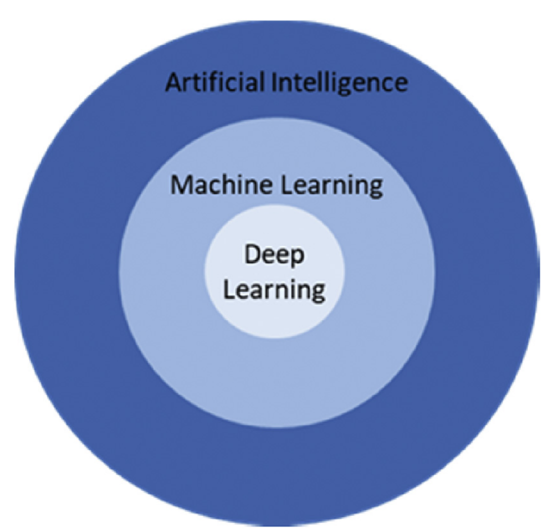
Three technological trends have driven the rapid diffusion of machine learning across industries: (1) the advent of big data; (2) advancement in computer technology in the areas of processing and data storage; and (3) advancement in machine-learning research. The adoption of machine learning is reaching its inflection point as technological, societal, and competitive pressures push enterprises to transform and innovate. As machine-learning technologies advance rapidly and more enterprises adopt the technology, managers at all levels need to familiarize themselves with machine-learning techniques to ensure that their portfolio of machine-learning projects—either in operation or under development—creates maximum value for their enterprises. A common challenge to the usability of machine-learning methods is the trade-off between interpretability and accuracy (Sarkar et al., 2016). Many machine-learning-based decisions need to be transparent and interpretable to understand the underlying decision logic. While many deep-learning methods are known to generate highly accurate outcomes, they often lack interpretability owing to their black-box nature. Therefore, in identifying the most appropriate machine-learning method for specific applications, managers need to consider the trade-off between interpretability and accuracy of various machine-learning methods.

To that end, we begin by briefly discussing the categories of machine learning, after which we present three types of machine-learning usage that have been adopted by leading companies: clustering, classification, and prediction. Then, we discuss the trade-off between accuracy and interpretability of machine-learning algorithms and selection considerations, followed by three machine-learning developments in financial services. Finally, we outline four challenges managers will face in deploying machine learning applications: the ethical challenge, the shortage of machine-learning engineers, the data-quality challenge, and the cost-benefit challenge.

2. Artificial intelligence, machine learning, and deep learning

Gartner's 2018 Hype Cycle places machine learning at the cycle's peak, which means it is being implemented in many industries, especially by early adopters (Gartner, 2018). Machine learning is a subset of artificial intelligence (AI) that can learn patterns from data without the need to define them a priori (Murphy, 2012). Applications can progressively improve their performance at specific tasks as they learn more about the data they are processing. Like many other applications, AI and machine learning also follow the input, process, and output stages in their development and operation (Canhoto & Clear, 2020; Paschen, Pitt, & Kietzmann, 2020). Figure 1 shows the relationship between AI, machine learning, and deep learning. AI is a broader domain than machine learning that includes speech and image recognition, natural-language processing (NLP), and object manipulation (Kaplan & Haenlein, 2018). The superset-subset relationship implies that while classification as machine learning is sufficient for classification as AI intelligence, the latter is necessary for the former. Deep learning is considered a subset of

Figure 1. Relationship between artificial intelligence, machine learning, and deep learning



machine learning. It brought about a renewed interest in machine learning in the 2000s, when computer processors and storage devices became sufficiently advanced to accommodate the massive, parallel neural networks necessary for machine learning. Unlike previous machine-learning methods, deep learning employs a complex structure with many nodes, hidden units, and learning algorithms, all of which influence training time and quality of learning. Deep learning is expected to be at the center of the machine-learning field, and we discuss it in more detail in Section 4.

3. Categories of machine-learning algorithms

Developing an effective machine-learning application requires suitable learning algorithms and domain knowledge in the area of the application. Several innovative machine-learning algorithms—which can be categorized into supervised, unsupervised, semisupervised, and reinforcement learning—have been developed recently.

Supervised machine learning is used to discover the mapping function with a data set that consists of input and output pairs for classification and prediction purposes. Machine learning for classification provides a discriminative ability for categorical purposes by estimating the posterior probabilities of classes with training data sets. By contrast, machine learning for prediction provides a predictive ability for decision making. Given the desired output and input data, supervised machine learning is trained to make decisions (e.g., classification or prediction), compare the calculated and desired outcomes, and reduce error rates through iterative learning by adjusting the parameter values of the internal mapping function. The learning process stops when it achieves an acceptable level of accuracy or meets other termination criteria. For supervised machine learning, labeling of input and output data for training and testing can be time-consuming and laborious, especially when big data are used. Some of the popular supervised machine-learning algorithms include k-nearest neighbors, naive Bayes classifiers, decision trees, linear regression, logistic regression, support-vector machines, random forests, and artificial neural networks (ANNs). The technical discussion of these algorithms is beyond the scope of this article. For more information about the technical details, see Jordan and Mitchell (2015).

Unsupervised machine learning is intended to capture the relationship or correlation among input data for theme analysis or grouping purposes when no information about desired outputs is available. Since there are no output classes that can relate to input data, unsupervised machine-learning algorithms attempt to identify similarities between the elements in the input data set and group the elements to gain meaningful insights. Automated multi-document summarization and customer clustering are some of the useful application areas of unsupervised machine learning. Popular unsupervised machine-learning algorithms include clustering, hierarchical clustering, k-means, principal component analysis, and association rules.

Semisupervised machine learning is a two-stage or multistage process that uses one learning type as preprocessing for the other, thus employing elements of both. For example, classification tasks via supervised machine learning (e.g., grouping customers into multiple classes) can be performed more accurately and efficiently with the help of seed-training data sets obtained from an unsupervised machine-learning algorithm (e.g., clustering) when no classification is known beforehand, or when labeling of output data would otherwise be time-consuming and costly. In another situation, where only a small portion of the data are labeled because of time constraints, the labeled data can be used to obtain classification rules with supervised machine learning, and these rules can then be used to label the remaining data by unsupervised machine learning. Finally, all labeled data can be used for training with supervised machine learning.

Reinforcement learning is used to train a software agent how to behave on the basis of environmental feedback and reward mechanisms. While supervised machine learning uses a set of input and output data pairs for training, reinforcement learning tests different actions to determine which ones maximize cumulative reward in an environment, as opposed to simply being told which actions to take. Some popular reinforcement-learning algorithms include Q-learning, hierarchical reinforcement-learning algorithms, temporal-difference learning, and policy-gradient algorithms. One real-world enterprise application of reinforcement learning is the summarization of lengthy texts. Paulus, Xiong, and Socher (2017) at Salesforce.com developed a reinforcement-learning algorithm that allows the model to generate its own summary, after which it uses an automated scorer called ROUGE (recall-oriented understudy for gisting evaluation) to compare the generated summary against the ground truth. This scorer then indicates how good the generated summary was. If the score is high, the model can update itself to improve the probability of such summaries appearing in the future. If the score is low, the model will be penalized and will change its generation procedure to prevent similar summaries.

4. Deep-learning methods

Deep learning uses multiple levels of representation and abstraction to perform complex learning tasks from a large amount of data. It has two key aspects: (1) models consisting of multiple layers or stages of nonlinear information processing; and (2) methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers (Deng & Yu, 2013). Multilayer ANNs in the 1970s and 1980s laid the foundations for deep-learning methods, which emerged in the 2000s as the most powerful machine-learning methods. High-performance graphics processors that far surpass the computational capabilities of multicore CPUs contributed to the applicability of large-scale deep-learning methods (Raina, Madhavan, & Ng, 2009).

Large data sets of diverse examples have become important sources of training in deep learning. For example, the Common Objects in Context (COCO) data set developed by Microsoft has been used for caption generation, object detection, keypoint detection, and object segmentation (Lin et al., 2014). The data set comprises more than 120,000 images for training and validation, and more than 40,000 images for testing. Innovative deep-learning methods have demonstrated impressive and often human-surpassing results in many pattern-recognition tasks, such as image recognition (Farabet, Couprie, Najman, & LeCun, 2013) and speech recognition (Sainath, Mohamed, Kingsbury, & Ramabhadran, 2013).

The deep neural network (DNN) has been the most widely used deep-learning method in the development of deep-learning applications. Unlike ANNs, which have a limited number of interconnected nodes and layers, DNNs employ a very complex structure with many nodes, hidden layers, and complex learning algorithms, all of which affect training time and accuracy. Other popular deep-learning methods include deep Boltzmann machines, deep-belief networks, and deep-reinforcement learning.

5. Machine-learning applications for enterprises

According to the IDC (2018), spending on cognitive and AI systems will reach $77.6 billion in 2022, which is more than three times the $24 billion forecast for 2018. The banking and retail industries made the largest investments in cognitive and AI systems in 2018. A recent O'Reilly Media (2018) survey indicates that about half of the world's 11,000 data specialists work for enterprises in the early stages of exploring machine learning, while the rest have moderate or extensive experience in deploying machine-learning models to production. According to Indeed.com (2019), the most advertised job is machine-learning engineer. Many leading enterprises across industries are actively recruiting machine-learning engineers, which reveals their seriousness about the potential of the field. But despite these trends, there are few articles on the current state of machine-learning applications at the enterprise level. We identify three major types of machine-learning applications used by large enterprises: (1) clustering, (2) classification, and (3) prediction.

5.1. Clustering

Clustering is used to group sets of objects on the basis of their similarities in a multidimensional space. Objects in the same cluster are more similar to each other than to those in different clusters. Clustering is considered unsupervised machine learning because the class labels of objects are not known beforehand. In clustering, objects are grouped within a high-dimensional space, and the similarity between two objects is measured with a similarity distance function. Clustering analysis reveals patterns of groups, provides insights into the key drivers of efficiency, and identifies best practices for business operations.

Clustering has been widely used to group customers to help provide personalized product and service recommendations, and to group similar documents into a predefined number of categories. For example, clustering has been used to help group customer comments to improve customer satisfaction. In social bookmarking or tagging, clusters of users that share certain tags are identified by their annotations. Netflix is a notable user of machine-learning and clustering techniques (Najafabadi, Mahrin, Chuprat, & Sarkan, 2017). It has used simple k-means and novel graphical approaches, such as affinity propagation, to divide its more than 130 million global members into over 1,000 'taste communities' of members with similar movie and TV show preferences. Netflix then recommends movies that share subjects popular within those communities.

Clustering has also been used to decide upon store layouts and locations on the basis of similarities in consumer purchase behavior or other criteria. Macy's uses clustering techniques to identify the statistical cluster membership of its stores, and it then uses an optimization model to recommend space adjustments that maximize sales potential on the basis of cross-sectional data (Carr, 2013). AutoZone also uses clustering for store segmentation to create a similar business model for stores with either similar product sales or customer brand preference (SAS, 2011).

5.2. Classification

Classification is the process of identifying the category or class of an observation. Unlike in clustering, the categories of objects are already known for purposes of training and testing. Once a classification application completes training for a given classification task, it can then assign a newly observed object to a category. One simple example of a binary classification task is the classification of a loan applicant as creditworthy or not. In this example, a machine-learning application can be trained using repayment and default data of past loan applicants. Machine-learning classifiers are also helpful for multiclass classification. For example, an application for sentiment analysis may be used to classify customers' product reviews into multiple categories (e.g., extremely positive, positive, neutral, negative, extremely negative).

Machine-learning classifiers make it possible for retailers to analyze vast amounts of customer-profile data to provide personalized customer services, and it creates opportunities for banks to serve customers more frequently and effectively. JPMorgan Chase, the largest investment bank in the U.S., has put machine-learning applications to use in several ways (JPMorgan Chase, 2017). For one, it uses a machine-learning classifier to identify clients that are best positioned for follow-on equity offerings through automated analysis of current financial positions, market conditions, and historical data. Classification is also useful for categorizing large numbers of documents. For example, State Street delivers customized, ranked newsfeeds derived from investors' portfolio holdings by using a combination of classification algorithms, NLP, and human editorial expertise. It gathers coverage from thousands of major global English-language news publications and combines machine-learning algorithms with portfolio data to curate newsfeeds for clients (State Street, 2018).

Classification is also helpful in optimizing other business operations. Anheuser-Busch, a brewing company, uses a machine-learning classifier to determine better routes for its drivers (Chandran, 2018). The classifier takes into account inputs from drivers regarding details such as parking and the best times for delivery to particular customers, and it incorporates these inputs to determine which delivery times minimize transportation costs and improve customer service. A machine-learning classifier is also transforming service operations in the retail industry. Walmart, the largest U.S. retail store, uses a classification system called Eden to categorize the freshness of fruit and vegetables. Walmart gathered food product specifications set by the USDA, layered them on their own rigorous product standards, and combined this information with more than a million photos to create a machine-learning-based freshness algorithm that prioritizes the flow of perishable goods worldwide (Musani, 2018).

5.3. Prediction

Machine learning is used to identify patterns in data and to predict future events. One major difference between classification and prediction is that classification is used to infer a rule or equation relevant to the current situation from the data (e.g., detecting credit card fraud or an internet security breach), while prediction is about anticipating what will happen in the future. For example, predictive machine learning may scrutinize data to detect market signals that will affect future market performance.

In the banking industry, machine-learning applications constantly search out the best strategies for investments and automation of financial services. JPMorgan has developed a proprietary machine-learning model to find the best execution strategy for trading orders. Using historical data and algorithms, the model calculates the probability of the best performing trade orders and executes them accordingly (McDowell, 2018). It currently provides recommendations to human traders, but is increasingly assuming an automated role in executing transactions. TD Bank also invests in machine-learning-based portfolio-management models to filter out the stocks with the highest forecasted volatility, and it then uses a traditional risk model to create low-volatility portfolios. This reduces the ex post volatility more than existing risk models (Bodjov, 2018).

Machine learning is also helpful in preventive maintenance. It enables manufacturers to learn from machine-generated data and to evolve predictive models over time. At General Electric Company (GE), machine learning for predictive analytics is used to identify anomalies, signatures, and trends in machine performance, develop a model of machine behavior, and predict when machines will need maintenance (Forbes, 2017). At Deere & Company, machine learning is a core competency that will facilitate competitive advantage in agricultural equipment (Grosch, 2018). The company is developing a predictive machine-learning system to identify plants in need of pesticides, fertilizers, and other chemicals, and to deliver optimal amounts of chemicals only where they are needed.

6. Algorithm selection: Accuracy versus interpretability

With a wide variety of machine-learning algorithms available, managers would find it challenging to select the best ones to solve their problems. It is widely known that a single machine-learning algorithm will not give the best performance across all problem instances (Kotthoff, 2016). Several variables, such as the volume, variety, and velocity of data, as well as the type of problem at hand, may affect the performance of machine-learning algorithms. One of the challenges in choosing the best algorithm is managing the trade-off between accuracy and interpretability (Operskalski & Barbey, 2016).

Accuracy is a measure of how well the algorithm will perform in practice. Interpretability refers to the ability to explain to users how a particular decision or response is made. A machine-learning algorithm is considered interpretable if its classification can be explained by a conjunction of conditional statements (i.e., if-then rules) about the collected data (Valdes et al., 2016). For example, standard decision trees with if-then rules can be explained and are thus considered interpretable. In the banking industry, the interpretability of decisions made by machine-learning algorithms can be essential to internal control and regulatory compliance. The trade-off between accuracy and interpretability arises for two reasons. First, by adding more parameters, a model's accuracy increases, but the outcomes become harder to understand and interpret. For example, a linear regression model is relatively easy to interpret since each input has one coefficient. But depending on the data, the accuracy of the model can be low. A more complex regression model with interaction terms may improve accuracy but would be more difficult to interpret. Second, certain assumptions about data distributions and model functions are instrumental to the interpretation of outcomes. But many highly accurate models, called black-box models, make few or no assumptions about the underlying data distributions and mathematical functions. While these black-box models learn the mathematical functions, the relations between model parameters are too complex for decision makers to interpret.

6.1. Accuracy of machine-learning algorithms

The accuracy of machine-learning algorithms is affected by both reducible and irreducible errors. Reducible errors originate from the fact that the chosen model will generally not be a perfect estimate of the true function, and this inaccuracy will introduce some errors. These errors can be reduced by using the best machine-learning method to estimate the function (Loeffel, 2017). But even if a machine-learning algorithm were able to capture perfectly the true function, some errors would be irreducible because of inaccurate data and missing predictor variables (James, Witten, & Hastie, 2013).

Reducible errors consist of bias error and variance error. Bias error refers to errors introduced through faulty assumptions about the nature of the function. For example, assume that there is a nonlinear relationship between input and output variables in the data set. If a company were to use a linear regression as a learning algorithm, the learning would be quick and easy, but the accuracy of the algorithm would suffer because of the high bias error that comes with using a linear regression on nonlinear data. The other type of reducible error, variance error, is the amount by which the estimate of the function learned from one training data set would change if a different training data set were used. In other words, variance error occurs because of sensitivity to small fluctuations in the training data set. To illustrate the trade-off between bias error and variance error, assume that the training data fit a quadratic regression, $Y = \beta_0 + \beta_1 x_1^2 + \beta_2 x_2^2$. A linear regression, $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, could be used for a prediction task with the same training data. This would increase the bias error of the model, though it would likely decrease the model's variance error. If the reduction in bias owing to the introduction of the quadratic term is greater than the increase in variance, then the quadratic model would be better. High variance error typically occurs when a model is overfitted to a training data set that contains random noise. Algorithms will typically display some degree of variance error caused by the difference in training data sets, and a small amount of variance error from one training data set to another is desirable as a safeguard against overfitting.

In general, decision trees and neural networks tend to overfit, and they therefore tend to have a high variance error but a low bias error. By contrast, linear regression, linear discriminant analysis, and logistic regression have a low variance error but tend to have a high bias error. To achieve the highest accuracy, machine-learning engineers need to find algorithms with the right combination of bias error and variance error in order to minimize the overall reducible error.

6.2. Interpretability of machine-learning algorithms

Ribeiro, Singh, and Guestrin (2016) suggested that an interpretable machine-learning model is one that "can be readily presented to the user with visual or textual artifacts." Interpretability of a machine-learning model is a means to engender trust (Kim, 2015). Many machine-learning algorithms, such as ANN and deep-learning algorithms, have been labeled black-box models because of their uninterpretable and complex inner workings—the very features that make these algorithms accurate. Accuracy is often prized, but interpretability is essential for investment justification, model documentation, regulatory oversight, and user acceptance and trust. White-box algorithms such as linear regressions, clustering, and decision trees can provide transparency and trust to clients at the expense of accuracy.

In the financial industry, the Equal Credit Opportunity Act requires banks to give a reason for denying a credit-card application. To prove that they are not discriminating against applicants and are complying with regulations, managers need to be able to explain their credit-issuing decisions. Because of these regulatory requirements and potential liability concerns, it is critical that recommendations made by certain types of machine-learning applications be transparent and interpretable. Furthermore, in the financial market, a single incorrect prediction can lead to a significant financial loss and liability to clients. Machine-learning algorithms should be able to explain how a given recommendation was reached. And in the healthcare industry, as more data- and privacy-protection regulations are enacted, providers will likewise need to ensure their medical services are not discriminatory or otherwise in violation of law. Across many large industries, the interpretability of machine-learning applications is already a pressing concern.

However, for some machine-learning applications, interpretability may not be a significant consideration, whether because the consequences of wrong decisions are only minor, the problems are well structured, or the algorithms have already been extensively studied and validated. For Netflix, the accuracy of movie recommendations is critical for their success and for customer satisfaction. Furthermore, transparency and interpretability of the recommendation system can facilitate customers' trust in it. But in the retail industry, say, customers may value the accuracy of chatbots but not their interpretability. So the pros and cons of interpretability and of accuracy must be weighed in each case. An improvement in one can often only be achieved at the expense of the other. Balancing the trade-off can be challenging, and managers may need to show flexibility depending on changes in solution requirements or user preferences.

7. Three cases of machine-learning development and applications in financial services

The financial-services industry is investing significantly in AI and machine learning, and it is realizing tangible benefits. In this section, we look at three large international companies—Allstate, ING, and Mizuho—and illustrate their machine-learning development activities and the interpretability of their models, drawing in part from news reports and the companies' websites.

7.1. Allstate

7.1.1. Digital transformation of Allstate's call centers with Amelia

Allstate, founded in 1931, is a publicly held personal-lines property and casualty insurer in the U.S. In September 2017, Allstate deployed Amelia, developed by IPsoft, to support customer-service representatives (Allstate, 2018). In contrast to completely automated chatbots, Amelia interacts with a customer-service representative to generate possible solutions to a variety of customer inquiries using NLP, machine-learning methods, and data analytics. Amelia leads customer-service representatives through step-by-step procedures to help find possible solutions (Allstate, 2018). In case Amelia does not know the answer to a specific inquiry, its machine-learning capability allows it to learn the answer to the customer inquiry by "listening" to interactions between representatives and customers, thus quickly expanding its knowledge base. Since it was first deployed, the average call duration has been reduced from 4.6 to 4.2 minutes (Dashevsky, 2019). In addition, 75% of inquiries have been resolved during the first call, up from 67% previously. Allstate has also employed Amelia to rapidly disseminate new corporate policies or other new information. Amelia has become an important component of Allstate's customer-service strategy, having assisted with more than three million customer inquiries.

7.1.2. Interpretability and accuracy of Amelia

Amelia's interpretability is high because its advice can be traced back step-by-step through a variety of algorithms to the source of the knowledge. The learning algorithms are well documented, and the input-output relationships are explainable. Amelia uses business process models, episodic memory, process memory, and intent recognition to enhance interpretability, and it can perform tasks according to precise specifications (Martinez, 2018). Allstate says Amelia can address more than 40 structured, industry-specific topics, such as policy, procedural, and regulatory information, and it continues to expand its knowledge over time (Allstate, 2018). This wide range of topics contributes to Amelia's accuracy.

7.2. ING

7.2.1. Using a machine-learning-based early-warning system (EWS) to assess credit risk

ING, a Dutch multinational banking and financial-services corporation, is known to use machine learning proactively—from deploying chatbots to help process customer requests, to using machine learning for bond trading and credit-risk management. As it is a lending institution, one of ING's core competencies is credit-risk management. ING must monitor its borrowers in order to rapidly assess whether repayment problems are likely. A credit-risk manager's portfolio can include dozens of corporate customers, and relevant news is easily lost in the deluge of information that results (PwC, 2018).

To improve credit-risk management, ING teamed up with Google and PwC in December 2018 to develop a machine-learning-based early-warning system (EWS) that enables credit-risk analysts to make quicker and more informed decisions by detecting whether customers are exposed to potential risks (ING, 2018). The EWS scans and analyzes large amounts of financial and nonfinancial information, including news items from all over the world. It uses Google's NLP and translation service to process approximately 80,000 items each day, including real-time market data from Refinitiv (formerly part of Thomson Reuters), as well as news from public sources (PwC, 2018). The EWS detects when a customer's share price declines by more than a preset percentage, or when a customer's media coverage becomes markedly negative. Using NLP and sentiment analysis, the EWS determines whether a news item concerns the customer in question, or whether the customer is only referred to incidentally. It then determines the news item's relevance for the credit-risk manager.

7.2.2. Interpretability and accuracy of EWS

ING developed its EWS in-house so it could more readily explain the system's underlying machine-learning models to regulators. The EWS's interpretability is high because the news-analysis algorithms, which include clustering and sentiment analysis, are relatively well structured and explainable. The system's NLP accuracy has also proven to be very high.

7.3. Mizuho

7.3.1. Machine-learning-driven solution for financial-market forecasts

Mizuho Financial Group is one of the largest financial-services companies in Japan. It manages subsidiaries and engages in ancillary operations related to their management. Over the last two decades, it has used multiple forecast models based on market data to identify prior dates similar to the prevailing market conditions. It then maps the price trends subsequent to each similar date to generate market predictions (Mizuho Financial Group, 2018). Mizuho decided to use machine-learning methods to help detect warning signs of market volatility and to generate more accurate forecasts. In March 2018, Mizuho and IBM teamed up to develop a new machine-learning-based solution for financial-market forecasts. The new solution harnesses the capabilities of a unique machine-learning technology called a dynamic Boltzmann machine (DyBM) (Mizuho Financial

Group, 2018). The DyBM is considered a time-series algorithm owing to its ability to extrapolate patterns outside of the domain of the training data (Dasgupta & Osogami, 2017). It was deployed within Mizuho's Global Markets Company to work on asset-liability management and treasury-portfolio operations (Nikolova, 2018). Mizuho (2018) noted that these two aspects are "always important for banks, especially when loan demands are sluggish, and forecasting sharp rises and falls in interest rates and stock prices is a top priority."

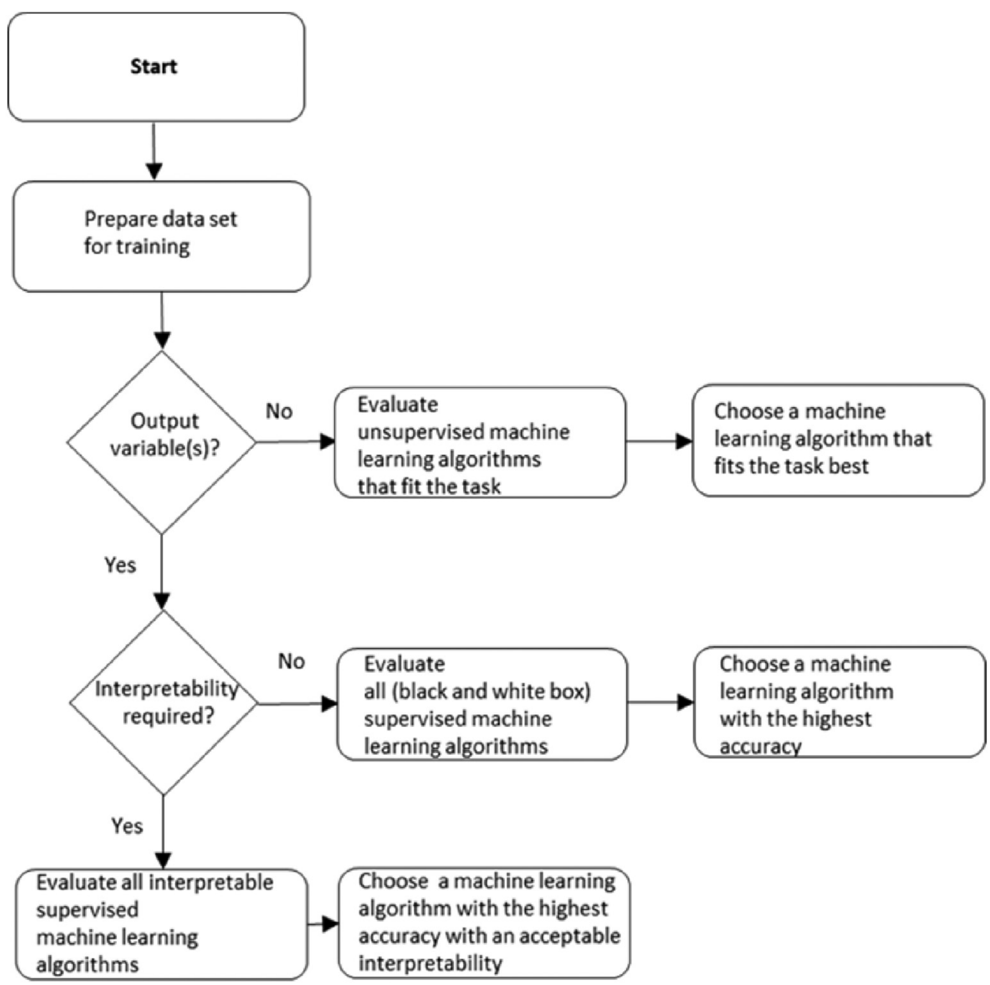
7.3.2. Interpretability and accuracy of the dynamic Boltzmann machine

The DyBM currently lacks the interpretability that is crucial to investment bankers who want to make transparent and explainable decisions. Furthermore, while the ability to derive confidence intervals is a default property of statistical time-series models, the DyBM lacks this ability because does not assume data distributions. Despite the distribution-free nature of the DyBM, its accuracy is high. The experimental results indicate that the DyBM achieves significant performance gains over other popular time-series forecasting models.

8. A process for selecting machine-learning algorithms

Selecting the best machine-learning algorithm for the data set is a challenge for any machine-learning project. No one algorithm works best for all problems, but several factors can be considered to help winnow the field. For example, some business problems require high interpretability above all. Figure 2 shows how managers can choose a machine-learning algorithm to fit the type of data set and the desired balance between

Figure 2. Choosing the right machine-learning algorithm



interpretability and accuracy. If the training data set contains input variables and one or more output variables, then supervised learning is possible. For example, if the input variables consist of market and economic factors, and the output variable is a stock price, a supervised learning algorithm can be used to develop a prediction model. If input variables are available but output variables are not, it may be necessary to use an unsupervised learning algorithm. For example, if customer demographic data are available, but customers have not been classified into groups, an unsupervised learning algorithm may be used for market segmentation.

Depending on the desired accuracy, desired interpretability, and the nature of the data set, the same firm might choose different algorithms for different purposes. Assume that a U.S. commercial bank plans to roll out an AI-driven virtual assistant for their retail banking customers. The bank has compiled a large amount of customer-service data (e.g., various customer-service inquiries and customers' personal and financial data as input variables, and service types as an output variable) with which to train a chatbot. They decide that interpretability of the chatbot is not a concern for customers but that the accuracy of its responses is critical to customer satisfaction. The company should therefore evaluate all supervised learning algorithms and select for the highest accuracy (recurrent neural networks would work well here).

Now assume that the same bank plans to develop a machine-learning application to identify client firms best positioned for various refinancing options, taking into account interest rates and economic conditions. This refinancing will enable the client firms to lower their overall financing costs by reconstructing their loan portfolio. A large data set of financial and economic parameters and refinancing decisions will be used to train the machine-learning application. Because the output variable (a refinancing recommendation) depends on the input variables (various financial and economic parameters), supervised learning is suitable for this application. As the bank needs to be able to explain its recommendations, the machine-learning application should be interpretable. Therefore, the bank should decide not to bother evaluating black-box machine-learning algorithms, such as random forests or DNNs. The bank should instead evaluate logistic regression and a decision tree for supervised learning—and it may then find the decision-tree model to be highly interpretable and more accurate than the logistic regression. As this scenario shows, the selection process can proceed iteratively, in parallel, and flexibly. Managers should ideally evaluate multiple machine-learning algorithms in parallel to determine the highest achievable accuracy, and they should use the information generated from this testing as a benchmark in selecting the most interpretable algorithm available. This benchmark can then serve as a target accuracy level for that algorithm.

9. Challenges in deploying machine learning at enterprises

As with any disruptive innovation, the adoption of machine learning will present multiple challenges. Jordan and Mitchell (2015) noted that the machine-learning field faces challenges in data privacy, data accessibility, and data sharing. However, their discussion focused on data-management concerns. We wish instead to highlight four technical and managerial challenges: the ethical challenge, the shortage of machine-learning engineers, the data-quality challenge, and the cost-benefit challenge.

9.1. The ethical challenge

While machine learning is touted as a great innovation, gaps between the design and operation of algorithms and our understanding of their ethical implications can have severe consequences for individuals, groups, and entire societies (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). Ethical concerns about machine learning may arise at any stage during the lifecycle of machine-learning application development, from early data collection to the final operation of the application. When collecting training data, a company must take care to comply with data privacy and protection rules (e.g., those governing personal medical data). During development, machine-learning algorithms will learn blindly from training data, and they will exhibit an inevitable bias if there are any latent biases in that data (e.g., biases related to gender, race, or age). During operations, users of machine-learning applications should find them equitably accessible and nondiscriminatory.

Ethical challenges abound for today's businesses, which operate in an environment characterized by unprecedented levels of complexity, intense competition, pressure from stakeholders, and social obligations (Stedham, Yamamura, & Beekunn, 2007). Under these conditions, managers in the banking industry subject to stringent performance pressures could be tempted to favor machine-learning applications that lead to biased investment decisions. Subtle discrimination in financial services or biased investment advice by machine-learning applications may be difficult to detect, especially when black-box algorithms are used. And when machine learning is used for automated job-applicant screening, an insufficiently trained machine-learning application may simply emulate the human preferences of those who designed it; biased candidate selection will result if it is not carefully trained.

Turilli (2007) argued that algorithms should reflect the same ethical principles as human workers to ensure consistency with an organization's ethical standards. Development of ethical machine-learning applications will require cooperation between researchers, developers, and policy makers. In the public sector, governments have started to regulate machine learning and automation. The U.S. Congress passed the SELF DRIVE Act and the AV START Act to address the safety of driverless vehicles, and the FUTURE of Artificial Intelligence Act to address AI concerns. In the private sector, Facebook, Google, Microsoft, and IBM aim to address ethical issues surrounding AI and to improve society's understanding of AI through the Partnership on Artificial Intelligence to Benefit People and Society (Wright & Schultz, 2018). Neubert and Montañez (2020) discussed virtue as a framework for ethical decision making and demonstrated the relevance of virtue for the ethical design and use of AI. Amid these public- and private-sector initiatives on ethical considerations of AI, managers must develop corporate ethical standards and practices for recognizing and minimizing the negative effects of biased training data sets, ensure compliance with ethical standards during application development, and conduct ethics audits.

9.2. The shortage of machine-learning engineers

According to Indeed.com (Culbertson, 2018), the demand for workers with AI talent has more than doubled over the past three years, with an increase of approximately 119% in AI-related vacancies. This is largely driven by an increase in vacancies for machine-learning and computer-vision engineers. The 2017 U.S. Emerging Jobs Report by LinkedIn (2017) also listed machine-learning engineers as the top emerging job. While the demand for machine-learning engineers has exploded and the skills gap has widened, only 3% of surveyed companies plan a significant increase in their investment in reskilling programs in the next three years (Woolf & McIntyre, 2018).

Academic institutions have been expanding machine-learning and AI-related programs, but it will take years to adequately meet the industry demand for machine-learning-educated job-seekers. When the demand for skilled machine-learning engineers increases rapidly, recruiting such talent becomes a challenge, and retraining of employees becomes increasingly important. For example, GE Global Research set up online programs that teach machine learning, and it has hosted symposia where scientists can explore new roles (Woyke, 2017). Hundreds of employees from across the company have completed GE's certification program for data analytics. Companies hoping to develop machine-learning engineers in-house will need to revise the career-development programs for their existing technical staff.

9.3. The data-quality challenge

In machine learning, data quality refers to the fitness of data for the specific purpose of machine-learning applications. Katz, Shabtai, and Rokach (2014) reported on the importance of high-quality data, as it is essential for the successful application of machine learning to real-world problems. As important as data quality is when training machine-learning applications, it is even more so during their operational use, as it is critical for ensuring confidence in applications' decisions. In the medical field, the quality of data for operational use is of paramount importance. For example, if a medical monitoring system's sensor were to generate erroneous data, the medical machine-learning application could make a wrong diagnosis, which may have fatal consequences.

Data quality is a particularly pressing issue for the performance of deep learning (Greenspan, van Ginneken, & Summers, 2016). When data are more unstructured and collected from more sources, data quality tends to decline (Lee, 2017). While structured data are essential for machine learning, more data are being created in unstructured forms (e.g., text, voice, and image data), which traditional data-management technologies are inadequate to process. Timely detection and processing of unstructured data will be necessary for the operation of machine-learning applications. Companies should establish a data-quality control process to develop quality metrics, collect new data, evaluate data quality, remove inaccurate data from the training data set, and assess the trade-off between quality-assurance costs and gains. To maintain data quality, data-management techniques such as extract, transform, and load (commonly known as ETL) can feed quality data to operational machine-learning applications.

9.4. The cost-benefit challenge

According to Deloitte's, 2018 Technology, Media, and Telecommunications Predictions report, the number of projects using machine learning will double from 2017 to 2018. The report states, however, that most enterprises have only a handful of deployments or pilots underway. Despite the touted benefits of machine learning, it can be tough to prove the value of investments to stakeholders owing to the delay between investment and reward.

Since machine learning is not a solution for all business problems, managers need to have a clear understanding of its value-generation mechanism. They should select efficacious machine-learning projects and be prepared to justify the investment to stakeholders, including senior management and users. When using emerging technologies such as the Internet of Things, there is a higher risk of project failure and higher irreversibility of investments than traditional technology projects (Lee & Lee, 2015). In addition, if tangible costs significantly outweigh tangible benefits, it will be hard to justify investment to senior management owing to the calculated negative financial returns, despite the potential for large intangible benefits. When a project is highly risky and irreversible, a real-option approach may be appropriate for investment justification (Lee & Lee, 2015). In a real-option approach, options such as postponement, expansion, shrinkage, and termination of a project represent a right rather than an obligation to execute the option. For example, in-house development of a machine-learning application may face a high risk of project failure and high irreversibility of investment. Because traditional investment-evaluation techniques may not capture the value of a right to execute the option, managers can instead use a real-option approach to reveal the value of options for machine-learning projects. An additional benefit of the real-option approach is that it enables managers to optimize the planning of machine-learning projects based on big-data platforms, such as Hadoop and Apache Spark.

10. Conclusion

Machine learning is becoming an essential part of business processes, and its diffusion across industries is expected to increase. Machine learning is changing how we work, communicate, and collaborate with our colleagues and customers. As machine learning pervades, managers who learn early on about machine-learning tools and techniques can quickly identify opportunities and potential benefits, effectively communicate their potential to stakeholders, and bring competitive advantages to their enterprises.

Because machine-learning applications are so novel, there is still a paucity of studies on their technical and managerial aspects. This makes it challenging for managers to make informed investment decisions on machine-learning application development. Our article helps managers understand the current state of machine-learning methods, types of enterprise applications, and the trade-off between interpretability and accuracy of machine-learning algorithms. Our work also presents a process model for selecting machine-learning algorithms depending on data type, desired interpretability, and desired accuracy. While machine-learning technologies have the potential to reduce costs and improve productivity, managers should prepare for the variety of challenges inherent to their design, implementation, and operation.

References

Accenture. (2018). *Redefine banking with artificial intelligence*. Available at https://www.accenture.com/_acnmedia/pdf-68/accenture-redefine-banking.pdf

Allstate. (2018, May 23). *Allstate elevates customer service through artificial intelligence*. Available at https://www.allstatenewsroom.com/news/allstate-elevates-customer-service-through-artificial-intelligence/

Bank of America. (2019). *Erica makes banking easier than ever*. Available at https://promo.bankofamerica.com/erica/

Bodjov, Y. (2018). Machine learning in risk forecasting and its application in low volatility strategy. *Toronto Dominion Bank*. Available at https://www.tdaminstitutional.com/tmi/pdfs/machlearn0918_E.pdf

Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183—193.

Carr, J. (2013). A simple approach to retail clustering. *Wilson Perumal and Company*. Available at http://cdn2.hubspot.net/hubfs/2088178/Wilson%20Perumal%20Nov2016%20Theme/PDF%20Link/Vantage_Point_2013_Issue3.pdf

Chandran, P. (2018, July 23). Disruption in retail: AI, machine learning, and big data. *Medium*. Available at https://towardsdatascience.com/disruption-in-retail-ai-machine-learning-big-data-7e9687f69b8f

Chase, JP Morgan (2017). *Innovations in finance with machine learning, big data, and artificial intelligence: Summary of the latest research and trends*. Available at https://www.jpmorgan.com/global/research/machine-learning

Culbertson, D. (2018, March 1). Demand for AI talent on the rise. *Indeed.com*. Available at https://www.hiringlab.org/2018/03/01/demand-ai-talent-rise/

Dasgupta, S., & Osogami, T. (2017). Nonlinear dynamic Boltzmann machines for time-series prediction. In *Proceedings of the thirty-first AAAI conference on artificial intelligence* (pp. 1833–1839). San Francisco, CA: AAAI.

Dashevsky, E. (2019, March 22). Allstate transforms its call centers with Amelia. *IPsoft*. Available at https://www.ipsoft.com/2019/03/22/allstate-transforms-its-call-centers-with-amelia/

Deloitte. (2018). *Technology, media, and telecommunications predictions 2018*. Available at https://www2.deloitte.com/content/dam/Deloitte/global/Images/infographics/technologymediatelecommunications/gx-deloitte-tmt-2018-predictions-full-report.pdf

Deng, L., & Yu, D. (2013). Deep learning methods and applications. *Foundations and Trends in Signal Processing, 7*(3/4), 197–387.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1915–1929.

Forbes. (2017, June 7). *How AI and machine learning are helping drive the GE digital transformation*. Available at https://www.forbes.com/sites/ciocentral/2017/06/07/how-ai-and-machine-learning-are-helping-drive-the-ge-digital-transformation/#4f3464321686

Gartner. (2018, July 23). *Hype cycle for data science and machine learning*. Available at https://www.gartner.com/doc/3883664/hype-cycle-data-science-machine

Gottsegen, G. (2019, May 13). Machine learning is changing the way retailers do business. *Built In*. Available at https://builtin.com/artificial-intelligence/machine-learning-ecommerce-retail

Greenspan, H., van Ginneken, B., & Summers, R. M. (2016). Deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging, 35*, 1153–1159.

Grosch, K. (2018). John Deere: Bringing AI to agriculture. *Digital Initiative*. Available at https://digital.hbs.edu/platform-rctom/submission/john-deere-bringing-ai-to-agriculture/

IDC. (2018). *Worldwide spending on cognitive and artificial intelligence systems forecast to reach $77.6 billion in 2022, according to new IDC spending guide*. Available at https://www.idc.com/getdoc.jsp?containerId=prUS44291818

Indeed.com. (2019, March 14). *The best jobs in the U.S.: 2019*. Available at http://blog.indeed.com/2019/03/14/best-jobs-2019/

ING. (2018, December 17). *Using AI to assess credit risk*. Available at https://www.ing.com/Newsroom/All-news/Using-AI-to-assess-credit-risk.htm

James, G., Witten, D., & Hastie, T. (2013). *An introduction to statistical learning with applications in R*. Cham, Switzerland: Springer.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260.

Kaplan, A., & Haenlein, M. (2018). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25.

Katz, G., Shabtai, A., & Rokach, L. (2014). Adapted features and instance selection for improving co-training. In A. Holzinger, & I. Jurisica (Eds.), *Knowledge discovery and data mining* (pp. 81–100). Heidelberg, Germany: Springer.

Kim, B. (2015). *Interactive and interpretable machine learning models for human machine collaboration*. PhD thesis. Boston, MA: Massachusetts Institute of Technology.

Kotthoff, L. (2016). Algorithm selection for combinatorial search problems: A survey. In C. Bessiere, L. De Raedt, L. Kotthoff, S. Nijssen, B. O'Sullivan, & D. Pedreschi (Eds.), *Data mining and constraint programming: Lecture notes in computer science*. Cham, Switzerland: Springer.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons, 60*(3), 293–303.

Lee, I., & Lee, K. (2015). The internet of things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons, 58*(4), 431–440.

LinkedIn. (2017, December 7). *LinkedIn's 2017 U.S. emerging jobs report*. Available at https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014, September). Microsoft COCO: Common objects in context. *Microsoft*. Available at https://www.microsoft.com/en-us/research/publication/microsoft-coco-common-objects-in-context/

Loeffel, P.-X. (2017). Adaptive machine learning algorithms for data streams subject to concept drifts. *HAL Archives*. Available at https://tel.archives-ouvertes.fr/tel-01812044/document

Martinez, J. (2018, May 24). The inner workings of AI systems. *IPsoft*. Available at https://www.ipsoft.com/2018/05/24/inner-workings-ai-systems/

McDowell, H. (2018, April 12). JP Morgan trading team develops machine learning model for equity desk. *The Trade*. Available at https://www.thetradenews.com/jp-morgans-trading-team-develop-machine-learning-model-equity-desk/

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society, 3*(2), 1–21.

Mizuho Financial Group. (2018, March 27). *Mizuho Financial Group, Mizuho Bank, and IBM develop new AI technology for financial market forecasting*. Available at https://www.mizuho-fg.com/release/pdf/20180327_2release_eng.pdf

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Musani, P. (2018, March 1). Eden: The tech that's bringing fresher groceries to you. *Walmart*. Available at https://corporate.walmart.com/newsroom/community/20180301/eden-the-tech-thats-bringing-fresher-groceries-to-you

Najafabadi, M. K., Mahrin, M. N., Chuprat, S., & Sarkan, H. M. (2017). Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior, 67*, 113–128.

Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. *Business Horizons, 63*(2), 195–204.

Nikolova, M. (2018, March 27). Mizuho, IBM team up to develop new AI-driven solution for financial market forecasts. *FinanceFeeds*. Available at https://financefeeds.com/mizuho-ibm-team-develop-new-ai-driven-solution-financial-market-forecasts/

Operskalski, J. T., & Barbey, A. K. (2016). Risk literacy in medical decision-making. *Science, 352*(6284), 413–414.

O'Reilly Media. (2018). *The state of machine learning adoption in the enterprise*. Available at https://www.oreilly.com/data/free/state-of-machine-learning-adoption-in-the-enterprise.csp

Paschen, J., Pitt, C., & Kietzmann, J. (2020). Artificial intelligence: Building blocks and an innovation typology. *Business Horizons, 63*(2), 147—155.

Paulus, R., Xiong, C., & Socher, R. (2017). A deep reinforced model for abstractive summarization. *Cornell University*. Available at https://arxiv.org/pdf/1705.04304.pdf

PwC. (2018). *Rapid shift to a digital credit risk alert system*. Available at https://www.pwc.nl/en/topics/about/articles-december-2018/rapid-shift-to-a-digital-credit-risk-alert-system.html

Raina, R., Madhavan, A., & Ng, A. (2009). Large-scale deep unsupervised learning using graphics processors. In L. Bottou (Ed.), *Proceedings of the 26th Annual international Conference on machine learning* (pp. 873—880). Montreal, Canada: Association for Computing Machinery.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *KDD*. Available at https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf

Rizzi, W., Wang, Z. M., & Zielinski, K. (2018, September). How machine learning can improve pricing performance. *McKinsey & Company*. Available at https://www.mckinsey.com/industries/financial-services/our-insights/how-machine-learning-can-improve-pricing-performance

Sainath, T., Mohamed, A.-R., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. In *Proceedings of the 2013 IEEE international conference on Acoustics, speech, and signal processing* (pp. 8614—8618). Vancouver, Canada: IEEE.

Sarkar, S., Weyde, T., d'Avila Garcez, A., Slabaugh, G., Dragicevic, S., & Percy, C. (2016). *Accuracy and interpretability trade-offs in machine learning applied to safer gambling*. Available at https://pdfs.semanticscholar.org/8f67/7a206e647c1a0bddf2d54175801633d5a5ec.pdf

SAS. (2011). *After market service: 2011 auto industry leadership forum*. Available at http://www.sas.com/images/landingpage/docs/AftermarketService-AutoForum.pdf

State Street. (2018). *State Street Verus: FAQ*. Available at http://www.statestreet.com/content/dam/statestreet/documents/SSGX/18-33087_Verus_FAQ.pdf

Stedham, Y., Yamamura, J. H., & Beekunn, R. I. (2007). Gender differences in business ethics: Justice and relativist perspectives. *Business Ethics: A European Review, 16*(2), 163—174.

Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology, 9*(1), 49—62.

Valdes, G., Luna, J. M., Eric Eaton, E., Simone, C. B., II, Ungar, L. H., & Solberg, T. D. (2016). MediBoost: A patient stratification tool for interpretable decision making in the era of precision medicine. *Nature*. Available at https://www.nature.com/articles/srep37854

Woolf, A., & McIntyre, A. (2018, April 4). Realizing the full value of AI in banking. *Accenture. Available at https://www.accenture.com/us-en/insights/banking/future-workforce-banking-survey*

Woyke, E. (2017). General Electric builds an AI workforce. *Technology Review*. Available at https://www.technologyreview.com/s/607962/general-electric-builds-an-ai-workforce/

Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823—832.