+Model
RETAIL-786; No. of Pages 18
ARTICLE IN PRESS

ELSEVIER

Journal of Retailing xxx (xxx, xxxx) xxx–xxx

Journal of Retailing

The role of machine learning analytics and metrics in retailing research

Xin (Shane) Wang [a,*], Jun Hyun (Joseph) Ryoo [b], Neil Bendle [c], Praveen K. Kopalle [d]

[a] Ivey Business School, Western University, 1255 Western Road London, Ontario N6G 0N1, Canada
[b] College of Business, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong
[c] Terry College of Business, University of Georgia, 630 S. Lumpkin St., Athens, GA, United States
[d] Tuck School of Business at Dartmouth, Dartmouth College Hanover, NH 03755, United States

Abstract

This research presents the use of machine learning analytics and metrics in the retailing context. We first discuss what is machine learning and explain the field's origins. We then demonstrate the strengths of machine learning methods using an online retailing dataset, noting key areas of divergence from the traditional explanatory approach to data analysis. We then provide a review of the current state of machine learning in top-level retailing and marketing research, integrating ideas for future research and showcasing potential applications for practitioners. We propose that the explanatory and machine learning approaches need not be mutually exclusive. Particularly, we discuss four key areas in the general scientific research process that can benefit from machine learning: data exploration/theory building, variable creation, estimation, and predicting an outcome metric. Due to the customer-facing nature of retailing, we anticipate several challenges researchers and practitioners might face in the adoption and implementation of machine learning, such as ethical prediction and customer privacy issues. Overall, our belief is that machine learning can enhance customer experience and, accordingly, we advance opportunities for future research.

Keywords: Machine learning; Prediction; Metrics; Analytics; Retailing; Trends

Introduction

In the late 1950s, the field of machine learning (ML) emerged from its parent fields of artificial intelligence (AI) and cognitive science. (For a brief history of AI, see Haenlein and Kaplan 2019). Early research largely attempted to replicate human thinking, but the influence of pattern recognition and statistics during the 1990s led to decreased interest in mapping human thought. Recent developments have centered on prediction—using historical data to better predict unobserved outcomes. With increasingly large volumes of data generated by both businesses and consumers, ML is projected to have a profound impact on retail.

ML is becoming increasing important to retailing. Some successful retail adopters include Walmart, which uses ML to group similar products from different merchants based on product features, images, and descriptions. The comprehensive online catalogue addresses the lack of universal identifiers for products in an efficient and scalable manner. Analyzing over 35 million products, Walmart's predictive model demonstrated an error rate below 0.01% (WalMart Labs 2017). At IKEA between 10% and 15% of returned items were discarded as waste but ML lowered this wastage by predicting where the returned items should be restocked (Council 2020).

These examples illustrate how ML can benefit retailers; however, marketplace adoption of ML remains nascent (Wladawsky-Berger 2018). Stéphane Bérubé, the chief marketing officer (CMO) of L'Oréal in Western Europe, suggests that rather than the technology being the largest hurdle, "the tough part [of AI] is finding the purpose" (Ives 2018). Industry reports suggest retailers have yet to fully realize their returns from investments in analytics (McKinsey Global Institute 2017). Indeed, the MSI Research Priorities describe a tier-one priority under *approaches to ingesting and analyzing data to drive marketing insights* as: "What are the current best practices in machine learning and large data to inform marketing decision making?" (Marketing Science Institute 2018).

* Corresponding author.
E-mail addresses: xwang@ivey.uwo.ca (X. Wang), jryoo.phd@ivey.ca (J.H. Ryoo), neil.bendle@uga.edu (N. Bendle), kopalle@dartmouth.edu (P.K. Kopalle).

https://doi.org/10.1016/j.jretai.2020.12.001
0022-4359/

Please cite this article as: Wang, Xin (Shane), et al, The role of machine learning analytics and metrics in retailing research, *Journal of Retailing*, https://doi.org/10.1016/j.jretai.2020.12.001

In response, we provide a conceptual overview of ML and demonstrate its valuable strengths in data analysis. We outline the current and future impact of ML on retailing with an emphasis on metrics and analytics, thus complementing prior work on the changing face of retail management (Shankar 2018). We propose that ML can be transformative for both retailing researchers and practitioners, providing new opportunities to improve customer experience and understand shopping behavior by changing how science is conducted. Retailing is not unique in facing a shifting research frontier; ML has already transformed fields such as computer science, physics, and medical science, and is currently making inroads in business. Yet, there will not be a simple mapping; ML will impact different fields in different ways. Particularly, retailers will have to confront contentious societal issues, such as ethical prediction and customer privacy concerns, that may complicate the adoption and implementation of ML.

This paper is organized as follows. First, we clarify terminologies related to ML and discuss important concepts that characterize ML methods. We then demonstrate ML's practical strengths using an online retailing dataset and provide a conceptual contrast with the traditional explanatory approach to data analysis. We next provide a review of existing ML research in retailing and the wider marketing literature, and outline how ML can impact the way research is conducted. We discuss implications, concluding that retailing research can benefit from using ML. Finally, we advance research questions related to analytics, metrics and retail organizations, as well as potential challenges and opportunities awaiting academics and practitioners.

Table 1
Independent variables used in the prediction of tablet prices on Amazon.

	Variable	Description
Product-related	RAM	Memory of the tablet in GB.
	GHz	Processing speed of the tablet in GHz.
	Storage	Storage capacity of the tablet in GB.
	Battery life	Battery life of the tablet in hours.
	Display	Display size of the tablet in inches.
	Age	Number of weeks observed on Amazon.
Marketing-related	Sales rank	Rank of the tablet, which is a function of sales on Amazon.
	Num new	Number of 'new condition' of tablet model available for sale
	Num used	Number of 'used condition' of tablet model available for sale.
	Num refurb	Number of 'refurbished condition' tablet model available for sale.
Review-related	Num reviews	Total number of reviews the tablet has garnered on Amazon.
	Avg rating	Average review rating of the tablet garnered in the week.
	Avg word	Average word count of the reviews posted in the week.
	Helpful ratio	Ratio of 'helpful' votes to total votes (i.e., both 'helpful' and 'not helpful') garnered by the reviews posted in the week.

Machine learning

Defining the terms

Achievements in ML and AI have been impressive (see Fig. 1 for a timeline), despite some setbacks (e.g., in the 1970s, the UK cut funding suggesting that AI would never reach the level of common sense reasoning; Haenlein and Kaplan 2019). ML's recent success and popularity has, however, meant that terms enter the zeitgeist and get used excessively and loosely. To clarify these terms, we first conceptualize ML as a subset of AI as noted below.

Artificial intelligence (AI) focuses on "intelligent agents," or machines that perceive their environment and perform tasks that human intelligence can tackle (Poole, Mackworth, and Goebel 1998). AI is "defined as a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Haenlein and Kaplan 2019). AI branches out into closely related fields of computer vision, natural language processing, speech recognition, and machine learning.

Machine learning (ML) (Samuel 1959) was first used to refer to a process whereby heuristics pruned possibilities in a checkers program. (Such heuristic processes never guarantee that the best approach is found; a heuristic that optimizes locally is computationally efficient but need not find the global optimum). "Machine Learning" is now the field of scientific study that concentrates on induction algorithms and on other algorithms that can be said to "learn" (Kohavi and Provost 1998). ML consists of "methods or algorithms designed to learn the underlying patterns in the data and make predictions based on these patterns" (Dzyabura and Yoganarasimhan 2018). Specifically, learning is through a *training experience* with respect to some *task*, and success is judged using some *performance metric*. The rapid emergence of ML has led to calls for human insight and interpretability (Ma and Sun 2020).

We illustrate these concepts in the retailing context by examining the substantive question: "can researchers effectively predict the prices of products carried by online retailers?" Doing so can provide retail managers suggestions for new products, and assist in pricing and promotional decisions. We use the Amazon tablet computer dataset made available by Wang, Mai, and Chiang (2014). We match product attributes, market price dynamics, and customer reviews to obtain 2,376 observations for 438 tablet computers at the weekly level. Our variables of interest are in Table 1.

The ML *task* is price prediction, our *substantive metric*. The *training experience* is our data, historical prices, along with product, marketing, and review-related variables. We focus on two *performance metrics*: r-squared and residual mean squared error (RMSE). We use five-fold cross validation, randomly splitting our data using four-fifths for training and the remainder as a hold-out sample. This process is repeated five times and the performance metrics averaged accordingly.

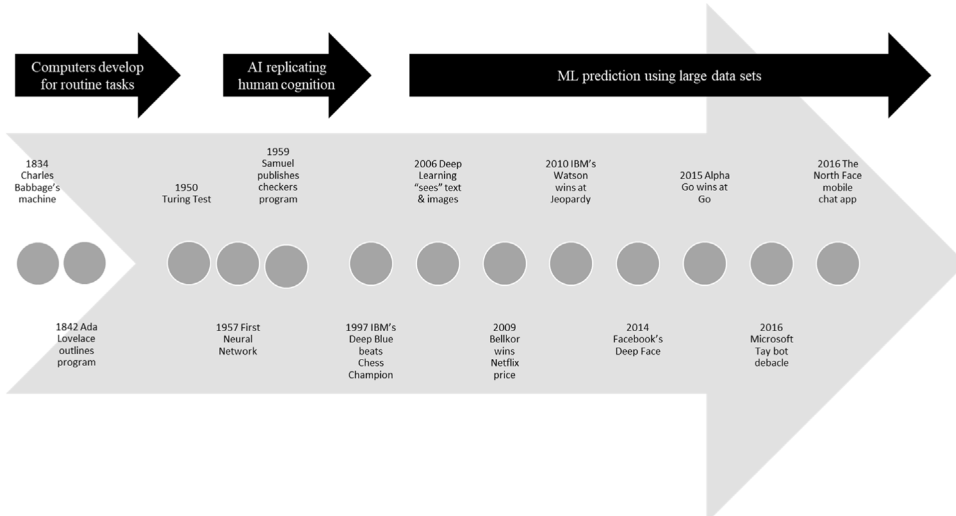

Fig. 1. History of machine learning.

Table 2
Model performances.

	Parametric	ML Non-parametric approaches		
	OLS	SVM	Regression trees	Bagged trees
R-squared	.76 (4)	.90 (3)	.91 (2)	.94 (2)
Residual mean squared error	.356 (4)	.231 (3)	.217 (2)	.177 (2)

We compare the predictive performances of four methods: ordinary least squares (OLS), support vector machine (SVM), regression trees, and bagged trees. OLS is a linear, parametric model traditionally used for testing relationships between variables. This is a benchmark to judge the performance of three non-linear and non-parametric ML algorithms: SVM, regression trees, and bagged trees. Per Table 2, OLS performs the worst, with an R-squared of .76 and RMSE of .356. SVM, regression trees, and bagged trees perform significantly better than OLS; bagged trees demonstrate the best performance with r-squared of .94 and RMSE of .177. This illustrates ML's strength in capturing non-linear relationships; by foregoing parametric assumptions, ML algorithms can offer significant predictive capabilities discussed in greater depth later.

The broadest categorization of ML methods is between supervised and unsupervised learning (Murphy 2012). Above is an illustration of supervised learning — an algorithm learns the mappings of input-output pairs from a "labeled" dataset to predict the output using inputs from an unlabeled dataset. Unsupervised learning, on the other hand, lacks an outcome of interest. Instead, it is concerned primarily with *knowledge discovery*; to uncover patterns and improve understanding of the data. Unsupervised algorithms often do not have standard performance metrics, but rather hyperparameters (which are set before the process bpings of input-output pairs from a "labeled" dataset to predict the output using inputs from an unlabeled dataset. Unsupervised learning, on the other hand, lacks an outcome of interest. Instead, it is concerned primarily with *knowledge discovery*; to uncover patterns and improve understanding of the data. Unsupervised algorithms often do not have standard performance metrics, but rather hyperparameters (which are set before the process begins). Users can calibrate hyperparameters like a miscroscope to focus in and out of the data. Common tasks unsupervised learning might address include data visualization, clustering, and dimension reduction.

For example, a manager may wish to "listen in" to what customers are discussing in their online reviews, exploring the data to obtain useful insights. Manually reading thousands of customer reviews can be expensive and time consuming. Also, having humans read and judge the reviews may introduce inconsistencies and biases in the coding. Unsupervised algorithms, such as Latent Dirichlet Allocation (LDA), can provide an efficient solution by summarizing text data into a set of human-interpretable topics (Bendle and Wang 2016). We apply LDA to our data of 37,562 tablet reviews and present the representative words of the uncovered eight topics in Fig. 2.

Categories of ML that have recently received attention from marketing researchers are semi-supervised learning, reinforcement learning, and deep learning. Semi-supervised learning combines elements from supervised and unsupervised learning and is often used where labeling the training data is expensive. It might well be the form of ML most used in retail. Semi-supervised algorithms are trained using both labeled and unlabeled data; the smaller set of labeled data is used to map the larger set of unlabeled data to labels. For example, automated attendants direct customer calls using speech recognition, which is often trained using semi-supervised learning. Collecting and labeling audio data can be expensive, so semi-supervised learn-

Fig. 2. Eight topics uncovered from tablet reviews on Amazon.

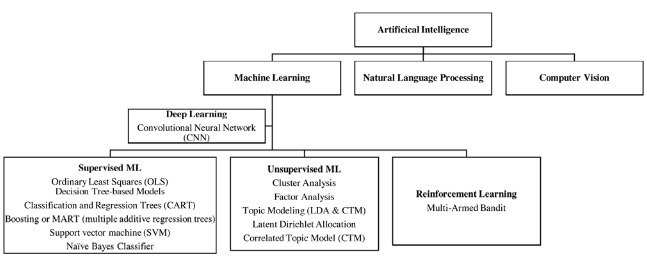

Fig. 3. A selected taxonomy related to machine learning.

ing uses a limited labeled dataset to better understand customers' speech by extrapolating accents, which are unlabeled.

Reinforcement learning involves an "agent" selecting and performing actions, often by trial and error at first, in return for rewards. The agent's objective is to learn the best strategy (called a policy), which is defined as the one that earns the greatest reward over time. An agent learns to balance exploration (testing new actions) and exploitation (gaining known rewards). An example in retailing that can utilize reinforcement learning is customer engagement. By specifying clicks as a reward for an agent that determines which advertisement to display, reinforcement learning can learn to present the optimal advertisements that maximize consumer engagement over time.

Deep learning is a branch of ML that focuses on learning data representations. For example, classifying what specific objects are contained in images was historically a complex task characterized by low accuracy. This was partly because of the numerous ways that one could quantify images (e.g., RGB, hue, saturation, brightness, etc.) but also because of a lack of procedures to meaningfully integrate and succinctly represent such information. Deep learning was a breakthrough that could represent images as low-dimensional vectors using Convolutional Neural Network (CNN), the layers of which equate to different levels of abstraction (LeCun, Bengio, and Hinton 2015). We present a taxonomy related to ML in Fig. 3.

Explanatory versus ML approaches to data analysis

ML can address a wide variety of empirical questions, but proper data collection remains important; the principle of "garbage-in, garbage-out," applies to all research. Poor-quality data will yield poor results regardless of an algorithm's complexity. We note that the methods per se do not define the conceptual approach of ML. For example, OLS is used for statistical inference, but it can also be used to predict a continuous outcome in ML studies. It is the approach used that sets ML apart.

When conducting data analysis we conceptualize an explanatory approach as one focused on the testing of hypotheses, the formulation of which is driven by either intuition or causal theory (Shmueli 2010). For example, a researcher might have a simple hypothesis of a positive relationship between valence, often operationalized using average review ratings (Liu 2006), and product sales. (Sales might also impact ratings but for purpose of illustration we will keep the hypothesis simple). The hypothesized causal relationship can be expressed as:

$$y_i = X_i\beta + \alpha + u_i \tag{1}$$

where i denotes products, y and X the dependent and independent variables, which in our example are sales and average review ratings respectively, and u the idiosyncratic error term. α and β are the parameters. Specifically, α is a constant and β the effect of average ratings on sales, which is a "single-issue causation" interpretation made possible by the linear functional form of Eq. (1). Parametric methods enable the estimation of α and β from data, and probabilistic assumptions allow researchers to test the estimates against the null hypothesis that there is no relationship between valence and sales (i.e., β is statistically insignificant).

Since internal validity, or the extent to which evidence supports the suggested causal relationship, is vital for explanatory analysis, endogeneity always poses an important concern. There are three potential sources of endogeneity: omitted variables, simultaneity, and measurement error (Sande and Ghosh 2018). Parametric models require the assumption that independent variables X and error term u are uncorrelated to retrieve unbiased and consistent estimates of β. If this assumption is violated, the model fails to accurately capture the data generating process (Rossi 2014).

With the proliferation of technology that enables retailers to track product and customers over multiple periods of time, researchers can more easily obtain panel data. For example, let i denote products and t denote weeks. A retailer can collect sales (Y_{it}), as well as average review ratings (X_{it}) over multiple weeks. When modeling the relationship between valence and sales, the researcher can now exploit the additional dimension of data as specified below:

$$y_{it} - \overline{y_i} = \left(X_{it} - \overline{X_i}\right)\beta + (\alpha_{it} - \overline{\alpha_i}) + (u_{it} - \overline{u_i}) \qquad (2)$$

where the bar notation denotes $\overline{X_i} = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, and $\alpha_{it} - \overline{\alpha_i} = 0$. Using this transformation, the researcher can control for time-invariant heterogeneity of i. High quality products are likely to garner higher review ratings online which in Eq. (1) would be captured in the error term because we have not included a measure of quality as an independent variable. This poses an endogeneity concern potentially invalidating the estimate of β as a causal effect. This concern is controlled for in estimating Eq. (2) because product quality does not vary over time.

Addressing endogeneity is thus a major topic of research in econometrics and statistics. e.g., applying vector-autoregression (for a use in retail see Curry et al. 1995). This makes sense as the objectives of the explanatory approach are hypothesis testing related to the relationships between variables (i.e., uncovering β), and addressing threats to internal validity.

What then differs with the ML approach? In many practical applications, retailers are not often interested in estimating β parameters or rejecting null hypotheses. Instead, retailers may want a model that makes a timely informative outcome prediction, not an explanation. The ML approach embraces this perspective; it tolerates more bias in understanding a particular dataset and expects to gain less variance across datasets (i.e., better out-of-sample prediction).

Consider instrumental variable regression, the use of an instrument reduces the model's predictive accuracy of an outcome metric on a testing dataset (Wedel and Kannan 2016), while conventional parametric models require linear functional form assumptions. Any loss of predictive accuracy from using instruments and functional form assumptions is implicitly accepted by those who care about explanation because they primarily want to identify a causal effect. Contrast this with ML which is agnostic to the "true" functional form. Multiple ML algorithms can be trained simultaneously and a "race" allows for selection of the best performing algorithm (e.g., Bagged Trees for our example in Table 2). Both explanation and prediction matter in different situations but retailers will often favor prediction. If the retailer cares more about an accurate prediction they may be unwilling to make the sacrifices academics accept to improve their explanation of causes.

To describe how ML impacts retailing, we will emphasize ML's upsides but we do not mean to minimize the challenges. Later we will discuss organizational issues, retailers are being faced with a fundamentally new way of analyzing data, as are academics. Plus, there are social downsides to ML, where concerns about privacy and discrimination may arise. ML has potential to improve retailing practice, but ML is not a costless boon.

Machine learning in retailing and marketing research

Machine learning's relevance to retail

In this section, we outline the current state of ML in retailing and marketing research. Our objective is not a comprehensive review, but to provide a broad overview of prior research's relevance to retailing hoping to inspire new applications. Focusing on the six journals, *Journal of Retailing, Marketing Science, Journal of Marketing Research, Journal of Marketing, Journal of Consumer Research,* and *International Journal of Research in Marketing*, we searched for ML articles on Google Scholar. We categorized articles on a customer experience journey – Awareness, Consideration, Purchase, Retention, and Advocacy (Coston 2016). Many papers can contribute at several stages but we simplify to show how ML could help solve major challenges for retailers (Table 3).

Awareness: A major challenge is "how can retailers make consumers aware of what they are offering?" Promotional activities using insights from ML can better attract customers. ML research is especially strong in market structure analysis. The abundance of user-generated content (UGC), such as customer reviews on websites and tags on social media platforms, allows retailers to conduct market structure analysis, which is an efficient alternative to the traditional methods of focus groups and surveys. In a prescient article in the *Journal of Retailing*, Aggarwal, Vaidyanathan, and Venkatesh (2009) develop an early approach to understanding brand positioning using Google search data. Lee and Bradlow (2011) extract specific phrases from the text of online customer reviews and conduct cluster analysis. Klostermann et al. (2018) cluster images, and use their associated social tags to study brand positions. Another area relevant to awareness where ML can contribute is advertising. For example, Xiao and Ding (2014) show how different faces and facial features impact consumers' attitudes toward advertisement. Li, Shi, and Wang (2019) use CNN to construct generalizable metrics of visual information in video data. Retailers can apply their metrics to optimize video content when competing for customer attention. Schwartz, Bradlow, and Fader (2017) apply reinforcement learning in a field experiment to select optimal banner advertisements. Such A/B testing is conducted by retailers in real-time without having to halt normal operations.

Consideration: With a plethora of online and bricks and mortars options why should a consumer even consider your store? To gain consideration retailers ideally offer perfectly tailored goods but what does a consumer want? Mining user generated and crowdsourced content on social media can show this (Ghose,

Table 3
Summary of Published Research in Retailing and General Marketing on Machine Learning.

Author(s)	Journal	Research question	Retail/Marketing domain	Machine learning contribution to the problem	Journey phase
Aggarwal, Vaidyanathan, and Venkatesh (2009)	JR	How can we analyze a brand's positioning relative to its competitors using Google brand–descriptor combinations.	Market Structure, Search Terms	Machine-based inferences from cooccurrence of adjectives and nouns.	Awareness
Lee and Bradlow (2011)	JMR	Can market structure be validly analyzed and visualized using customer reviews – User-Generated Content (UGC)?	Market Structure, UGC, Online WOM	K-means clustering to group extracted phrases from customer reviews.	Awareness
Li, Shi, and Wang (2019)	IJRM	How can we extract and measure meaningful information from videos?	Video Content, Visual Marketing	Uses a convolutional neural network to identify the content of frames sampled from videos.	Awareness
Schwartz, Bradlow, and Fader (2017)	MKSC	How can managers find the best online banner advertisement in field in real-time (i.e. "Earning while Learning")?	Advertising	Thomson sampling as a solution to the presented multi-armed bandit problem.	Awareness
Xiao and Ding (2014)	MKSC	How do faces in print advertisements affect consumers?	Visual Marketing, Advertising	CART to examine heterogeneity in response to facial features.	Awareness
Klostermann et al. (2018)	IJRM	How can various brand information and their associative networks be extracted from social media platforms?	Market Structure, UGC, Online WOM	Cluster Analysis (to group images), and Google Cloud Vision API (to tag images from Instagram).	Awareness
Schwartz, Bradlow, and Fader (2014)	MKSC	How can managers determine (in a data-driven way) which model to use given a certain dataset?	Prediction/Forecasting	Classification & Regression Trees test model performance, and Random Forest to test the classification tree itself.	Awareness
Chen, Iyengar, and Iyengar (2017)	MKSC	Can a multimodal continuous heterogeneity (MCH) distribution be modeled in a conjoint setting, given across- and within-segment heterogeneity? How should individual-level partworths be recovered?	Conjoint Analysis	Cross-validation used to determine the amount of shrinkage to recover individual-level partworths.	Consideration
Chen, Nelson, and Hsu (2015)	JMR	Are mental processes related to brands' personality traits connected with specific brain areas?	Neuroimaging	Cross-validation to predict neural response and brain activity in fMRI images.	Consideration
Cui and Curry (2005)	MKSC	How is prediction useful for marketing researchers, and what can SVM contribute if that is the case?	Prediction/Forecasting	Performance of SVM and its kernel method in non-linear and complex settings, such as non-compensatory structures.	Consideration
Dzyabura and Hauser (2011)	MKSC	Can an active-ML algorithm (adaptive question selection) be developed accounting for errors in non-compensatory decision heuristics?	Conjoint Analysis	Errors (from incorrect answers) in an adaptive conjoint analysis setting are framed as tuning parameters, which the algorithm learns after data collection.	Consideration
Evgeniou, Boussios, and Zacharia (2005)	MKSC	Can preference modeling be done without assuming a probabilistic model of the data?	Conjoint Analysis	Preference modeling as an optimization problem, extending optimization mechanics of SVM.	Consideration
Evgeniou, Pontil, and Toubia (2007)	MKSC	How can we better account for consumer heterogeneity in a conjoint analysis setting?	Conjoint Analysis	Data-driven ML cross-validation endogenously models heterogeneity and outperforms hierarchical Bayes (which uses priors for heterogeneity).	Consideration
Ghose, Ipeirotis, and Li (2012)	MKSC	How can a hotel ranking system incorporate the economic value of different locations and service-based characteristics of hotels?	Demand Modeling	SVM to classify images of cities to categories.	Consideration

Table 3 (*Continued*)

Author(s)	Journal	Research question	Retail/Marketing domain	Machine learning contribution to the problem	Journey phase
Hauser et al. (2010)	JMR	Can a method be developed for non-compensatory decision rules (in conjoint) account for many possible disjunctions-of-conjunctions rules?	Conjoint Analysis	Conceptual link between *cognitive simplicity* and *complexity control* (i.e. modeling and controlling for cognitive simplicity enhances prediction). Cross-validation determines exogenous parameters.	Consideration
Huang and Luo (2016)	MKSC	Can a computationally efficient method of preference elicitation for complex products (that can account for errors on the fly be developed?	Conjoint Analysis	Fuzzy SVM that can assign different weights to data points in customizing questions to refine customer-specific preference estimate.	Consideration
Timoshenko and Hauser (2019)	MKSC	Can customer reviews be used to identify customer needs in a more efficient way (relative to traditional methods such as interviews) to develop new products?	Preferences, Product Development	Supervised CNN used to extract sentences that contain customer needs from those that do not.	Consideration
Toubia, Iyengar, Bunnell, and Lemaire (2019)	JMR	Can psychological themes of entertainment products (movies) be extracted as features, to serve as input to predictive models for practitioners?	Feature Extraction, Prediction	Extends LDA topic modeling to create a supervised, nested algorithm that takes seed words (from experts and human coders) and generates topics from movie synopses that as inputs, improve model prediction.	Consideration
Trusov, Ma, and Jamal (2016)	MKSC	Can a method that predicts user profiles (interests and preferences) from online activity data (e.g. visitation intensity and dynamics) be developed which is efficient and scalable?	Preferences, Consumer Profiling	Topics that are generated from a correlated topic model reflect “roles” that consumers play while visiting various websites.	Consideration
Ansari, Li, and Zhang (2018)	MKSC	Can a model be developed to improve recommendations for customers, leveraging numerical ratings, texts, and product attributes?	Recommendation	Extends LDA as a supervised model using product features, user ratings, and movie tags to represent movies as latent topics. Develops a variational Bayesian framework for fast and scalable estimation.	Purchase
Jacobs, Donkers, and Fok (2016)	MKSC	Can a scalable model-based approach using customer characteristics provide accurate predictions of customer purchases in the context of large product assortments?	Feature Extraction, Prediction	An extension of LDA, a topic modeling algorithm, to model words as product purchases, a document as a customer’s purchase history, and a topic a certain preference for products in the assortment.	Purchase
Liu and Toubia (2018)	MKSC	Can a method be developed to uncover consumer preferences from the topics of search queries and webpages (for search engine optimization)?	Preferences, Consumer Profiling	A dual-LDA framework to extract a common set of topics from webpages and search queries.	Purchase
Liu, Singh, and Srinivasan (2016)	MKSC	Can TV ratings be predicted using user-generated content?	Big Data, UGC, Online WOM	Singular value decomposition for dimension reduction of data, LDA used to analyze content on Amazon Cloud computing.	Purchase

Table 3 (*Continued*)

Author(s)	Journal	Research question	Retail/Marketing domain	Machine learning contribution to the problem	Journey phase
Lu, Xiao, and Ding (2016)	MKSC	Can the service process of providing customer recommendations in-store be automated?	Recommendation, Video Content	Pre-trained classifier and collaborative filtering on real-time, in-store video data to calculate a “preference score”.	Purchase
Marchand and Marx (2020)	JR	Can an algorithm provide accurate recommendations and explanations of the reasoning for recommendations?	Recommendation	Combines content-based and collaborative filtering to two datasets with more than 100 million product ratings.	Purchase
Ascarza (2018)	JMR	Are people predicted to churn (using classification models) the most sensitive to firm intervention?	Churn	Random forest to estimate which individuals respond more favorably to intervention.	Retention
Bloemer et al. (2003)	IJRM	Can managers identify segments of customers most likely to churn?	Churn	Classification Trees, and Classification and Regression Trees.	Retention
Currim, Meyer, and Le (1988)	JMR	Can individual-level heterogeneity be estimated in customer choice models without hierarchical assumptions?	Choice Modeling	Decision trees used to build a system of disaggregate models.	Retention
Lemmens and Croux (2006)	JMR	Can we improve prediction of customer churn over binary logit model?	Churn	Bagging and boosting to improve the performance of the predictive model.	Retention
Neslin et al. (2006)	JMR	Which predictive model predicts churn the best?	Churn	Comparison of various classifiers in ‘tournament’ that is open to both academics and practitioners.	Retention
Netzer, Lemaire, and Herzenstein (2019)	JMR	Can words written by borrowers on a peer-to-peer lending platform increase the accuracy of predicting loan default?	UGC, Financial Lending	LDA visualizes and extract topics in loan requests. An ensemble of supervised algorithms test predictive accuracy of models with/without text variables.	Retention
Hartmann et al. (2018)	IJRM	Which methods perform the best in analyzing sentiment and content of unstructured text?	User-generated Content	An ensemble of classifiers (e.g. SVM, Naïve Bayes, and Random Forest) is used to classify sentiment and content.	Advocacy
Homburg, Ehm and Artz (2015)	JMR	How does firm engagement with consumers online affect consumer sentiment?	UGC, Online WOM	SVM used for sentiment analysis of user-generated content.	Advocacy
Ordenes et al. (2019)	JCR	What are the characteristics of content on social media platforms (Twitter and Facebook) and those that are most likely to be shared?	Online WOM, UGC	SVM to classify brand-generated messages as either assertive, expressive, or directive.	Advocacy
Tirunillai and Tellis (2012)	MKSC	Does user-generated content affect stock performance?	Marketing-Finance	Naïve Bayes and SVM to classify valence in text.	Advocacy
Tirunillai and Tellis (2014)	JMR	Can a method be developed to extract dimensions of customer satisfaction from customer reviews?	Market Structure, UGC, Online WOM	Extend the LDA framework to also capture dimensions of products, along with their corresponding valence from customer reviews.	Advocacy
Tirunillai and Tellis (2017)	MKSC	Do TV ads affect online chatter?	Advertising, Online WOM, UGC	SVM and Naïve Bayes to classify valence of reviews.	Advocacy
Zhang and Godes (2018)	MKSC	Do social ties (strong and weak) impact purchase decision quality?	Online WOM, UGC, Social Network	Classify whether the text in reviews are either informative or non-informative.	Advocacy

Legend: JR, Journal of Retailing; MKSC, Marketing Science; JMR, Journal of Marketing Research; JM, Journal of Marketing; JCR, Journal of Consumer Research; IJRM, International Journal of Research in Marketing.

Ipeirotis, and Li 2012). Chen, Nelson, and Hsu (2015) combine ML techniques with neuroimaging data to predict what brand people think about. Timoshenko and Hauser (2019) use deep learning to identify emerging customer needs from reviews. Topic modelling can also help us understand customer needs (Trusov, Ma, and Jamal 2016).

Conjoint analysis, can help retailers accurately measure customer preferences, potentially informing the selection of suppliers and manufacturers, or helping retailers develop their own private-label. Conjoint has seen a stream of ML research (Chen et al. 2017; Dzyabura and Hauser 2011; Evgeniou, Boussios, and Zacharia 2005; Evgeniou, Pontil, and Toubia 2007; Hauser et al. 2010; Huang and Luo 2016). The ability to cope with non-linear relationships is an important strength of ML (Cui and Curry 2005).

Purchase: A primary focus of ML research has been on the purchase stage, which includes activities such as choice, ordering, and payment. Securing a purchase is an obvious challenge in retail but this can be eased by improving the shopping experience with ML. Retailers can use virtual reality, showrooming, or photobooths to create a dynamic, stimulating experience for shoppers. ML can even identify what is selling in real-time to help keep popular SKUs in inventory. For example, Liu, Singh, and Srinivasan (2016) combine cloud computing, ML, and text mining to illustrate how online platform content may be used for more effective forecasting. The impact of ML at the purchase stage is most notable for online retailing, where customers can receive recommendations or discounts based on items in their baskets. While identifying missing, or complementary items, can increase retail revenues.

Studies of preferences can leverage ML's predictive capabilities. Liu and Toubia (2018) use an unsupervised algorithm to measure consumer preference from queries and webpages. Lu, Xiao, and Ding (2016) use in-store video to recommend garments in real time. Recommendations being a great strength of ML (Ansari, Li, and Zhang 2018). Marchand and Marx (2020) even provide reasons for the recommendations which is vital for human-interpretable insights.

Retention: ML can help at the post-purchase stage covering such interactions as item returns and service requests, mostly retention but might include other post-purchase customer analysis. (For ease of language we use the term retention for all post-purchase analysis). Delivering effective service, which includes additional product recommendations, can use ML driven insights.

Retailers often use loyalty programs to reward customers and provide sizeable discounts. Knowing which customers are vulnerable to ending their relationship, churning, can help retailers allocate retention resources (Bloemer et al. 2003; Lemmens and Croux 2006; Neslin et al. 2006). Supervised ML algorithms allows practitioners to efficiently identify customers who are most likely to churn using their observed characteristics and behaviour as a training dataset. We do caution that moving from the relationships between covariates to explanations of consumer behavior may not be straightforward (Ascarza 2018) and one should always be careful to understand a model's limitations (Currim, Meyer, and Le 1988).This highlights the importance of causal theory in guiding data analysis. Thus, the application of ML should augment, not replace, an explanatory approach to data analysis.

Advocacy: Retailers need to better understand, and ideally influence, what is being said about them. How they can do this is another major retailing challenge, this time at the Advocacy stage. Tirunillai and Tellis (2014) use a supervised extension of LDA with customer reviews and have other work investigating chatter (Tirunillai and Tellis 2012, 2017). Homburg, Ehm, and Artz (2015) develop a community matched measure of consumer sentiment by analyzing consumer posts from online forums. Zhang and Godes (2018) analyze the impact of online opinions on consumers' decision quality. Finally, Ordenes et al. (2019) text mine Facebook posts and Twitter tweets to demonstrate the impact of message intentions on message sharing.

In summary, ML can benefit both retailing researchers and practitioners. Next, we outline four key areas in the research process where we expect the greatest impact from ML: *data exploration/theory building, variable creation, estimation,* and *predicting an outcome metric.*

Machine learning in the research process

Data Exploration and Theory Building: Retailing research, similar to other scientific disciplines, often starts with the development of hypotheses by observing the real world (Mjolsness and DeCoste 2001). Thought-provoking and counter-intuitive phenomena give rise to managerially relevant research questions (Lynch et al. 2012). However, technological advances, and the subsequent rise of big data, have increased what we can observe. Bradlow et al. (2017) categorize modern sources of data for retailers into three main groups: traditional enterprise systems, customer-level data, and location-based data. Traditional enterprise data arise from UPC scanners and inventory data. Customer-level data arise from digital contact between the customer and retailer, such as loyalty programs, bonus cards, and emailing lists. Location-based data include customers' geo-location and navigation data enabled by RFID chips and smartphone apps. Retailers are confronted with what seems like limitless inflows of data from each of these sources. For example, if a retailer only had data on a customer's zip code, then segmentation would be imperfect but relatively easy. But when a retailer also has knowledge of the customer's musical tastes, holiday pictures, Facebook likes, Twitter feed, and a multitude of other information, this can leave the decision-maker overwhelmed.

The sheer volume and variety of observations can obscure what is truly deserving of attention, rendering some retail phenomena impossible for capacity-constrained humans to fully observe. To address this problem, there has been a stream of methodological research that focuses on feature extraction, which exploits the dimension-reduction capabilities of ML. For example, Jacobs, Donkers, and Fok (2016) extend the LDA algorithm in the context of large product assortments, summarizing large data into manageable "topics" that can accurately predict customer purchases. Toubia et al. (2019) also extend the LDA algorithm, extracting features from a dataset of movie synopses that capture psychological constructs to accurately predict movie-watching.

Research applying feature extraction includes Wang et al. (2015), who use LDA to summarize 40 years of the *Journal of Consumer Research*. They uncover 16 topics and visualize topic evolution. Novak and Hoffman (2018) leverage clustering and word2vec, a lingiuistic algorithm using an unsupervised neural network, to visualize massive Internet-of-Things data. Data exploration and visualization can lead to managerially relevant insights. Pilot studies captivating interest and motivating research questions are commonly observed in articles. Given the volume and variety of data retailers collect, we expect continuing development of unsupervised frameworks for data exploration, and applications of unsupervised learning to motivate hypotheses.

Variable Creation: Traditionally, researchers have relied on human coders to measure constructs from unstructured data. Such a process is often expensive and difficult to scale, prompting Wedel and Kannan (2016) to list the transformation of unstructured to structured data as a research prly measuring star ratings of website reviews, a retailer could also predict the text's sentiment; two reviews that rate the same product four stars could potentially evaluate individual attributes differently.

Other applied research has focused on applying ML to understand the impact of *images*. Liu, Dzyabura, and Mizik (2020) collect labeled images from Flickr to train a SVM and CNN to predict brand attributes from images shared on Instagram. Zhang et al. (2019) examine the impact of images on demand for properties on Airbnb.com, using a supervised algorithm to classify whether an image is high or low quality. This procedure of labeling images and predicting an output of interest can be generalized by retailers to examine a variety of questions. For example, retailers can efficiently label an image dataset using crowdsourcing platforms, such as Amazon Mechanical Turk, to predict which product image or packaging is more appealing to customers.

We expect to see an increase in the number of novel metrics for analyzing unstructured data, considerably decreasing the costs for retailers to implement such tools. Indeed, there exists commercial software to perform variable creation. LIWC predicts psychological constructs in text with high accuracy and validity (Humphreys and Wang 2017) providing opportunities to address questions that could not have been studied before due to the costs or expertise required.

Estimation: We note that ML is naturally determininistic, meaning that its predictions do not express any uncertainty. This can be a drawback in some settings because we often desire a likelihood or confidence intervals around our predictions. There is a stream of literature in econometrics and computer science that strives to combine the strengths from both explanatory and ML approaches. For example, Jiang, Zhang, and Cai (2008) extend the SVM classifier to yield consistent point estimates and confidence intervals using cross-validation. Such methods can potentially be helpful when presenting ideas. After predicting that a segment of customers is likely to churn, providing an upper and lower bound to the prediction could help persuade some managers to act and justfiy expenses in launching a retention campaign.

ML has also been incorporated to improve causal effect (Athey and Imbens 2017) where the research bifurcates into average causL has also been incorporated to improve causal effect (Athey and Imbens 2017) where the research bifurcates into average causal effects and heterogeneous causal effects. Average causal effects research involves estimating a treatment effect by comparing across the treated and control groups. As modern datasets can contain a number of covariates that exceed the sample size, difficulties may arise in selecting the approriate control variables. An ML-based method proposed by Susan, Imbens, and Wager (2016) allows for an efficient selection of covariates to configure the treated and control groups comparably. Where there are many instrumental variables, Chernozhukov et al. (2018) propose using ML to predict the optimal instrument, using the endogenous variable as an outcome and the instrument set as predictors. For retailers, this is practically useful for datasets with many ?Scolumns.?? For example, when estimating pricing?s̈ effect on purchase, there may exist many independent variables at the customer level (e.g., purchase history, demographic characterstics, etc.) that a retailer may be unsure of which to control a priori.

Research on heterogeneous causal effects involves estimating the causal effects for various subsets of the population. Consider that a treated group in an experiment can be partitioned into many smaller groups based on a covariate, in which the treatment effect could be different for each group. To address this, Athey and Imbens (2016) develop a framework based on regression trees that splits the dataset into two parts. The first part is used to train regression trees and the second part, which is assumed as exogeneous since it contains held-out data, is used to estimate treatment effects within each of the leaves. This method is relevant to retailers wishing to conduct supplementary analyses after estimating the average causal effect. For example, a retailer could be interested in price sensitivities for customers in different loyalty-point brackets.

Prediction of Outcome Metrics: With a labeled training sample of input-output pairs, supervised learning can learn mappings and predict outcomes in unobserved data. This enables retailers to address a range of tasks often limited only by the practitioner?s̈ creativity and ingenuity. For example, Matz et al. (2019) examine image characteristics, seeking to understand the personal characteristics of individuals liking them. Netzer, Lemaire, and Herzenstein (2019) examine text in loan requests on a peer-to-peer lending platform words demonstrate significant power in predicting defaults. Bradlow et al. (2017) outline potential applications of supervised learning to retail, such as incorporating psychological factors to predict retail pricing.

At a meta-level, ML can be used to evaluate ML algorithms. For example, Schwartz, Bradlow, and Fader (2014) demonstrate how to use ML to select the optimal algorithm for prediction given a specific dataset. Hartmann et al. (2018) compare which predictive algorithm performs the best when analyzing the sentiment of unstructured text. Given the different types of data and the potential non-linear relationships of variables that they might contain, an algorithm that performed the best when undertaking one task can potentially perform poorly in the next. For retailers, insights from these papers can potentially automate the very task of selecting and evaluating the algorithms.

Summary

Retailers have made significant investments in recent years to capture customer data. Coupled with decreasing barriers and costs of using ML, we anticipate touchpoints in each stage of the customer experience, from need recognition in the pre-purchase stage to retention management in the post-purchase phase, to be transformed in a data-driven manner. However, a potential drawback to the ML approach is related to explaining relationships between variables. Because ML cannot attribute its predictive performance to a single variable, explaining customer behavior based on any single variable has potential to mislead. This is where we believe the researcher-practitioner gap will move closer. We have discussed how the research process can benefit from ML and retailing data, and we believe researchers can provide new theories to help make sense of big data and interpret predictions. As retailing is at the frontline of where firms meet consumers, ML also introduces opportunities for further ?Shome-grown?? research that is unique to retailing.

We have so far examined where we are now in terms of ML and retailing. It is against this background that we will suggest new directions for research. We will build on the cautions of earlier scholars while maintaining an overall positive view of the possibilities.

Opportunities and challenges for machine learning analytics in retailing research

Much of basic ML research has been done in computer science. Given the customer facing nature of the retailing area, we envision retailing having a solid potential to advance its research. The increasing prominence of ML methods in industry and academia will provide novel directions and we outline implications for retailing research covering analytics, metrics, and organization.

ML and retail analytics

Here we discuss how ML applications shed light on retail analytics, particularly with respect to the prediction of, and causal linkages to, retail outcomes. Research can help us understand consumer and competitor behavior, and tackle the corresponding ethical challenges.

Prediction and Causation: Predicting an outcome fits well with what might be called an engineering approach to retailing, something that can be invaluable for managers. Yet, those who seek a theoretical focus on a single cause face challenges using ML. An algorithm might note that higher frequency mailings from a catalog retailer to its customers tend to presage churn, but this should not be interpreted as suggesting that higher frequency mailings cause churn. (The firm likely targets those thought to be at risk of churning). Understanding the business is vital to interpret ML findings, and it is important to not conflate predictions with causal studies.

Despite the challenges in establishing causation, ML techniques can move in a causal direction (Athey 2018). It may help to remember that single issue causation is a concept employed to make life manageable; it is not descriptive of the world. One can question whether any consumer buys at Best Buy solely because of any single offer. The consumer must also want electronics, value service, etc. Events have multiple ?Scauses and becauses.?? A benefit of ML is that it can review complex causal interactions leaving an understanding with caveats and forks in the road. Rather than seek boundary effects for a single cause, we look for a region where one relationship holds, and another region where a different relationship holds. ML?s̈ contextualized output is not necessarily betterit rarely permits universal assertionsbut it allows more nuance, leaving room for researchers to help in deploying and explaining ML. It would be especially interesting to see whether the benefits accrue disproportionately to certain firms. For example, do online retailers benefit more than brick- and-mortar retailers? An important future research direction is:

RQ1: How can retailers gain, and which retailers will gain most, from a more nuanced multi-causal view in ML retail analytics?

Demonstrating Success: Given the size of modern datasets, ML methods can be resource-intensive with algorithms taking a long time to converge. In addition to methodological research promoting computational efficiency, we expect future work to integrate new technology with ML methods to increase their efficiency for managerial use. For example, LDA is a computationally expensive topic-modeling algorithm, and is therefore difficult to use with large datasets. To bypass this problem, Liu et al. (2016) use Amazon?s̈ cloud computing service to analyze their dataset. Recursive algorithms will help retailers to better show their improvements. New technology, such as quantum computing, will further decrease computational costs. Researchers and managers alike will be able to analyze real world data and focus on phenomena that haven?ẗ been economically viable to tackle before. Each finding may be relatively modest but the proliferation of findings, each generating a small improvement, can create a significant boost in retail effectiveness. This could raise the perceived contribution of marketers (Whitler and Morgan 2017). Thus, we raise the following research issue:

RQ2: How can retail marketers using ML demonstrate the success of the improvements?

Understanding Customer and Competitor Behavior: ML holds the promise of much more detailed analysis of customers?s̈hopping habits. ML can delve into the data available from such innovations as Mwallets, analyzing their ability to cross-sell and up-sell (Kumar, Nim, and Sharma 2019). For example, analysis of paths of consumers around the store (virtual or brick-and-mortar) can only get more sophisticated with greater application of ML. Planograms can be designed given GPS tagged footfalls, and ML predictions can help optimize store layout. Paths taken conditional upon prior actions could give real time predictions for each customer?s̈ journey. The shopper?s̈ experience, and the pricing they see, could be personalized even in a brick-and-mortar environment. Of course, just because they can price discriminate does not mean managers will or even should. Significant challenges include legal and ethical issues. Should managers let algorithms price differently to consumers based upon factors correlated with personal traits? Even if the algorithm avoids sensitive issues, a consumer backlash to personalized pricing could occur if it is based upon hard to define ??unfair?c̈onditions (Guo 2015). ML?s̈ impact will rely not only on the technology but also its application. Thus,

RQ3: What will be the real-world application of ML driven insights, e.g., on planning and pricing?

Knowledge of the consumer, which permits personalization, could also soften competition, given its potential to create a differentiated monopoly. For example, Amazon, because of its data and algorithms, may know more about its customers than they know about themselves people forget, computers do not. Amazon not only competes with other retailers, it competes with third-party sellers on its platform, raising the possibility of leveraging ML analytics to predict competitive behavior. Sikdar, Kadiyali, and Hooker (2020) use a ML model within a product category to predict price changes by Amazon and its third-party competitors. They find that Amazon?s̈ price changes are significantly impacted by the price changes of established third-party sellers, particularly those with higher seller ratings. Future research could examine the following question:

RQ4: How can we leverage ML analytics to predict competitive behavior, for example in pricing decisions?

Ethics and Discrimination: The accuracy demonstrated by ML algorithms predicting consumer characteristics raises the issue of *ethical prediction*. Kosinski, Stillwell, and Graepel (2013) find that sensitive personal traitssuch as intelligence, sexual orientation, use of addictive substances, and political affiliationcan be predicted from Facebook ??likes.?Öpting not to provide personal information may be insufficient to protect a consumer?s̈ privacy from prying retailers. The retailer Target used of ML analytics to predict customers?p̈regnancies raised many ethical concerns about retail data usage and ML predictions (New York Times, February 19, 2012). Researchers can help develop methods for cloaking individuals (Chen, Fraiberger et al., 2017). Ethical use of data in retail can be of interest to firms, consumers, and public policy makers alike. This offers a major opportunity for retail academics to involve themselves in one of the most significant public policy debates of recent times. Which leads to our next question:

RQ5: How can research help navigate what data use is ethically permissible in retail ML analytics?

With respect to discrimination, widespread concerns exist about ML, namely that black box models can lead to discrimination (O?Ñeil 2016). An illustrative case is the launch, and withdrawal less than 24Oh later, of Microsoft?s̈ Tay. This bot interacted on Twitter and ?Slearned?? to be inconsistently bigoted; holding strident pro- and anti-views. Tay predicted Twitter behavior well enough to engage with humans, but who wants machines to learn humanity?s̈ worst qualities?

People discriminated long before they had computers, and limited evidence is often provided to suggest that ML causes increased prejudice. That said, Tay is a salutatory lesson of how a machine can adopt human bias. We do not want to bake society?s̈ evils into algorithms that can be opaque, making it hard to see whether any specific decision is impacted by prejudice. Retail algorithms have a black box quality (Haenlein and Kaplan 2019) but one could say the same of the human mind. Despite work to remove explicit human bias, prejudice exists but can be hidden (Banaji and Greenwald 2013). There may be value in treating ML algorithms similar to other forms of (non-artificial) intelligence. We cannot map exactly how a brain chemical?s̈ release governs an action, but we remain comfortable suggesting causal relationships in psychology. We can delve into what mediates outcomes even absent a perfect understanding of the process.

Recognizing our imperfect understanding is not to suggest that we cannot tackle discrimination related to ML (Williams et al. 2019). Even unsupervised methods are initiated by a human and, ultimately, a human chooses whether, and how, to use the output. Supervised methods involve human beings setting an objective, meaning we can impose societal values. There is no reason why non-prejudicial outputs, e.g., no differences between races in offers, cannot be an objective. To tackle discrimination, managers will need to be sophisticated in their thinking about prejudice. It is not enough to simply say ??our algorithm does not use dimensions typically associated with discrimination?.. Managers must be aware of issues that may correlate with discriminatory outcomes. To take a well-known example, given racial sorting in the U.S., using zip code in a prediction often leads to discriminatory outcomes even if no manager harbors ill-will.

ML may pick up any underlying patterns of discrimination in the real world in the learning process. What if the model?s̈ outputs result in decisions that are systematically biased against people with certain protected characteristics like race, gender, or religion, or even a retail loyalty program status (Stourm et al. 2020) Accordingly, a growing body of research is examining discrimination-aware data mining and fair ML, which attempts to detect and mitigate unfairness (Binns 2018). It is our belief that implementing ML predictions requires human input, and so the ethical responsibility falls not on the algorithm, but on the manager. Binns (2018) discusses ?Sluck based?? variables, or individual traits that are not determined by merit (e.g., race). This stresses the importance of data exploration, to see the extent to which predictions are influenced by these luck-based variables and determine whether to act on such predictions accordingly. Of course, the fact that managers can do things to mitigate discrimination in a world of ML does not mean managers will do so, but prejudice is not an inevitable consequence of algorithm use. It seems an abdication of managerial responsibility to merely say that the machine made me do it. Hence, an interesting future research issue would be to examine the following question:

RQ6: How should ML analytics be managed to mitigate any discriminatory application?

ML and metrics

ML presents ample opportunities for research that improves practical application, theoretical understanding, and the potential for an improvement in the use of metrics relevant to retail. In this section we note some key influences ML might have on retail metric use. There are many metrics that will be changed we will focus on a small number to show some potential impacts.

Market Share: A better understanding of retail market structure and competitor identification can be driven by algorithms that create consumer choice sets. Yet if a market can be partitioned as a myriad of sub markets, even by each individual, what use is overall market share? Should we consider personalized shares of wallet? Rather than a 20% market share, should we report a 20% share of those who bought Hunts for their last ketchup purchase? This leads to aggregation challenges, where customers?̈views of competitors differ. Hence, we ask:

RQ7: How to combine shares of wallet where market boundaries are less than clearly defined?

Sales Prediction: Share of past purchases are not what generally concern retailers. Instead, they are interested in predicting the next purchase. This returns to ML?s̈ strength: should we express, not historic sales or shares, but predictions of future purchases? Yet there will be no obvious paper trail and plenty of choices, by humans and machines, to create such predictions. This leads to the concern that (a) the predicted sales will vary considerably between different creators, and (b) that the numbers may be cherry picked. Hence, an interesting question is,

RQ8: How can managers, observers, even auditors, best use ML sales predictions?

Logistics: Retail is a complex business where context matters to delivering effective results (Kumar, Sunder, and Sharma 2015), which suggests ML could be especially valuable. For example, the most efficient retailer has just enough inventory to cover all purchases, but no more. ML prediction using the vast amount of data and processing power should help optimize inventory levels not just at the store level but across the entire system to ensure inventory is available. (One wonders how ML, with sufficient data access, could have helped predict personal protective equipment for the Covid-19 outbreak). Beyond basing estimates upon adjusted past purchases, ML can consider a plethora of appended data; social media chatter, sentiment analysis, footfall patterns near the store, trends in sales of competitor and complementary products, and numerous plausibly connected data points. Metrics can put a dollar amount on the value of limiting both out-of-stocks and excess inventory. Further, metrics grasping the multi- and omni-channel nature of modern retail will be invaluable. Indeed, attribution modeling should become more sophisticated. Can we create widely accepted measures of the value of showrooming and webrooming? Thus,

RQ9: How can ML help improve retail supply chain logistics and put a dollar value on such improvements?

Customer Lifetime Value (CLV): Use of CLV could be promoted by an increase in the accuracy of churn and retention prediction. ML will assist in CLV being used at an increasingly disaggregated level. (Cohort-based CLV, though often a reasonable response, is always a second best as such approaches overvalue some customers and under value others). Improved prediction has implications for use of customer-based valuation techniques (Gupta and Lehmann 2006). ML offers the opportunity to give greater specificity to the idea of the firm as a collection of customer relationships (plus other assets). Will ML do enough to convince the Financial Accounting Standards Board (FASB) to recognize customer-based assets? CLV is a valuable predictive measure for managerial action, e.g., it informs acquisition and retention decisions, which fits well with ML strengths. Algorithms can create real time valuations for each customer, allowing more sophisticated responses not only to acquisition but also to retention. ML is unlikely to make CLV a usable metric in every situation; using ML, CLV predictions are likely to be easiest for online retailers, given they can record all customer touchpoints. Still, store point-of- sale systems and customer systems can flash messages and offers to re-engage a customer predicted to be drifting away. In essence, ML will allow for greater use of valuable CLV methods.

RQ10: What are the most promising ways to use ML to make CLV a usable metric in every situation?

ML and retail organization

The effect of ML on retail organization is likely to be seen over time. Many of these problems are likely to have a longer time to fruition than changes to analytics, as it may take 10 years or more for organizations to evolve.

The Role of Retail Experience: As the barriers to entry in ML rapidly decrease, we expect an increasing number of retailers will adopt ML, but not all managers will embrace the change. Most obviously, we expect clashes between experienced retail buyersusing their years of experienceand algorithm users over how best to serve customers. Tried and tested practices may be overturned by ML prescribed methods. This is especially problematic as the shelf life of ML results are limited. A short shelf life does not mean the results are uninteresting, indeed we can learn a lot about market dynamics from changing results. Yet, where does this leave the retail buyer who has amassed experience over a long career? Meanwhile, retailers adopting ML techniques may find it increasingly challenging to explain their actions. Who will adopt a superior plan that might still fail and cannot be explained if it does fail? It is a prescription for bad retailing if buyers merely follow the algorithms?̈instructions. ML runs the risk of creating a world where human beings do not understand the rules of what is happening; one might argue this has always been the case, but some rules now will be stored in algorithms, not natural systems. There will likely be considerable effort devoted to understanding how models work by academics, but how will retail managers leverage both their experience and advances in ML models? More extensive theories as to how humans and machines interact to form managerial decisions will be required.

It might be important to make a distinction between automation and ML. The former creates a simple loop: if X happens, then do Y. ML, on the other hand, provides an input that is used by a decision process. While ML advice might often be acted on automatically, it is not necessary to do so. Using ML does not mean that its predictions must be automated; implementing insights from ML often requires managerial input. For instance, ML may simply augment human decision-making. What decision-making processes to automate, and whether managers are willing to accept the responsibility and forgo their agency of the decisions, are important research questions. ML advice may change the role of managers; they become gatekeepers and higher-level decision makers where they combine the art of retailing with the ML based science. Hence,

RQ11: How can retailers best alleviate the inevitable tensions between experienced retail managers and new ML-based management systems?

Managerial Use of Theory: ML algorithms are sometimes likened to a ?Sblack box,?? in which we observe only the input and the output of the algorithm but not its internal steps. This departs from traditional econometric models, in which the researcher can assess the effect size and confidence interval for each parameter. This poses problems for managers persuading others to act on the ML predictions. Surprisingly, this could mean that the burden placed on theory increases. Namely, an underlying theory could provide a gut check on ML predictions. Hence, we ask,

RQ12: Will managers rely more explicitly on theory to justify acting on ML predictions?

Changes in Retail Management: Direct marketing helped spawn online marketing with the rise of the internet. What retailing roles will develop most with the new ML approaches? The sub-fields of consumer insights and market research are already changing, but will store planning become increasingly ML driven? Firm behavior at the aggregate level will also gradually change. As executives increasingly delegate decisions to ML methods, researchers will observe less human, and more machine, decisions in firm-level data (Athey 2018). Thus, we pose,

RQ13: How will ML change retail management and how will this change firm level analysis?

Machine learning and the future of the retailing discipline

This final section examines what the widespread adoption of ML will mean for the academic retailing discipline. We expect these changes to take many years but academics have begun taking advantage of ML now. Some researchers see ML, and AI broadly, as not only providing research opportunities but a fundamentally new method of invention (Cockburn, Henderson, and Stern 2018). We expect a variety of scholars will need to collaborate with retail managers to answer these questions. Indeed, a key benefit of ML is giving new tools to study more managerially relevant questions.

Co-creating with Retailing Practice: The adoption of ML has the potential to better connect industry and academia. How retailers curate offerings to consumers may inform retailing research. A challenge linking academics and managers is that many academic methodse.g., laboratory experiments, and econometric analysis of public datadiffer from those that managers use. More constructive collaboration can occur if both researchers and managers speak the same ML language.

Academic questions could shift from understanding broad industry effects towards the specific challenges facing retailers. That a specific construct is linked with a specific outcome is useful background information, but no retail appeal features a single non-contextualized construct. Even digital A/B tests can only try so many permutations, yet retailers work with multiple interactions. Even with massive data capture, researchers might not be able to fully examine what they seek to understand. We might be interested in how construal level interacts with ad communication length, which interacts with communication media, which is in turn moderated by the prior customer-firm interactions etc. No traditional program of research can hope to document all the relevant main effects, never mind the interactions and interactions of interactions. ML techniques can theoretically consider a multiple of details and even with sufficient data, produce trees highlighting what is beneficial in any given context. We suggest,

RQ14: How can the flexibility of ML?š techniques create contextualized advice for retail action?

Encouraging ML Use: As the necessary skills are available to any who have trained on R, or similar programs, most quantitative researchers will adopt some ML techniques. Pre-packaged tools will even make ML techniques usable by those without quantitative expertise. How can appropriate ML use be encouraged? PhD courses should help develop more familiarity with ML, while linkages to economics or computer science improve technical skills. We would encourage brown bags to leverage the skills of more experienced researchers and relevant organizations, e.g., MSI, ISMS, ACR, AMA Retailing SIG, may be able to assist in best practice. That is,

RQ15: How can universities and umbrella organizations promote ML best practice in retailing?

Bridging the Practitioner Divide: As calls for responsible business research increase (Responsible Research in Business & Management 2018), modern marketers may learn from prior generations, invigorating research in the spirit of decision support systems. If a model is not used it is of little benefit (Little 1970). Research often occurs in a university setting, such as in a laboratory with student subjects, writing surveys for paid panels, or a lone researcher sequestered to interrogate secondary data. Ideally, an additional venue might increasingly be the workplace of practicing retailers. Working closely with managers will allow for testing models on up-to-date data and promote relevance for practice. To promote such collaboration, we would urge further attention be given to aligning the incentives of retail managers and academics. Academic research has often been seen as more important if it posits a near to appear more relevant. As ML generates weaker assertions about underpinning relationships, academics may need to embrace more modest, context-dependent findings. Such assertions, while retaining sophistication not always present in advice from consultants, can be more persuasive to managers than sweeping statements. Thus, we have,

RQ16: How can ML help to bridge the practitioner/academic divide?

Field Experimentation: Tasks best suited to ML are often those with large volumes of data and a myriad of interactions exploiting ML's flexibility in a messy world. Thus, one potential boon from ML is the better description of field experiments when managerial action has many different levels and levers. Retail experiments with randomized treatments and controls are expensive—managers must make numerous offers they do not believe will be effective. This happens, especially in online retailing, but such experimentation is still more limited than one might hope (Gneezy 2017). Field experiments often lack the control of a laboratory investigation, but ML methods can visualize dimensions that are uncontrolled, allowing the researcher to detect patterns and promote the value of field experiments. Do consumers of different ages appear to react similarly or not? Is using a ML technique boosting revenues? Can the possibilities of ML boost field experimentation? Thus,

RQ17: How can ML encourage retail field experimentation by retailers and aid its reporting?

Black Box Best Practice: Supervised ML means one must specify in advance what one considers success; this forces clarity about strategy (Agrawal, Gans, and Goldfarb 2018). Yet ML can search for the model that creates the best prediction, which contrasts with specifying a model a-priori. This sounds like ML gives the researcher more chances of a successful result—a worrying idea in an era of replication crisis (Honig et al. 2018). Given this, academics must include detailed descriptions of coding practices plus clear statements on division of the data for training and testing. The reader should understand the *recall*, i.e., (True Positives)/(True Positives + False Negatives), and *precision*, i.e., (True Positives)/(True Positives + False Positives), of the methods used.

One currently used approach that tries to unravel the black box nature of ML methods, or at least make the results more interpretable, is to simultaneously run a simple and a complex model. For example, Liu et al. (2020) apply both SVM and CNN to predict the same set of images, using the latter for prediction and the former for examining the weights (importance) of the different predictors. We expect the continuing development of *interpretable* machine learning methods to give greater confidence in ML's application. Therefore,

RQ18: What reporting methods give confidence that retail ML avoids inappropriate data mining?

Industry Knowledge: The ability to delve into the data allows for precise questions to be answered. Will a consumer react better to a specific type of display in a specific type of store at a particular time of the year? Or will different front-pages work for various customers at all times of the day? Letting the data speak requires the ability to listen, which may necessitate knowledge of how retail works on a day-to-day level. As ML skills become more abundant, the deficiency in academia may be in understanding the retail landscape. This might see the greater success of scholars with prior work experience, or those able to take sabbaticals to embed in industry. Hence,

RQ19: Will ML techniques help reward those with greater knowledge of specific retail practices?

Consumer Insight: In a world where tech giants (and others) pour vast resources into development, many contributions to advancing ML knowledge for consumer insight will come to academia from businesses allowing academics a translation and diffusion role. Academics can help managers from more traditional companies to understand the benefits of using the advances of e-commerce retailers. Investigating how ML can aid retailers in tier 3 and tier 4 cities, e.g., Scranton (U.S.) or Grimsby (U.K.), may not be as glamorous as advising Silicon Valley, but may benefit society more.

Personalization (products generated by a retailer to fit a consumer) and customization (options chosen by the consumer from a range of options) (Kumar et al. 2019) differ markedly in the active input expected from consumers. Customization requires consumers know themselves, personalization requires that retailer knows the consumers. ML-driven personalization may improve the retailer's understanding to a point where it better delivers what the consumer wants without asking the consumer. Retailing scholars need to outline when this applies and what this means for expressed and unexpressed needs and wants. Thus, we have,

RQ20: What are the implications of ML for understanding consumer preferences?

Manipulability: Will we start observing not natural behavior but attempts to game a machine's learning? If the algorithms identify price-sensitive consumers, consumers have an interest in managing the signals they send. There are already tools designed to beat the algorithms. A 'VPN' might actively manage interactions with retailers on the consumer's behalf. A Red Queen scenario arises; ML experts for firms and consumers each innovate, neither gaining a decisive advantage. What then is a consumer's preference given multiple layers of obfuscation? Social desirability infects survey data, but will this become a greater issue for "observed" behavior? Thus,

RQ21: Will scholars' primitives increasingly diverge from consumers' revealed preferences?

ML Prediction's Shelf life: Machine learning's contextualized advice highlights a fascinating problem. Findings hold only as long as the context holds, but everything constantly changes. Is today's prediction valid for activity next year? What is the shelf-life of predictions? Without universal statements of relationships, we only really know that a relationship held when tested. How do we know the relationship still holds without running the model again? This leads to a fundamental conflict between managerial and academic ML. Managers often seek advice on how to act now. Even if ML results hold only temporarily, it is relatively safe to act on them now. Furthermore, the manager can run the ML algorithm again; refreshing the model regularly makes a limited shelf life less of a problem. How do academics report limited shelf life results? So,

RQ22: How can academics learn to better recognize the value of limited shelf life results?

ML's approach means that we might rethink our research analogies. Currently, academic studies seem like mountain climbing expeditions. Scholars build on prior achievements to gain a little more altitude. ML uses a different logic; algorithms can fly the researcher between peaks. Research becomes not the inspection of a specific mountain top but can involve investigating numerous mountain tops across a range; often jumping

Table 4
Implications of ML for future research in retailing.

Prediction	Future research implication	Examples: ML might alter research
Falling Entry Barriers To ML	More ML applied in retailing. Most quantitative retailers will use ML techniques. Packages allow scholars from all approaches to use techniques.	Research into social networks, and how they influence consumer choice, can employ ML using it to predict both the social networks that will develop and the resulting changes in consumer choices.
Changing Retailing Models	Opportunity to revisit retailing models to better align them with novel findings both in theoretical work and practical application.	For discrete choice modelling ML techniques help identify which regressors serve as good predictors of choice while deriving a structural form of consumers' utility using the adaptive ML process.
Improving Retailing Best Practice	Problems with malpractice somewhat mitigated by algorithms' 'honesty". Clarifies upfront when model mining is occurring.	ML algorithms can help determine which controls are valid and robust in franchise loyalty and growth research. Helps resolve disputes over controls that seem implausible or imply conflicting results.
Change in Venues for Research	Potential for retailing academics to become more connected to practitioners in their workplace.	Firms engaging ML to optimize. Academics' ML implementable models applied giving them access to data generated by real transactions.
Move towards Metrics for Prediction	Develop new metrics for distribution. Planogramming based upon ML prediction. ML assisted Corporate valuation.	More effective prediction of CLV. Potential for research aiding managers understand and communicate more complex multi-casual results.
More Modest but More Useful Research	More modest claims in academic research. More credit given for clear implications in a specific context	ML techniques can quantify and identify photo characteristics. This allows identification of what qualities of a photo are important in the determination of consumer choice in a particular online marketplace, such as eBay or Airbnb.
Bridging the Practitioner Divide	More modest research that applies in specific contexts will help bridge the academic-practitioner divide	ML techniques can help estimate the quantitative effect of the warm glow effect, e.g., fair-trade products. Also predict future consumers' purchases.
Change in Research Stream Prominence	Expect decline in prominence of laboratory experiments; more interest in sub-disciplines using context, e.g., retail strategy, qualitative research.	Firms will likely maintain huge datasets that can be shared with researchers. General marketing researchers can rely less on laboratory experiments/panels.
Merging Retailing Research Streams	Greater incentive/opportunity for retailing scholars of diverse methodological backgrounds to collaborate.	ML can quantify and identify qualitative measures, e.g., media content and its emotional impact. Better connects quantitative and behavioral streams.

to a new knowledge area. This has obvious upsides in generating novel perspectives. We still will not know whether any peak crested is the highest, but the new challenge is that we visit more mountaintops than we can describe. How can researchers decide which are the most important questions to investigate? Table 4 provides a review of some implications of ML for future research in retailing.

Machine learning is already making a significant impact on retailing research and managerial behavior. We expect this to continue and, although challenges will certainly arise, the changes driven by the rise of ML can have significantly positive elements. It is important to understand that ML brings an altered way of thinking to research but that this has potential to solve many of the problems plaguing research, from model mining and an inability to connect with managers, to the fracturing of the discipline. Machine learning in retailing is here to stay, the problems are real, but there are very significant benefits we hope retailing will enjoy. Our belief is that machine learning can further enhance customer experience and retail research. Our contribution has been to raise the challenges researchers and practitioners might face in the adoption and implementation of machine learning and so help ML to fulfil its enormous potential.

Acknowledgements

We thank the Guest Co-Editor, V. Kumar, and two anonymous Journal of Retailing reviewers for their helpful comments and suggestions on an earlier version of the manuscript.

References

Aggarwal, Praveen, Rajiv Vaidyanathan and Alladi Venkatesh (2009), "Using Lexical Semantic Analysis to Derive Online Brand Positions: An Application to Retail Marketing Research," *Journal of Retailing,* 85 (2), 145–58.

Agrawal, Ajay K., Joshua Gans and Avi Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence,* Cambridge, Mass: Harvard Business Review Press.

Ansari, Asim, Yang Li and Jonathan Z. Zhang (2018), "Probabilistic Topic Model for Hybrid Recommender Systems: A Stochastic Variational Bayesian Approach," *Marketing Science,* 37 (6), 987–1008.

Ascarza, Eva (2018), "Retention Futility: Targeting High-Risk Customers Might Be Ineffective," *Journal of Marketing Research,* 55 (1), 80–98.

Athey, Susan (2018), "The Impact of Machine Learning on Economics," in *in The Economics of Artificial Intelligence: An Agenda* University of Chicago Press.

Susan, Athey and Guido Imbens (2016), "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Sciences,* 113 (27), 7353–60.

——— and ——— (2017), "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives,* 31 (2), 3–32.

Susan, Athey, Guido Imbens and Stefan Wager (2016), *Efficient Inference of Average Treatment Effects in High Dimensions via Approximate Residual Balancing.,.*

Banaji, Mahzarin R. and Anthony G. Greenwald (2013), *Blindspot: Hidden Biases of Good People,* New York, NY: Delacorte Press.

Bendle, Neil Thomas and XinShane Wang (2016), "Uncovering The Message From The Mess Of Big Data," *Business Horizons,* 59 (1), 115–24.

Binns, Reuben (2018), "Fairness in Machine Learning: Lessons from Political Philosophy," *Proceedings of Machine Learning Research,* 81, 1–11 (Conference on Fairness, Accountability, and Transparency).

Bloemer, Josée M.M., Tom Brijs, Koen Vanhoof and Gilbert Swinnen (2003), "Comparing Complete and Partial Classification for Identifying Customers at Risk," *International Journal of Research in Marketing,* 20 (2), 117–31.

Bradlow, Eric T., Manish Gangwar, Praveen Kopalle and Sudhir Voleti (2017), "The Role of Big Data and Predictive Analytics in Retailing," *Journal of Retailing,* 93 (1), 79–95.

Chen, Daizhuo, Samuel P. Fraiberger, Robert Moakler and Foster Provost (2017), "Enhancing Transparency and Control When Drawing Data-Driven Inferences about Individuals," *Big Data,* 5 (3), 197–212.

Chen, Yu-Ping, Leif D. Nelson and Ming Hsu (2015), "From 'Where' to 'What': Distributed Representations of Brand Associations in the Human Brain," *Journal of Marketing Research,* 52 (4), 453–66.

Chen, Yupeng, Raghuram Iyengar and Garud Iyengar (2017), "Modeling Multimodal Continuous Heterogeneity in Conjoint Analysis—A Sparse Learning Approach," *Marketing Science,* 36 (1), 140–56.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey and James Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal,* 21 (1), C1–68.

Cockburn, Iain M., Rebecca Henderson and Scott Stern (2018), *The Impact of Artificial Intelligence on Innovation.,.*

Coston, Rob (2016), "Five Stages of Your Customers' Buying Journey," *The Chartered Institute of Marketing Exchange,*, https://exchange.cim.co.uk/blog/five-stages-of-your-customers-buying-journey/

Council, Jared (2020), "IKEA Seeks to Cut Costs, Waste With Smarter Approach to Customer Returns," *Wall Street Journal Pro,*, https://www.wsj.com/articles/ikea-seeks-to-cut-costs-waste-with-smarter-approach-to-customer-returns-11583145000

Cui, Dapeng and David Curry (2005), "Prediction in Marketing Using the Support Vector Machine," *Marketing Science,* 24 (4), 595–615.

Currim, Imran S., Robert J. Meyer and Nhan T. Le (1988), "Disaggregate Tree-Structured Modeling of Consumer Choice Data," *Journal of Marketing Research,* 25 (3), 253–65.

Curry, David J., Suresh Divakar, Sharat K. Mathur and Charles H. Whiteman (1995), "BVAR as a Category Management Tool: An Illustration and Comparison with Alternative Techniques," *Journal of Forecasting,* 14 (3), 181–99.

Dzyabura, Daria and John R. Hauser (2011), "Active Machine Learning for Consideration Heuristics," *Marketing Science,* 30 (5), 801–19.

Dzyabura, Daria and Hema Yoganarasimhan (2018), "Machine Learning and Marketing," in *in Handbook of Marketing Analytics* Edward Elgar Publishing.255–79.

Evgeniou, Theodoros, Constantinos Boussios and Giorgos Zacharia (2005), "Generalized Robust Conjoint Estimation," *Marketing Science,* 24 (3), 415–29.

Evgeniou, Theodoros, Massimiliano Pontil and Olivier Toubia (2007), "A Convex Optimization Approach to Modeling Consumer Heterogeneity in Conjoint Estimation," *Marketing Science,* 26 (6), 805–18.

Ghose, Anindya, Panagiotis G. Ipeirotis and Beibei Li (2012), "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science,* 31 (3), 493–520.

Gneezy, Ayelet (2017), "Field Experimentation in Marketing Research," *Journal of Marketing Research,* 54 (1), 140–3.

Guo, Liam (2015), "Inequity Aversion and Fair Selling," *Journal of Marketing Research,* 52 (1), 77–89.

Gupta, Sunil and Donald R. Lehmann (2006), "Customer Lifetime Value and Firm Valuation," *Journal of Relationship Marketing,* 5 (2–3), 87–110. Taylor & Francis Group. http://resolver.scholarsportal.info/resolve/15332667/v5i2-3/87_clvafv

Haenlein, Michael and Andreas Kaplan (2019), "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *California Management Review,* 61 (4), 5–14.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp and Mark Heitmann (2018), "Comparing Automated Text Classification Methods," *International Journal of Research in Marketing,*,. Online.

Hauser, John R., Olivier Toubia, Theodoros Evgeniou, Rene Befurt and Daria Dzyabura (2010), "Disjunctions of Conjunctions, Cognitive Simplicity, and Consideration Sets," *Journal of Marketing Research,* 47 (3), 485–96.

Homburg, Christian, Laura Ehm and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research,* 52 (5), 629–41.

Honig, Benson, Joseph Lampel, Joel A.C. Baum, Mary Ann Glynn, Runtian Jing, Michael Lounsbury, Elke Schuessler, et al. (2018), "Reflections on Scientific Misconduct in Management: Unfortunate Incidents or a Normative Crisis?," *Academy of Management Perspectives,* 32 (4), 412–42.

Huang, Dongling and Lan Luo (2016), "Consumer Preference Elicitation of Complex Products Using Fuzzy Support Vector Machine Active Learning," *Marketing Science,* 35 (3), 445–64.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2017), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research,* 44 (6), 1274–306.

Ives, Nat (2018), "Marketers Learn to Work With the Machines," *Wall Street Journal,*, https://www.wsj.com/articles/marketers-learn-to-work-with-the-machines-11545079849

Jacobs, Bruno J., Bas Donkers and Dennis Fok (2016), "Model-Based Purchase Predictions for Large Assortments," *Marketing Science,* 35 (3), 389–404.

Jiang, Bo, Xuegong Zhang and Tianxi Cai (2008), "Estimating the Confidence Interval for Prediction Errors of Support Vector Machine Classifiers," *Journal of Machine Learning Research,* 9 (March), 521–40.

Klostermann, Jan, Anja Plumeyer, Daniel Böger and Reinhold Decker (2018), "Extracting Brand Information from Social Networks: Integrating Image, Text, and Social Tagging Data," *International Journal of Research in Marketing,* 35 (4), 538–56.

Kohavi, Ron and Foster Provost (1998), "Special Issue on Applications of Machine Learning and the Knowledge Discovery Process," *Journal of Machine Learning,* 30 (2–3), 271–4.

Kosinski, Michal, David Stillwell and Thore Graepel (2013), "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior," *Proceedings of the National Academy of Sciences,*, 201218772.

Kumar, V., Nandini Nim and Amalesh Sharma (2019), "Driving Growth of Mwallets in Emerging Markets: A Retailer's Perspective," *Journal of the Academy of Marketing Science,* 47 (4), 747–69.

Kumar, V., Bharath Rajan, Rajkumar Venkatesan and Jim Lecinski (2019), "Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing," *California Management Review,* 61 (4), 135–55.

Kumar, V., Sarang Sunder and Amalesh Sharma (2015), "Leveraging Distribution to Maximize Firm Performance in Emerging Markets," *Journal of Retailing,* 91 (4), 627–43.

LeCun, Yann, Yoshua Bengio and Geoffrey Hinton (2015), "Deep Learning," *Nature,* 521 (7553), 436–44.

Lee, T.Y. and E.T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research,* 48 (5), 881–94.

Lemmens, Aurélie and Christophe Croux (2006), "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research,* 43 (2), 276–86.

Li, Xi, Mengze Shi and XinShane Wang (2019), "Video Mining: Measuring Visual Information Using Automatic Methods," *International Journal of Research in Marketing,* 36 (2), http://dx.doi.org/10.1016/j.ijresmar.2019.02.004

Little, John D.C. (1970), "Models and Managers: The Concept of a Decision Calculus," *Management Science,* 16 (8), B466–485.

Liu, Jia and Olivier Toubia (2018), "A Semantic Approach for Estimating Consumer Content Preferences from Online Search Queries," *Marketing Science,* 37 (6), 930–52.

Liu, Liu, Daria Dzyabura and Natalie Mizik (2020), "Visual Listening In: Extracting Brand Image Portrayed on Social Media," *Marketing Science,* 39 (4), 669–86.

Liu, Xiao, Param Vir Singh and Kannan Srinivasan (2016), "A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing," *Marketing Science,* 35 (3), 363–88.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing,* 70 (3), 74–89.

Lu, Shasha, Li Xiao and Min Ding (2016), "A Video-Based Automated Recommender (VAR) System for Garments," *Marketing Science,* 35 (3), 484–510.

Lynch, John G. Jr, Joseph W. Alba, Aradhna Krishna, Vicki G. Morwitz and Zeynep Gürhan-Canli (2012), "Knowledge Creation in Consumer Research: Multiple Routes, Multiple Criteria," *Journal of Consumer Psychology,* 22 (4), 473–85.

Ma, Liye and Baohong Sun (2020), "AI and Machine Learning in Marketing Research: Call for Human Insights and Interpretability," *International Journal of Research in Marketing,*,. Forthcoming.

Marchand, André and Paul Marx (2020), "Automated Product Recommendations with Preference-Based Explanations," *Journal of Retailing,*, http://dx.doi.org/10.1016/j.jretai.2020.01.001. In advance

Marketing Science Institute (2018), *Research Priorities 2018-2020,* Cambridge: Mass. http://www.msi.org/research/2018-2020-research-priorities/

Matz, Sandra C., Cristina Segalin, David Stillwell, Sandrine R. Müller and Maarten W. Bos (2019), "Predicting the personal appeal of marketing images using computational methods," *Journal of Consumer Psychology,* 29, 370–90.

McKinsey Global Institute (2017), *Artificial Intelligence The Next Digital Frontier?,*.

Mjolsness, Eric and Dennis DeCoste (2001), "Machine Learning for Science: State of the Art and Future Prospects," *Science,* 293 (5537), 2051–5.

Murphy, K.P. (2012), *Machine Learning: A Probabilistic Perspective,* Cambridge, Mass: MIT Press.

Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research,* 43 (2), 204–11.

Netzer, Oded, Alain Lemaire and Michal Herzenstein (2019), "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications," *Journal of Marketing Research,*,. Forthcoming.

Novak, Thomas and Donna L. Hoffman (2018), "A Computational Social Science Framework for Learning and Visualizing the Latent Language of Structured IoT Interaction Data," *SSRN Working Paper,*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3278045

O'Neil, Cathy (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* New York, N.Y: Broadway Books.

Poole, David, Alan Mackworth and Randy Goebel (1998), *Computational Intelligence: A Logical Approach,* New York, N.Y: Oxford University Press.

Responsible Research in Business and Management (2018), *Vision 2030,* https://www.rrbm.network/position-paper/vision-2030/

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science,* 33 (5), 655–72.

Samuel, Arthur L. (1959), "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development,* 3 (3), 210–29.

Sande, Jon Bingen and Mrinal Ghosh (2018), "Endogeneity in Survey Research," *International Journal of Research in Marketing,* 35 (2), 185–204.

Schwartz, Eric M., Eric T. Bradlow and Peter S. Fader (2014), "Model Selection Using Database Characteristics: Developing a Classification Tree for Longitudinal Incidence Data," *Marketing Science,* 33 (2), 188–205.

———, ——— and ——— (2017), "Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments," *Marketing Science,* 36 (4), 500–22.

Shankar, Venkatesh (2018), "How Artificial Intelligence (AI) Is Reshaping Retailing," *Journal of Retailing,* 94 (4), vi–xi.

Shmueli, Galit (2010), "To Explain or to Predict?," *Statistical Science,* 25 (3), 289–310.

Sikdar, Sharmistha, Vrinda Kadiyali and Giles Hooker (2020), "Price Dynamics on Amazon Marketplace: A Multivariate Random Forest Variable Selection Approach," *SSRN,*, http://dx.doi.org/10.2139/ssrn.3518690

Stourm, Valeria, Scott Neslin, Eric Bradlow, Els Breugelmans, Yeon Chun So, Pedro Gardete, P.K. Kannan, et al. (2020), "Refining Loyalty Programs in the Era of Big Data: A Societal Lens Paradigm," *Marketing Letters,*,. In press.

Timoshenko, Artem and John R. Hauser (2019), "Identifying Customer Needs from User-Generated Content," *Marketing Science,* 38 (1), 1–20.

Tirunillai, Seshadri and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science,* 31 (2), 198–215.

——— and ——— (2014), "Mining Marketing Meaning from Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research,* 51 (4), 463–79.

——— and ——— (2017), "Does Offline TV Advertising Affect Online Chatter? Quasi-Experimental Analysis Using Synthetic Control," *Marketing Science,* 36 (6), 862–78.

Toubia, Olivier, Garud Iyengar, Renee Bunnell and Alain Lemaire (2019), "Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption," *Journal of Marketing Research,* 56 (1), 18–36.

Trusov, Michael, Liye Ma and Zainab Jamal (2016), "Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting," *Marketing Science,* 35 (3), 405–26.

Ordenes, Francisco Villarroel, Dhruv Grewal, Stephan Ludwig, De Ruyter Ko, Dominik Mahr and Martin Wetzels (2019), "Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research,* 45 (5), 988–1012.

Wal Mart Labs (2017), "Catalog-Matching," *Merchant Technology,*, https://www.walmartlabs.com/case-studies/catalog-matching

Wang, Xin (Shane), Neil Thomas Bendle, Feng Mai and June Cotte (2015), "The Journal of Consumer Research at Forty: A Historical Analysis," *Journal of Consumer Research,* 42 (1), 5–18.

Wang, Xin (Shane), Feng Mai and Roger H.L. Chiang (2014), "Database Submission—Market Dynamics and User-Generated Content about Tablet Computers," *Marketing Science,* 33 (3), 449–58.

Wedel, Michel and P.K. Kannan (2016), "Marketing Analytics for Data-Rich Environments," *Journal of Marketing,* 80 (6), 97–121.

Whitler, Kimberly A. and Neil Morgan (2017), "Why CMOs Never Last," *Harvard Business Review,* 95 (4), 47.

Williams, Jerome D., David Lopez, Patrick Shafto and KyungwonKyung Lee (2019), "Technological Workforce and Its Impact on Algorithmic Justice in Politics," *Customer Needs and Solutions,* 6 (3–4), 84–91.

Wladawsky-Berger, Irving (2018), "The Impact of Artificial Intelligence on the World Economy," *Wall Street Journal,*.

Xiao, Li and Min Ding (2014), "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science,* 33 (3), 338–52.

Zhang, Shunyuan, Dokyun Lee, Param Singh and Kannan Srinivasan (2019), "How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics," *Working Paper,*.

Zhang, Yuchi and David Godes (2018), "Learning from Online Social Ties," *Marketing Science,* 37 (3), 425–44.